An investigation into the function of WFDC2

Hannah Armes

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

School of Clinical Dentistry

The University of Sheffield

January 2019

(This page is left blank intentionally)

Acknowledgments

First and foremost, I would like to thank my supervisor Dr Lynne Bingle for her endless advice, encouragement and kind words. I have learnt so much from her extensive knowledge and experience. Under Lynne's supervision I have become a confident and independent researcher and for that I cannot thank her enough.

I would also like to thank Professor Colin Bingle for his support throughout both my PhD and my undergraduate degree. His guidance and encouragement made me realise that a career in science was something achievable for me.

I would like to give a huge thank you to the technical staff who helped me to make so much progress in my three years at the School of Clinical Dentistry: the wonderful Brenka McCabe and Jason Heath for teaching me molecular biology and microbiology techniques and to Hayley Stanhope for teaching me histological techniques and processing my tissue samples. I would also like to thank Jessica-Leigh Tallis for being an excellent SURE scheme student and for generating some brilliant immunohistochemistry analysis.

Thank you to the Bingle group for their support and kind words over the years (Apoorva, Priyanka, Renata, Debora, Chloe, Zulaiha and Esra). A huge thank you to my friends in the 3rd floor PGR office for making me laugh constantly for 3 years, even when faced with lab work disappointments: Amy Harding, Sven Niklander, Marianne Satur and Alex Bolger.

And finally, thank you to my lovely family and Chris, for their constant, unwavering love and encouragement. I could not have done this without you!

Abstract

Introduction

Whey-acidic-protein (WAP) four-disulfide core domain protein 2 (WFDC2) is a small, secretory glycoprotein that is characterised by possession of two cysteine-rich protein domains. In healthy tissues, WFDC2 is most abundantly expressed in the trachea and oral cavity, yet its function at these sites is unknown. Structural similarities to other family members suggests that WFDC2 may play a role in host defence, thus antimicrobial and anti-protease functions have been hypothesised. WFDC2 expression is exacerbated in ovarian cancer and as a result it is used clinically as a serum biomarker. WFDC2 is also reported to be upregulated in other malignancies, including lung cancer, in addition to benign diseases such as cystic fibrosis and renal fibrosis. It is unclear whether WFDC2 mediates tumorigenic effects. The aim of this study was to determine the function of WFDC2, with the primary objective of understanding whether it is involved in tumour progression.

Materials and methods

Recombinant human and murine WFDC2 were synthesised by gene cloning and transfection into HEK293 cells. WFDC2 protein was purified from conditioned media and utilised for protease inhibition and bacteria assays. The N-glycosylation site of human WFDC2 was elucidated via site-directed mutagenesis and enzymatic cleavage. The WFDC2 expression of lung and oral cancer-derived cell lines was analysed by RT-qPCR, ELISA and Western blotting. An oral cancer cell line was utilised for CRISPR gene editing to silence endogenous WFDC2 gene expression. Successful gene silencing was confirmed via sequencing, RT-PCR, RT-qPCR, Western blotting and ELISA. Changes in cell behaviour between wild-type and CRISPR edited cells were analysed via standard *in vitro* cell behaviour assays. WT and heterozygous Wfdc2-knockout mice were studied via H&E staining and

immunohistochemistry. Micro-CT scans of WT and homozygous knockout mice embryos at E14.5 and E18.5 were compared to elucidate differences in phenotype.

Results

Recombinant human and murine WFDC2 were unable to inhibit the growth of any of the bacterial strains tested, nor were they able to inhibit protease activity. CRISPR edited cells showed a significant reduction in their capacity to invade Matrigel-coated Transwells compared to controls. Homozygous Wfdc2-knockout mice died shortly after birth as a result of respiratory distress while heterozygous animals survived to adulthood and had no obvious histological phenotype. Micro-CT analysis showed that the trachea and bronchi of homozygote embryos were constricted at E14.5 and E18.5.

Conclusions

WFDC2 is not a host defence protein and is instead involved in development of the tracheal and bronchial lumina during embryogenesis. WFDC2 is involved in tumorigenesis by promoting cancer cell invasion. This suggests that WFDC2 is a prognostic biomarker and could represent an interesting therapeutic target. The role of WFDC2 in development requires further analysis.

Contents

List of figures

List of tables

List of abbreviations

ALI	Air-liquid interface
AMP	Antimicrobial peptide
BAL fluid	Bronchoalveolar lavage fluid
BPIFB1	BPI-fold containing family protein B1
CA125	Cancer antigen 125
CCSP	Club cell secretory protein
cDNA	Complementary DNA
CEA	Carcinoembryonic antigen
CF	Cystic fibrosis
CFBE	Cystic fibrosis bronchial epithelial cells
CFU	Colony forming units
CGH	Comparative genomic hybridisation
CRB3	Crumbs 3
CYFRA 21.1	Cytokeratin-19 fragment
ECM	Extracellular matrix
ELISA	Enzyme-linked immunosorbent assay
FACS	Fluorescence activated cell sorting
FBS	Foetal bovine serum
FDC	Four-disulfide core
FITC	Fluorescein isothiocyanate
FL	Full-length
gDNA	Genomic DNA
GFP	Green fluorescent protein
H&E	Haematoxylin and eosin

HE4	Human epididymis protein 4
HuSL	Human sublingual gland cells
IC^{50}	Inhibitory concentration to cause 50% decrease in CFU
IFN-γ	Interferon gamma
IHC	Immunohistochemistry
IL-1β	Interleukin 1 beta
LPS	Lipopolysaccharide
MEC	Mucoepidermoid carcinoma
Micro-CT	Micro computed tomography
MMP	Matrix metalloprotease
mNEC	Murine nasal epithelial cells
mTEC	Murine tracheal epithelial cells
MUC	Mucin
NSCLC	Non-small cell lung cancer
NSE	Neuron-specific enolase
PAGE	Polyacrylamide gel electrophoresis
PAM motif	Protospacer adjacent motif
PAMP	Pathogen-associated molecular pattern
PBS	Phosphate buffered saline
PI	Protease inhibitor
p-NA	p-nitroanilide
PSA	Prostate specific antigen
ROMA	Risk of Ovarian Malignancy Algorithm
RT-PCR	Reverse transcription polymerase chain reaction
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
SAGE technology	Serial analysis of gene expression technology

SAPNA	N-Succinyl-Ala-Ala-Ala-p-nitroanilide
SCC	Squamous cell carcinoma
SCLC	Small cell lung cancer
SD	Standard deviation
SDM	Site-directed mutagenesis
SDS	Sodium dodecyl sulfate
SEM	Scanning electron microscopy
SEM	Standard error of the mean
SEMG	Semenogelin
sgRNA	Single-guide RNA
siRNA	Small interfering RNA
SL	Sublingual
SLPI	Secretory leukocyte protease inhibitor
SMD	Submandibular
SNV	Single nucleotide variant
SPEM	Spasmolytic polypeptide-expressing metaplasia
TBE cells	Tracheobronchial epithelial cells
TCP	Tissue culture plastic
TNF-α	Tumour necrosis factor alpha
UTI	Urinary tract infection
V0-V4	Variant 0-4
WAP	Whey-acidic-protein
WFDC domain	Whey four-disulfide core (WFDC) domain
WFDC2	Whey-acidic-protein four-disulfide core domain protein 2
WT	Wild-type

Chapter 1
Literature review

1.1. An introduction to WFDC2

Whey-acidic-protein (WAP) four-disulfide core domain protein 2 (WFDC2) is a small, secretory glycoprotein characterised by its possession of two WAP four-disulfide core (WFDC) domains. The WFDC domain is a structural motif formed of approximately 50 amino acids, including 8 conserved cysteine residues that form 4 interlinked disulfide bonds (Hennighausen and Sippel, 1982). In related proteins, WFDC domains appear to confer anti-proteinase and antimicrobial activities, however the exact function of the domains in WFDC2 is unknown. At a nucleotide level, the WFDC2 gene is approximately 12 kb and consists of five exons which can undergo alternative splicing to produce five different protein isoforms.

In humans, WFDC2 is most abundantly expressed in the trachea and salivary glands where its expression pattern is largely restricted to the ducts and glands of mucosal surfaces. The presence of WFDC2 at mucosal surfaces, in addition to its structural similarity to other WFDC family members which exhibit host protective functions, has led to the hypothesis that WFDC2 is capable of bacterial killing and inhibition of host-derived proteases. This theory is yet to be extensively researched.

WFDC2 is primarily of interest to the scientific community because it is overexpressed in a number of malignancies including ovarian cancer where it is used clinically as a serum biomarker for disease detection. It has been reported that WFDC2 is also upregulated in a number of other cancer types, including lung cancer, although serum concentrations remain significantly lower than in patients with ovarian malignancies. Interestingly, WFDC2 appears to be overexpressed in some benign diseases, particularly in the case of cystic fibrosis. The role of WFDC2 in both malignant and benign diseases is not well understood. The effect of

WFDC2 on cancer cell behaviour has been studied using both WFDC2-overexpressing and -knockdown clones, but the resultant literature is conflicted and it remains unclear whether WFDC2 is involved in tumorigenesis.

1.1.1. The WFDC2 gene

Full-length WFDC2 is located at locus 20q13.12, spanning 11.8 kb, and is capable of undergoing alternative splicing to form five distinct splice variants. Although WFDC2 was first isolated and cloned in 1991 (Kirchhoff et al., 1991), its numerous splice variants were not identified until 2002, whereby alignment of expressed sequence tags (ESTs) and comparison to genomic DNA revealed distinct exonic arrangements between transcripts (Bingle et al., 2002). The full-length gene structure of WFDC2 consists of five exons, two of which are capable of existing in three forms (Figure 1.1). The gene also encodes a signal peptide, suggesting that its product is secretory in nature (Bingle et al., 2002).

Fig 1.1. The full length human WFDC2 transcript. A schematic diagram depicting the gene structure of full length human WFDC2, adapted from Bingle et al (2002). The gene is 11.8 kb in size. The intron between exon 3 and 4 is over 8 kb long. The estimated sizes (bp) of each exon are included. The 5' start sites of exon 3b and 4b are unknown and so the precise exon sizes are unknown. WFDC domains are encoded within exon 2 and exon 4.

1.1.2. The WFDC2 splice variants

Reverse transcription polymerase chain reaction (RT-PCR) analysis using multiple primer pairs and conceptual translation enabled the genomic arrangement of all five WFDC2 splice variants [variant (V) 0-4] to be determined. The exonic arrangements of each variant are depicted in Fig 1.2 (Bingle et al., 2002). The predominant isoform of WFDC2, V0, is referred to as full-length (FL) and is the only variant that encodes both an N- and C-terminal WFDC domain. V2 and V3 both encode C-terminal WFDC domains whilst V1 and V4 encode an N-terminal WFDC domain in addition to unique C-terminal sequences. All isoforms contain a 30

residue signal peptide aside from V3 which appears to have an alternative promoter region (Bingle et al., 2002; Kirchhoff et al., 1991). The biological role of the different protein isoforms is yet to be explored and it is unclear whether the splice variants are differentially regulated.
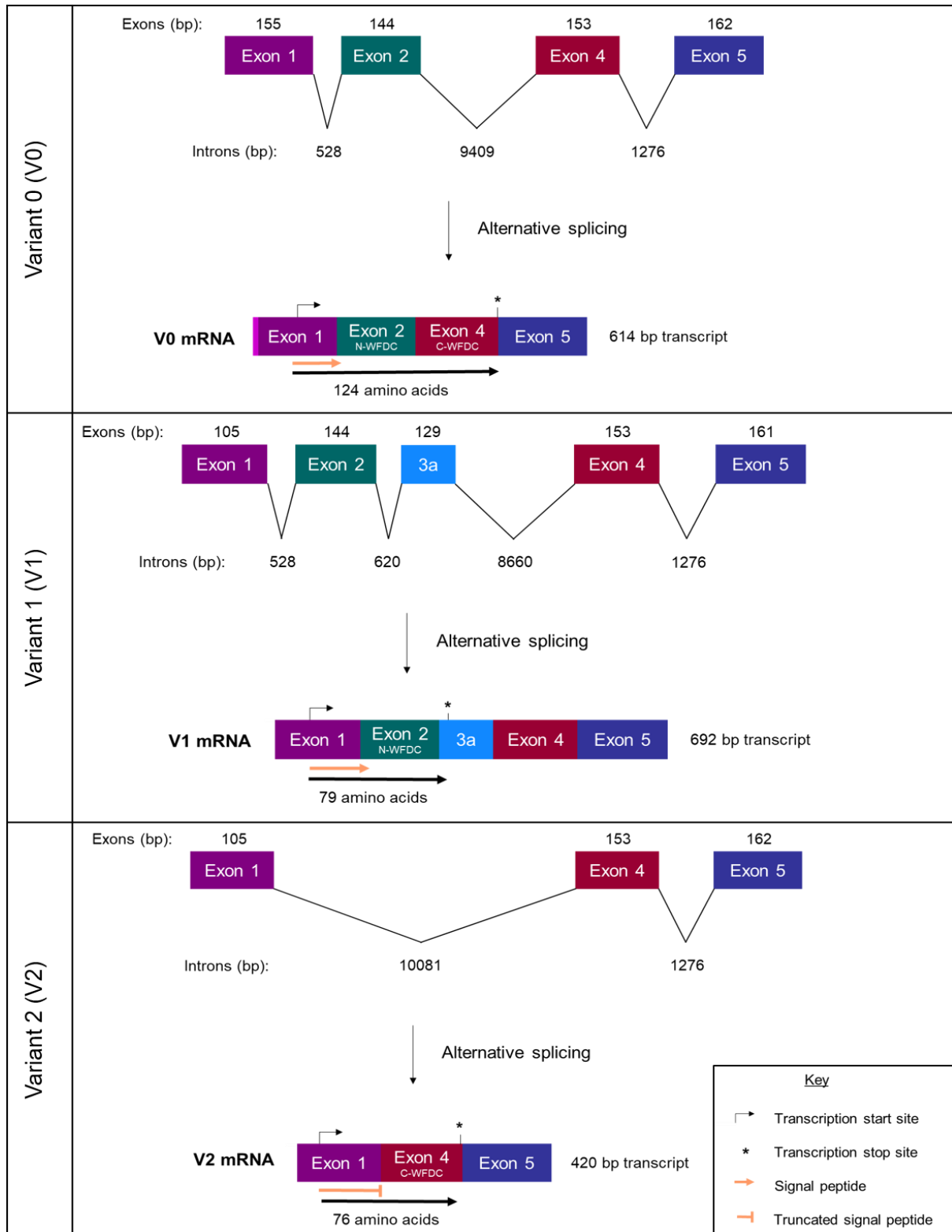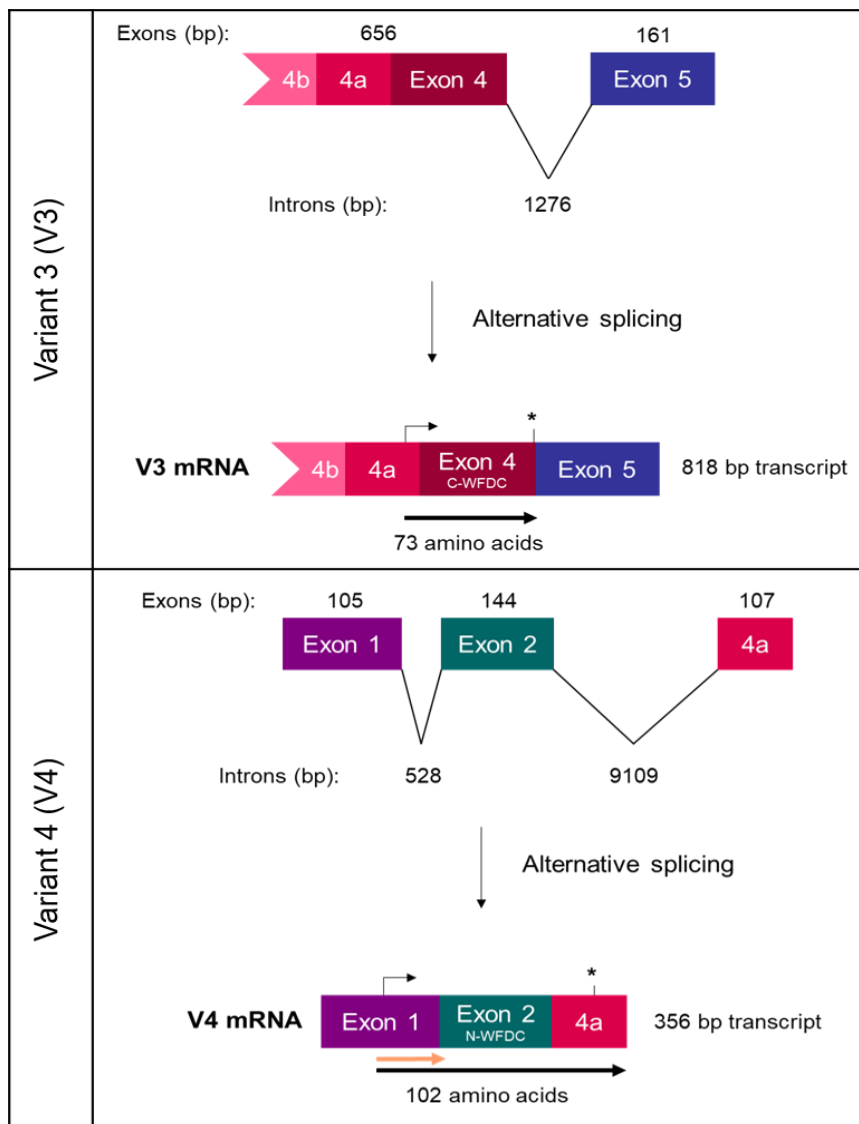
Fig 1.2. The genomic arrangement of the five WFDC2 splice variants. A schematic diagram to show the genomic organisation of the five WFDC2 splice variants (V0, V1, V2, V3 and V4), adapted from Bingle et al (2002). The approximate length (bp) of each exon and intron are indicated, as is the putative length of the resultant polypeptide after splicing. The putative transcription start and stop sites are shown (see Key in pane of Variant 3) and the presence or absence of a signal peptide is also indicated.

1.1.3. The WFDC protein family

Proteins that are known to express a motif with four conserved disulfide bridges are loosely termed four-disulfide core (FDC) containing proteins. The FDC grouping is subcategorised into the WFDC protein family, to which WFDC2 belongs (Bingle and Vyakarnam, 2008). There are eighteen known human WFDC genes, fourteen of which are located at a single locus on chromosome 20 (Bingle and Vyakarnam, 2008; Clauss et al., 2002).

Proteins containing a 40-50 residue motif with a conserved arrangement of four disulfide bridges were first collected into a family by Drenth et al (1980), who demonstrated that FDC motifs existed in a broad range of proteins with largely unrelated functions, including wheat germ agglutinin, snake venom neurotoxins and cardiotoxins, ragweed pollen allergen and hevein. The subsequent discovery and cloning of mouse whey acidic protein (WAP), the major whey protein in murine milk, led to its allocation to the FDC group due to the presence of two cysteine-rich domains (Drenth et al., 1980; Hennighausen and Sippel, 1982; Piletz et al., 1981).

The general structure of the FDC domain was further elucidated by Ranganathan et al (1999) whereby 84 FDC domains from a broad range of species were aligned in order to define a consensus motif. The study largely focused on WAP proteins from numerous species, therefore the FDC consensus was defined as a signature motif for identifying new members of the 'WAP-type' family, thus marking WAP as the prototypic member of the group (Ranganathan et al., 1999). Alignment analysis of the 18 human FDC-containing proteins by Bingle and Vyakarnam (2008) led to amendment of the 'WAP-type' signature sequence, thus defining a consensus motif specifically for the human proteins. The amended signature demonstrated that the only residues that remain constantly conserved in all 26 human domains are the 8 cysteine residues (Bingle and Vyakarnam, 2008). It is important to note that although the domain is frequently entitled the WAP motif, it is not found exclusively in whey acidic protein; therefore, in this review, the domain will be more accurately entitled the WFDC domain.

1.1.4. The WFDC domain

The X-ray crystal structure of the well-studied WFDC protein secretory leukocyte protease inhibitor (SLPI) has provided insight into the architecture of the WFDC domain (Grütter et al., 1988). SLPI is a protease inhibitor that contains two WFDC motifs: a C-terminal domain (domain 2) that has broad anti-protease activity and an N-terminal domain (domain 1) that is incapable of protease inhibition (Bingle and Vyakarnam, 2008; Eisenberg et al., 1990; Masuda

et al., 1995). Differences between the two motifs give an insight into the structural organisation required for successful protease inhibition. This should be carefully considered when analysing the domains of related WFDC proteins, including WFDC2.

According to Grütter et al (1988), SLPI is a boomerang shaped molecule with analogous WFDC domains at the end of each lobe. Each domain exists as an extended planar spiral, resulting in a flat, wedge-like conformation. Few non-covalent interactions exist between the two domains, suggesting that they function independently. The polypeptide chain of each domain is formed of four antiparallel strands: two twisted β-sheets joined by a hairpin loop form the internal part of the motif, while the two outer strands are connected by a wider loop (Figure 1.3). In domain 2, this outer loop contains the primary enzyme-binding segment and scissile peptide bond. The conserved intradomain disulfide bridges that characterise the domain serve to anchor the internal and external strands together, stabilising the structure and, in the case of domain 2, ensuring that the protease binding loop remains exposed (Grütter et al., 1988). It is hypothesised that the analogous loop in domain 1 is unable to exhibit protease inhibition as a result of its alternative cysteine spacing compared to that observed in domain 2, which is likely to affect the capacity of the loop to bind to target enzymes (Bingle and Vyakarnam, 2008).
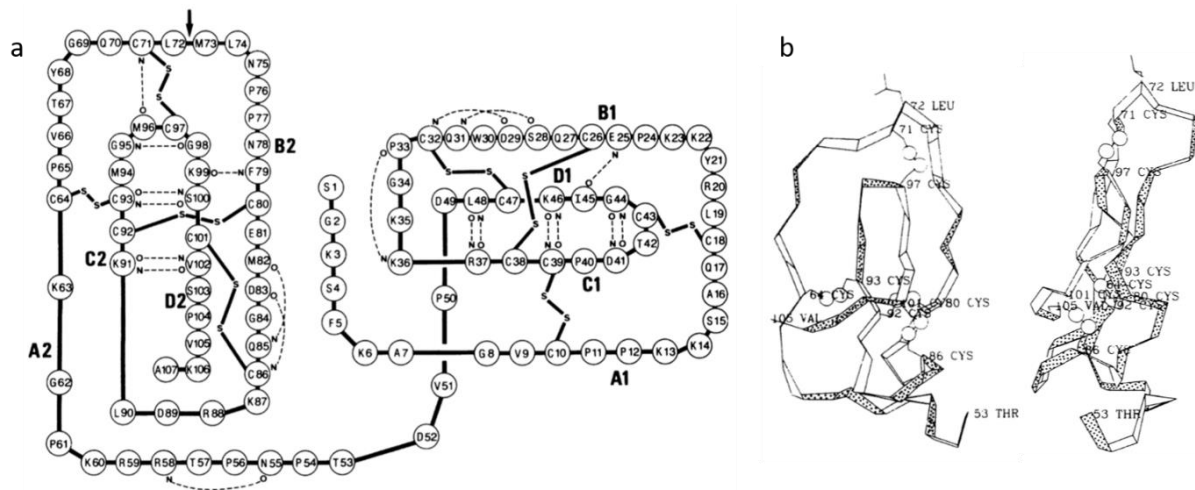
Fig 1.3. The architecture of the WFDC domains in human SLPI. (a) A summary of the polypeptide arrangement of SLPI, taken from Grutter et al (1988). Disulfide bridges are shown. The boundary between the two domains is residue 52. The four antiparallel segments are labelled (A1-D1 in domain 1 and A2-D2 in domain 2). Strand C and D form a twisted β-sheet joined by a hairpin loop surrounded by straight chains A and B. The exterior chains (A and B) are connected to the internal chains (C and D) by disulfide bridges. The arrow depicts the position of the scissile peptide bond (Leu72-Met73). The enzyme binding loop (residues 68-74) stacks upon the internal hairpin loop, forming a globular domain. Residues 13-24 in domain 1 would be analogous to residues 68-74 in domain 2. (b) A ribbon diagram of the polypeptide chain of domain 2. The 8 conserved cysteine residues and Leu72 are shown. The left diagram shows the front view of the structure, and the right-hand figure shows the side view after a 90° rotation.

1.1.5. Glycosylation

In addition to its two WFDC domains, WFDC2 is also characterised by a single N-linked glycosylation site, predicted to be positioned at Asn-44, which contributes around 5 kDa to the molecular weight of the protein (Drapkin et al., 2005; Hua et al., 2014; Wang et al., 2015) Glycosylation analysis carried out by Hua et al (2014) showed that WFDC2 has a complex, biantennary glycan structure. Fucose residues were identified at the glycan core whilst sialic acid residues were also identified at the terminus (see Figure 1.4). The complex glycan structure caused the molecular weight of WFDC2 to be larger than its theoretical mass (Hua et al., 2014). Glycosylation can influence a variety of key processes such as protein folding, secretion, turnover and trafficking, in addition to promoting protein-protein interactions and adhesion (Varki, 1993). The role of such extensive glycosylation in WFDC2 remains unclear.
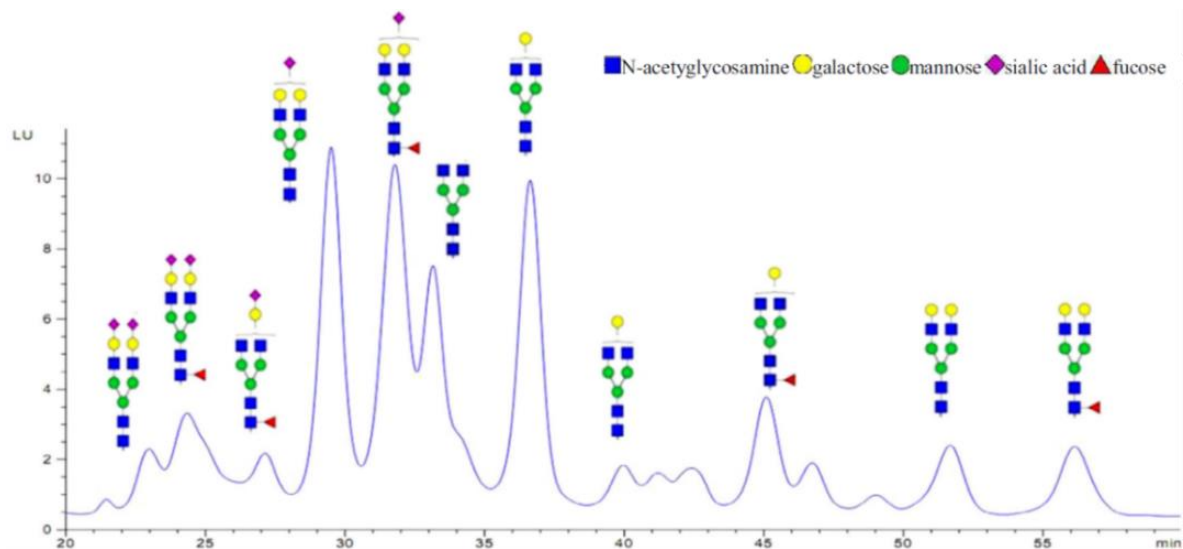
Fig 1.4. Mass spectrometry analysis of the N-glycan structure of WFDC2. Data taken from Hua et al (2014). Recombinant WFDC2 produced in mammalian cells was subjected to enzymatic cleavage to remove the N-glycans and the glycan structure was then analysed by liquid chromatography-electrospray ionisation mass spectrometry (LC-ESI-MS). The structure of the glycan from each chromatography peak is shown.

1.1.6. The WFDC locus

As previously mentioned, there are 18 human WFDC proteins, 14 of which are located on a 678 kb 'WFDC locus' which is located on chromosome 20 at position 20q12-13.1 (Clauss et al., 2002). WFDC genes tend to encode small, secretory proteins, expressing one or more WFDC domain. The majority of WFDC genes do not express a typical TATA-box, however, WFDC2 exhibits a TATA-like motif, which is an attribute prevalent in housekeeping genes (Clauss et al., 2002).

The WFDC locus is organised into two clusters of WFDC genes which are separated by a 215 kb region containing unrelated genes (Clauss et al., 2002) (Figure 1.5). The two subloci are thought to have evolved from a single gene duplication event that occurred more than 75 million years ago, with further gene multiplications leading to the clustering of many WFDC genes (Hurle et al., 2007; Schalkwijk et al., 1999). A typical WFDC domain is encoded within a single exon which enables chimeric genes to form via exon shuffling; for example, elafin encodes a WFDC domain and a transglutaminase substrate domain (Schalkwijk et al., 1999).
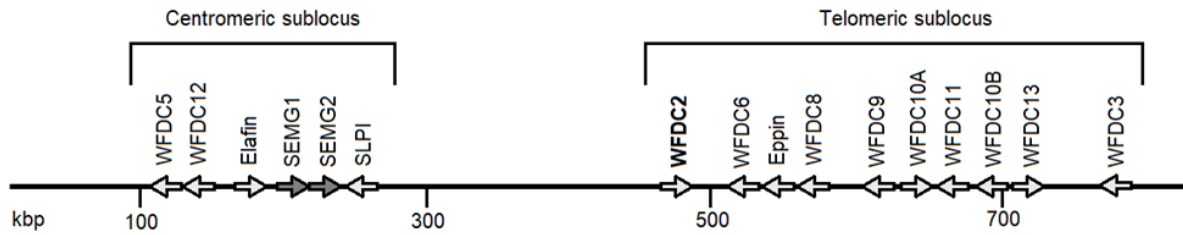
Fig 1.5. A schematic diagram of the WFDC locus. A summary of the organisation of the WFDC locus found on chromosome 20 at position 20q12-13.1. Adapted from Clauss et al (2002). The positions of each gene are indicated. WFDC2 is found in the telomeric sublocus, whilst SLPI, elafin and the semenogelins are located within the centromeric sublocus.

In addition to the WFDC genes, the WFDC locus also plays host to two genes that encode semen coagulum proteins, known as semenogelins (SEMG). The semenogelins are involved in crosslinking spermatozoa in human ejaculate; the matrix is subsequently broken down by prostate specific antigen (PSA) whereby the spermatozoa regain their motility. Interestingly, genomic analysis has shown that the SEMG genes share homology with the WFDC gene elafin, suggesting that the coagulum proteins evolved from an ancestral WFDC gene (Lundwall and Clauss, 2011).

The WFDC locus has been identified as one of the most rapidly diverging genomic segments between the human and chimpanzee genomes (Consortium., 2005). This accelerated divergence is characteristic of genes involved in the innate immune system, such as α- and β- defensins (Wiesner and Vilcinskas, 2010). Links have thus been drawn to suggest that WFDC genes provide host defence within the human reproductive system (Ferreira et al., 2013a; Ferreira et al., 2013b). An evaluation focusing on the extent of positive selection within the WFDC locus has shown that there is significant heterogeneity between remote human populations, suggesting that the heterogeneity results from a disparity in endemic pathogens (Ferreira et al., 2013b).

The functions of most of the genes encoded on chromosome 20 are largely unknown aside from SLPI, elafin and eppin which have been relatively well studied. SLPI and elafin are expressed at various mucosal surfaces, including the respiratory tract, and have been shown

to prevent harmful inflammation by acting as protease inhibitors in addition to exhibiting moderate antimicrobial activity against bacteria, fungi and viruses (reviewed in Sallenave, 2010). Eppin is expressed within the testis and epididymis where it binds to spermatozoa, providing antimicrobial protection and affecting sperm motility (Wang et al., 2007).

The four WFDC proteins located outside of chromosome 20 are more highly conserved than those found within the WFDC locus (even down to the level of invertebrates such as nematodes and insects). The genes tend to encode much larger proteins, with many other domains found expressed alongside the WFDC domain (Bingle and Vyakarnam, 2008).

1.2. WFDC2 expression

1.2.1. The expression profile of human WFDC2

Human WFDC2 was initially considered to be exclusively expressed in the epididymis after Northern blot analysis failed to identify its expression in any of the other six human tissues tested (Kirchhoff et al., 1991; Kirchhoff et al., 1990). Northern blotting and in situ hybridisation showed that WFDC2 was expressed largely at the distal portion of the epididymis and vas deferens, where it appeared to be localised to the epithelium. More recent studies have shown that WFDC2 is expressed in a much broader range of tissues aside from the male reproductive tract, meaning that its common alias 'human epididymis protein 4 (HE4)' is misleading.

High throughput sequencing techniques have identified the gene expression of WFDC2 to be highest in tracheal and salivary gland tissue, with significant expression also in the lung and kidney (Su et al., 2002; Uhlén et al., 2015) (Fig 1.6 and Fig 1.7). WFDC2 expression has also been identified within the female reproductive tract, particularly in the uterus, fallopian tube and endometrium, and in the male reproductive tract using high throughput sequencing (Uhlén et al., 2015).
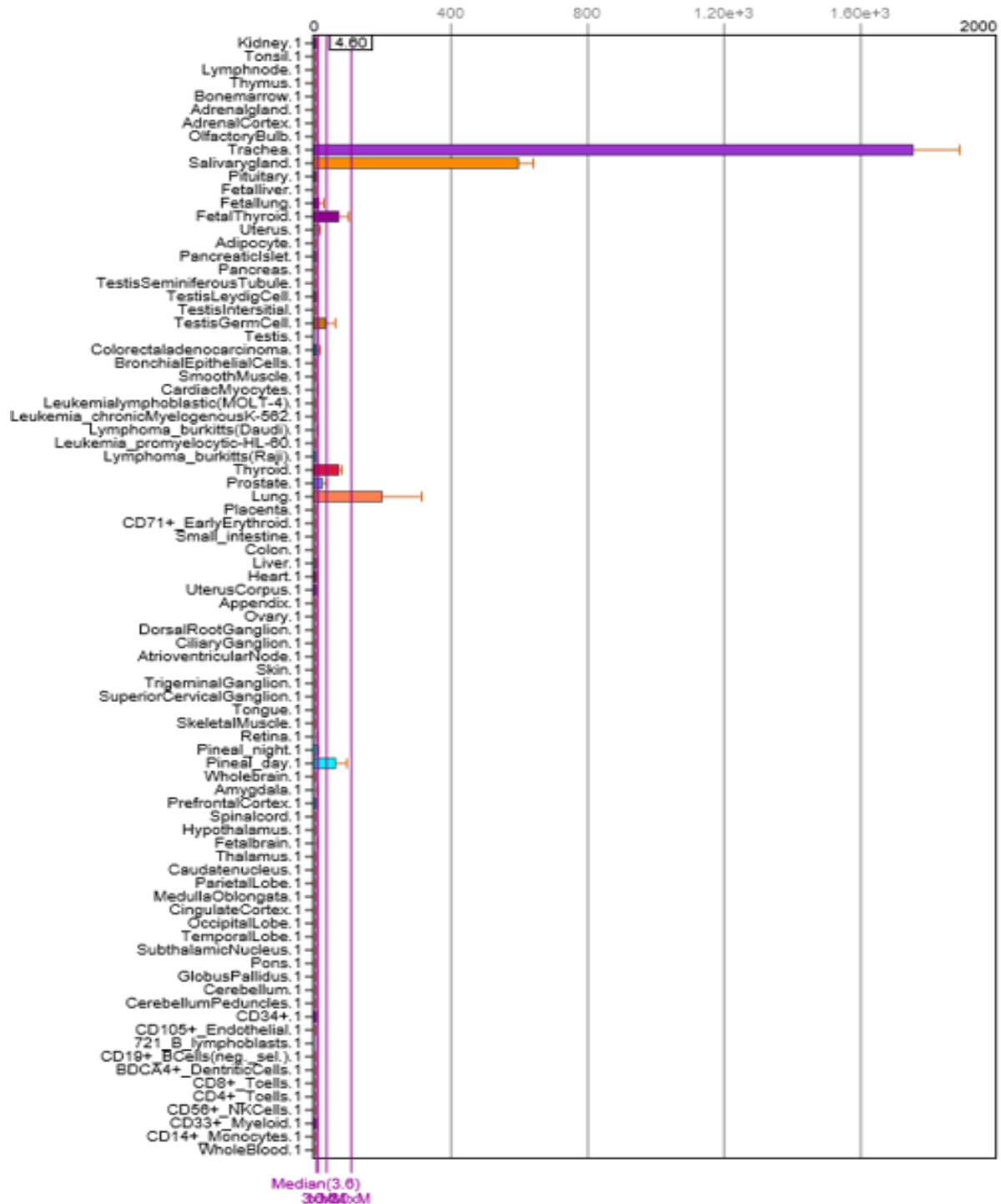
Fig 1.6. Gene expression profile of human WFDC2 from BioGPS. Array data showing relative WFDC2 expression in human tissues published on the BioGPS portal. Accessed November 2018 (Su et al., 2002).
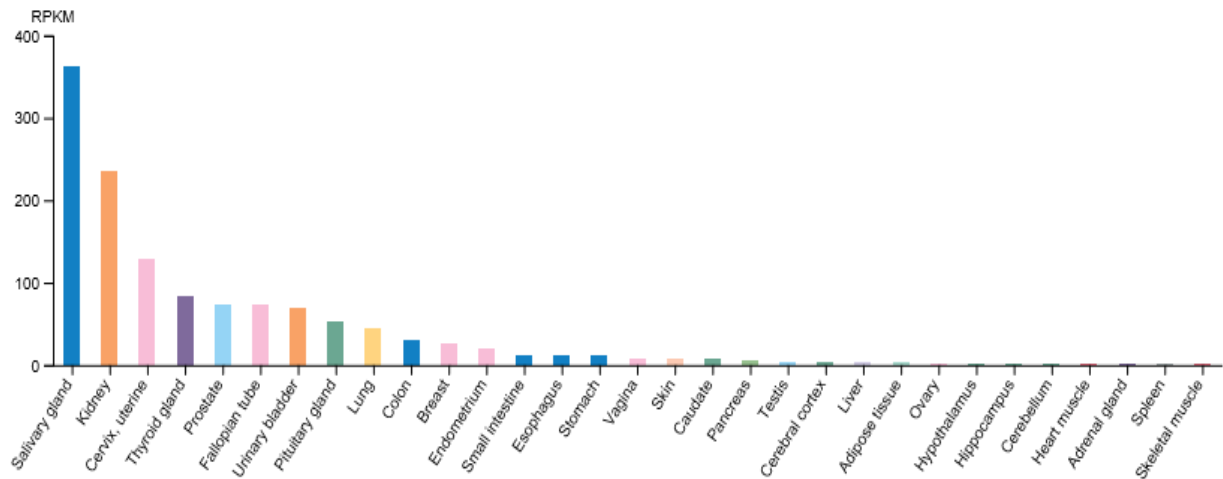
Fig 1.7. Gene expression profile of human WFDC2 from the Human Protein Atlas. RNA-seq data showing relative WFDC2 expression in human tissues published on the Human Protein Atlas database from the GTEx dataset, results are sorted according to expression levels: high to low. Accessed November 2018 (Uhlén et al., 2015).

At the protein level, the expression of human WFDC2 within the upper respiratory tract, nasopharynx and oral cavity was largely found to be localised to glands, particularly within the ductular epithelium of salivary glands and nasal passages, while expression was more specifically localised to serous cells of submucosal glands of the bronchi (Fig 1.8) (Bingle et al., 2006; Galgano et al., 2006; Uhlén et al., 2015). Moderate staining was also visible in other cell types, for example, weak epithelial staining was found in the nasal passage and bronchi, while some mucous cells within the maxillary sinus also stained positively (Bingle et al., 2006; Uhlén et al., 2015). In major and minor salivary glands, amongst the intense ductular staining, there was also some positivity within serous acini (parotid gland), striated and intercalated ducts (submandibular gland) and serous demilunes of mucous acini (sublingual gland) (Bingle et al., 2006). Peripheral lung tissue was negative for WFDC2 expression (Bingle et al., 2006). Mammary gland acini also stained positively for WFDC2 while the adipose tissue was found to be negative (Uhlén et al., 2015).
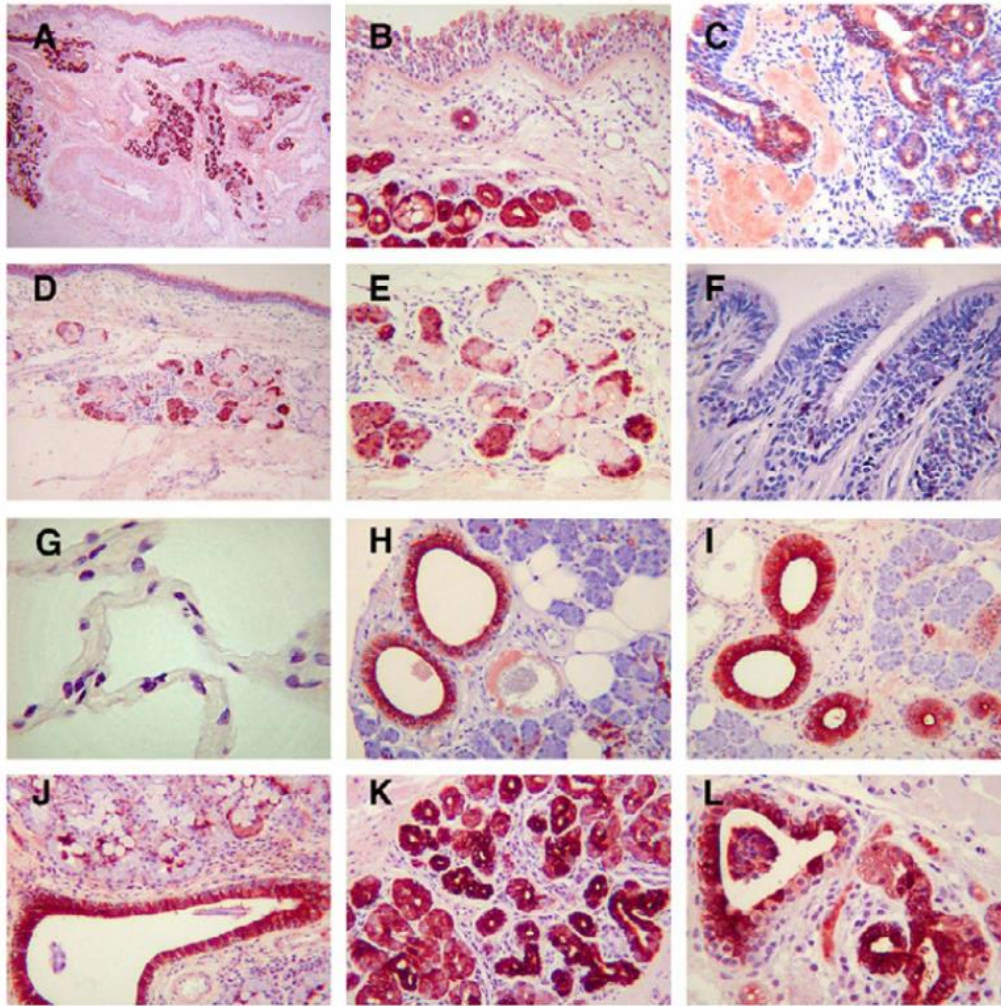
Fig 1.8. Analysis of WFDC2 staining in the respiratory tract and salivary glands. Immunohistochemical analysis of WFDC2 expression in the respiratory tract and salivary glands, taken from Bingle et al (2006). Staining was analysed in the nasal antral mucosa (A, B), nasal polyps (C), airway sub-mucosal glands (D, E), epithelial cells of the airways (F), peripheral lung (G), parotid gland (H), submandibular gland (I), sublingual gland (J), minor glands of the tongue (K) and minor glands of the tonsil (L). Original magnification of the images was X100 (A, B, C, D), X200 (E, F, H, I, J) and X400 (G, K, L). The antibody utilised was developed by Hellström et al (2003) using recombinant full-length WFDC2, synthesised in mammalian cells, as the immunogen.

Within the female reproductive tract, WFDC2 was abundant in the epithelium and glandular cells of the fallopian tube, endometrium and cervix, while the ovary was negative for WFDC2 expression (Drapkin et al., 2005; Galgano et al., 2006; Uhlén et al., 2015). The male reproductive tract also showed an abundance of WFDC2 in the epithelium and glandular cells of the epididymis, prostate and seminal vesicles (Drapkin et al., 2005; Galgano et al., 2006; Uhlén et al., 2015) (Fig 1.9). Positivity was detectable in the mucosal glands and goblet cells

of the duodenum and colon, whilst the tissues of the stomach and small intestine were largely negative for WFDC2 (Uhlén et al., 2015).
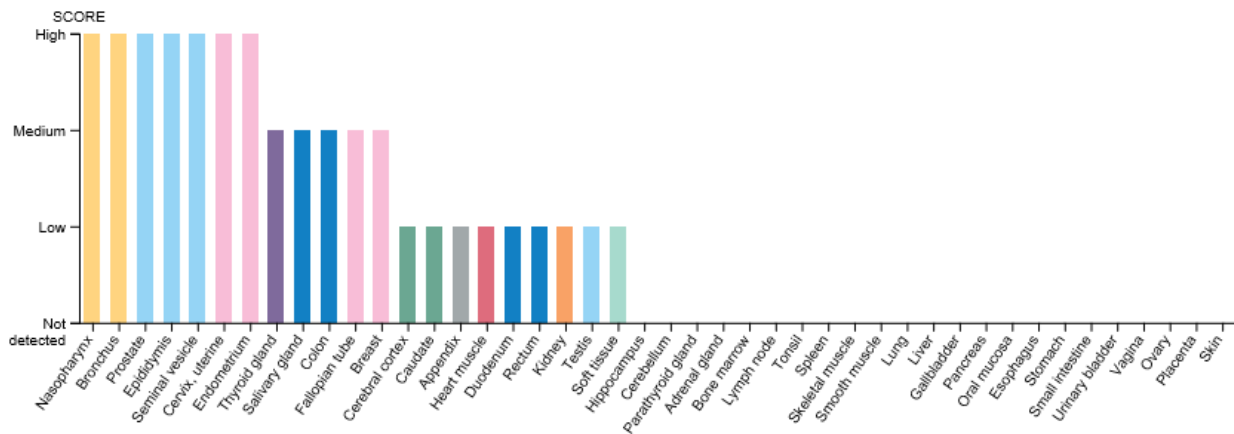
Fig 1.9. Protein expression profile of human WFDC2 from the Human Protein Atlas. Protein expression data derived from quantification of positive staining in tissues analysed by immunohistochemistry. Tissues are given a protein score of low, medium or high and results are sorted according to expression level: high to low. Data published on the Human Protein Atlas database. Accessed November 2018 (Uhlén et al., 2015). The antibody utilised was from Atlas Antibodies (HPA042302) and was developed using recombinant WFDC2 (Glu31-Phe124) as the immunogen.

1.2.2. Expression levels of WFDC2 splice variants

Using primers designed to span exon-exon boundaries, Jiang et al (2013) was able to study the gene expression of WFDC2 splice variants (V0-V4) within RNA from a range of human tissues, including prostate, ovary, endometrium, lung, trachea and kidney. Real time quantitative PCR (RT-qPCR) analysis demonstrated that although relative expression levels of different isoforms varied between tissues, V0 was always significantly more abundant than all other variants (Jiang et al., 2013). This is the only data published regarding the expression of WFDC2 splice variants. Further analysis is necessary to gain a thorough insight into the differential expression of the variants in human tissues.

1.2.3. WFDC2 orthologues

WFDC2 orthologues have been identified within the epididymis of various mammalian species, including cow (Kirchhoff et al., 1991; Uhlenbruck et al., 1993), horse (Uhlenbruck et

al., 1993), pig (Schäfer et al., 2003; Uhlenbruck et al., 1993), rabbit (Xu et al., 1996), dog (Ellerbrock et al., 1994; Pera et al., 1994) and rat (Bingle et al., 2002; Kappler-Hanno and Kirchhoff, 2003). The expression patterns of the majority of WFDC2 orthologues outside of the male reproductive tract are yet to be explored.

1.2.4. Rodent Wfdc2

According to microarray data published by Su et al (2002), murine Wfdc2 was found to be most abundant in the lung, uterus, bladder and large intestine, although the gene was also expressed moderately in the ovary, salivary gland, mammary gland and kidney (Fig 1.10). This expression data is supported by unpublished RT-PCR data, which showed positive expression of Wfdc2 at the aforementioned sites, in addition to the nasal cavity, trachea, testis and stomach (Bingle, unpublished). In murine tracheal epithelial cells (mTEC) grown at an air-liquid interface (ALI), Wfdc2 was the 6th most abundantly expressed gene when analysed by RNA-seq (Nemajerova et al., 2016). In the rat, Wfdc2 was found to be expressed in the lung, epididymis, kidney and liver as determined by Northern blotting (Kappler-Hanno and Kirchhoff, 2003). At the protein level, unpublished histological analysis reported that murine Wfdc2 was localised to the epithelia of the bronchi, trachea and nasal cavity, as well as being present in major salivary glands (Bingle, unpublished).

