Investigating the potential role for RBMY in cancer.

A thesis submitted to the University of Kent for the degree of

M.Sc. in Genetics

School of Biosciences

2019

Tanaka Gahadzikwa

Supervisor: Tim Fenton and Peter J. I. Ellis

I. Declaration Page

No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent, or any other University or Institution of learning

Tanaka Gahadzikwa
2020

II. Acknowledgements.

I would like to thank my supervisors Dr. Tim Fenton and Dr. Peter Ellis who made this project possible. I would like to note that guidance and never-ending patience for my endless questions has enhanced my abilities and research skills. I would also like to thank Dr. Tim Fenton's PhD students Nikki Smith and Mr Max and Dr. Nerrisa Kirkwood with whom I shared a lab and were always there to give me guidance in my experiments or in general research.

I would also like to thank my mother, my partner and all my friends, whom without their support, I would have not been able to achieve this thesis.

III. Contents.

IV. Figures and Tables list.

Tables

Figures

V. Abbreviations.

AKT	Protein kinase B
APC	Adenomatous polyposis coli
AR	Androgen Receptor
AZFb	Azoospermia factor b
BPH	benign prostatic hypertrophy
CDKN1A	Cyclin Dependent Kinase Inhibitor 1A
CSCs	Cancer stem cells
Cys-Cys	Cysteine dipeptides
DAPI	4',6-diamidino-2-phenylindole hydrochloride
DEN	Diethylnitrosamine
DMEM	Dulbecco's Modified Eagle Medium
EDTA	ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ERK	extracellular signal-regulated kinases
EpCAM	Epithelial cell adhesion molecule
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green fluorescent protein
GSK3β	Glycogen synthase kinase 3 beta
HBV	Hepatitis B
HCV	Hepatitis C
HCC	Hepatocellular carcinoma
HER2	human epidermal growth factor receptor 2
His-His	Histidine dipeptides
HNSCC	Head and Neck Squamous Cell Carcinoma
HPV	Human Papillomavirus
HPV+	Human Papillomavirus positive
HPVneg	Human Papillomavirus negative
hnRNP G	Heterogeneous Nuclear Ribonucleoprotein G
IGFBP-3	Insulin-like growth factor binding protein-3 (IGFBP-3)

IGF-1	Insulin-like growth factor 1
IGF-3	Insulin-like growth factor 3
MEK	Mitogen-activated protein kinase
MI SAC	Meiosis I spindle assembly checkpoint
MSCI	meiotic sex chromosome inactivation
MSY	Male Specific Y (MSY) chromosome (chromosome region)
MTT	3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium-bromide
OPSCC	Oropharyngeal squamous cell carcinoma
ORF	Open Reading Frame
OSCC	Oral squamous cell carcinoma
PAR	Population Attributable Risks
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PDGFB	Platelet Derived Growth Factor Subunit B
PVDF	Polyvinylidene fluoride
qPCR	quantitative Polymerase Chain reaction
RBMY	RNA-binding motif (RRM) gene on Y chromosome
RBMX	RNA binding motif protein X-Linked
RBMXL2	RNA-binding motif protein, X-linked-like-2
RRM	RNA Recognition Motif
SAC	spindle assembly checkpoint
SALL4	Spalt Like Transcription Factor 4
SAM68	Src associated substrate during mitosis of 68kDa
STAR	Steroidogenic acute regulatory protein
TBS	Tris-buffered saline
Tbst	Tris-buffered saline TWEEN

TCGA	The Cancer Genome Atlas
TGF-β	Transforming growth factor beta
Tra-2β	Transformer-2 protein homolog beta
TRA1	Transcription-associated protein 1
TRRAP	Transformation/transcription domain-associated protein
TSPY	Testis-specific protein on Y chromosome
UTR	Untranslated region
UTY	Ubiquitously Transcribed Tetratricopeptide Repeat Containing, Y-Linked
WT1	Wilms' tumour protein
ZFA	Zinc finger protein, autosomal
ZFX	Zinc finger X-linked protein
ZFY	Zinc finger Y-chromosomal protein

VI. Abstract

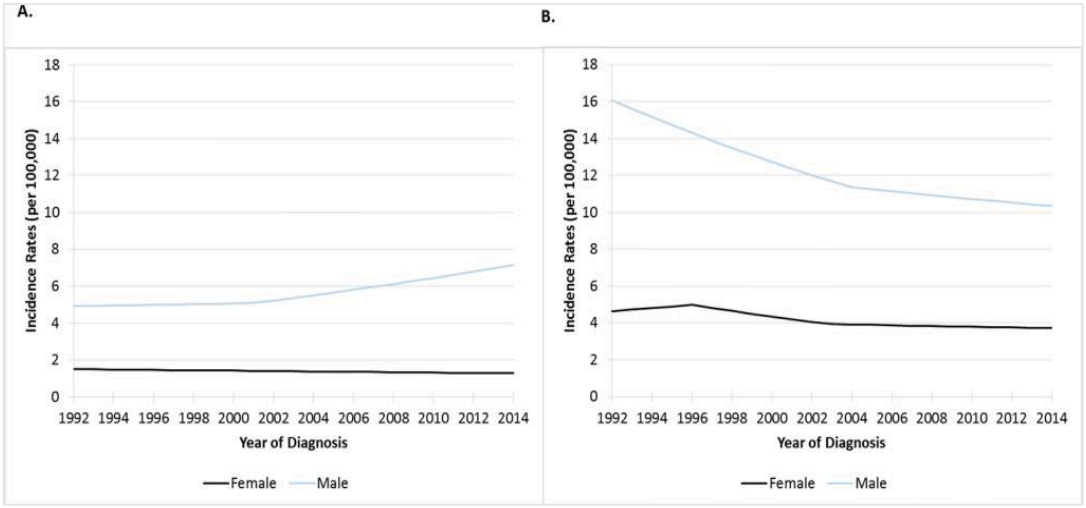
Introduction: Head and neck squamous cell carcinoma (HNSCC), is a major healthcare concern with a high male prevalence. We hypothesise that the testis specific mRNA splicing regulator, Y-linked *RBMY* gene, is aberrantly expressed in HNSCC in part promoting HNSCC through ZFY-short splicing. RBMY has been shown to enhance tumour development in male hepatocellular carcinoma (human tissue specimens and transgenic-mouse models) whilst ZFY-short is predicted to have anti-apoptotic properties and the deletion of *RBMY* locus on Y-chromosome resulted in lowered ZFY-short expression. Thus, we hypothesize that ZFY-short is generated by RBMY and exerts its anti-apoptotic effects to promote male HNSCC. **Methods**: Due to the coronavirus lockdown, bench work was restricted to 6 months, therefore, I conducted an extended analysis of *RBMY* expression in human cancer, including a computational analysis of *RBMY* gene expression with data from the cBioPortal database. In my bench-work, I attempted to establish GFP-RBMY expressing cell lines and conducted fluorescence microscopy, RT-PCR and qPCR to analyse *RBMY* expression in HNSCC cell lines and its impact on *ZFY-short* expression. **Results:** *RBMY* is expressed in several cancers, with no driver mutations. RBMY has nuclear localisation and is expressed in 93-UV-147T and UM-SCC-104 cell lines (both HPV16-positive HNSCC cell lines), with increased *ZFY-short* expression observed in UM-SCC-104. **Discussion:** Despite RBMY having been shown to be an oncogene in male liver cancer, our analysis of cBioPortal data suggests this activity may be restricted to the small minority of tumours of different cancer types that express RBMY. The paralleled expression of *RBMY* and *ZFY-short* in our cell lines indicate an association. UMSCC104 cell line originates from a highly an aggressive and recurrent tumour, RBMY is associated with tumour stemness, thus it is possible that via *ZFY-short*, RBMY could have promoted the aggressive phenotype in this, and in other HNSCCs.

1 INTRODUCTION

1.1 HNSCC Prevalence in Males.

Head and Neck Squamous Cell Carcinoma (HNSCC), the sixth most prevalent cancer globally, is now a major healthcare concern, causing yearly deaths of around 300 – 350 000 (since 2008) and about 650 – 700 000 new incidences per year (Cognetti, Weber and Lai, 2008) (Wondergem *et al.*, 2020). HNSCC covers anatomical regions of the upper respiratory tract, affecting the oral cavity, oropharynx, hypopharynx, and larynx. 80% of HNSCC incidences are as a result of Oral squamous cell carcinoma (OSCC) and Oropharyngeal squamous cell carcinoma (OPSCC), which both show a male prevalence, with a reported male to female cancer incidences ratio of 1.5:1 and 2.8:1 (Vigneswaran and Williams, 2014). This is in line with a lot of literature in which HNSCC is shown to have a high male prevalence, a point of focus for our study because we hypothesise that alongside other well-established high-risk factors that have been shown to drive the HNSCC male prevalence, the genetic distinction between males and females may also in part be the reason for this prevalence. Changing societal behaviours in association with high risk HNSCC carcinogens such as smoking, alcohol abuse and populations infection rate with oncogenic viruses like Human papilloma virus (HPV) induces a geographical and anatomical variation on the HNSCC cancer incidence (Vigneswaran and Williams, 2014). In many developing regions of the world, particularly in the Indian subcontinent, OSCC has emerged as the third and first most prevalent cancer in females and males respectively (Warnakulasuriya, 2009), whilst in developed EU and the USA, the decreased tobacco usage has been shown to be associated with the decline in OSCC incidence (Sturgis and Cinciripini, 2007). Conversely, OPSCC, as a result of changing sexual practices and increased population infections with high risk HPV infection, has shown an increasing incidence rate in the developed parts of the world (Vigneswaran and Williams, 2014). HPV infections in the developing parts of the world is a major issue because male OPSCC incidence associated with HPV infection has surpassed cervical cancer incidence in females, a trend which

is projected to become irreversible in 40 years' time (Chaturvedi *et al.*, 2011; Gillison *et al.*, 2015; Mourad *et al.*, 2017). Data from an American study showed that between 2002 to 2012, HNSCC incidence rate showed a stable 2.8% rise in males, whilst the female incidence rates during this previous showed a plateau (Fig.1A and 1B)(Fakhry *et al.*, 2018).



(Fakhry *et al.*, 2018)

Figure 1: 1992 - 2014 HNSCC incidence rate trends per 100 000 people. OPSCC (figure 1A) and Non-Oropharyngeal HNSCC (figure 1B) incidence trends. The male cancer incidence is indicated by the blue line, whilst the female cancer incidence is indicated by the black line. OPSCC in males appears to be increasing from 2001. In the same period, female OPSCC incidence was on the decline. In non-OP HNSCC, male incidence significantly declined from 1992-2014, with the greater extent of decline between 1992-2004 compared to the lesser decline rate between 2004-2014. The female non-OP HNSCC incidence rate showed a stable between 1992-1996 before a significant decline between 1996-2003 and eventually stabilising from 2003-2014 (Fakhry *et al.*, 2018).

HNSCC can be categorically divided as either HPV positive (HPV+) and HPV negative (HPVneg) HNSCC and in either case, the HNSCC risk factors affect males more than females. HPV (high risk viral strains) is an extremely powerful driver for initiation and progression of HNSCC, particularly OPSCC (St Guily *et al.*, 2011). HPV attributes to about 25% of all HNSCC (St Guily *et al.*, 2011), and as mentioned before, vast majority of HPV+ HNSCC incidences occur in developed countries, in which high risk HPV infection is detectable in 45-90% of

OPSCC incidences (D'Souza and Dempsey, 2011). High risk HPV-16, which accounts for over 80% of OPSCC associated with HPV infection was shown to have a significantly higher oral presence in males than in female, with a male to female HPV-16 oral presence ratio of 7:1[8].

Whilst the anti-smoking public health campaign has drastically seen a significant reduction in smokers across the world, there still remains a relatively large proportion of smokers in males, with global estimates of about 35% and 6% male and female smokers (Ritchie Hannah, 2019), and as a result, women are less likely to manifest certain risk factors from alcohol abuse (Ceylan-Isik, McBride and Ren, 2010). The combined use of alcohol and smoking, account for a large proportion of OPSCC incidences, and an even greater proportion of laryngeal cancer incidence (Hashibe *et al.*, 2009). As the use of either alcohol or tobacco varies between the 2 sexes, the consequent effect of the use of these drugs is also different between the sexes. In males, the combinational use of smoking and alcohol abuse poses the greatest risk to HNSCC oncogenesis rather than the individual use of either one of the drugs (Hashibe *et al.*, 2009) (Leemans, Snijders and Brakenhoff, 2018) (Blot *et al.*, 1988). However, in females, tobacco is shown to solely account for majority of HNSCC incidences rather than the combinational use of tobacco and alcohol or alcohol abuse alone (Hashibe *et al.*, 2009). Analysis of population attributable risks (PAR) (the measure of disease incidence proportion within a population that is attributable to a certain exposure) of tobacco and alcohol on HNSCC further illustrated the impact of the combinational use of tobacco and alcohol (Hashibe *et al.*, 2009), with a 72% PAR for the combined use of alcohol and tobacco smoke on HNSCC, 33% accounted for by the use of tobacco alone and 35% due to the joint use of both tobacco and alcohol (Hashibe *et al.*, 2009). Furthermore, the PAR for the combinational use of alcohol and tobacco for HNSCC in men is 74% whilst being 57% for females (Hashibe *et al.*, 2009). However, there is a substantial percentage of at least 28% of HNSCC incidences (42% and 26% for females and males respectively) that are largely unaccounted for by the abuse of alcohol or smoking (Hashibe *et al.*, 2009). This is particularly for OPSCC in which around 36% of OPSCC cancer incidences remain to be accounted for, as well as in females and young person's (Hashibe *et al.*, 2009).

When we consider the changing epidemiological outlook of the HNSCC, exposure to high-risk HNSCC carcinogenic factors such as smoking, alcohol and HPV infection offer insufficient explanations to why the HNSCC incidence exhibit such differences between the sexes. Thus, alongside the above-mentioned risk factors that contribute to the high male prevalence in HNSCC, we hypothesise that the underlying genetic differences between male and female could in part also explain and contribute to this gender disparity between male and female HNSCC incidence. This is an emerging concept in several diseases, as new reports regarding male specific genes are beginning to appear and offer explanations of the sexual dimorphism in diseases (Kido and Lau, 2015). Likewise, most cancers are shown to have an unusual male prevalence in both cancer incidence and mortality (American population data between 1975-2004) (Cook *et al.*, 2009). Given this data, we hypothesise that the most distinct thing between males and females, the Y chromosome, may play in part a significant role in HNSCC high prevalence and morbidity in males alongside the other well-established risk factors driving HNSCC male prevalence.

Males and females are genetically distinct as a result of their sex chromosome composition defined by males having the XY chromosomes whilst females have XX chromosomes (Kido and Lau, 2015). Despite evolving from the same ancestral autosome pair, the Y chromosome is smaller and contains significantly smaller set of functional genes than the X chromosome (Kido and Lau, 2015). This is as a result of millions of years of gene decay processes on the Y chromosome since its evolution from the ancestral autosome (Bachtrog, 2013). Despite this, Y chromosome genes are essential for the development of males and beyond this, there is increasing evidence that the Y chromosome is critical for biological functions that far supersedes its role in male sex developments, and its aberrant function is associated with various diseases including cardiovascular diseases (Maan *et al.*, 2017) (Hollows *et al.*, 2019).

The Y chromosome harbours proto-oncogenic features as recent studies have linked the loss of or aberrant expression of Y linked genes to several somatic tissue cancers, capable of having oncogenic roles as well as tumour suppressive

roles. Y chromosome genes *TSPY* and *RBMY* have been shown to be expressed in somatic cell cancers, with some recent papers describing them as male specific novel oncogenes (Kido and Lau, 2015), despite the fact that they are normally restricted to testicular germ cells only expression (Kido and Lau, 2015). Conversely, the Y chromosome gene UTY, is reported to be a tumour suppressor (suppressing leukemogenesis) alongside its X/Y homologous copy (Gozdecka *et al.*, 2018). In line with this, certain key protein players in meiosis have also garnered attention as the process of meiosis has been likened to resemble processes that occur in genomic instability (Nielsen and Gjerstorff, 2016). Genomic instability is a classic hallmark of cancer in which various genetic modifications drive the initiation and development of cancer. In meiosis, there is an exchange of genetic materials between homologous chromosomes, and similarly, genetic alterations are seen in cancers that have gone through genomic instability (Nielsen and Gjerstorff, 2016). It is possible that perhaps some of the proteins normally involved in meiosis are activated in oncogenesis and drive the genomic instability in somatic cell cancers (Nielsen and Gjerstorff, 2016).

Moreover, loss of Y chromosome in peripheral blood cells is reported to be associated with high risk of cancer and high cancer mortality in men, especially in non-haematological cancers (Kido and Lau, 2015) (Forsberg *et al.*, 2014). However, the molecular mechanisms by which ectopic expression in germline cells and association of loss of Y chromosome with cancer incidence and mortality are still yet to be elucidated. Nonetheless, these reports and observations certainly suggest that the roles of Y chromosome genes may extend further to influence disease state in males, particularly in some male somatic cancers like HNSCC (Kido and Lau, 2015).

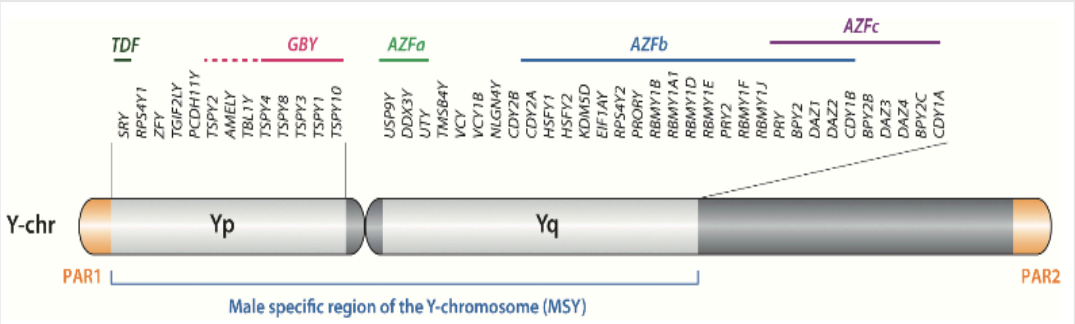
1.2 Project Aims.

As mentioned above, we hypothesise that HNSCC high male prevalence is in part, the result of RBMY ectopic expression in these tumours, triggering male specific oncogenic mechanisms that drives HNSCC in the male population alongside other risk factors. Thus, in this study we aim to show that *RBMY* is found ectopically expressed in male HNSCC cell lines UM-SCC-104 and 93-UV-147T, and that overexpression of *RBMY* induces *ZFY* splicing to produce *ZFY-short*.

As a result of the Covid outbreak, the subsequent lockdown and closure of business', our aims have somewhat changed. Prior to the lockdown, we had done about 6 month of wet-lab work and had many experiments lined up in the future. However, the closure of our labs as a result of the lockdown unfortunately became the last time that MSc students in our department were able to carry out any wet-lab experiments. Consequently, in addition to the small set of data gathered from wet-lab work, we decided to add a systematic literature review of what is known about *RBMY* and *ZFY*. Moreover, with the help of some computational analysis using some public domain resources we were also to analyse *RBMY* expression in all cancers. Thus, this thesis includes a substantial literature review (thesis) and computational analysis on the subject of the thesis.

Our overall aims are (1) to systematically review the known biology of both *RBMY* and *ZFY* genes, particularly with an eye to potential roles outside the testis, (2) look at in silico expression data to see if there is evidence for RBMY acting as an oncogene in cancers other than hepatocellular carcinoma and (3) at the bench, to move towards a model to test the hypothesis that RBMY regulates *ZFY* splicing patterns.

1.3 The Y chromosome.



(Kido and Lau, 2015)

Figure 2: Illustration of a single Arm of the Y chromosome. The diagram shows the structural segmentation of the Y chromosome and the protein coding genes carried on the Y chromosome.

The Y chromosome can be structurally divided into 3 segments. There is the male specific region of the Y chromosome (MSY) segment, which carries 23 protein coding and pseudo genes (Kido and Lau, 2015) (Jangravi *et al.*, 2013). MSY proteins are thought to be either gene expression regulators or protein stabilisers, these proteins have different expression patterns, as some of them are expressed ubiquitously, whilst the rest have predominant expression in the testis (Kido and Lau, 2015). The somatic expression of some MSY genes is thought to occur as a result of a balancing act of these MSY genes and their X-chromosome homologs, in order to maintain healthy functions in males (Bellott *et al.*, 2014). There are also the pseudo-autosomal regions (PAR1 and PAR2) (Skaletsky *et al.*, 2003), these regions are unique regions of the Y chromosome as the genes within this region are the only genes of the Y chromosome that undergo meiotic recombination in the same manner as autosomal genes (Kido and Lau, 2015). PARs contain a total of 20 protein coding genes, which also occur on the X-chromosome, these genes are reported to play key roles in male germ cell fertility and meiosis (Kido and Lau, 2015). The third structural segment is the heterochromatin region, which is a genetically inert region of the Y chromosome that contains repeating genetic sequences of the DYZ families. Consequently, the heterochromatin regions in different male groups is polymorphic in length (Quintana-Murci and Fellous, 2001).

1. RBMY

1.4.1 RBMY

RBMY is a highly conserved Y-chromosome gene that is found in all mammals, it is critical for normal male development as its microdeletion is associated with male infertility (Tsuei *et al.*, 2004) (Venables and Eperon, 1999). There are about 30 copies of *RBMY* genes and pseudogenes, however, only 6 are functional. The active copies of *RBMY* are all reported to be located on the AZFb locus on the Y chromosome (Chai, Salido and Yen, 1997) (Skaletsky *et al.*, 2003). Human *RBMY* (h*RBMY*) normal expression is restricted to nuclear germ cells, with *RBMY* active expression occurring during germ cell development from the spermatogonia to round spermatids. RBMY is essential for spermatogenesis as downregulated or abolished h*RBMY* expression or AZFb deletion was reported to induce meiotic spermatogenesis arrest of germ line cells (Elliott *et al.*, 1997) (Vogt *et al.*, 1996) (Elliott *et al.*, 1998). Mouse evidence shows that RBMY is a testis-specific splicing factor in spermatogenesis (Dreumont *et al.*, 2010) (Liu *et al.*, 2009), and experiments with transgenic mice revealed that *RBMY* knockout is associated with major spermatozoan structural abnormalities (Mahadevaiah *et al.*, 1998). The most distinctive feature of RBMY protein is the presence of an ancient N-terminal RNA Recognition Motif (RRM), that is characteristic of many RNA binding proteins. RBMY protein also contains a C-terminal auxiliary domain, which is interesting in that this domain contains a 37 amino acid repetitive sequence segments that are enriched in serine, arginine, glycine, and tyrosine residues, named the SRGY boxes (Chai *et al.*, 1998) (David J. Elliott, 2004). SRGY box amino-acid sequences corresponds to protein interaction domain (David J. Elliott, 2004). *RBMY* sequence, protein structure and cellular distribution of RBMY protein are consistent with an RNA splicing and processing role during spermatogenesis (David J. Elliott, 2004). Furthermore, RBMY could potentially be involved in the governance of the global splicing machinery that induces splicing changes seen in pre- and post-meiotic stages that enables

proper germ cell development (Schmid *et al.*, 2013). This assumption is based on the fact that the X homologues of *RBMY* genes have been described to participate in this process that promotes the meiotic global splicing changes (Schmid *et al.*, 2013) (Ehrmann *et al.*, 2008). Transcriptome-wide analysis using RNAseq of mouse testis tissues to identify expression changes from mitotic spermatogonia to meiotic spermatocytes showed that there is a change in expression from the X-linked Rbmx pre-meiosis, an Rbmy homologue to the retrogene RBMXL2 during and posts (Schmid *et al.*, 2013) (Ehrmann *et al.*, 2008) (David J. Elliott *et al.*, 2000). RBMXL2 is a key gene for meiosis as deletion on this gene blocks meiosis, completely inhibiting sperm production (Ehrmann *et al.*, 2019). This, alongside the already existing evidence of Rbmy knockout in mice inducing major sperm defects and infertility (in humans) shows that RBMY may have diverse roles in conducting and regulating the meiotic development (Mahadevaiah *et al.*, 1998) (Abid *et al.*, 2013).

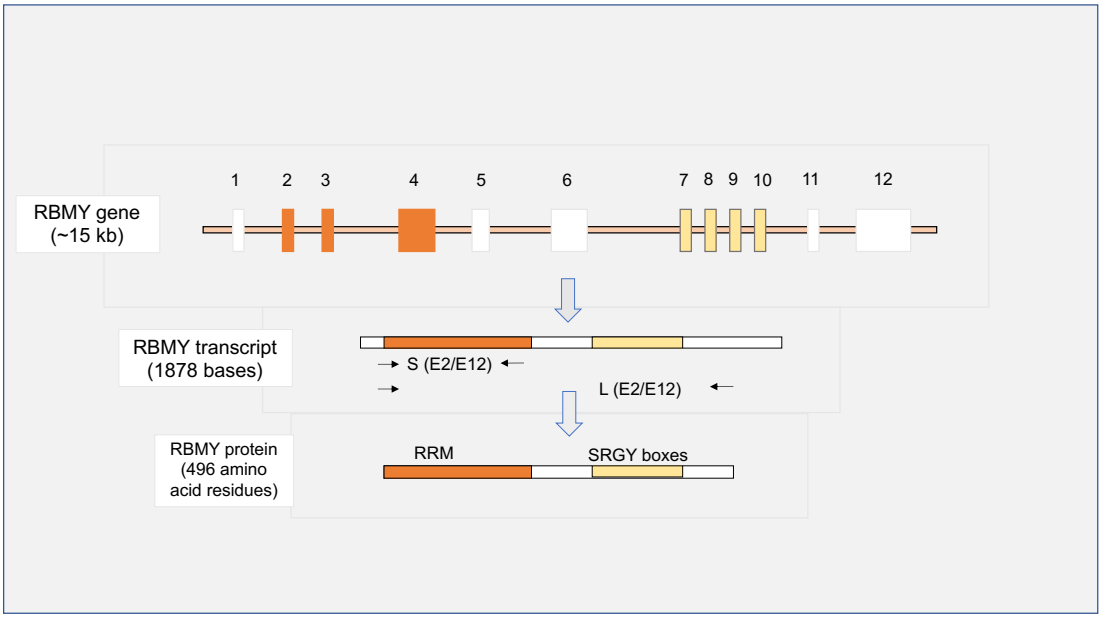


Figure 3: RBMY gene, mRNA and protein representation.

1.4.2 RBMY, A male specific splicing regulator.

Testes have a highly specialized gene expression program that is necessary to carry out the complex differentiation of germ cells from spermatogonia into mature spermatozoa (Skrisovska *et al.*, 2007). Alternative splicing, the most effective gene expression modelling mechanism is highly active in testis, although how these events are regulated at molecular level is still very much being investigated (Skrisovska *et al.*, 2007). HRBMY expression is nuclear restricted in adult male germ cells throughout all transcriptionally active stages of spermatogenesis and the expression block or deletion of functional RBMY induces spermatogenic block and cell arrest during the first meiotic division (Skrisovska *et al.*, 2007) (Elliott *et al.*, 1997). RBMY role as a splicing regulator is perhaps further supported by the protein structure on RBMY, which has the N-terminal RRM and a carboxy- terminal domain with SRGY boxes, features that associated with other RNA interacting proteins and splicing regulators (Skrisovska *et al.*, 2007) (Ma *et al.*, 1993).

Detailed protein interaction studies with RBMY have revealed several proteins that interact with RBMY (David J Elliott, 2004), with several RNA binding proteins having been identified through Yeast Two Hybrid screens to interact with RBMY. RBMY protein interaction with other proteins is revealed to not limited to the SRGY boxes which are sufficient to promote interaction but may also occur through RS dipeptides within the linker region (Dreumont *et al.*, 2010).

Despite the function of RBMY still being unclear, RBMY interaction with pre-mRNA splicing proteins suggest that it may act in conjunction with these proteins and regulate germ cell mRNA splicing, mediated by the SRGY boxes (Dreumont *et al.*, 2010) (David J. Elliott, 2004). Several pre-mRNA splicing regulators including RNA metabolism regulating factors, SR proteins and STAR proteins like SAM68 and T-STAR were identified to interact strongly with RBMY in Yeast two hybrid screens using RBMY (Venables and Eperon, 1999) (Venables, 2000) (Elliott, 2000). Furthermore, these yeast hybrid screens revealed that RBMY

interacts several RNA binding proteins of the SR family of proteins and tra-2*β* (David J Elliott, 2004). These interactions are quite interesting as RBMY is a male germline specific proteins whilst tra-2*β* and SR proteins show a ubiquitous expression pattern, suggesting that perhaps RBMY may be a cofactor for these splicing regulators that provides male germline spicing tissue specificity necessary for male germline development (D J Elliott *et al.*, 2000) (Venables, 2002). SR protein family is composed of many splicing regulators, which are contain the characteristic large domain of serine and arginine residues domain (RS domain) (David J Elliott, 2004). Tra-2*β* is a RS-rich protein essential for splicing activation, it binds to target protein in a sequence specific mechanism in which tra-2*β* binding domain locates and binds to the exon GAA sequence repeats (David J Elliott, 2004) (Tacke *et al.*, 1998). RBMY may be regulating gene expression at post transcriptional level by directly or indirectly interacting with proteins involved in pre-mRNA splicing, mRNA stability, and the transport of RNA between the nucleus and cytoplasm (Tsuei *et al.*, 2004) (Nagai *et al.*, 1995). Overall, based on yeast two-hybrid experiments, it is shown that RBMY potentially interacts with 3 different types of proteins belonging to the groups of SR family proteins, hnRNP G related proteins and STAR proteins, although the meaning of these interactions or how they occur remains to be solved (Tsuei *et al.*, 2004) (Venables and Eperon, 1999) (Venables, 2000).

1.4.3 RBMY structure and localisation.

Human *RBMY* gene carries localisation signalling codes in both the C- and N-regions, experimentation with full length hRBMY experimentation revealed that the C-terminal localisation signal is more dominant over the N-terminal localisation signal (Dreumont *et al.*, 2010). The specificity of hRBMY activity is the result of the N-terminal localisation, this is because the N-terminal localisation signal directs the protein into nuclear region that it vital for gene transcription (Dreumont *et al.*, 2010). The N-terminal signal localises hRBMY into the peri-speckle region of the nucleus, this is the region of active gene transcription and processing including splicing (Dreumont *et al.*, 2010). RBMY nuclear distribution in the peri-speckles has been revealed to be similar to that of two other proteins

that are part of the exon junction complex, protein complex involved in mRNA splicing (Dreumont *et al.*, 2010) (Degot *et al.*, 2004). These are the Magh and Y14 proteins, which are both shown to interact with hRBMY via co-immunoprecipitation studies (Degot *et al.*, 2004). RBMY is thus nuclear specific as a result of its role in potentially regulating transcription, splicing of pre-mRNA and its subsequent transport (Dreumont *et al.*, 2010). In the peri-speckle region RBMY is thought to physically interact with SR proteins and Exon Junction complex components in close proximity to regulate mRNA processing. Conversely, some studies suggest that it is rather unlikely that the RRM segment of the RBMY protein can conduct protein interactions (Venables, 2000) (D J Elliott *et al.*, 2000) (Dreumont *et al.*, 2010). These observations support the evidence that RBMY is an alternative splicing regulator or co-regulator (Dreumont *et al.*, 2010).

1.4.4 Oncogenic RBMY, hypotheses

As a result of the sexual dimorphism in somatic tissue cancer incidences and mortality rates, several ideas including sex hormones and sex chromosomes have been proposed to explain this disparity in disease development between the sexes (Kido *et al.*, 2020) (Clocchiatti *et al.*, 2016). The primary liver cancer hepatocellular carcinoma (HCC) which accounts for 80%-90% of all liver cancer cases (Yang *et al.*, 2019) (Kido *et al.*, 2020), significantly affects males more than females, affecting males at 5.4 times higher than in females (Kido *et al.*, 2020) (Wong *et al.*, 2017)**.** The Y linked gene *RBMY* has been reported to be aberrantly activated and highly expressed in male HCC tissues (Kido *et al.*, 2020). Thus, RBMY has been hypothesised to potentially induce oncogenic activity that promote HCC cellular transformation or tumour development.

Since the initial reporting in 2004, several other studies have shown RBMY expression in liver cancers. Despite normal expression of RBMY being largely restricted to male germ line cells, *RBMY* transcripts are reported to be detected in liver HCC and hepatoblastoma (Tsuei *et al.*, 2004), with *RBMY* expression in these liver cancers showing high specificity and expression as *RBMY* expression

was undetectable in other cancers tissues including cholangiocarcinoma, carcinomas of the colon, kidney, lung, stomach, and prostate (Tsuei *et al.*, 2004). Since the initial reporting of RBMY expression in liver cancers, RBMY has been shown to be aberrantly activated in other somatic cancers besides HCC, including lung adenocarcinoma and kidney renal papillary cell carcinoma (Kido *et al.*, 2020) (Kido and Lau, 2016). Although the role of RBMY in tumour development is still being debated for several reasons including the fact that RBMY is transcriptionally inactive in any testicular cancers (Cheung *et al.*, 2016), multiple studies have shown RBMY to be expressed in somatic cell cancers, with several hypothesis proposed to explain to potential oncogenic role of RBMY in human male somatic tissue cancers.

A follow up study in 2011 by Daw-Jen Tsuei et al to the original 2004 study that described *RBMY* as an oncogene investigated RBMY oncogenic role in the human liver cancer cell line HepG2 and the in vitro effect of RBMY in human livers and transgenic mice (Tsuei *et al.*, 2011). Results from this study prompted a hypothesis from the researchers, which suggests that RBMY oncogenic activity in male HCC may be acting via the androgen receptor (AR) pathway (Tsuei *et al.*, 2011). AR has been shown to mediate aberrant androgen activity that contributes to the initiation and development of HCC in both male mice and humans (HBV and HCV carriers) (Tsuei *et al.*, 2011)[53]. AR is a homo-dimeric nuclear receptor protein with several modulating co-activating and co-repressing androgen proteins. It is an upstream gene expression regulator of several genes including the essential Insulin-like growth factor genes like insulin-like growth factor 1 (IGF-1) and insulin-like growth factor binding protein 3 (IGFBP-3)[54][55].

A 2015 follow up study commented on the Daw-Jen Tsuei et al 2011 study for its inadequate explanation for male HCC prevalence resulting from RBMY oncogenic activity via the AR signalling pathway. Their issue with this hypothesis was that it does not explain the HCC high male prevalence in young people (Chua *et al.*, 2015). HCC incidence rate for individuals under 19 shows a high male prevalence, with as high male to female HCC incidence ratio of 3.74 reported in 6-9 year old and 2.19 ratio in 15 to 19 year olds (Chua *et al.*, 2015) (Chang *et al.*, 2009). The RBMY AR hypothesis cannot account for this because of low serum

androgen activity in young men especially below 9 years (Chua *et al.*, 2015). The 2015 study suggested a new hypothesis which would possibly explain RBMY oncogenic activity in HCC and explain HCC prevalence in young people. Their hypothesis suggests that RBMY may be associated with the dysregulation of signalling pathways controlling the regenerative progenitor cells to give rise to HCC, particularly the Wnt/β-catenin pathway, a pathways that is highly active in young people than in adults (Williams, Xu and Cancelas, 2006). The dysregulation of the self-renewal signalling pathways of stem cells like the Wnt/β-catenin induces initiation of HCC (Chiba *et al.*, 2007). In line with this, it is proposed that RBMY is linked to the Wnt/β-catenin pathway to induce HCC and emergence of cancer stem cells (CSCs) which gives rise to a highly stem male HCC subtype (Chua *et al.*, 2015). CSCs resemble progenitor cells in several ways including similar signalling pathway dysfunction leading to CSCs stemness properties including self-renewability and chemotherapeutic drug resistance (Yamashita and Wang, 2013). This study also showed that significant RBMY expression and RBMY cytoplasmic presence in HCC tissues induced a worse prognostic outcomes for male patients with HCC (Chua *et al.*, 2015).

Tastuo Kido et al, proposed a hypothesis (2020) that suggests RBMY to have dual functional cancer promoting activities via an initial role as a tumour suppressor that eventually switches promote tumour initiation and development (Kido *et al.*, 2020). In the study, *RBMY* is shown to be overexpressed at initial tissue transformation stage, where at this stage RBMY expression is reported to primarily act as a male specific tumour suppressor that impedes cellular proliferation and consequently tumour development (Kido *et al.*, 2020). HCC from this study is suggested to arise as a result of resistance mechanisms via tumour adaptation beyond RBMY proliferation block. This nature of RBMY to promote tumour development and also act as a tumour suppressor is not an exclusive ability to RBMY as duality of nature of oncogenes has been reported in various other studies that termed these genes as double-agent genes (Kido *et al.*, 2020). These genes including *P21/CDKN1A, TGF-β* and *WT1* act as both tumour suppressors and oncogenes (Kido *et al.*, 2020) (Shen, Shi and Wang, 2018) (Bashyam *et al.*, 2019).

The above-mentioned hypothesis although supported by data do have their limitations. Never-the-less, RBMY does have tumour promoting roles and has been shown to be able to drive cell transformation and in-vivo. We are going to analyse and criticise the available data that supports the proposed hypotheses: the available data from RBMY studies will be discussed as (i) RBMY expression in clinical samples (ii) cell line studies of RBMY function (iii) transgenic animal studies of RBMY function.