Mouse and rat Wfdc2 show 77% sequence alignment to one another; yet, the rodent orthologues share a much-reduced sequence identity to human WFDC2 due to the presence of an extra exon which is absent in human WFDC2 (Kappler-Hanno and Kirchhoff, 2003). The additional rodent exon forms a ~50 residue peptide 'linker' which separates exon 2 and 4, and is predicted to be heavily O-glycosylated (Kappler-Hanno and Kirchhoff, 2003). Although the most abundant rodent Wfdc2 gene transcript encodes exons 1-5 (including the additional exon 3 linker), other splice variants have also been identified by RT-PCR and sequencing (Bingle, unpublished) as well as Northern blotting (Kappler-Hanno and Kirchhoff, 2003).
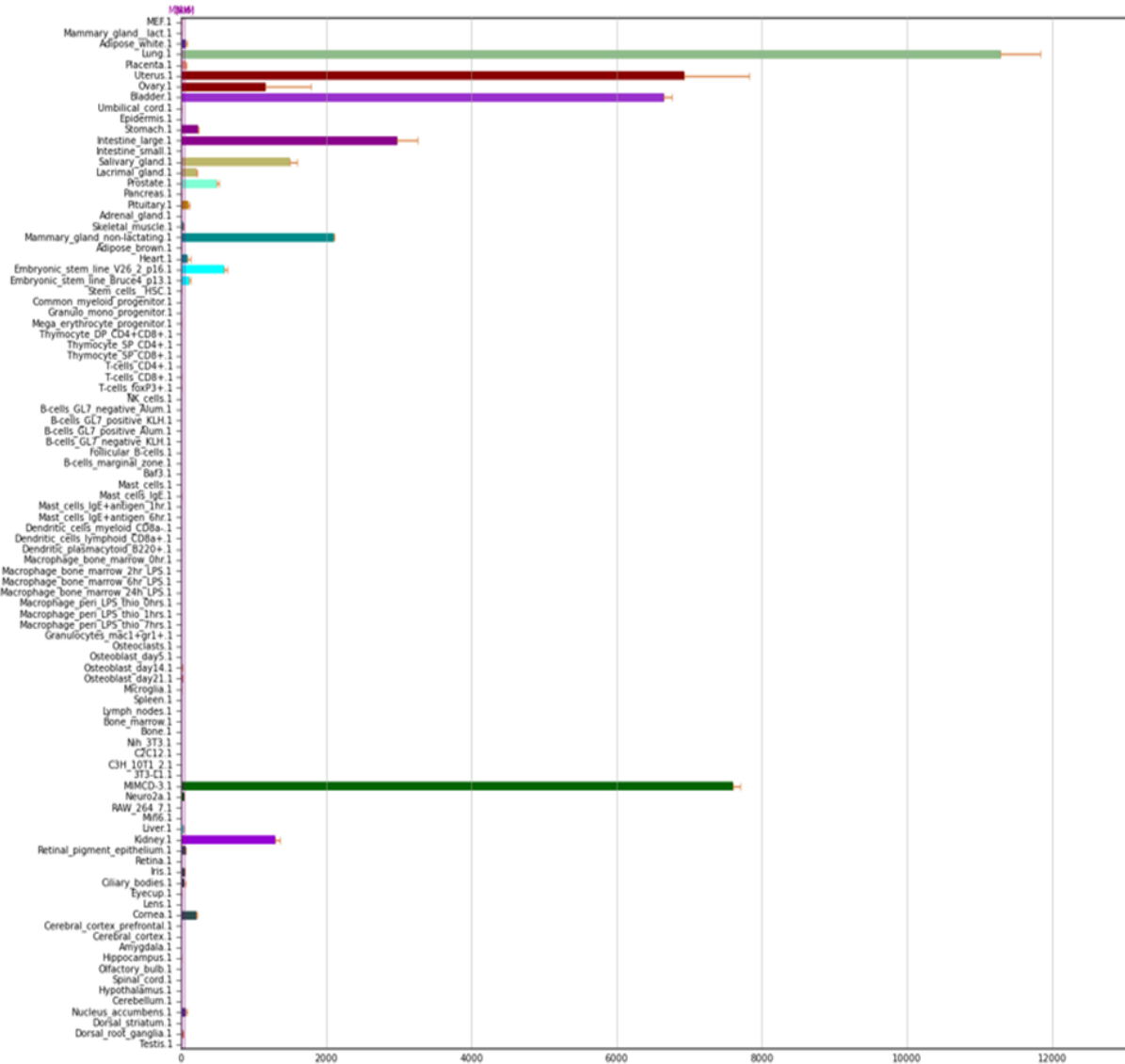
Fig 1.10. Expression profile of murine Wfdc2 from BioGPS. Array data showing relative Wfdc2 expression in adult mouse tissues published on the BioGPS portal. Accessed August 2018 (Su et al., 2002).

1.2.5. Human WFDC proteins in secretions

As discussed in Section 1.2.1, the expression of human WFDC2 within the respiratory tract, nasopharynx and oral cavity is largely localised to the glands, which suggests that WFDC2 is delivered into secretions at these sites e.g. into saliva or surface lining fluid. Mucosal surfaces within the oral cavity and respiratory tract are exposed to the external environment, so are concomitantly exposed to pathogens and allergens. As a result, glandular and epithelial

secretions are rich in proteins that provide a host-protective role to both pathogenic and host-derived threats. If WFDC2 is present in secretions, it is conceivable that it could exhibit antimicrobial and/or anti-inflammatory activity.

In the literature, there is some proteomic evidence to show that WFDC2 does indeed enter the glandular secretions at these sites. WFDC2 has been detected in bronchoalveolar lavage (BAL) fluid and in secretions from bronchial epithelial cells grown at an air-liquid interface (ALI), where it is the 13th most abundant protein (Peters-Hall et al., 2015; Wu et al., 2005). WFDC2 has also been identified in the proteome of nasal mucous (Tomazic et al., 2014), saliva (Denny et al., 2008; Fang et al., 2007; Ramachandran et al., 2006; Sivadasan et al., 2015; Yan et al., 2009) and tears (Zhou et al., 2012). Reproductive tissues also appear to secrete WFDC2, as the protein has been identified within seminal plasma (Pilch and Mann, 2006) and cervicovaginal lavage fluid where it is found within the top 10% of most abundantly secreted proteins (Birse et al., 2013). Additionally, enzyme-linked immunosorbent assay (ELISA) data identifies WFDC2 in vaginal secretions at a median concentration of 108 nM with a range of 2.6 to 429 nM (Orfanelli et al., 2014; Pilch and Mann, 2006).

Evidently WFDC2 protein is detectable in a number of glandular secretions, as determined by proteomic analysis. Unfortunately, there is no Western blot analysis available within the literature to support this data.

The expression patterns of human SLPI and elafin show some crossover with WFDC2. Both proteins are expressed in the oral cavity and airways (Bingle et al., 2006; Lee and Downey, 2001; Sallenave et al., 1994; Su et al., 2002; Uhlén et al., 2015; Wahl et al., 1997) yet the cell types expressing each protein are different; for example, while WFDC2 tends to be most abundant in ductular cells, SLPI is expressed in epithelial, serous and mucous cells and tends to be absent in ducts (Abbinante-Nissen et al., 1993; Bingle et al., 2006; Franken et al., 1989; Shugars, 1999; van Wetering et al., 2000). Elafin, on the other hand, is largely found in squamous epithelial cells, but can be found in some ducts (Lee et al., 2002; Pfundt et al., 1996; Uhlén et al., 2015).

Aside from the oral cavity and airways, SLPI is also found in the kidneys, intestine and male and female reproductive tracts, in addition to immune cells such as mast cells, neutrophils and macrophages (Franken et al., 1989; Helmig et al., 1995; Mihaila and Tremblay, 2001; Ohlsson et al., 1995; Ohlsson et al., 2001; Sallenave et al., 1997; Si-Tahar et al., 2000; Uhlén et al., 2015; Westin et al., 1999). SLPI is a secretory protein and resultantly can be detected in saliva, nasal mucous, BAL fluid, cervical mucous and seminal plasma (Helmig et al., 1995; Ohlsson et al., 1995; Saitoh et al., 2001; Shugars, 1999; Thompson and Ohlsson, 1986; van Wetering et al., 2000; Westin et al., 1994). SLPI is secreted at mucosal surfaces where it is reported to play a host protective role at these sites, which will be discussed in more detail in Section 1.3.

Elafin, on the other hand, is found in the skin (and is abundant in psoriatic skin), pharynx, oesophagus, female reproductive tract and small intestine, in addition to various sites within the oral cavity, including the gingiva, tonsil and tongue (Pfundt et al., 1996; Schalkwijk et al., 1999; Suzuki et al., 2000; Uhlén et al., 2015). It has also been found to be expressed in neutrophils and macrophages (King et al., 2003; Mihaila and Tremblay, 2001). Elafin can be detected in bronchial secretions and sputum (Sallenave et al., 1992; Sallenave and Ryle, 1991) and is hypothesised to provide a similar host-protective role to SLPI at mucosal surfaces exposed to the external environment (see Section 1.3).

Although the expression sites of WFDC2, SLPI and elafin show some correlation, each protein appears to exhibit some exclusivity in terms of the cell type it is expressed within. This may suggest that the proteins exhibit distinct functions and may indeed be under separate regulatory controls. As a result, it is important to recognise that WFDC2 may exhibit an entirely different biological function to SLPI and elafin, despite their collective possession of one or more WFDC domains.

1.3. The activities of WFDC proteins

1.3.1. The putative functions of human WFDC2

A role for WFDC2 in host defence is frequently hypothesised as a result of three key factors. Firstly, as discussed previously, WFDC2 is secreted at mucosal sites that are consistently exposed to the external environment and its concomitant pathogens. Secondly, the WFDC2 gene is encoded within the rapidly diverging WFDC locus, as discussed in Section 1.1.6, and accelerated evolution is characteristic of genes involved in the innate immune system (Hurle et al., 2007). Finally, WFDC2 shares structural similarities with two related WFDC proteins, SLPI and elafin (Bingle and Vyakarnam, 2008), which are widely reported to exhibit anti-inflammatory and antimicrobial activities; it is feasible that WFDC2 could exhibit similar functions.

1.3.2. Antimicrobial activity of SLPI and elafin

SLPI and elafin are small, non-glycosylated, cationic proteins, encoded around 200-300 kb upstream of WFDC2 on the WFDC locus (see Figure 1.5) (Clauss et al., 2002). SLPI consists of two WFDC domains (Grütter et al., 1988) whilst elafin encodes a single WFDC domain (Zeeuwen et al., 1997). Elafin initially exists as a precursor protein called pre-elafin (or trappin-2) which has an N-terminal transglutaminase substrate-binding domain to enable binding to extracellular matrix components via transglutaminase crosslinking. This N-terminal domain is then cleaved to yield mature elafin (Zeeuwen et al., 1997).

SLPI and elafin have both been reported to exhibit moderate antimicrobial activity, for instance, SLPI has been shown to reduce the *in vitro* growth of *Escherichia coli*, *Staphylococcus aureus*, *Staphylococcus epidermidis, Pseudomonas aeruginosa*, *Streptococcus pyogenes, Salmonella typhimurium* and *Mycobacterium tuberculosis* (Fernie-King et al., 2002; Gomez et al., 2009; Hiemstra et al., 1996; Si-Tahar et al., 2000; Wiedow et al., 1998) in addition to affecting the viability of fungi, including *Candida albicans* and *Aspergillus fumigatus* (Tomee et al., 1997; Wiedow et al., 1998). Similarly, elafin or pre-elafin

can impact the growth (albeit more weakly compared to SLPI) of *P. aeruginosa*, *S. aureus*, *Klebsiella pneumoniae*, *Haemophilus influenzae, Streptococcus pneumoniae* and *Branhamella catarrhalis* (Baranger et al., 2008; Meyer-Hoffert et al., 2003; Simpson et al., 1999), in addition to fungi such as *C. albicans* and *A. fumigatus* (Baranger et al., 2008). Elafin has also been proposed to exhibit anti-viral activity against human immunodeficiency virus 1 (HIV-1) by inhibiting virus binding (Drannik et al., 2012).

The *in vitro* assays used to determine the effect of SLPI on bacterial and fungal viability/growth were largely consistent, whereby the microorganisms were incubated with increasing concentrations of recombinant SLPI prior to serial dilution, plating out and counting of colony forming units (CFU). The single exception to this method was the experiment reported by Payne et al (2017) who used radial diffusion assays. The concentrations of protein required to cause a 50% decrease in CFU are summarised in Table 1.1. SLPI was able to induce significant killing of many species at concentrations close to the endogenous levels recorded in human secretions, as summarised in Table 1.2. Little data is available regarding the endogenous concentrations of elafin, but it seems unlikely that they would be as high as the IC^{50} values necessary to induce significant microbial killing.

The antibacterial activity of SLPI against *Mycobacteria* has received particular attention and has been established via *in vitro* and *in vivo* assays. Human and murine SLPI orthologues are reported to have IC^{50} values of 2.0 µM and 0.5 µM, respectively, against *M. tuberculosis* (Gomez et al., 2009; Nishimura et al., 2008). Interestingly, when mice were intranasally infected with *Mycobacterium bovis* BCG and later treated with intranasal SLPI (12.8 µM), a significantly lower bacterial load was recovered from the lungs in comparison to controls (Gomez et al., 2009). Further to this, Slpi knockout mice were found to be significantly more susceptible to *M. tuberculosis* infection compared to WT (Nishimura et al., 2008). Evidently SLPI has significant antimycobacterial activity; however, outside of this specific function, the broader antimicrobial activity of SLPI is yet to be determined using *in vivo* or knockout mouse

models. The absence of an elafin orthologue in mice makes it difficult for *in vivo* analysis of its function to be carried out.

		Bacteria	IC^{50}	Reference
SLPI		*Escherichia coli*	4.2 µM	Hiemstra et al., 1996
		Escherichia coli	2.5 µM	Wiedow et al., 1997
		Staphylococcus aureus	7.6 µM	Hiemstra et al., 1996
		Staphylococcus aureus	4.5 µM	Wiedow et al., 1997
		Staphylococcus epidermidis	3.0 µM	Wiedow et al., 1997
		Pseudomonas aeruginosa	3.0 µM	Wiedow et al., 1997
		Streptococcus pyogenes	~8.4 µM	Fernie-King et al., 2002
		Salmonella typhimurium	~5.1 µM	Si-Tahar et al., 2000
		Mycobacterium tuberculosis	~2.0 µM	Gomez et al., 2009
		Pseudomonas aeruginosa (n=4 clinical isolates plus reference strain)	NA[1]	Payne et al., 2017
		Staphylococcus aureus (n=4 clinical isolates plus reference strain)	NA[1]	Payne et al., 2017
		Streptococcus spp. (n=4 clinical isolates plus *Streptococcus anginosus* reference strain)	NA[1]	Payne et al., 2017
		Anchromobacter spp. (n=3 clinical isolates)	NA[1]	Payne et al., 2017
		Stenotrophomonas maltophilia (n=3 clinical isolates)	NA[1]	Payne et al., 2017
SLPI		Fungi	IC^{50}	Reference
		Candida albicans	5.8 µM	Tomee et al., 1997
		Candida albicans	3.5 µM	Wiedow et al., 1997
		Aspergillus fumigatus	10.0 µM	Tomee et al., 1997

	Bacteria	IC^{50}	Reference
Elafin/pre-elafin	*Pseudomonas aeruginosa* (n=2 strains)	~3.0 µM or ~16.0 uM (strain dependent)	Meyer-Hoffert et al., 2003
	Staphylococcus aureus	~25.0 µM	Simpson et al., 1999
	Staphylococcus aureus	~30.0 µM	Baranger et al., 2008
	Klebsiella pneumoniae	NA[2]	Baranger et al., 2008
	Haemophilus influenzae	NA[2]	Baranger et al., 2008
	Streptococcus pneumoniae	NA[2]	Baranger et al., 2008
	Branhamella catarrhalis	NA[2]	Baranger et al., 2008
	Fungi	IC^{50}	Reference
Pre-elafin	*Candida albicans*	~7.5 µM	Baranger et al., 2008
	Aspergillus fumigatus	NA[2]	Baranger et al., 2008

Table 1.1. The *in vitro* 50% inhibitory concentrations (IC^{50}) of SLPI, elafin and pre-elafin against bacteria and fungi. IC^{50} is the concentration of protein that causes a 50% decrease in colony forming units. NA[1] refers to cases which showed no demonstrable antimicrobial activity at any concentration tested. NA[2] refers to cases where some antimicrobial activity was evident, but 50% killing was not reached at any concentration tested. Approximated values were determined by extrapolation from graphs in the literature. References to the relevant literature are indicated.

Sample	Mean concentration of SLPI (µM)	Reference
Whole human saliva (unstimulated)	0.097	Shugars et al., 1999
Parotid saliva	0.047	Lin et al. 2004
Submandibular/sublingual	0.0966	Lin et al. 2004
Bronchoalveolar lavage fluid	0.011	Vogelmeier et al., 1991
Lavage of upper respiratory tract	8.7	Vogelmeier et al., 1991
Lavage of lower respiratory tract	0.6	Vogelmeier et al., 1991
Bronchoalveolar lavage fluid	0.0134	Rohde et al. 2014
Sample	Mean concentration of elafin (uM)	Reference
Bronchoalveolar lavage fluid	0.0134	Rohde et al. 2014

Table 1.2. The mean SLPI and elafin protein concentrations in human secretions. A summary of the endogenous concentrations of SLPI and elafin protein in human secretions quantified by ELISA. References to the relevant literature are indicated.

Although the majority of reports show that SLPI and elafin are capable of antimicrobial activity, there is some conflicting evidence which suggests the contrary. Payne et al (2017) used radial diffusion assays to determine the minimum inhibitory concentration (MIC) of various antimicrobial peptides (AMPs), including SLPI and β-defensins, against a number of different cystic fibrosis-derived clinical isolates and reference strains. The group reported that SLPI was incapable of inducing inhibitory activity against any isolate or reference strain tested (see Table 1.1). It has been suggested that the difference between the SLPI activity described in this report compared to the results described by Wiedow et al (1997) and others was due to the 'inherent antimicrobial resistance' following 'prolonged exposure to antibiotics' of the strains tested (Payne et al., 2017). The differences could, however, be due to the different experimental methods used.

The mechanisms through which SLPI and elafin exhibit antimicrobial activity are still under investigation. Both proteins were shown to retain antimicrobial activities when their protease inhibitory activity had been suppressed (i.e. by mutagenesis or heat-treatment), suggesting that the two functions are independent (Baranger et al., 2008; Gomez et al., 2009). In support

of this, preliminary evidence also suggested that the N-terminus of SLPI was responsible for its antimicrobial activity, although the isolated domain was much less active than the protein as a whole (Hiemstra et al., 1996; Tomee et al., 1997). This fits with the theory that the C-terminus is capable of protease inhibition and suggests that the two WFDC domains of SLPI exhibit separate functions (Bingle and Vyakarnam, 2008; Eisenberg et al., 1990).

It is feasible that SLPI and elafin impact bacterial and fungal viability by binding to cells and disrupting their membranes to induce lysis as a result of their cationic charge. Many AMPs carry out their killing via cationic lysis (Wiesner and Vilcinskas, 2010), including a member of the WFDC protein family known as eppin (Yenugu et al., 2004). Investigations into the bacterial-binding capacity of SLPI showed that it could coat the surface of *Mycobacteria* strains (Gomez et al., 2009). Similarly, elafin and pre-elafin were shown to bind to Gram-negative bacteria (*P. aeruginosa*), Gram-positive bacteria (*S. aureus*) and fungi (*C. albicans*), appearing to induce moderate membrane disruption, as visualised by scanning electron microscopy (SEM) (Baranger et al., 2008; Bellemare et al., 2010). After incubation with recombinant elafin and pre-elafin, *P. aeruginosa* was shown to exhibit membrane blebbing, blistering and the presence of pore-like structures (Bellemare et al., 2010). It should be noted, however, that the membrane disruption and subsequent depolarisation was moderate compared to the lytic effects exhibited by the control AMP, magainin 2 (Bellemare et al., 2010).

SLPI and elafin appeared to show much weaker antimicrobial effects when compared to other AMPs including lysozyme, defensins and magainin (Bellemare et al., 2010; Hiemstra et al., 1996; Payne et al., 2017; Tomee et al., 1997). The high concentrations of SLPI and elafin required to induce significant antimicrobial effects may suggest that such activities are not the primary functions of these proteins.

Interestingly there is evidence to show that both SLPI and elafin are capable of priming innate immune responses. Gomez et al (2009) demonstrated that SLPI was able to coat *M. bovis* and *M. tuberculosis* in a process that appeared to be facilitated by interactions between SLPI and pathogen-associated molecular patterns (PAMPs) such as surface-associated lipids.

This, in turn, appeared to lead to more effective phagocytosis (Gomez et al., 2009). Similarly, pre-elafin was reported to coat the surface of *P. aeruginosa* cells to induce phagocytosis by macrophages, in a process seemingly mediated by interactions between pre-elafin and CD14 (Wilkinson et al., 2009). Opsonisation with pre-elafin also appeared to induce neutrophil recruitment (Wilkinson et al., 2009). Clearly the antimicrobial activities of SLPI and elafin are complex and further analysis is required to advance the understanding of their role in the innate immune system.

1.3.3. Putative antimicrobial activity of WFDC2

In contrast to SLPI and elafin, evidence supporting an antimicrobial role for WFDC2 is sparse and largely consists of circumstantial evidence i.e. correlation between bacterial colonisation and an increase in WFDC2 expression. For instance, *in vitro*, when the endocervical cell line End1 was exposed to a bacterium associated with bacterial vaginosis (*Atopobium vaginae*), the concentration of WFDC2 in the conditioned media was significantly increased (Eade et al., 2015). In patient samples an upregulation in WFDC2 was also apparent: the vaginal secretions of healthy women whose vaginal microbiomes were predominated with *Gardnerella vaginalis* (a biofilm-forming bacterium responsible for causing bacterial vaginosis) exhibited raised concentrations of WFDC2 compared to microbiomes dominated by commensal species (Orfanelli et al., 2014). Additionally, in children with cystic fibrosis (CF), those with bacterial colonisation of the lung had significantly higher serum WFDC2 levels compared to those without bacterial colonisation (Nagy et al., 2016).

Thus far, the only analysis to directly determine whether WFDC2 is capable of antimicrobial activity is that described by Hua et al (2014) who used recombinant human WFDC2 for *in vitro* binding and growth inhibitory assays. Their results showed binding of a WFDC2-Fc fusion protein to Gram-positive strains (*S. aureus* and *Bascillus subtilis*) and minor binding to Gram-negative strains (*E. coli* and *P. aeruginosa*). Two recombinant forms of WFDC2 were used in the binding assay: one produced in yeast and the other in mammalian cells; the WFDC2 protein produced in mammalian cells appeared to have a greater propensity to bind to bacteria

compared to the protein produced in yeast cells, which the author suggests implicates the importance of glycosylation for interactions to occur.

Despite the positive binding between WFDC2 and *S. aureus* described by Hua et al (2014)*,* when recombinant WFDC2 was incubated with *S. aureus* for 12 hours, it appeared to have no significant effect on *S. aureus* growth, even at high concentrations (7.75 uM). Further analyses with other strains will be required to deduce the effect of human WFDC2 on bacterial growth.

Investigations into the functions of WFDC2 orthologues are also limited. The antimicrobial properties of murine Wfdc2 have been investigated against attenuated *Mycobacterium bovis* (BCG) using viability and membrane permeability assays. The data showed that murine Wfdc2 was unable to induce membrane permeabilization or killing of *M.* bovis (Nishimura et al., 2008). Within the same study, it was shown that murine Slpi (1 µM) was capable of almost complete inhibition of three *Mycobacteria* strains and that incubation of Slpi with *M. bovis* caused pronounced damage to the cell surface as shown by SEM and membrane permeabilization assays (Nishimura et al., 2008).

The tammar wallaby WFDC2 orthologue has also been used for *in vitro* assays against various bacteria strains. The recombinant protein elicited significant antimicrobial activity against *S. aureus*, *P. aeruginosa* and *Salmonella enterica* (Watt et al., 2012). Interestingly, when constructs encoding single WFDC domains were utilised in the same assays, the C-terminal domain alone was found to be capable of inhibiting growth at a similar rate to the full-length protein, while the N-terminus had no effect on bacterial growth. When the WFDC2 proteins were incubated with the commensal *Enterococcus faecalis*, no growth inhibition occurred with full-length WFDC2 or the single WFDC domains (Watt et al., 2012). It should be noted that the concentration of WFDC2 used in the assays was not described, so it is unclear whether the bactericidal effects occurred at physiological concentrations. Additionally, the molecular weight of full-length recombinant wallaby WFDC2 was reported to be 13 kDa, which is significantly smaller than human WFDC2 (despite being two amino acids longer than the human orthologue). The group have suggested that wallaby WFDC2 may be involved in

antimicrobial defence of the mammary gland in females during pregnancy and involution, but also to protect the young during lactation. An absence of the pathogenic bacterial strains (*S. aureus*, *P. aeruginosa*, *S. enterica*) in the gut of the developing young, but presence of *E. faecalis* (to which WFDC2 has no bactericidal activity) suggests that WFDC2 may be involved in the development and selection of commensals in the gut microbiota of developing young.

The limited data available in the literature appears to suggest that human and murine WFDC2 are not directly capable of antimicrobial activity, although further studies will be necessary to add substance to this claim.

1.3.4. Induction of expression of WFDC proteins

Both SLPI and elafin have been shown to exhibit increased mRNA expression and/or protein production in response to stimuli such as inflammatory cytokines (interleukin 1β; IL-1β and tumour necrosis factor α; TNF-α), bacterial PAMPs (lipopolysaccharide; LPS) and host proteases (neutrophil elastase), although the reported changes in expression are inconsistent between studies (see Table 1.3). For example, SLPI mRNA levels were shown to be upregulated in two pulmonary cell lines following exposure to TNF-α and IL-1β (Sallenave et al., 1994), and an intestinal epithelial cell line secreted higher concentrations of SLPI protein following TNF-α and IL-1β exposure (Si-Tahar et al., 2000). In contrast, primary tracheobronchial epithelial (TBE) cells and primary alveolar epithelial pneumocytes (Type II cells) showed no change in SLPI mRNA expression following exposure to TNF-α and IL-1β (Bingle et al., 2006), and primary murine macrophages showed no increase in murine Slpi expression after exposure to the same agents (Jin et al., 1998). These discrepancies may demonstrate cell type-specific differences in SLPI responses or may simply reflect differences in experimental techniques. Although many differences between results were apparent, some consistencies were also found; for example, elafin mRNA levels have been shown to be upregulated in response to IL-1β in all experiments described (Bingle et al., 2006; Bingle et al., 2001; Reid et al., 1999; Sallenave et al., 1994) (see Table 1.3). The type of mediators that can affect SLPI and elafin expression support the proteins' hypothesised dual role of defence

against pathogenic threat from the external environment and protection against excessive host-derived inflammation.

	Target	Treatment	Result	Reference
SLPI	Tracheal epithelial cell line (9HTEo-)	Elastase IL-8 TNF-α Trypsin	SLPI mRNA expression upregulated by elastase and trypsin; TNFα and IL-8 had no effect	Abbinante-Nissen et al., 1993
	Primary mouse macrophages*	IL-1β IL-6 IL-10 LPS TNF-α	SLPI mRNA expression upregulated by LPS, IL-10 and IL-6; TNF-α and IL-1β had no effect	Jin et al., 1998
	Two pulmonary cell lines (H322 and A549)	IL-1β TNF-α	SLPI mRNA expression upregulated by IL-1β and TNF-α	Sallenave et al., 1994
	Primary TBE cells and primary alveolar epithelial pneumocytes (Type II cells)	IL-1β LPS TNF-α	SLPI mRNA expression unaffected by IL-1β, LPS and TNF-α	Bingle et al., 2006
	Primary amniotic cells	IL-1α IL-1β TNF-α	SLPI protein production upregulated by IL-1α, IL-1β and TNF-α	Zhang et al., 2001
	Intestinal epithelial cell line (Caco2-BBE)	IFN-γ IL-1β TNF-α	SLPI protein production upregulated by IL-1β and TNF-α; IFN-γ had no effect	Si-Tahar et al., 2000

	Target	Treatment	Result	Reference
Elafin	Primary alveolar epithelial pneumocytes, two pulmonary cell lines (H322 and A549)	IL-1 LPS TNF-α	Elafin mRNA expression upregulated by IL-1 and TNF-α; LPS had no effect	Bingle et al., 2001
	Pulmonary cell line (A549)	Elastase	Elafin mRNA expression upregulated by elastase	Reid et al., 1999
	Primary TBE cells and primary alveolar epithelial pneumocytes (Type II cells)	IL-1β LPS TNF-α	Elafin mRNA expression upregulated by IL-1β; LPS and TNF-α had no effect	Bingle et al., 2006
	Two pulmonary cell lines (H322 and A549)	IL-1β TNF-α	Elafin mRNA expression upregulated by IL-1β and TNF-α	Sallenave et al., 1994

Table 1.3. The effects of pro-inflammatory stimuli on SLPI and elafin expression. SLPI/elafin mRNA expression or protein production were monitored following cell treatment with different pro-inflammatory stimuli. The target cell type, treatment and results are summarised for each study. References to the relevant literature are given. *Murine Slpi expression was assessed in the primary mouse macrophages.

When similar exposure assays were used to monitor WFDC2 expression, there was little evidence to suggest that its expression was stimulated in response to cytokines or LPS. Bingle

et al (2006) exposed primary TBE and Type II cells to IL-1β, TNF-α and LPS and studied the effect on WFDC2 expression via RT-PCR. No change in WFDC2 mRNA level was observed in TBE cells following any of the treatments, while the Type II cells showed a slight increase in WFDC2 following exposure to IL-1β only (but it is unclear if this change was statistically significant). In another exposure study, ovarian cancer cells were exposed to IL-1β and TNF-α and changes in gene expression were monitored by RT-PCR. No changes in WFDC2 mRNA levels were observed following cytokine exposure, while elafin expression was shown to be upregulated by both cytokines (Clauss et al., 2010).

Aside from the two semi-quantitative RT-PCR studies, the direct effect of inflammatory mediators upon WFDC2 expression has not been investigated. Further analysis is required to determine what factors influence its transcriptional regulation, given that the WFDC2 protein is overexpressed in a number of inflammatory diseases including various cancers, cystic fibrosis and renal fibrosis, as discussed in more detail in Sections 1.4 and 1.5.

1.3.5. Anti-protease activity in the WFDC family

Although SLPI and elafin have been shown to exhibit some antimicrobial activity, they are best known for their anti-inflammatory role as protease inhibitors in the respiratory tract. Anti-proteases impose a level of control over the host immune system by preventing inappropriate inflammation. SLPI and elafin are classified as 'alarm' protease inhibitors (Sallenave, 2000), and are thus produced locally in response to 'alarms', or danger signals, such as proteases, LPS and cytokines. The group of WFDC proteins encoded on chromosome 20 are assumed to inhibit proteases as a result of their similarity to SLPI and elafin, but this theory is yet to be extensively tested for the majority of WFDC proteins, including WFDC2 (Clauss et al., 2002).

SLPI and elafin have had their anti-protease activity thoroughly established over the last 40 years, and are proven to be active against a variety of proteases including elastase (Ying et al., 1994; Ying and Simon, 1993), cathepsin G (Smith and Johnson, 1985), trypsin (Smith and Johnson, 1985), chymotrypsin (Franzke et al., 1996; Smith and Johnson, 1985), proteinase-3

(Wiedow et al., 1991) and chymase (Walter et al., 1996). Both proteins exhibit identical positioning of their conserved cysteine residues within the active site, which appears to be a requirement for successful protease inhibition (Grütter et al., 1988; Kramps et al., 1990; Tsunemi et al., 1996). WFDC2 does not exhibit the same consensus of cysteines as SLPI's C-terminal domain and elafin; resultantly, it has been suggested that WFDC2 is incapable of protease inhibition (Bingle and Vyakarnam, 2008).

Nonetheless, some data in the literature argues that WFDC2 is indeed capable of protease inhibition despite its alternative structure. Hua et al (2014) showed that recombinant WFDC2 produced in mammalian cells was capable of limiting trypsin and elastase activity *in vitro*. Incubation with 6.4 μM of WFDC2 induced a 45% reduction in elastase activity, whereas 10.4 μM of WFDC2 caused a 30% decrease in trypsin activity. Within these assays, WFDC2 concentrations as high 29.9 μM were used to demonstrate increasing inhibition of enzyme activity. There is limited data in the literature describing the physiological concentrations of WFDC2 at sites in the human body, therefore it is difficult to determine whether the concentrations of protein used in these assays are physiologically relevant.

Zymography was also used to analyse WFDC2 inhibition towards matrix metalloproteases (MMPs). The results showed that MMP9 gelatin degradation could be inhibited by WFDC2, while WFDC2 had no effect on MMP2 activity; however, there is no information given regarding the concentration of protein utilised in the assay (Hua et al., 2014).

During the same study, Hua et al (2014) used the disk diffusion technique to determine whether WFDC2 is capable of inhibiting the activity of proteases derived from *B. subtilis* and *P. aeruginosa.* Bacteria plated upon a skim-milk plate incubated alongside WFDC2 exhibited a 40% reduction in proteolytic clearing upon the plate compared to controls.

Protease inhibitory activity was also ascribed to WFDC2 in another study, whereby Chhikara et al (2012) demonstrated that WFDC2 could inhibit trypsin, chymotrypsin, PSA, protease K, pepsin and papain. Although their data showed a greater than 70% reduction in activity for all

proteases tested, even at 0.4 uM, the results lack appropriate negative controls to show enzyme activity when WFDC2 is absent which impacts the reliability of the data. Additionally, the WFDC2 protein used in the study was isolated from seminal plasma and proposed to exist as a 14 kDa monomer and as a 42 kDa trimer in native conditions (Chhikara et al., 2012). Given that the glycosylated, monomeric form of WFDC2 is commonly observed as being around 20 kDa in size, an observed size of 14 kDa is unusual, however, the authors suggest that the discrepancies in size are due to tissue-specific differences i.e. between ovarian and prostate cell lines.

Fibrotic kidney samples express enhanced serine protease expression compared to healthy tissues which causes enhanced collagen degradation and thus exacerbates the symptoms of fibrosis. The relationship between WFDC2 and proteases in renal fibrosis was investigated by LeBleu et al (2013) using colorimetric assays to analyse residual protease activity in fibrotic tissue following incubation with recombinant WFDC2. The data showed that protease activity in fibrotic renal tissue was significantly higher than in non-fibrotic tissue, however, following incubation with recombinant WFDC2 the protease activity of fibrotic tissue decreased to a level close to the non-fibrotic baseline. Similarly, collagen digestion by trypsin, Prss35, Prss23, MMP2 and MMP9 was measured via hydroxyproline release and was found to be significantly reduced when WFDC2 (390 nM) was added to the assay (LeBleu et al., 2013).

The concentrations of WFDC2 required to induce successful protease inhibition vary greatly between studies. For example, in the case of trypsin, a 25-fold higher concentration of WFDC2 was used by Hua et al (2014) to significantly reduce protease activity to a level comparable to that described by LeBleu et al (2013). These variances perhaps reflect the experimental methods used, including differences in substrates.

Evidently there is some data supporting a protease inhibitory role for WFDC2, however, further analysis is required to give a greater understanding of WFDC2's effect on enzyme activity and to determine whether it is directly capable of inhibition.

1.4. WFDC2 and cancer

WFDC2 is markedly upregulated in a range of diseases. Its expression is known to be amplified in a number of cancers including ovarian, lung and endometrial cancer. In the case of ovarian cancer, WFDC2 is expressed to such an extent that it is used clinically as a biomarker for detection of the disease (Moore et al., 2010; Moore et al., 2009; Ruggeri et al., 2011). Such biomarkers are valuable clinical tools as serum assessment is simple, low cost, non-invasive and allows long-term monitoring of disease status.

1.4.1. WFDC2 as a biomarker for ovarian cancer

Ovarian cancer is the fifth most common cause of cancer death in women in the UK, with an expected 10-year survival rate of 35% (Cancer Research UK, 2018b). Around 3/4 of patients are diagnosed at an advanced stage, largely due to an absence of disease-specific symptoms, by which point anti-cancer therapies are markedly less effective and mortality rates are high (Thibault et al., 2014). As a result, biomarkers for cancer detection are an extremely valuable resource when used alongside transvaginal ultrasonography for detection of the disease at an early stage.

Cancer antigen 125 (CA125), an epitope of mucin 16 (MUC16), has been widely used over the past three decades as an FDA-approved biomarker for ovarian cancer detection (Bast et al., 1983). MUC16 is a high molecular weight glycoprotein that is expressed at healthy epithelial surfaces but is abundantly shed from ovarian cancer cells and can be subsequently detected in patient serum (Argüeso et al., 2003; Bast et al., 1983; Davies et al., 2007; Yin and Lloyd, 2001). Although CA125 has proven to be an effective biomarker, particularly for detection of advanced disease, its sensitivity is low and false positive results are frequently found in benign diseases such as endometriosis and due to natural fluctuations in CA125 associated with menstruation (Díaz-Padilla et al., 2012; Medeiros et al., 2009; Nustad et al., 1996). Subsequently, research has been ongoing to identify more sensitive and reliable biomarkers.

Serial analysis of gene expression (SAGE) technology has been a powerful tool in the identification of potential biomarkers for ovarian cancer, allowing the differential expression of thousands of genes to be quantified and compared between healthy and malignant cells (Hough et al., 2000; Ismail et al., 2000; Ono et al., 2000; Schummer et al., 1999; Welsh et al., 2001). This technology resulted in the identification of WFDC2 as a gene that is elevated in ovarian tumours and subsequently monoclonal antibodies were developed and used for WFDC2 quantification by ELISA (Hellström et al., 2003).

Analyses of WFDC2 concentrations in serum showed that approximately 80% of ovarian cancer patients have serum WFDC2 levels over 150 pM (Molina et al., 2011; Ruggeri et al., 2011). The cut-off point used to identify patients with an ovarian malignancy is generally around 70 pM (Braicu et al., 2013). High serum WFDC2 concentrations correlated with poor prognosis and surgical outcome (Bandiera et al., 2011; Braicu et al., 2013; Chen et al., 2014; Molina et al., 2011; Moore et al., 2014).

As a result of these findings, WFDC2 is currently being used as a predictive biomarker for ovarian cancer alongside CA125 in the FDA-approved algorithm known as the Risk of Ovarian Malignancy Algorithm (ROMA). ROMA takes into account the serum concentration of CA125 and WFDC2 alongside menopausal status to predict the likelihood of ovarian cancer (Moore et al., 2009; Moore et al., 2011). Studies analysing the efficacy of ROMA have reported mixed results, with some arguing that WFDC2 improves sensitivity compared to CA125 alone, particularly in detecting disease at an early stage and in differentiating between benign and malignant pelvic masses (Bandiera et al., 2011; Chan et al., 2013; Karlsen et al., 2012; Molina et al., 2011; Ruggeri et al., 2011). Conversely, other studies claim that the combination of biomarkers does not improve upon their individual sensitivity and specificity (Jacob et al., 2011; Van Gorp et al., 2011).

When histologically studying WFDC2 expression in ovarian carcinomas, staining was often diffuse throughout the tumour and localised to the cytoplasm/secretions (Georgakopoulos et al., 2012; Uhlen et al., 2017) (Fig 1.11). WFDC2 staining was more prevalent in serous and

endometrioid tumours, whereas more variable positivity was seen in clear cell and mucinous carcinomas (Drapkin et al., 2005; Galgano et al., 2006; Georgakopoulos et al., 2012; Schaner et al., 2003). Interestingly, this difference in positive WFDC2 expression appeared to be reflected in the serum concentrations of patients with different tumour subtypes (Hertlein et al., 2012). The variability in expression between ovarian cancer subtypes impacts its reliability as a biomarker; nonetheless, WFDC2 has proven to be effective in distinguishing between benign and malignant disease (Anastasi et al., 2013; Drapkin et al., 2005; Georgakopoulos et al., 2012; Ortiz-Muñoz et al., 2014; Zhang et al., 2014).
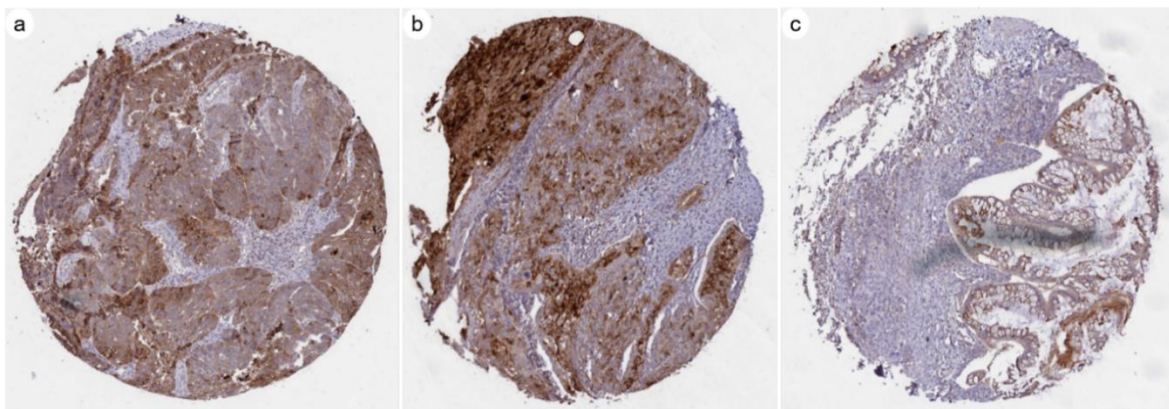
Fig 1.11. WFDC2 staining in ovarian malignancies. Immunohistochemical analysis of WFDC2 expression in ovarian malignancies, taken from Protein Atlas (Uhlén et al., 2017). Staining was analysed in serous (a), endometrioid (b) and mucinous (c) ovarian malignancies. Patient IDs: 2347 (a), 2568 (b) and 1844 (c). The antibody utilised was from Atlas Antibodies (HPA042302) developed using recombinant WFDC2 (Glu31- Phe124) as the immunogen. Accessed in December 2018. Data available from: https://www.proteinatlas.org/ENSG00000101443-WFDC2/pathology/tissue/ovarian+cancer#img

1.4.2. WFDC2 as a biomarker for other cancer types

WFDC2 was also found to be upregulated in the serum of patients with other types of cancer compared to controls, for instance, those suffering from endometrial (Brennan et al., 2014; Capriglione et al., 2015; Wang et al., 2017) and lung cancer (Choi et al., 2017; Liu et al., 2013). WFDC2 upregulation was also evident when endometrial and lung cancer tissues were analysed by immunohistochemistry (Fig 1.12). Although the serum levels in patients with ovarian cancer are often significantly higher than in patients with other types of cancer, WFDC2 still appears to show promise for use as a biomarker.
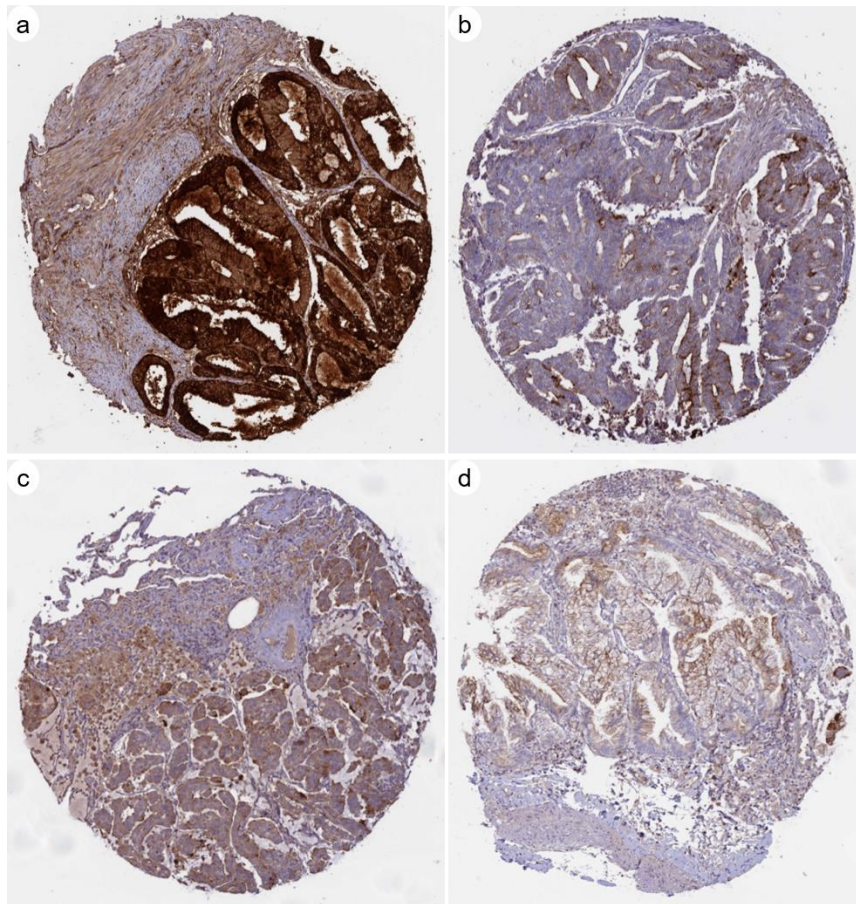
Fig 1.12. WFDC2 staining in endometrial and lung malignancies. Immunohistochemical analysis of WFDC2 expression in non-ovarian malignancies, taken from Protein Atlas (Uhlén et al., 2017). Staining was analysed in endometrial carcinoma (a, b) and adenocarcinoma of the lung (c, d). Patient IDs: 2607 (a), 2118 (b), 1847 (c) and 3391 (d). The antibody utilised was from Atlas Antibodies (HPA042302) developed using recombinant WFDC2 (Glu31- Phe124) as the immunogen. Accessed in December 2018. Endometrial cancer data available from:

https://www.proteinatlas.org/ENSG00000101443-WFDC2/pathology/tissue/endometrial+cancer#ihc

Lung cancer data available from:

https://www.proteinatlas.org/ENSG00000101443-WFDC2/pathology/tissue/lung+cancer#ihc

In endometrial carcinomas WFDC2 is frequently heightened in patient serum and concentrations have been reported to reflect disease stage (Abbink et al., 2018; Brennan et al., 2014; Capriglione et al., 2015; Dewan et al., 2017; Wang et al., 2017; Yılmaz et al., 2017; Zanotti et al., 2012). High WFDC2 levels have been hypothesised to correlate with poor prognosis. Unlike ovarian cancers, endometrial malignancies are often symptomatic and detected at an early stage, therefore the 10 year survival rate is 78% (Cancer Research UK,

2018c). Nonetheless, biomarkers are still a valuable tool for assessing patient prognosis and for customising treatment. CA125 has been investigated for its use as an endometrial cancer biomarker, but its sensitivity and specificity are often poor (Abbink et al., 2018; Moore et al., 2008; Powell et al., 2005), therefore WFDC2 is currently being studied as a potential candidate biomarker to be used alone or in combination with other biomarkers such as CA125 (Abbink et al., 2018; Brennan et al., 2014; Dewan et al., 2017; Wang et al., 2017).

WFDC2 is also being investigated as a potential biomarker for detection of lung cancer. As with ovarian cancer, lung cancer is often asymptomatic and prognosis is poor, with a 10 year survival rate of only 5% (Cancer Research UK, 2018a). A number of proteins are under investigation for use as serum biomarkers for detecting lung carcinomas, including carcinoembryonic antigen (CEA), neuron-specific enolase (NSE), cytokeratin-19 fragment CYFRA 21.1 and WFDC2 (Kulpa et al., 2002; Liu et al., 2017; Lou et al., 2014; Tas et al., 2000; Zhang et al., 2017). WFDC2 has been shown to be upregulated in both small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) (Jiang et al., 2014; Liu et al., 2013; Wang et al., 2014), and serum concentrations of WFDC2 appeared to correlate with disease progression and poor prognosis (Choi et al., 2017; Iwahori et al., 2012; Lamy et al., 2015; Yamashita et al., 2012). Additionally, high WFDC2 levels appeared to be associated with a lack of responsiveness to chemoradiotherapy in patients with advanced NSCLC (Lan et al., 2016).

Immunohistochemical staining of lung tumour tissue showed that NSCLC tissues, particularly adenocarcinomas and bronchioalveolar carcinomas, more frequently stained positively (>80%) for WFDC2 expression than any other pulmonary tumour type and staining tended to be distributed throughout the tumour (Bingle et al., 2006). WFDC2 may be a valuable biomarker for detection of NSCLC and SCLC to be used independently or in combination with other promising biomarkers such as CYFRA 21.1, CEA and NSE.

Oligonucleotide microarray analysis performed by Galgano et al (2006) to quantify the gene expression of WFDC2 in 11 different carcinoma types showed that, as may be expected,

WFDC2 mRNA levels were most abundant in serous ovarian cancers, followed by lung adenocarcinomas. Additionally, the microarray data showed moderate WFDC2 levels in pancreatic, breast, gastric and renal malignancies. This expression data is supported by histopathological analyses which showed an abundance of positive WFDC2 staining in tumour tissues derived from breast (Kamei et al., 2010), stomach (Guo et al., 2015) and salivary gland (Galgano et al., 2006), demonstrating a significant upregulation of WFDC2 in many non-gynaecological malignancies.

Although WFDC2 expression is upregulated to a moderate extent in many malignant diseases (i.e. gastric, pancreatic and breast cancer), it appears that the increase in gene expression does not result in an increase in serum WFDC2 levels to allow reliable differentiation between healthy and diseased patients. This is evidenced by Hertlein et al (2012), who measured serum WFDC2 concentrations from hundreds of patients with benign and malignant disorders and compared them with those of healthy patients (Fig 1.13). Patients with ovarian cancer showed a median WFDC2 level of 242 pM, which is over 6-fold higher than in healthy volunteers, while patients with lung cancer showed a 2-fold higher median concentration of WFDC2 than controls. Gastric and renal cancer show only slightly elevated WFDC2 levels compared to healthy patients, with concentrations being comparable to those reported in patients with benign gastrointestinal disease or renal insufficiency, thus evidencing WFDC2's lack of sensitivity or specificity in detecting such malignancies (Fig 1.14) (Hertlein et al., 2012).

WFDC2 serum levels in cancer patients
Hertlein et al. (2012)
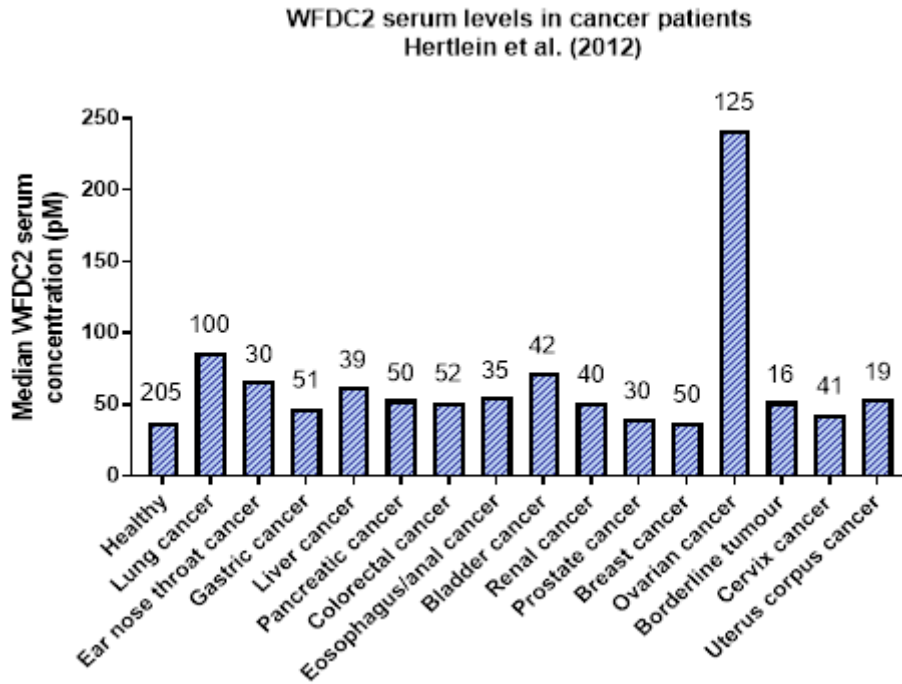
Fig 1.13. The serum concentration of WFDC2 in patients with malignant disease. A graph to show the median serum WFDC2 concentration (pM) in patients with malignant disease, as described by Hertlein et al (2012). The number of samples per group is listed (number above each bar).