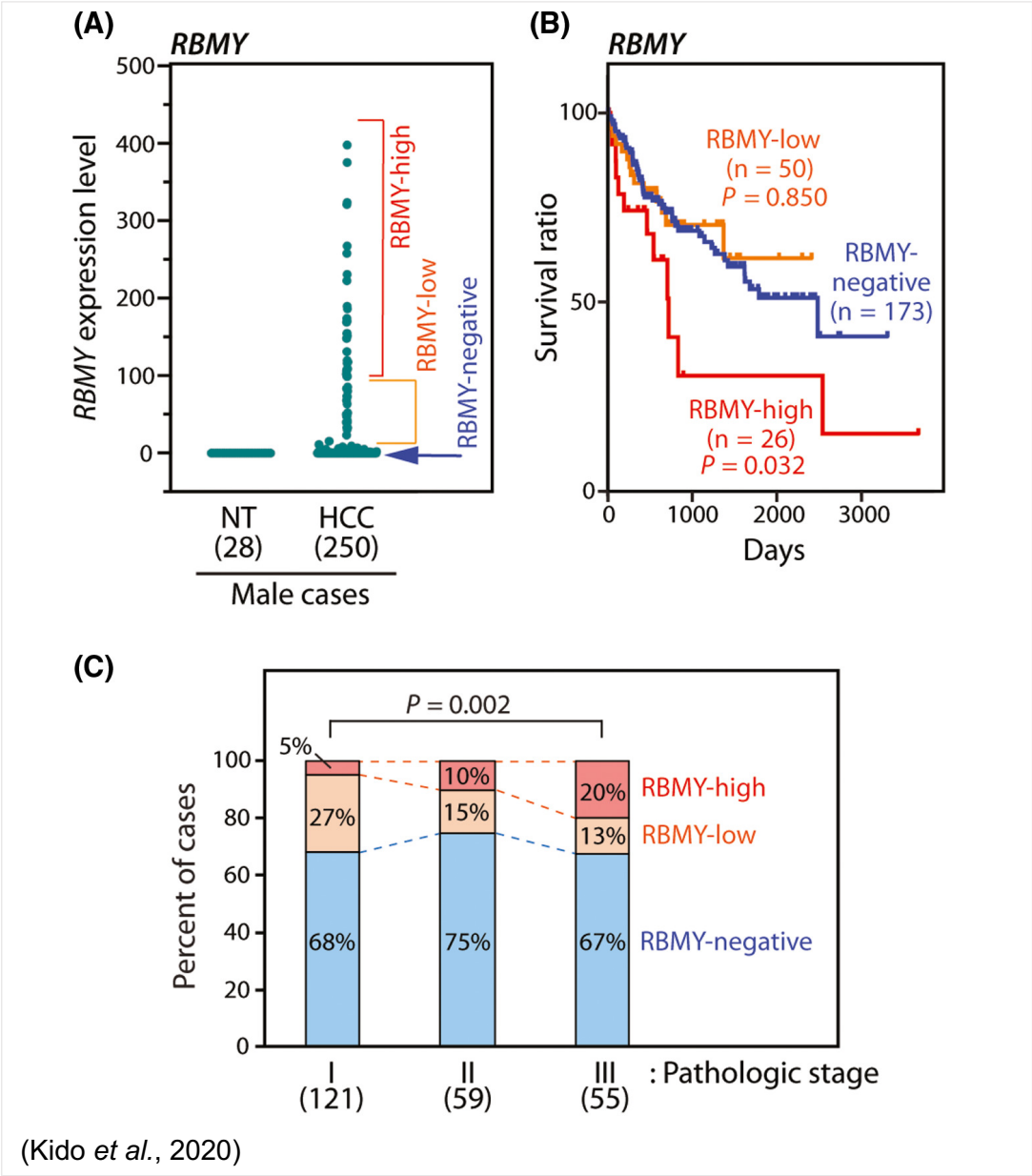
1.4.5 RBMY Expression in HCC Clinical Samples

The Daw-Jen Tsuei et al 2004 initial reporting of *RBMY* expression in human somatic tissue cancer, showed that there was a high expression of *RBMY* transcripts in male HCC tissues (Tsuei *et al.*, 2004). Despite normal expression of *RBMY* being largely restricted to male germ line cells, *RBMY* transcripts were reported to be detected in HCC and hepatoblastoma tissues (Tsuei *et al.*, 2004). Furthermore, *RBMY* expression in these cancer tissues was found to be exclusive to male cancer patients only, with no *RBMY* detection in female cancer tissues or non-cancer male/female tissues (Tsuei *et al.*, 2004). In data collected from an HCC sample from a patient group 90 total males of 13 young and 77 adult males, *RBMY* was shown to be expressed in 36% of those samples. Furthermore, *RBMY* expression was detected in 67% of a sample group of 6 young males with hepatoblastoma (Tsuei *et al.*, 2004). As *RBMY* is a Y lined gene, no *RBMY* expression was detected in any HCC and hepatoblastoma young and adult females tissue (Tsuei *et al.*, 2004). The absence of detectable *RBMY* transcripts in the females confirms the specificity of the detection method (and confirms none of the phenotypic females were chromosomally XY). Nevertheless, the high expression rate and specificity suggests strong association of RBMY with hepatocarcinogenesis.

Further investigation into RBMY expression in HCC tissues revealed that RBMY expression in HCC tissues is correlated with cellular proliferation in HCC (Kido *et al.*, 2020). Immunostaining of HCC specimens revealed that RBMY had differential expression pattern in HCC tissues, which expression ranging between

positive and dense RBMY expression, positive and sparse RBMY expression and negative RBMY expression (Kido *et al.*, 2020). RBMY expression restriction to tumour areas showed high specificity of RBMY expression within these tumours (Kido *et al.*, 2020), thus demonstrating that RBMY expression in non-germ line cells is ectopic and restricted to cells in disease state. When the expression pattern of RBMY was referenced with the expression of other tumour markers including the proliferation marker Ki-67, the reference showed that positive RBMY expression whether dense and sparsely expressed in HCC specimens overlapped significantly with Ki-67 immunohistochemical expression patterns, whilst HCC specimens that were RBMY negative did not (Kido *et al.*, 2020). These results suggests an association between high RBMY expression and the high proliferative activity in male HCC patients (Kido *et al.*, 2020) (Kido *et al.*, 2014), thus linking RBMY to male specific HCC stemness with poor prognosis.

The correlation between RBMY expression in HCC specimens and the high proliferative activity in male HCC patients led to the assumption that high *RBMY* expressing HCC male patient samples from The Cancer Genome Atlas might be associated with a poorer prognosis (Kido *et al.*, 2020). *RBMY* expression in HCC specimens and the clinical significance of this expression was investigated using patient survival information from The Cancer Genome Atlas datasets (Kido *et al.*, 2020). The data included data from tissue specimens of patients of both sexes and various age groups with HCC and patients with non-tumour specimens (Kido *et al.*, 2020) (Hutter and Zenklusen, 2018). With *RBMY* expression graded from negative, low to high and the expression grading compared against patient survival (Fig. 4A), high *RBMY* expression in male HCC samples was found to be associated with poor prognostic outcomes (Kido *et al.*, 2020). The data results showed that the overall survival of male patients with high *RBMY* expression in their tumours was significantly lower compared to male patients with low *RBMY* and negative *RBMY* expression in their HCC tumours, with little difference in survivability between the latter 2 categories (Fig. 4B). Thus, the evidence supports the conclusion in which high RBMY expression in male cancer tissues is associated with a poor prognostic result (Kido *et al.*, 2020)



(Kido *et al.*, 2020)

Figure 4: RBMY transcription analysis of TGCA HCC patient data. HCC specimen Transcriptome data from a cohort of HCC cancer patients of 250 and 121 male and females was acquired. Furthermore, tumour negative patient survival dataset from a cohort of 22 males and 28 females was also acquired (Kido *et al.*, 2020). (A). Relative *RBMY* expression levels in male HCC tissue specimens and non-tumour tissues (NT). The plot demonstrates the levels of RBMY expression, showing High, Low and Negative *RBMY* expression, (B). Keplan Meler survival plot of male HCC patients of varied RBMY tumour expression levels, patients were grouped as either showing high (indicated by the red line), Low (indicated by the orange line) and Negative *RBMY* expression (indicated by the blue line), (C). Bar plots showing the relationship between HCC pathological stage and RBMY expression levels (Kido *et al.*, 2020). The plots show ratios of HCC cancer specimens with *RBMY* high, low or negative expression in different pathological stages of their HCC cancer. *RBMY* high expression shows an increase from 5% in grade I pathological stage to 20% in grade III, whilst *RBMY* low expression detection in the cancer is decreasing as the pathological stages of HCC goes up (Kido *et al.*, 2020).

Most studies looking at RBMY expression in cancer mostly show a dominant nuclear expression of RBMY in liver cancer cell models (Tsuei *et al.*, 2004) (Venables, 2000) (Dreumont *et al.*, 2010) (Heinrich *et al.*, 2009). However, the study by Huey-Huey Chua et al showed RBMY extensive cytoplasmic presence in both human and mice liver cancer cells (Chua *et al.*, 2015). Furthermore, this study showed association of RBMY cytoplasmic presence with a more aggressive form of HCC in males (Chua *et al.*, 2015). RBMY cytoplasmic presence was reported before this papers' reporting, albeit under different circumstances. The 2004 study by Daw-Jen Tsuei et al reported of a cytoplasmic restricted RBMY protein that lacked the C terminus domain (Tsuei *et al.*, 2004). Furthermore, this RBMY protein displayed tumorigenicity in NIH3T3 cells albeit in a dose-dependent manner (Tsuei *et al.*, 2004). Never-the-less, Huey-Huey Chua et al show data of RBMY cytoplasmic presence in cancer cell lines and a novel hypothesis that may explain RBMY role in male liver cancers.

The presence of RBMY in the cytoplasmic is shown to elicit tumour promoting activities that are distinct from RBMY that has nuclear presence only (Chua *et al.*, 2015). In this study by Huey Huey Chua et al, patients with HCC that showed RBMY expression and presence in the cytoplasm were associated with a more aggressive male HCC subtype that was more highly tumorigenic with a worse prognostic outcome compared to patients HCC that showed only nuclear positive RBMY expression (Chua *et al.*, 2015). HCCs associated with exclusive nuclear RBMY expression lack ability to drive tumorigenic activity of CSCs and thus has limited tumour stemness properties like self-renewal capability, chemoresistance and vascular invasive properties (Chua *et al.*, 2015). This illustrates, as according to the data in this paper, that nuclear restricted RBMY does not induce as aggressive oncogenic effects as cytoplasmic RBMY and gives a reason to believe that perhaps cytoplasmic RBMY expressing HCC cell may arise from an alternative CSC driven origin in comparison to RBMY nuclear only expressing HCC cells, and thus explaining the differences in tumour properties both RBMY expression induces (Chua *et al.*, 2015).

To further investigate the prognostic effects of RBMY cytoplasmic presence, immunohistochemical experiments were undertaken (Chua *et al.*, 2015). Immunohistochemical experiments of 205 HCC patient tissue samples demonstrated that HCC tissues samples with cytoplasmic RBMY presence were associated with HCC tumours of a higher grade and larger in size than HCC tumour samples from RBMY nuclear only expression (Chua *et al.*, 2015). Further examination on tissue samples demonstrated that cytoplasmic RBMY presence in tissues increased the chances for liver cancer metastasis and the reappearance of liver tumours (Chua *et al.*, 2015). The high mortality rate and poor survivability of young patients between 5 to 10 years with HCC was shown to be linked to cytoplasmic RBMY presence (Chua *et al.*, 2015). As aforementioned, young persons of this age do not have a high active AR signalling and RBMY may act via an alternative signalling pathway to induce tumour initiation and development in men especially young male persons (Chua *et al.*, 2015).

The data discussed thus far supports a conclusion in which RBMY expression is highly specific and is ectopically expressed only in male liver cancer tissues as supported by *RBMY* transcripts being detected to be expressed in male cancer tissues only (Tsuei *et al.*, 2004). RBMY's male cancer specificity was further supported by immunostaining experiments which showed that RBMY expression in liver cancer tissues was restricted to tumour regions only (Kido *et al.*, 2020). The prognostic value of RBMY is however complicated based on the aforementioned results, Huey-Huey Chua et al describes RBMY nuclear-cytoplasmic presence to be associated with a more aggressive, highly tumorigenic, worse prognostic HCC compared to HCC cases with only nuclear positive RBMY expression (Chua *et al.*, 2015). On the contrary, Tatsuo Kido et al were unable to detect any cytoplasmic RBMY presence in their experiments and showed that the poor prognostic value seen in RBMY expressing HCC male patients is as the result of nuclear RBMY expression. Furthermore, In contrast of Huey-Huey Chua reporting, another study, using heterokaryon experiments with NIH3T3 cells, showed that human RBMY was shown to be a non-shuttling protein, with the inability unable to shuttle between the nucleus and the cytoplasm but restricted to the nucleus (Dreumont *et al.*, 2010). Another major limitation that

may offer an explanation to Huey Huey Chua study finding of cytoplasmic RBMY is the method of experimentation used in the study. Immunohistochemistry and Immunofluorescence imaging utilised in this study to show cytoplasmic RBMY measured relative rather absolute abundance of RBMY in the cytoplasm versus the nucleus, perhaps there may be substantial amounts of RBMY in the nucleus even if it looks by eye as though the vast majority is in the cytoplasm. Furthermore, if the nuclear import is a limiting factor, it may be that samples with relatively low RBMY levels it all ends up in the nucleus. For example, consider Sample A with 0 RBMY units in cytoplasm and 1 RBMY unit in nucleus and Sample B with 50 RBMY units in cytoplasm and 3 RBMY units in nucleus, although Sample B will appear by IF and IHC to be almost exclusively cytoplasmic, it still has more nuclear RBMY than sample A. Additionally, as the studies did not utilise the same antibodies, an evaluation of the antibodies used to detect cytoplasmic RBMY in Huey-Huey Chua study is necessary as their results are still to be confirmed by any other studies. Never-the-less, the data presented thus far shows that high RBMY expression rate and specificity suggests strong association of RBMY with hepatocarcinogenesis.

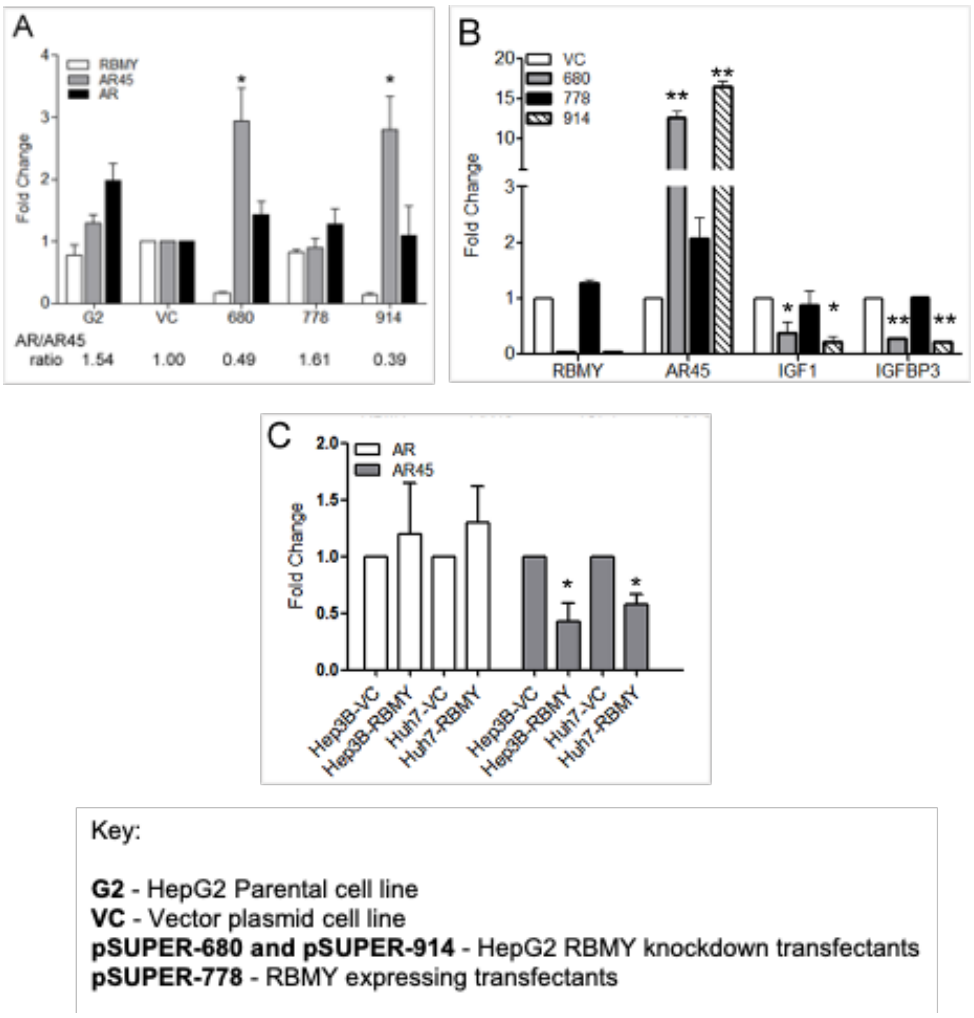
1.4.6 Studies of RBMY Function in Cell Lines

Using RBMY transfected NIH3T3 cells, RBMY was first described as a novel oncogene as combinational analysis of cell growth assessment and NIH3T3 inoculation in nude mice showed that RBMY could drive tumour formation (Tsuei *et al.*, 2004). This was the key study for what has become a significant research topic in cancer biology as the NIH3T3 experiment is a classic assay that was essential in identifying oncogenes including *RAS* and *HER2*, which were identified by their oncogenic ability to induce focus formation in NIH3T3 cells (Varmus, 1984).

Via conducting an anchorage independence assay with HepG2 cell lines, Daw-Jen Tsuei study demonstrated that *RBMY* knockdown was associated reduced anchorage-independent growth of HepG2 cells (Tsuei *et al.*, 2011). HepG2 cells with the *RBMY* knockdown displayed a reduced colony formation capability by

greater than 40% when compared to *RBMY* expressing colonies. This was significant in showing that RBMY expression in HCC might be associated with transformation efficiency of HepG2 cells (Tsuei *et al.*, 2011).

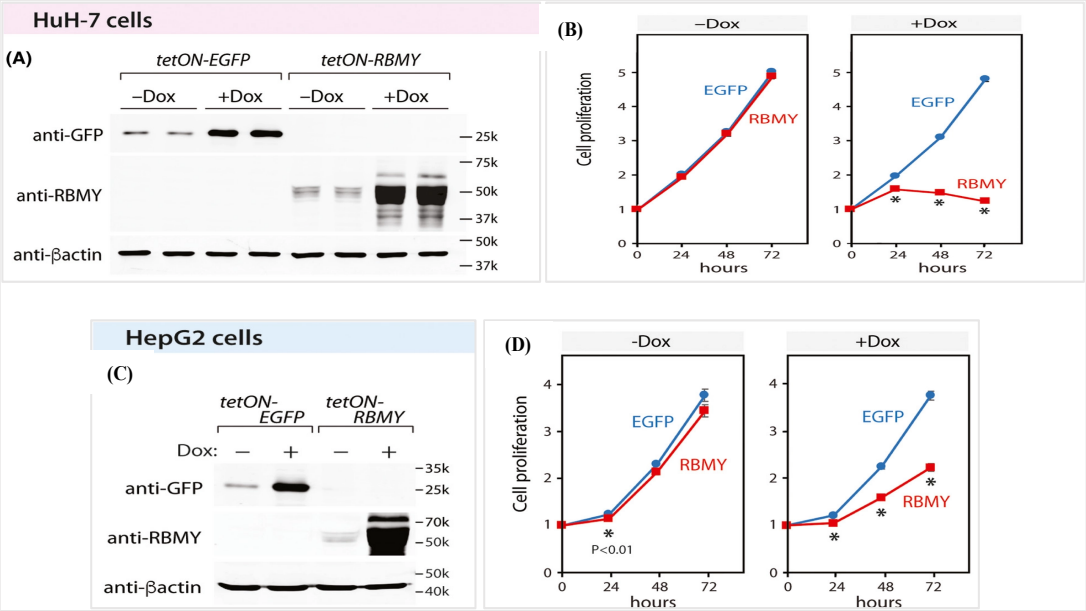
RBMY is thought to interact with AR to promote oncogenic activity, although this interaction when it occurs directly or indirectly is yet to be shown (Tsuei *et al.*, 2011). *RBMY* expression effect on the expression AR and its inhibitory isoform AR45 in *RBMY* knockdown HepG2 cell lines was analysed (Tsuei *et al.*, 2011). *RBMY* knockdown showed a three-fold increase in AR45 expression, however, the AR45/AR expression ratios reduced by ≥ 50% (Fig. 5A) (Tsuei *et al.*, 2011). Furthermore, AR downstream target genes IGF-1 and IGF-3 expression in HepG2 cell line were significantly reduced (Fig. 5B), illustrating the effect of *RBMY* on AR signalling activation (Tsuei *et al.*, 2011). Additionally, overexpression of *RBMY* in other male hepatocellular carcinoma cell lines Hep3B and Huh7 significantly reduced the expression of AR45 by 42% and 80% respectively (Fig. 5C) (Tsuei *et al.*, 2011).



(Tsuei *et al.*, 2011)

Figure 5: The effect of RBMY expression in HCC cells lines on the expression of AR and AR45. **(A).** Bar charts of a semi quantitative RT-PCR and Densitometric analysis of expression effect of RBMY on AR and AR45 (Tsuei *et al.*, 2011). The expression of RBMY, AR and AR45 in G2, VC, pSIPER-680, pSUPER-680 and -914 and pSUPER-778 HepG2 cell lines (Tsuei *et al.*, 2011). RBMY expression in HepG2 cell line had a drastic effect on the expression ratio of AR and its inhibitory isoform AR45 as demonstrated by densitometry analysis, as AR45 expression tripled in RBMY knockdown cells in comparison to VC (Tsuei *et al.*, 2011). (B). Bar chart plots of quantitative RT-PCR analysis on the knockdown effect of RBMY on the expression of AR45, IGF-1 and IGFBP-3 expression in RBMY knockdown pSUPER-680 and -914 cell lines and RBMY expressing VC and pSUPER-778 cell lines (Tsuei *et al.*, 2011). RBMY knockdown is shown to enhance the expression of AR45 whilst suppressing the expression of AR and consequently inhibiting transcription of AR target genes IGF-1 and AGFBP-3 (Tsuei *et al.*, 2011). (C). Quantitative RT-PCR of AR and AR45 expression changes in control and RBMY transfected Hep3B and Huh7. These experiments further illustrated that RBMY expression affects the expression equilibrium of AR/AR45 in HCC cell lines as there was a significant reduction in AR45 expression, with a 42% and 60% reduction in both RBMY transfected Hep3B and Huh7 cells respectively (Tsuei *et al.*, 2011).

In HuH-7 and HepG2 cells, inducible overexpression of *RBMY* impeded cellular proliferation upon initial *RBMY* overexpression (Fig.6A and Fig.6C) (Kido *et al.*, 2020). Fascinatingly, continuous *RBMY* overexpression actively induced adaptation of the tumour cell lines and re-establishment of tumour cellular proliferation properties beyond *RBMY* block (Kido *et al.*, 2020). Using RNAseq to examine the effect of *RBMY* overexpression on the expression of other genes, several genes including *EGF* and *PDGFB* were found downregulated as a result of *RBMY* overexpression (Kido *et al.*, 2020). These genes are involved in several signalling pathways including the frequently activated oncogenic RAS/RAF-MAPK and PI3K/AKT signalling pathways as well as other pathways linked to oncogenesis (Hippo and WNT signalling pathways) (Kido *et al.*, 2020) (Pan, 2010) (Manning and Cantley, 2007). RAS and AKT signalling pathways are frequently activated in many cancers including human HCC (Calvisi *et al.*, 2006) (Yang and Liu, 2017). The activation of these signalling pathways in cancers are is involved in the progression of many different cancers (Kido *et al.*, 2020), thus supporting the hypothesis in which RBMY has dual functions in liver cancer, acting as both a tumour suppressor that eventually switches to becoming a tumour promoter and driver (Kido *et al.*, 2020).

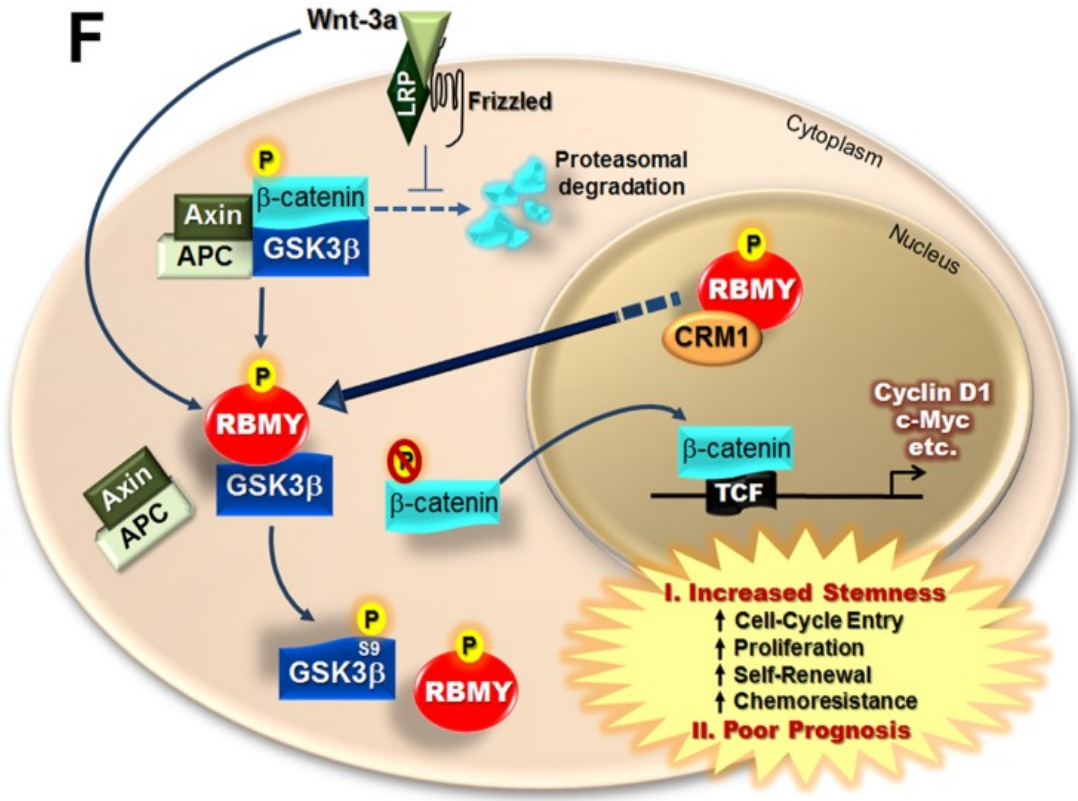


(Kido *et al.*, 2020)

Figure 6: RBMY overexpression induces initial cellular proliferation block in cell lines HuH-7 and HepG2. RBMY overexpression effects on the cellular proliferative activity of HuH-7 and Hep-G2 cell lines was studied using a lentiviral vector-mediated tet-ON mechanism(Kido *et al.*, 2020). This mechanism was used to induce a conditional RBMY overexpression in the HCC and hepatoblastoma cell lines. **(A), (C).** Western Blot images of TetON-EGFP and TetON-RBMY transfected HuH-7 and HepG2, blots taken 1-day post Dox induction. **(B), (D)** Cellular proliferation assays of Dox induced RBMY overexpression in HuH-7 and HepG2 cell lines. Overexpression of RBMY hinders cellular proliferation compared to EGFP induction cells (Kido *et al.*, 2020).

As aforementioned, one hypothesis to explain RBMY oncogenic activity in male liver cancers is RBMY being linked to the Wnt/β-catenin pathway to induce HCC via the emergence of CSCs. CSCs and progenitor stem cells share the same signalling pathway, In progenitor stem cells, this pathway regulates their development and gives progenitor stem cells their characteristics, and in CSCs, the same signalling pathway activates them (Yamashita and Wang, 2013). Consequently, this pathways also gives CSCs similar characteristics as progenitor stem cells in tumours, characteristics that promotes tumour stemness which include their ability to self-renew and chemo-resistance (Yamashita and Wang, 2013) (Chua *et al.*, 2015). In order to identify the oncogenic role of RBMY

via the Wnt/β-catenin signalling pathway, microarray experiments were conducted to identify the proteins that bind and interact with RBMY (Chua *et al.*, 2015). Active GSK3β was observed to interact with the SRGY boxes of cytoplasmic RBMY (Chua *et al.*, 2015). GSK3β is a key protein regulator of the Wnt/β-catenin signalling pathways as it creates β-catenin protein equilibrium whereby GSK3β associates itself with 2 other proteins APC and Axin-2 (Chua *et al.*, 2015). Analyses shows that RBMY is a competitive inhibitor of GSK3β, competitively inhibiting GSK3β from interactions with APC or Axin-2 to form the β-catenin destruction complex (Chua *et al.*, 2015). As cytoplasmic RBMY was observed to be activated and enriched in hepatospheres, promoting the self-renewability of these CSCs model, thus perhaps the growth of hepatospheres was dependant on GSK3β-β-catenin-RBMY (Chua *et al.*, 2015). Therefore, Wnt/β-catenin signalling is further enhanced via cytoplasmic RBMY phosphorylation activity to inactivate GSK3β, dysregulating stem cell renewability and promoting HCC initiation and progression.

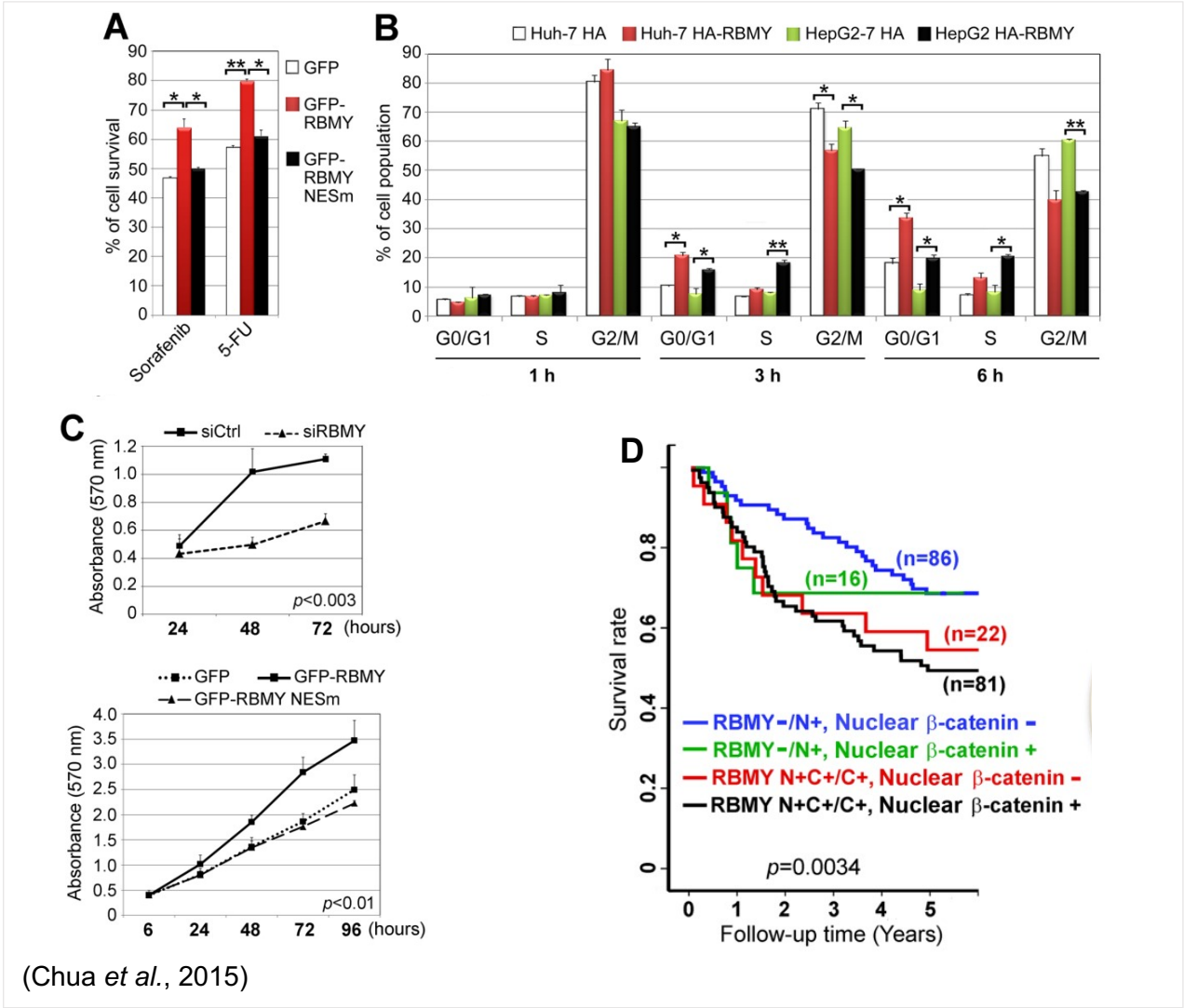


(Chua *et al.*, 2015)

Figure 7: Hypothetical model for Cytoplasmic RBMY oncogenic role on the β-catenin signalling pathway. Aberrant reactivation of RBMY during HCC tumorigenesis leads to RBMY protein expression. RBMY protein in the nucleus associates and binds with CRM1, which rapidly shuttles RBMY out of the nucleus into the cytoplasm. In the cytoplasm, active Wnt-3a maintains the cytoplasmic presence of RBMY (Chua *et al.*, 2015). RBMY is a competitive inhibitor of the β-catenin destruction complex, which directly or indirectly phosphorylates and inhibits GSK3β from binding to its β-catenin destruction complex binding partners Axin and APC (proteins with which RBMY competes with) (Chua *et al.*, 2015). Phosphorylation of GSK3β Ser9 residue as a result of RBMY binding to GSK3β (RBMY eventually dissociates from GSK3β, leaving phosphorylated GSK3β) blocks the phosphorylation and subsequent proteasomal degradation of β-catenin (Chua *et al.*, 2015). This is the stabilisation of β-catenin. β-catenin undergoes nuclear translocation and activate the transcription of Wnt target genes, including powerful oncogene proteins of Cyclin D1 and c-Myc (Chua *et al.*, 2015). Consequently, stabilisation of β-catenin induces HCC tumour stemness properties and drives poor prognosis of the disease (Chua *et al.*, 2015).

Cytoplasmic RBMY, via triggering stabilisation of β-catenin, promotes tumour stemness via chemoresistance and enhanced cellular proliferation (Chua *et al.*, 2015). HCC cells with cytoplasmic RBMY presence were shown to have acquired

chemoresistance and increased survivability *(Fig. 8A)* (Chua *et al.*, 2015). Additionally, *RBMY* expressing HuH-7 and HepG2 HCC cell lines demonstrated accelerated G_2/M phase and a sped-up entry to S-phase (Fig. 8B). Furthermore, expression of cytoplasmic RBMY is associated with expression of Ki-67, a proliferation signifier, and male patients expressing nuclear-cytoplamic and cytoplasmic RBMY had the worst prognosis (Fig. 8C) (Chua *et al.*, 2015). This supports the conclusion that poor prognosis of male HCC may be in part the result of RBMY activity that promotes tumour stemness in a Wnt–GSK3β–β-catenin signalling dependant manner. HCC cells that co-express CSCs markers acquire ability to express cytoplasmic RBMY, these tumours bore tumour stemness properties as a result of CSCs like enhanced ability to self-renewal, chemoresistance, proliferation, driven by cytoplasmic RBMY activities (Chua *et al.*, 2015). Consequently, these properties form an aggressive HCC subtype with poor prognosis, increased ability to metastasise and early remission (Lee *et al.*, 2006). CSCs markers, including EpCAM and SALL4 were transcriptionally active in this HCC subtype and were shown to have similar expression profile with cytoplasmic RBMY in same HCC tissues (Chua *et al.*, 2015). Based on the observed evidence, it is thought that cytoplasmic RBMY and CSCs markers are actively expressed in human and rodent fetal livers, albeit at very minute levels. This expression switched off in mature hepatocytes then reactivated in the tumour cells of this HCC subtype (Inada *et al.*, 2008). Unlike progenitor cell markers expression profile, RBMY expression in healthy tissues is scarce and predominantly expressed in high concentration in the testes, thus their high expression in HCC is indicative of a tumour promoting role.



(Chua *et al.*, 2015)

Figure 8: Cytoplasmic RBMY induces poorer prognosis of HCC in male patients.

(A), Analytical bar graphs of MMT (3-(4,5-dimethylthiazol-2-yl)−2,5-diphenyltetrazolium bromide) assays demonstrating the chemo-resistance effect that is induced by cytoplasmic *RBMY* presence in HuH-7 cells. In the assays, chemoresistance was measured by percentage of cells surviving after treatment with chemotherapeutic drugs (6μM sorafenib and 400ng/ml 5-fluorouracil). In the study, 3 HCC cell models groups of HuH-7 were used, GFP Hu7-7, *RBMY* HuH-7 and GFP *RBMY* NESm HuH-7. GFP-*RBMY* NESm HuH-7 lacks the nuclear export signal and thus are unable to shuttle *RBMY* into the cytoplasm. The results demonstrate that cytoplasmic *RBMY* induces enhanced survivability via chemo-resistance in HuH-7 cells compared to control cells (GFP) or cells with dysfunctional nuclear export of *RBMY* when incubated for 30 hours in chemo-drugs. **(B).** The effect of *RBMY* expression on the cell cycle progression of Huh-7 and HepG2. HA-*RBMY* and HA expressing Huh-7 and Hep-G2 cell lines were incubated for 20 hours in 0.2 μg/ml of the G2/M cell cycle arresting agent Nocodazole, allowing the cells to synchronise at the G2/M points of the cell cycle. Flow-cytometry results demonstrated that the HA-*RBMY* G2/M synchronised cells displayed advanced cell-cycle progression. Furthermore, HepG2 cells displayed ability for quicker entry into S-phase. **(C)** HepG2 and Huh7 growth curves. *RBMY* expression in HCC is associated with enhanced growth advantage of HCC tumour cells. **(D).** Kaplan-Meler 5-year survival graphs of *RBMY* and nuclear β-catenin expressing or non-expressing HCC patients. The data demonstrates that expression of cytoplasmic *RBMY* effectively associates poor prognosis of the disease in patients (Chua *et al.*, 2015).

β-catenin stabilisation under RBMY control not only promote tumour growth, but also promotes the CSCs phenotypic properties in this HCC subtype (Chua *et al.*, 2015), thus enhancing tumour stemness and promoting poor disease prognosis via increased self-renewability of tumour cells, acquired chemoresistance, enhanced proliferation and early recurrence of HCC in male patients (Zulehner *et al.*, 2010) (Mokkapati *et al.*, 2014). Thus, cytoplasmic RBMY is an important regulator and driver for development and progression of male HCC via the Wnt–RBMY–GSK3β regulatory circuit.