WFDC2 serum levels in patients with benign disease
Hertlein et al. (2012)
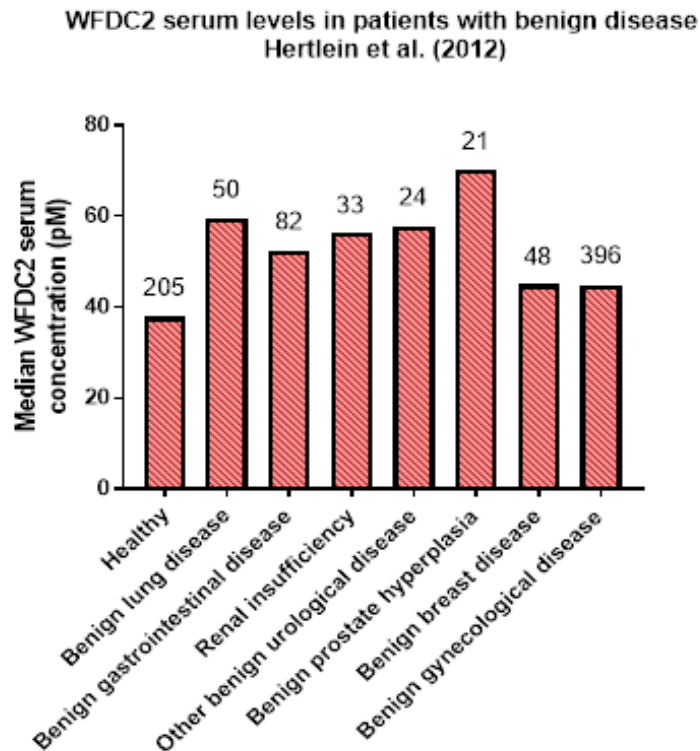
Fig 1.14. The serum concentration of WFDC2 in patients with benign disease. A graph to show the median serum WFDC2 concentration (pM) in patients with benign disease, as described by Hertlein et al (2012). The number of samples per group is listed (number above the bar).

1.4.3. Amplification of the 20q13 locus

In many different cancer types, the process of tumorigenesis appears to correlate with upregulated WFDC2 expression and elevated protein secretion. This process is particularly evident in ovarian and gastric malignancies, as healthy tissues at these sites appear to be completely devoid of WFDC2 expression, yet when gastric and ovarian cells become malignant they are shown to positively express WFDC2 (Guo et al., 2015; Uhlen et al., 2017; Uhlén et al., 2015). One explanation for the change in WFDC2 expression is the frequent amplification of the WFDC2-expressing locus, 20q13, which is one of the most common genomic aberrations that occurs in many cancer types.

In ovarian tumours, locus 20q13 has been consistently reported to be one of the most frequently amplified loci when analysed via comparative genomic hybridisation (CGH) (Israeli et al., 2005; Kiechle et al., 2001; Sonoda et al., 1997; Tanner et al., 2000). Frequent DNA amplification has been reported in low and high grade ovarian carcinomas but appears to be absent in benign masses (Iwabuchi et al., 1995; Tominaga et al., 2010).

Copy number aberrations at the 20q13 locus have also been linked to many non-ovarian cancers including gastric (Nishimura, 2008), lung (Zhu et al., 2007), breast (Larramendy et al., 2000), salivary gland (Uchida et al., 2010), prostate (Tabach et al., 2011) and cervical tumours (Scotto et al., 2008), in addition to brain metastases (Petersen et al., 2000), following CGH analysis of cancer cell lines or primary tumour tissue.

The amplification of locus 20q13 has been reported to result in genome instability, a factor that may go some way to explaining the frequent correlation between high WFDC2 expression and malignancy (Savelieva et al., 1997). The recurrent amplification suggests that genes encoded at this locus may be involved in tumour progression, however, candidate genes are yet to be identified.

1.4.4. WFDC2 and metaplasia

Interestingly, WFDC2 has also been implicated in cases of metaplasia. Microarray analysis of murine gastric cells induced to undergo metaplastic transition indicated that Wfdc2 was one of the most highly upregulated factors in metaplastic cells (Nozaki et al., 2008; Weis et al., 2014). Histological and microarray analysis has also shown that WFDC2 becomes upregulated in human gastric and intestinal metaplasias (Weis and Goldenring, 2009). It is possible that WFDC2 becomes upregulated in gastric metaplasia as a result of the instability of the 20q13 locus.

1.4.5. WFDC2 and cancer cell behaviours

Evidently WFDC2 is overexpressed in a number of cancers and some cases of metaplasia, but it is yet to be determined whether the gene is actively involved in downstream tumour-promoting activities or whether it is simply an innocent bystander.

WFDC2 is expressed in a number of cancer cell lines, including those derived from ovarian and endometrial tumours (Bingle et al., 2006; Drapkin et al., 2005), providing a suitable tool for knockdown studies. Various attempts to engineer stable WFDC2 overexpression and knockdown cell lines have been described in the literature, with the aim of ascertaining the potential impact of WFDC2 expression on the behaviour of cancer cells. Unfortunately, a considerable proportion of the literature fails to suitably evidence the successful knockdown/overexpression of WFDC2, or otherwise shows a mildly altered cell phenotype in the resultant clones. Within the better-evidenced studies, the reported effects of WFDC2 on cellular behaviour are described below.

Studies using WFDC2 overexpressing ovarian and endometrial cancer cell lines have shown enhanced proliferation rates in comparison to controls (Li et al., 2013; Zhu et al., 2016). Similarly, endometrial and pancreatic cancer cell lines treated endogenously with recombinant WFDC2 appear to proliferate more rapidly than untreated cells (Lu et al., 2016). This *in vitro* data is supported by *in vivo* analyses using xenograft mouse models, whereby animals

injected with WFDC2 overexpression clones developed tumours that were significantly larger than those derived from controls (Li et al., 2013; Zhu et al., 2016). Conversely, cancer cell lines that had endogenous WFDC2 expression knocked-down using RNA interference exhibited reduced proliferation/viability rates compared to controls (Guo et al., 2015; Zhu et al., 2016). Xenograft tumours derived from WFDC2-knockdown clones were also significantly smaller (Lu et al., 2012; Zhu et al., 2016).

Interestingly, within the scope of their xenograft studies, Zhu et al (2016) also monitored the formation of metastatic lung nodules and found that animals injected with WFDC2-overexpressing ovarian cancer cells had significantly greater metastatic nodules than control animals. Correspondingly, no metastatic lung nodules were found in animals that had been injected with WFDC2-knockdown ovarian cancer cells, whereas 40% of mock controls exhibited metastasis, and 100% of animals injected with WFDC2-overexpressing ovarian cancer cells had lung nodules (Zhu et al., 2016). These observations of changes in invasive capacity are further supported by *in vitro* wound healing and Transwell assays, which show that WFDC2-overexpressing ovarian cancer clones appear to be more migratory and invasive (Lu et al., 2012; Zhu et al., 2016). A WFDC2-knockdown gastric cancer cell line was also reported to exhibit reduced migration compared to controls (Guo et al., 2015).

Although the results from these studies are, for the most part, in agreement with each other, one study using WFDC2-overexpression clones reported opposing findings whereby mice injected with WFDC2-overexpressing ovarian cancer cells exhibited significantly smaller tumours than animals injected with mock-transfected cells. (Gao et al., 2011). Within the same study, *in vitro* assays showed that overexpression clones also had increased rates of apoptosis compared to controls, in addition to a reduced capacity to migrate and invade (Gao et al., 2011). Lu et al (2012) also showed that overexpression of WFDC2 did not have an effect on proliferation compared to the controls.

Largely, the *in vitro* evidence appears to be consistent in reporting an increase in cell proliferation, invasion and migration when WFDC2-overexpressing clones are analysed, with

downstream xenograft studies replicating this result *in vivo*, showing increased tumour growth and metastasis. The converse was frequently reported when WFDC2-knockdown clones were assessed in *in vitro* and xenograft models.

1.5. WFDC2 and benign diseases

Upregulation of WFDC2 is also evident in benign diseases such as cystic fibrosis (CF) (Bingle et al., 2006; Nagy et al., 2016), idiopathic pulmonary fibrosis (Raghu et al., 2018) and renal fibrosis (LeBleu et al., 2013). Although several fibrotic diseases do express high levels of WFDC2, its expression is not upregulated in liver fibrosis or cirrhosis, so cannot be viewed as a universal marker of fibrosis, *per se* (Zhang et al., 2018). Nonetheless, the wide range of benign diseases characterised by WFDC2 overexpression only serves to heighten the ambiguity surrounding its function.

1.5.1. WFDC2 and benign lung diseases

Elevated levels of WFDC2 have been identified in the serum of patients with a number of benign lung diseases, including pulmonary tuberculosis, pneumonia, idiopathic pulmonary fibrosis and CF (Liu et al., 2013; Nagy et al., 2016; Raghu et al., 2018; Yamashita et al., 2012). Consequently, if WFDC2 was to be considered as a potential biomarker for lung cancer, a higher serum cut-off concentration would be required in order to distinguish between patients with malignant and benign disease, compared with healthy individuals and cancer patients (Liu et al., 2013).

Histological studies carried out by Bingle et al (2006) have been used to further analyse the upregulation of WFDC2 that occurs in patients with CF. Chronically inflamed tissue derived from the peripheral lung of CF patients exhibited extensive staining compared to disease-free peripheral lung sections (Fig 1.15), yet little difference in expression was identified between CF and healthy tissue sections derived from the large airways (Bingle et al., 2006). The pattern of expression within these lung sections suggests that WFDC2 is secreted into the lumen of

diseased airways; a theory supported by the elevated level of WFDC2 in the sputum of CF patients compared to healthy controls (Nagy et al., 2016).
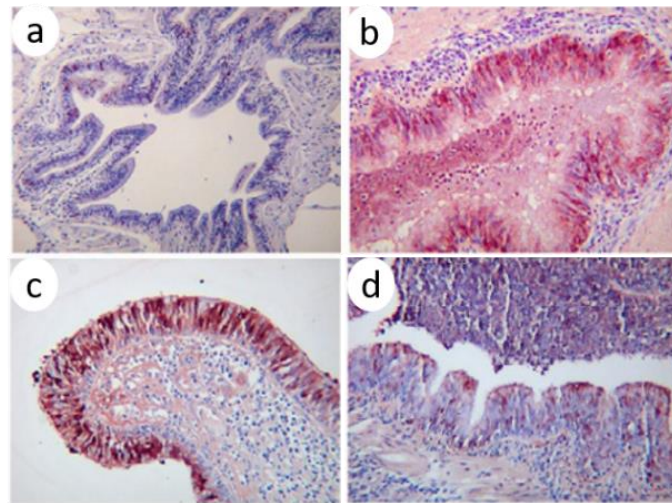
Fig 1.15. Analysis of WFDC2 staining in the cystic fibrosis lung. Immunohistochemical analysis of WFDC2 expression in normal lung tissue (a) compared to cystic fibrosis lung (b, c, d), taken from Bingle et al (2006). Original magnification of the images was X100 (a) and X200 (b, c, d). The antibody utilised was developed by Hellström et al (2003) using recombinant full-length WFDC2 synthesised in mammalian cells as the immunogen.

More recently the value of WFDC2 as a serum biomarker for CF has been under investigation. In a study involving 144 CF patients, graded with mild, moderate or severe disease, the serum concentration of WFDC2 was measured and compared to that of 94 patients with non-CF lung disease (e.g. patients with chronic bronchitis, asthma, pneumonia or bronchiectasis) and 117 healthy controls (Nagy et al., 2016). The authors reported elevated median serum levels in all CF patients compared to normal controls, regardless of age or genotype. Compared to healthy controls, patients without CF but afflicted with other lung diseases had higher average WFDC2 levels than healthy controls, however, serum concentrations were still significantly lower in non-CF lung disease patients compared to CF patients (Nagy et al., 2016). This hierarchy of WFDC2 expression enables the identification of two WFDC2 concentration cut-off points to be determined, with one value to discriminate between CF patients and normal controls (49.3 pM), and another cut-off value to distinguish between CF and non-CF diseased patients (98.4 pM). These values are comparable with those suggested by Liu et al (2013) when studying

differences in serum concentrations between patients with lung malignancies and those with pulmonary tuberculosis or pneumonia.

In a trend comparative to that seen in advanced stage cancer patients, WFDC2 concentration appeared to be positively correlated with disease severity in CF patients, as demonstrated in adult and child cohorts (Nagy et al., 2016).

Interestingly, the WFDC2 levels in juvenile CF patients diagnosed with chronic bacterial colonisation (n=54) were significantly raised in sera compared to patients lacking colonisation (n=23) (Nagy et al., 2016). This preliminary evidence could suggest that WFDC2 is overexpressed in CF lung tissue as a result of chronic infections, and that expression is induced by exposure to bacteria, although there are many other factors that could be involved, including disease severity.

When CF bronchial epithelial (CFBE) cells cultured *in vitro* were exposed to LPS, no change in WFDC2 secretion was induced (Nagy et al., 2016). This data supports the evidence discussed in Section 1.3.4 which describes a lack of WFDC2 induction following exposure of TBE cells to inflammatory stimuli (Bingle et al., 2006).

Further investigations will be necessary to determine the role that WFDC2 plays in CF and to understand why exacerbated WFDC2 levels correlate with increasing disease severity. It is possible that the upregulation results from an alteration in cellular phenotype in the inflamed, fibrotic pulmonary tissue.

1.5.2. WFDC2 and renal fibrosis

A significant upregulation in WFDC2 expression has been observed in patients with renal fibrosis or renal dysfunction compared to healthy controls and higher serum WFDC2 levels correlate with advanced kidney disease (Escudero et al., 2011; LeBleu et al., 2013; Nagy et al., 2012; Wan et al., 2016; Yuan and Li, 2017). Thus, WFDC2 has been proposed to be a potential biomarker of renal fibrosis.

Mouse models of renal fibrosis have been used to analyse Wfdc2 expression in murine fibrotic kidneys compared to normal kidneys using RT-qPCR and Western blotting (Chen et al., 2017; LeBleu et al., 2013). The studies using mouse models report an upregulation of Wfdc2 in fibrotic renal tissue. LeBleu et al (2013) specifically shows the abundant expression of Wfdc2 in murine fibrosis-associated fibroblasts compared to normal fibroblasts, although it should be noted that recombinant human WFDC2 was used as a positive control for Western blot analysis alongside murine samples, with the results showing both human and mouse orthologues to be the same size (approximately 20 kDa) which is not the case. Similarly, Chen et al (2017) used Western blotting to confirm the expression of murine Wfdc2 in fibrotic kidneys; according to the manufacturer the antibody detects human WFDC2 as a 22 kDa band, which is identical to that reported for murine Wfdc2 in this study. The reliability of the Western blotting data described is, therefore, not confirmed.

If WFDC2 is indeed capable of protease inhibition, as discussed in Section 1.3.5, this may provide sound reasoning as to why the protein is abundant in fibrotic tissues. Fibrosis is a self-perpetuating process, and secretion of protease inhibitors can prevent the degradation of collagen, which resultantly inhibits the turnover of collagen within the tissue. Accumulation of extracellular matrix (ECM) components and immune cells leads to the formation of scar-tissue, which limits kidney function and leads to kidney failure (LeBleu et al., 2013). Interestingly LeBleu et al (2013) use cell staining, protease assays and zymography to show that inhibition of WFDC2 with neutralising antibodies leads to a reduction in fibrosis severity in murine fibrotic kidneys. The authors suggest WFDC2 exacerbates fibrosis by inhibiting proteases that are required for the efficient turnover of collagen, resultantly causing type I collagen to accumulate (LeBleu et al., 2013).

1.6.   Main hypotheses and aims

Hypotheses:

1. WFDC2 is a host defence protein that is capable of antibacterial activity against bacteria endogenous to the oral cavity and lung.

2. Wfdc2‑knockout mice exhibit a similar phenotype to Slpi-knockout animals, whereby the mice appear phenotypically normal but are more susceptible to bacterial infection.

3. Silencing of WFDC2 expression causes significant changes to cancer cell behaviour, including increased cell proliferation and invasion.

The work described in this thesis aims to address the shortfalls and inconsistencies in the literature to elucidate the function of WFDC2 in both healthy and diseased tissues. Recombinant WFDC2 will be used to identify whether the protein is capable of antibacterial or anti-protease activity, as is frequently hypothesised. To identify whether WFDC2 expression exhibits tumorigenic activity, CRISPR gene edited cell lines will be used to analyse the effect of WFDC2 silencing on cancer cell behaviour. Finally, the phenotype of Wfdc2-knockout mice will be analysed, with a view to identifying the role of Wfdc2 in development and physiology. Utilisation of these molecular and genetic tools will provide insights into the function of WFDC2 in humans and mice.

Chapter 2

Materials and methods

2.1. Cell culture

Cancer cell lines and primary human sublingual gland cells were routinely cultured in T75 flasks. The culture media was changed 3 times a week and cells were passaged once a week. All cells were incubated at 37°C with 5% CO_2 in a humidified chamber. The derivation and source of the cell lines used in this study are summarised in Table 2.1.

Cell line name:	Tissue origin:	Provided by:
H357	Established from human squamous cell carcinoma of the tongue	Dr. Simon Whawell (Academic Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, UK)
CAL27	Established from human squamous cell carcinoma of the tongue	Dr. Craig Murdoch (Academic Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, UK)
FaDu	Established from human squamous cell carcinoma of the hypopharynx	Dr. Craig Murdoch (Academic Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, UK)
D20	Established from human dysplastic cells of the lateral tongue	Dr. Keith Hunter (Academic Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, UK)
NCI-H292	Established from human mucoepidermoid carcinoma of the lung	Prof. John Carmichael (University of Nottingham, UK)
NCI-H647	Established from human adenosquamous carcinoma of the lung	Prof. John Carmichael (University of Nottingham, UK)
NCI-H226	Established from human squamous cell carcinoma of the lung	Prof. John Carmichael (University of Nottingham, UK)
NCI-H358	Established from human adenocarcinoma of the lung	Prof. John Carmichael (University of Nottingham, UK)
NCI-H441	Established from human papillary adenocarcinoma of the lung	Prof. John Carmichael (University of Nottingham, UK)
NCI-H322	Established from human bronchioalveolar carcinoma of the lung	Prof. John Carmichael (University of Nottingham, UK)
HEK293	Established from human embryonic kidney	Prof. James Stewart (the Institute of Infection and Global Health, University of Liverpool, UK)
HuSL	Established from explanted tissues of human sublingual gland	Zulaiha Rahman (Academic Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, UK). Ethics code STH 17021.

Table 2.1. Cell lines.

2.1.1. Culture media

H357, FaDu, NCI-H292, NCI-H647 and HEK393-T cell lines were cultured in low glucose DMEM (Sigma-Aldrich) supplemented with 10% (v/v) foetal bovine serum (FBS; Gibco), 2 mM L-glutamine (Sigma-Aldrich), 100 µg/ml of Penicillin and 100 U/ml of Streptomycin (Sigma-Aldrich).

The CAL27 cell line was cultured in high glucose DMEM (4500 mg/L glucose) with 10% (v/v) FBS, 2 mM L-glutamine, and 100 µg/ml of Penicillin and 100 U/ml of Streptomycin.

NCI-H322, NCI-H441, NCI-H358 and NCI-H226 cell lines were grown in RPMI-1640 (Sigma-Aldrich) with 10% (v/v) FBS, 2 mM L-glutamine, and 100 µg/ml of Penicillin and 100 U/ml of Streptomycin.

The D20 cell line and HuSL primary cells were cultured in KGM media. KGM media consists of DMEM supplemented with 23% Ham's F12 (Sigma-Aldrich), 10% (v/v) FBS, 100 µg/ml Penicillin and 100 U/ml Streptomycin, 2 mM L-glutamine, 180 µM adenine (Sigma-Aldrich), 0.5 µg/ml of hydrocortisone (Sigma-Aldrich), 100 pM cholera toxin (Sigma-Aldrich) and 10 ng/ml of epidermal growth factor (EGF; Sigma-Aldrich).

2.1.2. Collection of conditioned media

Cultured cells were seeded at a density of 1 x 10^6 cells per T75 flask and grown for 3-5 days to reach ~70% confluency. Cells were then washed three times in PBS (Sigma-Aldrich) and incubated with 6 ml serum-free media with 2 mM L-glutamine for 24 hours. The media was then collected and stored at -20°C for use within 1 month. Total protein content was measured via BCA assay (see Section 2.3.2) and protein concentrations were standardised prior to analysis by Western blot or ELISA.

2.1.3. Harvesting cells

Cultured cells were washed twice in PBS and harvested using trypsin/EDTA (Sigma-Aldrich). Trypsin was neutralised with media containing 10% (v/v) FBS. Cells were collected by

centrifugation at 1,000 x g for 5 minutes. For crude Western blot analysis, pelleted cells were lysed directly in 2X SDS loading buffer and stored at -20°C. For RNA extraction the cell pellet was flash frozen at -80°C for up to 1 month.

2.2. DNA analysis

2.2.1. RNA extraction

Total RNA was extracted from cultured cells using TRI reagent according to the manufacturer's protocol. The resultant RNA pellet was resuspended in 50 µl of nuclease-free H_2O and the yield was determined using the NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific). RNA was stored at -80°C.

Total RNA to be used for downstream RT-qPCR was extracted from cultured cells using the Monarch Total RNA Miniprep Kit (NEB) as per the manufacturer's instructions. Optional DNase I treatment steps for removal of residual genomic DNA were carried out.

2.2.2. Reverse transcription

Complementary DNA (cDNA) was synthesised from 1 µg of total RNA template using the High-Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific) as per the manufacturer's protocol. cDNA was stored at -20°C.

2.2.3. Polymerase chain reaction

Polymerase chain reaction (PCR) was performed using 1 µl of template cDNA, 100 ng of forward and reverse oligonucleotide primers and 2X DreamTaq Green PCR Master Mix (ThermoFisher Scientific) in a total reaction mixture of 20 µl. The primer pair sequences are summarised in Table 2.2. Cycling conditions were as follows: 5 minutes at 95°C, 35 cycles of 95°C for 1 minute, 60°C for 1 minute, 72°C for 1 minute, followed by 7 minutes at 72°C. Annealing temperatures at 55°C, 60°C and 65°C were originally tested; 60°C was found to be optimal and used for all subsequent PCRs. Resultant PCR products were analysed by agarose gel electrophoresis.

Gene target		Sequence (5'-3')	Product size (bp)
Human WFDC2 (V0)	F	CCGACAACCTCAAGTGCTG	105
	R	CGAGCTGGGGAAAGTTAAT	
Human WFDC2 (V1)	F	AATGGCCAACTGGCTGAG	132
	R	TTTGAGAGAGTCCCCAGCTC	
Human WFDC2 (V2)	F	CCATGCCTGCTTGTCGCC	97
	R	CAGGAACCCTCCTTATCTGA	
Human WFDC2 (V3)	F	GCCATGCTGCAGGTACAAGT	75
	R	ATCTGGGTAGAAAAAGGAGTAAGG	
Human WFDC2 (V4)	F	CCCAATGCACTGTTCCACT	85
	R	AGTCCCAAGTGGGCCTTC	
Human WFDC2 (full-length)* *can amplify multiple splice variants	F	CGGCTTCACCCTAGTCTCAG	V0: 450 V1: 579 V2: 306
	R	AAAGGGAGAAGCTGTGGTCA	
Human OAZ1	F	CCAACGACAAGACGAGGGATT	164
	R	AGCGAACTCCAGGAGAACTG	
Murine Wfdc2 (FL)	F	AGGTGGACAGCCAGTGTTCT	196
	R	GACCAGGAAGAAATGCAA	

Table 2.2. Primer sequences utilised for RT-PCR. Primer sequences and the expected PCR product size (bp) are shown.

2.2.4. DNA agarose gel electrophoresis

DNA samples and PCR products were visualised on a 1 or 2% agarose gel (percentage dependent on expected product size) containing 0.2 µg/ml of ethidium bromide. Gels were run in 1X TAE buffer at 80 V and visualised using the InGenius3 gel documentation system (Syngene). See Appendix for further details.

2.2.5. Quantitative polymerase chain reaction

Quantitative polymerase chain reaction (qPCR) was performed using a 20X TaqMan probe mix for detection of human WFDC2 variant 1 exons 2-3 (ThermoFisher, Hs00899484_m1) and 2X qPCRBIO Probe Blue Mix Lo-ROX master mix (PCR Biosystems). A 20X β-2 macroglobin (B2M) control probe mix (Life Technologies) was used as the endogenous control. Each reaction was performed in triplicate using 0.5 µl of template cDNA in a total reaction mixture

of 10 µl. Thermal cycling was carried out using a 7900 HT Fast Real-Time PCR machine. Cycling conditions were as follows: 40 cycles of 95°C for 15 seconds, 60°C for 1 minute, 95°C for 15 seconds. Data was shown as fold-change in gene expression compared to the endogenous control or values were normalised to the untreated controls.

Quantification of expression was calculated using the delta delta C_T (2-$\Delta\Delta C_T$) method (Livak and Schmittgen, 2001). The cycle threshold (C_T) is the cycle number that causes the fluorescent signal to reach a fixed threshold. The ΔC_T value represents the difference in C_T values between WFDC2 target gene and B2M endogenous control. The calculated 2-$\Delta\Delta C_T$ values represent the relative fold change in gene expression between samples.

2.3. Protein analysis

2.3.1. Human and mouse samples

Whole human saliva samples were collected from healthy, non-smoking volunteers from the School of Clinical Dentistry (University of Sheffield). Parotid gland secretions were collected from one healthy, non-smoking volunteer using a Carlson-Crittenden cup. Submandibular/sublingual secretions were collected from the same volunteer using a custom-made device. Samples were centrifuged at 13,000 rpm to remove cells and debris prior to analysis. Collection of saliva samples had full ethical approval (ethics number: 003166).

Human lung epithelial secretions were kindly donated by Dr Michael Campos (Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, University of Miami, USA). Secretions were collected from healthy bronchial cells (derived from bronchial brushings) grown at an air-liquid interface (ALI). Samples were centrifuged at 13,000 rpm to remove cells and debris prior to analysis.

Mouse tracheal epithelial secretions were kindly donated by Dr Khondoker Akram (Department of Infection, Immunity and Cardiovascular Disease, Medical School, University of Sheffield, UK). Murine tracheal epithelial cells were isolated from mice, cultured at an ALI and apical secretions were collected. Mouse nasal epithelial secretions were kindly donated

by Priyanka Anujan (Department of Infection, Immunity and Cardiovascular Disease, Medical School, University of Sheffield, UK). Cells were isolated from mice and cultured (as outlined in Akram et al., 2018 and Mulay et al., 2016). Samples were centrifuged at 13,000 rpm to remove cells and debris prior to analysis.

Antibodies

The antibodies used for Western blotting are summarised in Table 2.3.

Primary antibody	Manufacturer and catalogue number	Description	Working concentration or dilution	Secondary antibody	Immunogen
Anti-WFDC2	R&D Systems AF6274	Sheep polyclonal to human WFDC2	1.75 µg/ml	Anti-sheep-HRP	NS0-derived recombinant human WFDC2 Thr28-Phe124
Anti-WFDC2	Abcam ab200828	Rabbit monoclonal to human WFDC2	1.5 µg/ml	Anti-rabbit-HRP	Recombinant fragment within human WFDC2 Met1–Phe124
Anti-Wfdc2	Antiserum kindly donated by Ronny Drapkin[1]	Rabbit polyclonal to mouse Wfdc2	1:1000	Anti-rabbit-HRP	Not disclosed
Anti-FLAG	Sigma F3165	Mouse monoclonal to FLAG-tagged proteins	1.0 µg/ml	Anti-mouse-HRP	FLAG peptide sequence DYKDDDDK
Anti-MUC7	Novus NBP2-50391	Mouse monoclonal to human MUC7	1.5 µg/ml	Anti-mouse-HRP	KLH conjugated peptide EGRERDHELRHRRHHHQ in the N-terminus of MUC7

Table 2.3. Antibodies for Western blotting. Catalogue number, working concentration, secondary antibody required and immunogen are summarised. [1]Ronny Drapkin (Perelman School of Medicine, University of Pennsylvania, USA).

2.3.2. BCA assay

Total protein concentration in samples was determined using the Pierce BCA Protein Assay Kit (ThermoFisher Scientific) according to the manufacturer's instructions. Bovine serum albumin (BSA) (ThermoFisher Scientific) was used to generate a standard curve and the absorbance of standards and samples was measured at 570 nm using the POLARstar Galaxy Spectrophotometer (BMG LabTech). Only phenol red-free conditioned media was analysed via BCA assay as phenol red is known to interfere with the assay.

2.3.3. Western blotting

For general details of Western blotting reagents see the Appendix. Samples to be analysed by Western blotting were lysed in 2X SDS loading buffer or 5X loading buffer (National Diagnostics) and heated at 95°C for 5 minutes. As standard, proteins were separated on a 12% acrylamide gel against an EZ-Run Prestained Rec protein ladder (ThermoFisher Scientific). In some instances, alternative gel percentages were used depending on the molecular weight of the sample being resolved. Proteins were transferred to a nitrocellulose membrane (GE Healthcare) sandwiched between 3 sheets of filter paper soaked in semi-dry transfer buffer. Transfer conditions were as follows: 1.0 A and 25 V for 30 minutes, using a Western blotting transfer system (BioRad, Trans-Blot Turbo). The presence of transferred protein bands was detected using Ponceau stain (Sigma Aldrich). Membranes were blocked for 1 hour using 5% skimmed milk in Tris buffered saline (1X TBS) on a rolling platform at room temperature. Primary antibodies were diluted in 5% skimmed milk in 1X TBS plus 0.1% Tween 20 (1X TBS-Tween). Diluted antibodies were incubated with the membrane overnight at 4°C on a rolling platform. The membrane was washed with 1X TBS-Tween 3 times for 10 minutes and then incubated with diluted secondary antibody conjugated with horseradish peroxidase (HRP) for 1 hour at room temperature on a rolling platform. Membranes were washed again with 1X TBS-Tween 3 times for 10 minutes. Protein was visualised using enhanced chemiluminescence (ECL) Western blotting substrate (ThermoFisher Scientific). Signal was detected by exposure to X-ray film (ThermoFisher Scientific) and subsequent passage through a Compact X4 automatic X-ray film processor (Xograph) or alternatively using a Li-Cor C-Digit Western Blot Scanner and Image Studio Software.

2.3.4. Densitometry

Western blots developed using the Li-Cor C-Digit Western Blot Scanner were analysed by densitometry using Image Studio Software. Bands were manually drawn around and the signal within each area of focus was measured.

2.3.5. SDS-PAGE analysis

Twelve percent acrylamide gels were used for SDS PAGE analysis. To visualise protein bands, gels were incubated in InstantBlue protein stain (Sigma-Aldrich) overnight at room temperature with agitation. The following day, gels were washed in H_2O and then imaged using the InGenius3 gel documentation system (Syngene).

2.3.6. Native PAGE analysis

Ten percent acrylamide discontinuous native PAGE gels were used for Western blotting of native protein samples. The PAGE gel, running buffer and sample buffer lacked SDS to prevent protein denaturation. The standard Western blotting protocol (see 2.3.3) was carried out using these native reagents. See the Appendix for more details.

2.3.7. Enzyme-linked immunosorbent assay

Human WFDC2 protein concentration was quantified using the Human HE4/WFDC2 Quantikine ELISA Kit (R&D Systems), a sandwich ELISA that uses a monoclonal and polyclonal antibody, both raised against recombinant WFDC2. Saliva, secretions from primary lung cell culture and conditioned media from cell lines were prepared by centrifugation at 13,000 x g for 5 minutes to remove cells/debris and diluted 1,000-fold, 300-fold and 250-fold, respectively, in Calibrator Diluent RD5-26 (1X) provided. The ELISA was carried out following the manufacturer's protocol. WFDC2 standards (made using provided recombinant NS0-derived WFDC2) were used to generate a standard curve in GraphPad Prism 7. The WFDC2 concentration was extrapolated from the standard curve and resultant values were multiplied by the dilution factor to determine the concentration of WFDC2 in the original sample.

2.4. Synthesis of recombinant WFDC2

2.4.1. Primer design and polymerase chain reaction

Primers were designed to amplify full length human WFDC2 V0 and mouse Wfdc2 V0. A NotI enzyme restriction site and a Kozak consensus sequence were incorporated into the forward

primer. A BglII restriction site and a FLAG-tag sequence (DYKDDDDK) were incorporated into the reverse primer. Primer sequences are summarised in Table 2.4. Primers amplified the desired cloning insert from the template cDNA via PCR as described in Section 2.2.3. cDNA was derived from RT-PCR and provided by Prof. Colin Bingle (the Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, UK). PCR conditions were as follows: 4 minutes at 95°C, 35 cycles of 94°C for 1 minute, 60°C for 1 minute, 72°C for 1 minute, followed by 7 minutes at 72°C. PCR products were visualised on a 1% agarose gel.

Gene transcript		Sequence (5'-3')	Product size (bp)
Human WFDC2	F	ATGCGGCCGCCGCCGCCACCATGCCTGCTTGTCGCCTAGGCCCGCT	427
	R	GCAGATCTTTACTTGTCATCGTCGTCCTTGTAGTCGAAATTGGGAGTGACACAGGA	
Mouse Wfdc2	F	ATGCGGCCGCCGCCGCCACCATGCCTGCCTGTCGCCTCTGCTT	577
	R	GCAGATCTTTACTTGTCATCGTCGTCCTTGTAGTCGAATTTGGGTGTGGTGCAGGCCAT	

Table 2.4. Primer sequences utilised for gene cloning. Primer sequences and the expected PCR product size (bp) are shown.

2.4.2. TOPO TA subcloning

TOPO TA cloning (ThermoFisher Scientific) was performed to sub-clone amplified WFDC2 PCR products directly into a pCR®II-TOPO® 4.0 kb vector (ThermoFisher Scientific). Cloning was carried out according to the manufacturer's protocol. Plasmids were transformed into NEB 5-alpha competent *E. coli* cells (NEB) and successful transformants were selected on LB-agar plates containing ampicillin (50 µg/ml). Transformed colonies were picked, inoculated into 2 ml of LB-broth containing ampicillin (50 µg/ml) and grown overnight at 37°C and 220 rpm. Plasmid DNA was isolated from the bacterial culture using the Isolate II Plasmid Mini Kit (Bioline) and DNA yield was measured using a NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific).

2.4.3. Isolation of insert DNA

To isolate the WFDC2 inserts, TOPO vector DNA was digested with BglII and NotI restriction enzymes (NEB) in 10X NEBuffer 3.1 (NEB). Restriction digestion reactions were incubated overnight at 37°C. Digested products were separated on a 1% agarose gel to confirm expected fragment sizes of the human WFDC2 (416 bp) and mouse Wfdc2 (566 bp) inserts. The WFDC2 inserts were extracted from the gel using a GeneJET Gel Extraction Kit (ThermoFisher Scientific) and kept at 4°C for short-term storage.

2.4.4. Vector preparation

The target vector for WFDC2 cloning was the pVR1255 plasmid (from Prof. James Stewart, the Institute of Infection and Global Health, University of Liverpool), mapped in Figure 2.1.
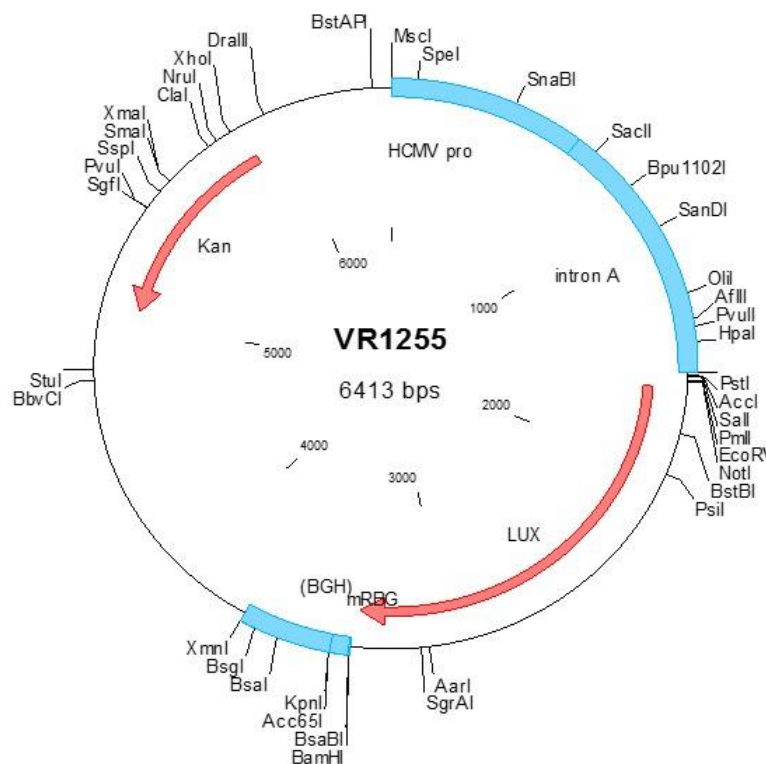
Fig 2.1. The pVR1255 vector. A plasmid map of the target vector, pVR1255. Labelled cassettes include the promoter (HCMV pro), the kanamycin resistance gene (Kan) and the luciferase gene (LUX) which was excised and replaced with the gene of interest (WFDC2). Restriction sites are also shown. The plasmid vector is 6413 bp in size.

pVR1255 was prepared by digestion with NotI and BglII to excise the luciferase cassette and produce compatible sticky ends for the insertion of the WFDC2 insert. The double-cut vector (~4.7 kb) with luciferase (~1.6 kb) excised was resolved on a 1% agarose gel and isolated using a GeneJET Gel Extraction Kit (ThermoFisher Scientific). DNA yield was determined using a NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific). Vector cut with BglII only (single-cut) was isolated in the same way and used as a ligation control.

2.4.5. Ligation

A sample of WFDC2 insert isolated by gel extraction (described in Section 2.4.3) was resolved on a 1% agarose gel and quantified by comparison against PCR Ranger 100 bp DNA ladder (Norgen) reference bands to estimate the DNA concentration. WFDC2 insert was ligated into the double-cut VR1255 vector using T4 DNA Ligase (Promega) and Ligase 10X Buffer (Promega) following the manufacturer's protocol in a total reaction volume of 10 µl.

2.4.6. Transformation

Five microlitres of ligated pVR1255-WFDC2 vector was mixed with 10 µl of NEB 5-alpha competent *E. coli* cells (NEB) and the mixture was left on ice for 30 minutes. Cells were then heat-shocked for 45 seconds at 42°C and immediately placed back on ice for 5 minutes. One hundred microlitres of SOC medium (NEB) was added to the mixture, followed by incubation for 1 hour at 37°C. Cells were plated onto LB-agar plates containing kanamycin (30 µg/ml) and incubated overnight at 37°C in order to select for successful transformants. Single colonies of transformants were picked and inoculated into 2 ml LB-broth with kanamycin (30 µg/ml) and incubated at 37°C and 220 rpm overnight.

The overnight broth culture was then used for plasmid isolation. Glycerol stocks were also made from 0.5 ml of bacterial culture with 0.5 ml of 50% glycerol and kept at -80°C.

2.4.7. Plasmid isolation 'miniprep'

Vector DNA was isolated from 1 ml of transformed cells using the Isolate II Plasmid Mini Kit (Bioline). DNA yield was determined using a NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific) and DNA stored at -20°C.

2.4.8. Plasmid isolation 'maxiprep'

For higher DNA yields, 100 ml of LB-broth with kanamycin (30 µg/ml) was inoculated with 0.5 ml of transformant culture from stocks kept at 4°C. The bacterial culture was incubated at 37°C with shaking (220 rpm) overnight. Plasmid DNA was harvested using a Plasmid Maxi Kit (Qiagen). DNA yield was determined using a NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific) and then aliquoted and stored at -20°C.

2.4.9. Confirmation of target insertion

Plasmid DNA was digested using restriction enzymes and expected fragment sizes were confirmed by agarose gel electrophoresis to show that the insert was successfully incorporated into the vector.

2.4.10. Sequencing of clones

The nucleotide sequence of the WFDC2 insert in the VR1255 vector was further validated by DNA sequencing (Research Core Facility, the Medical School, University of Sheffield, UK).

2.4.11. FuGENE transfection

HEK293 cells were seeded into a 24-well plate at a density of 3×10^5 cells per well and incubated overnight at 37°C and 5% CO_2. Prior to transfection, cells were washed and media was changed to Opti-MEM reduced serum media (Gibco). Cells were co-transfected with pVR1255 with human or mouse insert (500 ng) and pEGFP-N1 (100 ng; ClonTech Laboratories) using 3 µl of FuGENE HD transfection reagent (Promega) in 200 µl of Opti-MEM media as per the manufacturer's recommendation and incubated overnight at 37°C and 5%

CO_2. Media was changed the following day. After a further 24 hours, cells were lysed in 2X SDS lysis buffer directly in the well and stored at -20°C for downstream analysis.

2.4.12. Calcium phosphate transfection

HEK293 cells were seeded into a 24-well plate at a density of 3×10^5 cells per well and incubated overnight at 37°C and 5% CO_2. Two hours prior to transfection, media was changed with fresh DMEM supplemented with 10% (v/v) FBS. Cells were co-transfected with pVR1255-WFDC2 (500 ng) and pEGFP-N1 (100 ng). DNA was mixed with 3.75 µl of $CaCl_2$ (2.5 M) and made up to 30 µl with H_2O. The DNA-$CaCl_2$ mixture was added to 30 µl of 2X HEPES buffered saline (HBS; see Appendix) and mixed by pipetting. The reaction mixture was incubated for 15 minutes at room temperature and then added drop-wise to the cells. Transfection reactions were incubated overnight at 37°C and 5% CO_2. Media was changed the following day. After a further 24 hours, cells were lysed in 2X SDS lysis buffer directly in the well and stored at -20°C for downstream analysis.

For transfection of HEK293 cells in T75 flasks, cells were seeded at a density of 4.5×10^6 cells per flask and incubated at 37°C and 5% CO_2 for 48 hours. Media was changed two hours prior to transfection with fresh DMEM supplemented with 10% (v/v) FBS. Cells were co-transfected with pVR1255-WFDC2 (15 µg) and pEGFP-N1 (3.75 µg). DNA was mixed with 140.6 µl of $CaCl_2$ (2.5 M) and made up to 1.125 ml with H_2O. The DNA-$CaCl_2$ mixture was added to 1.125 ml of 2X HEPES buffered saline (HBS), mixed by pipetting, and incubated for 15 minutes at room temperature. The mixture was then added drop-wise to the cells. Transfection reactions were incubated overnight at 37°C and 5% CO_2. Media was changed for fresh DMEM supplemented with 10% (v/v) FBS the following day. After 24 hours, cells were washed 3 times with PBS and then incubated with serum-free DMEM. Media was collected and replaced with fresh serum-free DMEM after 24 hours. This was repeated for a further 3 days. All collected conditioned media was stored at -20°C. For analysis by Western blotting, the serum-free conditioned media was concentrated 6-fold using Vivaspin 6 centrifugal concentrators with a

10 kDa molecular weight cut-off point (Sartorius). Serum-free conditioned media was not concentrated when used for affinity gel purification.

2.4.13. Anti-FLAG M2 affinity gel purification

FLAG-tagged recombinant WFDC2 proteins were purified from the serum-free conditioned media of transiently transfected HEK293 cells using Anti-FLAG M2 affinity gel (Sigma) as per the manufacturer's instructions. Conditioned media was prepared for purification by adjusting the salt concentration to 0.15 M with NaCl. Debris was removed by centrifugation at 15,000 x g for 15 minutes followed by filtration through a 0.22 µm sterile filter (Sartorius). The affinity gel was washed in 0.1 M glycine-HCl (pH 3.5) and then TBS prior to use. Typically, 100 µl of affinity gel was added to 6 ml of conditioned media and kept at 4°C for 2 hours. Bound protein was eluted from the affinity gel using 150 µl of FLAG peptide (150 ng/µl) (Sigma). Eluted protein was aliquoted and stored at -20°C. The presence of WFDC2 protein was confirmed by Western blotting and protein yield was quantified by BCA. A standard 150 µl elution would produce ~25 µg of purified protein. Aliquots of affinity gel were reused for purification of the same protein construct up to 3 times before disposal, by washing in 0.1 M glycine-HCl (pH 3.5) and storage in 50% glycerol with TBS containing 0.02% sodium azide.

2.4.14. Immunofluorescence microscopy

NCI-H292 cells were seeded into a 12-well plate at a density of 1.6×10^6 cells per well and were transfected the following day with pVR1255-WFDC2 or empty vector using the FuGENE transfection method (Section 2.4.11). Cells were transfected with pVR1255 (1 µg) using 6 µl of FuGENE HD transfection reagent (Promega) in 400 µl of Opti-MEM media. Following transfection, the cells were incubated overnight at 37°C and 5% CO_2. Cells were washed in PBS and then incubated in DMEM supplemented with 10% (v/v) FBS for 6 hours. Transfected cells were trypsinised and seeded into Nunc Lab Tek II chamber slides (ThermoFisher Scientific) at a density of 3.5×10^4 and incubated overnight at 37°C and 5% CO_2.

The following day, media was aspirated from each chamber of the slide and cells were fixed by immersion in ice-cold 100% methanol (ThermoFisher Scientific) for 10 minutes. Cells were rinsed in PBS and incubated for 5 minutes in 0.5% Triton-X-100 in PBS (Sigma-Aldrich) to permeabilise the cells. Cells were then washed 3 times in PBS.

For the blocking step, cells were incubated in 10% serum (Sigma-Aldrich) in PBS for 1 hour at room temperature with gentle rocking. The blocking buffer was aspirated and cells were washed 3 times in PBS. Primary antibody was diluted in 10% serum and cells were incubated with diluted antibody for 2 hours in a humidified chamber at room temperature with gentle rocking. See Table 2.5 for details of the appropriate blocking serum and antibody concentrations used. The primary antibody was aspirated and cells were washed 3 times in PBS. Alexa Fluor-488 conjugated secondary antibodies were diluted in PBS and cells were incubated for 1 hour at room temperature with gentle rocking and were protected from light. The secondary antibody was aspirated and cells were washed 3 times in PBS. The media chamber frame was then removed and the slide was inverted onto a coverslip with VECTASHIELD Mounting Media with DAPI (Vector Laboratories). The edges of the coverslip were sealed using transparent nail polish and left to dry in the dark.

Primary (1°) antibody	1° antibody concentration (µg/ml)	Secondary (2°) antibody	2° antibody concentration (µg/ml)	Appropriate serum for blocking step
Anti-WFDC2 polyclonal (R&D systems)	3.5	Rabbit anti-goat IgG (H+L) Alexa Fluor 488 conjugate (Invitrogen)	2.0	10% rabbit serum (Sigma-Aldrich)
Anti-FLAG monoclonal (Sigma-Aldrich)	2.0	Goat anti-mouse IgG (H+L) Alexa Fluor 488 conjugate (Life Technologies)	2.0	10% goat serum (Sigma-Aldrich)

Table 2.5. Antibodies for immunofluorescence microscopy. Working concentration, secondary antibody and blocking serum required are summarised.

When staining cells transfected with murine Wfdc2, a Mouse-On-Mouse (M.O.M) Immunodetection Kit (Vector Laboratories) was required to reduce background staining as anti-FLAG is raised in mouse. M.O.M Mouse IgG blocking reagent (90 µl of M.O.M Mouse IgG Blocking Reagent in 2.5 ml of PBS) was used to block for 1 hour at room temperature and the

cells were then washed with PBS and incubated with diluted M.O.M Protein (600 µl of Protein concentrate in 7.5 ml of PBS) for 2 minutes. Anti-FLAG primary antibody was also diluted in the M.O.M Protein concentrate. The rest of the staining protocol remained as standard (described above).

Cells were imaged using an Axiovert 200M fluorescent microscope (Zeiss) and AxioCam camera (Zeiss). Each field was imaged using a DAPI and GFP filter, in addition to an image taken in white light.

The images captured were in black and white so to visualise the DAPI and GFP staining together within a single field, the two images were merged and colour was imitated using Fiji by ImageJ software (Schneider et al., 2012).

2.5. Glycosylation analysis

2.5.1. Site-directed mutagenesis

Site directed mutagenesis (SDM) was used to remove the predicted N-glycosylation site (Asn-44) from the human WFDC2 gene. Mutagenic primers were designed with two nucleotide substitutions to produce an amino acid change [Asn (AAC) to Gln (CAA)]. Primer sequences are summarised in Table 2.6.

Gene target	Sequence (5'-3')	
SDM targeting	F primer	CTCCAGGCTGACCAGCAATGCACGCAAGAGTGC
	R primer	GCACTCTTGCGTGCATTGCTGGTCAGCCTGGAG

Table 2.6. Primer sequences utilised for SDM.

SDM was carried out using the QuikChange II Site-Directed Mutagenesis Kit (Agilent) and the manufacturer's protocol was followed accordingly. pVR1255 containing the human WFDC2 insert was used as template DNA for the PCR. Plasmid DNA from successful SDM transformants was isolated by miniprep (see Section 2.4.7) and sequenced to confirm the

presence of the desired mutation (Research Core Facility, the Medical School, University of Sheffield, UK).

2.5.2. PNGase F enzyme treatment

PNGase F enzyme (NEB) was used to cleave N-linked glycans from target proteins. Reactions were carried out as per the manufacturer's recommendation, using 1 µl of PNGase F enzyme and 1-20 µg of glycoprotein substrate in a total reaction volume of 20 µl. PNGase F-treated products were lysed in 2X SDS lysis buffer and analysed by Western blotting.

2.5.3. Sialidase enzyme treatment

Recombinant sialidase enzyme with the sequence derived from *Tannerella forsythia* was kindly donated by Marianne Satur and Dr Andrew Frey (School of Clinical Dentistry, University of Sheffield; method used to produce recombinant sialidase described in Frey et al., 2018). Sialidase was used to cleave sialic acid residues from WFDC2. Twenty microlitres of each purified WFDC2 orthologue were incubated with sialidase (500 nM) diluted in PBS for 1 hour at 37°C. Samples were then lysed in 2X SDS lysis buffer and analysed by Western blotting to visualise molecular weight shifts due to cleavage of sialic acid.

2.6. Protease inhibition assays

2.6.1. Enzyme optimisation for protease inhibition assay

Trypsin from porcine pancreas (Sigma) was solubilised in 1 mM HCl solution and then its activity at 0, 260, 520, 780, 1040 and 1300 nM was assessed by incubation with 830 µM of Nα-Benzoyl-DL-arginine 4-nitroanilide hydrochloride (BAPNA; Sigma). BAPNA was solubilised in DMSO and then diluted in triethanolamine (TEA) buffer (125 mM TEA with 13 mM $CaCl_2$, pH 7.8).

In a 96-well plate, 25 µl of trypsin (0-1300 nM) and 75 µl of BAPNA (830 µM) were incubated together in 50 µl of TEA buffer for 15 minutes at 37°C. Fifty microlitres of 30% acetic acid was added to stop the reaction and then the absorbance was read at 405 nm using a POLARstar

Galaxy Spectrophotometer (BMG LabTech). The mean absorbance value for BAPNA alone was subtracted from all readings to account for background.

Elastase from porcine pancreas (Sigma) was solubilised in HEPES buffer (100 mM HEPES, 500 mM NaCl, 10% DMSO) and its activity at 0, 65, 130 and 195 nM was assessed by incubation with 465 µM of N-Succinyl-Ala-Ala-Ala-p-nitroanilide (SAPNA; Sigma). SAPNA was solubilised and diluted in HEPES buffer. Enzyme activity was measured using the same protocol as for trypsin, but HEPES buffer was used in place of TEA buffer.