The male liver cancer cell lines HuH7, HepG2 and Hep3B have been instrumental in showing the effect of RBMY expression male HCC cell lines (*Hep G2 [HEPG2] ATCC® HB-8065™*, no date; *Hep 3B2.1-7 [Hep 3B, Hep-3B, Hep3B] ATCC ® HB-8064™ Homo sap*, no date; *HuH-7*, no date). Based on the several experiments with these cell lines, RBMY is shown to be linked with key oncogenic signalling molecular pathways that may in part drive male HCC prevalence alongside other risk factors. RBMY molecular interaction with the AR signalling pathway has been debated for several reasons and remains to be elucidated (Tsuei *et al.*, 2011). Part of the issue arise as RBMY is also reported to have interaction with Sam68, a protein involved in the Scr signalling pathway which is also a co-activator of AR (Venables and Eperon, 1999) (Matter, Herrlich and König, 2002). Sam68 is a well-established oncogenic protein that drives prostate cancer tumorigenesis in males by transactivating AR signalling (Rajan *et al.*, 2008). Thus, whether AR45 expression regulation by AR occurs directly or via Sam68 interaction remains to be elucidated (Tsuei *et al.*, 2011). Huey-Huey Chua study showing credible data of cytoplasmic RBMY associated with the Wnt/B-Catenin pathway to induce the emergence of CSCs (Chua *et al.*, 2015). However, their hypothesis is none-the-less currently a weak one as cytoplasmic RBMY presence is still being debated having been shown to not be seen in cells before and after the release of this hypothesis, in response to Huey-Huey Chua findings.

The most convincing hypothesis, albeit not completely established, is from the Tatsuo kido et al study in a ‘double agent’ role is proposed for RBMY in male

HCC cancer; acting as both a tumour suppressor and a proto-oncogene dependant on spatiotemporal changes in expression levels (Fig. 7) (Kido *et al.*, 2020).

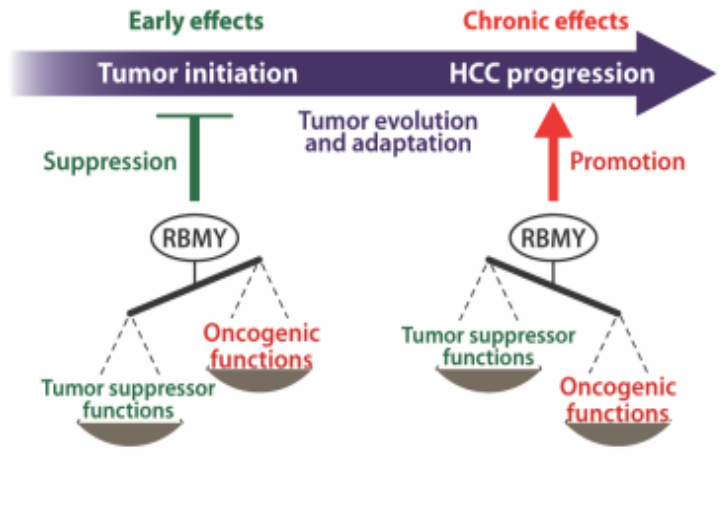


Figure 9: The double agent roles of RBMY protein in male HCC. (Kido *et al.*, 2020).

Supported by results in which RBMY overexpression was linked to *EGF* and *PDGFB* downregulation, initial RBMY overexpression in liver cancer tissues acts as a tumour initiation block, inhibiting tumour initiating activities of oncogenic signalling pathways like PIP3/AKT and RAS/RAF/MAP (Kido *et al.*, 2020). Clonal evolution of surviving cancer cells driven by the cellular proliferation block imposed by RBMY eventually promotes aggressive HCC development and progression. Thus, male specific HCC may be driven by persistent RBMY overexpression, this notion is reinforced by the correlation of high RBMY expression patterns and high expression pattern of the cell proliferation marker Ki-67 in male HCC specimens (Kido *et al.*, 2020). Currently, of the suggested hypothesis explaining RBMY role in male HCC tumours, we think this is the best one so far. However, it remains quite unclear why the RBMY block in male HCC leads to excess tumours in males as females have no RBMY and no block to overcome. Never-the-less, RBMY HCC expression is associated with poor

prognosis and its expression in these cell lines and the associated effects discussed above suggests that RBMY does have a tumorigenic role in male HCC (Kido *et al.*, 2020) (Chua *et al.*, 2015).

1.4.7 Transgenic animal studies of RBMY function.

Cytoplasmic RBMY expression pattern was also demonstrated in transgenic mice with HCC (Tsuei *et al.*, 2011), further supporting the idea that perhaps it is the cytoplasmic RBMY presence that is responsible in part for the high prevalence of HCC development in males.

An extensive number of experiments has been carried out in transgenic mice models to demonstrate RBMY's ability to drive tissue transformation in vivo and further support RBMY's oncogenic ability (Tsuei *et al.*, 2004). Daw-jen Tsuei initial observation that RBMY was being expressed in tumours was further supported by examining the tumorigenicity of *RBMY* transfected NIH3T3 cells in nude mice (Table 1) (Tsuei *et al.*, 2004). High and low dosages of *RBMY* transfects and parental NIH3T3 cell (refer to Table 1 for the dosage numbers) were inoculated into nude mice (Tsuei *et al.*, 2004). The results from this inoculation revealed that *RBMY* was potentiating liver tumorigenesis in nude mice as all the nude mice injected with wild type *RBMY* developed tumours, regardless of the dosage (Tsuei *et al.*, 2004). Moreover, mice inoculated with a low dosage of *RBMY* variants with an incomplete C terminus domain, less than four SRGY boxes were unable to develop tumours. However, inoculation with these *RBMY* variants at high dosage levels induced liver tumours, albeit inducing significantly smaller and slower growing tumours compared to full length wild type *RBMY* induced tumours (Table 1) (Tsuei *et al.*, 2004). The results from this transgenic mouse experiments in this study demonstrated that *RBMY* C-terminal domain as shown in cellular studies, alongside being important for the nuclear localisation for RBMY is also fundamental for RBMY transformation capability, possibly resulting from the former role (Tsuei *et al.*, 2004).

T[illegible]umor formation of NIH3T3 transfectants in nude mice

Transfectants	*Inoculated cell no.*	*Days to tumor initiation*	*Tumor size[a] (in cm³)*	*Tumor incidence*
NIH3T3/–	5×10^6	—	—	0/2
NIH3T3/pcDNA3	5×10^6	40	0.48	1/5
NIH3T3/pcDNA3-RBM384[b]	10^6	—	—	0/5
	5×10^6	35.2±7.9	2.15±1.81	5/5
NIH3T3/pcDNA3-RBM496	10^6	36.8±13.1	0.23±0.17	5/5
	5×10^6	24.0±11.6	1.57±0.84	5/5
SK-OV-3	5×10^6	10.0±1.4	4.53±3.33	2/2

(Tsuei *et al.*, 2004).

Table 1: NIH3T3 cell transfected nude mice tumour establishment. Nude mice tumour establishment measured (tumour mass measured in cm^3) and analysed 7 weeks post inoculation. 2 *RBMY* variants were utilised in these experiments: RBM496 encode for wild-type *RBMY* and RBM383 encode for RBMY protein with a single base deletion on exon 8 resulting in loss of 2 SRGY boxes (Tsuei *et al.*, 2004)

Further transgenic mouse experiments illustrated RBMYs oncogenicity as RBMY was shown to enhance HCC in transgenic mice with liver-specific RBMY expression (Tsuei *et al.*, 2011). Daw-Jen Tsuei et al established RBMY liver specific expressing transgenic mice models, the ability of these mice to spontaneous form HCC tumours was analysed (Tsuei *et al.*, 2011). Spontaneous liver tumours developed at twice the rate in *RBMY* transgenic mice than in wild type mice, with 8.7% males and 4.5% female mice developing HCC tumours (Tsuei *et al.*, 2011). In order to further explore transformative activity of RBMY, diethylnitrosamine (DEN) chemical hepatocellular carcinogenesis induced mouse models were utilised. In DEN-induced male mice, RBMY was a significant factor in driving tumorigenesis in transgenic mice as tumours were observed as early as 14 weeks post chemical carcinogenesis induction, with 86% of 14 RBMY DEN-induced transgenic mice having developed tumours whilst only 42% of 12 DEN-induced control group developed tumours (Tsuei *et al.*, 2011). At 34 weeks post DEN-induction, significantly more enlarged tumours with diameter ≥3 mm were observed with 52 RBMY DEN-induced transgenic mice compared to 19 in DEN-induced control group (Table 2) (Tsuei *et al.*, 2011). As a result of their oestrogen regulated immune system, wild type female mice have resistance to DEN carcinogenesis (Naugler *et al.*, 2007). However, RBMY ability to transform liver tissue in transgenic mouse model was further illustrated by RBMYs ability to significantly drive cancer lesions in female transgenic mouse model (Tsuei *et al.*,

2011), demonstrating the male specific RBMY enhanced HCC development even in female transgenic mouse models (Tsuei *et al.*, 2011).

Gender	RBMY	14 wks	26 wks	34 wks	
		Incidence of cancerous lesions (No. mice)	Incidence of cancerous lesions (No. mice)	Incidence of cancerous lesions (No. Mice)	No. tumor ≥3 mm (No. mice)
F	+	ND	78% (7/9)**	90% (9/10)**	0 (10)
F	−	ND	10% (1/10)	23% (3/13)	0 (13)
M	+	86% (12/14)*	100% (5/5)	100% (6/6)	52 (6)*
M	−	42% (5/12)	90% (9/10)	100% (7/7)	19 (7)

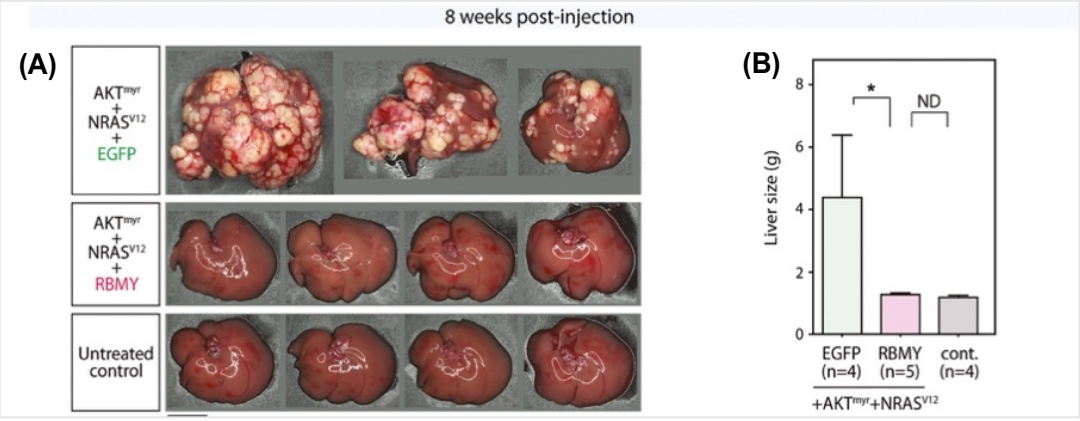
RBMY transgenic or control mice (14-day old) received a single intra-peritoneal injection of diethyl nitrosamine (DEN) 10 mg/kg of body weight. They were euthanized by cervical dislocation at 26 and 34 weeks for females, and at 14, 26, and 34 weeks for males after injection.
ND, not determined.
*$p<0.05$;
**$p<0.01$, RBMY-positive vs. RBMY-negative mice; by Student's t test.
doi:10.1371/journal.pone.0026948.t002

(Tsuei *et al.*, 2011).

Table 2: Hepatocarcinogenesis incidence (DEN-induced) in RBMY transgenic and control mice.

As aforementioned, RBMY overexpression was demonstrated to be associated with downregulation of genes involved in the oncogenic signalling pathways of RAS/RAF/MAP and PI3K/AKT (Kido *et al.*, 2020) (Pan, 2010) (Manning and Cantley, 2007). RAS and AKT signalling pathways are frequently activated in many cancers including human HCC (Kido *et al.*, 2020). After transcriptome analysis demonstrated that RBMY overexpression directly affects the RAS and AKT signalling pathways, RBMY overexpression effects were explored in vivo (Kido *et al.*, 2020) with AKT and RAS induced hepatic cancer transgenic mouse models. Two sets of mice models were designed to study the in vivo effects of RBMY overexpression, (i) a RBMY mouse model of *RBMY* carrying *pT3-* vector that constitutively express Ras and AKT under CAG and EF1α promoters respectively (*AKT^{myr}/$NRAS^{V12}$/RBMY* mice), which leads to spontaneous development of liver cancers and (ii) a AKT/Ras positive control mouse model with pT3-EGFP expression vector (*AKT^{myr}/$NRAS^{V12}$/EGFP* mice) (Fig. 5A) (Kido *et al.*, 2020). After 8 weeks post hydrodynamic tail injection (refer to supp. Fig. 1 for more information on how this methods was conducted in this study) with the plasmids carrying the constitutively active Ras (*$NRAS^{V12}$*) and AKT (*AKT^{myr}*) oncogenes, the mice livers were collected and the analysis showed that the

positive control mice group had developed significantly large HCC tumours, whilst the livers from RBMY mice model group had no HCC tumours developing (Kido *et al.*, 2020). RBMY mice models results of no tumour development 8 weeks post injection were very similar to the negative control mice group (untreated mice). This results suggests that the addition of a tetracycline inducible *RBMY* transgene was able to suppress RAS and/or AKT induced cancer development in the RBMY model, showing that RBMY functions in this context as a tumour suppressor gene (Kido *et al.*, 2020). This experiment was repeated on 3 different occasions with multiple animal groups (4-6 mice in each group) (Fig. 10) (Kido *et al.*, 2020). Thus, these results are suggestive of a RBMY role in which RBMY acts as a tumour suppressor, inhibiting the oncogenic pathways of RAS and AKT (Kido *et al.*, 2020). There is a limitation to the interpretation of this experiment as the experiment puts human RBMY into a mouse, therefore, the effects may be different between mice and humans due to differential expression or recognition of the downstream target genes and/or their splice sites.



(Kido *et al.*, 2020)

Figure 10: RBMY overexpression suppresses the Ras and AKT oncogenes induced liver cancer in mouse models. (A), Selected morphological images of mouse livers at 8-week post injection of livers with tumorigenesis driving plasmid vectors carrying constitutively active Ras and AKT oncogenes. The active AKT and RAS oncogenes promoted the initiation and development of tumours in the control mice (top row), however, in the RBMY expressing control mice, RBMY inhibited any tumour formation (middle row), with results alike to the untreated control mice(Kido *et al.*, 2020). (B). Mean average weights of the mice livers of mice utilised and shown in Fig. 10(A) (Kido *et al.*, 2020).

Dysregulation of signalling pathways regulating the self-renewal ability of stem cells is assumed to potentiate the rise of CSCs during tumour development, cytoplasmic RBMY is thought to be able to aberrantly active this signalling pathway, giving rise to male specific HCC that has stem-like properties as a result of CSCs (Chua *et al.*, 2015). By studying F344 rat model livers using an in vitro nonadherent spheroid in stem cell culture to explore tumour stemness via studying tumour initiation, several observations indicated the association between RBMY nuclear-cytoplasmic and cytoplasmic expression patterns with CSCs signifiers and self-renewal factors (Chua *et al.*, 2015). F344 rat models were utilised in an attempt to show exclusive cytoplasmic RBMY expression in the rat bile ductules and oval cells. Tumours resembling hepatospheres began to appear from the 5th day (Chua *et al.*, 2015). Thus, rat models illustrated this idea of CSC emergence in HCCs linked with cytoplasmic RBMY expression in these tumours. Thus, like the other studies looking at RBMY oncogenesis in vivo, increased cytoplasmic RBMY expression pattern was shown in HCC rat models

(Chua *et al.*, 2015), further supporting the idea that perhaps it is cytoplasmic RBMY presence that is responsible for the most part of RBMY induced hepatic cancer development in males.

The data from the in vivo studies clearly show RBMYs' ability to promote HCC tumour initiation and development. However, there seem to be a contradictory analysis between 2 studies that showed RBMY effect on tissue transformation in vivo. Daw-Jen Tsuei et al (2011) showed that in both RBMY transgenic mouse models and DEN induced mouse models, RBMY was a significant driving force for tumorigenesis and tumour development as cancer lesions in the transgenic mice was observed as early as 14 weeks post DEN-induction (Tsuei *et al.*, 2011). However, Tatsuo Kido showed data to support their hypothesis in which RBMY is hypothesised to act as a tumour suppressor that inhibits oncogenic effects of signalling pathways associated with RAS and AKT proteins (Kido *et al.*, 2020). Clonal evolution of surviving cancer cells driven by the cellular proliferation block imposed by RBMY eventually promotes aggressive HCC development and progression (Kido *et al.*, 2020). Tatsuo Kido et al results showed that RBMY was able to suppress the tumorigenic effects of the RAS/RAF/MAP and PI3K/AKT signalling pathways (Kido *et al.*, 2020). Daw-Jen Tsuei et al analysed the liver tissues on their DEN induced mice models 14 weeks and 34 weeks post induction whilst the liver tissue from their transgenic mice models were analysed at 15, 21 and 24 months, whilst Tastuo Kido et al transgenic mice model liver cancer tissues were analysed at 8 weeks post Hydrodynamic tail injection (Tsuei *et al.*, 2011) (Kido *et al.*, 2020). It is very difficult to try and compare results from 2 different studies because of the various variables that differ within each study but, we stand on the aforementioned conclusion in which we believe Tatsuo Kido et al offers the best hypothesis to explain RBMY oncogenic activity in liver cancers. We therefore do not think that these results contradict each other, but rather shows the in vivo effect on RBMY at 2 different time frames. It is therefore possible if left longer, the RBMY block proposed by Tastuo kido et al would have been eventually overcome in their transgenic mice models and there would have eventually seen tumour development in the RBMY transgenic mice models. Regardless, the experiments carried out in this study further illustrates the

carcinogenic effect of RBMY in HCC cell lines and its ability to promote tumour initiation and development in vivo in both transgenic and DEN-induced mice.

1.5 *ZFY* Gene:

1.5.1 ZFY, an Evolutionary Story.

Once considered as a candidate gene for the testis' determining factor, the *ZFY* family of genes are an interesting phenomenon that we are still defining since the initial identification of the mouse *Zfy1* and *Zfy2* gens in the late 1980s (Yamauchi *et al.*, 2015). When the actual testis-determining factor Sry was discovered, research on *Zfy* gene halted for over 2 decades, eventually emerging in research as a novel spermatogenic factor (Yamauchi *et al.*, 2015). Deriving its name from having a protein structure that is typical of Zinc finger transcription factors (Decarpentrie *et al.*, 2012), *ZFY* gene expression is conserved in all eutherian mammals with expression patterns varying between each species (Decarpentrie *et al.*, 2012). Humans express a homologous pair of zinc finger proteins of ZFY family, a Y encoded *ZFY* gene with 2 splice variants, and its X chromosome homologue ZFX (Decarpentrie *et al.*, 2012). However, mice express a paralogous pair of Zfy genes named *Zfy1* and *Zfy2*, alongside *Zfx* and *Zfa* (an autosomal copy of *ZFX* which may or may not be a pseudogene) (Decarpentrie *et al.*, 2012) (Peter Koopman, Ashworth and Lovell-Badge, 1991). These mice *Zfy* genes are predicted to have arisen from a single gene duplication event (Peter Koopman, Ashworth and Lovell-Badge, 1991).

ZFY and *ZFY* related genes are in a unique subgroup of the zinc finger proteins, their primary structure has a very distinctive structural feature that even in this large group of transcription factors is unique between all zinc finger proteins, this is a two-finger repeated structural pattern. This unique feature is suggested to be reflective of the evolution of *ZFY* from their primordial ancestral gene that shared the same structure had had two zinc fingers (Peter Koopman, Ashworth and Lovell-Badge, 1991). Mammalian *ZFY* genes encode homologous proteins that are structurally similar, with these proteins characterised by a highly conserved

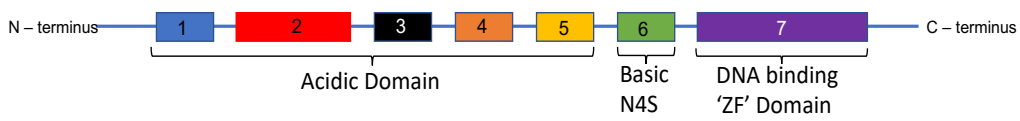
12-13 Cys-Cys and His-His form Zinc finger domain on the C terminus, a localisation small nuclear code that separates the Zinc finger domain from the proceeding activating large acidic N-terminus domain (Peter Koopman, Ashworth and Lovell-Badge, 1991) (Page *et al.*, 1987) (Schneider-Gädicke *et al.*, 1989). The negatively charged acidic activating N-terminus domain is predicted to bind to target protein (Decarpentrie *et al.*, 2012): the structure of ZFY and experiments with mouse Zfy has led to the suggestion that zinc finger domain in ZFY proteins is utilised to recruit regulatory or transcription machinery (Decarpentrie *et al.*, 2012).

1.5.2 *ZFY* gene expression.

As aforementioned, *ZFY* expression varies between species: humans express a single *ZFY* gene with 2 splice variants, of which the full length variant is reported to have transactivation activity in yeast reporter system (Mardon *et al.*, 1990). We propose that along with a number of high-risk factors including lifestyle and HPV infection, RBMY carcinogenic effect also drive the high male prevalence of HNSCC. We propose that RBMY splices *ZFY* into *ZFY*-short which promotes the carcinogenic effect of RBMY. This proposal is the result of several evidences that we will summarise below:

- Our hypothesis for ZFY is based mostly on circumstantial evidence in testis and some unpublished results of cancer cells, in which *ZFY* splicing is shown in UM-SCC-104 cell line as a result of both *ZFY*-long and *ZFY*-short bands seen on gel (see supp. Fig.3). This is a cell line we have shown to also express RBMY.
- *ZFY*-short as will be mentioned further has predominant expression in cells pre-meiotic germ line cells (Decarpentrie *et al.*, 2012), cells in which the splicing factor RBMY protein is preferentially translated (Abid *et al.*, 2013).
- Deletion of Y gene AZFb region in a patient (Ste-363) in a 2012 study, the same region where the functioning RBMY genes are clustered on resulted in lowered expressions of *ZFY*-short , whilst a man (Apop12) with a similar

stage meiotic arrest but no AZFb deletion had no reduction in *ZFY*-short levels (Decarpentrie *et al.*, 2012).



Gene	Exons	Expression	Pre- or post- meotic
Human ZFY-Long	All	Widespread but Higher in Testis	Both, the ratio is unclear
Human ZFY-Short	Excludes 2	Testis specific	Both, Perhaps more pre-meiosis
Zfy1-Long	All	Testis specific	Both, mainly post meiotic – minor isoform in the testis
Zfy1-Short	Excludes 2	Testis specific	Both, mainly pre-meiotic, major testis isoform
Zfy2-Long	All	Testis specific	Almost exclusively post meiotic – major testis isoform
Zfy2-Short	Excludes 2	Low to none	Very little if at all, minor isoform

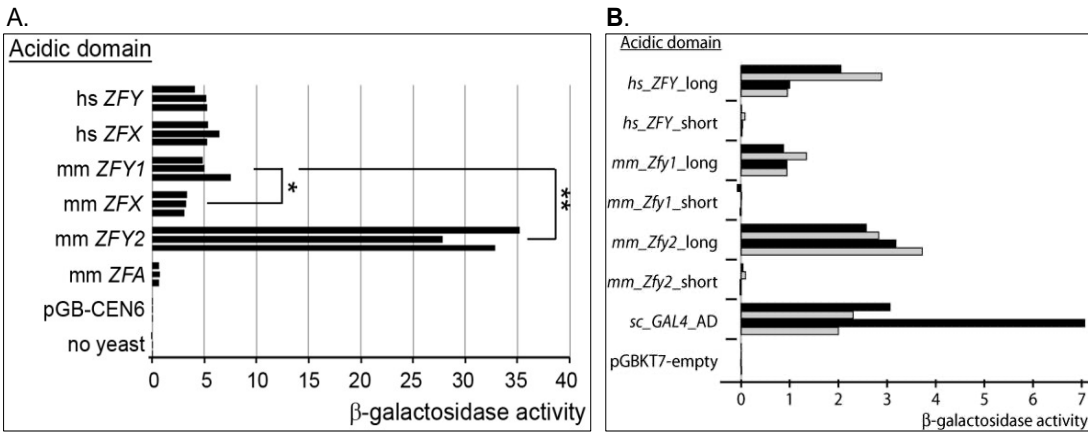
Figure 11: *ZF* genes coding exon arrangement.

The short form of human *ZFY* lacks the important acidic activating domain and as a consequence has been shown to be unable to conduct any transactivation activity (Mardon *et al.*, 1990). Human *ZFY* gene expression is unique to that of mouse *Zfy* genes, as human *ZFY* expression is dominant in the adult male testes and with a significantly low expression profile in other somatic tissues including the liver (Nagamine *et al.*, 1989), whilst mouse paralogous pair of *Zfy* genes are both exclusively expressed in mice testis (Decarpentrie *et al.*, 2012). *ZFY*-long expression is seen in most other mammals (Shimmin, Chang and Li, 1994), however, with the exception of rodent *Zfy*-long tissue distribution (restricted to germline cells), the tissue expression of the *ZFY*-long in other mammals remains undefined. *ZFY* short remains to be characterised, with its expression reported in the testis alone in human, mouse and sheep, the only species in which *ZFY*-short expression was assessed in (Decarpentrie *et al.*, 2012)[105].

Mouse *Zfy1* is predominantly transcribed in its short inactive variant whilst mouse *Zfy2* is principally transcribed as a long active variant (Vernet *et al.*, 2014).

Consequently, mouse *Zfy2* is predicted to play a more significant role in *Zfy* transactivation as full length *Zfy2* was shown to have a significantly greater transactivation activity than that of full-length *Zfy1* (Vernet *et al.*, 2014). Interestingly, human *ZFY* expression does not mirror that of mouse *Zfy2* but is similar to that of mouse *Zfy1* as both human *ZFY* and mouse *Zfy1* have the capability to produce transcripts with or without the second coding exon in germ cells whilst *Zfy2* almost always is expressed as full-length transcripts (Vernet *et al.*, 2014). One take suggested by Michael Mitchell is that this possibly may indicate that Zfy1 is essential for assuring core eutherian ZFY functions, whilst Zfy2 is specialised for elimination of spermatocytes with failed sex chromosome pairing (Decarpentrie *et al.*, 2012). However, another possibility and one we prefer is that it's possible that ZFX has fully taken over the function in non-testis tissues, while Zfy1 has specialised for pre-meiotic and Zfy2 for post-meiotic functions.

When bound as a fusion protein with to the DNA binding domains of Gal4 transcription activator, the acidic domains Zfy2 and Zfx are both reported to activate transcription in yeast Saccharomyces cerevisiae (Mardon *et al.*, 1990). A follow up study used constructs expressing short and long acidic domains of Zfy1, Zfy2 and human ZFY bound to the DNA binding domain of Gal 4 (Decarpentrie *et al.*, 2012). They showed that full length Zfy2, long acidic chain of Zfy1 and human ZFY are able to transactivate the Gal-4 promoters in yeast Y187 strain , whilst the short acidic domains fails to promote the transactivation short acidic domain ZFY1, ZFY2 and human ZFY (Decarpentrie *et al.*, 2012). Based on this, it is assumed that the long and short acidic domains of ZFY, Zfy1 and Zfy2 are functionally distinct, with the long, not short acidic domains able to promote Gal-4 dependent promoters transactivation (Decarpentrie *et al.*, 2012). This is supportive evidence for a proposed model that explains ZFY being able to specifically recruit and bind to TRAPP sequences as a result of its long acidic domain (Decarpentrie *et al.*, 2012).



(Vernet *et al.*, 2014) (Decarpentrie *et al.*, 2012).

Figure 12: Transactivation activity of 'Zf' genes acidic domain. (A). Transactivation activity of Gal-4 DNA binding domain fused to the short and long acidic domains of mouse (denoted by mm in the figure) Zfy1, Zfy2, Zfx and human ZFY (denoted by hs in the figure) and ZFX. Transactivation activity is judged upon the β-galactosidase levels produced in yeast under Gal-4 promoter. The ZFY1, ZFY2 and human ZFY long and short acidic domains were fused of Gal-4 DNA binding domain in a vector utilised to transform Y187 yeast with a LacZ system under the Gal-4 control. Transcription activity measured via analysis of β-galactosidase production. The graphs demonstrate that Zfy2 acidic domains has the most powerful transactivation activity in comparison to Zfy1 or human ZFY: pGB-CEN6 is the negative control(Vernet *et al.*, 2014). ***(B).*** long and short acidic domains of human ZFY, Zfy1 and Zfy2 are functionally distinct, with the long acidic domains able to transactivate transcription in a Gal-4 promoter system in yeast. Transcription activity measured via analysis of β-galactosidase activity (Decarpentrie *et al.*, 2012).

The protein structure of ZFY proteins and the ability of ZFY proteins to transactivate expression activity in Gal-4 model systems alongside other experiments has led to the predictions that these proteins might be able to bind to specific DNA sites and via the N-terminus acidic domain to recruit transcriptional machinery (Decarpentrie *et al.*, 2012). Yeast Gal-4 system works via recruitment of the large pre-initiation complex proteins to the target gene promoter domain (Decarpentrie *et al.*, 2012). These pre-initiation complex proteins include histone acetyltransferases and TATA binding protein whose recruitment is contingent upon the interaction of Gal-4 acidic terminus with Tra1

protein, a homologue of the mammalian Transformation/transcription domain-associated protein (TRRAP) (Decarpentrie *et al.*, 2012) (Traven, Jelicic and Sopta, 2006). TRRAP is key protein in DNA replication, DNA repair and transcription processes, it is vital for the recruitment of chromatin histone acetyltransferases, which are essential for the function of the aforementioned DNA replication and repair and transcription (Murr *et al.*, 2007). Based on the Gal-4 experiments with ZFY, it is assumed that ZFY-long during meiosis actively recruits TRRAP complexes onto chromosomes (Decarpentrie *et al.*, 2012). Although ZFY-short is unable to transactivate the Gal-4 promoters, it still maintains a 183 amino acid acidic domain, it is possible that during spermiogenesis, ZFY-short may be able to recruit other transcription regulatory proteins to the chromatin (Decarpentrie *et al.*, 2012). Alternatively, we suggest that the function of ZFY-short may be to compete with ZFY-long for DNA binding sites, and thus prevent the downstream target genes from being activated too early.

Human ZFY and mouse Zfy despite having structural resemblances display functional differences which are reflected upon by the differences in their genetic sequences (P Koopman, Ashworth and Lovell-Badge, 1991) (Nagamine *et al.*, 1989). Furthermore, although both mouse *Zfy* and human *ZFY* genes share an matching putative start codon, the mouse *Zfy* genetic code as a result of in frame genetic coding deletions, encode for a Zfy protein that is 783 amino acids whilst the human *ZFY* gene encode a larger ZFY protein with 801 amino acids (Nagamine *et al.*, 1989) (Lau and Chan, 1989). Moreover, although human ZFY which has been likened to have similar expression profile with Zfy-2, the 2 genes only share 82% and 73% sequence homology at the zinc finger domain and their C terminus domain respectively (P Koopman, Ashworth and Lovell-Badge, 1991) (Nagamine *et al.*, 1989). The protein homology of these genes is even lower at these sites, with 80% and 59% protein homology at the zinc finger domain and their C terminus domain respectively between the human ZFY and mouse Zfy-2 (Nagamine *et al.*, 1989). Thus, this illustrates that Human ZFY and mouse Zfy-2, although having been derived from the same ancestral protein, have divulged into distinct roles necessary for their species function.

Unlike mouse *Zfy* genes which are exclusively expressed in the gonadal tissues, human *ZFY* is predominantly but not exclusively expressed to the testes (Lau and Chan, 1989). Human *ZFY* gene expression is interesting as transcripts are expressed in a discrete-size manner in both somatic and testes tissues, with the testes expressing the major *ZFY* splice variant that is 3-kb whilst somatic tissues express the minor 5.7kb slice variant (Lau and Chan, 1989). The testis variant of human *ZFY* has exon 3 spliced out, this is a 573 bp long coding region that codes for about 50% of ZFY putative activating domain. Interestingly, and importantly to function of *ZFY* genes in both mice and human, human *ZFY* exon 3 resembles the exon 6 of the mouse *Zfy1* (Decarpentrie *et al.*, 2012). Thus, we assume that the majority of this variation in the *ZFY*/*Zfy* variants is in their UTRs, and that the ORF is the same between the 3kb and 5.7kb variants. This pattern of expression is reported to occur in various other proto-oncogenes(Nagamine *et al.*, 1989) (Mutter1', Grills' and Wolgemuth1', 1988) (Zakeri, Ponzetto and Wolgemuth, 1988). It is suggested that the difference in the human *ZFY* transcripts may be reflective of ZFY separate functions in somatic and gonadal tissues (Nagamine *et al.*, 1989).

Mouse Zfy1 and Zfy2 are shown to be absolutely vital for spermatogenesis and the normal development of spermatozoa, and the disruption of *Zfy1* and *Zfy2* by Crispr Cas9 led to complete destruction of sperm development and function (Nakasuji *et al.*, 2017). The disruption of *Zfy1* and *Zfy2* is shown to impact sperm development via inducing substantial genetic and morphological defects on the developing sperm: the disruption of these genes led to developing spermatozoa with chromosomal aberrations, complete defect of spermatozoan in capacitation, destruction to acrosome reaction, and defect in the oocyte activation (Nakasuji *et al.*, 2017). Zfy2 is a regulator of male germline precursor to sperm, as it plays the significant role in regulating the transformation of precursor cells into spermatozoa as well as proper development of sperm cells (Yamauchi *et al.*, 2015). Despite Zfy2 having more significant activity in spermatogenesis than Zfy1, they are both essential proteins at various phases of the spermatogenesis, and proper levels of both proteins at these stages is critical for the progression of sperm cell development (Yamauchi *et al.*, 2015). Puzzlingly, the CRISPR Cas9

technique in this study was used to target the second coding exon, thus only disrupting the function of ZFY-long but not ZFY-short. This explains why the team did not observe any premeiotic phenotype in their knockout but makes it difficult to see how they arrived at the conclusion that disruption of both Zfy1 and Zfy2 has adverse effects on spermatocyte health. In spermatocytes, Zfy1 and Zfy2 have very distinct roles as Zfy2 rather than Zfy1 is the one necessary for the efficient apoptotic elimination of spermatocytes with univalent X chromosome during the first meiotic metaphase checkpoint (Decarpentrie *et al.*, 2012) (Nadge Vernet *et al.*, 2011). The functional differences between Zfy1 and Zfy2 may be the result of Zfy exon 6 which encodes for the acidic domain required effective transactivation activity of Zfy2 (Decarpentrie *et al.*, 2012). It is therefore hypothesised that for apoptotic elimination at the first meiotic metaphase checkpoint to occur, Zfy protein with an acidic domain that is able to recruit TRRAP-containing complexes to the chromatin (Decarpentrie *et al.*, 2012).

A lot of our understanding of *ZFY* gene has come from substantial amount of mouse studies. However, the inability to disrupt Y gene functions and lack of genetic sequence of the mouse Y chromosome initially impeded the research on uncovering roles of mouse Y-encoded genes on spermatogenesis (Vernet *et al.*, 2014). With the emergence of the Y transgene rescue strategy, whereby normal Y chromosome genes are added to Y deletion variants whose spermatogenic defects are known, this field of research was immense again (Vernet *et al.*, 2014). This strategy has since been used to study mouse *Zfy* genes with several mouse models having been designed to study Zfy.

1.5.3 Functions of Zfy1 and Zfy2 in Spermiogenesis

Gametogenesis is an extremely important mechanism for the survival, progression and evolution of any species and as such, is subject to extremely high levels of quality control and regulation (Nadége Vernet *et al.*, 2016). The male spermiogenesis has a more stringent and harsher quality control mechanism and as a result, spermiogenesis consequently produces very little aneuploidies in comparison to the female oogenesis (Nadége Vernet *et al.*, 2016)

(Nagaoka, Hassold and Hunt, 2012) (Hunt and Hassold, 2002). *Zfy* genes have been reported to be the reason for the highly efficient spermiogenesis quality control as a result of Zfy playing a major role in the efficient apoptotic aneuploidy elimination mechanism in the spermiogenesis (Royo *et al.*, 2010) (Vernet *et al.*, 2014).

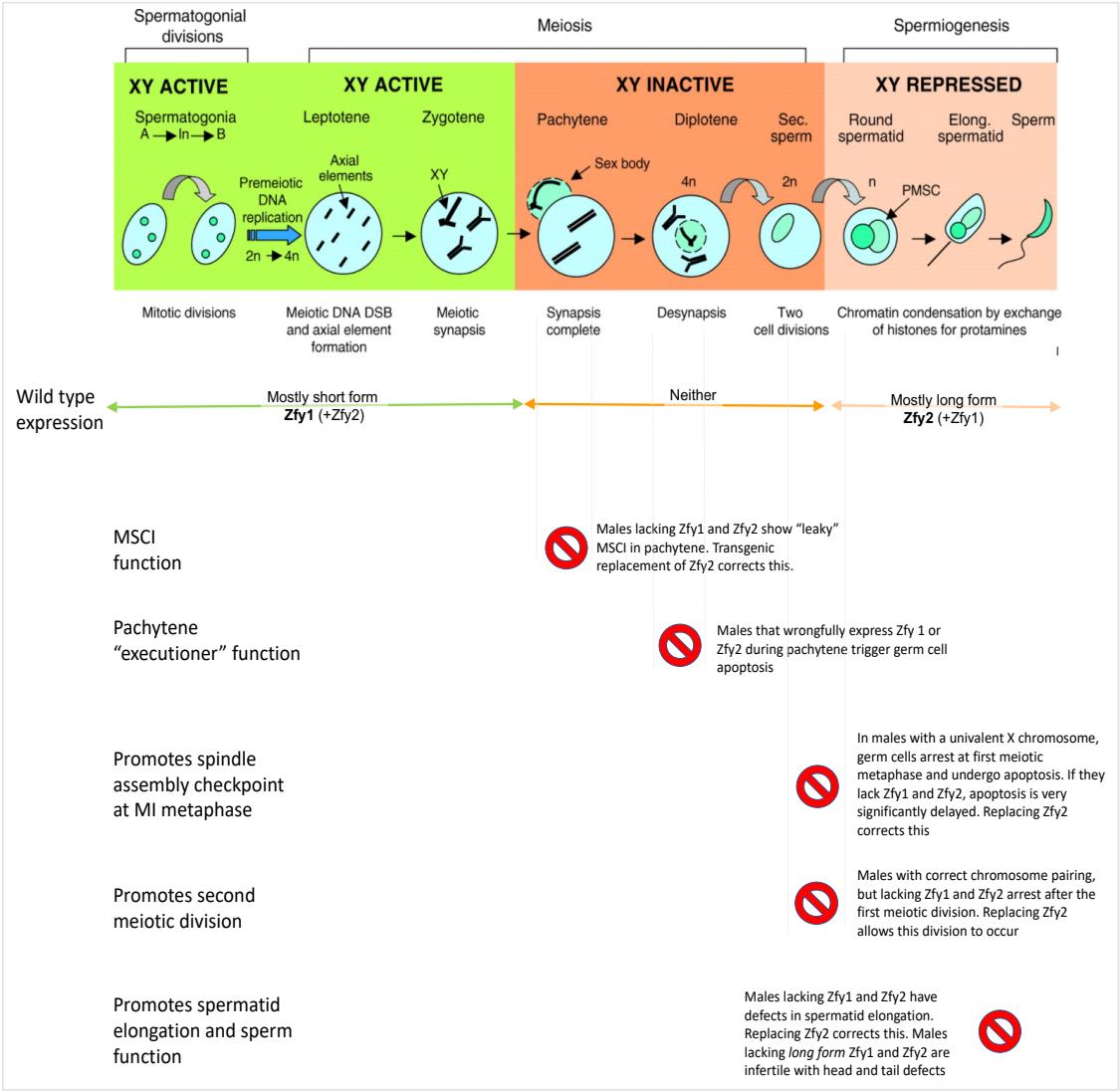


Figure 13: ZFY roles during spermatogenesis. (Cloutier and Turner, 2010)(Decarpentrie *et al.*, 2012)(Nadége Vernet *et al.*, 2016) (Royo *et al.*, 2010) (Nadge Vernet *et al.*, 2011)(Vernet *et al.*, 2014)(Nadege Vernet *et al.*, 2016) (Nakasuji *et al.*, 2017).