2.6.2. Serine protease inhibition assay

Purified recombinant human and mouse WFDC2, recombinant human SLPI (R&D Systems) and solubilised cOmplete™, Mini, EDTA-free Protease Inhibitor (PI) tablets (Sigma) were analysed for their capacity to inhibit trypsin and elastase activity. WFDC2 proteins and SLPI were diluted in TBS and used at a final concentration of 0, 100, 200, 500 and 1000 nM. Each PI tablet was solubilised in 1.16 ml of TBS to give a 6X solution and sequentially increasing volumes were assayed for enzyme inhibition (0, 1.25, 2.5, 6.25 and 12.5 µl).

In a 96-well plate, 25 µl of inhibitor was pre-incubated with 25 µl of trypsin (1.3 µM) or elastase (130 nM) for 15 minutes at room temperature. Seventy-five microlitres of appropriate substrate was added (830 µM of BAPNA for trypsin, 465 µM SAPNA for elastase) and the reaction made up to 150 µl with buffer (TEA buffer for trypsin, HEPES buffer for elastase). The 96-well plate was incubated at 37°C for 15 minutes and then the reaction was quenched with 30% acetic acid. The absorbance was read at 405 nm using a POLARstar Galaxy Spectrophotometer (BMG LabTech). The mean absorbance value for substrate alone was subtracted from all readings to account for background. Results are expressed as a percentage reduction in enzyme activity whereby enzyme and substrate with no inhibitor represents 100% activity.

2.6.3. Zymography

Samples to be analysed by zymography were incubated at 37°C for 10 minutes prior to resuspension in 5X non-reducing sample buffer. Samples were separated on a 7.5%

acrylamide gel containing gelatin for ~2 hours at 150 V in 1X Tris-glycine running buffer (Bio-rad). Resolved gels were washed twice for 30 minutes in washing buffer to remove SDS from the gel. Gels were rinsed for 10 minutes in incubation buffer at 37°C with agitation. Gels were placed in fresh incubation buffer and left for 24 hours at 37°C to enable the gelatinase reaction to occur. Gels were stained in InstantBlue stain (Sigma-Aldrich) for 1 hour, rinsed in H_2O and then imaged using the InGenius3 gel documentation system (Syngene). Areas of enzyme activity appear as white bands. See the Appendix for details of zymography reagents.

2.7. Bacteria assays

2.7.1. Bacterial strains

Bacterial species	Strain	Provided by:
Streptococcus mutans	Ingbritt	Academic Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, UK
Streptococcus gordonii	Challis	Academic Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, UK
Streptococcus pneumoniae	D39	Department of Infection, Immunity & Cardiovascular Disease, University of Sheffield, UK
Escherichia coli	Clinical isolate from urinary tract	Academic Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, UK

Table 2.7. Bacterial strains.

2.7.2. Bacterial culture

All strains were maintained on Colombia blood agar (BA) plates. Liquid cultures of all strains were grown using Brain-Heart Infusion (BHI) broth. Bacterial assays were performed using bacteria grown overnight in BHI. *E. coli* was incubated at 37°C in aerobic conditions. *S. mutans* and *S. gordonii* were incubated at 37°C and 5% CO_2. *S. pneumoniae* was incubated at 37°C in 80% N_2, 10% H_2, 10% CO_2 using an anaerobic cabinet (miniMACS Anaerobic Workstation, Don Whitley Scientific, UK).

BA plates were made by adding 7.8 g of Colombia blood agar powder (Oxoid) to 200 ml of distilled H_2O. Agar was sterilised by autoclaving at 121°C for 15 minutes, cooled to 55°C, supplemented with 7% oxalated horse blood (ThermoFisher Scientific) and then poured into Petri dishes to set. Plates were stored at 4°C.

BHI broth was made by adding 14.8 g of BHI powder (Oxoid) and 2.5 g of yeast extract (Oxoid) to 400 ml of distilled H_2O and 20 ml aliquots were sterilised by autoclaving.

2.7.3. Colony polymerase chain reaction and 16S sequencing

To confirm the species of bacteria being cultured, colony PCR and 16S sequencing were used. A colony of bacteria was picked from a plate, resuspended in 25 µl nuclease-free H_2O and heated at 95°C for 10 minutes.

PCR was performed using 5 µl of lysed pellet solution alongside final concentrations of 0.5 µM of forward and reverse oligonucleotide primers, 5X Phusion buffer, 200 µM of dNTPs and 0.02 U/µl of Phusion High-Fidelity DNA polymerase. Universal primers sequences for targeting 16S rDNA are summarised in Table 2.8.

Gene target	Sequence (5'-3')		Product size (bp)
16S rDNA	Universal primer 27 F	AGAGTTTGATYMTGGCTCAG	~500 bp
	Universal primer 519 R	GWATTACCGCGGCKGCTG	

Table 2.8. Primer sequences utilised for 16S sequencing. Primer sequences and the expected PCR product size (bp) are shown.

The cycling conditions were as follows: initial denaturation of 98°C for 30 seconds, 30 cycles of 98°C for 10 seconds, 65°C for 30 seconds and 72°C for 30 seconds, final extension of 72°C for 2 minutes. Resultant PCR products were analysed by agarose gel electrophoresis.

Five microlitres of amplified PCR products with 5 µl of forward or reverse 16S primer were analysed by Sanger sequencing (GATC Biotech, Eurofins Genomics). Bacterial species were confirmed by BLAST searching using the NCBI blastn Suite. Glycerol stocks were made from the cultures that were confirmed by 16S sequencing.

2.7.4. Pull down assay

The optical density of an overnight broth culture was adjusted to an OD_{600} of 1.0. One millilitre of culture was used per sample to be analysed. Each 1 ml aliquot of cells was washed 3 times

in sterile PBS by centrifugation (all spins at 13,000 rpm for 3 minutes). The cell pellet was then resuspended in 20 µl of purified protein (diluted 1:1 in PBS) or 20 µl of control FLAG peptide elution solution (150 ng/µl; diluted 1:1 in PBS) or 20 µl of PBS alone and incubated at 37°C for 2 hours. After 2 hours had elapsed the cells were pelleted and the unbound protein was removed and lysed in 2X SDS lysis buffer. The pellet was washed once in PBS and then transferred to a fresh Eppendorf. The bacteria were washed twice more and the pellet was resuspended in 10 µl of PBS and lysed in 20 µl of 2X SDS lysis buffer. The unbound protein and bacterial pellet in 2X SDS lysis buffer were heated at 95°C for 5 minutes and then analysed by Western blotting (see 2.3.3).

2.7.5. Miles-Misra viability assay

Ten microlitres of bacterial culture with an OD_{600} of 0.05 was incubated with 1.0 µM of purified WFDC2 protein diluted in PBS, in a total volume of 20 µl. An equal volume of FLAG peptide elution solution (150 ng/µl) or 10 µl of PBS alone were used as controls. The culture was incubated at 37°C for 3 hours and then plated out using the Miles-Misra technique (10 µl of serial dilutions) to determine the number of colony forming units (CFU) in the bacterial suspension. When plated, *E. coli* and *S. gordonii* plates were maintained at room temperature overnight whereas slower growing *S. mutans* and *S. pneumoniae* were kept at 37°C. Colonies were counted at the lowest dilution factor where distinct colonies were visible.

2.7.6. TECAN growth curve viability assay

Wells of a 96-well plate were loaded with 5 µl of bacterial culture adjusted to an OD_{600} of 1.0 and either 1.0 µM or 3.0 µM of purified WFDC2 protein in a total volume of 200 µl of BHI broth. An equal volume of FLAG peptide elution buffer (150 ng/µl) and PBS alone were used as controls. Bacteria with BHI only and bacteria in BHI with ampicillin (60 µg/ml) were also used as controls. Bacterial culture optical density was measured at 600 nm every 30 minutes for 12 hours using a microplate reader (TECAN).

2.7.7. Overnight bacterial incubation analysis

One millilitre of bacterial culture adjusted to an OD_{600} of 1.0 was pelleted and resuspended in 10 μl of purified WFDC2 and incubated overnight at 37°C. The following day the bacteria were pelleted and WFDC2 protein in the supernatant was collected. The protein was lysed in 2X SDS lysis buffer and analysed by Western blotting to determine whether the proteins were digested by bacteria when incubated overnight.

2.7.8. Agglutination assay optimisation

The optical density of an overnight broth culture was adjusted to an OD_{600} of 1.0. Aliquots of 0.3 ml, 0.5 ml, 0.7 ml and 0.9 ml of each bacterial strain were pelleted, washed in sterile PBS by centrifugation (13,000 rpm for 3 minutes) and then resuspended in 1 ml of PBS. Each 1 ml aliquot of cell suspension was inoculated with 1 μl of fluorescein isothiocyanate (FITC; 0.1 μg/ml; Sigma) and kept in darkness for 15 minutes at 4°C with agitation. Cells were washed 4 times with 1.5 ml of PBS to remove excess FITC and then pelleted. Each pellet was resuspended in 1 ml of PBS and then 100 μl of cell suspension was added to a U-bottomed 96-well plate (Greiner). Plates were kept at 4°C in darkness overnight. Images were captured using the InGenius3 gel documentation system (Syngene). The smallest starting volume of bacteria with a visible pellet of non-agglutinated cells was used for future assays.

2.7.9. Agglutination assay

The optical density of an overnight broth culture was adjusted to an OD_{600} of 1.0. Aliquots of the following volumes were used for each bacterial strain: 0.5 ml of *E. coli*, 0.9 ml of *S. gordonii* and *S. mutans* and 0.7 ml of *S. pneumoniae.* FLAG peptide elution buffer (150 ng/μl) and PBS were used as negative controls. Whole saliva centrifuged at 13,000 rpm to remove cells and debris was used as an agglutinating positive control. The same protocol was used as described in Section 2.7.8. After the washing steps to remove excess FITC, cells were pelleted and resuspended in 1 ml of PBS. One hundred microlitres of FITC-stained bacterial suspension was aliquoted (per sample or control to be tested) and pelleted. The pellet was

resuspended in 100 µl of purified WFDC2 (1 µM) diluted in PBS or in 100 µl of control. Each 100 µl suspension was added to a U-bottomed 96-well plate (Greiner) and kept at 4°C in darkness overnight. Images were taken using the InGenius3 gel documentation system (Syngene) to visualise FITC-stained bacteria. A pellet of cells shows no agglutination has occurred, while no visible pellet suggests agglutination has occurred and cells instead exist as a film of agglutinated bacteria.

2.7.10. Biofilm formation assay

The optical density of an overnight broth culture was adjusted to an OD_{600} of 1.0. Two hundred microlitres of culture was added to wells of a 96-well plate. Sterile PBS was added to the peripheral wells to avoid water loss in wells containing bacteria as this can affect biofilm formation. Purified WFDC2 protein (1 µM) or an equal volume of FLAG peptide elution buffer (150 ng/µl), PBS or saliva (healthy whole saliva centrifuged at 13,000 rpm to remove cells and debris) controls were immediately added to the bacteria. Sterile BHI was used as a negative control. Plates were incubated at 37°C for 48 hours to allow biofilms to form. After 48 hours, the broth culture containing planktonic bacteria was aspirated and wells were washed twice with PBS. The remaining biofilms were stained with 200 µl/well of 0.1% crystal violet for 10 minutes at room temperature. Wells were washed 3 times with PBS to remove excess dye and plates were left to air dry. The bacteria-bound crystal violet was then solubilised in 200 µl/well of 33% acetic acid and mixed by pipetting up and down. Plates were read at OD_{570} using a microplate reader (Tecan) to measure biofilm formation. The absorbance readings from wells incubated with sterile BHI broth alone were used to determine background staining and were subtracted from all other readings.

2.7.11. Biofilm disruption assay

Biofilm disruption was measured as described in Section 2.7.10 but the purified WFDC2 protein or control was added 24 hours after bacteria were seeded into the 96 well plates i.e. when biofilms had already begun to form. Following addition of protein or control, the plates

were incubated for a further 24 hours at 37°C and the biofilms were stained with crystal violet and analysed as before.

2.8. Cytokine exposure assay

HuSL, CAL27 and H357 cells were seeded into a 6-well plate at a density of 3 x 10^5, 4 x 10^5, 5 x 10^5 respectively, and incubated for 24 hours at 37°C and 5% CO_2. Cells were washed twice in PBS and then serum-starved for 24 hours. Cells were then exposed to TNF-α (20 ng/ml), IFN-γ (5 ng/ml), LPS (50 µg/ml) and IL-1β (20 ng/ml) (all purchased from ThermoFisher Scientific) in 3 ml of serum-free high glucose DMEM for 24 hours. Untreated cells were used as a control. After 24 hours had elapsed, the media was removed from the cells and 0.6 ml of 1X Monarch RNA Protection Reagent (NEB) was added to each well. Plates were stored at -80°C. RNA was extracted from the cells using the Monarch Total RNA Miniprep Kit (NEB) as per the manufacturer's instructions. One microgram of RNA was reverse transcribed to cDNA as previously described (Section 2.2.2). cDNA was used for RT-qPCR with TaqMan probes against human WFDC2 and B2M was used as the endogenous control probe as previously described (Section 2.2.5).

2.9. CRISPR/Cas9 gene editing

CRISPR/Cas9 gene editing was performed based on the protocol outlined by Ran et al (2013).

2.9.1. Single-guide RNA design

Oligos (~20 bp) were designed to be complementary to target sites within human WFDC2 genomic DNA. The target sites were required to be directly upstream of a 5'-NGG sequence which is the protospacer adjacent motif (PAM site). An additional 5'CACC sequence was added to the F primer and a 5'AAAC sequence was added to the R primer: when the oligos are annealed to form the single-guide RNA (sgRNA) these additional sequences provide overhangs which are complementary to the overhangs of the cut target plasmid. The sgRNA sequences are summarised in Table 2.9.

Oligos		Sequence (5'-3')
Construct 1 (sgRNA1)	F	CACCGGCACAGGAGCAGAGAAGAC
	R	AAACGTCTTCTCTGCTCCTGTGCC
Construct 2 (sgRNA2)	F	CACCGGGCGTGTGCCCCGAGCTCC
	R	AAACGGAGCTCGGGGCACACGCCC

Table 2.9. Single-guide RNA sequences.

2.9.2. Single-guide RNA preparation

Oligos were resuspended in nuclease-free H_2O to a final concentration of 100 µM. Both sgRNA1 and sgRNA2 were prepared by annealing 1 µl of forward oligo and 1 µl of reverse oligo together in the presence of 1X T4 ligation buffer and nuclease-free H_2O. Annealing was carried out using the following PCR conditions: 5 minutes at 95°C, 75°C for 10 minutes, 50°C for 10 minutes and 25°C for 10 minutes.

2.9.3. Vector preparation

The target plasmid, pSpCas9(BB)-2A-GFP (abbreviated to pX458) was a gift from Feng Zhang (AddGene) and encodes the guide RNA scaffold, in addition to the Cas9 endonuclease, GFP and ampicillin resistance cassettes as depicted in Fig 2.2 (Ran et al., 2013). To prepare the vector, pX458 was cut with BbsI-HF (NEB) restriction enzyme and resolved on a 1% agarose gel. The cut plasmid was then extracted from the gel and isolated using a GeneJET Gel Extraction Kit (ThermoFisher Scientific). DNA yield was determined using a NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific).
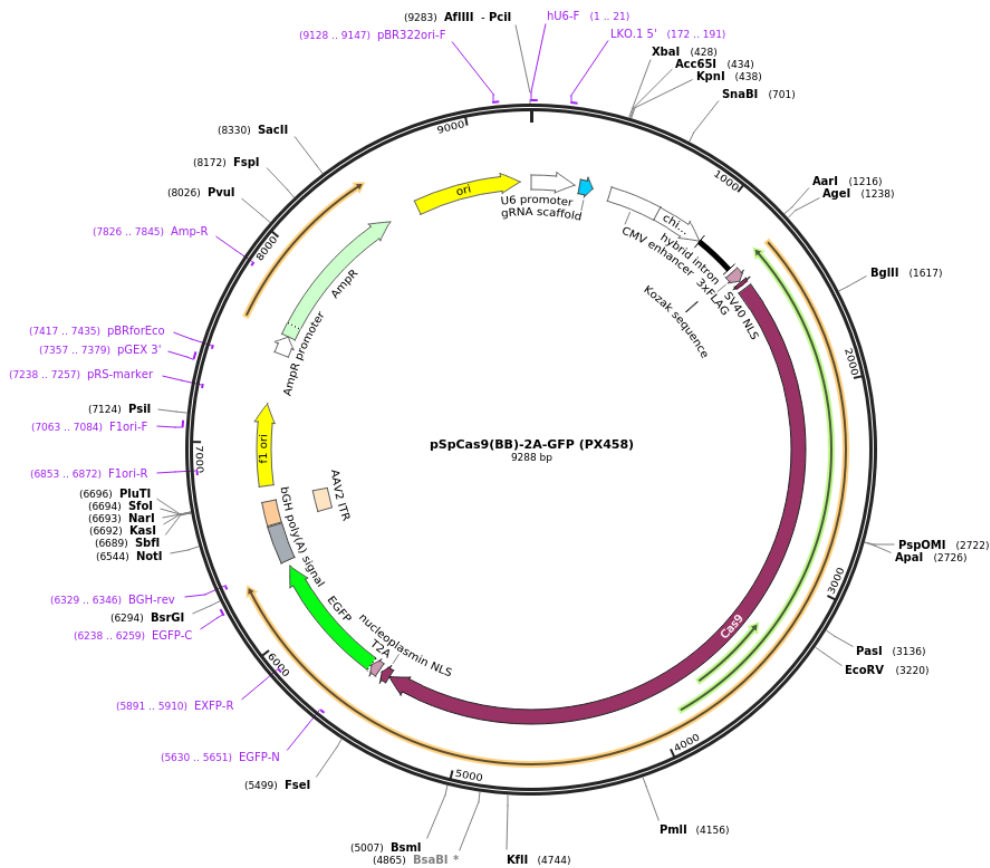
Fig 2.2. The pX458 vector. A plasmid map of the target vector, pX458. Labelled cassettes include the U6 promoter, the gRNA scaffold, Cas9 endonuclease gene, GFP gene and ampicillin resistance gene. Restriction sites are also shown. The plasmid vector is 9288 bp in size. Taken from Addgene: plasmid #48138; http://n2t.net/addgene:48138; RRID: Addgene_48138.

2.9.4. Ligation

The annealed oligos, sgRNA1 and sgRNA2, were diluted 1:200 in nuclease-free H_2O and then ligated into the cut pX458 vector using T4 DNA Ligase (Promega) and Ligase 10X Buffer (Promega) following the manufacturer's protocol. The total 10 µl reaction was incubated overnight at 4°C. Vector incubated without sgRNA was included as a control.

2.9.5. Transformation

Following overnight incubation, ligated vectors (pX458-sgRNA1 and pX458-sgRNA2) were transformed into competent *E. coli* cells as described in Section 2.4.6 and selected using LB-agar plates containing ampicillin (100 µg/ml). Successful transformants were picked, inoculated into 2 ml LB-broth with ampicillin (100 µg/ml) and incubated at 37°C, 220 rpm

overnight. The overnight broth culture was then used for subsequent plasmid isolation (see Section 2.4.7). Glycerol stocks were also made and kept at -80°C for future use.

2.9.6. Confirmation of target insertion

Plasmid DNA was digested using restriction enzymes. Expected fragment sizes were confirmed by agarose gel electrophoresis to show that the plasmid had been successfully transformed into bacteria. The pX458 plasmid was then confirmed to contain the expected sgRNA sequences by DNA sequencing (Research Core Facility, the Medical School, University of Sheffield, UK).

2.9.7. Clonal isolation test

CAL27 cells were serially diluted in a 96-well plate in order to test their capacity to grow from a single cell. One hundred microlitres of media was added to all wells aside from well A1. Well A1 was seeded with 4,000 CAL27 cells in 200 µl of media. One hundred microlitres of media from well A1 was transferred to well B1, mixed by pipetting, and the process repeated down the column. The 100 µl of media transferred from well H1 was discarded. An additional 100 µl of fresh media was added to all wells of column A. One hundred microlitres of media from column A was transferred to column B, mixed by pipetting, and repeated across all columns of the plate. The 100 µl of media from column 12 was discarded. All wells had 100 µl of media added to bring the final well volume to 200 µl. A half media change was carried out every 2 days. After 7 days at 37°C and 5% CO_2 the cells were fixed in 70% ethanol and stained using 0.1% crystal violet to visualise single cell colonies.

2.9.8. Transfection using JetPrime

CAL27 cells were seeded into a 6-well plate at a density of 3.75 x 10^5 cells per well and incubated overnight at 37°C and 5% CO_2. Prior to transfection, a media change was performed. Four wells of CAL27 cells were co-transfected with pX458 with sgRNA1 insert (1 µg) and pX458 with sgRNA2 insert (1 µg) using 200 µl of JetPrime buffer (Polyplus) and 4 µl

JetPrime transfection reagent (Polyplus) per the manufacturer's recommendations. Media was changed the following day.

2.9.9. Fluorescence activated cell sorting

Transfected CAL27 cells were trypsinized and then sorted using fluorescence activated cell sorting (FACS). A BDFACSAria IIu cell sorter (Research Core Facility, the Medical School, University of Sheffield, UK) was used to separate and collect the GFP positive and GFP negative cells from the transfected population. Cell sorting was performed by Sue Clark (Flow Cytometry Core Facility technician). Non-transfected wild-type CAL27 cells were used to calibrate the machine and set a baseline for fluorescence. The GFP negative sorted cells were used as an internal control for all downstream assays.

2.9.10. Single cell colony growth and expansion

GFP positive and GFP negative FACS sorted cells were serially diluted for clonal isolation in 96-well plates, as described in Section 2.9.7. A half media change was carried out every 2-3 days. After 7 days growth, rounded colonies radiating from a central point were identified as potential clonal colonies and were marked. Clonal colonies were grown for a further 7 days, passaged into 24-well plates and then expanded into T25 flasks. Stocks of cells were made in 10% DMSO and stored in liquid nitrogen for future use.

2.9.11. Confirmation of positive CRISPR edited colonies

Genomic DNA (gDNA) was extracted from the clonal colonies (multiple GFP positive colonies and a GFP negative control) using the Wizard Genomic DNA Purification Kit (Promega). DNA was quantified using a Nanodrop 1000 Spectrophotometer (ThermoFisher Scientific) and stored at 4°C.

To determine whether the clonal colonies had been successfully gene edited, primers were designed to amplify the nucleotide sequence surrounding and including the sgRNA target site. The GFP negative FACS sorted colony was used as a negative control. Two different primer

pairs were designed, both with an F primer in intron 1, 50-100 bp upstream of the sgRNA target site, and an R primer in exon 2, 50-80 bp downstream of the sgRNA target. Primer sequences are summarised in Table 2.10.

Gene target		Sequence (5'-3')	Product size (bp)
CRISPR target site in WFDC2	F1	GACTCCTAGGGCCAGAGACT	226 bp
	R1	CTGCAGCACTTGAGGTTGTC	
	F2	TGGGGTTAAGGTTTGGAGCA	223 bp
	R2	AGAGAGCAGAAGGTGGCAC	

Table 2.10. Primer sequences utilised for gene editing analysis. Primer sequences used to probe upstream and downstream of the CRISPR target site are summarised. The expected PCR product sizes (bp) are shown.

Five hundred nanograms of gDNA was used in the PCR reaction with Phusion High-Fidelity PCR Master Mix (NEB). Thermocycling conditions were as follows: 30 seconds at 98°C, 30 cycles of 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds, followed by 2 minutes at 72°C. PCR products were resolved on a 2% agarose gel for 80 minutes at 80 V and visualised using the InGenius3 gel documentation system (Syngene).

Colony 11 produced a smaller PCR product compared to the negative control and was thus identified as a potential gene edited colony. The PCR product was 'cleaned-up' using an Isolate II PCR and gel kit (Bioline) according to the manufacturer's instructions and the purified product was resolved on a 2% agarose gel for 120 minutes at 80 V and visualised using the InGenius3 gel documentation system. The resultant bands were individually extracted from the gel using an Isolate II PCR and gel kit (Bioline) and then a sample of each eluted product was resolved on a 2% agarose gel for 120 minutes at 80 V to ensure single products had been successfully extracted.

Five microlitres of each gel extracted and purified PCR product with 5 µl of forward or reverse primer was analysed by Sanger sequencing (GATC Biotech, Eurofins Genomics). Successful

74

CRISPR gene editing was confirmed by BLAST searching using NCBI blastn Suite to compare the CRISPR target sequence with the WT sequence.

2.9.12. Analysis of WFDC2 expression in CRISPR edited cells

RNA was extracted from gene edited colony 11 cells and the FACS sorted negative control cells at passage 1, 2 and 3 (post-FACS) using a Monarch Total RNA Miniprep Kit (NEB) as per the manufacturer's instructions. A batch of WT CAL27 cells was also used as a control and had RNA extracted at the same time-points, however, this batch had not been FACS sorted. RNA yield was quantified using a Nanodrop 1000 Spectrophotometer (ThermoFisher Scientific) and cDNA was synthesised from 1 µg of total RNA template using the High-Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific) according to the manufacturer's protocol.

Resultant cDNA was used for downstream qPCR analysis using a TaqMan probe for human WFDC2 variant 1 and a B2M probe as the endogenous control (for method see Section 2.2.5).

cDNA was also used for PCR using the protocol described in Section 2.2.3. Primers against the full length WFDC2 gene, V0, V2 and OAZ1 were used (see Table 2.2 for primer pairs). PCR thermocycling conditions are described in Section 2.2.3. PCR products were resolved on a 2% agarose gel (see Section 2.2.4) and visualised using the InGenius3 gel documentation system (Syngene).

Conditioned media from the colony 11 cells, negative control cells and WT cells was collected as described in Section 2.1.2 and concentrated 10-fold using a Vivaspin 6 centrifugal concentrator with a 10 kDa molecular weight cut-off point (Sartorius). Protein concentration was quantified via BCA assay as described in Section 2.3.2 and then 45 µg of total protein was analysed by Western blotting (see Section 2.3.3) using a monoclonal antibody against human WFDC2 (Abcam). Protein was detected by exposure to X-ray film and passage through a Compact X4 automatic X-ray film processor (Xograph).

The concentration of WFDC2 protein in the conditioned media was also quantified using a Human HE4/WFDC2 Quantikine ELISA Kit (R&D Systems). Samples were diluted 250-fold and assayed as described in Section 2.3.7.

2.10. CRISPR cell functional assays

All functional assays were performed in triplicate, with three technical repeats within each experiment. Results are shown as the mean of the triplicate experiments and error bars represent standard error of the mean (SEM). Statistical significance was analysed by one-way analysis of variance (ANOVA).

2.10.1. Cell morphology

Cell morphology was visualised using a light microscope (Olympus CKX41) with a 10X lens and four random fields of view were imaged using a Nikon Coolpix camera.

2.10.2. Viability assay

Cell viability was determined using CellTiter 96 Aqueous One Solution (Promega). Cells were serum-starved overnight and then seeded into a 96-well plate at a density of 5×10^3 cells per well in 100 μl of high glucose DMEM supplemented with 10% (v/v) FBS. Cells were incubated for 72 hours at 37°C and 5% CO_2. After 72 hours had elapsed, media was changed to 100 μl of serum-free high glucose DMEM and 20 μl of CellTiter 96 Aqueous One Solution was added to each well avoiding exposure to direct light. Empty wells were treated the same way and used as a negative control. The plate was incubated for 2 hours at 37°C, protected from the light. The absorbance was read at 490 nm using a microplate reader (TECAN). The mean absorbance of the empty control wells was subtracted from each reading to account for background.

2.10.3. Manual cell counting

Cells were seeded in a 6-well plate at a density of 3×10^5 cells per well in 3 ml of high glucose DMEM supplemented with 10% (v/v) FBS and incubated for 48 hours. Cells were then

trypsinized, resuspended in 2 ml of media and manually counted using a haemocytometer to give a total cell count.

2.10.4. Adhesion assay

Cell adherence was measured by coating 96-well plates with 100 µl/well of rat tail collagen type I (30 µg/ml; Corning) or bovine fibronectin protein (30 µg/ml; ThermoFisher Scientific) for 1 hour at 37°C. Plates were then blocked for 1 hour at 37°C using 100 µl/well of BSA in PBS (10 mg/ml). Each well was seeded with 25,000 cells per well in serum-free media for 1 hour at 37°C. Cells had been serum-starved overnight. After 1 hour elapsed, unbound cells were removed by washing each well with 200 µl of PBS. Cells remaining bound to the plate were quantified by adding 100 µl of serum-free media and 20 µl of CellTiter 96 Aqueous One Solution per well, avoiding exposure to direct light. Empty wells were treated the same way and used as a negative control. The plate was incubated for 2 hours at 37°C, protected from the light. The absorbance was read at 490 nm using a microplate reader (TECAN). The mean absorbance of the empty control wells was subtracted from each reading to account for background.

2.10.5. Migration assay

Cell migration was measured using a Transwell assay. Cells were serum-starved overnight. The following day, cells were trypsinized and resuspended in 0.1% BSA in serum-free high glucose DMEM. Each 12 mm, 8 µm pore Transwell insert (Falcon) was seeded with 50,000 cells in 200 µl of 0.1% BSA serum-free media and the well beneath the insert was filled with 0.5 ml of high glucose DMEM supplemented with 10% (v/v) FBS. Cells were left to migrate for 7 hours. Non-migrated cells within the insert were wiped away with a cotton bud. Cells that had migrated through the insert pores were fixed using 0.5 ml of ice-cold 100% methanol for 10 minutes. Inserts were rinsed with PBS and then migrated cells were stained using 0.1% (w/v) crystal violet for 20 minutes. Inserts were washed multiple times with distilled H_2O to remove excess stain and were left to dry. Each insert was imaged across 7 fields of view and

the number of migrated cells were manually counted using a plugin available on ImageJ named 'Cell Counter'. The mean number of migrated cells per field of view was calculated for each insert.

2.10.6. Invasion assay

Cell invasion was measured using a Matrigel assay. Cells were serum-starved overnight. Matrigel aliquots (Corning) were pre-thawed for 2 hours at 4°C prior to use and were subsequently kept on ice. Pre-chilled pipette tips and media were used when diluting the Matrigel. Each 12 mm, 8 µm pore Transwell insert (Falcon) was filled with 100 µl of Matrigel (0.25 mg/ml) in serum-free high glucose DMEM and left to set overnight at 37°C. The following day, cells were trypsinized and resuspended in 0.1% BSA in serum-free high glucose DMEM. Each Matrigel-coated insert was seeded with 100,000 cells in 200 µl of 0.1% BSA serum-free media and the well beneath the insert was filled with 0.5 ml of high glucose DMEM supplemented with 10% (v/v) FBS. Cells were left to invade for 24 and 48 hours, respectively. Non-invaded cells and Matrigel within the insert were wiped away with a cotton bud. Cells that had invaded through the insert pores were fixed using 0.5 ml of ice-cold 100% methanol for 10 minutes. Inserts were rinsed with PBS and then invaded cells were stained using 0.1% (w/v) crystal violet for 20 minutes. Inserts were washed multiple times with distilled H_2O to remove excess stain and were left to dry. Stained cells were solubilized using 200 µl of 33% acetic acid for 20 minutes. The absorbance was measured at 570 nm using a microplate reader (Tecan).

2.11. Tissue specimen preparation and histological analysis

2.11.1. Wfdc2-knockout mouse tissues

Homozygous and heterozygous WFDC2-knockout mice were established at the MCR Harwell Institute under the International Mouse Phenotyping Consortium (IMPC) project (http://www.mousephenotype.org/data/genes/MGI:1914951). Mice were bred and maintained by the Mary Lyon Centre, MRC Harwell and were housed in specific-pathogen free conditions.

All animal experimentation was approved by the Animal Welfare and Ethical Review Body at MRC Harwell. The humane care and use of mice in this study was under the authority of the appropriate UK Home Office Project Licence.

WFDC2 silencing was established by CRISPR-induced mutation using two sgRNAs. A 2431 nucleotide region of genomic DNA spanning intron 2 to intron 4 was deleted to induce a null allele in C57BL/6N mice. Heterozygote animals were culled between ages of 112-115 days. Six WT and six heterozygote animal tissues were donated to this study, with a ratio of 1:1 of male and female animals in each phenotypic group. Salivary gland, trachea, lung, stomach, gastrointestinal tract, kidney, testis, epididymis and ovary tissues from WT and heterozygote animals were histologically analysed.

Homozygote animals were analysed at the MRC Harwell Institute and the results were communicated to researchers within this study. The volume and crown-to-rump length of embryos at embryonic day E14.5 and E18.5 were measured. Micro computed tomography (micro-CT) analysis was used to perform 3D visual assessments of the embryos. Morphological differences between WT and homozygote embryos were recorded. Trachea diameter was also determined from the micro-CT scans.

2.11.2. Wfdc2-knockout mouse tissue processing

Mouse tissues were received from the MRC Harwell Institute in 10% neutral buffered formalin. Kidneys and stomach were cut with a scalpel in the 'butterfly' manner while the 'Swiss rolled' intestines were also cut into smaller slices. All other tissues were small enough to be processed whole so were omitted from the grossing step. Grossed and whole tissues were placed into labelled cassettes. Smaller tissues were held in place within the cassette with sponge. Tissues were processed overnight using an automated Shandon Citadel 2000 tissue processor (ThermoFisher Scientific). The processing schedule is summarised in Table 2.11.

Stage	Solution	Processing time
Dehydration	70% alcohol	1 h
	80% alcohol	1 h
	90%alcohol	1 h 30 mins
	Absolute alcohol I	1 h 30 mins
	Absolute alcohol II	1 h 30 mins
	Absolute alcohol III	2 h
Clearing	Toluene I	1 h 30 mins
	Toluene II	2 h
	Xylene	PASS
Wax infiltration	Wax I	2 h
	Wax II	2 h

Table 2.11. Tissue processing schedule.

2.11.3. Tissue sectioning

After processing, tissues were infiltrated with paraffin wax (Paraplast) using an embedding centre (Leica EG1160H). Care was taken to orientate each tissue to ensure optimal positioning of the plane when sectioning. Tissues were left to solidify in wax on a cooling plate (Leica EG1150C). A rotary microtome (Leica RM2235) was used to cut 4 µm thick ribbons from the paraffin blocks which were 'floated out' onto a floatation bath (Electrothermal) set at 38°C. Tissue ribbons were then mounted onto SuperFrost Plus microscope slides (ThermoFisher Scientific) and then oven-dried for 15 minutes at 60°C. Sectioning was carried out by Hayley Stanhope (School of Clinical Dentistry, University of Sheffield).

2.11.4. Human tumour tissue

Paraffin-embedded human oral carcinoma tissue was derived from the Oral Pathology tissue bank (School of Clinical Dentistry, University of Sheffield). Tissue blocks were sectioned using a microtome as described in Section 2.11.3.

2.11.5. Haematoxylin and eosin staining

Tissues were haematoxylin and eosin (H&E) stained using an automated linear staining machine (Shandon). The staining schedule is summarised in Table 2.12.

Step	Solution	Step	Solution
1	Xylene	15	Running tap water
2	Xylene	16	Scott's tap water substitute
3	Xylene	17	Running tap water
4	99% IDA	18	Eosin Y (Shandon)
5	99% IDA	19	Eosin Y (Shandon)
6	70% IDA	20	Eosin Y (Shandon)
7	Distilled water	21	Running tap water
8	Distilled water	22	99% IDA
9	Harris' haematoxylin (Shandon)	23	99% IDA
10	Harris' haematoxylin (Shandon)	24	99% IDA
11	Harris' haematoxylin (Shandon)	25	Xylene
12	Harris' haematoxylin (Shandon)	26	Xylene
13	Running tap water	27	Xylene
14	0.1% acid alcohol	28	Xylene

Table 2.12. Haematoxylin and eosin staining schedule.

Following staining, slides were mounted in DPX mountant (ThermoFisher Scientific) with a coverslip. Tissues were visualised and imaged using an Olympus BX51 microscope, Colour View Illu camera and Cell^D software (Olympus Soft Imaging Solutions).

2.11.6. Immunohistochemistry

Slides were de-waxed by immersion in xylene (2 x 5 minutes) and rehydrated in 100% ethanol (2 x 5 minutes). Endogenous peroxidase activity was blocked by immersion in 3% hydrogen peroxide in methanol for 20 minutes and then slides were rinsed in PBS. Antigen retrieval was carried out by heating in 0.01 M sodium citrate tribasic for 8 minutes in the microwave on full

power. Slides were washed for 5 minutes in PBS and then blocked for 30 minutes in 100%

normal serum (Sigma) at room temperature in a humidified chamber. The blocking serum used

was derived from the same animal that the secondary antibody was raised in. Slides were

incubated in primary antibody diluted in 100% normal serum overnight at 4°C in a humidified

chamber. Primary antibodies and the appropriate secondary antibody and blocking serum

used are summarised in Table 2.13. All antibodies required antigen retrieval.

Primary antibody	Manufacturer or source	Description	Working concentration or dilution	Secondary antibody	Blocking serum
Anti-WFDC2	Abcam (ab200828)	Rabbit monoclonal to human WFDC2	2.30 µg/ml	Anti-rabbit-HRP	100% goat serum (Sigma-Aldrich)
Anti-Wfdc2	Antiserum kindly donated by Ronny Drapkin[1]	Rabbit polyclonal to mouse Wfdc2	1:800		
Anti-Slpi	RayBiotech (DS-PB-02937)	Rabbit polyclonal to human/mouse SLPI	1.10 µg/ml		
Anti-CCSP	Antiserum developed by Lynne Bingle[2]	Rabbit polyclonal to mouse CCSP	1:1000		
Anti-MUC5B	Santa Cruz (sc-21768)	Rabbit polyclonal to human MUC5B	0.4 µg/ml		

Table 2.13. Antibodies for immunohistochemistry. Catalogue number, working dilution, secondary antibody and blocking serum required are summarised. [1]Ronny Drapkin (Perelman School of Medicine, The University of Pennsylvania, USA), [2]Lynne Bingle (School of Clinical Dentistry, University of Sheffield, UK).

The following day slides were washed in PBS (2 x 5 minutes) with agitation and then incubated

with secondary biotinylated antibody (Vector Laboratories) for 30 minutes at room temperature

in a humidified chamber. Slides were washed in PBS (2 x 5 minutes) with agitation and then

incubated with Vectastain ABC solution (Vector Laboratories) for 30 minutes at room

temperature in a humidified chamber. Slides were washed in PBS (2 x 5 minutes) with

agitation. Positive antibody reaction was detected following incubation with DAB substrate

(Vector Laboratories) for 3 minutes and the reaction stopped by submerging the slides in

distilled H_2O for 5 minutes. Slides were counterstained in haematoxylin using an automated

linear staining machine (Shandon; steps 11 to 17 and 21 to 28 only, as described in Table 2.12). Slides were mounted in DPX mountant (ThermoFisher Scientific) with a coverslip. Tissues were visualised and imaged using an Olympus BX51 microscope, Colour View Illu camera and Cell^D software (Olympus Soft Imaging Solutions).

IHC analysis of Wfdc2, Slpi and CCSP expression in heterozygote knockout mouse tissues was performed by Jessica-Leigh Tallis (Sheffield Undergraduate Research Experience scheme student, University of Sheffield) under the supervision of Hannah Armes.

2.11.7. Picrosirius red staining

To visualise collagen, slides were de-waxed by immersion in xylene (2 x 5 minutes), rehydrated in 100% ethanol (2 x 5 minutes) and then counterstained in haematoxylin (Table 2.12, steps 11-13 only). Slides were incubated with picrosirius red stain (0.1% w/v of Sirius red F3B in saturated aqueous picric acid) for 1 hour at room temperature in a humidified chamber. Slides were washed in 0.1% acid alcohol (Table 2.12, step 14, passed through twice) and Scott's tap water substitute and then dehydrated (Table 2.12, steps 15-16 and 21-28). Slides were mounted and imaged as described in Section 2.11.5.

2.11.8. Quantification of immunohistochemical staining

Stained slides were scanned and digitalised (SITraN, University of Sheffield). DAB and Picrosirius staining were quantified using Aperio ImageScope software (Leica Biosystems) whereby sections of tissue to be analysed were highlighted using a drawing tool and then analysed using the Positive Pixel Count v9 algorithm (see Appendix for visual representation of this process). When analysing the stomach and intestines, chyme was excluded from the analysis due to it staining non-specifically, to avoid skewing the results. Positively stained pixels were counted and used to determine percentage of positivity from all pixels. Sites on the slide without tissue are not recognised as pixels due to the pre-set colour saturation threshold which blank space does not meet. Percentage positivity was determined using the following equation: $\left(\frac{Number\ of\ positively\ stained\ pixels}{Total\ numer\ of\ positvely\ and\ negatively\ stained\ pixels} \right) x 100$

2.12. Online resources

Human and murine WFDC2 DNA and peptide sequences were established from Ensembl genome browser [www.ensembl.org].

The presence of N-linked glycans and O-linked glycans in human and murine WFDC2 were predicted using NetNGlyc and NetOGlyc online glycan predictor software, respectively [http://www.cbs.dtu.dk/services/NetNGlyc/] [http://www.cbs.dtu.dk/services/NetOGlyc/]. The inputted WFDC2 peptide sequences were recovered from Ensembl genome browser.

Nucleotide BLAST was used to identify regions of similarity between the WT WFDC2 sequences and Sanger sequencing results during WFDC2 gene cloning and WFDC2 CRISPR gene editing. Nucleotide BLAST was also used to identify bacterial strains derived during 16S sequencing [https://blast.ncbi.nlm.nih.gov/Blast.cgi].

ExPASy translate tool was used to identify changes in peptide sequence resulting from CRISPR gene editing. Sequencing data from PCR products of colony 11 were inputted into ExPASy to identify any CRISPR-induced premature stop codons or other mutations [https://web.expasy.org/translate/].

ExPASy Compute pI/Mw tool was used to compute the theoretical isoelectric point (pI) and molecular weight (MW) from peptide sequences (Bjellqvist et al., 1994). The tool was used to identify differences in pI and MW between human/murine WFDC2 and SLPI. [https://web.expasy.org/compute_pi/].

2.13. Statistics

GraphPad Prism 7 software was used to determine statistical significance. An unpaired t-test was used to compare two groups. One-way analysis of variance (ANOVA) was used to analyse more than two groups. A P-value of <0.05 was considered to be statistically significant. The number of biological repeats are shown as 'N='. Technical repeats are represented as 'n='.

Chapter 3

Synthesis and analysis of recombinant WFDC2

3.1.   Introduction

Several reports in the literature have utilised recombinant human or murine WFDC2 to carry out *in vitro* functional assays with the aim of understanding more about the protein's function (Chhikara et al., 2012; Hua et al., 2014; LeBleu et al., 2013; Nishimura et al., 2008). The results frequently vary between reports, for instance, significantly different concentrations of protein are used to exhibit similar effects (as described in Section 1.3.5). One of the potential reasons for discrepancies between reports is the inconsistency in the size of the recombinant protein being used, which suggests that the protein is either not WFDC2, or that it lacks the appropriate post-translational modifications, both of which could significantly affect the results. The recombinant WFDC2 orthologues synthesised in this chapter will be subjected to multiple analyses to ensure that they are the correct protein and that they are appropriately glycosylated. Importantly, the recombinant protein will be compared to endogenous protein via Western blotting to ensure that it is the correct size. To further the understanding of the structure of human WFDC2, the gene construct will be subjected to site-directed mutagenesis (SDM) to confirm the N-glycosylation site. Finally, the concentration of WFDC2 in saliva and bronchial cell secretions will be analysed to identify the endogenous levels of WFDC2 as the literature is lacking in quantitative analysis of WFDC2 protein expression outside of the serum values identified in diseased patients (Molina et al., 2011; Ruggeri et al., 2011).

3.2.   Aims and objectives

The overall aim of the work presented in this chapter was to synthesise recombinant human and mouse WFDC2 by gene cloning of their predominant splice variants followed by

transfection into mammalian cells. The glycosylated proteins were synthesised with the main objective of analysing their activities in downstream *in vitro* studies. SDM and enzymatic glycan cleavage were used to elucidate the presence of the correct post-translational glycan modifications. The molecular weight of the recombinant protein was compared against endogenous WFDC2 in oral- and lung-derived secretions via Western blot.

3.3.   Materials and methods

To access the relevant materials and methods utilised in this chapter, please refer to the sections below:

- Gene cloning of human and murine WFDC2 was carried out as described in Section 2.4.

- The transfection method was optimised as described in Sections 2.4.11 and 2.4.12. Transfections were subsequently carried out as per Section 2.4.12. Transfected cells were analysed by immunofluorescence microscopy as described in Section 2.4.14.

- Protein was purified from conditioned media using a FLAG affinity resin as defined in Section 2.4.13. The purified products were analysed by Western blotting and native PAGE gel as described in Sections 2.3.3 and 2.3.6, respectively. Protein concentration was quantified via BCA assay as per Section 2.3.2.

- Glycosylation analysis was carried out via SDM as outlined in Section 2.5. Enzymatic analysis coupled with Western blotting was also used for analysis of glycans, as described in Section 2.5.2 and 2.5.3.

- Endogenous WFDC2 protein was analysed by Western blotting and also quantified using ELISA analysis using the method described in Section 2.3.7.

3.4. Synthesis of recombinant WFDC2 orthologues

3.4.1. Cloning

Recombinant human and murine WFDC2 were synthesised via conventional cloning techniques. Primers were designed to span the whole coding region of the predominant splice variant (V0) of both orthologues. The primers were also designed to incorporate a FLAG tag sequence at the terminus of each gene immediately before the stop codon and enzyme restriction sites upstream and downstream of the coding DNA. PCR was used to amplify target cDNA encoding the V0 splice variant of both WFDC2 orthologues. The resultant PCR products were confirmed to be of the expected size by DNA gel electrophoresis (Fig 3.1).
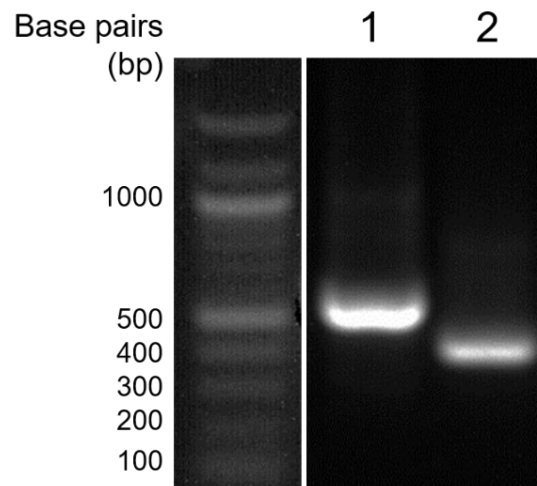
Figure 3.1. Resolution of PCR products for cloning of WFDC2 orthologues. Primers were designed to amplify the whole coding region of human and mouse WFDC2 orthologues in addition to attaching a FLAG tag at the gene terminus. Amplified PCR products were resolved on a 1% agarose gel. Lane 1: mouse Wfdc2 PCR product (516 bp); Lane 2: human WFDC2 PCR product (416 bp).

PCR products were sub-cloned into a pCR®II-TOPO® vector to remove the 3' adenine overhang resulting from Taq polymerase amplification. The DNA inserts were excised from the vector by restriction digestion and subsequently ligated into the pVR1255 target vector. Following transformation into competent *E. coli* cells, the plasmid DNA was isolated from single colonies and digested with restriction enzymes to confirm that the insert was ligated into the plasmid. Successful incorporation of WFDC2 was confirmed by the presence of bands of

the expected sizes when analysed by DNA gel electrophoresis. The results show that multiple

plasmid preparations contained a successfully incorporated insert of the expected size (Fig

3.2). The WFDC2 inserts were further validated by DNA sequencing which confirmed the

nucleotide sequences were identical to the V0 sequences published in online genome

databases (Ensembl).
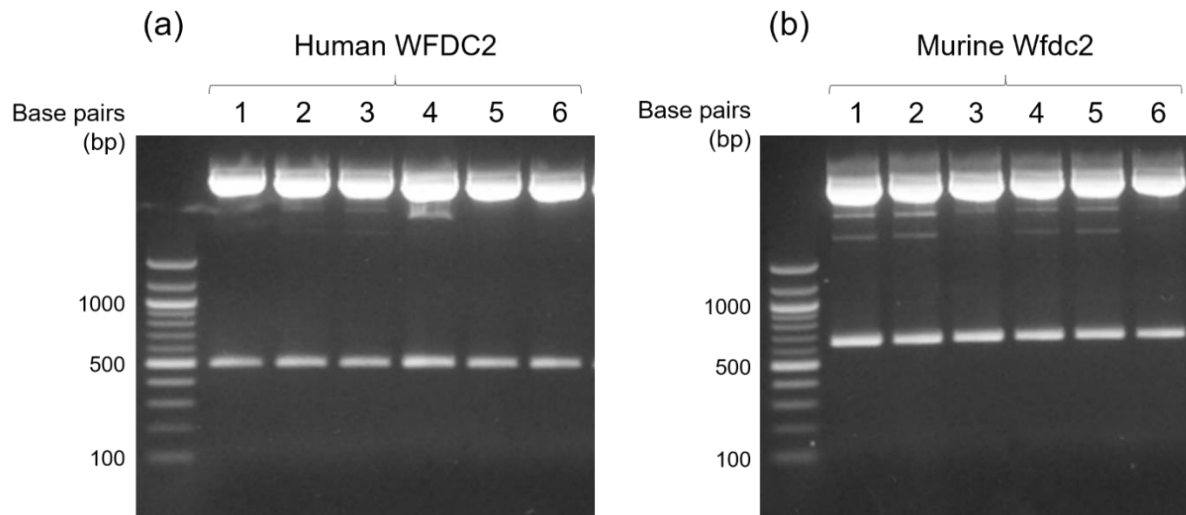
Figure 3.2. Resolution of digested pVR1255 vector for validation of gene insertion. pVR1255 vector DNA was isolated from individual colonies of *E. coli* transformants and digested with restriction enzymes (KpnI and EcoRV). The resultant DNA fragments were resolved on a 1% agarose gel to confirm successful incorporation of the human (a) and mouse (b) WFDC2 insert. Lanes 1-6 represent digested vector DNA isolated from different colonies of successful *E. coli* transformants. The expected band size for successful insertion of the gene of interest was 493 bp for human WFDC2 and 643 bp for murine Wfdc2. The upper band represents the rest of the pVR1255 vector (~4.6 kb).

3.4.2. Protein synthesis and analysis

To produce recombinant protein, pVR1255 vector containing either human or murine WFDC2

was transfected into a mammalian cell line. The calcium phosphate and FuGENE® methods

were used to co-transfect a human embryonic kidney cell line (HEK293), seeded in a 24-well

plate, with the pVR1255 vector (500 ng) and a vector encoding green fluorescent protein

(pEGFP-N1; 100 ng). Co-transfection with pEGFP-N1 was carried out to give an indication of

transfection success and efficiency. Following transfection, cells were lysed and analysed by Western blotting.