As a consequence of either being prone to recombination errors or/and lack of an efficient mechanism able to efficiently eliminate recombination errors like in male spermiogenesis, female oogenesis is more prone to aneuploidies than

spermiogenesis (Nadége Vernet *et al.*, 2016). Huge advancements in scientific technology has led to improved ability to examine spermatocyte and oocyte crossover associated proteins, it has been reported that almost all chromosomes in spermatocytes have at least a single crossover whilst in oocytes (Nadége Vernet *et al.*, 2016) (Lynn *et al.*, 2002). Consequently, the lack of an efficient apoptotic elimination system during oogenesis means that more aneuploidies, with about half of aneuploidies estimated to result from with more than 10% of oocytes carrying no less than a single 'crossover-less' bivalent (Nadége Vernet *et al.*, 2016) (Cheng *et al.*, 2009).

Synaptic defects during spermiogenesis are not tolerated both in mice and humans and as such almost always lead to spermatocyte elimination at the pachytene stage or first meiotic metaphase (Nagaoka, Hassold and Hunt, 2012). Females are able to retain fertility despite multiple genetic mutations which in male gametes cause meiotic arrest and gamete apoptosis and sterility (Nagaoka, Hassold and Hunt, 2012). Mouse studies have shown that synaptic defects that lead to meiotic blockade and arrest in spermiogenesis, such as chromosomal translocations and meiotic mutations, do not hinder the progression of meiosis of all oocytes (Nagaoka, Hassold and Hunt, 2012). However, this comes at a cost as most females who undergo this have a limited reproductive lifespan (Nagaoka, Hassold and Hunt, 2012). Chromosomal synapsis in male gamete meiosis occurs with the small region of the pseudoautosomal region, with un-synapsed chromosomes having their transcription completely suppressed at the spermatocyte pachytene stage (Turner, 2007) (Burgoyne, Mahadevaiah and Turner, 2009), this is the meiotic sex chromosome inactivation (MSCI) (Nagaoka, Hassold and Hunt, 2012). MSCI occurs by various mechanisms, however, the inactivation of the sex chromosome is a highly conserved mechanism in heterogametic animals (Wojtasz *et al.*, 2012) (33). MSCI is a male exclusive mechanism that is crucial for male fertility: explains the differences observed in male and female gametogenesis as a result of unsynapsed chromatin (Nagaoka, Hassold and Hunt, 2012).

This difference is attributed to the apoptotic machinery operating in developing gametes that with univalent chromosomes, an exclusively male meiosis feature

(Nadége Vernet *et al.*, 2016) (Kurahashi *et al.*, 2012) (LeMaire-Adkins, Radke and Hunt, 1997): this apoptotic elimination is regulated at 2 checkpoints, with one operating during mid-pachytene, monitoring homologous chromosome synapsis as well as meiotic sex chromosome inactivation (Barchi *et al.*, 2005) and the other spindle assembly checkpoint (SAC) operating at metaphase of the first meiotic division (Gorbsky, 2015). *Zfy* genes have been implicated in the regulation of this apoptotic machinery (Royo *et al.*, 2010) (Vernet *et al.*, 2014).

1.5.4 ZFY regulatory roles in the development of sperm cells

The apoptotic elimination of unsynapsed germ cells has 2 separate checkpoints, the first one which overlooks gametogenesis at both the synapsis of homologous chromosomes and the MSCI (Burgoyne, Mahadevaiah and Turner, 2009), and the second checkpoint has operations at first meiotic metaphase on the SAC (Gorbsky, 2015).

During pachytene, protein coding genes including *Zfy* located on the non-pairing regions of sex chromosomes are periodically silenced by the MSCI in order to allow for meiotic progression without errors (Nadége Vernet *et al.*, 2016) (Monesi, 1965). *Zfy* or *Zfx* autosomal transgenes have been shown in mice to be able to evade MSCI and present during pachytene as a result of ectopic expression. The presence of Zfy1 and Zfy2 at the pachytene is shown to be associated with mass cellular apoptotic elimination (Nadége Vernet *et al.*, 2016).

Silencing of *Zfy1* and *Zfy2* is crucial for progression of germ cells beyond pachytene as displayed in research models like XYY males, *H2afx -/-, Hormad2 -/-* Nadége Vernet *et al.*, 2016). Apoptotic elimination of germ cells was reported in these cells at the tubule stage that is associated with germ cell apoptotic elimination in cells with compromised MSCI or Y chromosome function (Royo *et al.*, 2010) (Fernandez-Capetillo *et al.*, 2003) (Wojtasz *et al.*, 2012), thus showing the essentialness of *Zfy* genes silencing for the development of germ cells past the mid-pachytene checkpoint. Nadége Vernet *et al.*, 2016). Zfy proteins play important executioner roles at the MSCI checkpoint, with abberent expression of either Zfy protein leads to apoptotic elimination of germ cells (Burgoyne,

Mahadevaiah and Turner, 2009) (Turner, 2015). Furthermore, they are predicted to have dual detection capability to detect for MSCI success or failure. It is thought that because of their genomic location on the Y chromosome, *Zfy* genes reported to likely control their own expression enabling them to exert their own negative feedback sensory role (via carefully inhibiting their own expression) that regulates germ cells transition at the mid pachytene checkpoint (Nadége Vernet *et al.*, 2016). In essence, Zfy proteins have a regulatory role at the start of MSCI, the progression and development at MSCI and elimination of cells that are not able to achieve MSCI.

Zfy2 has an extensive regulatory role at the first meiotic metaphase. The developmental direction of arrested spermatocytes with a univalent X chromosome ($X^{E}O$,*Sry)* at the first meiotic metaphase is determined by whether Zfy2 protein is present or absent (Nadége Vernet *et al.*, 2016). Zfy2 absence leads to the univalent X chromosome spermatocytes recovery and progression past the first meiotic division (Nadége Vernet *et al.*, 2016), however, due the X chromosome univalence, these germ line cells arrest as interphasic secondary spermatocytes, eventually undergoing apoptosis (Nadége Vernet *et al.*, 2016). The situation is different when Zfy2 is present as the univalent X chromosome germ line cells drive the cells into arrest at the meiotic metaphase I (MI SAC), followed by apoptosis (Nadège Vernet *et al.*, 2011) (Nadége Vernet *et al.*, 2016). The mechanism by which Zfy2 regulates the MI SAC is still to be elucidated but the research so far shows that Zfy2 acts as a regulator for germ cell progression at the MI SAC.

To summarise the roles of ZFY:

- ZFY is predicted to have several roles to the proper functioning of spermatogenesis, with its roles out of testis almost completely unknown.
- *ZFY* gene expression shows splicing shifts during the development of germ cells, with the testis-specific *ZFY-short* dominantly expressed in the high mitotically proliferating spermatogonia cells and the splicing pattern

shifting to dominant expression of *ZFY-long* in post mitotic cells (Decarpentrie *et al.*, 2012).

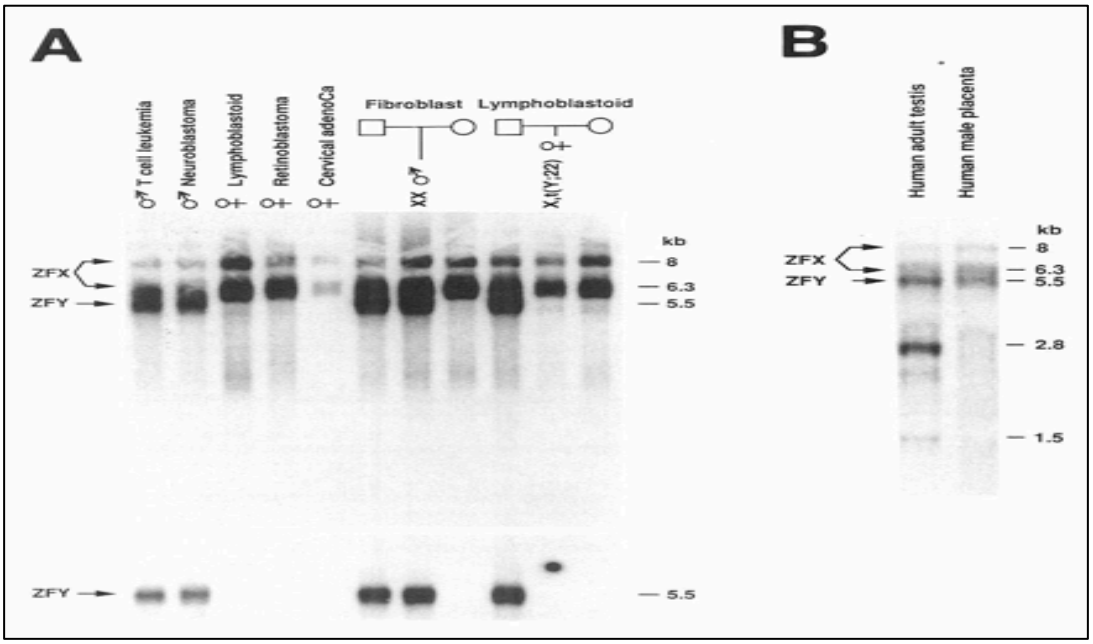
- The evidence from mouse Zfy studies suggests that the long form (preferentially produced by ZFY2) has a pro-apoptotic role at two different checkpoints (Nadége Vernet *et al.*, 2016).
- The roles of ZFY-short (preferentially produced by Zfy1) are still yet to be determined, we can predict based on what we know of ZFY-long expression and its functions that ZFY short form is expressed dominantly in high mitotically active cells and may potentially have anti-apoptotic function via the repression of pro-apoptotic capabilities of ZFY long form (possibly repressing pro-apoptotic genes activated by the long form). potentially has an anti-apoptotic function (by repressing

1.5.5 ZFY expression in tumours.

Because of the nature of Zfy protein structure and predicted role as a transcription regulator, the potential of ZFY protein having a role in malignancy has been explored. One study analysed samples of human male specific cancer of prostate adenocarcinoma for expression of both human ZFY and its X chromosome homologue ZFX (Tricoli and Bracken, 1993). Northern blot showed that 3 out of 31 prostate adenocarcinomas expressed *ZFY* transcripts. However, when reverse transcriptase PCR (RT-PCR) was utilised to detect *ZFY* transcripts in the 31 prostate adenocarcinoma samples, 19 out of 31 samples were found to be expressing *ZFY* transcripts, with 12 of the 19 exclusively expressing *ZFY* transcripts whilst the other 7 expressed both *ZFY* and *ZFX* (Tricoli and Bracken, 1993). Additionally, 6 benign prostatic hypertrophy (BPH) tissue samples were analysed for *ZFY* expression using RT-PCR. 2 of the 6 BPH samples showed weak, hardly detectable *ZFY* expression whilst the rest of the samples showed no expression of *ZFY* (Tricoli and Bracken, 1993).

ZFY genes are shown to be transcriptionally active in malignant prostate tumours. The activation of *ZFY* genes in these tumours may possibly be as a result of the

malignancy, or perhaps ZFY protein is actively playing a role in the development and progression of the tumour, either as an oncogene itself or indirectly activated by an oncogene (Tricoli and Bracken, 1993).



(Lau and Chan, 1989)

Figure 14: Northern blot image of *ZFY* gene hybridised B-lymphoblastoid, leukemic T cells and neuroblastoma male cell lines and female fibroblasts, B-lymphoblastoid, retinoblastoma, cervical carcinoma, and osteosarcoma cell lines. (Lau and Chan, 1989)

Importantly, expression of *ZFY* in other tumours and tumour derived cell lines have been shown in other studies (Fig. 14) (Lau and Chan, 1989) (Schneider-Gädicke *et al.*, 1989) and *ZFY* expression is a favourable prognostic marker for head and neck cancers as detailed on the protein atlas page (*Tissue expression of ZFY - Summary - The Human Protein Atlas*, no date). In line with this, ZFY is also hypothesised to have tumour suppressor ability based on analysis of mutational signatures from over 8200 tumour and normal tissue samples (Davoli *et al.*, 2013). As we predict ZFY long and short to have opposing roles, with the long form shown to be associated with pro-apoptotic genes in post-mitotic cells (Nadége Vernet *et al.*, 2016), it is critical to understand both the expression levels of ZFY as well as the splicing patterns in cancer cells. Perhaps RBMY activity

could be the inactivation of male specific tumour suppressor activities to promote Head and neck cancers.

Although very little is known or has been researched on the role of ZFY in human cancers, however, ZFX, a member of the *ZFY* gene family and the X-linked homologue of *ZFY* has been linked by substantial amount of research and data to several cancers including laryngeal, HCC and prostate cancers. Although this is not the focus of this thesis, it does highlight the important roles regulated by this gene's family such as in regulating cell cycle progression and cell fate commitment, which can be altered to promote cancer development. *ZFX* has been described as a potent oncogene that is associated with the initiation and/or progression of various cancers including cancers of the prostate, breast, colorectal and gastric cancers, as well as renal cell carcinoma, glioma, gallbladder adenocarcinoma, non-small cell lung carcinoma, and laryngeal squamous cell carcinoma (Fang *et al.*, 2012a, 2014), (Zhou *et al.*, 2011; Nikpour *et al.*, 2012), (Yin *et al.*, 2015), (Li *et al.*, 2013), (Yang *et al.*, 2014; Weng *et al.*, 2015). In these cancers, ZFX is overexpressed and associated with the tumorigenesis of the cancers, and consequently the downregulation or suppression of *ZFX* is shown to directly influence the suppression of cellular proliferation and pro-apoptotic processes, highlighting the importance of ZFX to cancer establishment (Fang *et al.*, 2014), (Jiang and Liu, 2015), (Yang *et al.*, 2015), (Yan *et al.*, 2016). Furthermore, ZFX is a powerful oncogene of which its expression in these cancers is shown to drive a poor prognostic outcome for cancer patients (Jiang and Liu, 2015), (Li *et al.*, 2015), (Yang *et al.*, 2015), (Yan *et al.*, 2016). ZFX is also involved in the regulation of self-renewability of hematopoietic and embryonic stem cells, and thus has been suggested to potentially drive tumour stemness of laryngeal squamous cell carcinoma (Palmer *et al.*, 2014) (Yin *et al.*, 2015).

In squamous cell carcinoma, experiments with Tca-8113 and CAL-27 cell lines showed that *ZFX* knockdown cells were subjected to cell cycle arrest and reduced cellular proliferation activity (Yin *et al.*, 2015). Furthermore, experiments with these cells lines have yielded results that are consistent with other reports in which *ZFX* knockdown is associated with compromised cancer development, with

various cancer experiments *in-vivo* and *in-vitro* demonstrating that *ZFX* knockdown is associated with impaired cancer cellular growth, cellular proliferation and cancer cell stemness (Yin *et al.*, 2015).

ZFX is implicated in several molecular pathways that are classically activated in various cancers. In-vivo and in-vitro knockdown of *ZFX* is shown to inhibit the activating phosphorylation processes of the ERK1/2 and MEK1/2 oncogenic pathways. Furthermore, *ZFX* knockdown is shown to suppress expression of key cell cycle genes cyclin D1 and B1, thus, inducing cell cycle arrest (Zhu *et al.*, 2013) (Yang *et al.*, 2014) (Li *et al.*, 2013). ZFX expression promotes cancer development that is able to bypass apoptotic processes via the suppression of pro-apoptotic genes as demonstrated via *ZFX* knockdown inducing suppression of anti-apoptotic protein expression whilst enriching the expression of pro-apoptotic genes of *bax* and *caspases*. (Fang *et al.*, 2012b) (Jiang *et al.*, 2012) (Wu *et al.*, 2013).

Genome wide profiling of *ZFX* expression in various cancer cell lines (cell lines acquired from kidney cancer, colon cancer, prostate cancer and breast cancer) demonstrated that ZFX protein activity in these cancers occurs via ZFX binding and potentially activating promoter regions of CpG islands as the knockdown of *ZFX* in these cell lines is associated with significant down regulation of the activated promoters of CpG islands (Rhie *et al.*, 2018). A hypothesis was presented to explain ZFX association with CpG island promoters, the hypothesis suggests that based on the analysis of genome wide profiling data, ZFX is predicted to act as a transcription activator with a regulatory role on these CpG promoter regions (Rhie *et al.*, 2018).

Taken together, ZFX is a key oncogene that may act to promote cancer developments via several different pathways and perhaps this could be extrapolated and suggestive of possible roles that ZFY is playing in male cancers (Rhie *et al.*, 2018). It is worth noting that in laryngeal cancer patients, ZFX expression was seen to have a higher male prevalence than in females, perhaps alluding to a link between the ZFX cancer upregulation and sex (Yin *et al.*, 2015).

Cancer expression studies with ZFY and ZFX highlight that these proteins potentially have important regulatory roles that can be altered to promote cancer development. We that RBMY, a male germ line specific splicing regulator is ectopically expressed in male HNSCC cancer tissues, alters the splicing of ZFY to give ZFY-short, consequently promoting male specific cancer inducing mechanisms.

2. MATERIALS AND METHODS

2.1 Tissue Culture

University of Michigan-Squamous Cell Carcinoma-104 (UM-SCC-104), a HNSCC (oral cavity tumour) cell line that is HPV16 positive, originating from a 56 year old male (Tang *et al.*, 2012), which was obtained from Dr Tom Carey, University of Michigan, USA(Tang *et al.*, 2012). 93VU147T, also HPV16+ HNSCC cell line which was obtained from Peter Snijders, Free University Medical Centre Amsterdam, Netherlands (Busch *et al.*, 2013). These cell lines were grown in a CO_2 incubator at 37°C and 5% CO_2 until 70% confluency, at which point the cells would be split for passaging (every 3 days). The growth medium used to maintain the cell lines was Dulbecco's Modified Eagle Medium (DMEM) containing 1% L-Glutamine (PAN Biotech), supplemented with 10% Bovine serum and 1% Penicillin/streptomycin to make up the total growth medium. Cells were maintained in t75 growth flask with 8ml growth medium.

At 70-80% confluency, cells were harvested after washing twice (washing once if cells are just being split) with phosphate buffered saline (PBS), followed by trypsinising cells by adding 2ml 0.25% trypsin-EDTA for 5 minutes in 37°C incubator to detach cells from the culturing flask. 6ml of DMEM was added to neutralise the trypsin and the cells are transferred and centrifuged in a 15ml falcon tube at approximately 300-500g (1200-1500g max on a table-top swinging bucket centrifuge rotors) for 5 minutes at 4°C. Supernatant was removed and cell pellet stored for future use at -80°C. If splitting cells, supernatant was removed after spinning and fresh DMEM added to resuspend the pellet in the falcon tube, fraction of the cells is pipetted into a fresh DMEM media in a tissue culture to make up 8ml total volume.

2.2 RNA extraction

At 70-80% confluency, cells were harvested after washing the cells twice with phosphate buffered saline (PBS), followed by tryptonising cells by adding 2ml 0.25% trypsin-EDTA for 5 minutes in 37°C incubator to dissociate adherence of cells from the culturing flask. Cells were centrifuged at approximately 300-500g for 5 minutes, at 4°C and the supernatant removed. The pellet was rewashed with PBS twice and pelleted again. RNA extraction was done using the New England Biolabs Inc. Monarch® Total RNA Miniprep Kit. Cells were resuspended in RNA lysis buffer according to the volumes suggested by the kit protocol. The mixture was transferred to a gDNA removal column and centrifuged (all centrifugal steps in this are done at 16000xg) and flow through collected. Equal volume ethanol (>95%) was added to the flow-through and mixed thoroughly. The mixture was transferred into an RNA purification column and spun for 30 seconds to remove flow-through. RNA wash buffer was pipetted onto the purification column and column spun for 30 seconds. DNAse 1 reaction mixture (5µl DNase I reconstituted enzyme with 75µl DNase I Reaction Buffer) was then pipetted onto the purification column and incubated for 15 minutes at room temperature. The purification column then washed by 500µl RNA Priming Buffer, centrifuged for 30 seconds and flow-through discarded. This RNA priming buffer wash was repeated. This was followed by addition of 500µl RNA Wash Buffer to the purification column, column was spun for 30 seconds and flow-through is discarded. This RNA wash buffer step was then repeated and centrifuged for 2 minutes. Purification column then transferred to an RNase-free microfuge tube and 50µl nuclease-free water was pipetted directly to the centre of column matrix and spin for 30 seconds, collecting RNA in the water(*Monarch ® Nucleic Acid Purification*, no date). RNA was stored at -20°C for short -term storage (less than 1 week), or at -80°C for long-term storage. RNA was quantified using a Nanodrop.

2.3 Protein extraction

At 70-80% confluency, cells were harvested after washing the cells twice with PBS. This was followed by the trypsinisation of cells by adding 2ml 0.25% trypsin-EDTA for 5 minutes in 37°C incubator to dissociate adherence of cells from the culturing flask. Cells were centrifuged at approximately 300-500g for 5 minutes, at 4°C and the supernatant removed. The pellet was rewashed with PBS twice and pelleted again and put on ice. The cells were resuspended in a lysis buffer (30µl) containing 1x protease inhibitor cocktail (ThermoFisher), 1x phosphatase inhibitor cocktail (ThermoFisher) and benzonase (0.8µl/ml). This mixture was incubated for 15 minutes on ice and spun in a cold centrifuge for 20 minutes at 4°C. The supernatant was pippeted into a clean Eppendorf and stored at -20°C. Protein is quantified using BSA method.

2.4 PCR

Table 3: Reverse Transcription PCR primers.

Target	Sequence	Product Size (bp)	Annealing temperature	Relevant information
RBMY forward	5'-GAAACCAATGAGAAGATGCTT-3'	473 (spanning exon 2 to exon 5)	53°C	As *RBMY* family of genes are clustered on the Y chromosome, the major products from several *RBMY* genes in 473bp.
RBMY reverse	5'-TTGCTTCTTGCCACAGCAG-3'			

				However, they may occur minor products of 459bp and 295bp, all of which are *RBMY* family transcripts.
ZFY-long forward	5'-GAATTGCAGCCACAAGAGCC-3'	729 (Spanning exon 2 to exon 4)	59°C	As the ZFY-long reverse primers are the same as ZFY-short primers, some ZFY-short bands may appear on the gel.
ZFY-long reverse	5'-CACCTTGATGACTTCAGGAC-3'			
ZFY-short forward	5'-GATGGAATAGTGGATGATGC-3'	123 (Exon 2 -4 splice junction)	53°C	
ZFY-short reverse	5'-CACCTTGATGACTTCAGGAC-3'			

RNA was reverse transcribed into cDNA using the LunaScript™ RT SuperMix Kit. This was done by mixing the following components in Table 4.

Table 4: Reaction components and measurements for reverse transcription.

Components	20µl REACTION	Final concentration
LunaScript RT SuperMix (5X)	4 µl	1X
RNA Sample	Variable	(up to 1 µg) *
Nuclease-free Water	Make up to 20 µl	

The reactions were boiled in a PCR machine following a single cycle program in Table 5.

Table 5: PCR cycle steps.

Cycle Step	Temperature (°C)	Time (minutes)
Primer Annealing	25	2
cDNA Synthesis	55	10
Heat Activation	95	1

A Taq polymerase PCR mastermix was prepared following Table 6 instructions. Each PCR reaction had total volume of 10µl and cDNA concentration of 10ng. The reaction was run for 30 cycles.

Table 6: Components in a single PCR reaction.

Components	Volume (µL)
Water	5.46
10X Buffer	1
50mM $MgCl_2$	0.3
10mM dNTP Mix	0.2
10µM Forward Primer	0.5
10µM Reverse Primer	0.5
Taq Polymerase	0.04
cDNA (5ng)	2

Following the PCR, the reactions were run on a 1.5 Agarose gel electrophoresis set and visualised using a lightbox or G-box. DNA was dyed using 5% ethidium bromide.

2.5 SDS Page and Western Blot

Protein extracts were quantified using the BCA Assay (Antoniw *et al.*, 1978). The protein extracts were prepared with Dichlorodiphenyltrichloroethane (100mM), 4x Laemmli sample buffer (62.5 mM Tris-HCl, pH 6.8 sample buffer 10% glycerol, 1% LDS and 0.005% Bromophenol Blue) and heated at 70°C for 10-mins. The protein extracts were run on SDS Page on 30% polyacrylamide gels with Bio-rad SDS running buffer containing 25 mM Tris, 192 mM Glycine, 0.1% SDS (pH 8.3u). Electrophoresis was conducted at 100 volts for 90 minutes. After electrophoresis, the gel was extracted from the casting shell and transferred onto a Polyvinylidene Difluoride (PVDF) membrane. Blocking buffer containing Tris-buffered saline (TBS) with 0.1% Tween-20 and 5% BSA was used to block the PVDF membrane by incubating membrane for 60 minutes. PVDF membrane was washed with TBS with 0.1% Tween-20 (Tbst) 3 times and then incubated overnight with anti-rabbit (1:1000 skimmed milk). The PVDF membrane was then washed 4 times Tbst, with each wash after 15-minute intervals, then incubated for an hour and room temperature with HRP-conjugated secondary anti-rabbit GFP antibody (1:10,000), raised against the primary antibodies. The PVDF membrane was then washed for a final 4 times at 15-minute intervals. The PVDF membrane was finally incubated in enhanced luminol-based detection (ECL) reagent and visualised on Syngene G:Box-fluorescence imager.

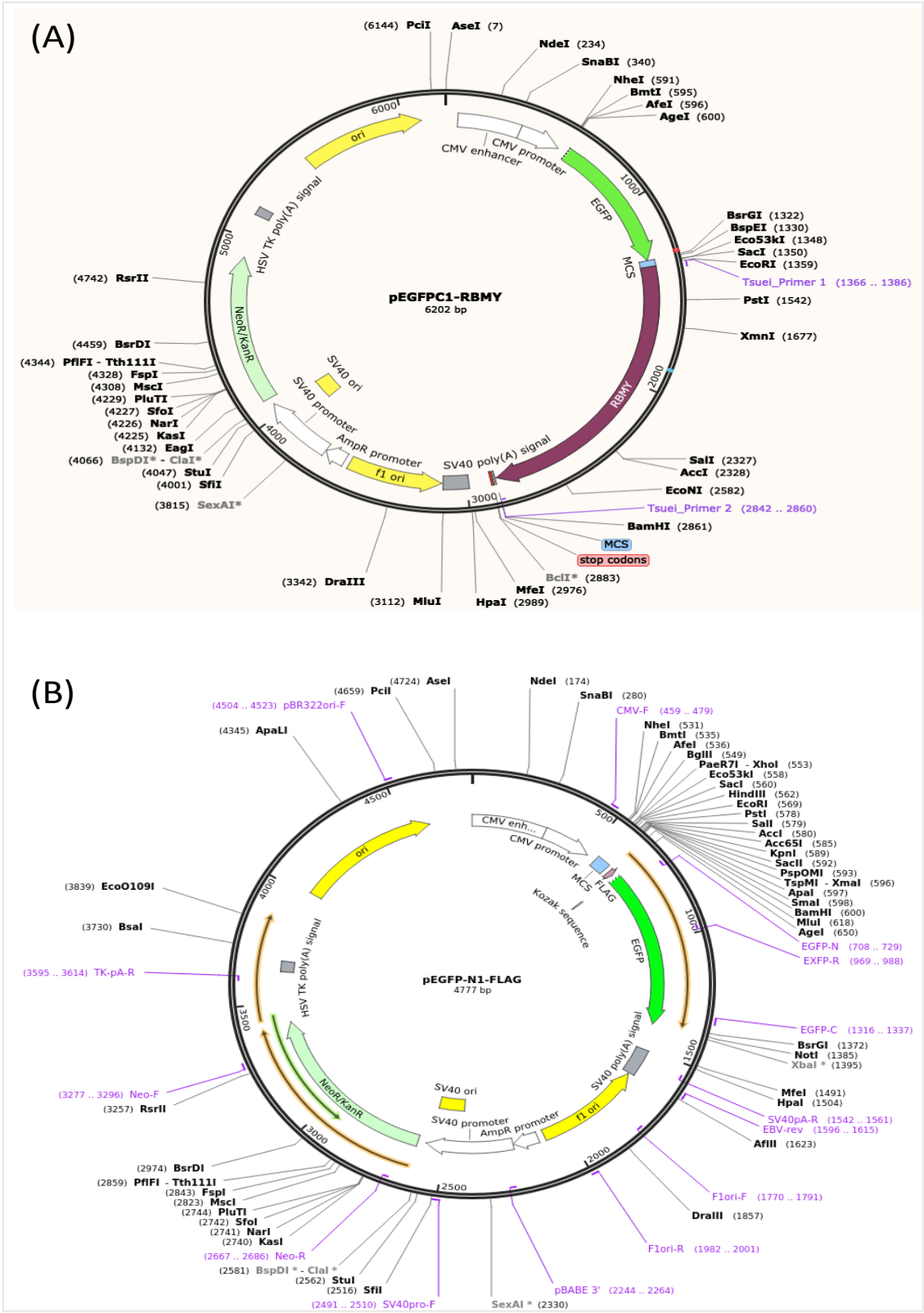


Figure 15: Transfection vector plasmids. (A) pEGFPC1-*RBMY* plasmid carrying GFP-*RBMY* fusion gene. (B) pEGFP-N1 plasmid.

2.6 Transfection Success Analysis

To assess successfully transfected cells in the cultures, cells grown on microscopy slides were collected 2 weeks post transfection and the slides developed. Successfully transfected UM-SCC-104 and 93-UV-147T PEGFP-N1 type cells have green fluorescence in the cytoplasm whilst UMSCC-104 and 93-UV-147T of GFP-RBMY type cells would show green fluorescence over the nuclear stained blue. High powered fluorescence microscopy images were obtained and using standard Microsoft PowerPoint, the image contrast was enhanced to 67%. A 24 x 18 (Fig.16) quadrat was overlayed onto the icture dividing the image into small squared which were utilised to count cells showing green florescence (nuclear and cytoplasm) and those only showing blue fluorescence for the nucleus. Percentages of cells with green fluorescence were calculated, mean average per cell type calculated and results displayed on a graph.

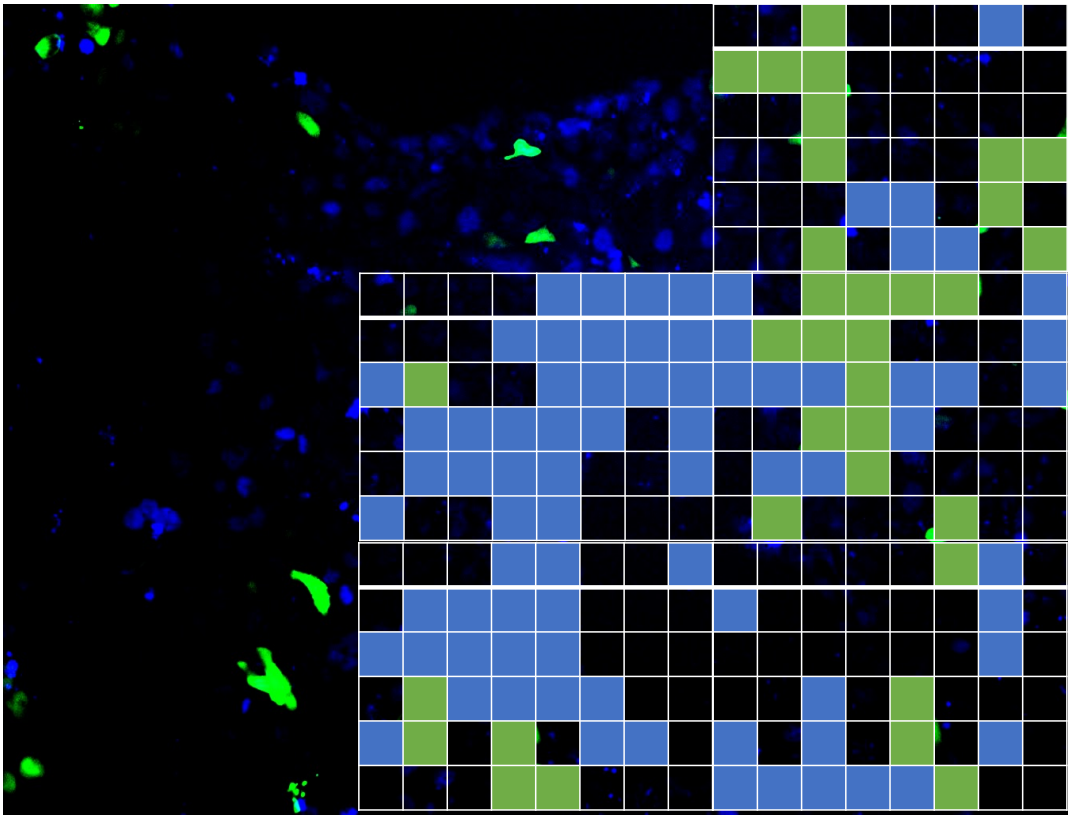


Figure 16: Image demonstrating the quadrat technique utilised to count successfully transfected cells. A 24 x 16 square quadrat was overlayed on the images and utilised to analyse colour in each image representing cell type.

2.7 *RBMY* expression in Cancer Computational analysis.

The computational analysis was conducted using 3 science research public domains, GTEx Portal (https://gtexportal.org/home/), cBioPortal (https://www.cBioPortal.org/) and TGCA Data from the National Cancer Institute (https://portal.gdc.cancer.gov) (data not shown as it corresponded with data from cBioPortal).

On GTEx, *RBMY* expression in normal tissues was queried for. A list of *RBMY* genes and pseudogenes appeared and their expression was analysed with the graphs of *RBMY* genes expression added into this thesis.

On cBioPortal, all protein coding *RBMY* genes were queried for against all the patient case files on the cBioPortal database, with relevant analytical data and graphs added into this thesis.

On the Database, a “Quick Search” was conducted. Figure 17 demonstrates Quick Search results for RBMY genes.

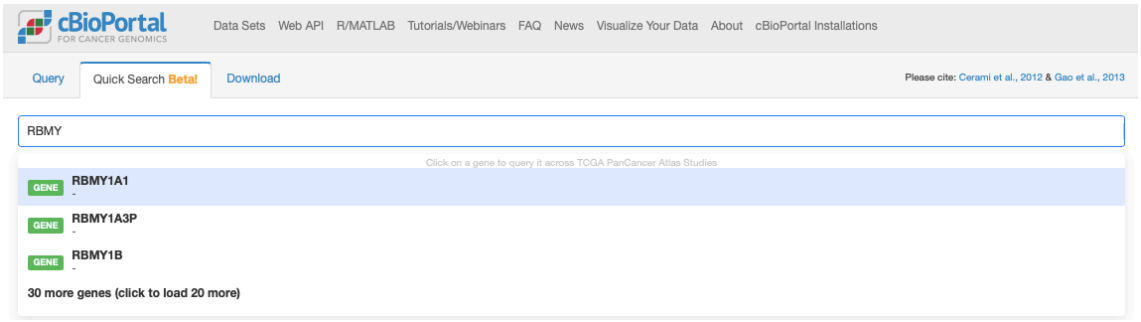


Figure 17: cBioPortal Quick Search displaying results for RBMY genes.

Once the data was queried, a new tab appears with several options (Fig.18), which includes OncoPrint; Cancer Type Summary; Plots; Mutations; Comparison/Survival; CN Segments; Pathways; and Download.

Figure 18: Query Tabs for data analysis maps and graphs on cBioPortal. Depending on the data available from studies, the tabs available will differ as some data may be vast enough to allow for more data analytical methods.

The Cancer type Summary was analysed; Plots of Data were analysed looking at the relationship between mRNA expression vs Mutation type and mRNA expression vs Cancer tissue. Parameters demonstrated in the figure 19.

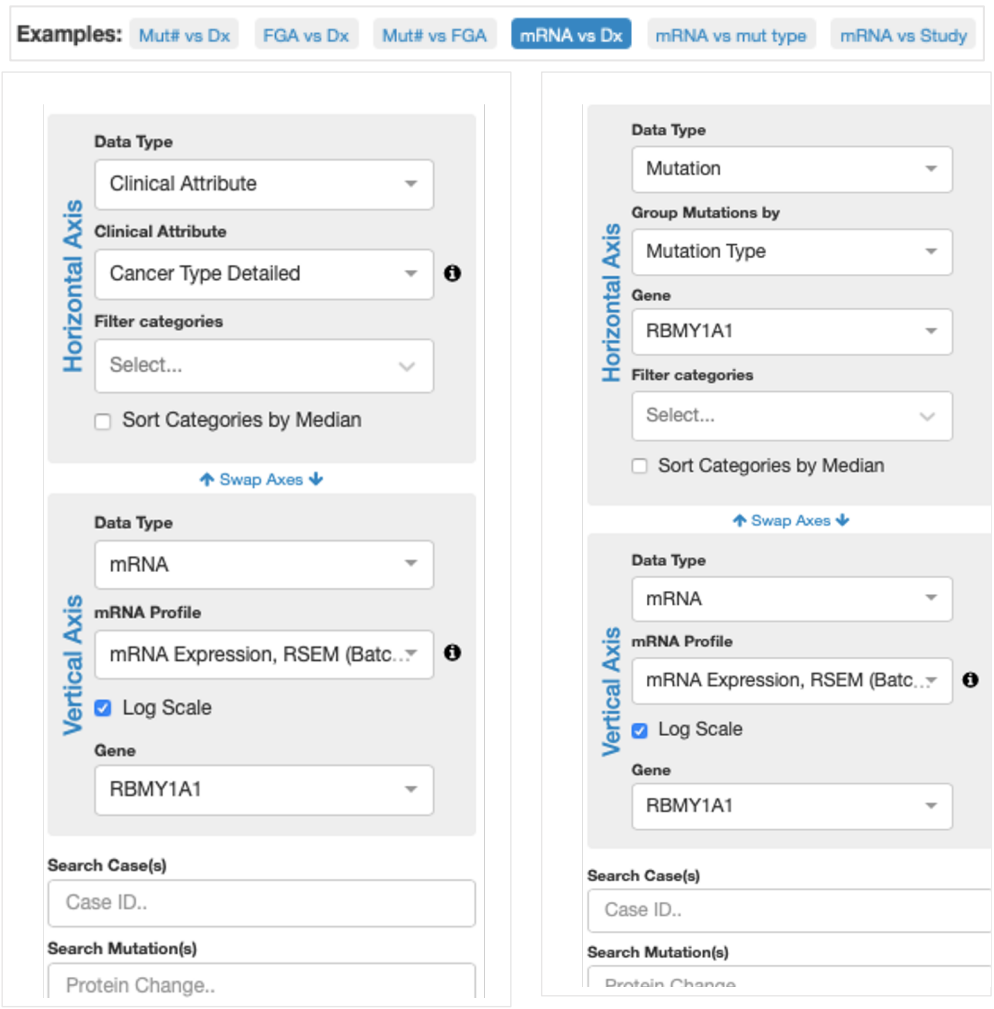


Figure 19: cBioPortal Data parameters utilised obtain cancer type summary showing mRNA expression vs Cancer tissue and mRNA expression vs mutation type. A. Example bar for plots available for given data, B. mRNA vs Dx plot parameters used for RBMY genes in this thesis and C. mRNA vs mut type parameters utilised to analyse data in this thesis.

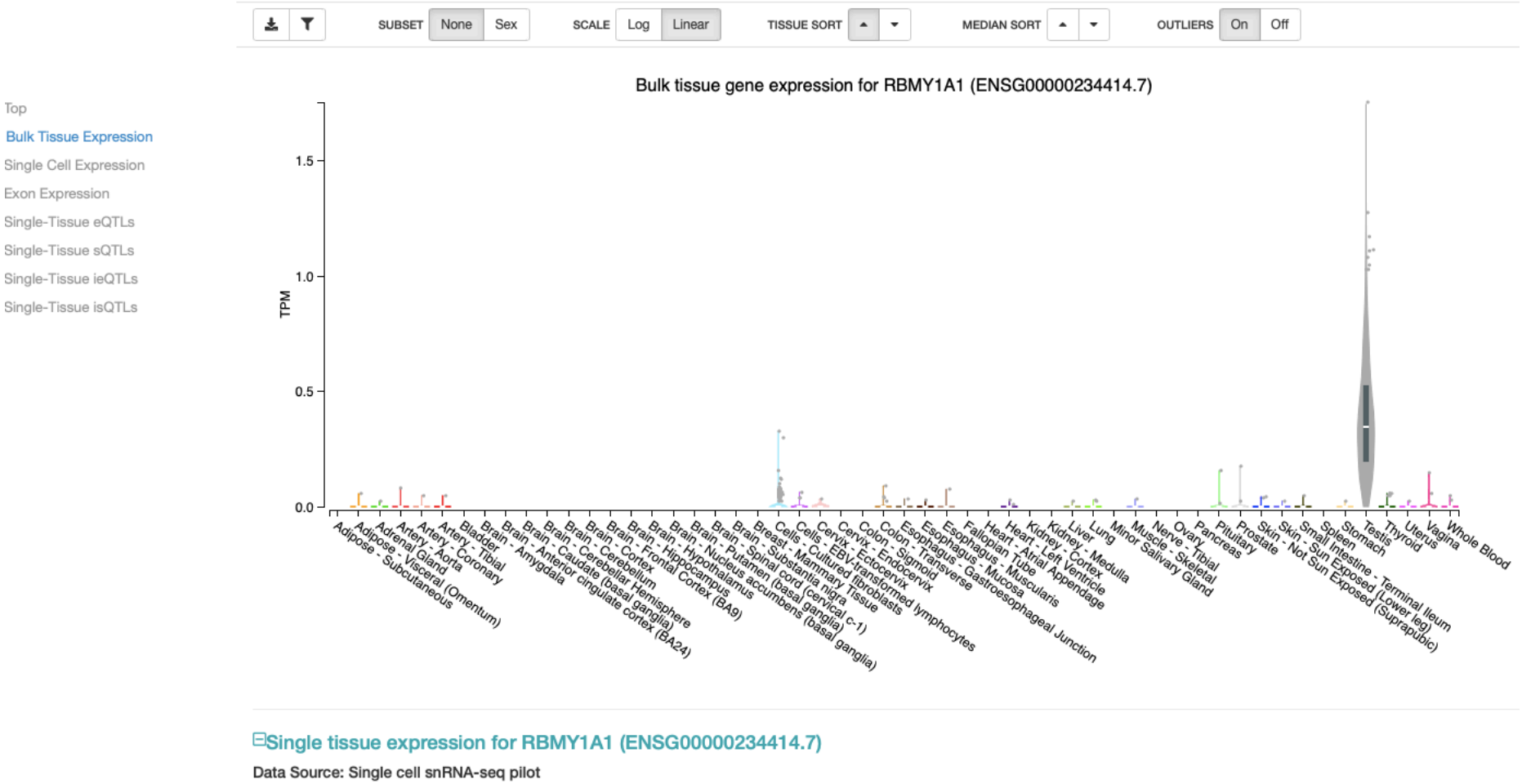
Furthermore, the Mutations plot were assessed to analyse if the RBMY genes were accumulating driver mutations, which as a result might be driving HNSCC in males. Default parameters were used in the data set.

3. **RESULTS**

3.1 RBMY expression in Cancers: A general cancer survey

As aforementioned, there has been a huge increase in the interest of the possible role that the Y chromosome may play in the differences in disease incidence and disease state between male and females. The demonstration that Y linked genes like *RBMY* and TSPY, are overexpressed in some cancers arising from somatic cells has really opened up a new space for research (Kido and Lau, 2015). As such, we have become interested in RBMY, a recently described novel male specific oncogene that may explain the high prevalence of hepatocellular carcinoma in males. We hypothesise that RBMY expression in HNSCC is the driving factor triggering male specific oncogenic mechanisms that drive the HNSCC high male prevalence. In line with this hypothesis, we are also interested in doing a general survey of the RBMY expression in cancers, to explore the behaviour of RBMY ectopic expression in cancers.

To explore the expression of *RBMY* in cancer, we want to understand which cancer tissues *RBMY* is expressed in and whether *RBMY* is expressed in those cancers that display a higher incidence in males. Furthermore, we would like to explore if *RBMY* is highly mutated in cancers, investigate the proportion of cancers in which *RBMY* mutations have been detected in, and whether *RBMY* acquires activating mutations in cancers.



(Gtex portal)

Figure 20. ***RBMY* expression in normal tissues.** The graphs from GTex public resource site showing expression of the functional *RBMY* genes in normal tissues. The graphs show that *RBMY* expression is restricted to the testes' only. This is the result of *RBMY* genes being Y chromosome linked and male specific roles.

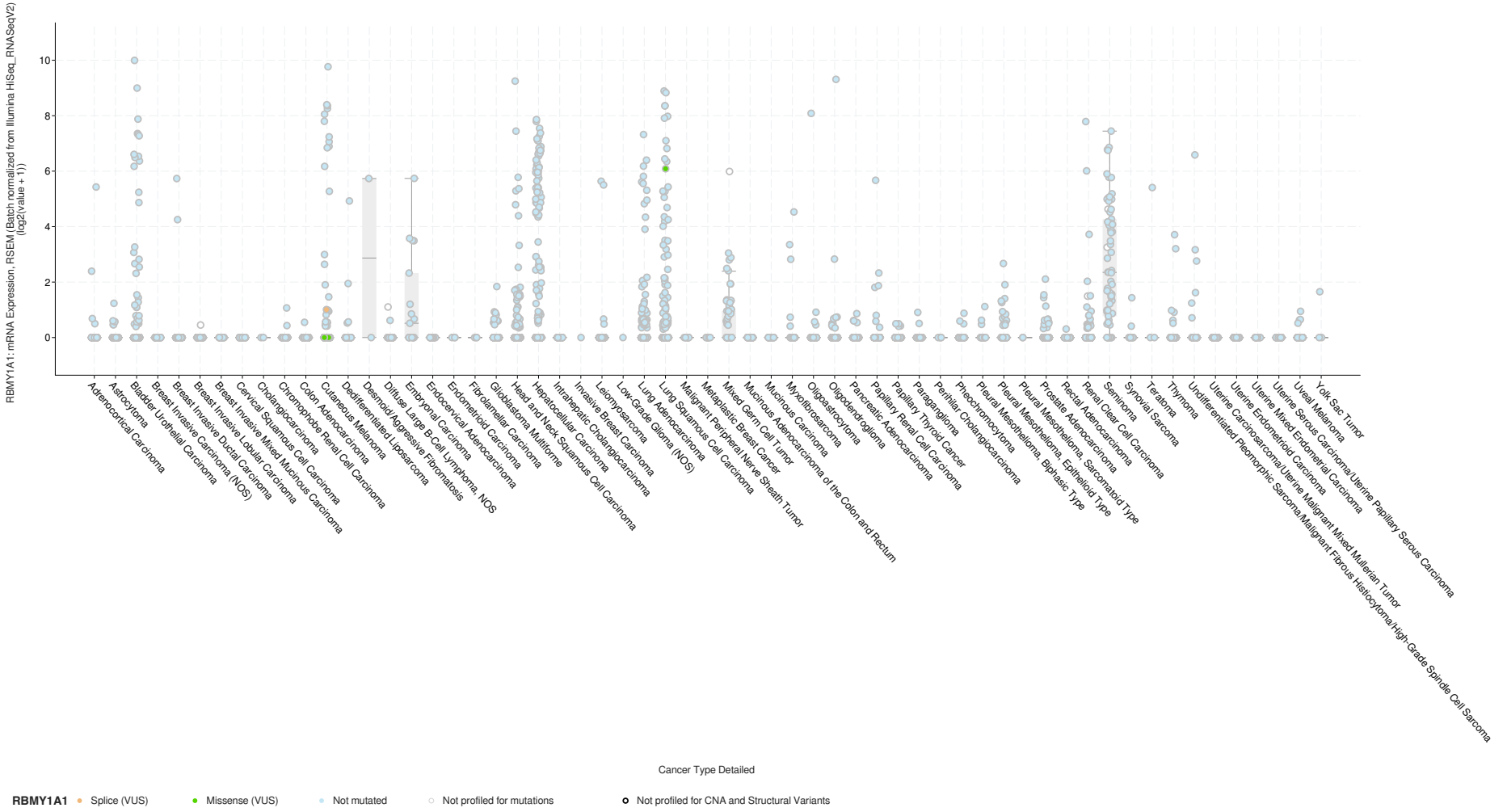
Normal expression of *RBMY* as demonstrated in figure 20, is restricted to the male germ line cells. The immediate problem faced when looking at *RBMY* expression is that *RBMY* is a family of genes with 6 predicted functional genes, therefore, at every step it is only possible to analyse a single *RBMY* gene and not as an entire group. Using data from Genotype-Tissue Expression (**GTEx**) public resource site, we are able to use RNA seq data to confirm that *RBMY* is exclusively expressed in the testis in normal tissues. Of the 6 predicted functional *RBMY* genes, *RBMY1A1*, *RBMY1B*, *RBMY1D*, *RBMY1E*, *RBMY1F, RBMY1J*, the data found on Gtex public domain show expression (via RNA Seq) of *RBMY1A1*, *RBMY1E*, *RBMY1F* and *RBMY1J*, all displaying an exclusive expression in the testes in graphs same as figure 20.

Using cBioPortal, a large-scale Cancer Genomics data set public resource site, we studied the expression of *RBMY* genes in different cancers. The data on cBioPortal showed *RBMY* cancer expression for *RBMY1A1*, *RBMY1B*, *RBMY1E*, *RBMY1F* and *RBMY1J*. This data from cBioPortal is acquired from a combination of studies and collected from 10967 samples of cancer patients covering a wide range of different cancer types.

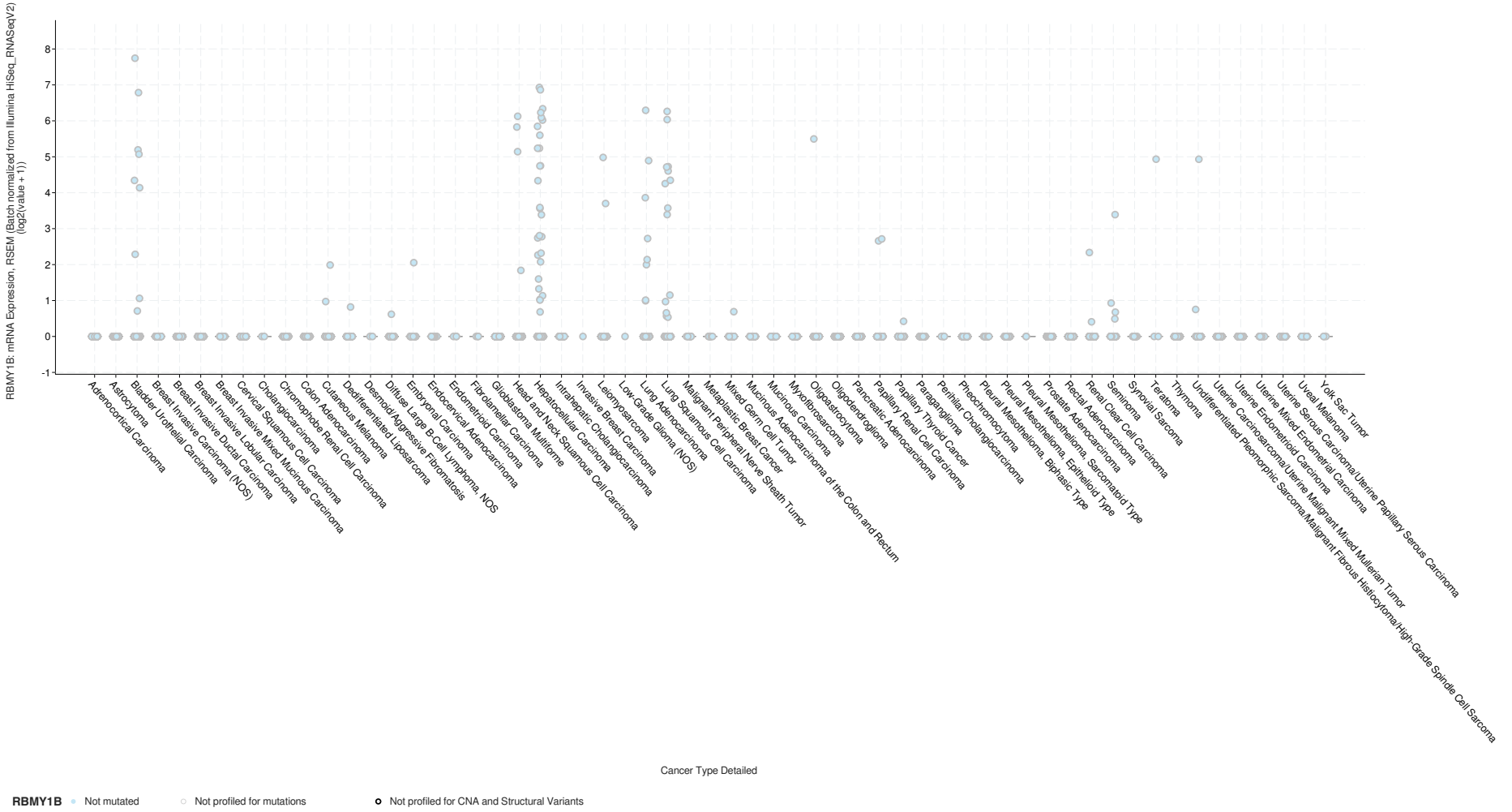
Figure 21 shows *RBMY* genes expression across various cancers. There is a variation in the expression of *RBMY* genes in different cancers, with *RBMY1A1* being the most dominantly expressed of the *RBMY* genes in all the cancers based on RNA seq V2 (Fig. 21). Compared to *RBMY1A1* expression in cancers, *RBMY1B* and *RBMY1E* are expressed at about 1/3 the expression of *RBMY1A1* and *RBMY1F* and *RBMY1E* are the least expressed in all cancers, with none of these genes expressed above 50 RSEM RNA seq mRNA values. Bladder Urothelial Carcinoma, Hepatocellular carcinoma, Lung Adenocarcinoma and Lung Squamous Cell Carcinoma are all substantially expressed in *RBMY1A1*, *RBMY1B*, *RBMY1E*, *RBMY1F* and *RBMY1J* (Fig. 21). With the exception of *RBMY1E*, the rest of the *RBMY* genes are expressed in HNSCC and with the exception of *RBMY1B*, all the other *RBMY* genes are expressed in Cutaneous Melanoma and Seminoma (Fig. 21). Overall, there is a pattern of all the *RBMY* genes being expressed in the same various cancers of Bladder Urothelial

Carcinoma, Hepatocellular carcinoma, Lung Adenocarcinoma, Lung Squamous Cell Carcinoma, Seminoma and Cutaneous Melanoma, albeit with varied expression with *RBMY1A1* having the highest and possibly most significant expression levels in these cancers.

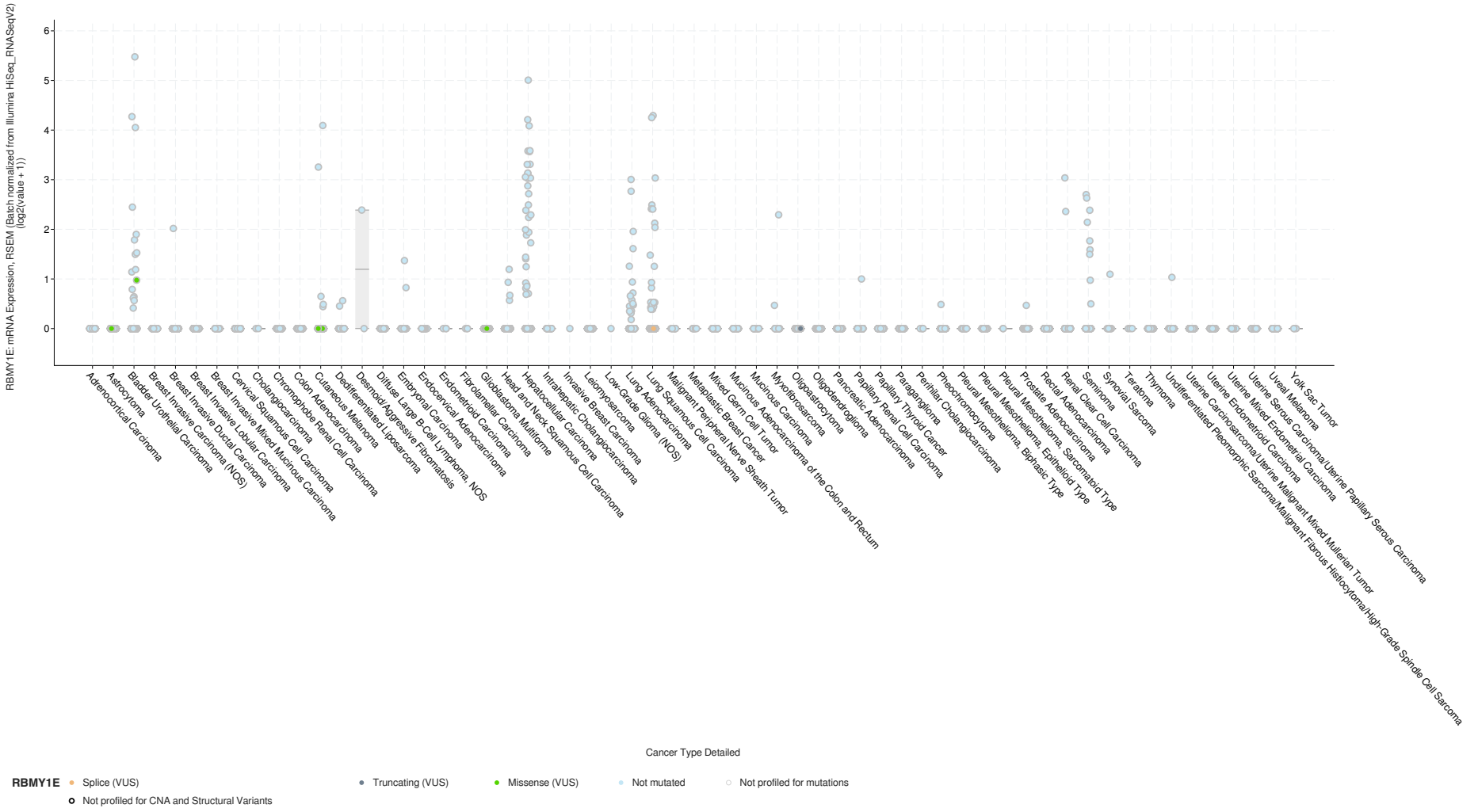
A. RBMY1A1


RBMY1A1: mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2) (log2(value + 1))
10
8
6
4
2
0
Adrenocortical Carcinoma
Astrocytoma
Bladder Urothelial Carcinoma
Breast Invasive Carcinoma (NOS)
Breast Invasive Ductal Carcinoma
Breast Invasive Lobular Carcinoma
Breast Invasive Mixed Mucinous Carcinoma
Cervical Squamous Cell Carcinoma
Cholangiocarcinoma
Chromophobe Renal Cell Carcinoma
Colon Adenocarcinoma
Cutaneous Melanoma
Dedifferentiated Liposarcoma
Desmoid/Aggressive Fibromatosis
Diffuse Large B-Cell Lymphoma, NOS
Embryonal Carcinoma
Endocervical Adenocarcinoma
Endometrioid Carcinoma
Fibrolamellar Carcinoma
Glioblastoma Multiforme
Head and Neck Squamous Cell Carcinoma
Hepatocellular Carcinoma
Intrahepatic Cholangiocarcinoma
Invasive Breast Carcinoma
Leiomyosarcoma
Low-Grade Glioma (NOS)
Lung Adenocarcinoma
Lung Squamous Cell Carcinoma
Malignant Peripheral Nerve Sheath Tumor
Metaplastic Breast Cancer
Mixed Germ Cell Tumor
Mucinous Adenocarcinoma
Mucinous Carcinoma
Myxofibrosarcoma
Oligoastrocytoma
Oligodendroglioma
Pancreatic Adenocarcinoma
Papillary Renal Cell Carcinoma
Papillary Thyroid Cancer
Paraganglioma
Perihilar Cholangiocarcinoma
Pheochromocytoma
Pleural Mesothelioma
Pleural Mesothelioma, Biphasic Type
Pleural Mesothelioma, Epithelioid Type
Prostate Adenocarcinoma
Rectal Adenocarcinoma, Sarcomatoid Type
Renal Clear Cell Carcinoma
Seminoma
Synovial Sarcoma
Teratoma
Thymoma
Undifferentiated Pleomorphic Sarcoma/Malignant Fibrous Histiocytoma/High-Grade Spindle Cell Sarcoma
Uterine Carcinosarcoma/Uterine Malignant Mixed Mullerian Tumor
Uterine Endometrioid Carcinoma
Uterine Mixed Endometrial Carcinoma
Uterine Serous Carcinoma/Uterine Papillary Serous Carcinoma
Uveal Melanoma
Yolk Sac Tumor
Cancer Type Detailed
RBMY1A1
Splice (VUS)
Missense (VUS)
Not mutated
Not profiled for mutations
Not profiled for CNA and Structural Variants

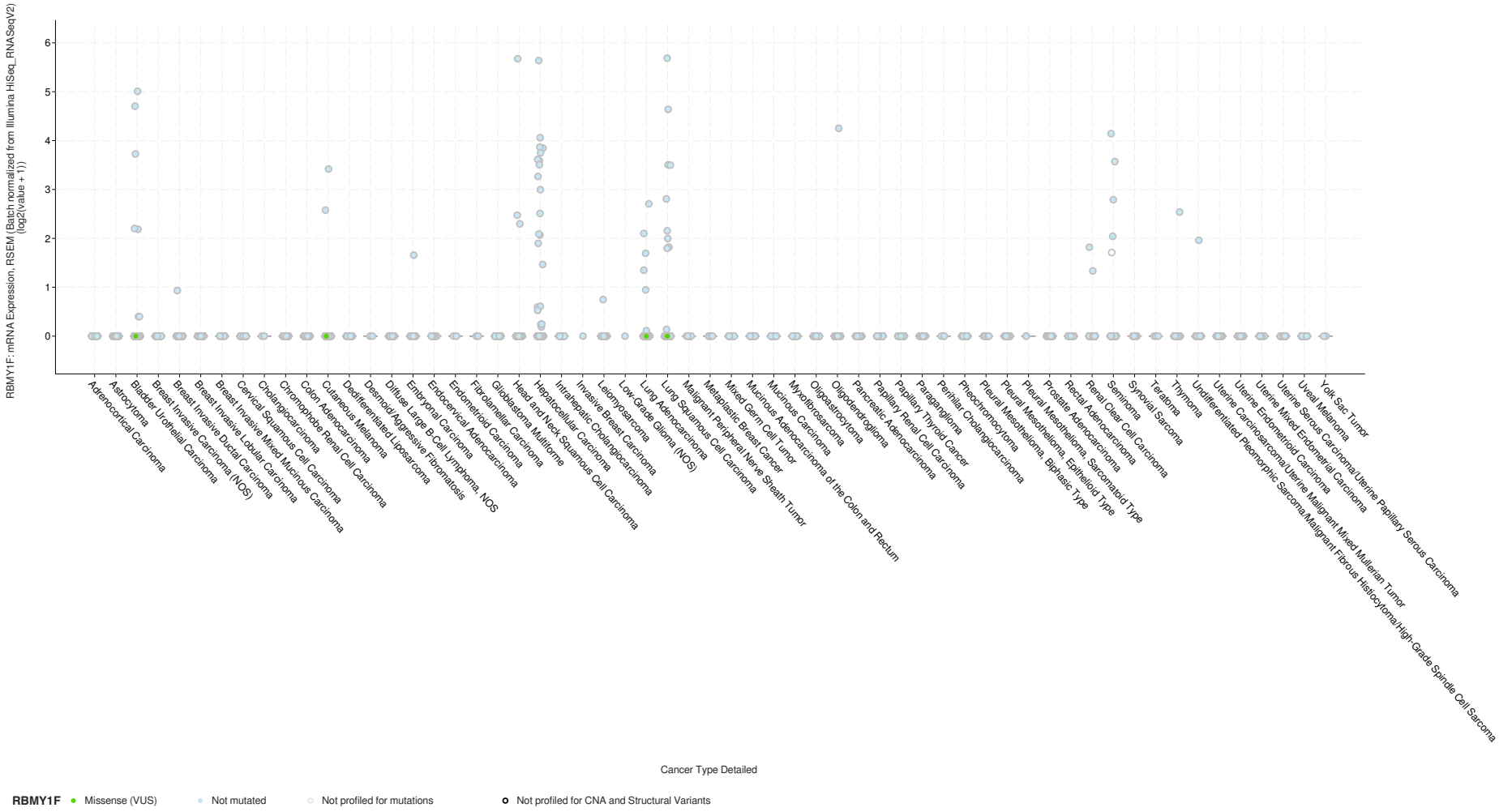
B. RBMY1B


RBMY1B: mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2) (log2(value + 1))
8
7
6
5
4
3
2
1
0
-1
Adrenocortical Carcinoma
Astrocytoma
Bladder Urothelial Carcinoma
Breast Invasive Carcinoma (NOS)
Breast Invasive Ductal Carcinoma
Breast Invasive Lobular Carcinoma
Breast Invasive Mixed Mucinous Carcinoma
Cervical Squamous Cell Carcinoma
Cholangiocarcinoma
Chromophobe Renal Cell Carcinoma
Colon Adenocarcinoma
Cutaneous Melanoma
Dedifferentiated Liposarcoma
Desmoid/Aggressive Fibromatosis
Diffuse Large B-Cell Lymphoma, NOS
Embryonal Carcinoma
Endocervical Adenocarcinoma
Endometrioid Carcinoma
Fibrolamellar Carcinoma
Glioblastoma Multiforme
Head and Neck Squamous Cell Carcinoma
Hepatocellular Carcinoma
Intrahepatic Cholangiocarcinoma
Invasive Breast Carcinoma
Leiomyosarcoma
Low-Grade Glioma (NOS)
Lung Adenocarcinoma
Lung Squamous Cell Carcinoma
Malignant Peripheral Nerve Sheath Tumor
Metaplastic Breast Cancer
Mixed Germ Cell Tumor
Mucinous Adenocarcinoma of the Colon and Rectum
Mucinous Carcinoma
Myxofibrosarcoma
Oligoastrocytoma
Oligodendroglioma
Pancreatic Adenocarcinoma
Papillary Renal Cell Carcinoma
Papillary Thyroid Cancer
Paraganglioma
Perihilar Cholangiocarcinoma
Pheochromocytoma
Pleural Mesothelioma
Pleural Mesothelioma, Biphasic Type
Pleural Mesothelioma, Epithelioid Type
Prostate Adenocarcinoma
Rectal Adenocarcinoma, Sarcomatoid Type
Renal Clear Cell Carcinoma
Seminoma
Synovial Sarcoma
Teratoma
Thymoma
Undifferentiated Pleomorphic Sarcoma/Malignant Fibrous Histiocytoma/High-Grade Spindle Cell Sarcoma
Uterine Carcinosarcoma/Uterine Malignant Mixed Mullerian Tumor
Uterine Endometrioid Carcinoma
Uterine Mixed Endometrial Carcinoma
Uterine Serous Carcinoma/Uterine Papillary Serous Carcinoma
Uveal Melanoma
Yolk Sac Tumor
Cancer Type Detailed
RBMY1B
Not mutated
Not profiled for mutations
Not profiled for CNA and Structural Variants

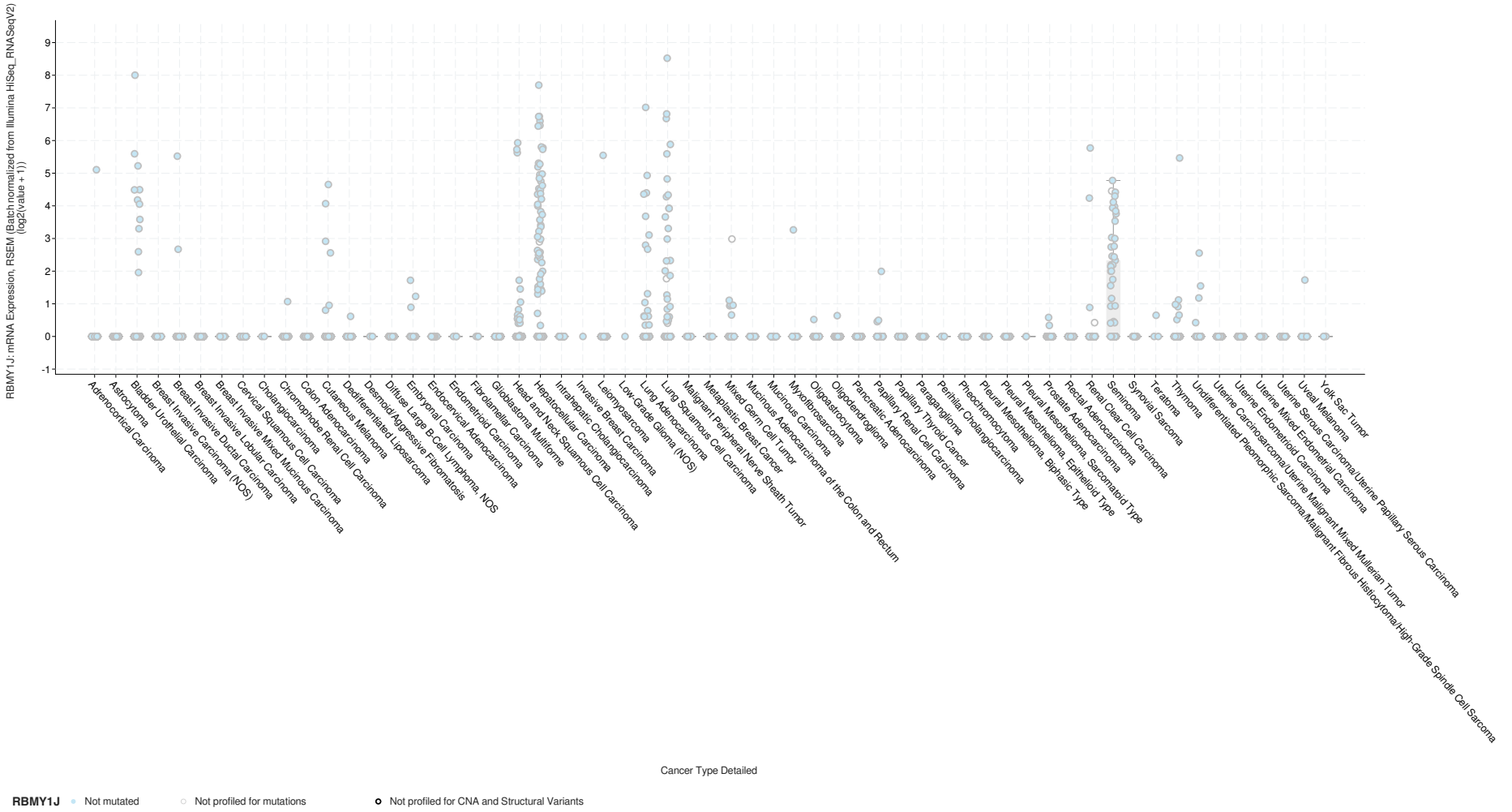
C. RBMY1E


RBMY1E: mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2) (log2(value + 1))
6
5
4
3
2
1
0
Adrenocortical Carcinoma
Astrocytoma
Bladder Urothelial Carcinoma
Breast Invasive Carcinoma (NOS)
Breast Invasive Ductal Carcinoma
Breast Invasive Lobular Carcinoma
Breast Invasive Mixed Mucinous Carcinoma
Cervical Squamous Cell Carcinoma
Cholangiocarcinoma
Chromophobe Renal Cell Carcinoma
Colon Adenocarcinoma
Cutaneous Melanoma
Dedifferentiated Liposarcoma
Desmoid/Aggressive Fibromatosis
Diffuse Large B-Cell Lymphoma, NOS
Embryonal Carcinoma
Endocervical Adenocarcinoma
Endometrioid Carcinoma
Fibrolamellar Carcinoma
Glioblastoma Multiforme
Head and Neck Squamous Cell Carcinoma
Hepatocellular Carcinoma
Intrahepatic Cholangiocarcinoma
Invasive Breast Carcinoma
Leiomyosarcoma
Low-Grade Glioma (NOS)
Lung Adenocarcinoma
Lung Squamous Cell Carcinoma
Malignant Peripheral Nerve Sheath Tumor
Metaplastic Breast Cancer
Mixed Germ Cell Tumor
Mucinous Adenocarcinoma
Mucinous Carcinoma
Myxofibrosarcoma
Oligoastrocytoma
Oligodendroglioma
Pancreatic Adenocarcinoma
Papillary Renal Cell Carcinoma
Papillary Thyroid Cancer
Paraganglioma
Perihilar Cholangiocarcinoma
Pheochromocytoma
Pleural Mesothelioma
Pleural Mesothelioma, Biphasic Type
Pleural Mesothelioma, Epithelioid Type
Prostate Adenocarcinoma
Rectal Adenocarcinoma, Sarcomatoid Type
Renal Clear Cell Carcinoma
Seminoma
Synovial Sarcoma
Teratoma
Thymoma
Undifferentiated Pleomorphic Sarcoma/Malignant Fibrous Histiocytoma/High-Grade Spindle Cell Sarcoma
Uterine Carcinosarcoma/Uterine Malignant Mixed Mullerian Tumor
Uterine Endometrioid Carcinoma
Uterine Mixed Endometrial Carcinoma
Uterine Serous Carcinoma/Uterine Papillary Serous Carcinoma
Uveal Melanoma
Yolk Sac Tumor
Cancer Type Detailed
RBMY1E
Splice (VUS)
Truncating (VUS)
Missense (VUS)
Not mutated
Not profiled for mutations
Not profiled for CNA and Structural Variants

D. RBMY1F


RBMY1F: mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2) (log2(value + 1))
0
1
2
3
4
5
6
Adrenocortical Carcinoma
Astrocytoma
Bladder Urothelial Carcinoma
Breast Invasive Carcinoma (NOS)
Breast Invasive Ductal Carcinoma
Breast Invasive Lobular Carcinoma
Breast Invasive Mixed Mucinous Carcinoma
Cervical Squamous Cell Carcinoma
Cholangiocarcinoma
Chromophobe Renal Cell Carcinoma
Colon Adenocarcinoma
Cutaneous Melanoma
Dedifferentiated Liposarcoma
Desmoid/Aggressive Fibromatosis
Diffuse Large B-Cell Lymphoma, NOS
Embryonal Carcinoma
Endocervical Adenocarcinoma
Endometrioid Carcinoma
Fibrolamellar Carcinoma
Glioblastoma Multiforme
Head and Neck Squamous Cell Carcinoma
Hepatocellular Carcinoma
Intrahepatic Cholangiocarcinoma
Invasive Breast Carcinoma
Leiomyosarcoma
Low-Grade Glioma (NOS)
Lung Adenocarcinoma
Lung Squamous Cell Carcinoma
Malignant Peripheral Nerve Sheath Tumor
Metaplastic Breast Cancer
Mixed Germ Cell Tumor
Mucinous Adenocarcinoma of the Colon and Rectum
Mucinous Carcinoma
Myxofibrosarcoma
Oligoastrocytoma
Oligodendroglioma
Pancreatic Adenocarcinoma
Papillary Renal Cell Carcinoma
Papillary Thyroid Cancer
Paraganglioma
Perihilar Cholangiocarcinoma
Pheochromocytoma
Pleural Mesothelioma
Pleural Mesothelioma, Biphasic Type
Pleural Mesothelioma, Epithelioid Type
Prostate Adenocarcinoma
Rectal Adenocarcinoma, Sarcomatoid Type
Renal Clear Cell Carcinoma
Seminoma
Synovial Sarcoma
Teratoma
Thymoma
Undifferentiated Pleomorphic Sarcoma/Malignant Fibrous Histiocytoma/High-Grade Spindle Cell Sarcoma
Uterine Carcinosarcoma/Uterine Malignant Mixed Mullerian Tumor
Uterine Endometrioid Carcinoma
Uterine Mixed Endometrial Carcinoma
Uterine Serous Carcinoma/Uterine Papillary Serous Carcinoma
Uveal Melanoma
Yolk Sac Tumor
Cancer Type Detailed
RBMY1F
Missense (VUS)
Not mutated
Not profiled for mutations
Not profiled for CNA and Structural Variants