Both human and murine WFDC2 were detectable in the transfected HEK293 cell lysate when analysed by Western blotting (Fig 3.3). Human WFDC2 resolved at a molecular weight of ~20-25 kDa, appearing to exist as two distinct bands. Murine Wfdc2 resolved at a molecular weight of ~26-34 kDa. For both human and mouse WFDC2, the band intensity appeared to be stronger in cells that had been transfected using the FuGENE® method compared to those transfected with calcium phosphate.
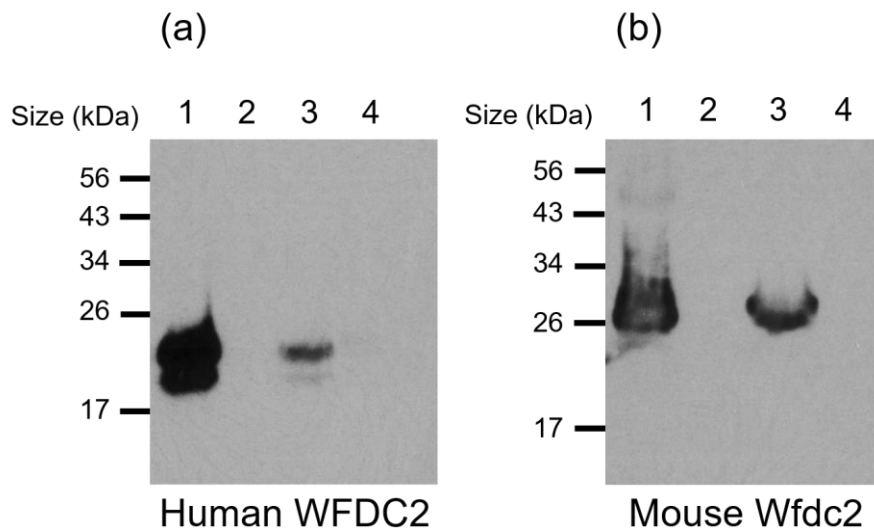
Figure 3.3. Western blot analysis of cell lysate from HEK293 cells transfected with human and mouse WFDC2 using two different transfections methods. HEK293 cells were transfected with pVR1255 vector containing human (a) or mouse (b) WFDC2. FuGENE® and calcium phosphate transfection methods were tested. Expression of recombinant WFDC2 in the cell lysate was analysed by Western blotting using anti-FLAG monoclonal antibody. Lane numbers represent the following: 1 – HEK293 cells transfected with WFDC2 using FuGENE®, 2 – HEK293 cells transfected with empty vector using FuGENE®, 3 – HEK293 cells transfected with WFDC2 using calcium phosphate, 4 – HEK293 cells transfected with empty vector using calcium phosphate.

Following successful calcium phosphate transfection of HEK293 cells in a 24-well plate, the transfection method was scaled up in order to collect larger quantities of recombinant protein. HEK293 cells were seeded into a T75 flask and transfected using the calcium phosphate method with pVR1255 (18.75 µg) and pEGFP-N1 (3.75 µg). Culture media was changed to

serum-free media (6 ml) 48 hours post-transfection and then conditioned media was collected the following day. This process of conditioned media collection and replacement with fresh serum-free media (6 ml) was continued for a further 3 days (i.e. 6 days post-transfection). A representative image of WFDC2 transfection efficiency in HEK293 cells is shown in Fig 3.4. After 4 days of media collection had elapsed cells were lysed for analysis by Western blotting. Conditioned media was concentrated 6-fold using centrifugal concentrator columns and then denatured for downstream analysis.
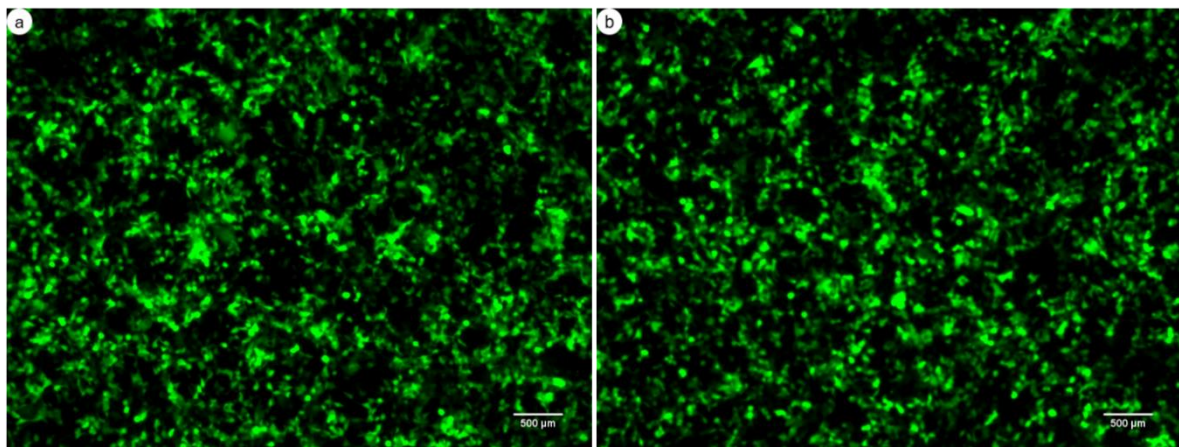
Fig 3.4. HEK293 cells transfected with WFDC2 and eGFP-N1 in a T75 flask using the calcium phosphate method. Representative images show human (a) and murine (b) WFDC2 transfected HEK293 cells 48 hours post-transfection. Cells were co-transfected with eGFP-N1 and the calcium phosphate method was used. Cells were cultured in a T75 flask. Original magnification was 5X.

When conditioned media from transfected HEK293 cells was analysed by Western blotting it was evident that both human and mouse WFDC2 protein were secreted in abundance into the media. Human WFDC2 was visualised as two separate bands, ~18 kDa and ~24 kDa in size respectively. Murine Wfdc2 appeared as a single smeared band from ~26 to 34 kDa. A second ~50 kDa band could be seen when analysing murine Wfdc2 which potentially represents a protein dimer.

The transfected cells continued to secrete protein up to 6 days post-transfection as shown by the continued presence of a positive band when analysed by Western blotting (Fig 3.5). Expression of recombinant WFDC2 in the cell lysate was significantly lower than in the

conditioned media and was often undetectable. All further transfection experiments were carried out to the same time-scale, whereby conditioned media was collected from transfected cells for up to 4 days. HEK293 cells were also transfected with empty pVR1255 vector and the conditioned media collected to act as a control.
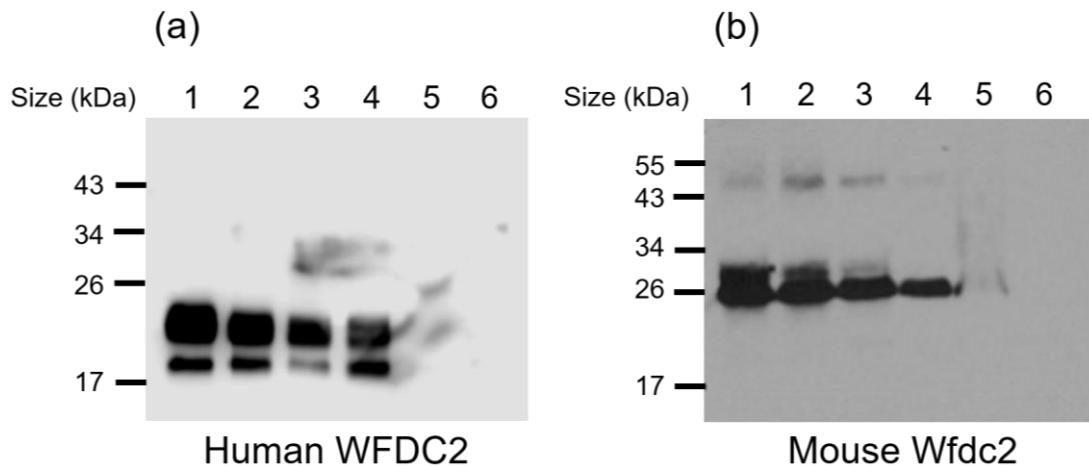
Figure 3.5. Western blot analysis of conditioned media from HEK293 cells transfected with human and mouse WFDC2. HEK293 cells were seeded into a T75 flask and transfected with pVR1255 vector containing the human or mouse WFDC2 insert. Serum-free conditioned media from transfected cells was collected 3 to 6 days post-transfection and analysed by Western blot using a polyclonal antibody against human WFDC2 (R&D Systems) or mouse Wfdc2 (Drapkin), respectively. Lane numbers represent the following: 1 – media collection 1 (day 3), 2 – media collection 2 (day 4), 3 – media collection 3 (day 5), 4 – media collection 4 (day 6), 5 – transfected HEK293 cell lysate, 6 – conditioned media from cells transfected with empty pVR1255 vector.

In order to enrich and purify WFDC2 protein from transfected HEK293 conditioned media, a FLAG affinity resin was used. Conditioned media was incubated with the resin for 2 hours and then FLAG-tagged WFDC2 was eluted using FLAG peptide buffer (150 ng/µl). Each batch of eluted protein was analysed by Western blotting and quantified by BCA assay. The results show that the amount of protein was significantly enriched in purified human WFDC2 (P=0.007) and mouse Wfdc2 (P=0.005) compared to the original media sample (Fig 3.6), as shown by a significant increase in signal when it was quantified by densitometry (Fig 3.7).
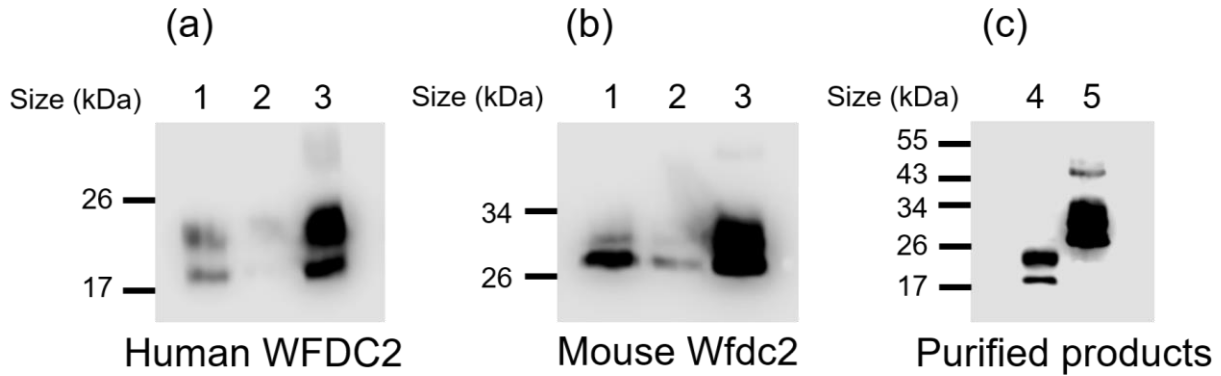
Figure 3.6. Western blot of WFDC2 enrichment from conditioned media to purified protein.
Human (a) and mouse (b) WFDC2 was purified and enriched from conditioned media using a FLAG
affinity resin and the eluted product was analysed by Western blotting using anti-FLAG monoclonal
antibody. Equal volumes of sample were loaded per lane. Lane numbers represent the following: 1 –
unconcentrated conditioned media before purification, 2 – unconcentrated conditioned media after
purification, 3 – purified protein (N=2). (c) Purified human (4) and murine (5) WFDC2 loaded on the
same Western blot for size comparison (N=3).
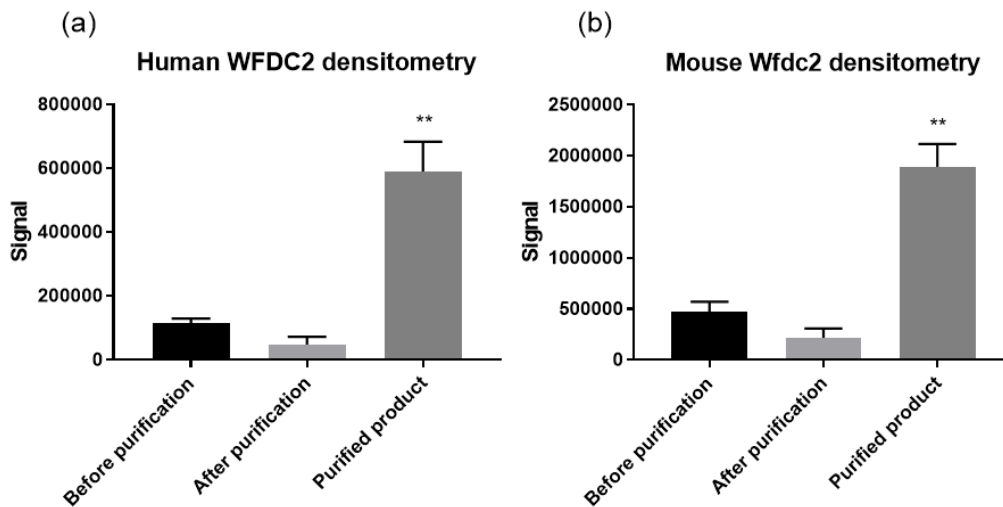
Figure 3.7. Densitometry of Western blot for purified protein. Densitometry results derived from
analysis of Western blots comparing the protein expression in media before purification, after
purification and the purified product for human (a) and mouse (b) recombinant protein (mean ± SD,
N=2); ** $P \leq 0.01$.

To determine the intracellular localisation of WFDC2 when transfected into mammalian cells,

NCI-H292 cells were transfected with human and mouse WFDC2 constructs. Cells were fixed

and stained with anti-FLAG monoclonal antibody or anti-WFDC2 polyclonal antibody and

fluorescently labelled secondary antibody. Mock transfected cells were used as a negative control. NCI-H292 cells, rather than HEK293 cells, were chosen for transfection and imaging due to their more adherent nature. The cells are also larger so protein localisation within the cytoplasm is easier to visualise.

The transfected cells showed an abundance of WFDC2 staining within the cytoplasm when analysed by immunofluorescence microscopy (Fig 3.8). Both human and mouse WFDC2 appeared to exhibit a similar staining pattern.
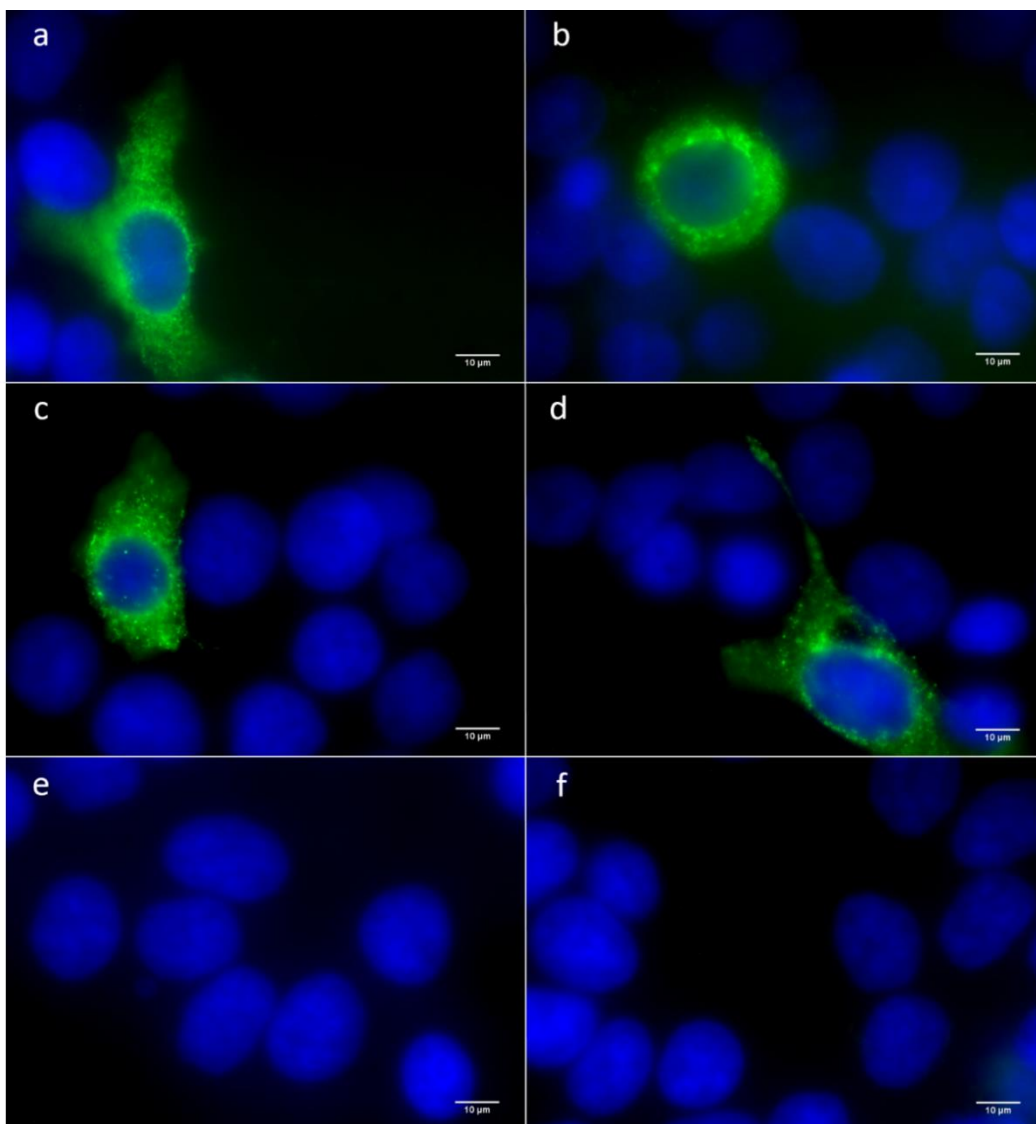
Figure 3.8. Immunofluorescence microscopy of NCI-H292 cells transfected with human and mouse WFDC2 constructs. Cells were transfected with human (a, c, d) or mouse (b) WFDC2 or empty vector (e, f) and then fixed and stained with anti-FLAG monoclonal antibody (Sigma) (a, b, e) or polyclonal anti-human WFDC2 (R&D systems) (c, d, f). Original magnification is 100X.

To determine the native conformation of the recombinant WFDC2 protein, both human and mouse WFDC2 were analysed by native PAGE gel and then analysed by Western blotting. Reaction conditions permitted the protein to retain its native conformation.

The Western blot results show that recombinant human WFDC2 migrates through the native gel at ~26 kDa which is similar to denatured glycosylated WFDC2 protein (~24 kDa) (Fig 3.9a). Recombinant mouse Wfdc2 resolves at ~70 kDa in its native conformation (Fig 3.9b) compared to ~30 kDa in the denatured form.
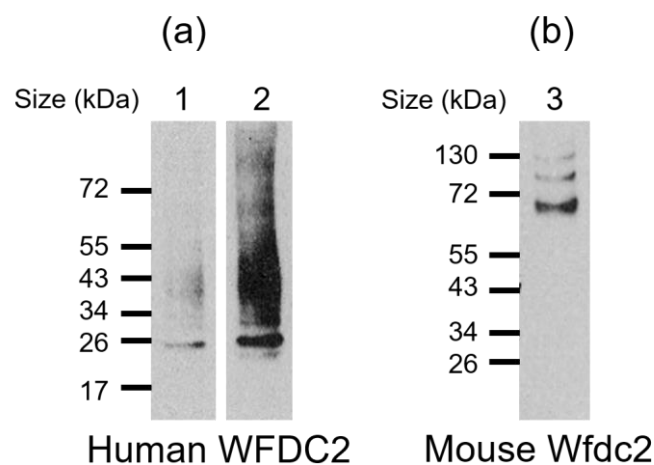
Figure 3.9. Western blot of recombinant WFDC2 analysed by native PAGE gel. Western blot analysis of recombinant human (a) and mouse (b) WFDC2 following resolution on a native PAGE gel. Western blots were probed using polyclonal antibodies against human WFDC2 and murine Wfdc2, respectively. Lane 1 and 2 represent the same gel of human WFDC2 but with different exposure times. Lane 3 shows murine Wfdc2 (N=1).

3.5.   WFDC2 glycosylation analysis

Recombinant human WFDC2 resolved as two separate bands when analysed by Western blotting which is suggestive of glycosylated and non-glycosylated protein isoforms. To identify possible glycosylation sites, the peptide sequence of human WFDC2 (as described in the Ensembl database) was inputted into NetNGlyc and NetOGlyc glycan predictor software. Asparagine-44 was identified as a potential N-glycosylation site, as well as Threonine-33 as an O-glycosylation site – both glycosylated residues reside in exon 2 (see Appendix).

To confirm whether the ~24 kDa band of human WFDC2 represented glycosylation at Asn-44, site-directed mutagenesis (SDM) was carried out using pVR1255 plasmid with human WFDC2 gene insert as the target sequence to substitute asparagine-44 for a glutamine residue. The SDM modified plasmid was isolated, sequenced and then subsequently transfected into mammalian cells. Conditioned media containing mutated WFDC2 was collected and analysed by Western blotting.

The results show that SDM-modified WFDC2 resolves at the same molecular weight as the smaller ~18 kDa band of wild-type (WT) WFDC2 (Fig 3.10). WT WFDC2 digested with a glycosidase enzyme, PNGase F, to remove all N-glycans also resolves at a molecular weight of ~18 kDa with no upper 24 kDa visible. SDM-treated WFDC2 resolves at the same molecular weight as the glycosidase-treated control.
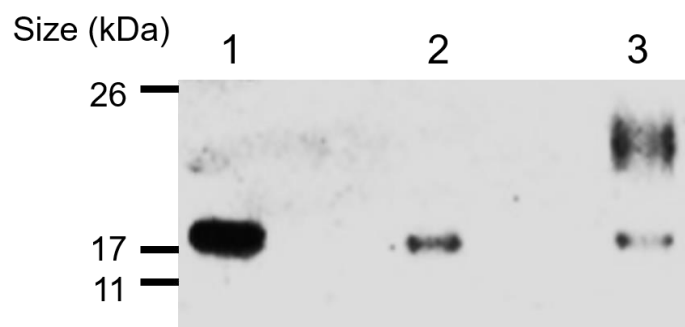
Figure 3.10. Western blot of SDM-modified human WFDC2 protein compared to wild-type protein. SDM-modified WFDC2 compared to wild-type WFDC2 analysed by Western blotting using anti-FLAG monoclonal antibody. WFDC2 digested with PNGase F used as a control. Lane numbers represent the following: 1 – human WFDC2 digested with PNGase F, 2 – SDM-modified human WFDC2, 3 – wild-type human WFDC2 (N=2).

To further analyse the glycan structure of human WFDC2, the recombinant protein was digested with sialidase enzyme to remove any sialic acid residues present. The digested product was analysed by Western blotting. The results show that incubation of WFDC2 with sialidase induces a shift in molecular weight compared to the untreated control (Fig 3.11).
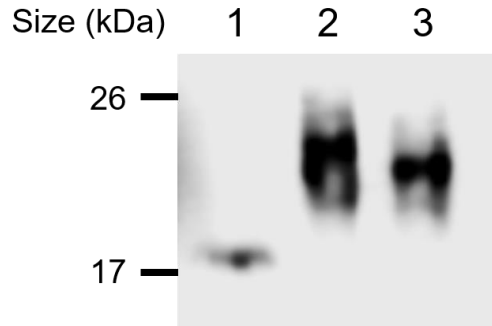
Figure 3.11. Western blot of human WFDC2 protein digested with sialidase enzyme compared to untreated control. Human WFDC2 digested with sialidase enzyme compared to untreated WFDC2 analysed by Western blotting using anti-FLAG polyclonal antibody. Lane numbers represent the following: 1 –untreated SDM-modified WFDC2, 2 – untreated wild-type WFDC2, 3 – wild-type WFDC2 digested with sialidase enzyme (N=3).

To determine whether the intracellular expression pattern of SDM-modified WFDC2 was affected compared to the WT protein, transfected NCI-H292 cells were analysed using immunofluorescence microscopy. SDM-modified WFDC2 was detected using anti-WFDC2 polyclonal antibody followed by a fluorescently labelled secondary antibody. Mock transfected cells were used as a negative control. No change in intracellular localisation was visualised between cells transfected with WT WFDC2 and those transfected with SDM-modified WFDC2 (Fig 3.12).
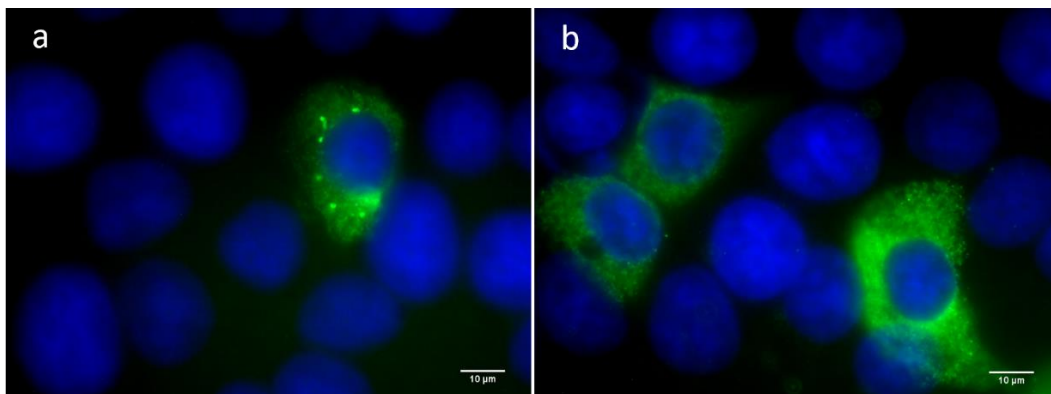
Figure 3.12. Immunofluorescence microscopy image of NCI-H292 cells transfected with human SDM-modified WFDC2. Cells were transfected with SDM-modified WFDC2 and then fixed and stained with anti-FLAG monoclonal antibody (Sigma) (a) or polyclonal antibody reactive against human WFDC2 (R&D systems) (b). Original magnification is 100X.

Recombinant murine Wfdc2 resolves as a single smeared band 26-34 kDa in size (see Fig 3.6). This is distinctive to human WFDC2, which forms two separate bands due to the presence of an N-glycosylation site. The peptide sequence of murine Wfdc2 was analysed using NetNGlyc and NetOGlyc online glycan predictor software which predicted that the protein has 17 O-glycosylation sites, 15 of which are found in exon 3. There are no predicted N-glycosylation sites. When murine Wfdc2 was incubated with sialidase and analysed by Western blotting there was a visible shift in molecular weight between the digested and non-digested forms (Fig 3.13).
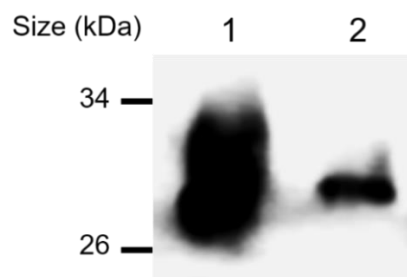
Figure 3.13. Western blot of murine Wfdc2 protein digested with sialidase enzyme compared to untreated control. Mouse Wfdc2 digested with sialidase enzyme compared to untreated Wfdc2 analysed by Western blotting using anti-FLAG monoclonal antibody. Lane numbers represent the following: 1 – untreated Wfdc2, 2 – Wfdc2 digested with sialidase enzyme (N=3).

3.6. Endogenous expression of human WFDC2

To assess whether WFDC2 is present in human saliva, 4 whole saliva samples from healthy non-smoking volunteers were analysed by Western blotting. Parotid gland saliva and submandibular/sublingual gland saliva from 1 healthy non-smoking volunteer were also analysed.

The results show that human WFDC2 is present in all 4 human saliva samples, existing as what appears to be a glycosylated isoform at a molecular weight of ~25 kDa (Fig 3.14). Some non-glycosylated WFDC2 also appears to be present in saliva sample 2 (Lane 2). The parotid and submandibular/sublingual saliva are negative for WFDC2 as no band is visible at the expected molecular weight (~25 kDa).

A second band is also present at a molecular weight of ~150 kDa in all samples except for the parotid saliva. This may be due to non-specific antibody binding to another large protein within the saliva. Alternatively, it could represent binding of WFDC2 to a large protein causing the protein band for WFDC2 to shift. Mucins are abundant in saliva and have a high molecular weight. Mucin 7 (MUC7), a mucin abundant in saliva, has a predicted molecular weight of 130-180 kDa. To determine whether the 150 kDa band in Fig 3.14 represented MUC7, the four WFDC2-positive whole saliva samples were also probed with anti-MUC7. The result shows that positive bands of ~150 kDa size are visible in blots probed with anti-MUC7 and anti-WFDC2 (Fig 3.15).
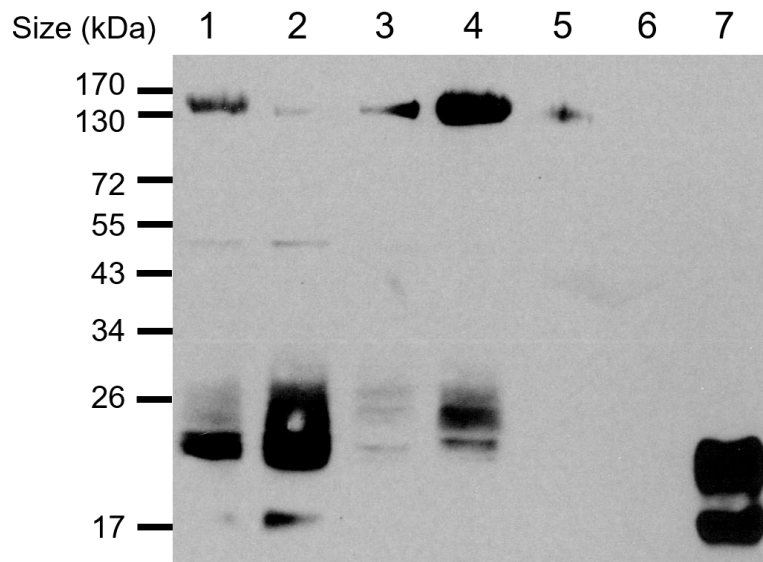
Figure 3.14. Western blot of WFDC2 expression in human saliva samples. Four whole saliva samples (1-4), a submandibular/sublingual saliva sample (5) and a parotid gland saliva sample (6) analysed by Western blotting against a control of recombinant human WFDC2 (7). Anti-WFDC2 polyclonal antibody used for detection of protein (N=2).
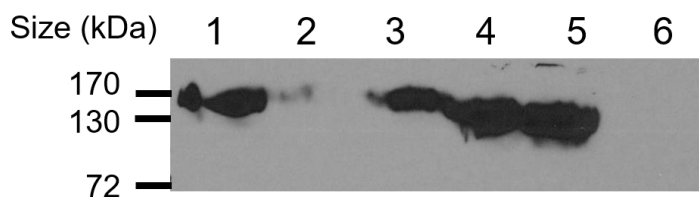
Figure 3.15. Western blot of MUC7 expression in human saliva samples. Four whole saliva samples (1-4), a submandibular/sublingual saliva sample (5) and a parotid gland saliva sample (6) analysed by Western blotting. Anti-MUC7 polyclonal antibody used for detection of protein (N=2).

To ensure that the ~25 kDa band in Western blots of saliva truly represented human WFDC2, a saliva sample was digested with PNGase F to remove all N-glycans.

The Western blot analysis shows that glycosidase-digested saliva resolves on the gel at the same molecular weight as the non-glycosylated isoform of recombinant WFDC2 (~18 kDa), while the untreated saliva sample resolves at a similar molecular weight to the N-glycosylated isoform of recombinant WFDC2 (Fig 3.16).
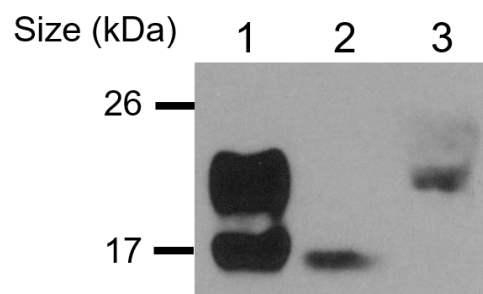
Figure 3.16. Western blot of human saliva digested with PNGase F compared to untreated saliva and recombinant WFDC2. Saliva digested with PNGase F was analysed by Western blotting using anti-WFDC2 polyclonal antibody and compared to an untreated saliva control and recombinant WFDC2. Lane numbers represent the following: 1 – recombinant human WFDC2, 2 – human saliva digested with PNGase F, 3 – untreated human saliva (N=2).

To quantify the endogenous salivary concentration of WFDC2, saliva samples from 21 healthy, non-smoking volunteers were analysed using a WFDC2 ELISA. The mean concentration of WFDC2 in whole saliva was 139.3 nM with a range of 62.9 - 338.2 nM (Fig 3.17). The sample size for saliva analysis was not large enough to identify any differences in WFDC2 levels based on the age or gender of volunteers.

WFDC2 concentration in secretions
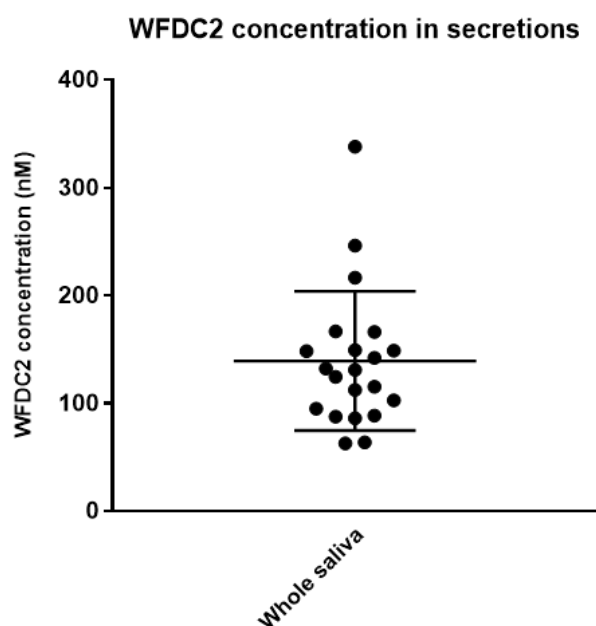
Figure 3.17. ELISA analysis of the concentration of WFDC2 in human saliva. ELISA analysis was used to indicate the endogenous expression levels of WFDC2 in saliva (n=21) derived from healthy volunteers. Samples were centrifuged to remove cells/debris and diluted according to the manufacturer's recommendation prior to ELISA analysis. The mean ± SD of each data set are shown.

To determine whether WFDC2 is secreted from healthy human lung cells, apical secretions from bronchial cells (derived from bronchial brushings) grown at an air-liquid interface (ALI) were analysed by Western blotting. Two apical secretions derived from cells of independent donors were analysed.

The results show that WFDC2 is present in secretions from bronchial cells grown at an ALI (Fig 3.18). Both samples appeared to be abundantly glycosylated as the protein resolved at a molecular weight of ~25 kDa, which is comparable with the larger N-glycosylated band in the recombinant control. To support this hypothesis, an apical secretion sample (Lane 2 in Fig 3.18) was digested using PNGase F to cleave away all N-glycans. The results show that the glycosidase-treated apical wash has the same molecular weight as non-glycosylated recombinant WFDC2 (~18 kDa) (Fig 3.19) which supports the theory that the positive bands accurately represent endogenous WFDC2 in bronchial secretions.
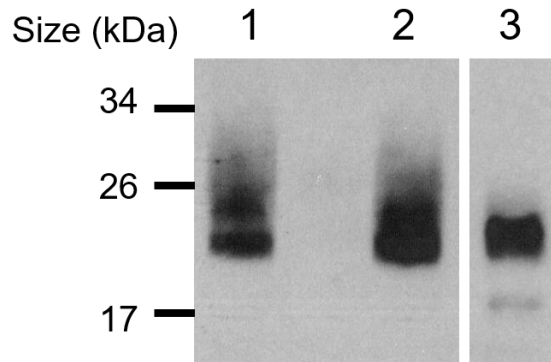
Figure 3.18. Western blot of WFDC2 expression in apical secretions from human bronchial cells. Two apical secretions from human bronchial cells grown at an air-liquid interface (1, 2) analysed by Western blotting against a control of recombinant human WFDC2 (3). Anti-WFDC2 polyclonal antibody used for detection of protein (N=2).
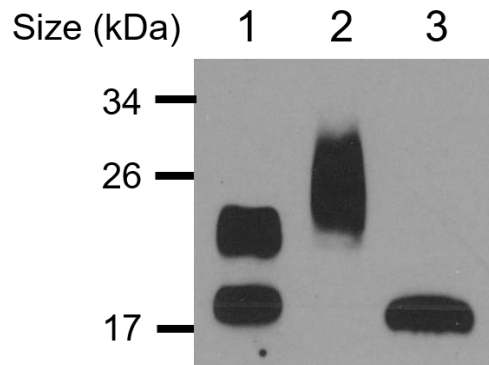
Figure 3.19. Western blot of apical secretions from human bronchial cells digested with PNGase F compared to untreated secretions and recombinant WFDC2. An apical secretion of bronchial cells digested with PNGase F was analysed by Western blotting using anti-WFDC2 polyclonal antibody and compared to an untreated control and recombinant WFDC2. Lane numbers represent the following: 1 – recombinant human WFDC2, 2 – untreated apical secretion, 3 – apical secretion digested with PNGase F (N=2).

To quantify the concentration of WFDC2 in apical secretions from human bronchial cells, secretions from cells derived from 11 different donors were analysed using a WFDC2 ELISA (n=11). The mean concentration of WFDC2 in bronchial secretions was 86.4 nM with a range of 82.3 - 92.1 nM (Fig 3.20).
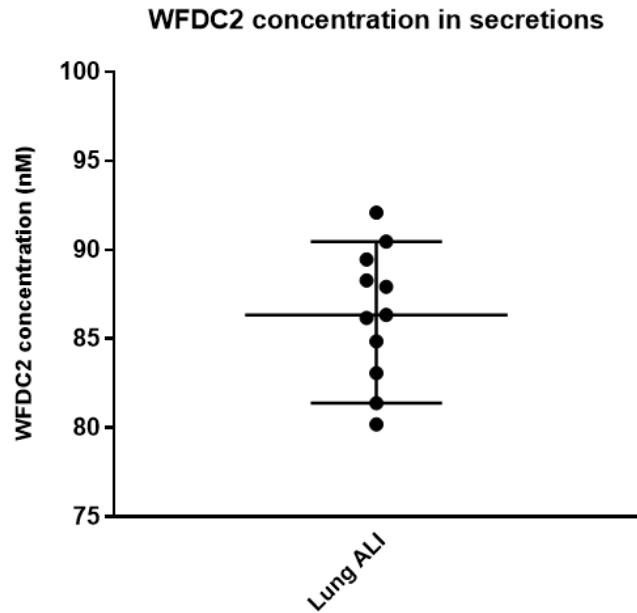
Figure 3.20. ELISA analysis of the concentration of WFDC2 in human lung ALI secretion. ELISA analysis was used to indicate the endogenous expression levels of WFDC2 in secretions from lung cells grown at an ALI (n=11). Samples were centrifuged to remove cells/debris and diluted according to the manufacturer's recommendation prior to ELISA analysis. The mean ± SD of each data set are shown.

3.7. Endogenous expression of murine Wfdc2

To determine whether Wfdc2 is secreted from murine epithelial cells of the trachea (mTEC) and nasal cavity (mNEC), apical secretions from primary cells grown at an ALI were analysed by Western blotting. The results show that Wfdc2 is present in secretions from mTEC and mNEC cells grown at an ALI (Fig 3.21). The endogenous protein resolved slightly higher than that of recombinant Wfdc2, with a molecular weight of ~28-40 kDa compared to the 26-34 kDa recombinant protein.
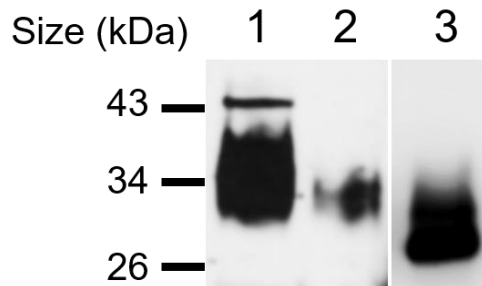
Figure 3.21. Western blot of apical secretions from murine tracheal and nasal epithelial cells compared to recombinant Wfdc2. Apical secretions of murine tracheal epithelial cells (1) and nasal epithelial cells (2) analysed by Western blotting using anti-Wfdc2 polyclonal antibody, compared to recombinant murine Wfdc2 (N=2).

3.8. Discussion

Recombinant proteins are an effective resource for analysing protein function *in vitro* and are extensively used in research. To ensure that desired target proteins are produced to a high standard, cloning and expression systems should be carefully adapted with the end-product in mind. During this study, when synthesising WFDC2 orthologues for *in vitro* functional assays, the key considerations were that the recombinant proteins were correctly glycosylated and were also tagged with a marker to enable downstream purification.

The vector utilised for WFDC2 production encoded a HCMV promoter and intron A upstream of the target gene insertion site to stimulate a high level of transcription (Chapman et al., 1991). A FLAG tag was incorporated into the reverse primer used for amplification of WFDC2 so that the tag was inserted immediately upstream of the stop codon. The inclusion of a FLAG tag, an eight-residue peptide (DYKDDDDK), was important for the downstream purification of WFDC2, as monoclonal antibodies raised against the FLAG epitope can be used for affinity purification using commercially available resins (Einhauer and Jungbauer, 2001).

The most appropriate gene target for cloning of human WFDC2 was splice variant 0 (V0), often described as 'full-length' (FL) WFDC2. V0 is the only isoform that encodes both WFDC protein domains. Considering the potential for WFDC domains to establish distinct functions

within a single protein, as seen in SLPI (described in Section 1.3.2), it was important to produce a recombinant protein expressing both domains.

Most of the expression analysis available in the literature utilises probes/antibodies that target FL WFDC2; consequently, very little is known about the expression profile of isoforms V1-V4 (Bingle et al., 2002; Bingle et al., 2006; Drapkin et al., 2005; Kirchhoff et al., 1991). Preliminary RT-qPCR analysis using isoform-specific probes has shown that V0 is by far the most abundant isoform in the epididymis, breast, trachea and endometrium (Jiang et al., 2013). It is therefore likely that the lack of expression data pertaining to V1-V4 is due to their relative absence in comparison to V0. This data further supported the decision to discount V1-V4 as cloning targets.

In the case of murine Wfdc2, very little is known about the expression of any of the splice variants. The limited data available in the literature and in unpublished data (Bingle, unpublished; Su et al., 2002; Kappler-Hanno and Kirchhoff, 2003) suggests that the FL murine Wfdc2 transcript (V0), encoding exons 1-5, is the most abundantly expressed variant (Section 1.2.4). The fact that the rodent orthologue of Wfdc2 appears to have evolved an extra exon which serves to separate the two WFDC domains is interesting and could result in the protein exhibiting distinct functions to the human orthologue. Consequently, FL Wfdc2 was also targeted for protein synthesis.

Following the successful insertion of both WFDC2 orthologues into vectors, the plasmids were utilised for transfection experiments. WFDC2 was initially transiently transfected into mammalian cells (HEK293) in a 24-well plate using two methods: FuGENE (a non-liposomal, lipid-based reagent) and calcium phosphate (a chemical method based on formation of a co-precipitate between calcium phosphate and DNA). Western blot analysis showed that both transfection methods effectively induced expression of WFDC2 in the cell lysate. The band intensity was greater in lysate from cells transfected using FuGENE, suggesting that this method has a greater transfection efficiency and results in a greater protein yield (Fig 3.3). Nonetheless, the calcium phosphate method was also able to successfully induce production

of WFDC2 protein. Given the large yield of protein required for the intended downstream *in vitro* assays, the calcium phosphate method was taken forward as the chosen transfection method as it has been optimised for large scale experiments (Jordan et al., 1998) and is more cost-effective. The calcium phosphate transfection protocol was scaled-up for WFDC2 production in T75 flasks. Co-transfection of the WFDC2 plasmid alongside a GFP-expressing vector gave a visual indication of the success/efficiency of the DNA delivery.

The transient nature of the transfection method meant that WFDC2 secretion was finite, therefore an appropriate time-frame for protein collection was required. Conditioned media from the transfected cells was collected for 4 days and then analysed by Western blotting. The results demonstrated that human and murine WFDC2 secretion continued for 4 days of collection, although the band intensity was beginning to decrease by days 3-4 (Fig 3.5). By collecting the media and replacing it with fresh serum-free media, cells were maintained in nutrient-rich conditions and the protein was better protected from degradation by host-derived proteases due to daily harvesting. Serum-free media reduced the number of contaminating, undefined proteins in the conditioned media.

HEK293 cells were chosen as the target cell line for transfection as mammalian cells are able to synthesise proteins with similar molecular structures to endogenous human/mouse proteins. The cells can produce complex glycosylated proteins, an important factor in the case of WFDC2 which has been shown to exhibit a complex glycan profile (Hua et al., 2014). Glycosylation can influence factors such as protein yield, clearance and stability, in addition to bioreactivity and immunogenicity (Moremen et al., 2012). HEK293 cells are used commercially to synthesise biotherapeutic proteins which demonstrates their suitability for production of recombinant peptides (Dumont et al., 2016).

Recombinant WFDC2 was purified from the conditioned media of the transfected HEK293 cells using an affinity resin with immobilised monoclonal antibody which binds to FLAG-tagged fusion proteins. The purified products were analysed by Western blotting using an anti-FLAG

monoclonal antibody for detection, and the results show a significant enrichment of WFDC2 when compared to the original conditioned media (Fig 3.6).

When the purified products were analysed using polyclonal antibodies raised against human and murine Wfdc2, a positive band of an identical molecular weight to that detected by anti-FLAG antibody was visualised, supporting the identity of the recombinant protein.

The molecular signature of Western blotted recombinant human WFDC2 showed two distinct bands, 18 kDa and 24 kDa in size, respectively (Fig 3.6c). The expected molecular weight of the protein is 13 kDa, however this discrepancy in size has already been reported in the literature (Drapkin et al., 2005; Hellström et al., 2003; Hua et al., 2014). It is hypothesised that the 24 kDa band results from the addition of N-glycans, predicted to be present at asparagine-44. Human WFDC2 also has a predicted O-glycosylation site at threonine-33 and the presence of the FLAG peptide (1 kDa) would also increase the protein molecular weight above the predicted size.

To confirm whether the 24 kDa band of human WFDC2 represented the presence of N-glycans, the protein was subjected to digestion by PNGase F, an amidase enzyme capable of removing N-linked sugars from glycoproteins. Analysis of the cleaved product by Western blotting showed complete removal of the 24 kDa band, suggesting that it did indeed correspond to N-glycosylation. Although PNGase F treatment of WFDC2 has been carried out previously showing similar results (Drapkin et al., 2005), there is no finite evidence that the site of N-glycosylation is Asn-44, as is predicted using glycan predictor software. To confirm this theory, the human WFDC2 gene construct was targeted for site-directed mutagenesis (SDM), whereby the Asn-44 residue was converted to a Gln-44 amino acid. The substitution from one polar residue to another means it is unlikely that the protein folding of WFDC2 would be affected, while removal of the Asn residue prevents the post-translational addition of N-linked sugars. When SDM-modified WFDC2 was analysed by Western blotting, the 24 kDa band was absent, thus showing that the upper band resulted from N-glycan linkages at Asn-44 which add ~5 kDa to the molecular weight (Fig 3.10).

When PNGase F catalyses cleavage of N-glycans, the sugars that are cleaved away remain largely intact and this enables them to be further analysed. Analysis by Hua et al (2014) using mass spectrometry demonstrated that the N-linked sugars of recombinant human WFDC2 consisted of complex biantennary structures with sialic acid residues present at the terminus of the sugar chains. To show that the protein produced in this study was correctly post-translationally modified during synthesis, it was incubated with a sialidase enzyme (cloned from the *Tannerella forsythia* gene) which catalyses cleavage of sialic acid residues from the terminus of glycan chains. Western blot analysis demonstrated that the molecular weight of the protein was reduced following incubation with sialidase compared to the untreated control (Fig 3.11), suggesting that the protein is modified with complex oligosaccharides terminating with sialic acids, thus resembling its native structure.

To further assess the quality of the protein, Western blot analysis was used to compare the recombinant protein with endogenous WFDC2. When whole saliva samples from healthy volunteers were analysed by Western blotting using an anti-WFDC2 polyclonal antibody, positive bands were present with a molecular weight of 24-32 kDa (Fig 3.14). The slightly higher molecular weight of salivary WFDC2 was hypothesised to represent differences in glycosylation, whereby endogenous WFDC2 exhibits more elaborate glycan structures than the recombinant protein. Similarly, when apical secretions derived from human bronchial cells cultured at an ALI were analysed by Western blotting, the same result was observed: positive bands were 24-30 kDa in size (Fig 3.16).

To ensure that the ~30 kDa bands in saliva and ALI secretions represented WFDC2, the samples were digested with PNGase F and subjected to Western blot. The digested products resolved at the same molecular weight as the non-glycosylated (19 kDa) band of recombinant WFDC2 showing that the positive bands in saliva and ALI secretions did indeed represent WFDC2, albeit with a slightly more complex glycan profile.

Interestingly, the saliva samples analysed by Western blotting were positive for WFDC2 at the expected molecular weight (~24 kDa) but also expressed an additional band that resolved at

a molecular weight of ~150 kDa (Fig 3.14). In contrast, when submandibular/sublingual (SMD/SL) and parotid secretions were analysed, a 24 kDa band was absent, but the SMD/SL secretions produced a positive band at 150 kDa. Given the abundance of mucins in saliva and their corresponding extensive molecular weight, it was hypothesised that the large band represented mucin 7, a secretory mucin abundant in saliva with a known molecular weight of ~150 kDa. This theory was supported by the presence of the large band in SMD/SL secretions, which have a serous and mucinous component, whereas parotid serous secretions failed to express the large band. By probing the samples with an anti-MUC7 polyclonal antibody, a similar positive 150 kDa band was visible in all samples, aside from the parotid secretion which was negative for MUC7 (Fig 3.15).

It remains unclear whether the 150 kDa band in the WFDC2-probed blot resulted from WFDC2 binding to MUC7 or whether the anti-WFDC2 antibody simply bound non-specifically to MUC7 itself – further analysis would be required to elucidate any interactions between the two proteins. It should be noted, however, that there is evidence in the literature suggesting that WFDC2 forms interactions with mucins. During proteomic analysis of apical secretions from tracheobronchial epithelial (TBE) cells and induced sputum from healthy donors (Kesimer et al., 2009; Radicioni et al., 2016) WFDC2 was identified within the mucin-rich pool of proteins rather than in the protein-rich fraction of lower molecular weight proteins. Mucins are extensively glycosylated and have the capacity to bind to smaller glycosylated proteins even following treatment with strong denaturants (Radicioni et al., 2016). Other low molecular weight proteins with putative host defence functions were detected in the high-density fraction, including SLPI and BPI-fold containing family protein B1 (BPIFB1) (Radicioni et al., 2016). BPIFB1 is already known to co-localise with mucin 5B (MUC5B), which perhaps explains its presence in the high-density pool (Bingle et al., 2013; Bingle et al., 2012). It is hypothesised that these proteins associate with mucins to form an 'interactome', consequently altering the properties of the mucosal surface and enhancing the defensive properties of the mucosa

(Radicioni et al., 2016). It is feasible that WFDC2 interacts with mucins in order to contribute to mucosal properties.

When recombinant murine Wfdc2 was analysed using the same Western blotting and enzyme digestion techniques, it resolved as a smeared band ranging from 26 kDa to ~34 kDa. The predicted molecular weight of murine Wfdc2 is 18 kDa, however this discrepancy can again be attributed to the presence of an abundance of glycans, identified using glycan predictor software. Wfdc2 has 17 predicted O-glycosylation sites, 15 of which are found within the exon 3 encoded linker region found in rodents (see Appendix). No N-glycosylation sites were predicted. To confirm that the smeared band did indeed result from the presence of a multitude of sugars, the protein was subjected to sialidase digestion and was analysed by immunoblot. The digested protein showed a significant shift in molecular weight compared to the undigested control (Fig 3.13), suggesting that the recombinant protein is heavily glycosylated with complex O-glycans and that sialic acid residues are present at the termini of the sugar chains.

Denatured murine Wfdc2 also shows a secondary band at ~50 kDa when analysed by Western blotting. Given that the molecular weight of murine Wfdc2 can be as low as ~24 kDa (presumably when in its deglycosylated form), it is possible that the protein is able to form a dimer. The recombinant protein was thus resolved on a native PAGE gel, which permits analysis of proteins in their natural conformation, avoiding disruption of protein complexes. The resultant immunoblot showed that murine Wfdc2 resolved at ~68 kDa, suggesting that the protein was capable of forming a dimer. Further analysis is required to determine whether this theory is correct. Human WFDC2 resolves on a native gel at ~26 kDa, suggesting that it does not form any protein complexes. Given that both proteins are extensively glycosylated, it seems unlikely that murine Wfdc2 resolves at a higher molecular weight due to the presence of oligosaccharides, as a similar result would be expected with human WFDC2. If murine Wfdc2 does indeed form a dimer, it is possible that the exon 3 linker is involved in the interaction, as this is the key difference between the structure of the two orthologues.

To more closely understand whether the recombinant murine protein resembles endogenous Wfdc2, Western blot analysis was performed using apical secretions derived from murine tracheal and nasal epithelial cells grown at an ALI. The immunoblot results showed that endogenous Wfdc2 resolved at a slightly higher molecular weight than recombinant protein, as was seen with human WFDC2. This analysis shows that recombinant proteins produced by HEK293 cells formed complex glycans with terminal sialic acids, however the glycan profile of endogenous proteins remained more extensive.

The aim of synthesising human and murine WFDC2 protein was to use the recombinant protein in *in vitro* antimicrobial and anti-protease functional assays. In order to make the assays as biologically relevant as possible, it was important to identify a suitable concentration of protein to use. The literature is devoid of any quantitative analysis of WFDC2 levels at mucosal sites, aside from a study which tested the WFDC2 concentration in vaginal secretions from 18 healthy women. The ELISA analysis described a median WFDC2 concentration of 108 nM (and a range of 2.6 to 429 nM). All remaining quantitative assessments of WFDC2 are limited to serum levels in diseased patients, whereby WFDC2 serum concentrations tend to exist within the picomolar range.

To enhance the understanding of WFDC2 expression in human mucosal secretions, ELISA analysis was used to quantify WFDC2 at sites hypothesised to express the protein. For instance, IHC has demonstrated that WFDC2 is abundantly produced in human salivary glands, where it is largely localised to the ducts (Fig 1.8). This ductular expression suggests that WFDC2 is secreted from salivary glands and is subsequently present in saliva – a theory that has been confirmed using Western blotting within this study (Fig 3.14). When analysed by ELISA, saliva samples contained a mean WFDC2 concentration of 139 nM, a level that is comparable to the concentration of SLPI detected in saliva as described in the literature (Lin et al., 2004; Shugars, 1999)

IHC has also shown that WFDC2 is expressed in the airway epithelium and in submucosal glands, suggesting that the protein is secreted at the mucosal surface of the airways. In ALI

secretions from lung cells, a mean concentration of 86 nM of WFDC2 was identified. Quantification of WFDC2 in the airways and salivary glands was important for determining a biologically relevant concentration of recombinant protein to use in the downstream functional assays.

Chapter 4

Investigating the putative host defence functions of WFDC2

4.1. Introduction

The antimicrobial and anti-protease activities of recombinant SLPI and elafin have been studied extensively in the literature using standard *in vitro* assays (Fernie-King et al., 2002; Gomez et al., 2009; Hiemstra et al., 1996; Si-Tahar et al., 2000; Wiedow et al., 1998). In comparison, WFDC2 has received very little attention, with only four functional analyses reported. The evidence so far suggests that neither recombinant human or murine WFDC2 are incapable of exhibiting antibacterial activity, however, further analysis is required to substantiate this theory (Hua et al., 2014; Nishimura et al., 2008). The anti-protease activity of recombinant human WFDC2 has undergone limited analysis, which appears to show that the protein is capable of protease inhibition against serine proteases and matrix metalloproteases (LeBleu et al., 2013; Chhikara et al., 2012). However, the concentration of protein required to elicit such effects differs widely between studies and often reaches high micromolar levels that are not physiologically relevant. Consequently, the assays described in this chapter will be carried out using levels of WFDC2 that are considered to be biologically relevant. Finally, two reports have suggested that WFDC2 is not significantly upregulated in response to pro-inflammatory mediators, as shown by RT-PCR (Bingle et al., 2006; Clauss et al., 2010). To verify these results, primary human salivary gland cells will be exposed to pro-inflammatory cytokines in the same way but changes in gene expression will instead be quantified using RT-qPCR, to improve the sensitivity of the assay.

4.2.   Aims and objectives

The aim of the work presented in this chapter was to identify whether WFDC2 is capable of significant antimicrobial and/or anti-protease activity, as hypothesised in the literature. Recombinant human and murine WFDC2 orthologues were used for *in vitro* assays to assess bacterial binding, killing and agglutination activity, as well as the capacity to inhibit biofilm formation or disrupt formed biofilms. The serine protease inhibitory activities of WFDC2 orthologues were assessed using chromogenic substrates, while the effect of WFDC2 on MMP activity was measured by zymography. Additionally, RT-qPCR was used to quantify changes in WFDC2 expression in human primary cells when exposed to pro-inflammatory mediators.

4.3.   Materials and methods

To access the relevant materials and methods utilised in this chapter, please refer to the sections below:

- Protease inhibition assays were optimised as per Section 2.6. Serine protease assays and zymography analysis were performed as described in Section 2.6.2 and 2.6.3, respectively.

- Bacteria were confirmed to be the expected strains as described in Section 2.7.1 via colony PCR and 16S sequencing (Section 2.7.3). Culture methods are defined in Section 2.7.2. Bacterial assays, including viability, binding, agglutination and biofilm assays are described in Section 2.7.4 to 2.7.11.

- Culture conditions for propagation of primary salivary gland cells are defined in Section 2.1.1. Cytokine exposure assays were performed as described in Section 2.8. Changes in gene expression were quantified using RT-qPCR, as outlined in Section 2.2.5.

4.4. Analysis of protease inhibitory activity

To establish whether WFDC2 is capable of inhibiting serine proteases, both WFDC2 orthologues were incubated with trypsin and elastase. The effect on enzyme activity was monitored by measuring chromogenic substrate digestion via the subsequent accumulation of cleaved p-nitroanilide (p-NA).

Firstly, trypsin and elastase were incubated with substrate alone (without WFDC2 or inhibitor controls) for 15 minutes at 37°C to ensure that the enzymes were active against their chosen chromogenic substrate. Different concentrations of enzyme (as used in Erlanger et al., 1961 and Gauthier et al., 1982) were tested against a fixed concentration of substrate to identify suitable reaction parameters. Enzyme concentrations that induced a change in absorbance of ~0.3 were chosen for downstream protease inhibition experiments to allow scope for visualising any reduction in activity and so that the maximum substrate cleavage was a similar value for both trypsin and elastase experiments (Fig 4.1).
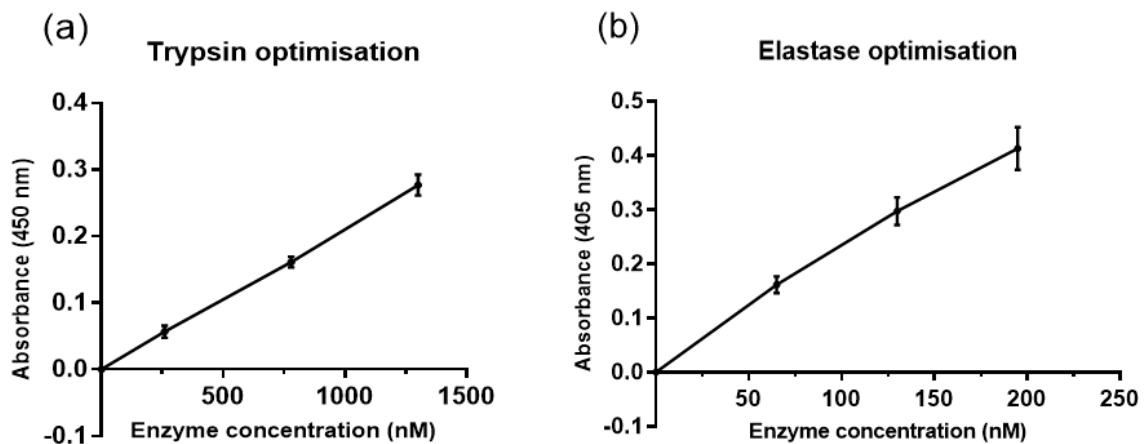
Figure 4.1. Optimisation of enzyme concentration for use in protease inhibition assays. The rate of chromogenic substrate hydrolysis following exposure to increasing concentrations of trypsin (a) and elastase (b) was assessed by measuring the absorbance at 405 nm. The mean absorbance value of substrate alone (without enzyme) was subtracted from all other readings to account for background. Data represents the mean ± SD (n=3).

To assess the effect of WFDC2 on enzyme activity, WFDC2 protein (0, 100, 200, 500 and 1000 nM) was pre-incubated with a fixed concentration of trypsin or elastase for 15 minutes to allow for binding between the potential inhibitor and enzyme substrate. Chromogenic substrate was added and the reaction was incubated for 15 minutes at 37°C. The reaction was quenched and the absorbance measured at 405 nm. Recombinant human SLPI (0, 100, 200, 500 and 1000 nM) and a broad-spectrum protease inhibitor (PI) solution (added in sequentially increasing volumes) were used as controls for the assay.

The results show that neither human or murine WFDC2 has any significant effect on trypsin activity at concentrations as high as 1000 nM (Fig 4.2a) ($P=0.86$ and $P=0.2$, respectively). The positive controls caused a significant reduction in trypsin activity, with SLPI causing a 50% decrease in enzyme activity at 500 nM ($P=0.0003$). The broad-spectrum PI solution induced an 80% reduction in trypsin activity even at a dilution 20-fold lower than the working concentration recommended by the manufacturer for total protease inhibition (see data point in line with 100 nM) ($P=0.0001$). Increasing the concentration of PI solution further had no greater inhibitory effect on trypsin.

When incubated with elastase, the broad-spectrum PI solution appeared to cause a significant decrease in enzyme activity and, at the manufacturer's recommended working concentration, elastase activity decreased by 60% (see data point in line with 1000 nM) ($P=0.0001$). On the other hand, human WFDC2, murine Wfdc2 and SLPI all failed to inhibit the activity of elastase. Instead, all three WFDC proteins appeared to cause an increase in substrate digestion (by 21%, 28% and 13% respectively) compared to the no inhibitor control (Fig 4.2b).
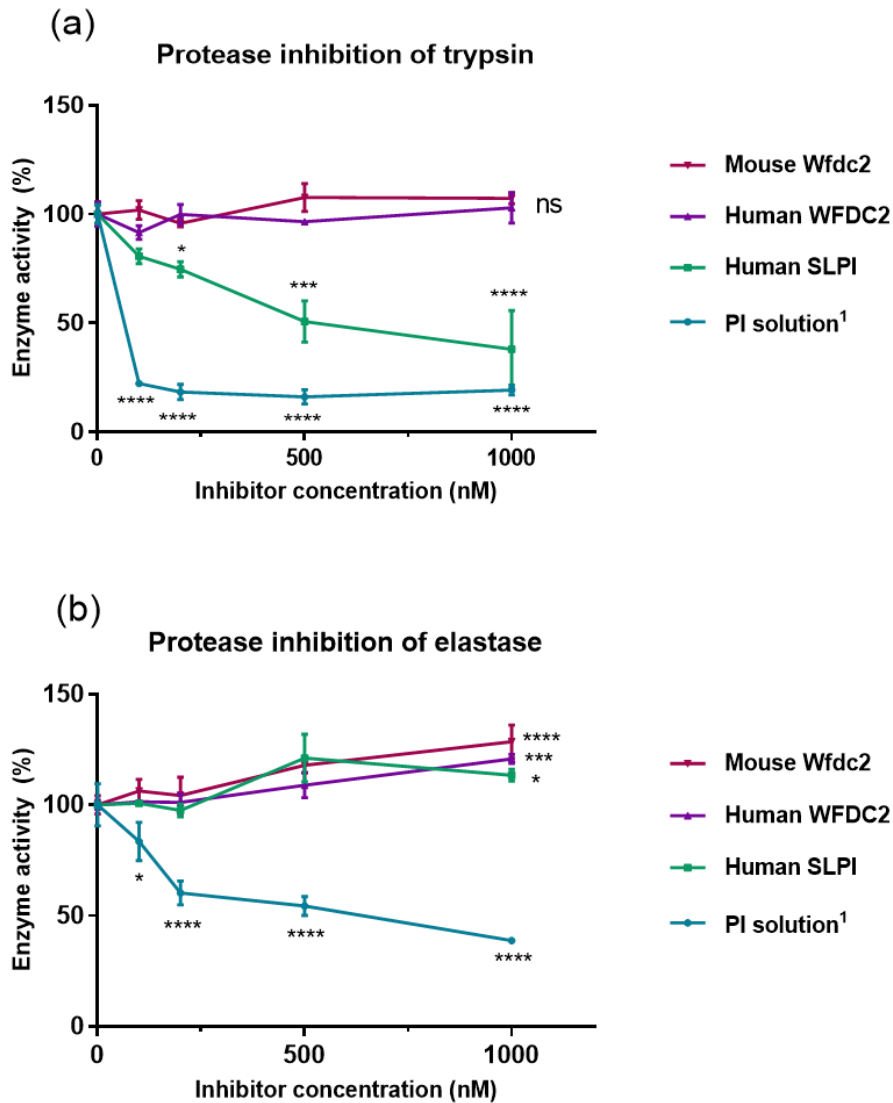
Figure 4.2. Serine protease inhibitory activity of human and mouse WFDC2. The effect of WFDC2 orthologues, SLPI and PI solution on trypsin (a) and elastase (b) activity were measured by monitoring changes in absorbance at 405 nm which represents chromogenic substrate hydrolysis. The mean absorbance of substrate alone was subtracted from all readings to account for background. Results are expressed as a percentage reduction in enzyme activity with reactions without inhibitor representing 100% enzyme activity. Data represents the mean ± SD (n=3). Significance value is calculated as activity compared to enzyme without inhibitor; ** P ≤ 0.01; *** ≤ 0.001; ≤ 0.0001. [1]The PI solution was of an unknown concentration but was added in increasing volumes. Data point in line with 1000 nM represents the working concentration of PI solution recommended by the manufacturer.

To determine whether WFDC2 orthologues are capable of gelatinase inhibition, increasing amounts of recombinant WFDC2 were incubated with fixed amounts of matrix metalloproteinases, MMP2 and MMP9, and subsequently analysed using zymography. MMP

activity was measured by analysing the bands of gelatin digestion which are visible following Coomassie staining of the gelatin gel.

The results show that neither human or mouse WFDC2 were capable of significantly inhibiting MMP2 or MMP9 activity, even when an enzyme-to-WFDC2 ratio of 1:50 was explored i.e. 500 ng of WFDC2 (Fig 4.3). No difference in band size resulting from gelatin digestion was visible at any WFDC2 concentration tested.
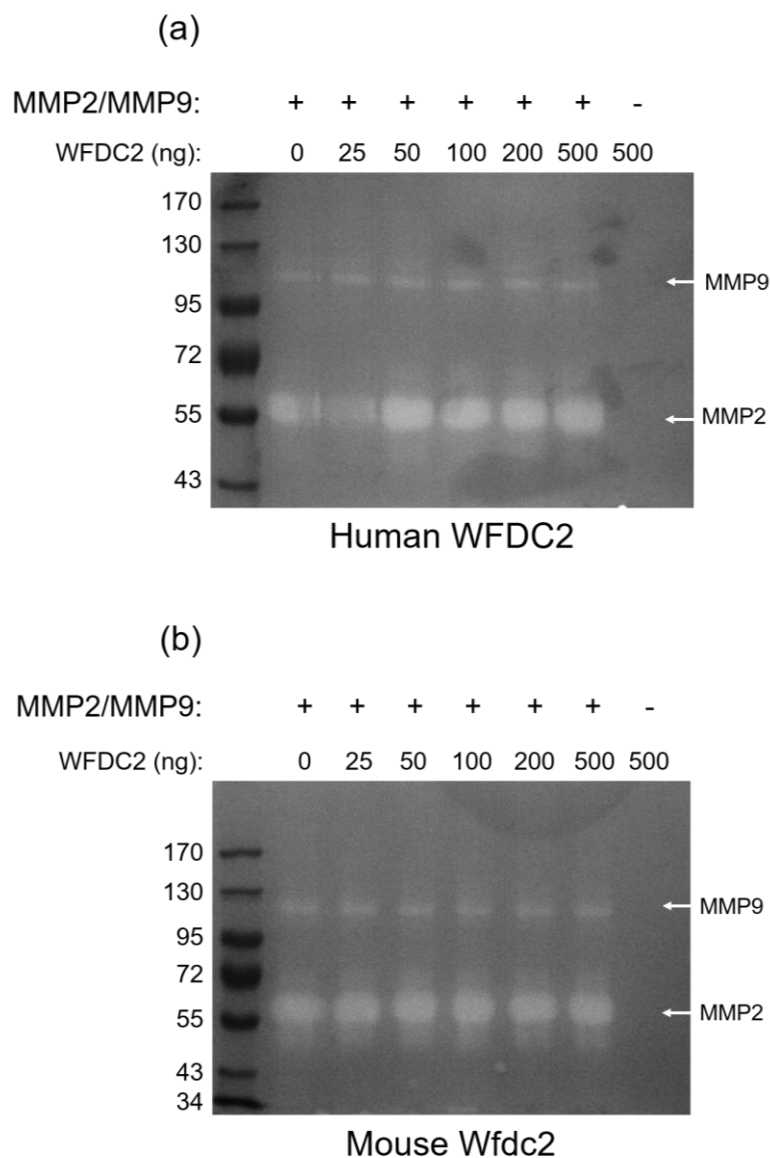
Figure 4.3. Gelatinase inhibitory activity of human and mouse WFDC2. The effect of increasing amounts of human (a) and mouse (b) WFDC2 on the activity of MMP2 and MMP9 were measured using zymography. Gelatin digestion is visible as clear bands in the Coomassie-stained gelatin gel. MMP digestion is indicated (white arrows). Representative images shown (N=3).

4.5.   Analysis of antimicrobial activity

The antimicrobial properties of human and murine WFDC2 were assessed against *Streptococcus mutans, Streptococcus gordonii, Streptococcus pneumoniae* and *Escherichia coli*. Each bacterial strain cultured was confirmed to be the expected species by colony PCR followed by 16S rDNA sequencing. In all assays, FLAG peptide elution solution was used as a control.

4.5.1. Binding assay

The capacity of WFDC2 to bind to bacteria was measured using traditional pull down assays. Purified human and murine WFDC2 proteins were incubated with bacteria for 2 hours at 37°C and then cells were washed extensively to remove unbound protein. The bacterial pellet was analysed by Western blotting to determine whether any protein had bound to the bacteria.

Human and murine WFDC2 appeared to be incapable of binding to the bacteria tested, as indicated by the absence of WFDC2 protein in the lysed bacterial pellet (Fig 4.4). Similarly, no binding was detected in the FLAG elution buffer control. The anti-FLAG monoclonal antibody appeared to bind non-specifically to *S. gordonii, S. mutans* and *S. pneumoniae* as bands of the incorrect size were present in lanes lacking WFDC2 protein.
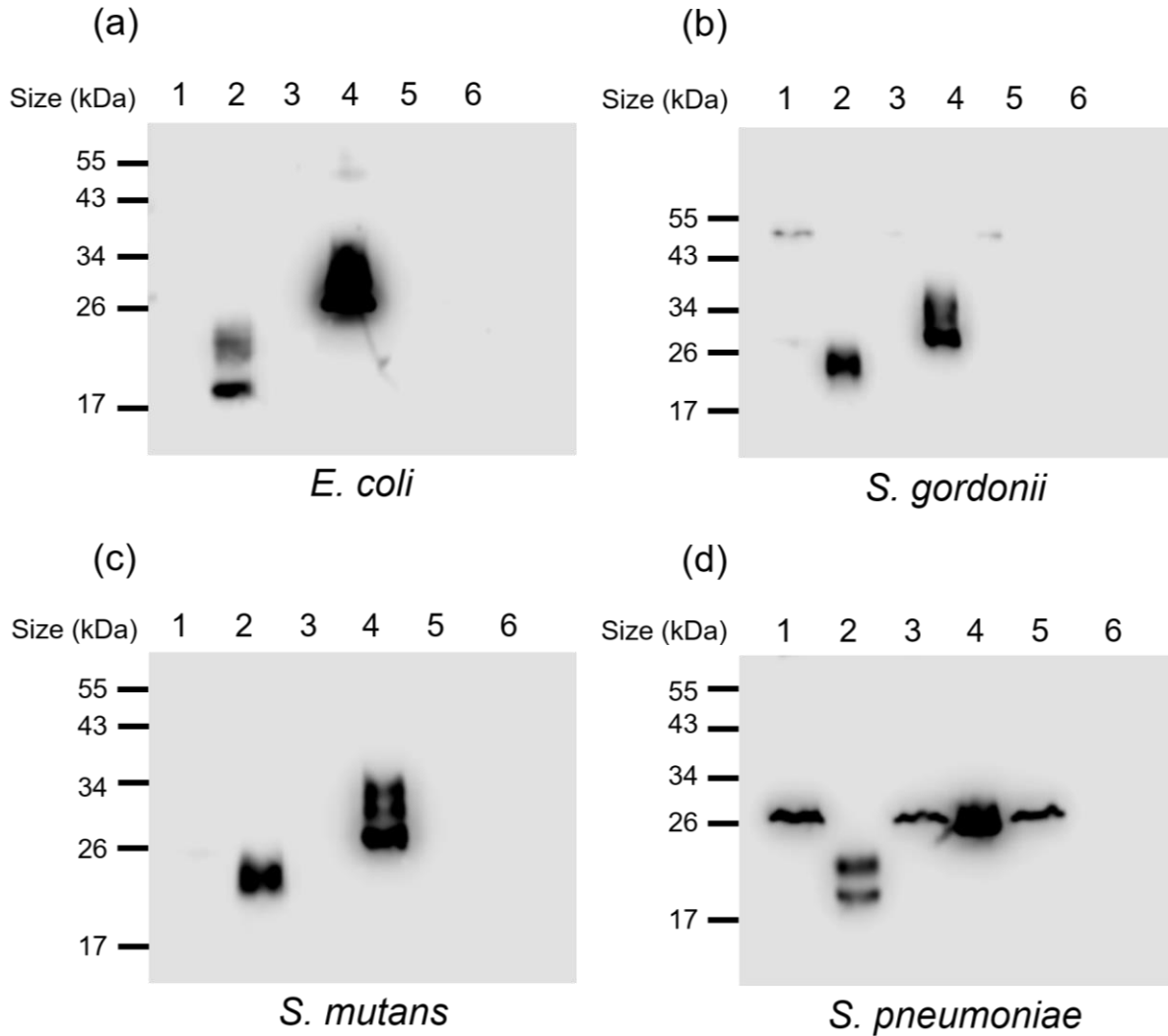
Figure 4.4. Binding capacity of WFDC2 orthologues to bacteria. The capacity of WFDC2 to bind to *E. coli* (a), *S. gordonii* (b), *S. mutans* (c) and *S. pneumoniae* (d) was measured by pull down assay and subsequent Western blotting using anti-FLAG monoclonal antibody. Lane numbers in each blot represent the following: 1 – bacterial pellet with bound human WFDC2, 2 – unbound human WFDC2, 3 – bacterial pellet with bound mouse Wfdc2, 4 – unbound mouse Wfdc2, 5 – bacterial pellet with bound FLAG elution buffer, 6 – unbound FLAG elution buffer. Representative images shown (N=3).

4.5.2. Viability assays

The effect of WFDC2 on bacterial viability was analysed using two methods. Firstly, a traditional method was used which involved incubating bacterial suspension with purified WFDC2 (0.5 and 1.0 µM). The bacterial suspension was incubated at 37°C for 3 hours and

then plated out via the Miles-Misra technique. The following day the number of colony forming units (CFU) in the original bacterial suspension were determined.

The results show that incubation of human or mouse WFDC2 with any of the four bacteria strains tested had no effect on the bacterial viability either at 0.5 µM (data not shown) or at 1 µM (Fig 4.5). There was no significant difference in the number of CFU per millilitre (CFU/ml) formed in the presence of WFDC2 compared to FLAG peptide or PBS negative controls.
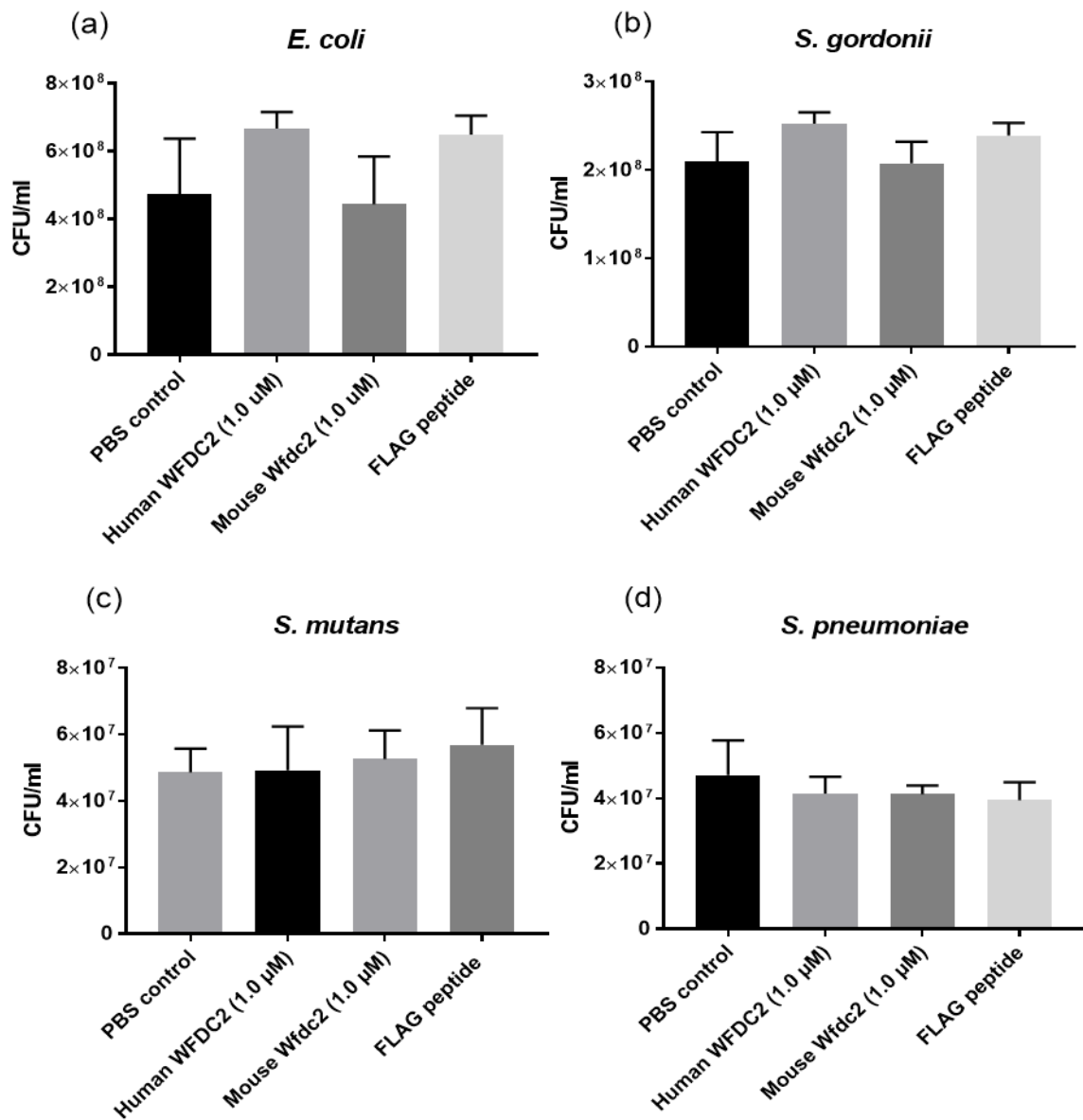
Figure 4.5. The effect of WFDC2 orthologues on bacterial viability analysed using the Miles-Misra technique. Bacterial suspensions of *E. coli* (a), *S. gordonii* (b), *S. mutans* (c) and *S. pneumoniae* (d) were incubated with WFDC2 (1 µM) or controls and plated out using the Miles-Misra technique to determine the number of CFU/ml formed in the presence of WFDC2. Data represents the mean ± SEM (N=3; n=3).

The effect of WFDC2 on bacterial viability was also monitored by measuring changes in absorbance of bacterial suspension over time to produce a growth curve. Briefly, bacterial culture was inoculated with WFDC2 protein (1.0 µM) and seeded into a 96 well plate. The optical density of each well was measured at 600 nm every 30 minutes for 12 hours using a microplate reader.

The results show that the addition of human or mouse WFDC2 to bacterial culture had no significant effect on bacterial growth when compared to untreated bacteria, or bacteria incubated with FLAG peptide elution buffer or PBS (Fig 4.6). Conversely, bacterial growth was significantly inhibited when cells were exposed to ampicillin (60 µg/ml). No change in absorbance was visible in wells loaded with sterile broth, showing that wells were not contaminated during incubation in the plate reader overnight.
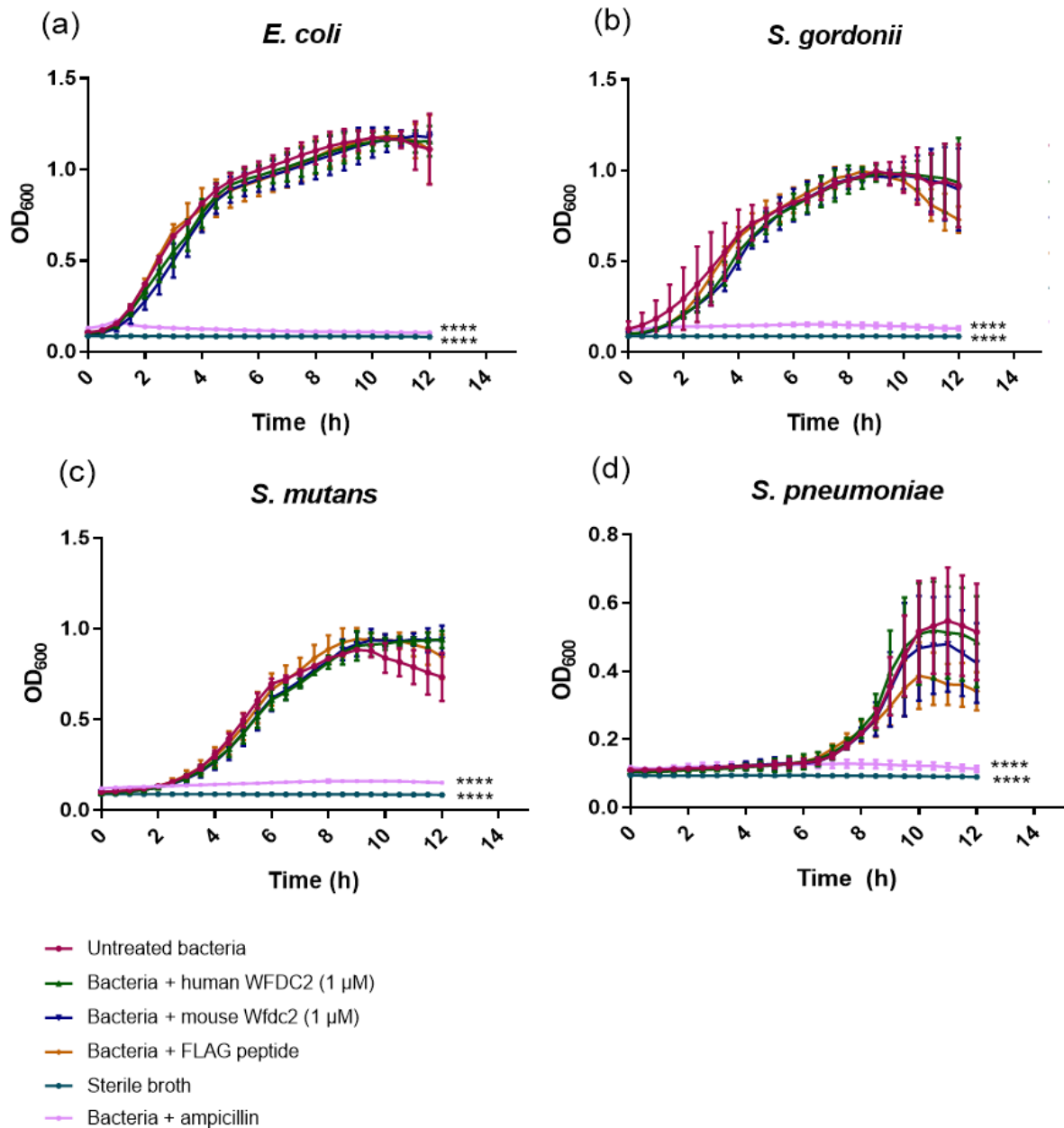
Figure 4.6. The effect of WFDC2 orthologues (1 μM) on bacterial viability analysed using the growth curve technique. Suspensions of *E. coli* (a), *S. gordonii* (b), *S. mutans* (c) and *S. pneumoniae* (d) were incubated with WFDC2 protein overnight. Changes in suspension optical density at OD_{600} were measured every 30 minutes for 12 hours using a plate reader to produce a growth curve. Data represents the mean ± SD (n=3); * P ≤ 0.0001.

To determine whether bacteria were exposed to insufficient WFDC2 to induce a reduction in bacterial viability, the assay was repeated against *S. gordonii* and *S. mutans* with a greater concentration of WFDC2 (3 µM). The same controls were used as before but volumes were adjusted to account for the increased volume of WFDC2 added.

The data shows that exposure of *S. gordonii* and *S. mutans* to a 3-fold higher concentration of the WFDC2 orthologues caused no change in bacterial growth (Fig 4.7) as no significant difference in bacterial viability was identified following exposure to WFDC2 or the negative controls. Incubation with ampicillin induced a significant reduction in bacterial viability.
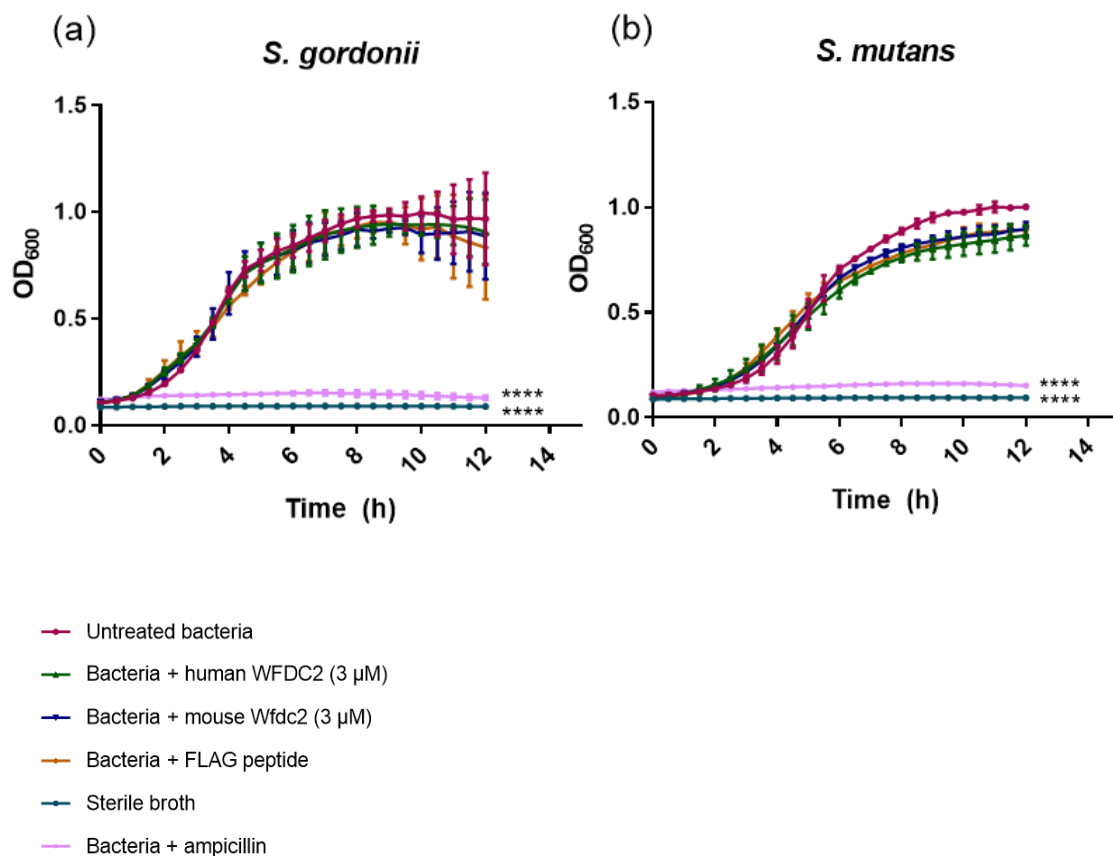
Figure 4.7. Analysis of the effect of WFDC2 orthologues (3 µM) on bacterial viability using the growth curve technique. The growth curve viability assay was repeated using *S. gordonii* (a) and *S. mutans* (b) suspensions which were incubated with a 3-fold higher WFDC2 concentration (3 µM). Data represents the mean ± SD (n=3). * P ≤ 0.0001.

To determine whether recombinant WFDC2 was digested by bacteria during the 12-hour growth curve incubation, the protein was incubated with bacteria overnight at 37°C and then analysed by Western blotting to determine whether any obvious digestion had occurred.

Overnight incubation with bacteria and WFDC2 appeared to induce some protein degradation (Fig 4.8). Glycan cleavage was identifiable in human WFDC2 incubated with *S. gordonii* and *S. pneumoniae* as evidenced by the reduction in band intensity of the upper glycosylated band of human WFDC2 (Fig 4.8a). The glycans of murine Wfdc2 also appeared to be cleaved when incubated with *S. pneumoniae,* as visualised by a reduction of the molecular weight of the smeared murine Wfdc2 band (Fig 4.8b).

Protein degradation was analysed semi-quantitively using densitometry. The results show that the signal derived from human WFDC2 following incubation with *S. gordonii* and *S. mutans* was significantly reduced compared to WFDC2 that had been incubated with PBS only (Fig 4.8c). All other bacteria failed to significantly reduce the signal, suggesting that the majority of the WFDC2 protein remained intact. The densitometry results for murine Wfdc2 show that overnight incubation with all four strains fails to produce a significant reduction in signal compared to the PBS control, suggesting that the protein is not degraded by the bacteria (Fig 4.8d).
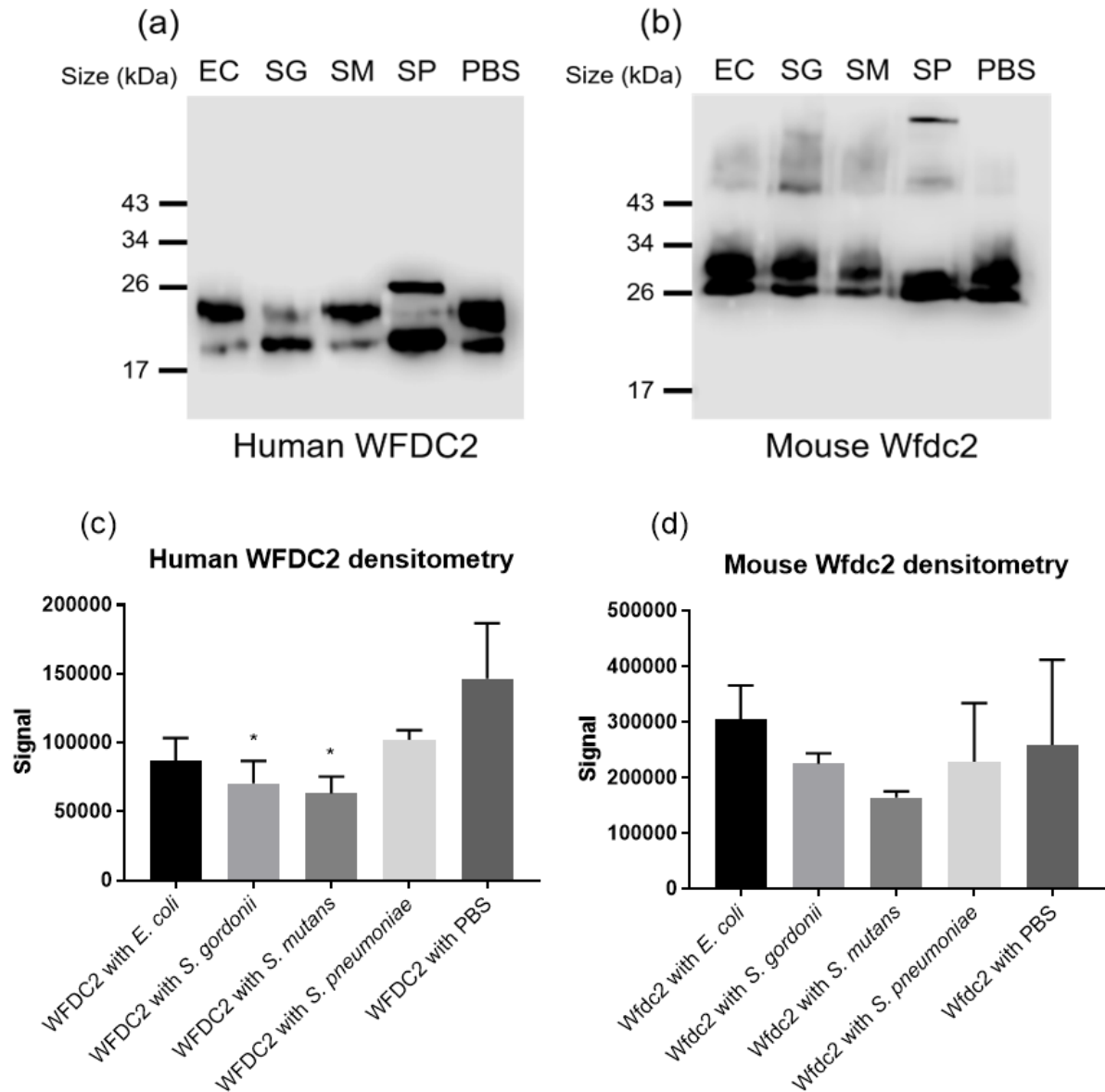
Figure 4.8. The degradative capacity of bacteria against WFDC2 proteins. Bacteria were incubated with both WFDC2 orthologues overnight at 37°C. The bacteria were pelleted and the human (a) and mouse (b) WFDC2 protein in the supernatant was analysed by Western blotting using anti-FLAG monoclonal antibody. Lane references represent the following: EC – WFDC2 incubated with *E. coli*, SG – WFDC2 incubated with *S. gordonii,* SM – WFDC2 incubated with *S. mutans,* SP – WFDC2 incubated with *S pneumoniae*, PBS – WFDC2 incubated with PBS (N=2; representative images shown). Densitometry values from Western blots of human (c) and mouse (d) WFDC2 are shown (mean ± SD, N=2); * P ≤ 0.05.

4.5.3. Agglutination assay

The ability of WFDC2 to induce bacterial agglutination was measured by incubating purified WFDC2 with fluorescein isothiocyanate (FITC) labelled bacteria. In brief, optimised numbers of bacteria (see Appendix) were labelled with FITC, washed and resuspended with WFDC2 (1 µM) overnight to test the capacity of WFDC2 to agglutinate bacteria. Saliva was used as a positive control for agglutination.

The results show that the WFDC2 orthologues are unable to induce agglutination in the bacterial strains tested, as determined by the 'pellet' of non-agglutinated cells visible in each well (Fig 4.9). The saliva positive control induced agglutination in all four strains tested as indicated by the film of bacteria and lack of bacterial pellet. The negative controls also failed to induce agglutination.
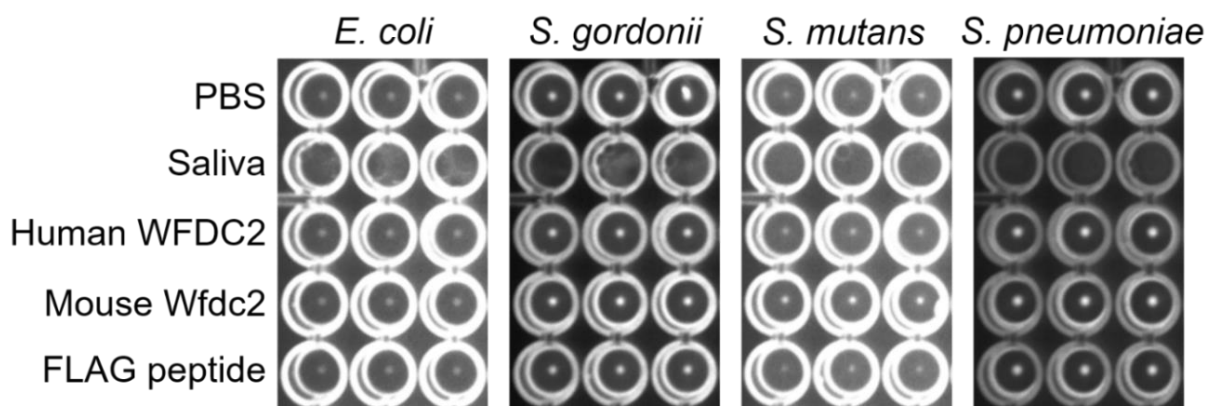
Figure 4.9. The agglutination capacity of WFDC2 orthologues. FITC-stained *E. coli, S. gordonii*, *S. mutans* and *S. pneumoniae* were incubated with human and mouse WFDC2 (1 µM) overnight to determine the capacity of WFDC2 to induce agglutination. Saliva was used as a positive control. FLAG peptide and PBS were used as negative controls (N=2, n=3; representative images shown).

4.5.4. Biofilm formation and disruption assays

The ability of WFDC2 to inhibit the formation of bacterial biofilms was tested by incubating recombinant WFDC2 (1 µM) with bacterial culture for 48 hours in a 96 well plate. The resultant biofilms were stained with 0.1% crystal violet and then solubilised in acetic acid, enabling the biofilms to be quantified using a microplate reader. Wells incubated with sterile broth were

stained and the absorbance reading was subtracted from all other values to account for background.

The results show that human and mouse WFDC2 are incapable of significantly inhibiting the formation of biofilms in all bacteria tested compared to control wells containing untreated bacteria (Fig 4.10). Similarly, the FLAG peptide elution solution, saliva and PBS controls had no significant effect on biofilm formation.
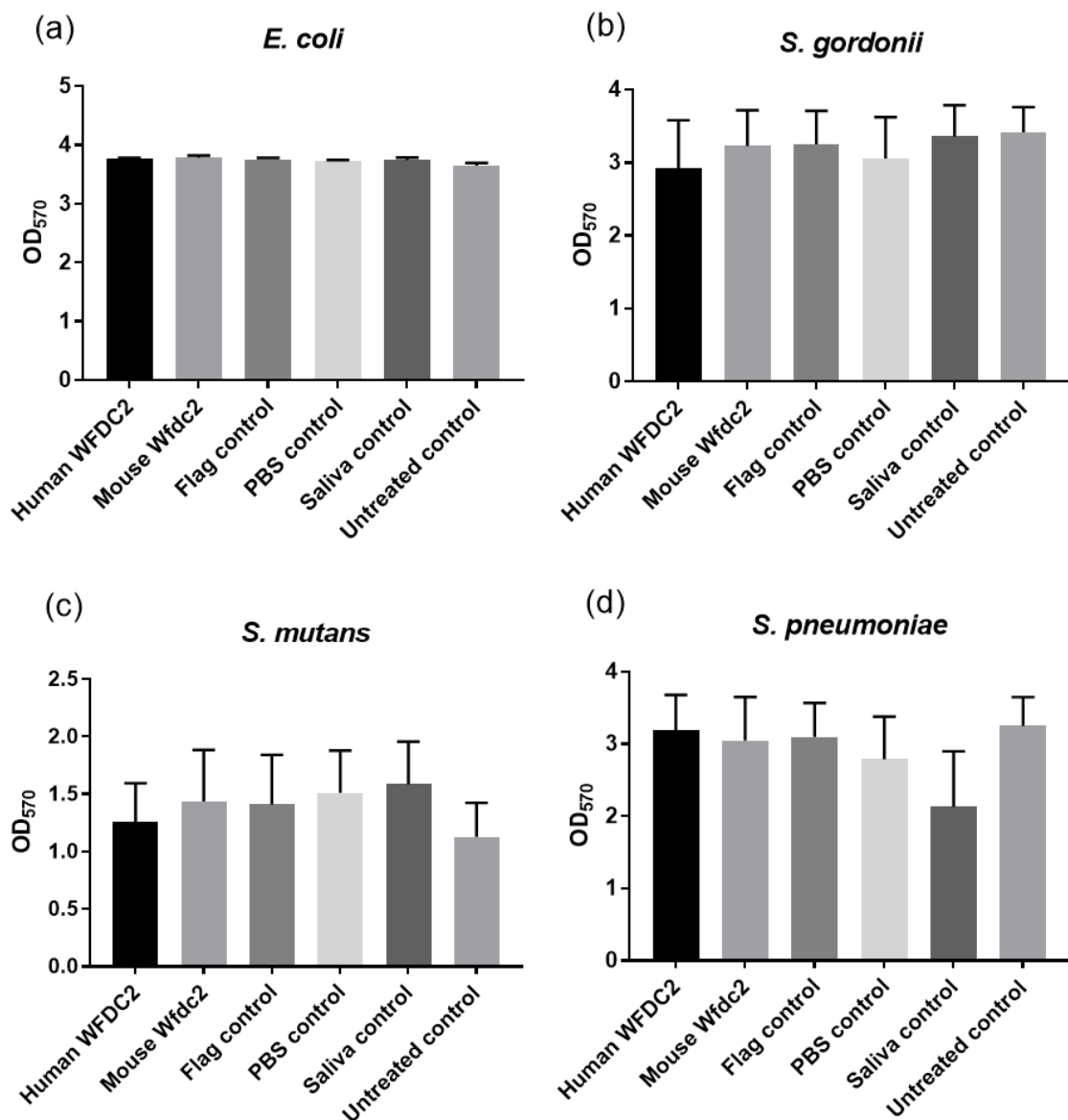
Figure 4.10. The effect of WFDC2 orthologues on bacterial biofilm formation. *E. coli* (a), *S. gordonii* (b), *S. mutans* (c) and *S. pneumoniae* (d) were exposed to human and mouse WFDC2 for 48 hours. The biofilms that formed in the presence of WFDC2 were stained and quantified. Statistical significance was calculated against untreated control. Data represents the mean ± SEM (N=3; n=3).