E. RBMY1J

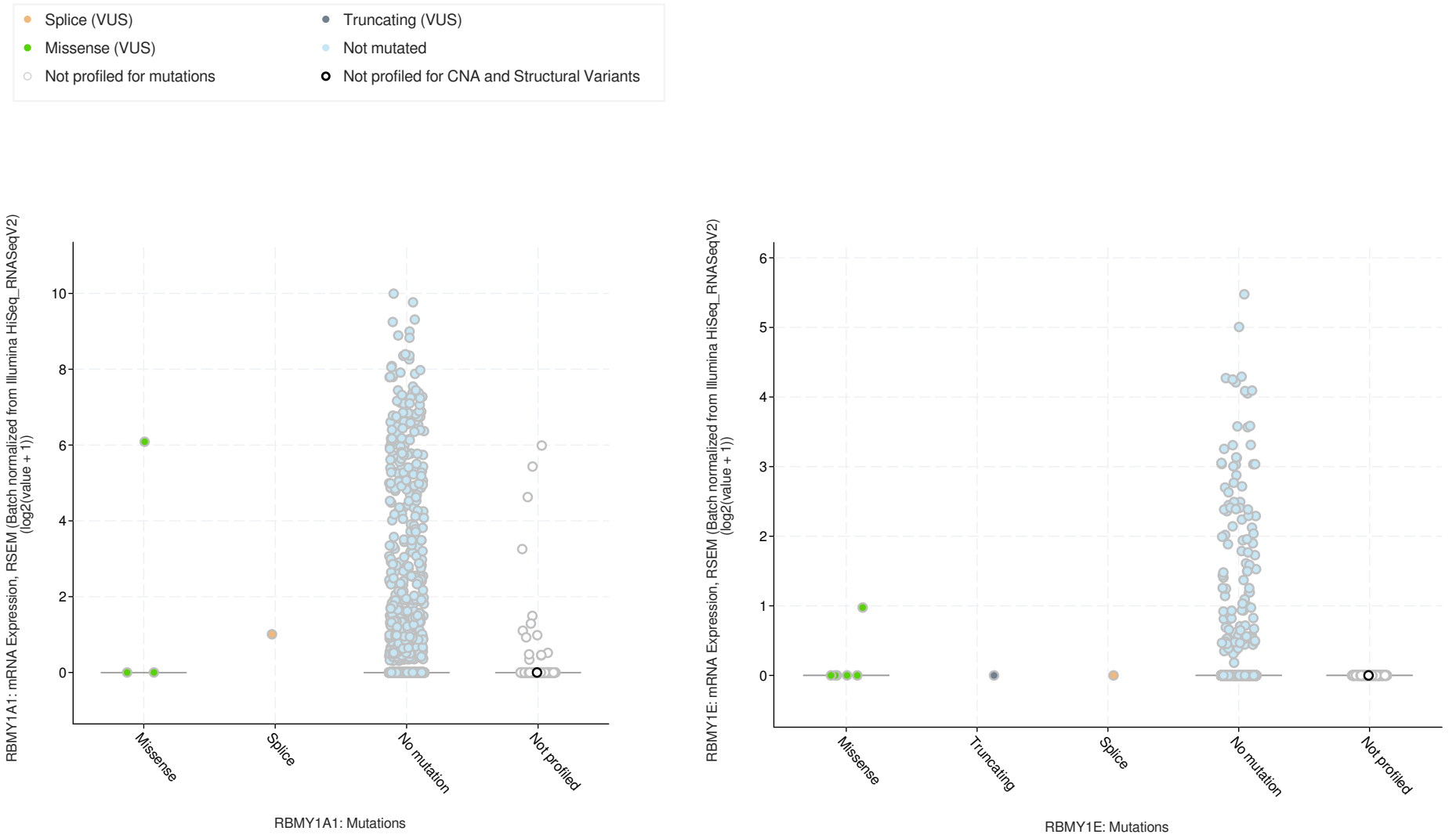


(Gao *et al.*, 2013)

Figure 21: RBMY expression in different cancers. *Expression was assessed through RBMY mRNA expression (RSEM (Batch normalized from Illumina HiSeq_RNASeqV2)(log2(value + 1)). Fig. 21A-E, RBMY is shown to be expressed largely in Bladder Urothelial Carcinoma, HNSCC, Hepatocellular carcinoma, Lung Adenocarcinoma, Lung Squamous Cell Carcinoma, Cutaneous Melanoma and Seminoma across the various RBMY genes, although with varied expression depending on the gene.*

The patient data of *RBMY* expressing cancer patients from the studies on cBioPortal was analysed and we were able to determine that *RBMY* in these cancers was being expressed exclusively in male patient cancers, by definition this has to be true as females don't have a Y-chromosome. When we looked at the literature of the cancer types in which *RBMY* is seen expressed, the cancers all show a high incidence rate in males than in females (see Table 1). RBMY has been shown to have oncogenic activity that account in part for the high prevalence in male HCC and play a key role in the aggressive nature of these HCC tumours. It remains to be shown if *RBMY* may in part also be an oncogene in these other cancers as only a handful of tumours from each of these cancer types is expressing RBMY. Nevertheless, the specificity of expression of *RBMY* genes in the cancer types combined with what is known about RBMY expression in HCC suggests that RBMY may potentially pay a key role in development of certain male cancers.

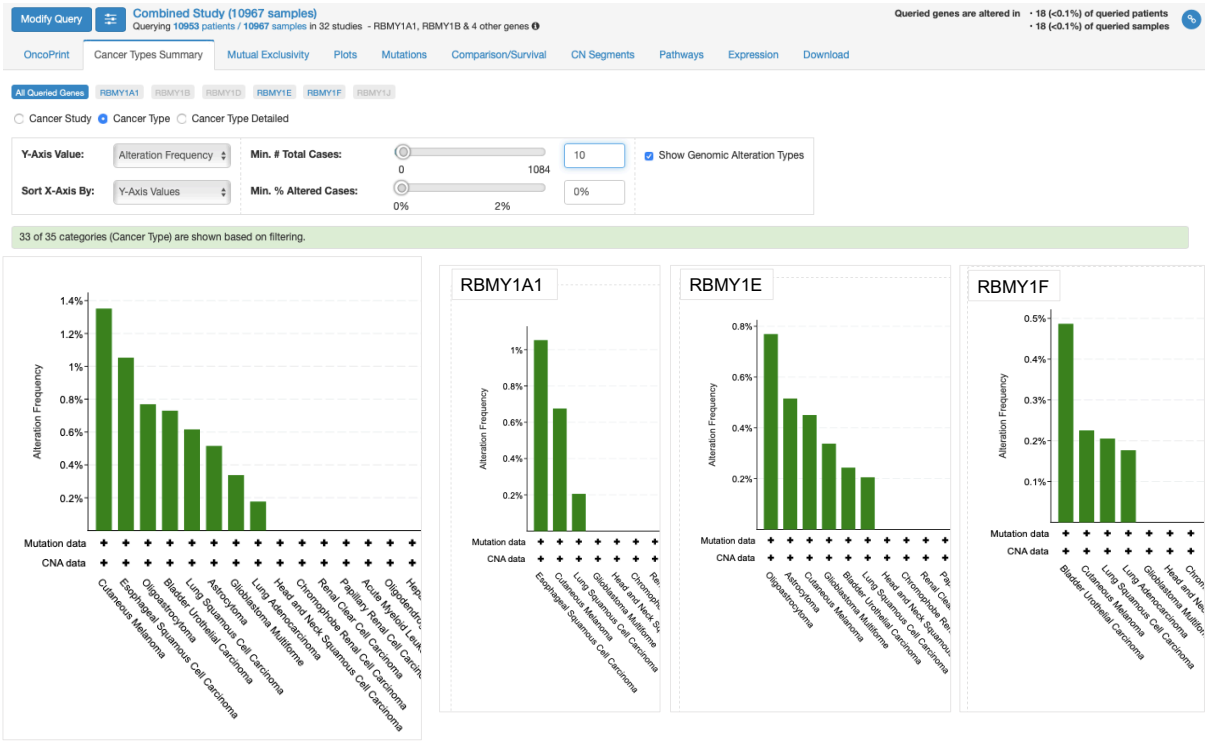
Following from this, we wondered if maybe RBMY is mutated and if mutated, in what proportion of the cancers is *RBMY* mutations detected. We also wondered if RBMY, described as a novel oncogene driving male specific HCC acquires driver mutations that potentates its oncogenic activity. Using cBioPortal to analyse large scale studies in which *RBMY* expression is detected in cancers, *RBMY* genes are shown to be hardly mutated when expressed in these cancers and the mutated *RBMY* genes are largely inactivated in cancers, with very minute mRNA expression of the mutated *RBMY* genes detected in the cancers (Fig. 22).



(Gao *et al.*, 2013).

Figure 22: RBMY mutations plots in cancers tissues. RBMY genes are scarcely mutated, with most the few mutations observed being mostly missense mutations shown in RBMY1A1, RBMY1E and RBMY1F (supplementary data). In RBMY1A1 and RBMY1E, splice mutation have been observed in Skin cutaneous melanoma and Lund Squamous Cell Carcinoma respectively. This is of interest as splice mutations are noted in literature as associated with cancers (Anczuków and Krainer, 2016).

Moreover, mutations of *RBMY* genes to date have only been recorded in *RBMY1A1*, *RBMY1E* and *RBMY1F* and not in the other *RBMY* genes. The data suggest that *RBMY* genes only exhibit missense and truncating mutations (Fig. 22). The alterations of *RBMY1A1*, *RBMY1E* and *RBMY1F* genes are all commonly seen in Cutaneous and Lung cancers with varied expression of the 3 RBMY genes in other cancers of the Esophageal Squamous Cell Carcinoma, Ogligoastrocytoma, Bladder Eurothelial Carcinoma, Astrocytoma and Glioblastoma (Fig. 23). These mutations do however occur at about 0.1% of somatic mutation frequency (Fig. 23), therefore it is very unlikely that *RBMY* genes acquire driver mutations in these cancers to act as an oncogene.



(Gao *et al.*, 2013)

Figure 23: RBMY mutation quantity in cancers. The Graphs show the alteration frequency of the RBMY genes in the Different type of Cancers. As shown in fig. 22, RBMY genes are hardly mutated, and this is also displayed here with the highest alteration frequency for any RBMY gene being RBMY1A1 in Cutaneous Melanoma at less that 1.4% alteration frequency.

From this general survey using data from over 10000 cancer patient data queries, we can see that *RBMY* expression is seen in several cancers that have high prevalence in males. The ectopic expression of *RBMY* genes in Bladder Urothelial Carcinoma, Hepatocellular carcinoma, Lung Adenocarcinoma, Lung Squamous Cell Carcinoma, Seminoma and Cutaneous Melanoma highlights the possibility of an oncogenic role of *RBMY* genes in male prevalent cancer.

3.2 Western blot of UM-SCC-104 RBMY GFP and WT

This experimentation was conducted to detect RBMY protein expression in the transfected UM-SCC-104 GFP-RBMY cells as the fluorescence signal had been thought to be diminishing overtime.

UM-SCC-104 cells that were initially used in the project had been prepared by the student whose project I am continuing. The cells had been transfected with a plasmid encoding *GFP-RBMY* and selected with G418 to generate a stable, GFP-RBMY-expressing cell line (Fig. 15A). When viewed under a fluorescence microscope however, the cells did not show a strong GFP signal and over time the GFP signal appeared to weaken. We hypothesised that the GFP-RBMY expression was being lost with increasing passaging, possibly due to a deleterious effect on cell proliferation or viability. To test for this, I conducted a western blot using UM-SCC-104 GFP-RBMY and WT UM-SCC-104. Figure 24 western blot image showed that the UM-SCC-104 GFP-RBMY were expressing GFP-RBMY, however, using fluorescence microscopy on live cells, we observed a dramatic loss of GFP fluorescence over subsequent passages. We suspect this might reflect an active suppression of GFP-RBMY expression or simply selection of cells with a reduction or complete loss of expression due to its previously discussed negative effects on mitogenic signalling. As previously mentioned, RBMY is hypothesised to act as a tumour suppressor that initially inhibits oncogenic effects of signalling pathways associated with RAS and AKT proteins (Kido *et al.*, 2020). It is possible that the reduced GFP-RBMY signalling is as the result RAS/RAF/MAP and PI3K/AKT signalling pathways overwhelming RBMY

expression as the pathways overcome the RBMY block to promote tumorigenic effects. However, due to a malfunction of our liquid nitrogen storage over Christmas 2019, the stored stock vials of the GFP-RBMY expressing UM-SCC-104 were unfortunately lost.

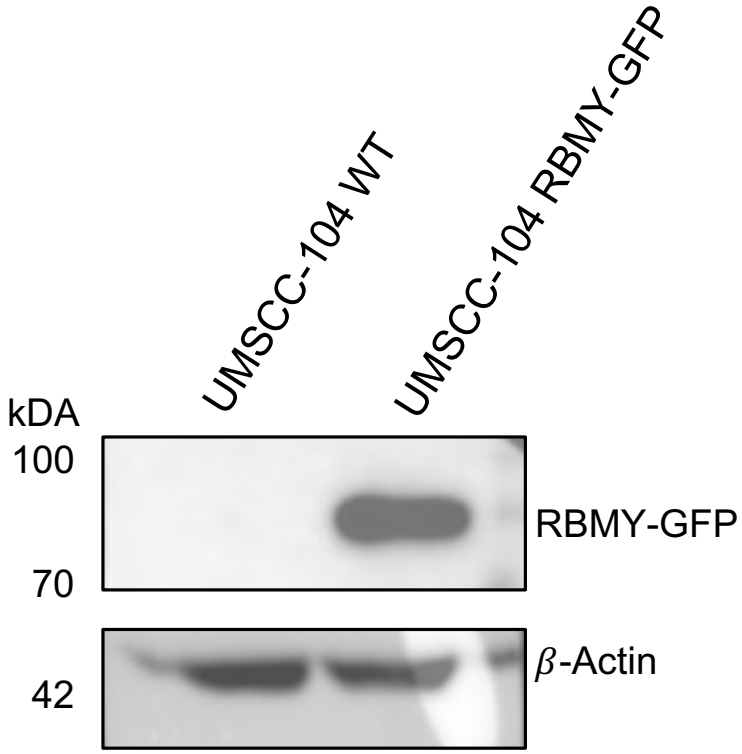


Figure 24: GFP-RBMY expressed in mutant UM-SCC-104 cells. Western blot showing GFP-RBMY protein is being expressed in UM-SCC-104 GFP-RBMY cells. GFP-RBMY protein was detected by rabbit anti-GFP primary antibody and HRP-conjugated secondary anti-rabbit GFP antibody.

3.3 *RBMY* and *ZFY* expression in UM-SCC-104 and 93-UV-147T cell lines.

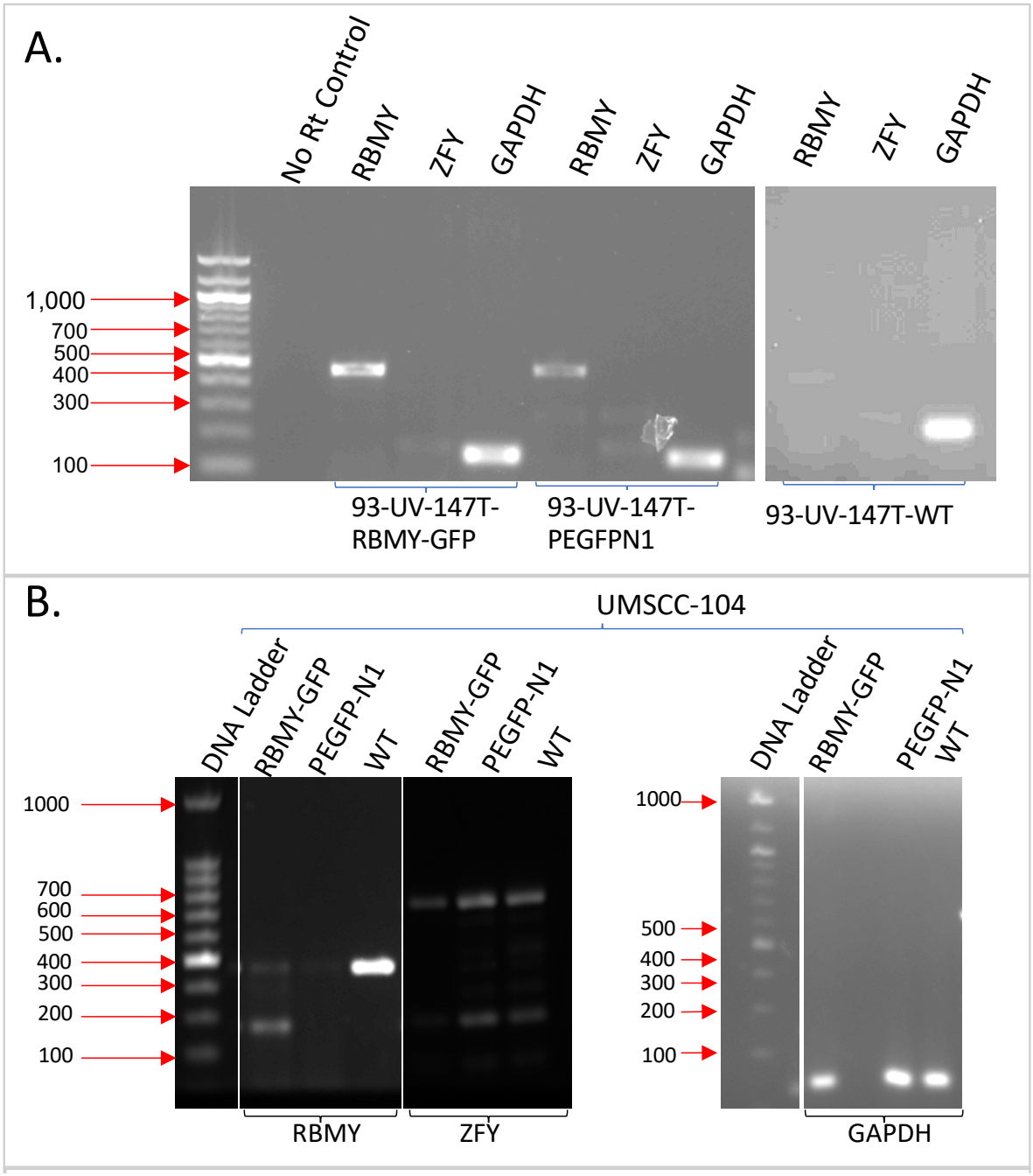


Figure 25 *RBMY* is expressed in HNSCC cell lines UM-SCC104 and 93-UV-147T. **(A)** Agarose gel image showing *RBMY* and *ZFY* transcription in 93-UV-147T GFP-RBMY, 93-UV-147T pEGFP and 93-UV-147T WT. **(B)** Agarose gel image showing RBMY and ZFY transcription in UM-SCC-104 GFP-RBMY, UM-SCC-104 pEGFP and UM-SCC-104 WT. Images were taken on a syngene G:box imager.

These experiments were conducted to attempt in detection of positive RBMY transcription in the newly transfected GFP-RBMY 93-UV-147T and UM-SCC-104 cells.

The presence of *RBMY* was detected in both UM-SCC-104 and 93-UV-147T (Fig. 25A and Fig. 25B). For this project, we wanted to examine the localisation and activity of *RBMY* and thus we transfected UM-SCC-104 and 93-UV-147T with *GFP-RBMY* and *GFP* (negative control) in order to visualise RBMY within cells. Given the problems with loss of GFP-RBMY expression in stable lines, transiently transfected cells were collected 3 days post transfection and RT-PCR was conducted to confirm the presence of RBMY (primers used were targeting RBMY exon 2 and 5) and ZFY expression. *GAPDH* gene (123 bp), which is ubiquitously expressed by all cells was amplified for control. Both 93-UV-147T GFP-RBMY and GFP showed presence of *RBMY*, although 93-UV-147T GFP-RBMY showed higher expression of *RBMY*, predictably as a result of the transfection. Given time, we would have been confirmed this using qRT-PCR. Presence of *ZFY-short* (156bp) was also shown in both *GFP-RBMY* and *pEGFP-N1* (vector control)-transfected 93-UV-147T, although the expression was very low as seen by very faint bands. Primer blast software analysis would predict that we were detecting RBMY1D, which is a functional copy on RBMY. Interestingly, we were able to show expression of both *ZFY-long* and *ZFY-short* in untransfected UM-SCC104, GFP-RBMY and pEGFP. In 93-UV-147T, faint bands of *ZFY-short* were observed GFP-RBMY and PeGFP lines. RBMY expression was observed in untransfected UM-SCC-104, showing that there is endogenous *RBMY* expression in this cell line. With these results we were able to show expression of RBMY in two HPV16+ HNSCC cell lines: UM-SCC-104 and 93-UV-147T.

3.4 RBMY localisation

Fluorescence microscopy experiments were conducted to examine the localisation of RBMY in the transfected cells.

To visualise the localisation of RBMY, we tried to select GFP-RBMY and PEFGP-N1 transfected UM-SCC-104 and 93-UV-147T cells for 2 weeks using G418 selection, majority of the cells kept dying. Transfection success rate for any of our transfections excluding 93-UV-147T were almost always less than 10% as visualised in figures 22 and 26. We think this is probably because despite the cells successfully undergoing transient transfection where the plasmid inserts into cells, very few cells were forming stable integration of the plasmid into the cellular DNA. Therefore, we could not form stable UM-SCC-104 GFP-RBMY or 93-UV-147T GFP-RBMY. Despite this, we were able to get images of RBMY localisation in both UM-SCC-104 GFP-RBMY and 93-UV-147T GFP-RBMY cells 6 days post transfection. Figure 22 shows that RBMY expression in both UM-SCC-104 and 93-UV-147T is restricted to the nucleus. RBMY is reported to be a splicing factor, therefore its expression would be expected to be within the nucleus.

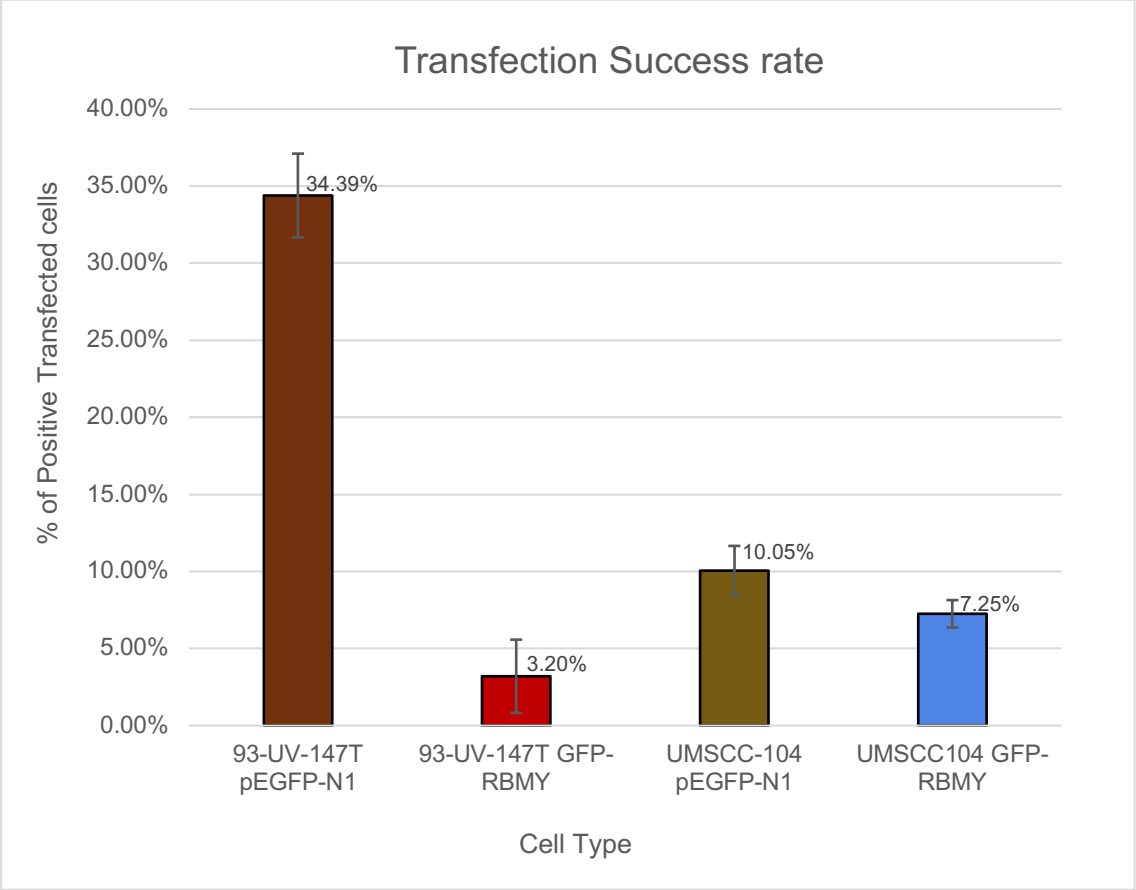
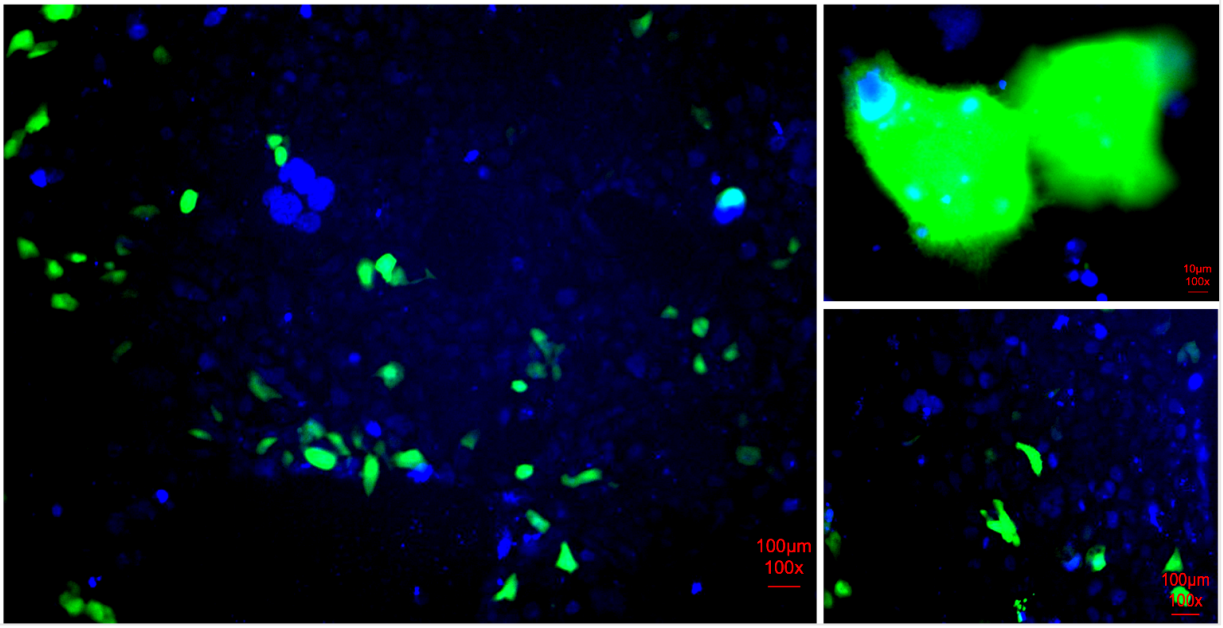
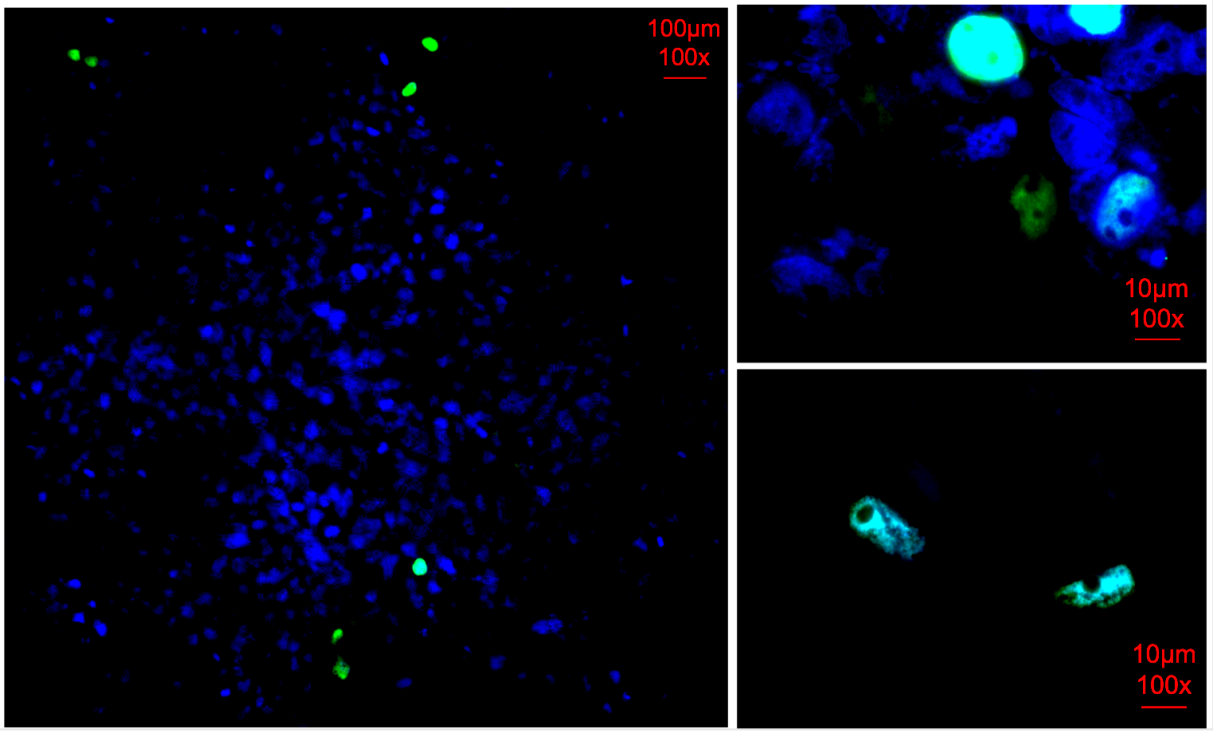


Figure 26: Transfection success rate of the transfected cells. The graph displays the success rate of transfected cells. Microscopy images (taken 2 weeks post transfection) were utilised to count cells with successful transfection Data shows that 93-UV-147T had the most successful transfection rate at 6 days post transfection whilst the rest of the cell types failed to have transfection rate at greater than 15% of cells with successful transfection.

93-UV-147T-PEGFPN1



93-UV-147T GFP-RBMY



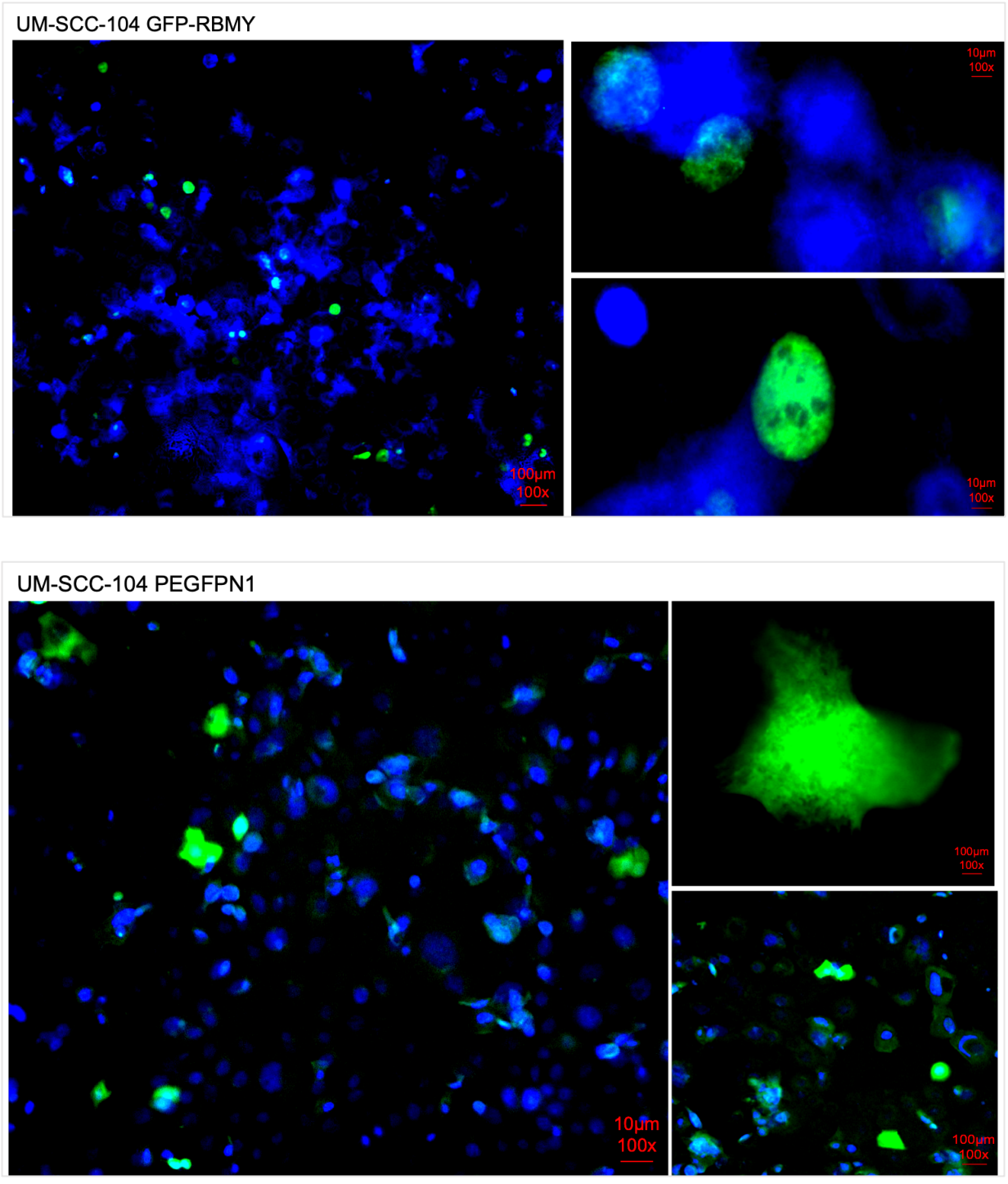
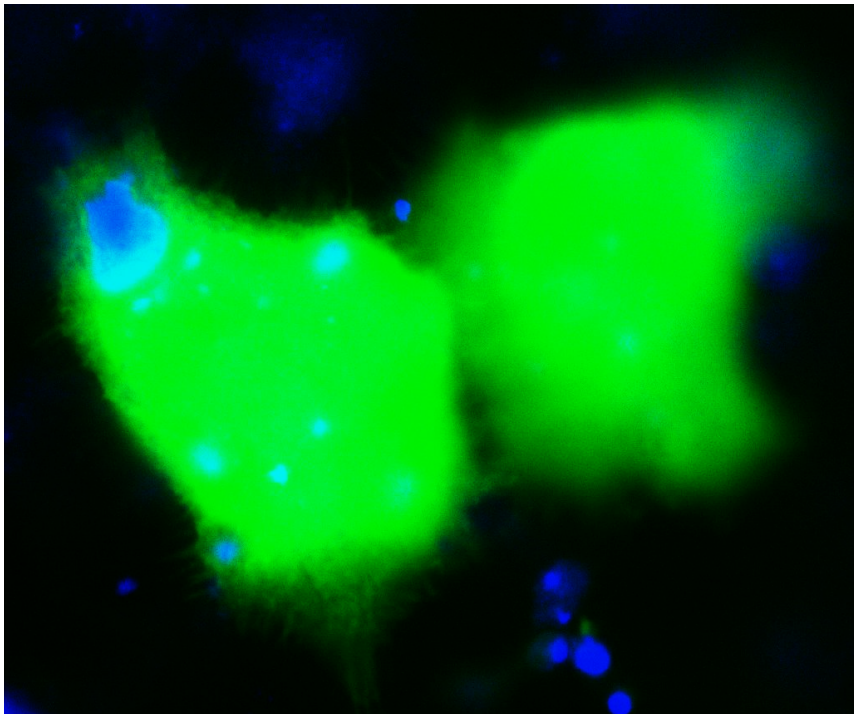
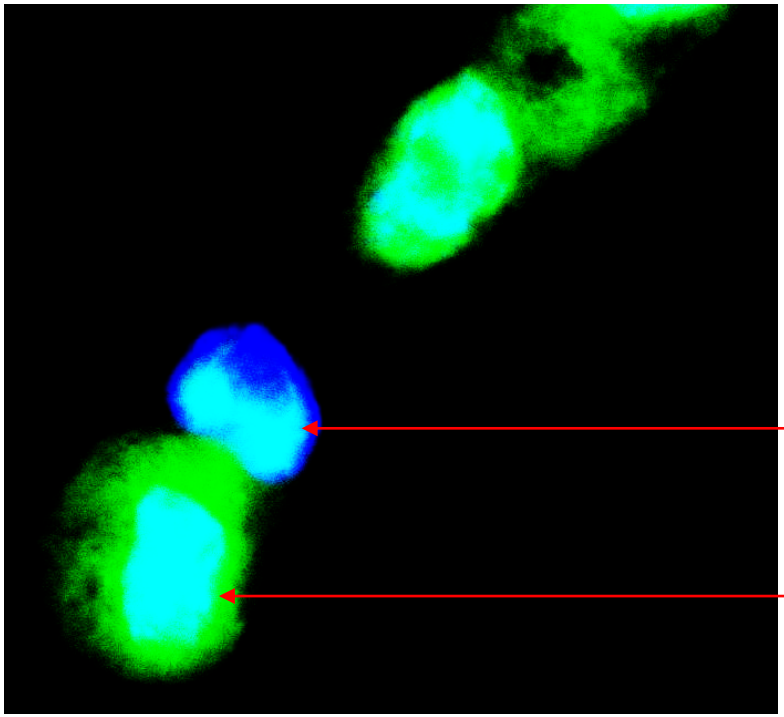


Figure 27: RBMY expression is localised within the nucleus. Fluorescent microscopy was used to image the expression of RBMY in 93-UV-147T and UM-SCC-104 cell lines using the GFP tag attached to the 5' end of RBMY. PEGFP-N1 plasmid was used as negative control in order to show GFP expression without RBMY. Fluorescence signals were observed at both 10x and 100x magnification. Cells were stained with DAPI to show nucleus staining.