The capacity of WFDC2 to disrupt bacterial biofilms that are already established was also tested by adapting the method described above. Bacterial cultures were incubated in a 96 well plate for 24 hours, allowing time for biofilms to begin to form within the well. After 24 hours, recombinant WFDC2 (1 µM) was added to the bacteria and incubated for a further 24 hours. Adherent biofilms were then stained and quantified as before.

The results show that the WFDC2 orthologues and controls are incapable of significantly disrupting biofilms in *E. coli, S. gordonii* and *S. pneumoniae* when compared to the wells containing bacteria alone (Fig 4.11). Interestingly, incubation of WFDC2 orthologues with *S. mutans* appears to significantly enhance the biofilms produced compared to the untreated control, however this enhancement is also evident when *S. mutans* is incubated with FLAG peptide and PBS controls. Saliva had no significant effect on the biofilms of *S. mutans* compared to the untreated control.
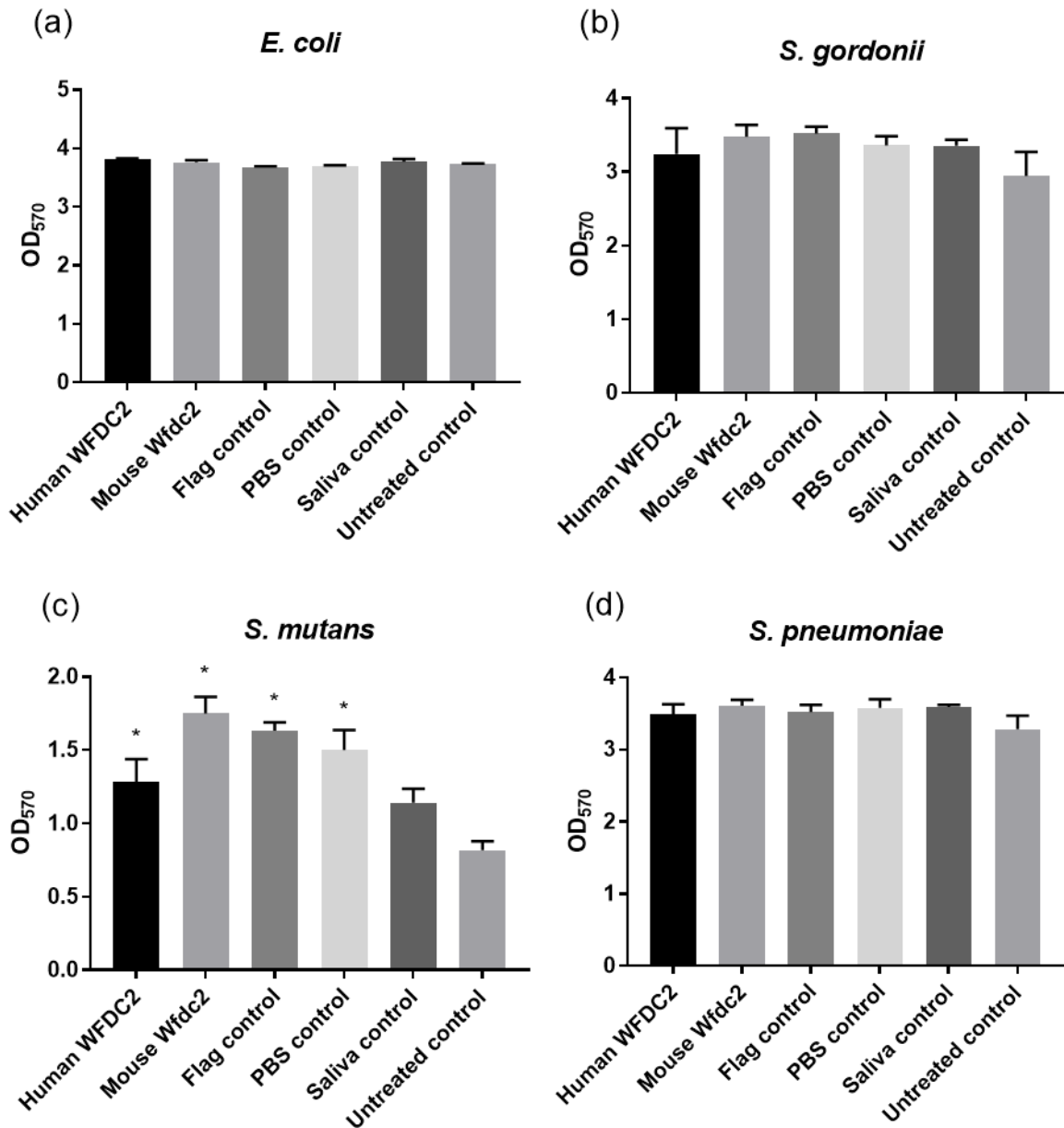
Figure 4.11. The effect of WFDC2 orthologues on bacterial biofilm disruption. *E. coli* (a), *S. gordonii* (b), *S. mutans* (c) and *S. pneumoniae* (d) biofilms that had grown over 24 hours were exposed to human and mouse WFDC2 and incubated with the protein for a further 24 hours. The resultant biofilms were stained and quantified to determine the effect of WFDC2 on biofilm disruption. Statistical significance was calculated against untreated control. Data represents the mean ± SEM (N=3; n=3); *P ≤ 0.05.

4.6. WFDC2 expression in salivary gland cells

To analyse the endogenous expression of all five WFDC2 splice variants using RT-PCR, primers were designed to span exon-exon boundaries and amplify individual isoforms. PCR conditions for the multiple primer pairs were optimised by testing three different annealing

temperatures (55°C, 60°C and 65°C) while the length of the annealing step and the number of cycles was kept constant (1 minute and 35 cycles, respectively).

The results demonstrated that 60°C was the optimal annealing temperature as all bands were present at the expected size with minimal primer dimers forming (Fig 4.12). At an annealing temperature of 55°C and 65°C, the V1 and V4 primers produce PCR products of the incorrect size. All subsequent splice variant RT-PCR reactions were carried out using an annealing temperature of 60°C.
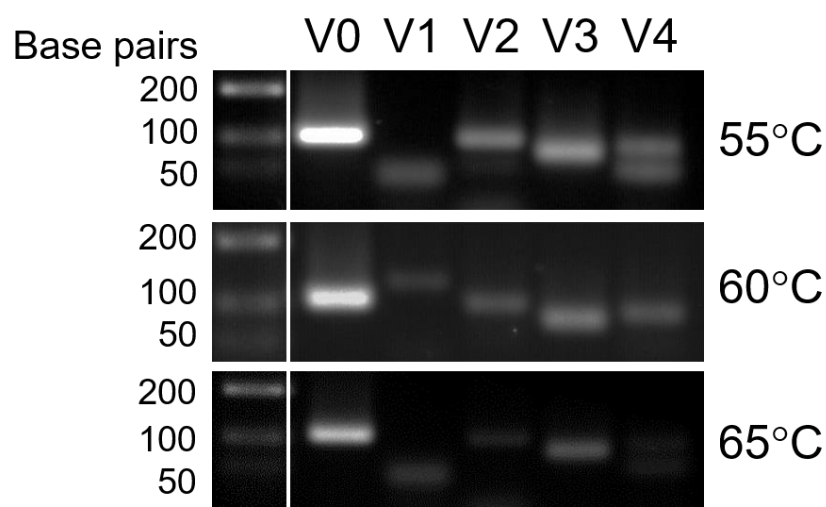
Fig 4.12. Optimisation of PCR conditions for splice variant-specific primers. Different annealing temperatures for five primer pairs targeting WFDC2 splice variants were tested using PCR to identify optimal thermal cycling conditions. CAL27 cancer cell line cDNA was used as the template as it expresses high levels of WFDC2 (see Section 5.5). PCR products were resolved on a 2% agarose gel. Expected band sizes were V0: 105 bp; V1: 132 bp; V2: 97 bp; V3: 75 bp; V4: 85 bp.

The endogenous WFDC2 expression in primary human sublingual gland (HuSL) cells derived from explanted human sublingual tissue was analysed by RT-PCR. Expression of all five splice variants was assessed semi-quantitatively using the aforementioned exon-spanning primers and the housekeeping gene, ornithine decarboxylase antizyme 1 (OAZ1), was used as the endogenous control. PCR products were resolved by agarose gel electrophoresis.

The RT-PCR results show that there is very low endogenous expression of V0, V1, V3 and V4 WFDC2 splice variants in HuSL cells when compared to the OAZ1 control (Fig 4.13).
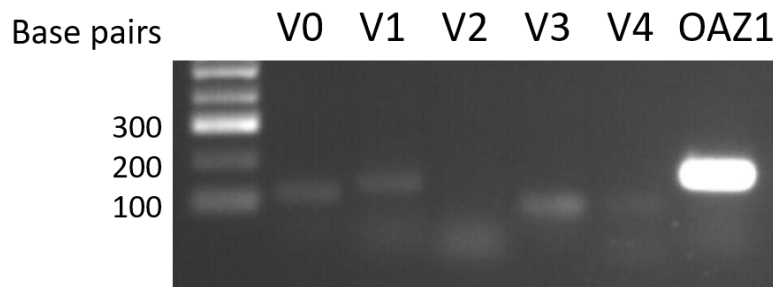
Fig 4.13. RT-PCR analysis of WFDC2 splice variant expression in HuSL cells. Primers specific to WFDC2 splice variants were used to probe primary human sublingual gland (HuSL) cell cDNA using RT-PCR to detect WFDC2 expression. PCR products were resolved on a 2% agarose gel. Expected band sizes were V0: 105 bp; V1: 132 bp; V2: 97 bp; V3: 75 bp; V4: 85 bp; OAZ1 endogenous control: 164 bp (N=2).

4.6.1. Cytokine exposure assay

To determine whether the expression of WFDC2 in HuSL cells could be upregulated in response to inflammatory mediators, the cells were exposed to LPS (50 μg/ml), TNF-α (20 ng/ml), IL-1ß (20 ng/ml) and IFN-γ (5 ng/ml) for 24 hours. RT-qPCR was then carried out to quantify WFDC2 expression and determine whether expression was affected following cytokine treatment when compared against the untreated control cells.

The RT-qPCR analysis shows that when HuSL cells were exposed to LPS, TNF-α, IL-1ß and IFN-γ for 24 hours, no significant change in WFDC2 expression was seen when compared to the untreated controls (Fig 4.14).
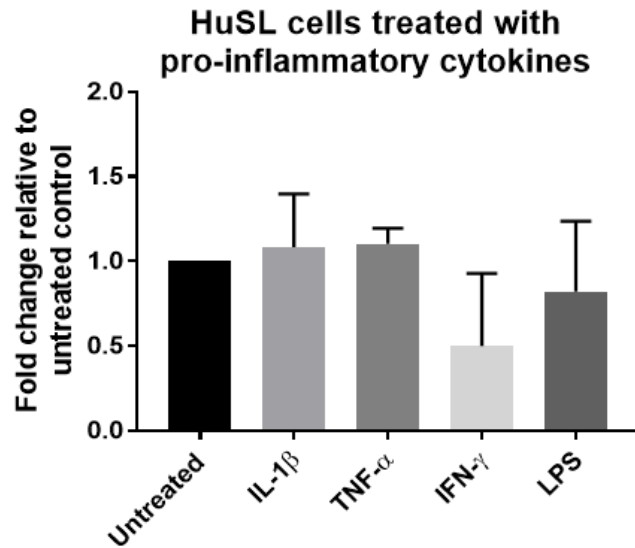
Fig 4.14. RT-qPCR analysis of WFDC2 expression in HuSL cells following exposure to inflammatory mediators. WFDC2 expression in human primary sublingual gland (HuSL) cells was tested by RT-qPCR following exposure to inflammatory mediators: LPS (50 µg/ml), TNF-α (20 ng/ml), IL-1ß (20 ng/ml) and IFN-γ (5 ng/ml). WFDC2 expression is represented as the fold change in expression compared to the untreated HuSL cell controls. Data represents the mean ± SD (N=3; n=3).

4.7.   Discussion

The aim of the work described in this chapter was to address the shortfalls in the literature regarding the putative anti-protease and antimicrobial functions of WFDC2. The development of recombinant human and murine WFDC2 protein, described in Chapter 3, provided tools to assess the biological activities of WFDC2 using *in vitro* functional assays.

The experiments were performed using concentrations of WFDC2 that were determined to be biologically relevant. Assessment of endogenous protein levels present at sites with the highest predicted WFDC2 expression, namely the salivary glands and upper airways, exhibited mean concentrations of 139 nM and 86 nM respectively. These values were similar to the median level of WFDC2 detected in vaginal secretions, as reported by Orfanelli et al (2014). The highest recorded concentration of WFDC2 in secretions was that defined by Orfanelli in vaginal secretions of a patient where WFDC2 levels reached 429 nM. With these values in mind, the assays described in this chapter were largely performed using 1 µM of

WFDC2, which takes into account that the endogenous concentration may be elevated in response to factors such as pro-inflammatory mediators and infectious agents, or due to binding to mucosal/tooth surfaces where the protein may become enriched. The literature is lacking in quantitative expression data for WFDC2, aside from the serum concentrations of cancer patients (which are found at picomolar levels), which suggests that previous studies have utilised an arbitrary concentration of recombinant WFDC2 which may not be representative of the biological levels found in the body.

Since its discovery in 1991, WFDC2 has been proposed to function as a host-derived serine protease inhibitor due to its possession of two WFDC domains which, in SLPI and elafin, have been shown to confer protease inhibitory properties (Kirchhoff et al., 1991). Within this study, recombinant WFDC2 was used in anti-trypsin and anti-elastase *in vitro* assays to determine whether it is indeed capable of inhibitory activity.

Interestingly, the results from this study showed that WFDC2 orthologues had no significant effect on trypsin activity when compared to the activity of trypsin alone (without inhibitor). SLPI, on the other hand, was able to reduce the activity of trypsin by ~50% at a concentration of 500 nM, an IC^{50} similar to that described in the literature (Masuda et al., 1995). At a concentration of 1 µM, SLPI was able to induce similar levels of trypsin inhibition to a broad-spectrum protease inhibitor solution (PI).

Contrary to the anti-trypsin assay, analysis of elastase inhibition showed SLPI to have no inhibitory effect on elastase activity. It is unclear why SLPI failed to inhibit elastase within the parameters of this experiment, given its well documented potency as an elastase inhibitor (Masuda et al., 1995; Smith and Johnson, 1985; Ying et al., 1994; Ying and Simon, 1993). Two different sources of elastase were assayed, to discount the possibility that the problems resulted from an impure source of enzyme, but the same result was seen. Evidence in the literature suggests that SLPI is sensitive to oxidation, whereby exposure to oxidative agents can cause the inhibitor to be inactivated (Boudier and Bieth, 1994). Given that elastase and trypsin both bind to the same active site (Eisenberg et al., 1990), the discrepancy in SLPI

activity between the anti-trypsin and anti-elastase assays suggests that the problem resides specifically in the reaction conditions when using elastase, as a disrupted active site would affect trypsin inhibition too. It is possible that the buffer used in the elastase assay has affected SLPI activity as it contained 10% DMSO which was required for solubilisation of the substrate. DMSO may have induced oxidation of SLPI and inhibited its activity as a protease inhibitor.

It is also unclear why the presence of SLPI and WFDC2 not only failed to inhibit elastase activity, but instead appeared to cause its activity to significantly increase. Interestingly, the PI control retained its capacity to inhibit elastase. The fact that the PI solution continued to elicit inhibitory activity within the experiment parameters adds complexity to the issue, as all potential inhibitors (PI, SLPI and WFDC2) were exposed to identical reaction conditions.

When considering the results from the trypsin inhibitory assay, the disparity in activity between SLPI and WFDC2 can perhaps be explained by the differences in the structure of their WFDC domains. As described in Section 1.1.4 and 1.3.5, SLPI has a highly conserved pattern of cysteine residues within its C-terminal WFDC domain which confers protease inhibitory activity. This domain has an identical cysteine spacing to that seen in the WFDC motif of elafin, another proven protease inhibitor. No other members of the WFDC protein family have been shown to exhibit anti-protease activity as a result of their WFDC domain [although some members encode other domains which confer protease inhibition, such as eppin's Kunitz domain (Wang et al., 2007)] and this corresponds with an absence of this cysteine pattern. Consequently, it has been suggested that the conserved domain structure of SLPI and elafin is an essential requirement for protease inhibition (Bingle and Vyakarnam, 2008). The fact that WFDC2 has an alternative, more closely aligned cysteine pattern suggests that the active site would not permit binding of a protease ligand.

Murine Wfdc2 has a different structure to human WFDC2, with the aforementioned exon 3 linker serving to separate the two WFDC domains (Section 1.2.4). This, however, does not appear to have sufficiently altered the structure in a way that permits the orthologue to exhibit anti-trypsin activity. This is perhaps because the cysteine-spacing within its domains is almost

identical to human WFDC2 and thus induces the same functional limitations (Bingle et al., unpublished).

Analysis of WFDC2's inhibitory capacity was extended to include matrix metalloproteases, MMP2 and MMP9, due to their prevalence in diseases also known to exhibit upregulated WFDC2, such as cancer and cystic fibrosis. When incubated with increasing concentrations of WFDC2, no significant reduction in MMP2 or MMP9 gelatin digestion was observed, suggesting that WFDC2 is not capable of gelatinase inhibition. It should be noted that SLPI is not reported to inhibit MMPs so was not an appropriate control for use in the zymography assays.

The results from this study suggest that WFDC2 is incapable of trypsin and MMP inhibition; however, this is inconsistent with some data found in the literature, such as that described by Hua et al (2014), Chhikara et al (2012) and LeBleu et al (2013). Because of their conflicting results, the methods described in these studies were closely analysed and are discussed below.

The protease inhibitory assay described by Hua et al (2014) utilises concentrations of protein that perhaps exceed levels that would be found in the human body, especially in the case of trypsin, which is exposed to a minimum concentration of 10.4 µM of WFDC2. If these concentrations cannot be achieved endogenously, then the biological relevance of the results should be called into question. Nonetheless, when 0.8 µM of WFDC2 was exposed to elastase, the enzyme activity did reduce by 15%, but it is unclear whether this result was statistically significant compared to negative controls as the P values were only shown for the highest WFDC2 concentration. The zymography results described in Hua et al (2014) also show WFDC2 inhibition of MMP9, however the concentration of WFDC2 protein required to produce this result is not described, which calls the reliability of the data into question. It is possible that a higher level of WFDC2 was used compared to the concentration used in this study which could, in part, explain the discrepancy in results.

In Chhikara et al (2012), the purified WFDC2 utilised for protease inhibition experiments was described as a 14 kDa glycosylated monomer which forms a 42 kDa trimer in native conditions. This result greatly contrasts with the molecular weight of recombinant and endogenous WFDC2 protein reported within this study and in the literature when analysed in denaturing and native conditions (Drapkin et al., 2005; Hellström et al., 2003; Hua et al., 2014). The authors attribute the difference in protein size to the fact that it was purified from seminal plasma and represents a different splice variant, yet tissues of the male reproductive tract are known to abundantly express WFDC2 V0 as shown by IHC (Uhlén et al., 2015), and V0 has been shown to be ~100-fold more abundantly expressed in the epididymis compared to V1-V4 (Jiang et al., 2013). With this in mind, it seems unlikely that another isoform would be purified. These factors call the reliability of the results into question.

The assays carried out by LeBleu et al (2013) appear to use a more appropriate concentration of WFDC2 and the human recombinant protein is closer to the expected size (~18 kDa). Their results show that human WFDC2 is able to inhibit the collagen digestion of MMPs and Prss enzymes, which contrasts with the results reported in this study (although their methods differ so results cannot be directly compared). The result also contrasts with that described by Hua et al (2014) as both MMP2 and MMP9 are shown to be inhibited by WFDC2. Evidently further experiments will be required to resolutely elucidate whether WFDC2 is capable of protease inhibition, given the ongoing inconsistencies in the data.

During this study, the putative antibacterial functions of human and murine WFDC2 were also assessed. The functional assays were carried out using 1 μM of recombinant protein. FLAG peptide elution solution was used as a negative control as it is possible that, during the elution step of WFDC2 purification, some of the FLAG peptide used for elution could also be present in the purified WFDC2 product.

The bacteria strains used within the *in vitro* assays were selected for various reasons. Firstly, Gram-positive and Gram-negative bacteria were desired as some antimicrobial agents are more effective against a particular Gram group, so *Escherichia coli* (Gram-negative) and

species of *Streptococcus* (Gram-positive) were chosen. Secondly, given that WFDC2 is abundantly expressed in the oral cavity and respiratory tract, and its concentration in secretions at these sites had been quantified, strains endogenous to the mouth and lungs were identified: *S. gordonii* and *S. mutans* which colonise the mouth, and *S. pneumoniae* which colonises the respiratory tract. Finally, the four chosen microbes represent different degrees of virulence. *S. gordonii* is commensal-like and represents the least harmful of the panel of bacteria chosen. *S. mutans*, though occurring naturally as part of the normal oral flora, is an opportunistic pathogen and a major cause of dental caries. The strain of *E. coli* used was derived from a clinical isolate of a patient with a urinary tract infection (UTI) therefore is likely to express virulence factors which enable urinary tract colonisation. UTIs can vary in severity, from mild bladder infection to potentially fatal renal failure and sepsis. Finally, *S. pneumoniae* is an invasive pathogen which can cause mild diseases such as otitis media and sinusitis or much more severe illnesses including pneumonia and meningitis. By using a variety of bacteria for the functional assays in this study, it is more likely that any antimicrobial activities of WFDC2 will be elucidated.

When bacterial binding was tested using a traditional pull down assay, it was clear that neither WFDC2 orthologue was capable of binding to any of the four bacterial species. Due to the cationic nature of many WFDC family proteins, their putative mode of bactericidal activity is via cationic lysis. This 'coating' of bacteria and subsequent induction of membrane damage and lysis is known to occur with eppin and elafin, as described in Section 1.3.2. The fact that WFDC2 fails to interact with the bacteria strains tested suggests that, if it is capable of bacterial killing, it achieves lysis via a different route, without binding. This negative result is supported by data showing that incubation of murine Wfdc2 with *Mycobacterium bovis* fails to induce membrane permeabilization and killing, whereas murine Slpi is able to induce surface damage and bacterial killing when analysed in the same assay (Nishimura et al., 2008).

Conversely, a positive interaction between WFDC2 and four different bacterial species (including *S. aureus*) has been reported in the literature, as determined by pull down assay

(Hua et al., 2014). Given that the recombinant WFDC2 used by Hua et al (2014) was an Fc-fusion protein, the positive binding with *S. aureus* is probably due to the well documented binding interaction between the Fc domain of IgG and the *S. aureus* surface protein, Protein A. It is probable that the Fc-fusion changes the binding properties of WFDC2 to enhance its affinity with other bacterial surface proteins in addition to Protein A.

The effect of WFDC2 on bacterial viability was tested using two different methods: a traditional assay which involved plating out WFDC2-exposed bacteria via Miles-Misra and counting the CFU, and a technique involving an overnight incubation of bacteria and protein, whereby absorbance of the suspension was monitored to analyse cell density. Both assays showed that WFDC2 orthologues had no effect on the viability of any of the four microbes tested. Even when the WFDC2 concentration was raised 3-fold higher, no bactericidal effects were seen. Data published by Hua et al (2014) also shows that incubation of recombinant WFDC2 with *S. aureus* has no significant effect on bacterial viability, even at a concentration as high as 7.75 µM.

This result supports the lack of interaction between WFDC2 and bacteria described in the pull-down experiments: it seems unlikely that WFDC2 would elicit bacterial killing via an entirely different mechanism to that of its protein family members given their similar structure and its lack of other functional domains. It should also be noted that when the bacteria were incubated with WFDC2 overnight, the orthologues appeared to undergo protein degradation and glycan harvesting, as visualised by shifts in the Western blotting bands and quantified by densitometry. Glycosylation can affect protein stability, structure and activity, therefore it is possible that this cleavage of sugars by bacteria may negatively impact the biological function of WFDC2.

Host defence roles of salivary glycoproteins also include agglutination and biofilm disruption. Agglutination is achieved by glycoproteins such as salivary agglutinin and mucins, which can induce bacterial aggregation, promote their clearance and prevent biofilm formation (Ericson and Rundegren, 1983). Although known agglutinins tend to be large proteins (>250 kDa) with

an abundance of glycosylation (Cross and Ruhl, 2018), other smaller glycoproteins have been shown to be involved in aggregation, such as members of the BPI fold containing family of proteins (Lunn, 2013). Like WFDC proteins, BPI fold containing family proteins are expressed in the oral cavity and upper respiratory tract, have putative innate immune functions and exhibit rapid molecular evolution (Bingle et al., 2009).

To investigate whether WFDC2 is involved in such defence processes, the recombinant orthologues were used in agglutination and biofilm disruption assays. The results showed that WFDC2 orthologues were incapable of bacterial agglutination, with a clear pellet of bacteria forming in the well. In contrast, the saliva positive control clearly induced a film of bacteria to form.

Biofilm formation and disruption assays were also carried out, which demonstrated that incubation with WFDC2 caused no significant disruption in the biofilms that formed, regardless of whether the protein was present while biofilms were still in the process of forming, or if protein was added to an already established biofilm. Host proteins with proposed anti-biofilm activity tend to elicit their effects by binding to bacteria; for example, salivary IgA binds to bacterial adhesins to prevent their initial attachment to oral surfaces and subsequent biofilm formation (Childers et al., 2006; Smith and Taubman, 1996). As WFDC2 appears unable to bind to bacteria, this perhaps explains their inability to affect biofilms. It should, however, be noted that tissue culture plastic (TCP) is not representative of the tooth surface/enamel pellicle, therefore it is possible that this method of analysing biofilm formation could prevent endogenous anti-biofilm functions of WFDC2 to be recognised e.g. if bacteria have a greater propensity to bind to TCP than the tooth. Nonetheless, these preliminary investigations suggest that WFDC2 is incapable of compromising the biofilms or inducing agglutination to contribute to bacterial clearance.

To augment the functional assay data collected using recombinant protein, endogenous WFDC2 expression was also examined using RT-qPCR. Changes in WFDC2 expression in

primary sublingual gland (HuSL) cells were measured in response to exposure to pro-inflammatory mediators.

The endogenous WFDC2 expression in HuSL cells was low, as shown by RT-PCR and gel electrophoresis. The difference in expression between whole sublingual glands (Fig 1.8) and the primary cells was most likely due to the explanted cells being grown as a monolayer and in a defined cell culture medium, which affects the cell phenotype. Monolayer cells are devoid of complex structures such as acini and ducts, therefore the primary sites known to abundantly express WFDC2 were absent, which was reflected in the RT-PCR result.

When the primary cells were exposed to pro-inflammatory mediators for 24 hours, there was no significant change in WFDC2 gene expression when measured by RT-qPCR and compared against the untreated control. This result is interesting as it corroborates the functional assay data which appears to show that WFDC2 is incapable of anti-protease or antimicrobial activity. If WFDC2 was capable of protease inhibition, its expression would most likely be upregulated in response to pro-inflammatory cytokines such as IL-1β and TNF-α, which are produced at sites of inflammation resulting from excessive protease activity. This is the case for both SLPI and elafin, which are largely upregulated in response to cytokines and proteases themselves (see Section 1.3.4). As a result, SLPI and elafin are frequently upregulated in diseases characterised by inflammation and dysregulated protease activity such as emphysema and psoriasis (Alkemade et al., 1994; Betsuyaku et al., 2002). If WFDC2 was capable of antimicrobial activity, WFDC2 gene expression levels would be expected to become upregulated in response to pro-inflammatory cytokines and bacterial PAMPs such as LPS.

These results suggest that the correlation between bacterial colonisation and WFDC2 upregulation, described in Section 1.3.3, is not due to conventional upregulation of WFDC2 expression in response to common alarm signals such as IL-1β, TNF-α and LPS.

It should also be noted that within the scope of this study, two oral-derived cancer cell lines with moderate to high endogenous WFDC2 expression were used in a similar exposure assay, as described in Chapter 5. The cancer cell lines also showed no significant change in gene expression in response to IL-1β and TNF-α, supporting the results seen when monitoring HuSL cellular responses. Exposure assays published in the literature using ovarian cancer cells also showed that IL-1β and TNF-α had no observable effect on WFDC2 expression, whereas elafin was upregulated by both cytokines (Clauss et al., 2010; described in Section 1.3.4.).

Chapter 5

Determining the role of WFDC2 in tumorigenesis

5.1.   Introduction

Although an abundance of histological evidence is available in the literature to show WFDC2 expression in gynaecological and lung tumours (Bingle et al., 2006; Georgakopoulos et al., 2012; Uhlen et al., 2017), there is limited information regarding expression of WFDC2 in tumours of the oral cavity (Galgano et al., 2006). To address this, two different subtypes of oral tumour will be analysed by immunohistochemistry (IHC) to elucidate the WFDC2 staining pattern.

The overexpression of WFDC2 in cancer is well documented, however, it is unclear whether WFDC2 plays an active role in tumorigenesis and the literature remains conflicted. Many attempts have been made to knock-down WFDC2 expression in cancer cell lines using small interfering RNA (siRNA), with varying success (Gao et al., 2011; Guo et al., 2015; Lu et al., 2012; Zhu et al., 2016), yet CRISPR gene editing has not been used to assess the effect of WFDC2 silencing on cell behaviour. To address this, an appropriate cancer cell line with high endogenous WFDC2 expression will be targeted for CRISPR gene editing, and the assays described in this chapter will be the first experiments to use CRISPR cells to assess the effects of WFDC2 on cell behaviour. Finally, cancer cell lines will also be used to observe changes in WFDC2 expression in response to pro-inflammatory mediators, to expand upon the exposure assays described in the literature (Clauss et al., 2010).

5.2.   Aims and objectives

The aim of the research presented in this chapter was to further elucidate the expression profile of WFDC2 in oral cancer and to determine whether its expression affects cancer cell

behaviour. Immunohistochemistry was used to analyse WFDC2 expression in oral tumours. Analysis of WFDC2 expression in oral and lung cancer cell lines was used to determine the level of gene expression and any subsequent secretion of WFDC2 protein into conditioned media. The effect of pro-inflammatory mediators on WFDC2 expression in oral cancer cell lines was assessed using RT-qPCR. Finally, an oral cancer cell line identified to express high levels of endogenous WFDC2 was targeted for CRISPR/Cas9 gene editing. Successful gene silencing was evidenced by the absence of WFDC2 protein in the conditioned media when analysed by Western blotting and ELISA. The downstream effect of WFDC2 silencing on cancer cell behaviour was monitored using standard functional assays.

5.3. Materials and methods

To access the relevant materials and methods utilised in this chapter, please refer to the sections below:

- The expression of WFDC2 protein in oral cancers was assessed using immunohistochemistry as described in Section 2.11.
- Oral- and lung-derived cancer cell lines were cultured as outlined in Section 2.1.1. The endogenous WFDC2 gene expression in cancer cell lines was analysed by RT-PCR and RT-qPCR as described in Section 2.2.2 to 2.2.5. The secretion of endogenous WFDC2 protein from the cancer cell lines was analysed by Western blotting and ELISA, as described in Section 2.3.3 and 2.3.7, respectively.
- Cancer cell lines were exposed to pro-inflammatory mediators and the changes in WFDC2 gene expression were monitored by RT-qPCR, as described in Section 2.8.
- CRISPR gene editing was performed as per the methods in Section 2.9. Successful gene editing was confirmed by RT-PCR, RT-qPCR, Western blotting and ELISA, as outlined in Section 2.9.12.
- The CRISPR edited cells were subjected to conventional cell behaviour assays as described in Section 2.10.

5.4. Expression of WFDC2 in oral tumours

To assess whether WFDC2 protein is produced in tumours of the oral cavity, five cases of squamous cell carcinoma (SCC) of the tongue and a single case of lymph node metastasis were analysed by immunohistochemistry (IHC) using a monoclonal antibody raised against human WFDC2.

Positive WFDC2 staining was found to be exclusively localised to phenotypically normal minor salivary glands present within the tongue, with intense positivity being visible in the epithelium of the ducts (Fig 5.1). All other cell types present in the sections, particularly tumour cells, were negative for WFDC2 expression, suggesting that the protein is not expressed in oral SCC. A single case of SCC lymph node metastasis was also analysed and was negative for WFDC2 staining (data not shown).

The expression of WFDC2 in mucoepidermoid carcinoma (MEC), the most common type of salivary gland tumour, was also assessed by IHC. In contrast to SCC, many MEC sections exhibited positive, focal WFDC2 expression (21/58 sections were positive). Staining was notably abundant in mucin-producing cells, with intense staining also visible in cystic spaces (Fig 5.2a-f). WFDC2 staining was largely absent in sites of normal mucous acini within the MEC tissue (Fig 5.2g and Fig 5.2h).
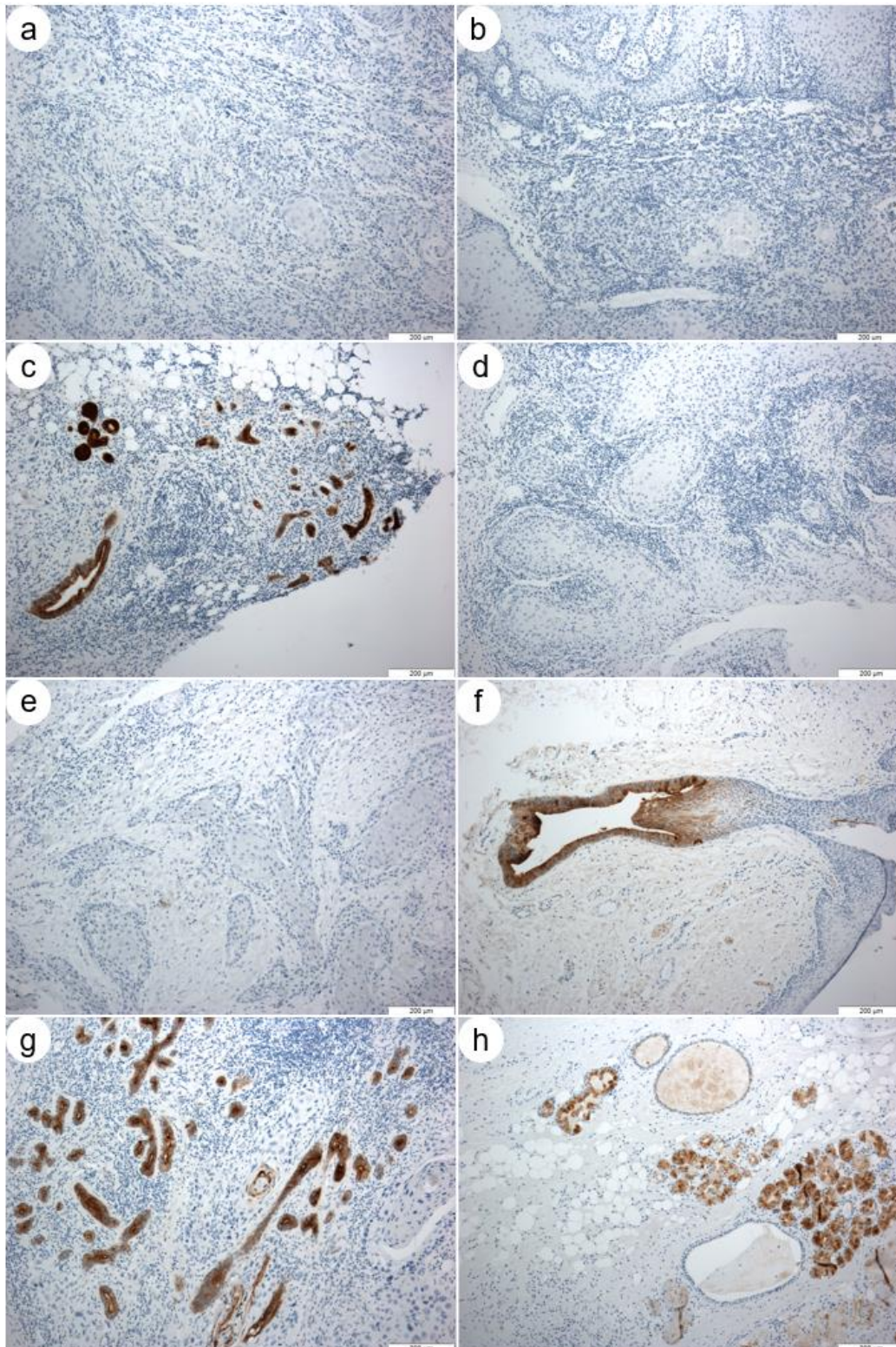
Fig 5.1. Localisation of WFDC2 in human squamous cell carcinoma of the tongue. Sections of squamous cell carcinoma (SCC) of the tongue were stained for WFDC2 by IHC. Cancer cells were negative for WFDC2, while minor salivary glands were positive for WFDC2 (a-g). Healthy parotid gland tissue was used as a control for positive staining (h). Original magnification of the images was 10X.
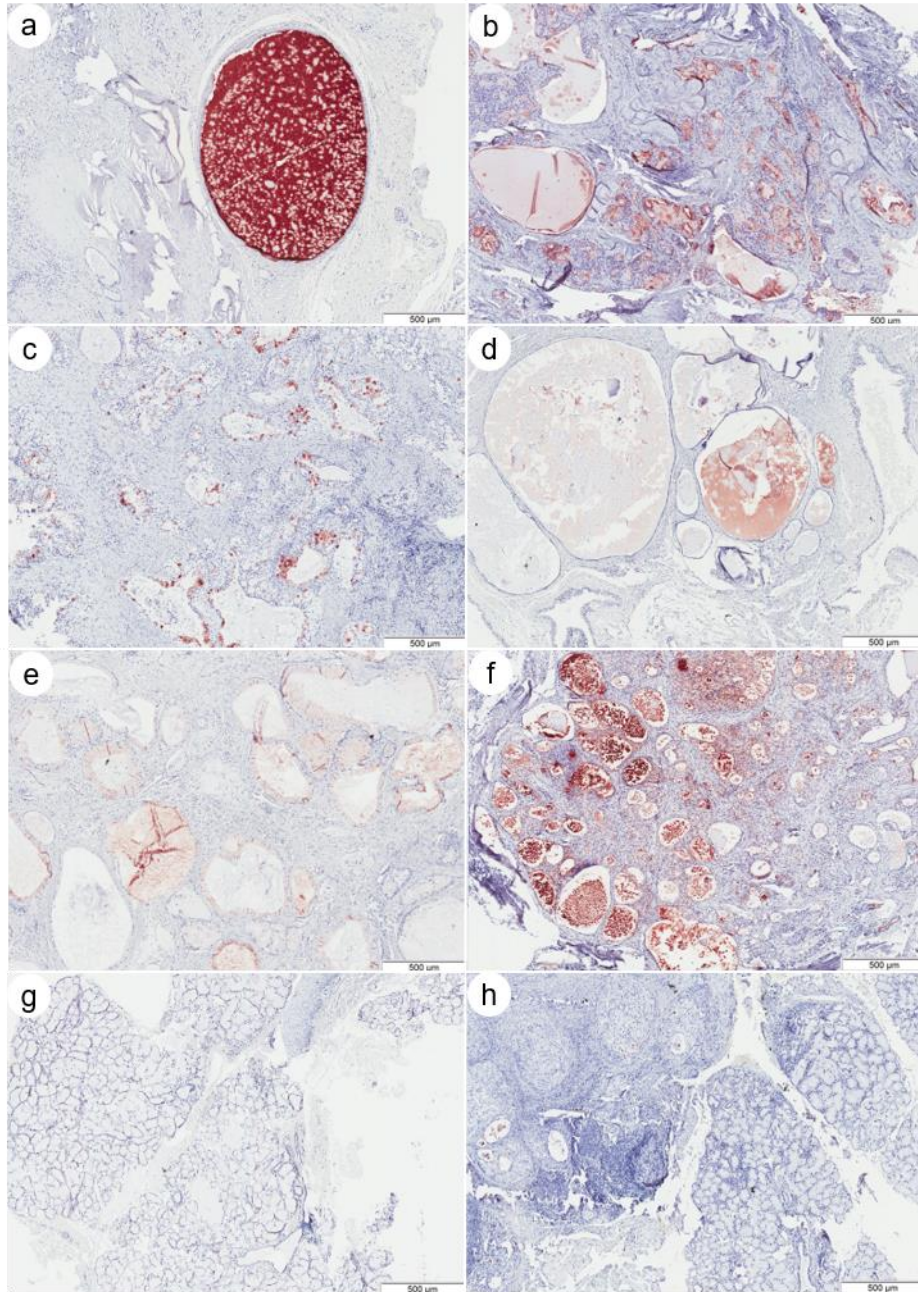
Fig 5.2. Localisation of WFDC2 in human mucoepidermoid carcinoma. Sections of mucoepidermoid carcinoma (MEC) were stained for WFDC2 by IHC (a-h). Positive staining was seen in mucous-producing cells and cystic ductal spaces (a-f). Normal mucous acini were mostly negative for WFDC2 (g, h). Original magnification of the images was 4X. Staining performed by Debora Pereira (visiting student from the University of Campinas, Brazil).

5.5.   Expression profile of WFDC2 in cancer cell lines

Data in the literature has shown that cancer cell lines derived from lung and ovarian tumours endogenously express WFDC2 (Bingle et al., 2006; Drapkin et al., 2005), thus providing a valuable tool for knockdown studies. To further analyse the expression profile of WFDC2 in cancer cell lines, four oral cancer cell lines and six lung cancer cell lines (see Table 2.1) were analysed by RT-PCR and RT-qPCR to identify and quantify positive gene expression. Downstream protein secretion into culture media was also analysed using Western blotting and ELISA.

To semi-quantitatively assess whether all five splice variants of WFDC2 are expressed by the ten cell lines in question, primers for PCR were designed to target individual splice variants. OAZ1 was used as an endogenous control. Resultant PCR products were resolved on a 2% agarose gel.

The results appear to show that all five splice variants are expressed in CAL27, H357, NCI-H292, NCI-H441 and NCI-H358 cell lines. CAL27 and NCI-H441 cells appeared to express the highest level of WFDC2 variants when compared to the band intensity of the endogenous control, OAZ1 (Fig 5.3). FaDu, D20 and NCI-H647 cell lines expressed low levels of WFDC2 and failed to express variant 2 and/or variant 4. NCI-H322 and NCI-H226 lacked expression of all five splice variants.
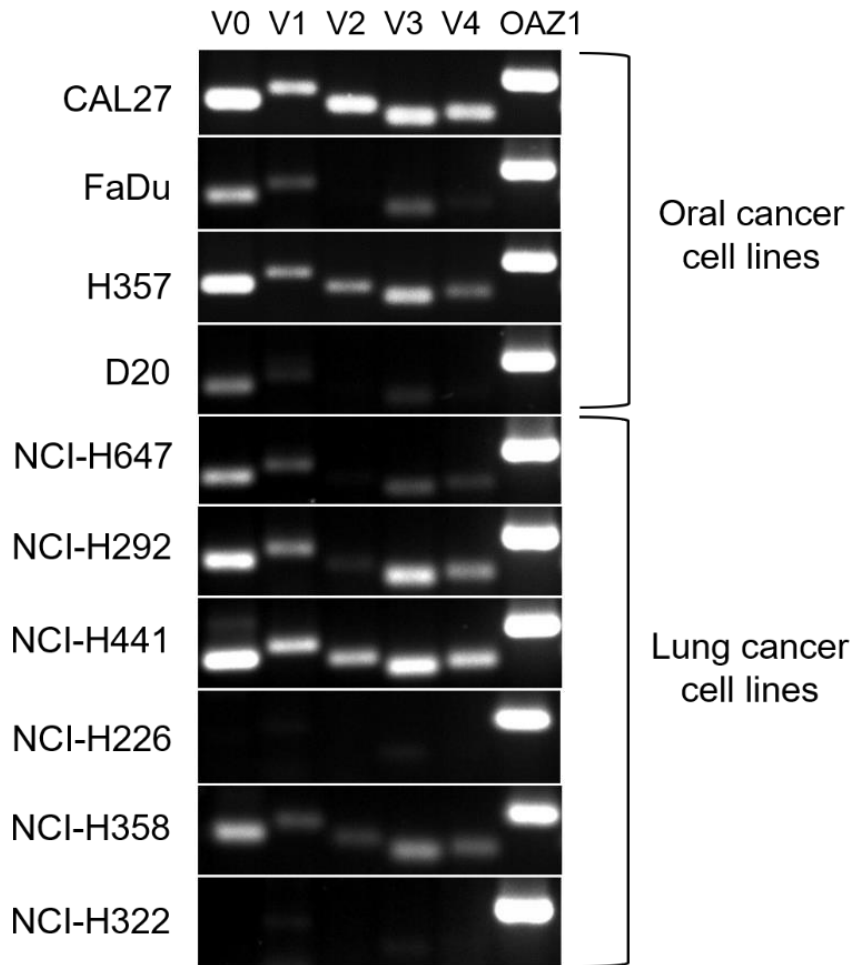
Fig 5.3. Differential expression of WFDC2 splice variants analysed by RT-PCR. Primers specific to WFDC2 splice variants (V0-V4) were used to probe 10 cancer cell line target cDNAs using RT-PCR to detect splice variant expression. PCR products were resolved on a 2% agarose gel. Expected band sizes were V0: 105 bp; V1: 132 bp; V2: 97 bp; V3: 75 bp; V4: 85 bp; OAZ1 endogenous control: 164 bp (N=3).

RT-qPCR was utilised to quantify expression of WFDC2 using TaqMan primers that detected isoform V1. Expression was determined as the fold-change compared to the B2M endogenous control. The results show that expression of WFDC2 in CAL27 cells, a cell line derived from a squamous cell carcinoma of the tongue, was significantly higher than in all other cancer cell lines tested (P=<0.0001) (Fig 5.4). NCI-H441 cells, derived from a papillary adenocarcinoma of the lung, were the second most highly expressing cells. This supports the semi-quantitative RT-PCR analysis of the cell lines seen in Fig 5.3.
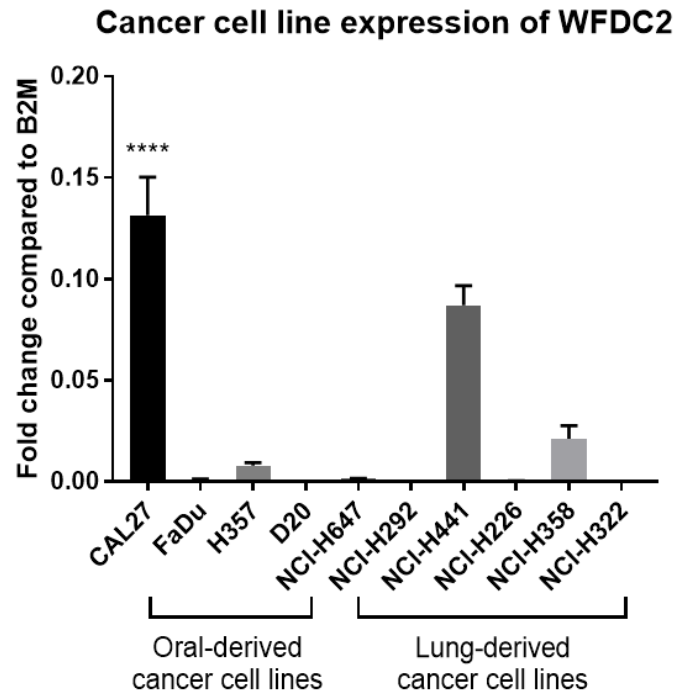
Fig 5.4. RT-qPCR quantification of WFDC2 gene expression in 10 cancer cell lines. WFDC2 expression in 4 oral-derived and 6 lung-derived cancer cell lines was tested using RT-qPCR. WFDC2 expression is represented as the fold change in expression compared to the B2M endogenous control. NCI-H441 is significantly more highly expressed than all cell lines (P=<0.0001) apart from CAL27. The CT values are presented in the Appendix. Data represents the mean ± SD (N=3; n=3). **** P ≤ 0.0001.

To further analyse the expression levels of WFDC2 in cancer cell lines, the presence of WFDC2 protein in conditioned media was analysed by Western blotting. Serum-free conditioned media from all 10 cancer cell lines was collected, concentrated and analysed by BCA to determine total protein concentration. The protein content was standardised and 45 µg of total protein was subjected to Western blot analysis.

The Western blotting result shows that WFDC2 is only detectable in media derived from CAL27 cells (Fig 5.5). All other cancer cell lines failed to express detectable levels of WFDC2. The molecular weight of the positive band is ~25 kDa.
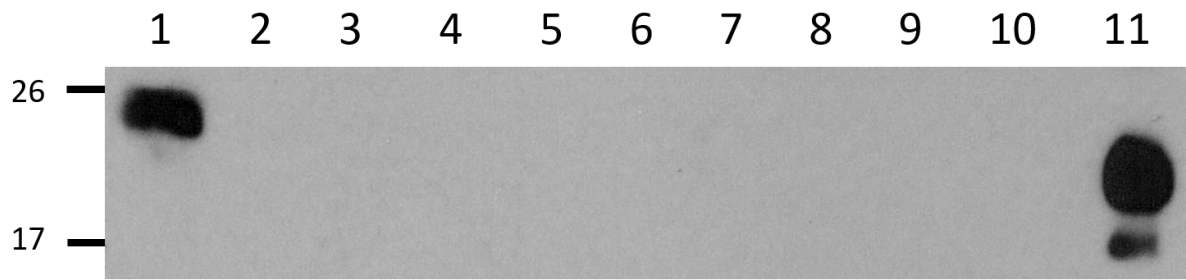
Fig 5.5. Western blot analysis of WFDC2 protein secretion in 10 cancer cell lines. The presence of WFDC2 protein in concentrated conditioned media from four oral-derived and six lung-derived cancer cell lines was analysed by Western blotting using an anti-WFDC2 monoclonal antibody. Lane numbers represent the following: 1 – CAL27, 2 – FaDu, 3 – H357, 4 – D20, 5 – NCI-H647, 6 – NCI-H292, 7 – NCI-H441, 8 – NCI-H226, 9 – NCI-H358, 10 – NCI-H322, 11 – recombinant human WFDC2 (N=2).

To quantify the endogenous concentration of WFDC2 secreted by cancer cell lines, conditioned media from the ten cancer cell lines was analysed using a WFDC2 ELISA. An equal concentration of total protein was analysed for each sample which was subsequently diluted according to the manufacturer's recommendation.

CAL27 cells secreted significantly more WFDC2 protein than all other cell lines tested (P=<0.0001) with a mean concentration of 147.5 nM in the conditioned media (Fig 5.6). This result supports both the RT-qPCR and Western blot analysis. NCI-H441 cells exhibited the second most abundant level of WFDC2 of protein (16.1 nM), followed by H357 (10.5 nM). All other cell lines secreted less than 3 nM of WFDC2. It should be noted that although the expression levels were extremely low for many cell lines, concentrations of WFDC2 were still detectable (>0 nM).