PEGFPN1 UMSCC104



93-UV-147T GFP-RBMY

DAPI staining showing the nucleus (blue)

GFP-RBMY localisation shown as nucleus with the GFP fluorescence superimposing the nucleus stain

Figure 28: Images displaying florescence signalling of nuclear GFP-RBMY and cytoplasmic GFP. Fluorescence microscopy images of PEGFP-N1 UMSCC104 and (3-UV-147T GFP-RBMY cells. The images show the contract between cytoplasmic GFP expression and nuclear GFP which shows the cellular localisation of RBMY gene.

3.5 qPCR data: Quantification of RBMY expression

These experiments were conducted to attempt at quantifying the RBMY expression in the transfected cells.

In order to study the expression profiles of *RBMY* and *ZFY* gene variants, we studied the effect of RBMY overexpression on the expression of *ZFY* variants. We decided to conduct a series of qPCR experiments. Unfortunately, as a result of the Covid lockdown, we were only able to run a small set of preliminary tests and as such lack sufficient and significant data. However, despite this, we believe it is beneficiary to show these results because they do in part show some data worth noting post this thesis. The results from these preliminary tests show overexpression of RBMY protein in UM-SCC-104 cells did indeed affect the expression of *ZFY-short* as the results show a significant difference in *ZFY-short* expression between UM-SCC-104 GFP RBMY and pEGFP cells (Fig. 18). We unfortunately were not able to do the same experiment with 93-UV-147T cells as well as run a series of repeats of qPCR with UM-SCC-104 to solidify the above statement. However, these preliminary results do indicate a likelihood of RBMY overexpression being associated with an increase of *ZFY-short* expression.

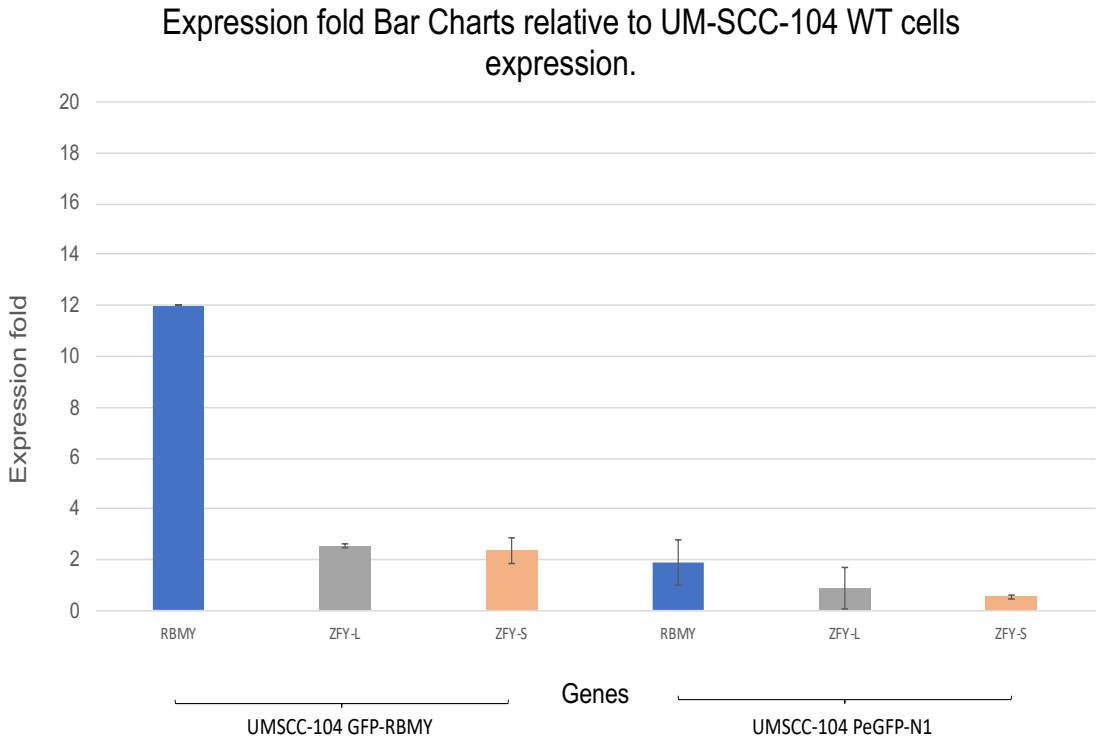


Figure 29: RBMY overexpression is associated with the upregulation of ZFY-short expression. The bar chart represents the expression fold values (ΔCt) of RBMY, ZFY-long and ZFY-short genes in UM-SCC-104 GFP-RBMY and UM-SCC-104 pEGFP relative to the expression of UM-SCC-104 WT.

4. **DISCUSSION.**

4.1 The motivations of our study

HNSCC has become an overwhelming healthcare burden, and its high prevalence in males has long warranted a discussion and research. Changing societal behaviours has meant a massive shift in exposure to high-risk carcinogens like smoking, alcohol abuse or infections to oncogenic viruses or bacteria (Fakhry *et al.*, 2018). And as such, over the last decade, new research has shown that influence of sex hormones and the intrinsic gender differences in susceptibility to chemical carcinogens may in part be driving male prevalence in cancers (Kadekar *et al.*, 2012). In line with this, we believe that the genetic distinction between males and females may also account for this male prevalence in HNSCC. The Y chromosome, the major genetic distinction between males and females may potentially enable male specific cancer inducing activities that may drive in part the high male HNSCC prevalence. Testis specific Y linked genes have been shown to be aberrantly activated in somatic cell cancers (Kido and Lau, 2015). This is interesting because testis specific Y linked genes, particularly those that are involved in or interact with proteins that have a role in meiosis, could potentially exert meiotic roles during oncogenesis. This is because the process of meiosis and oncogenesis share analogous processes, interchanging of genetic materials between homologous chromosomes occurring in meiosis, which is analogous to cancer genomic instability, a classic hallmark of cancer which potentiates various genetic modifications that drive the initiation and development of cancer (Nielsen and Gjerstorff, 2016). RBMY, the testis specific Y linked genes have been described as a novel oncogene in male HCC (Tsuei *et al.*, 2011), is the gene of interest in our research partly because of its literature as an oncogene in HCC, but also because of its role in spermatogenesis which we think may be associated with ZFY-short splicing. RBMY is a spicing factor that is translated preferentially in pre-meiotic germ line cells which also pre-dominantly expresses ZFY-s (Abid *et al.*, 2013). Furthermore, the link between

RBMY and ZFY-short was drawn from a study that showed that a man (Ste-363) with an AZFb deletion had greatly reduced levels of ZFY-short , while a man (Apop12) with a similar stage meiotic arrest but no AZFb deletion had no reduction in ZFY-short levels (Decarpentrie *et al.*, 2012). With this literature combined with our unpublished results in which ZFY-short is shown expressed in HNSCC cell line UM-SCC-104, we want to move towards a model to test the hypothesis in which RBMY induced oncogenic activity via regulation of splicing of ZFY-s, a gene we deduce as a result of its expression in highly mitotically active cells, potentially has an anti-apoptotic function, a key function in oncogenesis.

4.2 RBMY expression in Cancers.

RBMY expression in tumours is well documented in liver cancers, and the literature on RBMY cancer expression in HCCs has described RBMY as a male specific novel oncogene, which has been shown to be associated with aberrant activation of molecular pathways that drive cancer development (Tsuei *et al.*, 2004, 2011; Chua *et al.*, 2015; Kido *et al.*, 2020). RBMY status as an oncogene in HCCs suggests the ability of RBMY to drive in part the male prevalence in liver cancers, and thus, it might be possible that RBMY is also significantly seen expressed in various other cancers, perhaps exerting the same oncogenic effects in those cancers as described in male HCC. Our computational analysis using cancer patient data on cBioPortal to study the expression of *RBMY* in cancer (via cBioPortal) revealed that *RBMY* is expressed in various different cancers, in Bladder Urothelial Carcinoma, Hepatocellular carcinoma, Lung Adenocarcinoma, Lung Squamous Cell Carcinoma, Seminoma and Cutaneous Melanoma. There was a huge variation to the extent in which *RBMY* genes were being expresses in these cancers, with *RBMY1A1* showing the highest expression in cancers, *RBMY1B* and *RBMY1E* being expressed at around 1/3 the expression of *RBMY1A1* and *RBMY1F* and *RBMY1E* showing the weakest expression, with neither of the genes being expressed above 50 RSEM RNA seq mRNA values. We suggest that is it necessary to validate this expression of *RBMY* in these cancers by defining an expression threshold which indicates significant *RBMY*

expression in cancer tissues. This would enable us to calculate the percentage of tumours in each of these cancer types in which *RBMY* expression is above this threshold.

When we analysed the literature around RSEM-normalised count, we couldn't find any definition for RSEM-normalised read count value that defines high or significant expression. However, a TCGA paper on HNSCC discounted HPV E6/E7 mRNA expression with counts below 1000 (Lawrence *et al.*, 2015). Despite this, it is worth nothing that meaningful gene expression in cancer tissues is gene-dependant, therefore it is possible that significantly less *RBMY* transcript is necessary to give meaningful protein expression than E6/E7 transcript. Our assumption is that, because RBMY is shown to be oncogenic in male liver tissues by multiple studies (Tsuei *et al.*, 2004) (Tsuei *et al.*, 2011) (Chua *et al.*, 2015) (Kido *et al.*, 2020), the level of *RBMY* transcripts observed in HCC tissues must be meaningful, therefore at least in tumours from other cancer types that show similar levels of expression are of interest.

Based on our analysis of the data from cBioPortal (Fig. 21), in which *RBMY* expression is seen in the above-mentioned cancers, the data suggests that the RBMY oncogenic activity based on the expression pattern in these cancers may be restricted to the small minority of tumours of different cancer types that expressed *RBMY*. This is because only a handful of tumours in these cancer types groups is expressing *RBMY*. Furthermore, due to limited expression of *RBMY* genes in the patient data on cBioPortal, we were unable to analyse whether these *RBMY* genes are co-expressed or mutually exclusively expressed in these cancers, further showing that *RBMY* expression in these cancers of this patient data was only limited to a handful of cancers.

Because studies have shown RBMY to have oncogenic activity in male HCC tumours, our computational analysis of in silico cBioPortal data perhaps serves as pre-liminary data that warrants further research to explain the reason RBMY is ectopically expressed in these somatic cell cancers. The ectopic expression of *RBMY* genes in Bladder Urothelial Carcinoma, Hepatocellular carcinoma, Lung Adenocarcinoma, Lung Squamous Cell Carcinoma, Seminoma and Cutaneous

Melanoma highlights the possibility of an oncogenic role of *RBMY* genes that may in part drive male specific oncogenic processes driving the male prevalence of these cancers. Another possibility that could explain the expression pattern of *RBMY* in these cancers would be that *RBMY* perhaps was expressed earlier during the oncogenesis of these cancers, but the time of detection, these tumours may have already developing limiting the expression of *RBMY*. However, this is very difficult to prove. The data showed that *RBMY* is rarely mutated in tumour environment and does not acquire an activating mutation (Fig. 22).

4.3 RBMY expression in HNSCC.

Although we were unable to successfully achieve our aim to move towards a model to test the hypothesis in which RBMY regulates *ZFY* splicing pattern, the series of experiments we ran were able to show that *RBMY* is expressed in our 2 male HNSCC cell models and may be associated with the expression of ZFY-s. This is important to our hypothesis as ZFY-short is expressed mainly in high mitotically active cells, and potentially has an anti-apoptotic function, activities that the gene could potentially exerting in male HNSCC. Interestingly, UM-SCC-104, which preliminary qPCR data (Fig. 29) suggest that *RBMY* expression in this cell line may be associated with ZFY-short expression, originated from HNSCC tumour that is highly aggressive and recurrent. This is in agreement with the literature on RBMY expression in male liver cancers in which RBMY expression is show to induce a highly aggressive form of HCC with a poor prognosis (Chua *et al.*, 2015; Kido *et al.*, 2020). However, as this is just preliminary data, more qPCR experiments are necessary to confirm the conclusion. One issue we faced with our PCR experiments was the detection of endogenous *RBMY* in both our UM-SCC-104 and 93-UV-147T cells as our primers which span the *RBMY* exon 2 to exon 4 were not specific enough to distinguish endogenous *RBMY* expression from *GFP-RBMY* fusion gene *RBMY* expression. To improve this, we were in the process of acquiring primers that would span the *GFP RBMY* junction, which would enable detection of specific to the expression of our target gene.

Through western blot (Fig. 24), we were able to show RBMY protein expression in our UM-SCC-104 cell lines. The GFP-RBMY protein band we detected showed a band with protein mass within the region of protein mass that we expected, with the expected mass being 82.7kDA and the band being detected able the 70kDA ladder band and bellow the 100kDA protein ladder band. It is worth noting that RBMY protein is known to have 3 protein isoforms (Alikhani *et al.*, 2013), but without protein structural experiments, we are unable to tell which isoform we expressed. Despite this, the specificity of the GFP tag assures us that our detected band was GFP-RBMY.

Furthermore, we were able to show that RBMY expression in cancer lines follows the same pattern as normal tissue RBMY expression. In our study with a fluorescence microscope, we showed that RBMY expressed was restricted in the nucleus by studying GFP fluorescence and DAPI staining on fixed cells (Fig. 22). Nuclear expression of RBMY is well documented in most studies looking at RBMY expression in cancer cell lines or tissues, with the exception of Huey-Huey Chua study which reports cytoplasmic RBMY presence.

The combined results of our bench work showed RBMY nuclear expression in 2 male HPV+ HNSCC cell lines, that may be associated with splicing and expression of ZFY-short. However, both RT-PCR and qPCR experiments are essential to the future of this study to determine whether or not *RBMY* expression is associated with the splicing control and expression of ZFY-s. This is an encouraging early set of results to a promising research idea that could explain RBMY oncogenic activity in HNSCC and how it adds to the high male prevalence of male HNSCC.

4.4 Issues that impeded or affected our research outcome

We faced several problems that limited our research project in the lab besides Covid lockdown, including a lab wide contamination as well the malfunction of the liquid nitrogen stores during the 2019 Christmas break. This damaged all of our

stored vials of our mutant cell lines and most of our wild-type cell lines, consequently putting us behind our work by 2 months.

The major limitation of our project was our inability to form stable GFP-RBMY cell line with UMSCC-104 or 93-UV-157T cells post 2019 Christmas break. We faced 2 issues with our transfections, the first was low percentage of successful transfections (stable DNA integration) as shown in fig. 27, and the second issue we faced (main reason for unsuccessful transfection) was the loss of transfected cells at the selection stage (using G418 selection antibiotic), with less than 5% of cells surviving overall.

Several experimental changes including thawing out vials of cells that had not been overly exhausted through many passages, changing the ratio of plasmid DNA: Serum during transfection were tried to no effect on success of transfections. One solution that we were about to explore and one we were confident would improve our chances of creating GFP-RBMY stable integration cell lines was to try retroviral transduction, a method that is widely utilised in cancer and stem cell research. This method utilises retroviral vectors, which carry pro-viral sequences that enable the integration of our gene into target cells with high probability of success rate [11]. The reason this method has high success rate is that in retrovirus induction, target plasma DNA is transfected into ‘packaging cells’ which expresses viral gag and pol that allow for the production of virions carrying target gene into cell culturing medium [11]. The cell culturing medium with the virions is used to infect the target cells, with the virions targeting the cells, and gene of interest stably integrated into the target cell DNA(*Addgene: Retrovirus Guide*, no date; Kurian, Watson and Wyllie, 2000).

5. **Bibliography**

Abid, S. *et al.* (2013) 'Cellular ontogeny of RBMY during human spermatogenesis and its role in sperm motility', *Journal of Biosciences*. Springer, 38(1), pp. 85–92. doi: 10.1007/s12038-012-9281-8.

Addgene: Retrovirus Guide (no date). Available at: https://www.addgene.org/guides/retrovirus/ (Accessed: 30 November 2020).

Alikhani, M. *et al.* (2013) 'Expression analysis of RNA-binding motif gene on Y chromosome (RBMY) protein isoforms in testis tissue and a testicular germ cell cancer-derived cell line (NT2)', *Iranian Biomedical Journal*. Pasteur Institute of Iran, 17(2), pp. 54–61. doi: 10.6091/ibj.1148.2013.

Anczuków, O. and Krainer, A. R. (2016) 'Splicing-factor alterations in cancers', *RNA*. Cold Spring Harbor Laboratory Press, 22(9), pp. 1285–1301. doi: 10.1261/RNA.057919.116.

Antoniw, J. *et al.* (1978) *Purification and Characterization of Protein Synthesis Initiation Factors EIF-1, EIF-4C, EIF-40, and EIF-5 From Rabbit Reticulocytes*, *J. Gen. Virol*.

Bachtrog, D. (2013) 'Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration', *Nature Reviews Genetics*, pp. 113–124. doi: 10.1038/nrg3366.

Barchi, M. *et al.* (2005) 'Surveillance of different recombination defects in mouse spermatocytes yields distinct responses despite elimination at an identical developmental stage.', *Molecular and cellular biology*, 25(16), pp. 7203–15. doi: 10.1128/MCB.25.16.7203-7215.2005.

Bashyam, M. D. *et al.* (2019) 'The Yin and Yang of cancer genes', *Gene*. Elsevier B.V., pp. 121–133. doi: 10.1016/j.gene.2019.04.025.

Bellott, D. W. *et al.* (2014) 'Mammalian y chromosomes retain widely expressed dosage-sensitive regulators', *Nature*. Nature Publishing Group, 508(7497), pp. 494–499. doi: 10.1038/nature13206.

Blot, W. J. *et al.* (1988) *Smoking and Drinking in Relation to Oral and Pharyngeal Cancer*, *CANCER RESEARCH*.

Burgoyne, P. S., Mahadevaiah, S. K. and Turner, J. M. A. (2009) 'The consequences of asynapsis for mammalian meiosis', *Nature Reviews Genetics*. Nat Rev Genet, pp. 207–216. doi: 10.1038/nrg2505.

Busch, C. J. *et al.* (2013) 'HPV-positive HNSCC cell lines but not primary human fibroblasts are radiosensitized by the inhibition of Chk1', *Radiotherapy and Oncology*,

108(3), pp. 495–499. doi: 10.1016/j.radonc.2013.06.035.
Calvisi, D. F. *et al.* (2006) 'Ubiquitous Activation of Ras and Jak/Stat Pathways in Human HCC', *Gastroenterology*. W.B. Saunders, 130(4), pp. 1117–1128. doi: 10.1053/j.gastro.2006.01.006.
Ceylan-Isik, A. F., McBride, S. M. and Ren, J. (2010) 'Sex difference in alcoholism: Who is at a greater risk for development of alcoholic complication?', *Life Sciences*. NIH Public Access, pp. 133–138. doi: 10.1016/j.lfs.2010.06.002.
Chai, N. N. *et al.* (1998) 'Structure and organization of the RBMY genes on the human Y chromosome: Transposition and amplification of an ancestral autosomal hnRNPG gene', *Genomics*, 49(2), pp. 283–289. doi: 10.1006/geno.1998.5255.
Chai, N. N., Salido, E. C. and Yen, P. H. (1997) 'Multiple functional copies of the RBM gene family, a spermatogenesis candidate on the human Y chromosome', *Genomics*. Academic Press Inc., 45(2), pp. 355–361. doi: 10.1006/geno.1997.4944.
Chang, M. H. *et al.* (2009) 'Decreased incidence of hepatocellular carcinoma in hepatitis B vaccinees: A 20-year follow-up study', *Journal of the National Cancer Institute*. J Natl Cancer Inst, 101(19), pp. 1348–1355. doi: 10.1093/jnci/djp288.
Chaturvedi, A. K. *et al.* (2011) 'Human papillomavirus and rising oropharyngeal cancer incidence in the United States', *Journal of Clinical Oncology*. American Society of Clinical Oncology, 29(32), pp. 4294–4301. doi: 10.1200/JCO.2011.36.4596.
Chen, X. and Calvisi, D. F. (2014) 'Hydrodynamic transfection for generation of novel mouse models for liver cancer research', *American Journal of Pathology*. Elsevier Inc., pp. 912–923. doi: 10.1016/j.ajpath.2013.12.002.
Cheng, E. Y. *et al.* (2009) 'Meiotic Recombination in Human Oocytes', *PLoS Genetics*. Edited by R. S. Hawley. Public Library of Science, 5(9), p. e1000661. doi: 10.1371/journal.pgen.1000661.
Cheung, H. H. *et al.* (2016) 'Hypermethylation of genes in testicular embryonal carcinomas', *British Journal of Cancer*. Nature Publishing Group, 114(2), pp. 230–236. doi: 10.1038/bjc.2015.408.
Chiba, T. *et al.* (2007) 'Enhanced Self-Renewal Capability in Hepatic Stem/Progenitor Cells Drives Cancer Initiation', *Gastroenterology*. W.B. Saunders, 133(3), pp. 937–950. doi: 10.1053/j.gastro.2007.06.016.
Chua, H. H. *et al.* (2015) 'RBMY, a novel inhibitor of glycogen synthase kinase 3β, increases tumor stemness and predicts poor prognosis of hepatocellular carcinoma', *Hepatology*, 62(5), pp. 1480–1496. doi: 10.1002/hep.27996.
Clocchiatti, A. *et al.* (2016) 'Sexual dimorphism in cancer', *Nature Reviews Cancer*. Nature Publishing Group, pp. 330–339. doi: 10.1038/nrc.2016.30.
Cloutier, J. M. and Turner, J. M. A. (2010) 'Meiotic sex chromosome inactivation',

Current Biology. Cell Press. doi: 10.1016/j.cub.2010.09.041.
Cognetti, D. M., Weber, R. S. and Lai, S. Y. (2008) 'Head and neck Cancer an evolving treatment paradigm', *Cancer*. NIH Public Access, pp. 1911–1932. doi: 10.1002/cncr.23654.
Cook, M. B. *et al.* (2009) 'Sex disparities in cancer incidence by period and age', *Cancer Epidemiology Biomarkers and Prevention*. NIH Public Access, 18(4), pp. 1174–1182. doi: 10.1158/1055-9965.EPI-08-1118.
D'Souza, G. and Dempsey, A. (2011) 'The role of HPV in head and neck cancer and review of the HPV vaccine', *Preventive Medicine*. NIH Public Access, p. S5. doi: 10.1016/j.ypmed.2011.08.001.
Davoli, T. *et al.* (2013) 'XCumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome', *Cell*. Cell Press, 155(4), p. 948. doi: 10.1016/j.cell.2013.10.011.
Decarpentrie, F. *et al.* (2012) 'Human and mouse ZFY genes produce a conserved testis-specific transcript encoding a zinc finger protein with a short acidic domain and modified transactivation potential', *Human Molecular Genetics*, 21(12), pp. 2631–2645. doi: 10.1093/hmg/dds088.
Degot, S. *et al.* (2004) 'Association of the breast cancer protein MLN51 with the exon junction complex via its speckle localizer and RNA binding module', *Journal of Biological Chemistry*, 279(32), pp. 33702–33715. doi: 10.1074/jbc.M402754200.
Dreumont, N. *et al.* (2010) 'Human RBMY regulates germline-specific splicing events by modulating the function of the serine/arginine-rich proteins 9G8 and Tra2-β', *Journal of Cell Science*. J Cell Sci, 123(1), pp. 40–50. doi: 10.1242/jcs.055889.
Ehrmann, I. *et al.* (2008) 'Haploinsufficiency of the germ cell-specific nuclear RNA binding protein hnRNP G-T prevents functional spermatogenesis in the mouse', *Human Molecular Genetics*. Hum Mol Genet, 17(18), pp. 2803–2818. doi: 10.1093/hmg/ddn179.
Ehrmann, I. *et al.* (2019) 'An ancient germ cell-specific RNA-binding protein protects the germline from cryptic splice site poisoning', *eLife*. NLM (Medline), 8. doi: 10.7554/eLife.39304.
Elliott, D. J. *et al.* (1997) 'Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 94(8), pp. 3848–3853. doi: 10.1073/pnas.94.8.3848.
Elliott, D. J. *et al.* (1998) 'Dynamic changes in the subnuclear organisation of pre-mRNA splicing proteins and RBM during human germ cell development.', *Journal of cell science*, 111 (Pt 9), pp. 1255–65. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/9547301 (Accessed: 29 October 2019).
Elliott, D J *et al.* (2000) 'A mammalian germ cell-specific RNA-binding protein interacts with ubiquitously expressed proteins involved in splice site selection.', *Proceedings of the National Academy of Sciences of the United States of America*, 97(11), pp. 5717–22. doi: 10.1073/pnas.97.11.5717.
Elliott, David J. *et al.* (2000) 'An evolutionarily conserved germ cell-specific hnRNP is encoded by a retrotransposed gene', *Human Molecular Genetics*. Oxford University Press, 9(14), pp. 2117–2124. doi: 10.1093/hmg/9.14.2117.
Elliott, D. J. (2000) 'RBMY genes and AZFb deletions', *Journal of Endocrinological Investigation*, 23(10), pp. 652–658. doi: 10.1007/BF03343789.
Elliott, David J (2004) 'The role of potential splicing factors including RBMY, RBMX, hnRNPG-T and STAR proteins in spermatogenesis.', *International journal of andrology*, 27(6), pp. 328–34. doi: 10.1111/j.1365-2605.2004.00496.x.
Elliott, David J. (2004) 'The role of potential splicing factors including RBMY, RBMX, hnRNPG-T and STAR proteins in spermatogenesis', *International Journal of Andrology*, 27(6), pp. 328–334. doi: 10.1111/j.1365-2605.2004.00496.x.
Fakhry, C. *et al.* (2018) 'Head and neck squamous cell cancers in the United States are rare and the risk now is higher among white individuals compared with black individuals', *Cancer*. John Wiley and Sons Inc., 124(10), pp. 2125–2133. doi: 10.1002/cncr.31322.
Fang, J. *et al.* (2012a) 'Knockdown of zinc finger protein, X-linked (ZFX) inhibits cell proliferation and induces apoptosis in human laryngeal squamous cell carcinoma', *Molecular and Cellular Biochemistry*. Mol Cell Biochem, 360(1–2), pp. 301–307. doi: 10.1007/s11010-011-1069-x.
Fang, J. *et al.* (2012b) 'Knockdown of zinc finger protein, X-linked (ZFX) inhibits cell proliferation and induces apoptosis in human laryngeal squamous cell carcinoma', *Molecular and Cellular Biochemistry*. Springer, 360(1–2), pp. 301–307. doi: 10.1007/s11010-011-1069-x.
Fang, Q. *et al.* (2014) 'Knockdown of ZFX suppresses renal carcinoma cell growth and induces apoptosis', *Cancer Genetics*. Elsevier Inc., 207(10–12), pp. 461–466. doi: 10.1016/j.cancergen.2014.08.007.
Fernandez-Capetillo, O. *et al.* (2003) 'H2AX is required for chromatin remodeling and inactivation of sex chromosomes in male mouse meiosis.', *Developmental cell*, 4(4), pp. 497–508. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12689589 (Accessed: 2 October 2019).
Forsberg, L. A. *et al.* (2014) 'Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer.', *Nature genetics*, 46(6), pp.

624–8. doi: 10.1038/ng.2966.
Gao, J. *et al.* (2013) 'Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal', *Science Signaling*. Sci Signal, 6(269). doi: 10.1126/scisignal.2004088.
Gillison, M. L. *et al.* (2015) 'Epidemiology of human papillomavirus-positive head and neck squamous cell carcinoma', *Journal of Clinical Oncology*. American Society of Clinical Oncology, pp. 3235–3242. doi: 10.1200/JCO.2015.61.6995.
Gorbsky, G. J. (2015) 'The spindle checkpoint and chromosome segregation in meiosis.', *The FEBS journal*, 282(13), pp. 2471–87. doi: 10.1111/febs.13166.
Gozdecka, M. *et al.* (2018) 'UTX-mediated enhancer and chromatin remodeling suppresses myeloid leukemogenesis through noncatalytic inverse regulation of ETS and GATA programs', *Nature Genetics*. Nature Publishing Group, 50(6), pp. 883–894. doi: 10.1038/s41588-018-0114-z.
Hashibe, M. *et al.* (2009) 'Interaction between tobacco and alcohol use and the risk of head and neck cancer: Pooled analysis in the international head and neck cancer Epidemiology consortium', *Cancer Epidemiology Biomarkers and Prevention*. Cancer Epidemiol Biomarkers Prev, 18(2), pp. 541–550. doi: 10.1158/1055-9965.EPI-08-0347.
Heinrich, B. *et al.* (2009) 'Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regionsin pre-mRNA', *Journal of Biological Chemistry*. J Biol Chem, 284(21), pp. 14303–14315. doi: 10.1074/jbc.M901026200.
Hep 3B2.1-7 [Hep 3B, Hep-3B, Hep3B] ATCC ® HB-8064™ Homo sap (no date). Available at: https://www.lgcstandards-atcc.org/products/all/HB-8064.aspx (Accessed: 14 October 2020).
Hep G2 [HEPG2] ATCC® HB-8065™ (no date). Available at: https://www.lgcstandards-atcc.org/products/all/HB-8065.aspx?geo_country=gb (Accessed: 14 October 2020).
Hollows, R. *et al.* (2019) 'Association between loss of Y chromosome and poor prognosis in male head and neck squamous cell carcinoma', *Head & Neck*, 41(4), pp. 993–1006. doi: 10.1002/hed.25537.
Hu, J. *et al.* (2016) 'Co-activation of AKT and c-Met triggers rapid hepatocellular carcinoma development via the mTORC1/FASN pathway in mice', *Scientific Reports*. Nature Publishing Group, 6. doi: 10.1038/srep20484.
HuH-7 (no date). Available at: https://www.ebi.ac.uk/ols/ontologies/efo/terms?short_form=EFO_0005384 (Accessed: 14 October 2020).
Hunt, P. A. and Hassold, T. J. (2002) 'Sex matters in meiosis.', *Science (New York, N.Y.)*, 296(5576), pp. 2181–3. doi: 10.1126/science.1071907.
Hutter, C. and Zenklusen, J. C. (2018) 'The Cancer Genome Atlas: Creating Lasting

Value beyond Its Data', *Cell*. Cell Press, pp. 283–285. doi: 10.1016/j.cell.2018.03.042.
Inada, M. *et al.* (2008) 'Stage-specific regulation of adhesion molecule expression segregates epithelial stem/progenitor cells in fetal and adult human livers', *Hepatology International*. Hepatol Int, 2(1), pp. 50–62. doi: 10.1007/s12072-007-9023-4.
Jangravi, Z. *et al.* (2013) 'A fresh look at the male-specific region of the human y chromosome', *Journal of Proteome Research*. American Chemical Society, pp. 6–22. doi: 10.1021/pr300864k.
Jiang, H. *et al.* (2012) 'Knockdown of zinc finger protein X-linked inhibits prostate cancer cell proliferation and induces apoptosis by activating caspase-3 and caspase-9', *Cancer Gene Therapy*. Nature Publishing Group, 19(10), pp. 684–689. doi: 10.1038/cgt.2012.53.
Jiang, J. and Liu, L. Y. (2015) 'Zinc finger protein X-linked is overexpressed in colorectal cancer and is associated with poor prognosis', *Oncology Letters*. Spandidos Publications, 10(2), pp. 810–814. doi: 10.3892/ol.2015.3353.
Kadekar, S. *et al.* (2012) 'Gender differences in chemical carcinogenesis in national toxicology program 2-year bioassays', *Toxicologic Pathology*. NIH Public Access, 40(8), pp. 1160–1168. doi: 10.1177/0192623312446527.
Kanda, T. *et al.* (2008) 'Hepatitis C Virus Core Protein Augments Androgen Receptor-Mediated Signaling', *Journal of Virology*. American Society for Microbiology, 82(22), pp. 11066–11072. doi: 10.1128/jvi.01300-08.
Kemp, C. J., Leary, C. N. and Drinkwater, N. R. (1989) 'Promotion of murine hepatocarcinogenesis by testosterone is androgen receptor-dependent but not cell autonomous', *Proceedings of the National Academy of Sciences of the United States of America*, 86(19), pp. 7505–7509. doi: 10.1073/pnas.86.19.7505.
Kido, T. *et al.* (2014) 'The potential contributions of a Y-located protooncogene and its X homologue in sexual dimorphisms in hepatocellular carcinoma', *Human Pathology*. Hum Pathol, 45(9), pp. 1847–1858. doi: 10.1016/j.humpath.2014.05.002.
Kido, T. *et al.* (2020) 'Potential dual functional roles of the Y-linked RBMY in hepatocarcinogenesis', *Cancer Science*. Blackwell Publishing Ltd, 111(8), pp. 2987–2999. doi: 10.1111/cas.14506.
Kido, T. and Lau, Y. F. (2015) 'Roles of the Y chromosome genes in human cancers', *Asian Journal of Andrology*. Medknow Publications, pp. 373–380. doi: 10.4103/1008-682X.150842.
Kido, T. and Lau, Y. F. C. (2016) 'Identification of a TSPY co-expression network associated with DNA hypomethylation and tumor gene expression in somatic cancers', *Journal of Genetics and Genomics*. Institute of Genetics and Developmental Biology, 43(10), pp. 577–585. doi: 10.1016/j.jgg.2016.09.003.

Koopman, P, Ashworth, A. and Lovell-Badge, R. (1991) 'The ZFY gene family in humans and mice.', *Trends in genetics : TIG*, 7(4), pp. 132–6. doi: 10.1016/0168-9525(91)90458-3.
Koopman, Peter, Ashworth, A. and Lovell-Badge, R. (1991) 'The ZFY gene family in humans and mice', *Trends in Genetics*. Elsevier Current Trends, pp. 132–136. doi: 10.1016/0168-9525(91)90458-3.
Kurahashi, H. *et al.* (2012) 'Failure of homologous synapsis and sex-specific reproduction problems.', *Frontiers in genetics*, 3, p. 112. doi: 10.3389/fgene.2012.00112.
Kurian, K. M., Watson, C. J. and Wyllie, A. H. (2000) 'Retroviral vectors', *Journal of Clinical Pathology - Molecular Pathology*. BMJ Publishing Group, pp. 173–176. doi: 10.1136/mp.53.4.173.
Lau, Y. F. C. and Chan, K. (1989) 'The putative testis-determining factor and related genes are expressed as discrete-sized transcripts in adult gonadal and somatic tissues', *American Journal of Human Genetics*. Elsevier, 45(6), pp. 942–952. Available at: /pmc/articles/PMC1683462/?report=abstract (Accessed: 8 October 2020).
Lawrence, M. S. *et al.* (2015) 'Comprehensive genomic characterization of head and neck squamous cell carcinomas', *Nature*. Nature Publishing Group, 517(7536), pp. 576–582. doi: 10.1038/nature14129.
Lee, J. S. *et al.* (2006) 'A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells', *Nature Medicine*. Nat Med, 12(4), pp. 410–416. doi: 10.1038/nm1377.
Leemans, C. R., Snijders, P. J. F. and Brakenhoff, R. H. (2018) 'The molecular landscape of head and neck cancer', *Nature Reviews Cancer*. Nature Publishing Group, pp. 269–282. doi: 10.1038/nrc.2018.11.
LeMaire-Adkins, R., Radke, K. and Hunt, P. A. (1997) 'Lack of checkpoint control at the metaphase/anaphase transition: a mechanism of meiotic nondisjunction in mammalian females.', *The Journal of cell biology*, 139(7), pp. 1611–9. doi: 10.1083/jcb.139.7.1611.
Li, C. *et al.* (2015) 'ZFX is a strong predictor of poor prognosis in renal cell carcinoma', *Medical Science Monitor*. International Scientific Literature Inc., 21, pp. 3380–3385. doi: 10.12659/MSM.894708.
Li, K. *et al.* (2013) 'ZFX knockdown inhibits growth and migration of non-small cell lung carcinoma cell line H1299', *International Journal of Clinical and Experimental Pathology*. e-Century Publishing Corporation, 6(11), pp. 2460–2467. Available at: www.ijcep.com/ (Accessed: 12 October 2020).
Liu, Y. *et al.* (2009) 'The germ cell nuclear proteins hnRNP G-T and RBMY activate a testis-specific exon', *PLoS Genetics*, 5(11). doi: 10.1371/journal.pgen.1000707.

Lynn, A. *et al.* (2002) 'Covariation of synaptonemal complex length and mammalian meiotic exchange rates', *Science*. American Association for the Advancement of Science, 296(5576), pp. 2222–2225. doi: 10.1126/science.1071220.
Ma, K. *et al.* (1993) 'A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis.', *Cell*, 75(7), pp. 1287–95. doi: 10.1016/0092-8674(93)90616-x.
Maan, A. A. *et al.* (2017) 'The y chromosome: A blueprint for men's health?', *European Journal of Human Genetics*. Nature Publishing Group, pp. 1181–1188. doi: 10.1038/ejhg.2017.128.
Mahadevaiah, S. K. *et al.* (1998) 'Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities', *Human Molecular Genetics*, 7(4), pp. 715–727. doi: 10.1093/hmg/7.4.715.
Manning, B. D. and Cantley, L. C. (2007) 'AKT/PKB Signaling: Navigating Downstream', *Cell*. Cell, pp. 1261–1274. doi: 10.1016/j.cell.2007.06.009.
Mardon, G. *et al.* (1990) 'Mouse Zfx protein is similar to Zfy-2: each contains an acidic activating domain and 13 zinc fingers.', *Molecular and cellular biology*, 10(2), pp. 681–8. doi: 10.1128/mcb.10.2.681.
Matter, N., Herrlich, P. and König, H. (2002) 'Signal-dependent regulation of splicing via phosphorylation of Sam68', *Nature*. Nature Publishing Group, 420(6916), pp. 691–695. doi: 10.1038/nature01153.
Mokkapati, S. *et al.* (2014) 'β-Catenin activation in a novel liver progenitor cell type is sufficient to cause hepatocellular carcinoma and hepatoblastoma', *Cancer Research*. American Association for Cancer Research Inc., 74(16), pp. 4515–4525. doi: 10.1158/0008-5472.CAN-13-3275.
Monarch ® Nucleic Acid Purification (no date).
Monesi, V. (1965) 'Differential rate of ribonucleic acid synthesis in the autosomes and sex chromosomes during male meiosis in the mouse.', *Chromosoma*, 17(1), pp. 11–21. doi: 10.1007/bf00285153.
Mourad, M. *et al.* (2017) 'Epidemiological Trends of Head and Neck Cancer in the United States: A SEER Population Study', *Journal of Oral and Maxillofacial Surgery*. W.B. Saunders, 75(12), pp. 2562–2572. doi: 10.1016/j.joms.2017.05.008.
Murr, R. *et al.* (2007) 'Orchestration of chromatin-based processes: mind the TRRAP.', *Oncogene*, 26(37), pp. 5358–72. doi: 10.1038/sj.onc.1210605.
Mutter1', G. L., Grills', G. S. and Wolgemuth1', D. J. (1988) *Evidence for the involvement of the proto-oncogene c-mos in mammalian meiotic maturation and possibly very early embryogenesis*, *The EMBO Journal*.

Nagai, K. *et al.* (1995) 'The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA', *Trends in Biochemical Sciences*, pp. 235–240. doi: 10.1016/S0968-0004(00)89024-6.
Nagamine, C. M. *et al.* (1989) 'Chromosome mapping and expression of a putative testis-determining gene in mouse.', *Science (New York, N.Y.)*, 243(4887), pp. 80–3. doi: 10.1126/science.2563174.
Nagaoka, S. I., Hassold, T. J. and Hunt, P. A. (2012) 'Human aneuploidy: mechanisms and new insights into an age-old problem.', *Nature reviews. Genetics*, 13(7), pp. 493–504. doi: 10.1038/nrg3245.
Nakasuji, T. *et al.* (2017) 'Complementary Critical Functions of Zfy1 and Zfy2 in Mouse Spermatogenesis and Reproduction', *PLOS Genetics*. Edited by W. Yan. Public Library of Science, 13(1), p. e1006578. doi: 10.1371/journal.pgen.1006578.
Naugler, W. E. *et al.* (2007) 'Gender disparity in liver cancer due to sex differences in MyD88-dependent IL-6 production', *Science*. American Association for the Advancement of Science, 317(5834), pp. 121–124. doi: 10.1126/science.1140485.
Nielsen, A. and Gjerstorff, M. (2016) 'Ectopic Expression of Testis Germ Cell Proteins in Cancer and Its Potential Role in Genomic Instability', *International Journal of Molecular Sciences*. MDPI AG, 17(6), p. 890. doi: 10.3390/ijms17060890.
Nikpour, P. *et al.* (2012) 'Differential expression of ZFX gene in gastric cancer', *Journal of Biosciences*. J Biosci, 37(1), pp. 85–90. doi: 10.1007/s12038-011-9174-2.
Page, D. C. *et al.* (1987) 'The sex-determining region of the human Y chromosome encodes a finger protein.', *Cell*, 51(6), pp. 1091–104. doi: 10.1016/0092-8674(87)90595-2.
Palmer, C. J. *et al.* (2014) 'Zfx facilitates tumorigenesis caused by activation of the Hedgehog pathway', *Cancer Research*. American Association for Cancer Research Inc., 74(20), pp. 5914–5924. doi: 10.1158/0008-5472.CAN-14-0834.
Pan, D. (2010) 'The hippo signaling pathway in development and cancer', *Developmental Cell*. Dev Cell, pp. 491–505. doi: 10.1016/j.devcel.2010.09.011.
Pandini, G. *et al.* (2005) 'Androgens up-regulate the insulin-like growth factor-I receptor in prostate cancer cells', *Cancer Research*, 65(5), pp. 1849–1857. doi: 10.1158/0008-5472.CAN-04-1837.
Peng, L. *et al.* (2006) 'Growth inhibitory concentrations of androgens up-regulate insulin-like growth factor binding protein-3 expression via an androgen response element in LNCaP human prostate cancer cells', *Endocrinology*, 147(10), pp. 4599–4607. doi: 10.1210/en.2006-0560.
Quintana-Murci, L. and Fellous, M. (2001) 'The human Y chromosome: The biological role of a "functional wasteland"', *Journal of Biomedicine and Biotechnology*. Hindawi

Limited, pp. 18–24. doi: 10.1155/S1110724301000080.
Rajan, P. *et al.* (2008) 'The RNA-binding and adaptor protein Sam68 modulates signal-dependent splicing and transcriptional activity of the androgen receptor', *The Journal of Pathology*. John Wiley & Sons, Ltd, 215(1), pp. 67–77. doi: 10.1002/path.2324.
Rhie, S. K. *et al.* (2018) 'ZFX acts as a transcriptional activator in multiple types of human tumors by binding downstream from transcription start sites at the majority of CpG island promoters', *Genome Research*. Cold Spring Harbor Laboratory Press, 28(3), pp. 310–320. doi: 10.1101/gr.228809.117.
Ritchie Hannah (2019) *Who smokes more, men or women? - Our World in Data*, *Our World Data*. Available at: https://ourworldindata.org/who-smokes-more-men-or-women (Accessed: 29 September 2020).
Royo, H. *et al.* (2010) 'Evidence that meiotic sex chromosome inactivation is essential for male fertility.', *Current biology : CB*, 20(23), pp. 2117–23. doi: 10.1016/j.cub.2010.11.010.
Schmid, R. *et al.* (2013) 'The splicing landscape is globally reprogrammed during male meiosis', *Nucleic Acids Research*. Oxford University Press, 41(22), pp. 10170–10184. doi: 10.1093/nar/gkt811.
Schneider-Gädicke, A. *et al.* (1989) 'ZFX has a gene structure similar to ZFY, the putative human sex determinant, and escapes X inactivation.', *Cell*, 57(7), pp. 1247–58. doi: 10.1016/0092-8674(89)90061-5.
Shen, L., Shi, Q. and Wang, W. (2018) 'Double agents: genes with both oncogenic and tumor-suppressor functions', *Oncogenesis*. Springer Nature, 7(3). doi: 10.1038/s41389-018-0034-x.
Shimmin, L. C., Chang, B. H. J. and Li, W. H. (1994) 'Contrasting rates of nucleotide substitution in the X-Linked and Y-Linked zinc finger genes', *Journal of Molecular Evolution*. J Mol Evol, 39(6), pp. 569–578. doi: 10.1007/BF00160402.
Skaletsky, H. *et al.* (2003) 'The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes', *Nature*. Nature Publishing Group, 423(6942), pp. 825–837. doi: 10.1038/nature01722.
Skrisovska, L. *et al.* (2007) 'The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction', *EMBO Reports*, 8(4), pp. 372–379. doi: 10.1038/sj.embor.7400910.
St Guily, J. L. *et al.* (2011) 'Human papillomavirus genotype distribution in tonsil cancers', *Head and Neck Oncology*. BioMed Central, 3(1), p. 6. doi: 10.1186/1758-3284-3-6.
Sturgis, E. M. and Cinciripini, P. M. (2007) 'Trends in head and neck cancer incidence in relation to smoking prevalence: An emerging epidemic of human papillomavirus-

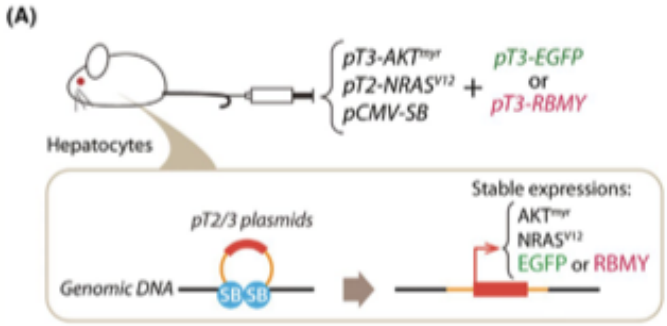
associated cancers?', *Cancer*. Cancer, pp. 1429–1435. doi: 10.1002/cncr.22963.
Tacke, R. *et al.* (1998) 'Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing', *Cell*. Cell Press, 93(1), pp. 139–148. doi: 10.1016/S0092-8674(00)81153-8.
Tang, A. L. *et al.* (2012) 'UM-SCC-104: A New human papillomavirus-16-positive cancer stem cell-containing head and neck squamous cell carcinoma cell line', *Head and Neck*. NIH Public Access, 34(10), pp. 1480–1491. doi: 10.1002/hed.21962.
Tissue expression of ZFY - Summary - The Human Protein Atlas (no date). Available at: https://www.proteinatlas.org/ENSG00000067646-ZFY/tissue (Accessed: 14 August 2020).
Traven, A., Jelicic, B. and Sopta, M. (2006) 'Yeast Gal4: a transcriptional paradigm revisited.', *EMBO reports*, 7(5), pp. 496–9. doi: 10.1038/sj.embor.7400679.
Tricoli, J. V. and Bracken, R. B. (1993) 'ZFY gene expression and retention in human prostate adenocarcinoma', *Genes, Chromosomes and Cancer*. John Wiley & Sons, Ltd, 6(2), pp. 65–72. doi: 10.1002/gcc.2870060202.
Trujilo, B. C. (2019) *RBMY and ZFY: Identification of two testis- specific genes in HPV-oropharyngeal squamous cell carcinomas.* University of Kent, School of Biosciences.
Tsuei, D. J. *et al.* (2004) 'RBMY, a male germ cell-specific RNA-binding protein, activated in human liver cancers and transforms rodent fibroblasts', *Oncogene*, 23(34), pp. 5815–5822. doi: 10.1038/sj.onc.1207773.
Tsuei, D. J. *et al.* (2011) 'Male germ cell-specific RNA binding protein RBMY: A new oncogene explaining male predominance in liver cancer', *PLoS ONE*. PLoS One, 6(11). doi: 10.1371/journal.pone.0026948.
Turner, J. M. A. (2007) 'Meiotic sex chromosome inactivation', *Development*. Development, 134(10), pp. 1823–1831. doi: 10.1242/dev.000018.
Turner, J. M. A. (2015) 'Meiotic Silencing in Mammals', *Annual Review of Genetics*. Annual Reviews Inc., 49, pp. 395–412. doi: 10.1146/annurev-genet-112414-055145.
Varmus, H. E. (1984) *THE MOLECULAR GENETICS OF CELLULAR ONCOGENES*. Available at: www.annualreviews.org (Accessed: 6 October 2020).
Venables, J. P. (2000) 'RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing', *Human Molecular Genetics*, 9(5), pp. 685–694. doi: 10.1093/hmg/9.5.685.
Venables, J. P. (2002) 'Alternative splicing in the testes.', *Current opinion in genetics & development*, 12(5), pp. 615–19. doi: 10.1016/s0959-437x(02)00347-7.
Venables, J. P. and Eperon, I. (1999) 'The roles of RNA-binding proteins in spermatogenesis and male infertility.', *Current opinion in genetics & development*, 9(3), pp. 346–54. doi: 10.1016/s0959-437x(99)80052-5.

Vernet, Nadège *et al.* (2011) 'The Y-encoded gene zfy2 acts to remove cells with unpaired chromosomes at the first meiotic metaphase in male mice.', *Current biology : CB*, 21(9), pp. 787–93. doi: 10.1016/j.cub.2011.03.057.
Vernet, Nadge *et al.* (2011) 'The Y-encoded gene Zfy2 acts to remove cells with unpaired chromosomes at the first meiotic metaphase in male mice', *Current Biology*, 21(9), pp. 787–793. doi: 10.1016/j.cub.2011.03.057.
Vernet, N. *et al.* (2014) 'Mouse Y-Linked Zfy1 and Zfy2 Are Expressed during the Male-Specific Interphase between Meiosis I and Meiosis II and Promote the 2nd Meiotic Division', *PLoS Genetics*, 10(6). doi: 10.1371/journal.pgen.1004444.
Vernet, Nadege *et al.* (2016) 'Mouse Y-encoded transcription factor Zfy2 is essential for sperm head remodelling and sperm tail development', *PLoS ONE*. Public Library of Science, 11(1). doi: 10.1371/journal.pone.0145398.
Vernet, Nadége *et al.* (2016) 'Zfy genes are required for efficient meiotic sex chromosome inactivation (MSCI) in spermatocytes', *Human Molecular Genetics*, 25(24), pp. 5300–5310. doi: 10.1093/hmg/ddw344.
Vigneswaran, N. and Williams, M. D. (2014) 'Epidemiologic trends in head and neck cancer and aids in diagnosis', *Oral and Maxillofacial Surgery Clinics of North America*. W.B. Saunders, pp. 123–141. doi: 10.1016/j.coms.2014.01.001.
Vogt, P. H. *et al.* (1996) 'Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11', *Human Molecular Genetics*, 5(7), pp. 933–943. doi: 10.1093/hmg/5.7.933.
Warnakulasuriya, S. (2009) 'Global epidemiology of oral and oropharyngeal cancer', *Oral Oncology*. Oral Oncol, pp. 309–316. doi: 10.1016/j.oraloncology.2008.06.002.
Weng, H. *et al.* (2015) 'Zinc finger X-chromosomal protein (ZFX) is a significant prognostic indicator and promotes cellular malignant potential in gallbladder cancer', *Cancer Biology and Therapy*. Taylor and Francis Inc., 16(10), pp. 1462–1470. doi: 10.1080/15384047.2015.1070994.
Williams, D. A., Xu, H. and Cancelas, J. A. (2006) 'Children are not little adults: Just ask their hematopoietic stem cells', *Journal of Clinical Investigation*. J Clin Invest, pp. 2593–2596. doi: 10.1172/JCI30083.
Wojtasz, L. *et al.* (2012) 'Meiotic DNA double-strand breaks and chromosome asynapsis in mice are monitored by distinct HORMAD2-independent and -dependent mechanisms.', *Genes & development*, 26(9), pp. 958–73. doi: 10.1101/gad.187559.112.
Wondergem, N. E. *et al.* (2020) 'The Immune Microenvironment in Head and Neck Squamous Cell Carcinoma: on Subsets and Subsites', *Current Oncology Reports*. Springer, 22(8), pp. 1–14. doi: 10.1007/s11912-020-00938-3.

Wong, M. C. S. *et al.* (2017) 'International incidence and mortality trends of liver cancer: A global profile', *Scientific Reports*. Nature Publishing Group, 7. doi: 10.1038/srep45846.
Wu, S. *et al.* (2013) 'Knockdown of ZFX inhibits gastric cancer cell growth in vitro and in vivo via downregulating the ERK-MAPK pathway', *Cancer Letters*. Cancer Lett, 337(2), pp. 293–300. doi: 10.1016/j.canlet.2013.04.003.
Yamashita, T. and Wang, X. W. (2013) 'Cancer stem cells in the development of liver cancer', *Journal of Clinical Investigation*. J Clin Invest, pp. 1911–1918. doi: 10.1172/JCI66024.
Yamauchi, Y. *et al.* (2015) 'Mouse Y-Encoded Transcription Factor Zfy2 Is Essential for Sperm Formation and Function in Assisted Fertilization', *PLOS Genetics*. Edited by M. S. Bartolomei. Public Library of Science, 11(12), p. e1005476. doi: 10.1371/journal.pgen.1005476.
Yan, X. *et al.* (2016) 'High expression of Zinc-finger protein X-linked promotes tumor growth and predicts a poor outcome for stage II/III colorectal cancer patients', *Oncotarget*. Impact Journals LLC, 7(15), pp. 19680–19692. doi: 10.18632/oncotarget.7547.
Yang, F. *et al.* (2015) 'Zinc finger protein x-linked (ZFX) contributes to patient prognosis, cell proliferation and apoptosis in human laryngeal squamous cell carcinoma', *International Journal of Clinical and Experimental Pathology*. E-Century Publishing Corporation, 8(11), pp. 13886–13899. Available at: www.ijcep.com/ (Accessed: 12 October 2020).
Yang, H. *et al.* (2014) 'shRNA-mediated silencing of ZFX attenuated the proliferation of breast cancer cells', *Cancer Chemotherapy and Pharmacology*. Springer Verlag, 73(3), pp. 569–576. doi: 10.1007/s00280-014-2379-y.
Yang, J. D. *et al.* (2019) 'A global view of hepatocellular carcinoma: trends, risk, prevention and management', *Nature Reviews Gastroenterology and Hepatology*. Nature Publishing Group, pp. 589–604. doi: 10.1038/s41575-019-0186-y.
Yang, S. and Liu, G. (2017) 'Targeting the RAS/RAF/MEK/ERK pathway in hepatocellular carcinoma', *Oncology Letters*. Spandidos Publications, pp. 1041–1047. doi: 10.3892/ol.2017.5557.
Yin, J. *et al.* (2015) 'Overexpression of ZFX and its involvement in squamous cell carcinoma of the tongue', *Oncology Reports*. Spandidos Publications, 33(1), pp. 141–148. doi: 10.3892/or.2014.3572.
Zakeri, Z. F., Ponzetto, C. and Wolgemuth, D. J. (1988) 'Translational regulation of the novel haploid-specific transcripts for the c-abl proto-oncogene and a member of the 70 kDa heat-shock protein gene family in the male germ line', *Developmental Biology*.

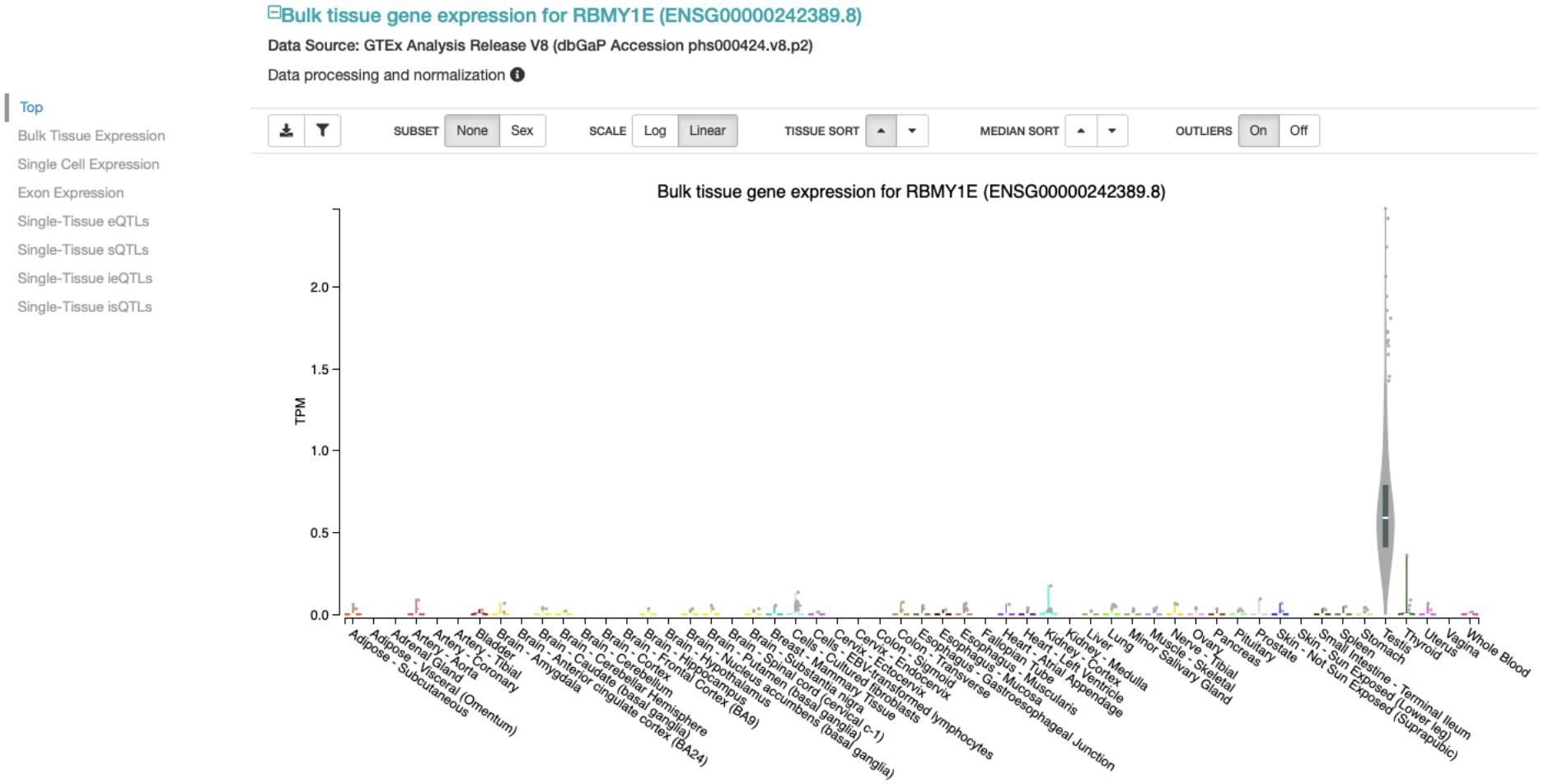
Academic Press, 125(2), pp. 417–422. doi: 10.1016/0012-1606(88)90222-9.
Zhou, Y. *et al.* (2011) 'The Zfx gene is expressed in human gliomas and is important in the proliferation and apoptosis of the human malignant glioma cell line U251', *Journal of Experimental and Clinical Cancer Research*. J Exp Clin Cancer Res, 30(1). doi: 10.1186/1756-9966-30-114.
Zhu, Z. *et al.* (2013) 'ZFX regulates glioma cell proliferation and survival in vitro and in vivo', *Journal of Neuro-Oncology*. J Neurooncol, 112(1), pp. 17–25. doi: 10.1007/s11060-012-1032-z.
Zulehner, G. *et al.* (2010) 'Nuclear β-catenin induces an early liver progenitor phenotype in hepatocellular carcinoma and promotes tumor recurrence', *American Journal of Pathology*. Elsevier Inc., 176(1), pp. 472–481. doi: 10.2353/ajpath.2010.090300.

6. **Supplementary Data**

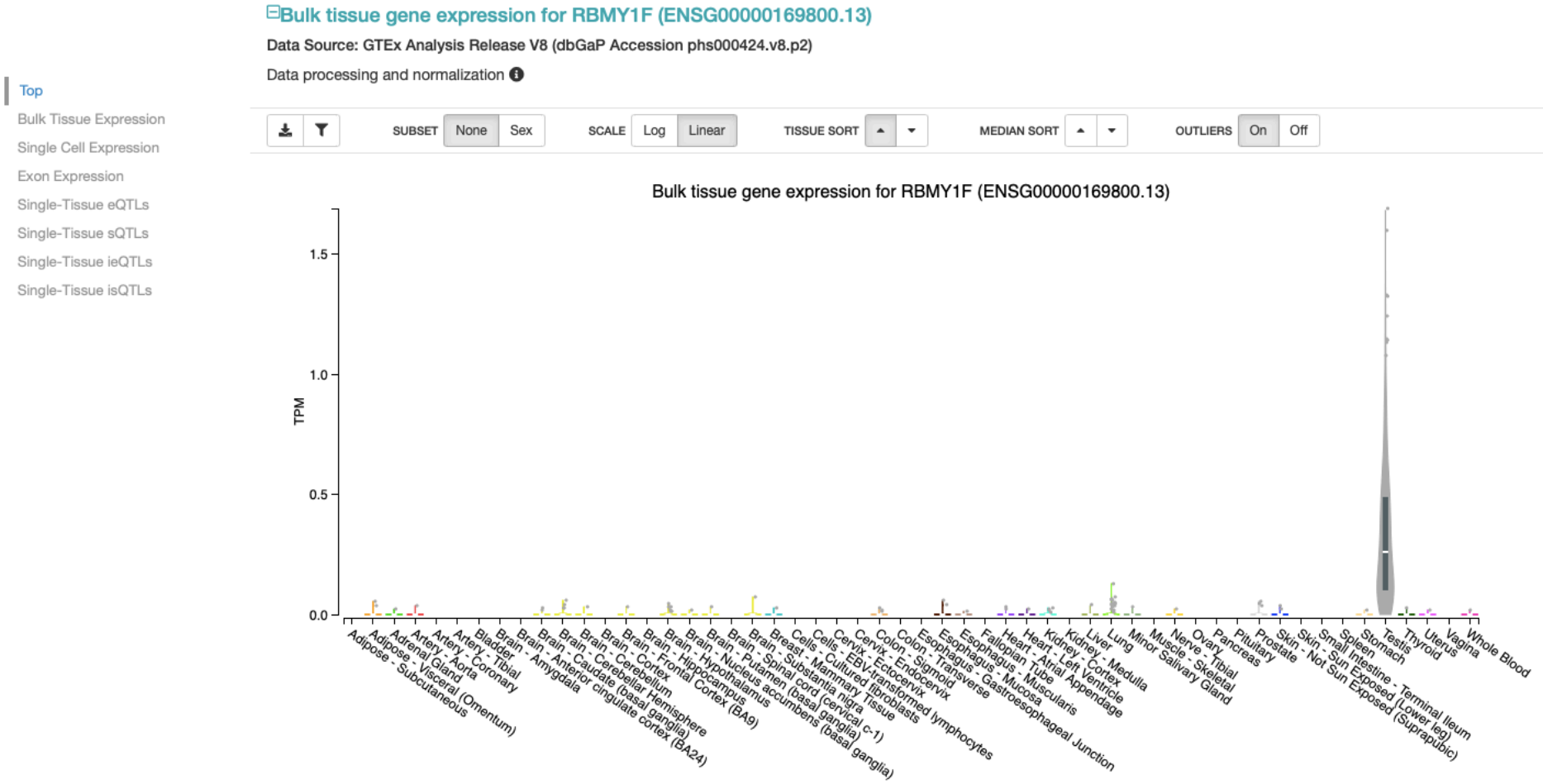


Supplementary Figure 1: Hydrodynamic tail co-injection of mice to create desired mouse models for the study. Hydrodynamic tail injection was utilised to create AKT and Ras induced hepatic cancer mouse models to study the in vivo effects of RBMY overexpression. Using this technique, Injected DNA primarily targets the mouse recipient liver(Kido *et al.*, 2020). Efficient and stable integration of these DNAs was achieved by using direct repeat sequences as flanks as well as co-injection of target DNA with sleeping beauty transposase expression cassette(Chen and Calvisi, 2014)(Hu *et al.*, 2016). To create the mouse models for this study, expression vectors carrying AKT (pT3-AKT^{myr}) and NRAS (pT2-NRAS^{V12}) oncogenes with SB vector (pCMV-SB) were utilised to create the desired mouse models: by co-injecting the expression vectors for the experimental models, 2 mouse models were designed, a RBMY mouse model carrying pT3-RBMY vector (AKT^{myr}/NRAS^{V12}/RBMY mice) and a AKT/Ras positive control mouse model with pT3-EGFP expression vector (AKT^{myr}/NRAS^{V12}/EGFP mice)(Kido *et al.*, 2020).

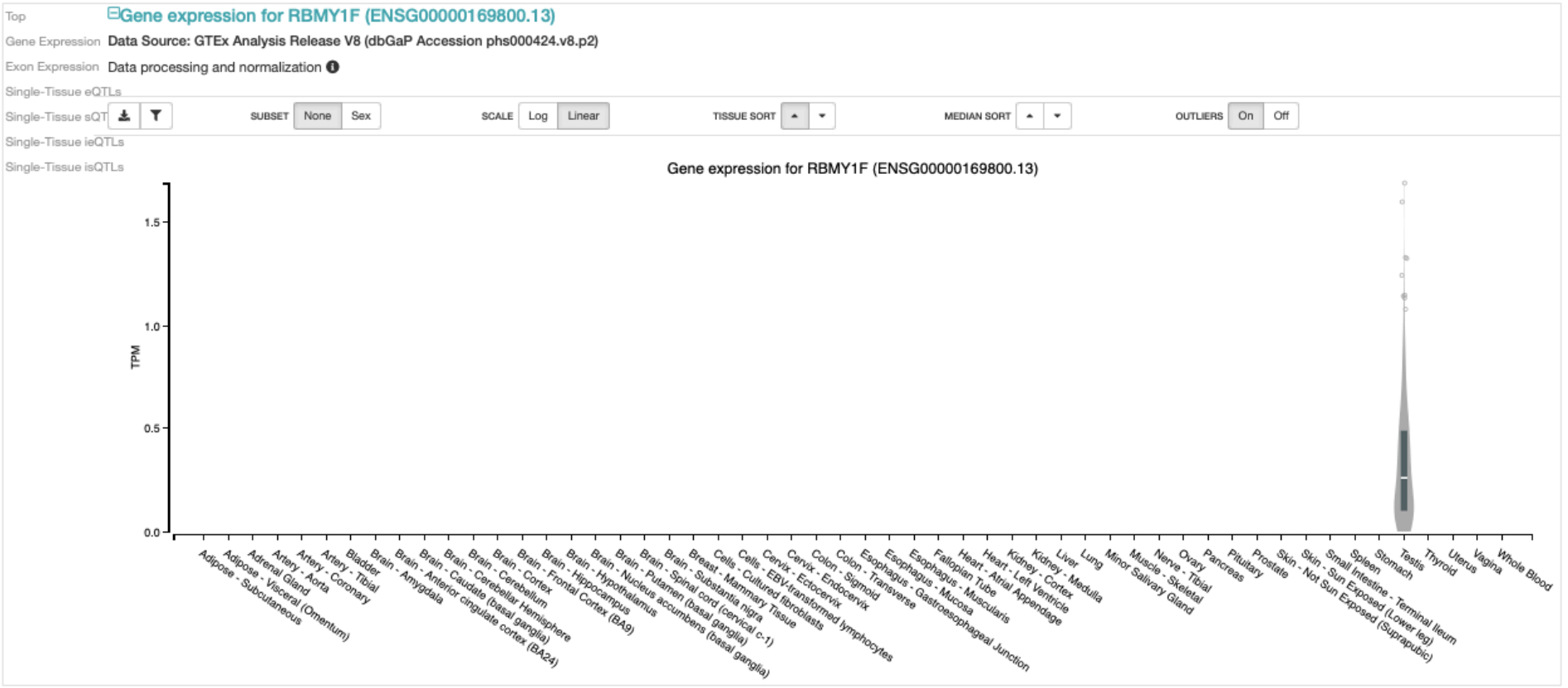
A. RBMY1E


Bulk tissue gene expression for RBMY1E (ENSG00000242389.8)
Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)
Data processing and normalization
Top
Bulk Tissue Expression
Single Cell Expression
Exon Expression
Single-Tissue eQTLs
Single-Tissue sQTLs
Single-Tissue ieQTLs
Single-Tissue isQTLs
SUBSET None Sex
SCALE Log Linear
TISSUE SORT
MEDIAN SORT
OUTLIERS On Off
Bulk tissue gene expression for RBMY1E (ENSG00000242389.8)
TPM
2.0
1.5
1.0
0.5
0.0
Adipose - Subcutaneous
Adipose - Visceral (Omentum)
Adrenal Gland
Artery - Aorta
Artery - Coronary
Artery - Tibial
Bladder
Brain - Amygdala
Brain - Anterior cingulate cortex (BA24)
Brain - Caudate (basal ganglia)
Brain - Cerebellar Hemisphere
Brain - Cerebellum
Brain - Cortex
Brain - Frontal Cortex (BA9)
Brain - Hippocampus
Brain - Hypothalamus
Brain - Nucleus accumbens (basal ganglia)
Brain - Putamen (basal ganglia)
Brain - Spinal cord (cervical c-1)
Brain - Substantia nigra
Breast - Mammary Tissue
Cells - Cultured fibroblasts
Cells - EBV-transformed lymphocytes
Cervix - Ectocervix
Cervix - Endocervix
Colon - Sigmoid
Colon - Transverse
Esophagus - Gastroesophageal Junction
Esophagus - Mucosa
Esophagus - Muscularis
Fallopian Tube
Heart - Atrial Appendage
Heart - Left Ventricle
Kidney - Cortex
Kidney - Medulla
Liver
Lung
Minor Salivary Gland
Muscle - Skeletal
Nerve - Tibial
Ovary
Pancreas
Pituitary
Prostate
Skin - Not Sun Exposed (Suprapubic)
Skin - Sun Exposed (Lower leg)
Small Intestine - Terminal Ileum
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
Whole Blood

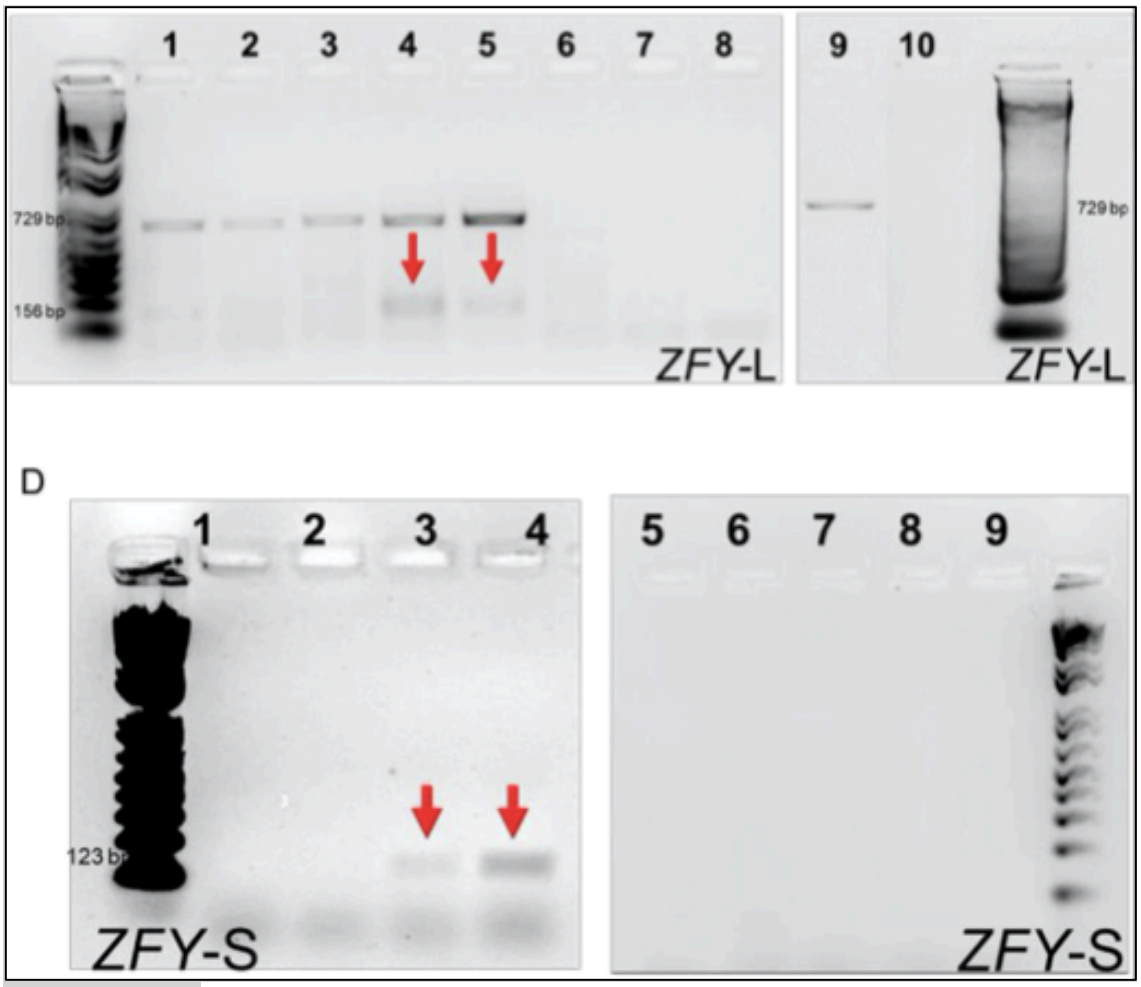
B. RBMY1F


Bulk tissue gene expression for RBMY1F (ENSG00000169800.13)
Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)
Data processing and normalization
Top
Bulk Tissue Expression
Single Cell Expression
Exon Expression
Single-Tissue eQTLs
Single-Tissue sQTLs
Single-Tissue ieQTLs
Single-Tissue isQTLs
SUBSET None Sex
SCALE Log Linear
TISSUE SORT
MEDIAN SORT
OUTLIERS On Off
Bulk tissue gene expression for RBMY1F (ENSG00000169800.13)
TPM
1.5
1.0
0.5
0.0
Adipose - Subcutaneous
Adipose - Visceral (Omentum)
Adrenal Gland
Artery - Aorta
Artery - Coronary
Artery - Tibial
Bladder
Brain - Amygdala
Brain - Anterior cingulate cortex (BA24)
Brain - Caudate (basal ganglia)
Brain - Cerebellar Hemisphere
Brain - Cerebellum
Brain - Cortex
Brain - Frontal Cortex (BA9)
Brain - Hippocampus
Brain - Hypothalamus
Brain - Nucleus accumbens (basal ganglia)
Brain - Putamen (basal ganglia)
Brain - Spinal cord (cervical c-1)
Brain - Substantia nigra
Breast - Mammary Tissue
Cells - Cultured fibroblasts
Cells - EBV-transformed lymphocytes
Cervix - Ectocervix
Cervix - Endocervix
Colon - Sigmoid
Colon - Transverse
Esophagus - Gastroesophageal Junction
Esophagus - Mucosa
Esophagus - Muscularis
Fallopian Tube
Heart - Atrial Appendage
Heart - Left Ventricle
Kidney - Cortex
Kidney - Medulla
Liver
Lung
Minor Salivary Gland
Muscle - Skeletal
Nerve - Tibial
Ovary
Pancreas
Pituitary
Prostate
Skin - Not Sun Exposed (Suprapubic)
Skin - Sun Exposed (Lower leg)
Small Intestine - Terminal Ileum
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
Whole Blood

C. RBMY1J



Supplementary Figure 2. *RBMY* genes expression in normal tissues. A. *RBMY1E*. B. *RBMY1F*. C. *RBMY1J*. There is no data for normal tissue expression of *RBMY1B* and *RBMY1D*.



(Trujilo, 2019)

Supplementary Figure 3: PCR test for ZFY expression in HPV+ and HPVneg cell lines.