WFDC2 production in cancer cell lines
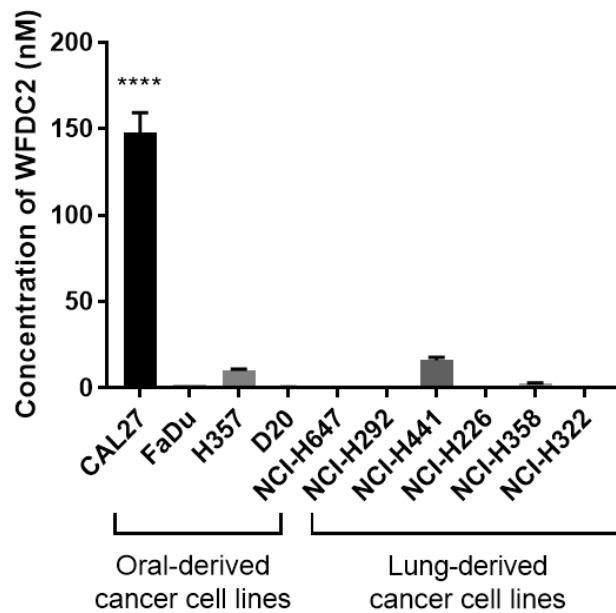
Fig 5.6. ELISA quantification of WFDC2 protein production in 10 cancer cell lines. WFDC2 production in 4 oral-derived and 6 lung-derived cancer cell lines was tested using ELISA. Data is presented as WFDC2 concentration (nM), with the mean ± SD shown (n=3) **** $P \leq 0.0001$.

5.6.   Cytokine exposure assay

To determine whether WFDC2 gene expression in cancer cell lines could be upregulated in response to inflammatory mediators, two oral-derived cancer cell lines (CAL27 and H357) were exposed to TNF-α (20 ng/ml), IL-1ß (20 ng/ml) and IFN-γ (5 ng/ml) for 24 hours. RT-qPCR was carried out to quantify WFDC2 expression levels in the cytokine-exposed cells and determine whether expression was affected following treatment when compared to untreated control cells.

The RT-qPCR analysis shows that exposure of CAL27 cells to IL-1ß and TNF-α for 24 hours had no significant effect on WFDC2 gene expression compared to the untreated control (P=0.997 and P=0.998 respectively). Interestingly, IFN-γ appeared to significantly reduce WFDC2 expression by approximately 0.5-fold (P=0.027) (Fig 5.7a).

Similarly, when H357 cells were exposed to IFN-γ they displayed a significant 0.7-fold reduction in WFDC2 expression (P=0.008). Treatment with IL-1ß and TNF-α caused no statistically significant change in WFDC2 expression compared to the untreated H357 cells, (P=0.830 and P=0.168 respectively) although TNF-α did cause an ~1.5-fold increase in WFDC2 expression (Fig 5.7b).
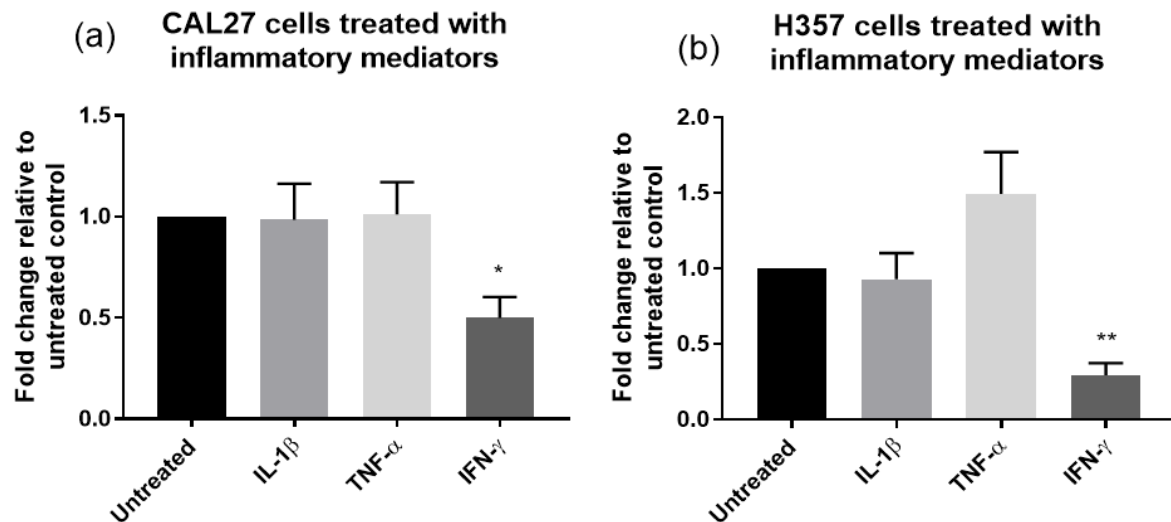
Fig 5.7. RT-qPCR analysis of WFDC2 expression in CAL27 and H357 cells following exposure to inflammatory mediators. WFDC2 expression in CAL27 (a) and H357 (b) cells was tested by RT-qPCR following exposure to inflammatory mediators: LPS (50 μg/ml), TNF-α (20 ng/ml), IL-1ß (20 ng/ml) and IFN-γ (5 ng/ml). WFDC2 expression is represented as the fold change in expression compared to the untreated cell control. Data represents the mean ± SD (N=3; n=3); ** P ≤ 0.01, * P ≤ 0.05.

5.7. CRISPR/Cas9 gene editing of WFDC2 in an oral cancer cell line

CRISPR/Cas9 gene editing was utilised to knock-out the expression of WFDC2 in CAL27 cells which exhibit high endogenous WFDC2 expression. The editing event was carried out using a plasmid encoding a guide RNA and Cas9 endonuclease, in addition to GFP which was used for downstream selection of transfected mammalian cells. Extensive tests were carried out to confirm that a successful gene silencing event had occurred. The CRISPR edited cells were subsequently used for downstream functional assays.

5.7.1. CRISPR editing

In order to recruit a Cas9 nuclease to the gene of interest and induce silencing, a single guide RNA (sgRNA) must be synthesised: the sgRNA is a 20 nucleotide DNA sequence complementary to the target gene.

To silence WFDC2, two different sgRNAs were designed and synthesised, each targeting the 5' end of exon 2 (Fig 5.8). Each sgRNA (sgRNA1 and sgRNA2) was ligated into a pX458 target vector and was subsequently transformed into competent *E. coli* cells. The pX458-sgRNA1 and pX458-sgRNA2 plasmids were isolated from ampicillin-resistant transformants. A sample of each plasmid was subjected to restriction digestion to confirm the presence of the vector, however at this stage the presence or absence of the ligated sgRNA cannot be confirmed due to its small size (~20 bp) which cannot be visualised on an agarose gel.
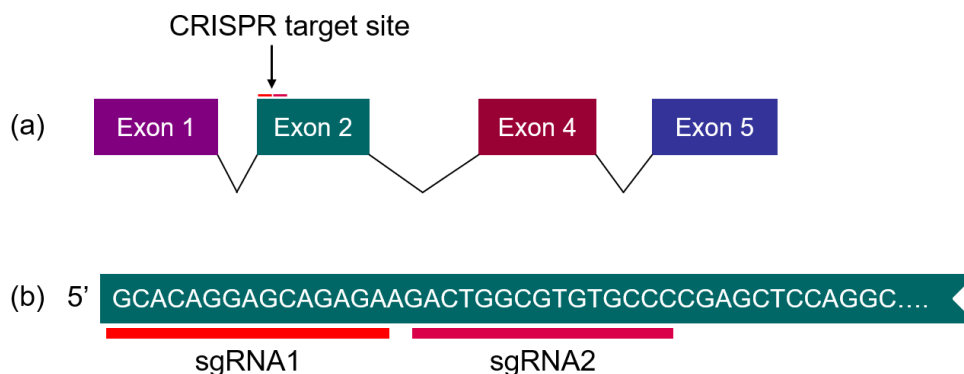
Fig 5.8. The CRISPR target site. A schematic diagram depicting the CRISPR target site within exon 2 of human WFDC2 isoform V0 (a). The exact target site of the two sgRNAs within exon 2 are shown (b).

Target CAL27 cells were seeded into a 6-well plate and co-transfected with pX458-sgRNA1 (1 µg) and pX458-sgRNA2 (1 µg) plasmids using JetPrime transfection reagent in order to induce a dual cleavage by Cas9 and remove ~40 nucleotides from exon 2 of WFDC2 (i.e. both sgRNA1 and sgRNA2 targets).

Transfected cells were sorted via fluorescence activated cell sorting (FACS) according to the GFP expression of positive transformants. Negative transformants were also collected to be used as a negative control in downstream assays. Following FACS sorting, cells were serially

diluted in a 96-well plate for clonal isolation, with the aim of growing colonies of cells derived from a single progenitor cell.

Colonies were expanded and genomic DNA (gDNA) was extracted and analysed by PCR using two primer pairs which were designed to be 50-100 bp upstream/downstream of the sgRNA target. PCR products were used to identify populations of clones with the desired CRISPR knockout mutation. Genomic DNA derived from CAL27 cells that were not successfully transfected with the pX458 vector (i.e. FACS sorted but GFP negative) was included in the PCR to act as a negative control.

PCR products derived from gDNA of 12 different clonal colonies were resolved on an agarose gel. The results showed that colony 11 had a visible shift in expected band size (Fig 5.9) with both primer pairs which suggested that a deletion resulting from a successful gene editing event had occurred.
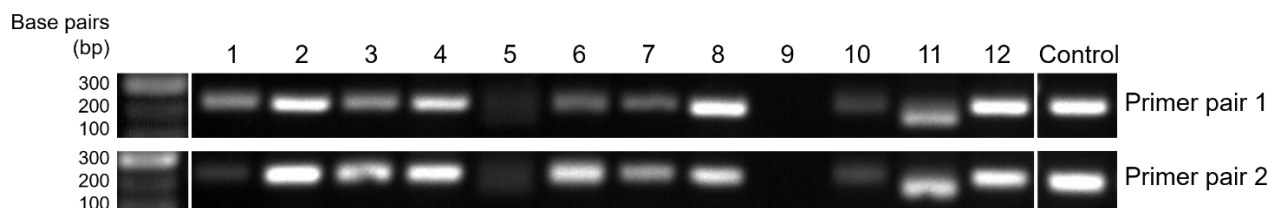
Fig 5.9. Resolution of PCR products derived from gDNA of potential gene edited CAL27 colonies. Primer pairs flanking the sgRNA target within gDNA of CRISPR gene edited CAL27 cell colonies were used for PCR. Products were resolved on a 2% agarose gel. In the absence of CRISPR-induced deletions, expected band sizes were 226 bp for primer pair 1 and 223 bp for primer pair 2. Lanes 1-12 represent different potential gene edited colonies (N=1).

The PCR product from colony 11 and the negative control were subjected to a PCR clean-up and resolved for longer on a 2% agarose gel to identify whether multiple products were present. More extensive resolution demonstrated that three separate bands were present within the PCR product from colony 11, although all three bands appeared to be smaller than the 223 bp control (Fig 5.10).
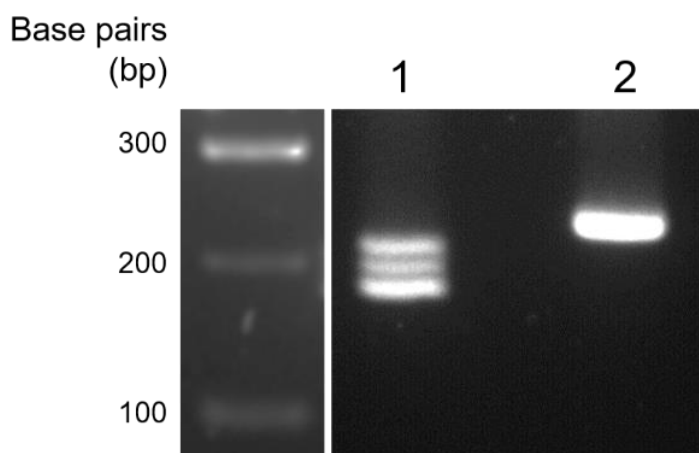
Fig 5.10. Higher resolution of the PCR product from colony 11. The PCR product visualised in Fig 5.9 was purified and resolved further on a 2% agarose gel against the negative control PCR product to identify the presence of multiple bands. The expected band size for the negative control was 223 bp. Lane numbers represent the following: 1 – colony 11 PCR product, 2 – negative control (N=1).

The three bands present within the PCR product of colony 11 were extracted from the gel and sent for Sanger sequencing. A sample of the resultant gel extracted, purified products was also analysed by gel electrophoresis to ensure the targeted bands had been successfully isolated (Fig 5.11).
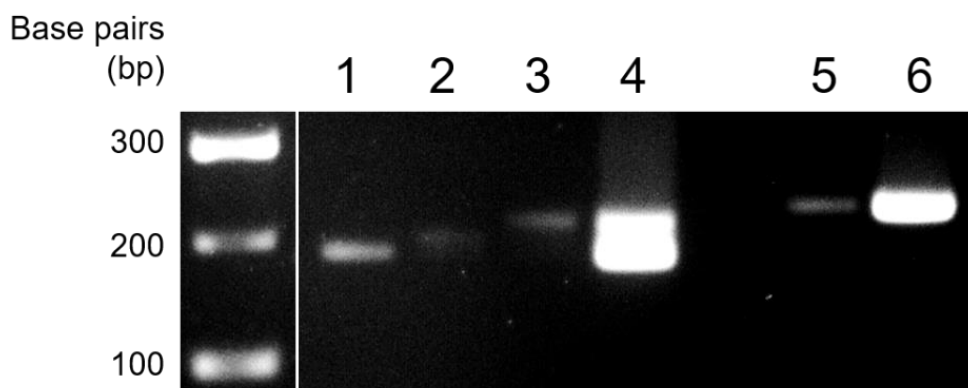
Fig 5.11. Resolution of three gel extracted bands from colony 11. The PCR product visualised in Fig 5.10 was subjected to a gel extraction to separate individual products from the same sample. Extracted products were resolved on a 2% agarose gel to ensure they had been successfully isolated. Lane labels represent the following: 1 – colony 11 fragment 1, 2 – colony 11 fragment 2, 3 – colony 11 fragment 3, 4 – unseparated bands of colony 11 PCR product, 5 – PCR clean-up of negative control, 6 – original negative control PCR product (N=1).

The sequencing results showed that fragment 1 i.e. the smallest band (Fig 5.11, lane 1) had been successfully gene edited, as a 46 bp deletion in the gDNA was identifiable (see schematic Fig 5.12). Approximately 25 bp of exon 2 was excised inducing a premature stop codon (Fig 5.13), in addition to 21 bp at the 3' end of intron 1. The deletion of 3' intronic DNA included the conserved 3' AG splice site.

The sequences derived from fragment 2 and 3 i.e. the intermediate and largest bands (Fig 5.11, lane 2 and 3) also showed that a deletion event had occurred, with 5 and 2 nucleotides being removed from the 3' terminus of intron 1, respectively (see schematic Fig 5.12). As a result, both transcripts lacked the 3' AG splice site. When the sequence identity between fragments 2 and 3 compared to WT disappeared, the remaining sequence showed no sequence similarity with WFDC2; in fact, when the sequence was inputted into BLAST software (NCBI blastn Suite), no sequence identity with any known sequence could be identified. This suggests that the additional nonsense sequence is an artefact of the PCR reaction and is due to the random assignment of nucleotides to the nascent PCR product.
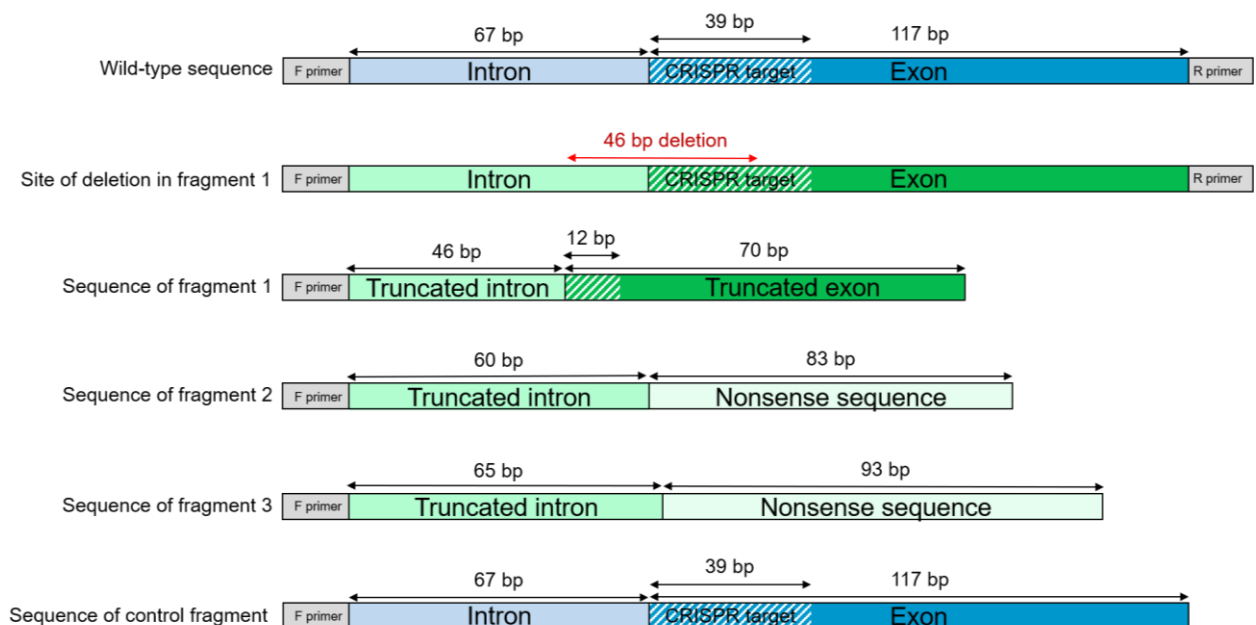
Fig 5.12. Schematic diagram of sequencing results of CRISPR edited colony 11 cells. Sequencing results of individual PCR products derived from colony 11 are summarised in a schematic diagram. Fragment 1-3 represent the individual bands shown in Fig 5.11, Lane 1-3, and the WT and control sequences are included. Sizes of relevant sections of the DNA in question are shown (bp).

Fig 5.13. Schematic diagram to show the truncation of exon 2 in fragment 1. The sequencing result of the fragment 1 PCR product derived from colony 11 was inputted into ExPASy translate tool (https://web.expasy.org/translate/) to identify the effect of the DNA truncation on the peptide sequence. WT peptide sequence is shown for reference.

The sequencing results derived from the CRISPR negative control showed that the gDNA was identical to that of wild-type WFDC2, thus showing that no mutations were present in WFDC2 of the negative control CAL27s.

5.7.2. WFDC2 expression in CRISPR edited cells

The CRISPR gene edited CAL27 cells were subjected to a number of assays to determine whether the deletions that occurred in intron 1 and exon 2 caused silencing of the WFDC2 gene. By carrying out RT-PCR using a forward primer targeting exon 1 and a reverse primer in exon 5, three different PCR products can potentially be amplified (V0, V1 and V2) depending on the cDNA profile derived from the CRISPR edited cells.

The PCR products derived from the CRISPR edited cells showed a ~150 bp shift in band size compared to WT and negative control cells (Fig 5.14). The band present in WT and control cells is consistent with the expected product for V0 (450 bp), whereas the band in CRISPR edited cells is similar to the expected band for V2 (306 bp).
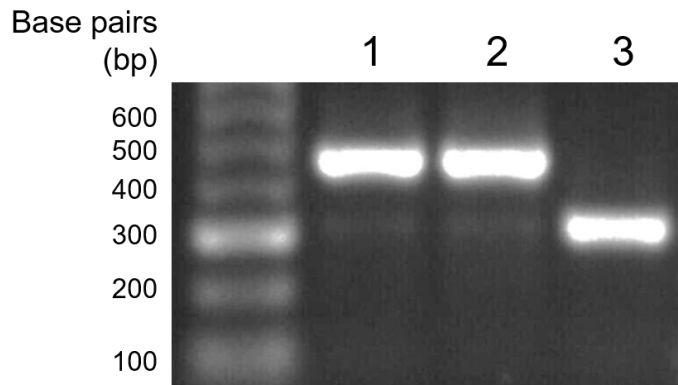
Fig 5.14. RT-PCR analysis of splice variants in CRISPR edited cells. Primers spanning the full-length of WFDC2 were used to probe cDNA derived from CRISPR edited cells and controls using RT-PCR. PCR products were resolved on a 1% agarose gel. Lanes represent the following: 1 – WT CAL27, 2 – CRISPR negative control cells, 3 – CRISPR edited cells. Splice variants that can be amplified by the primer pair include V0 (450 bp), V1 (579 bp) and V2 (306 bp) (N=3).

RT-PCR was also performed using primers specific to splice variant V0 and V2. OAZ1 was used as an endogenous control. The PCR products show that there is higher V0 expression in the WT and negative control compared to CRISPR cells (Fig 5.15). There appears to be greater expression of V2 in CRISPR cells, although positive expression of V2 is also visible in the control cells (Fig 5.15).
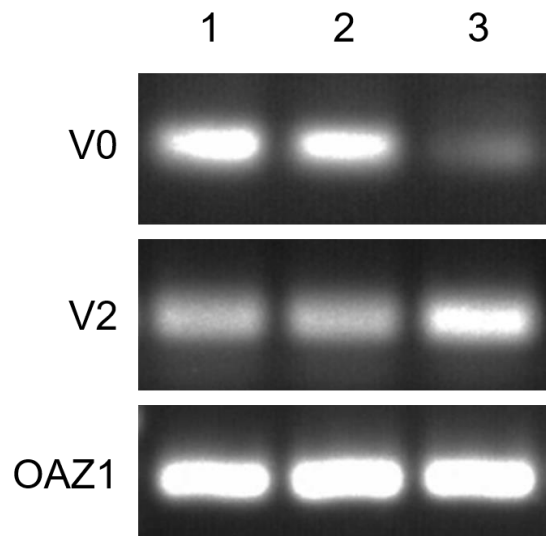
Fig 5.15. V0 and V2 expression in CRISPR edited cells by RT-PCR. Primers specific to splice variants V0 and V2 were used to probe cDNA from CRISPR edited cells and controls using RT-PCR. Products were resolved on a 2% agarose gel. Lanes represent the following: 1 – WT CAL27, 2 – CRISPR negative control cells, 3 – CRISPR edited cells. Band sizes were as follows: V0 – 105 bp, V2 – 97 bp, OAZ1 – 164 bp (N=3).

When WFDC2 expression was analysed by RT-qPCR using TaqMan primers against splice variant V1, the CRISPR gene edited cells were shown to express significantly less WFDC2 compared to both the CRISPR negative control (P=0.0137) and the WT CAL27 cells (P=0.002) evidencing that a considerable knockdown had been induced (Fig 5.16).
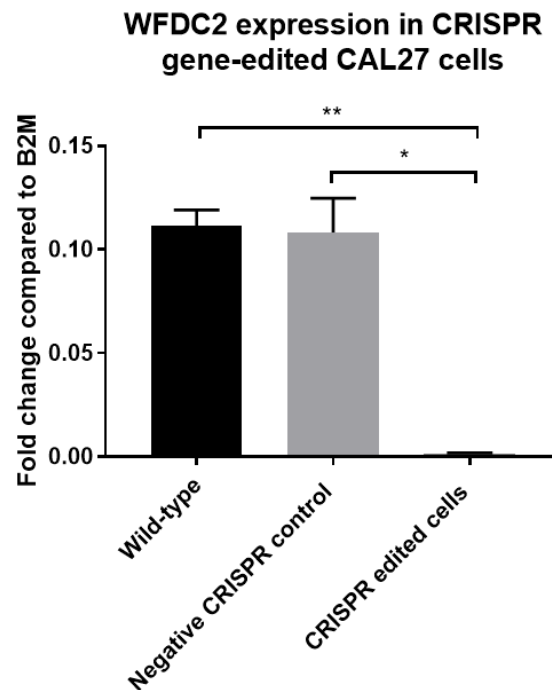
Fig 5.16. RT-qPCR quantification of WFDC2 gene expression in CRISPR gene edited CAL27 cells. WFDC2 expression in CRISPR gene edited cells compared to the negative control and WT CAL27 cells was analysed by RT-qPCR. WFDC2 expression is represented as the fold change in expression compared to the B2M endogenous control. Data represents the mean ± SD (N=3; n=3); ** $P \leq 0.01$, * $P \leq 0.05$.

Conditioned media derived from CRISPR gene edited cells was analysed by Western blotting and ELISA to determine whether the cells were still capable of successfully secreting WFDC2 protein. The serum-free conditioned media from gene edited, negative control and WT cells was concentrated, analysed by BCA and then protein content was standardised so that the same amount of total protein was analysed per cell phenotype.

The Western blot results demonstrated that the WT and CRISPR control cells were capable of producing WFDC2 protein, while the CRISPR gene edited cells were unable to secrete a detectable level of WFDC2 protein into the conditioned media (Fig 5.17).

The ELISA results showed that CRISPR edited CAL27 cells secreted no detectable WFDC2 protein into the conditioned media and therefore express significantly less protein than the WT and negative controls (P=0.007 and P=0.002, respectively) (Fig 5.18). There was no significant difference in protein production between the WT and negative control cells.
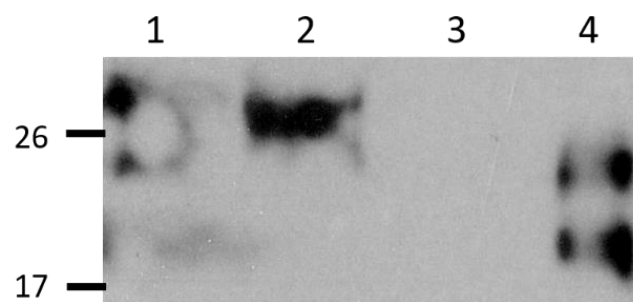
Fig 5.17. Western blot analysis of WFDC2 protein production in CRISPR gene edited CAL27 cells. Expression levels of WFDC2 in concentrated conditioned media from CRISPR gene edited cells compared to control cells were analysed by Western blotting using an anti-WFDC2 monoclonal antibody (N=2). Lane numbers represent the following: 1 – WT CAL27 conditioned media; 2 – CRISPR negative control conditioned media, 3 - CRISPR edited cell conditioned media, 4 – recombinant human WFDC2 control.
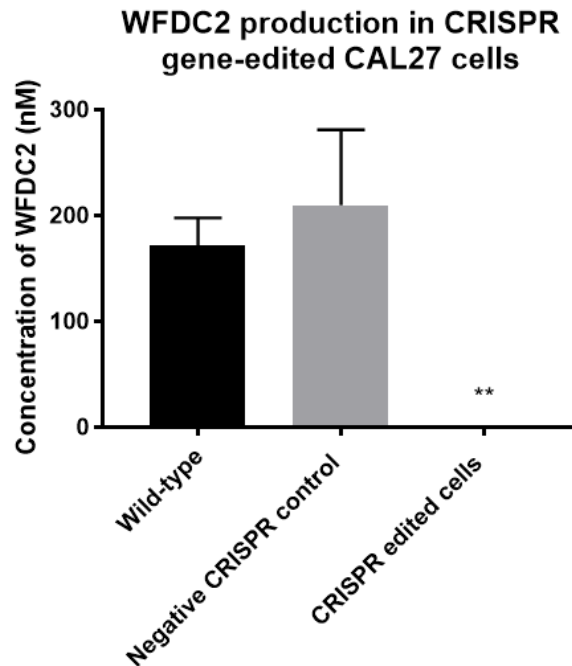
Fig 5.18. ELISA quantification of WFDC2 protein production in CRISPR gene edited CAL27 cells.
WFDC2 production in conditioned media from CRISPR gene edited cells compared to WT and CRISPR
negative control cells was analysed using an ELISA. Protein production is shown as the concentration
of WFDC2 (nM), with the mean ± SD (N=3, n=2). ** $P \leq 0.01$.

5.8.   CRISPR cell functional assays

The CRISPR gene edited CAL27 cells were subsequently used for a number of functional
assays, including proliferation, adhesion, migration and invasion assays, to determine whether
the CRISPR edited cells behaved differently to the CRISPR negative control and WT cells.

5.8.1. Cell morphology

The cells were visually assessed to determine whether the CRISPR editing had affected the
morphology of the cells. Comparison of the three cell phenotypes showed that there was no
striking difference in the cell appearance (Fig 5.19). The majority of cells in each group had a
rounded morphology and grew as colonies, while some cells showed a stellate morphology
and appeared to bridge gaps between colonies.

| Wild-type | Negative CRISPR control | Positive CRISPR edited cells |
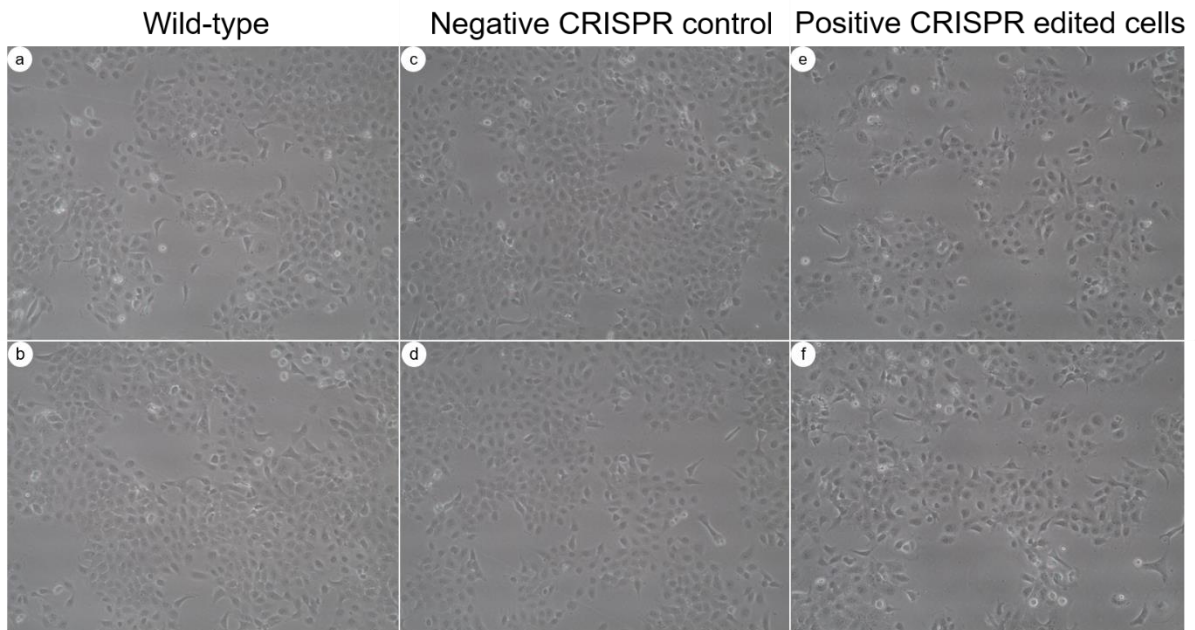
Figure 5.19. Phase-contrast microscopy images of CRISPR edited cell morphology versus controls. The morphology of CRISPR edited cells was analysed via phase-contrast microscopy and compared to WT and CRISPR negative controls. WT (a, b), CRISPR negative (c, d) and positive CRISPR edited cells (e, f) were imaged at an original magnification of 10X (N=3, n=4 fields of view, representative images shown).

5.8.2. Viability assays

To determine whether the CRISPR cells were less viable then the control cells, an MTS assay was performed. Cells were seeded into a 96-well plate at a density of 5,000 cells per well and cultured for 72 hours. MTS solution was then added and the linear relationship between the absorbance value and metabolic activity meant cell viability could be assessed.

The results show that significantly more viable WT cells were present in culture compared with the negative control cells, which showed an average reduction in absorbance of 24% (P=0.020) (Fig 5.20). There was no significant difference in absorbance between CRISPR edited cells and WT (P=0.064), or CRISPR edited cells and negative controls (P=0.61). CRISPR cells had an absorbance value that was 17% lower than the WT cells.
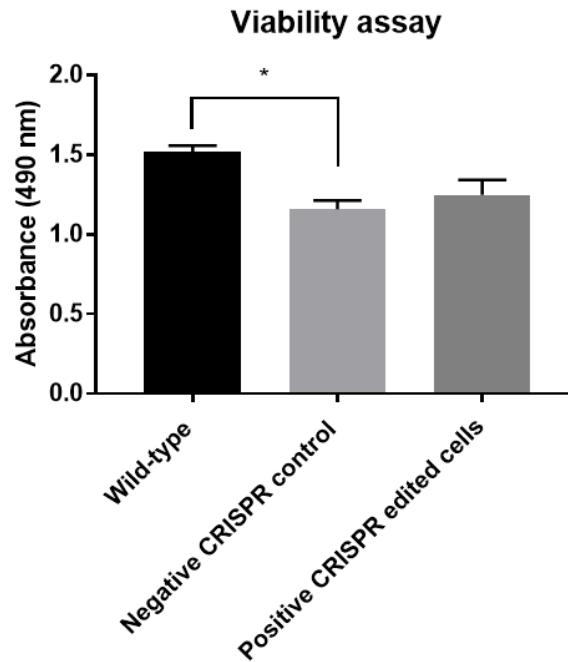
Fig 5.20. Viability assay using CRISPR gene edited CAL27 cells. The viability of CRISPR gene edited cells compared to WT and CRISPR negative control cells was analysed using an MTS assay. Absorbance correlates with the number of viable cells. Results show the mean ± SEM (N=3; n=3) * P ≤ 0.05.

To support the results determined from the MTS assay, cells were also manually counted. Cells were seeded into a 6-well plate at a density of 3×10^5 cells and incubated for 48 hours. The total number of cells were then counted using a haemocytometer. The results show that after 48 hours, significantly more WT cells were counted compared to the negative control (P=0.036). There was no significant difference in the number of CRISPR cells and either WT or negative control cells (P=0.0667 and P=0.876, respectively) (Fig 5.21).
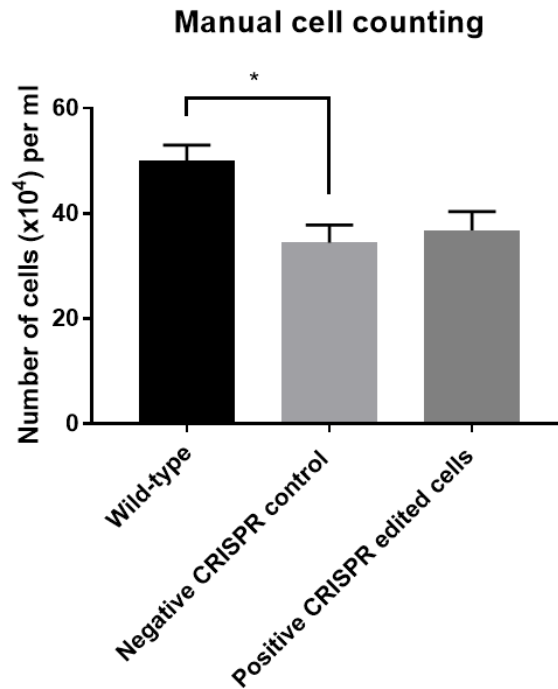
Fig 5.21. Manual counting of CRISPR gene edited CAL27 cells. The proliferation of CRISPR gene edited cells compared to WT and CRISPR negative control cells was analysed by counting the total cell number using a haemocytometers. The number of cells per ml are shown and results represent the mean ± SEM (N=3) * P ≤ 0.05.

5.8.3. Adhesion assay

The adhesive capacity of CRISPR edited cells was also tested by coating a 96-well plate with extracellular matrix (ECM) components, fibronectin and collagen I. Adherent cells that remained bound to the plate following a PBS wash were quantified using MTS reagent to determine the number of viable cells remaining in culture.

The results show that CRISPR edited cells exhibit no significant difference in their adhesive capacities compared to the WT and negative control (Fig 5.22). Both the CRISPR edited and CRISPR negative control cells appear to be slightly less adherent to the ECM components tested than their WT counterpart, however this difference is not significant.
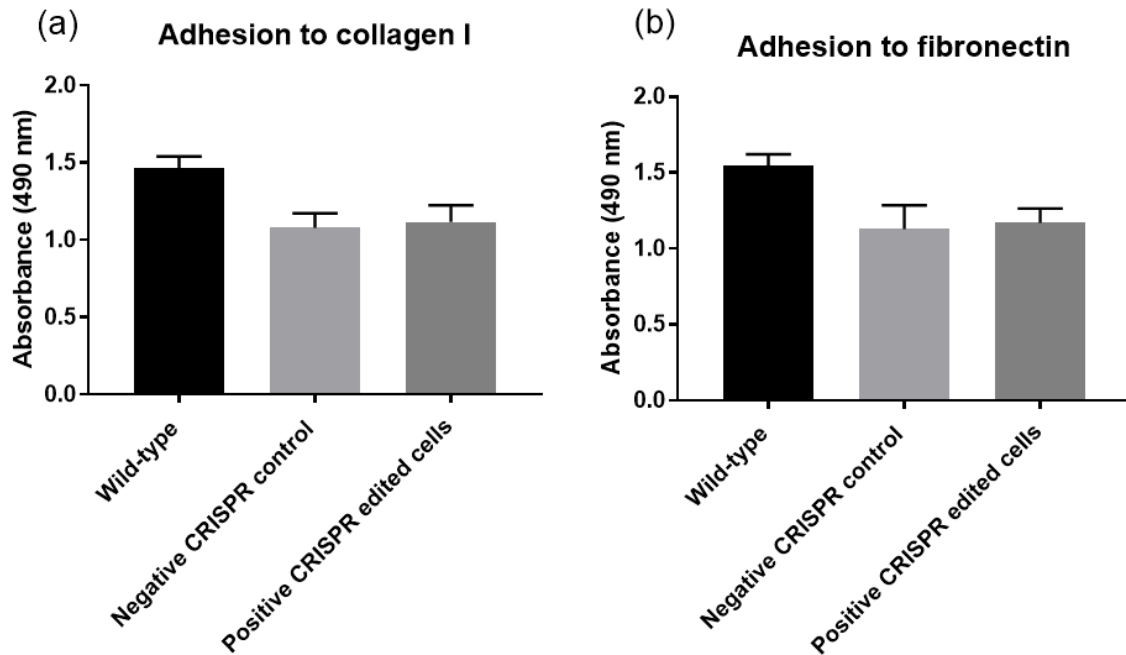
Fig 5.22. Adherence assay using CRISPR gene edited CAL27 cells. The adherence of CRISPR gene edited cells compared to WT and CRISPR negative control cells was analysed using collagen I (a) and fibronectin (b) coated 96-well plates. The number of adherent cells was measured via MTS assay. Absorbance correlates with the number of remaining (i.e. adhered) viable cells following a PBS wash. Results show the mean ± SEM (N=3; n=3).

5.8.4. Migration assay

The CRISPR edited cells were also assessed for any changes in migratory behaviour using Transwell assays. Cells were seeded into 8 µm pore Transwells and allowed to migrate for 7 hours toward serum-containing media beneath the Transwell which acted as a chemoattractant. Non-migrated cells were swabbed away and the cells that had migrated through the membrane were stained with crystal violet and manually counted.

The CRISPR negative control cells showed the highest level of migration, with significantly more cells per field of view compared to WT cells (P=0.011) and the positive CRISPR edited cells (P=0.0012) (Fig 5.23). Although the CRISPR edited cells appeared to show some reduction in migration compared to WT, this difference was not significant (P=0.114).
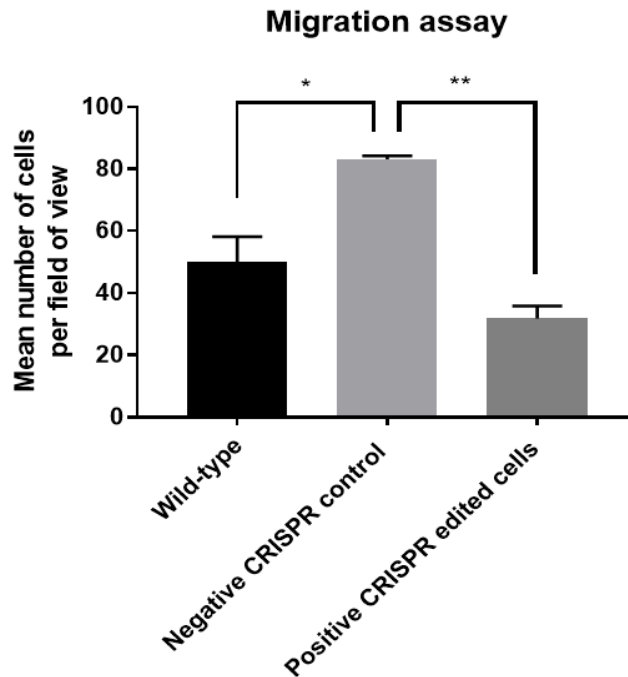
Fig 5.23. Migration assay using CRISPR gene edited CAL27 cells. The migration of CRISPR gene edited cells compared to WT and CRISPR negative control cells was analysed using Transwells, whereby migrated cells were stained and counted. The results are shown as the mean number of cells present per field of view taken from a total of 7 fields of view. Results show the mean ± SEM; (N=3; n=3) * P ≤ 0.05; *** P ≤ 0.001.

5.8.5. Invasion assays

To analyse the invasive capacity of the cells, 8 µm pore Transwells were coated with Matrigel, a basement membrane matrix, and cells were left to invade towards serum-containing media in the well below for 24 and 48 hours. Non-invaded cells were swabbed away and invaded cells were stained with crystal violet and then solubilised. The solubilised stain was measured via a microplate reader, the absorbance value represents the number of stained cells per insert.

The results show that the CRISPR edited cells had a slightly reduced invasive capacity after 24 hours compared to WT (P=0.660) and negative CRISPR control cells (P=0.364) but this difference was not significant (Fig 5.24a). After 48 hours, the CRISPR edited cells showed significantly reduced invasion compared to WT (P=0.0011) and negative control (P=0.0002)

cells (Fig 5.24b). The mean percentage reduction in invasion after 24 hours between CRISPR edited cells and WT or negative control cells was 13% and 17%, respectively.
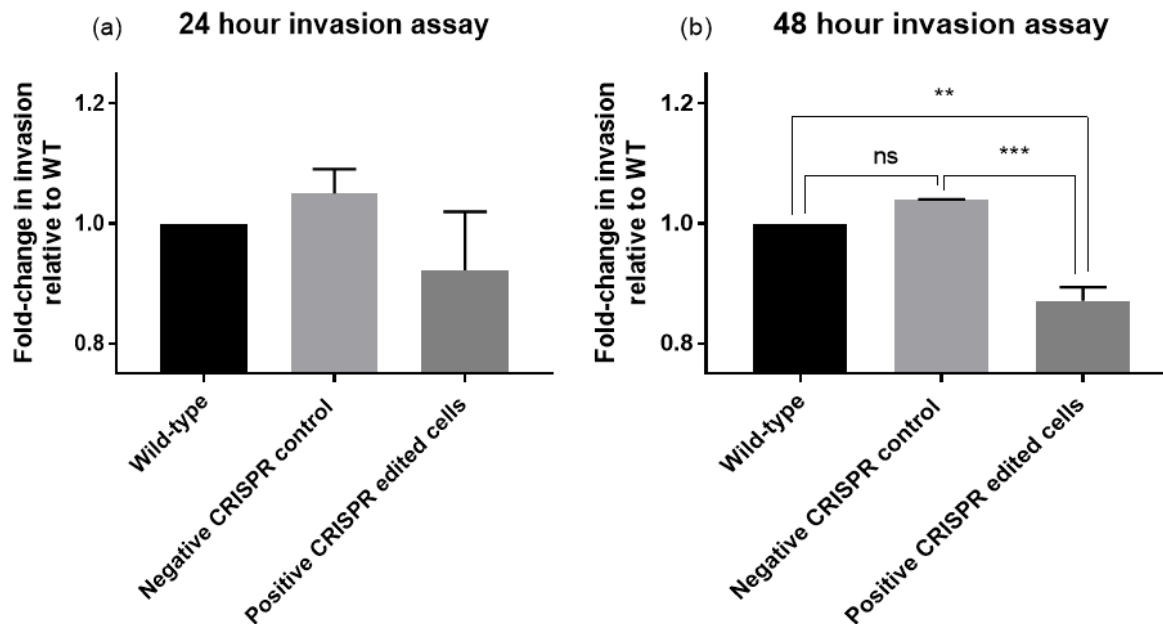
Fig 5.24. Invasion assay using CRISPR gene edited CAL27 cells. The invasion of CRISPR gene edited cells compared to WT and CRISPR negative control cells was analysed using Matrigel-filled Transwells, whereby cells were left to invade for 24 (a) or 48 (b) hours. Cells were then stained with crystal violet and solubilized. The absorbance of the solubilized cells was measured using a plate reader. Results show the mean ± SEM (N=3; n=3).

5.9. Discussion

The primary reason that WFDC2 is of interest to the scientific community is that its expression becomes exacerbated in ovarian cancer in addition to other malignancies including those of the lung and endometrium (see Section 1.4). When analysed histologically, the majority of tissues derived from gynaecological and pulmonary cancers show intense positive WFDC2 staining throughout the tumour. Some tissues exhibit more focal staining and, in these cases, positivity almost always aligns with abnormal ductular tissue (see Fig 1.11 and 1.12). Ovarian malignancies have been extensively analysed for WFDC2 expression, both in terms of quantifying serum levels and by histopathology, since the early 2000s. The expression of WFDC2 in other malignancies has drawn attention more recently, as serum levels of the

protein were also found to be elevated in lung cancer, calling into question the value of WFDC2 as a biomarker for non-gynaecological diseases.

WFDC2 expression in oral cancers is yet to be substantially defined in the literature. Within this study, IHC was used to analyse tissue sections of squamous cell carcinoma (SCC) of the tongue and mucoepidermoid carcinoma (MEC), with the aim of contributing to the understanding of WFDC2 expression in non-gynaecological malignancies. Oral SCCs make up 90% of the tumours of the oral cavity and the tongue is one of the most common sites, while MEC is the most common salivary gland malignancy in young adults and is most commonly found in the parotid gland (El-Naggar et al., 2017).

When analysed histologically, SCC of the tongue tissue was negative for WFDC2, with islands of tumour cells appearing entirely devoid of staining (Fig 5.1). The only sites of positive WFDC2 expression were the minor glands of the tongue, which appeared to have a normal ductal structure and most likely represented untransformed healthy regions of tissue. The SCC analysis suggests that transition from a healthy phenotype to a malignant one does not necessarily correlate with an upregulation of WFDC2. On the other hand, when WFDC2 expression in SCC of the lung was analysed by Bingle et al (2006) the results showed that many sections exhibited focal positive staining (18/60 cases), although the majority of cases (41/60) were negative. The comparison of SCC of the tongue and lung suggests that the presence of WFDC2 in SCC is site specific.

When MEC tissue was analysed in the same way, positive WFDC2 staining was identified in 36% of the tumour sections. Positivity was particularly abundant in mucous-secreting cells and in cystic spaces. Regions of phenotypically normal acini were identifiable and found to be mostly negative for WFDC2 which supports the evidence in the literature (Bingle et al., 2006). It is of interest to note that another salivary gland tumour that is characterised by cystic spaces and abundant mucous-producing cells, known as mucinous adenocarcinoma (MAC), experiences frequent amplification of the locus 20q13, upon which WFDC2 is expressed (Uchida et al., 2010; Wyszyńska-Pawelec et al., 2013). It is possible that the upregulation of

WFDC2 that occurs in MEC results from a similar amplification of the WFDC locus on chromosome 20.

It is interesting that healthy mucous cells within salivary glands are negative for WFDC2 (with positivity instead being almost entirely ductular) but following neoplastic transformation, the MEC tumour cells with a mucous-like phenotype express abundant WFDC2. Nevertheless, not all mucous-rich tumours exhibit high levels of WFDC2; for example, mucinous ovarian tumour sections exhibit negative or low focal WFDC2 staining and subsequently release comparatively low levels of WFDC2 into serum compared to other ovarian cancer types (Drapkin et al., 2005; Hertlein et al., 2012; Galgano et al., 2006).

The histological analysis carried out in this study has contributed to the understanding of WFDC2 expression in oral cancers but has also shown that the extent of WFDC2 positivity in cancer largely depends on the tumour site and not necessarily the phenotype of the tumour itself. The differences in WFDC2 expression between SCC of the tongue versus the lung, or mucinous tumours of the salivary gland versus the ovary, support this theory.

Within this study, analysis of WFDC2 expression was extended to investigate expression in a panel of ten cancer cell lines derived from the lung and oral cavity. Cancer cell lines are well established tools for analysis of cancer *in vitro* and are ideal targets for genome modifications given their homogeneity and indefinite replicative capacity. Cell lines derived from a number of tumour types, including ovarian, pancreatic and endometrial cancers, have been utilised for making WFDC2-overexpression and knockdown clones (see Section 1.4.5). Neither oral- or lung-derived cancer cell lines have been used as tools to analyse WFDC2 function prior to this study.

When analysed by traditional RT-PCR, the V0 splice variant showed the most abundant expression in CAL27, H357, NCI-H292 and NCI-H441 cell lines. A similar trend was found when analysing splice variants, where the aforementioned cell lines expressed the highest levels of V1-V4. Cell lines with low V0 expression, such as FaDu and NCI-H647, also exhibited

lower expression of the V1-V4 isoforms, with some variants such as V2 being absent altogether. In the case of cell lines that failed to express V0, such as NCI-H226 and NCI-H322, no other variants were expressed either. These results support data described by Jiang et al (2013) who report that V0 is significantly more abundant in human tissues samples (trachea, epididymis, breast and endometrium) than the other four isoforms (Section 1.2.2).

When RT-qPCR was used to quantify WFDC2 expression in the cancer cell line panel, CAL27 cells were found to express the gene most abundantly, with NCI-H441 cells exhibiting the second highest levels. When conditioned media was analysed by ELISA, CAL27 cells were shown to secrete significantly more WFDC2 than all other cells, secreting ~10-fold more WFDC2 than NCI-H441s. This result was supported by Western blot analysis, whereby WFDC2 was only detectable in the conditioned media of CAL27 cells.

The CAL27 cell line was established in 1982 from a squamous cell carcinoma of the tongue but has since been reclassified as an adenosquamous carcinoma following histopathological analysis and xenograft studies (Jiang et al., 2009). Adenosquamous carcinomas are characterised by their high invasive and metastatic capacity, with a worse prognosis than other types of SCC (El-Naggar et al., 2017).

There is no obvious correlation between WFDC2 expression levels and the cell line tissue origin. CAL27 and H357 cells are both derived from SCC of the tongue but exhibit disparate WFDC2 expression. NCI-H441 (the second highest WFDC2-expressing cell line) is derived from a papillary adenocarcinoma of the lung, a tumour which is characterised by early lymph node metastasis, so is similar to CAL27 in its enhanced capacity for invasion/migration (Aida et al., 2004; Miyoshi et al., 2003). NCI-H358, which expressed the third highest amount of WFDC2, is derived from bronchioalveolar carcinoma of the lung, which, unlike CAL27 and NCI-H441, exhibits a good prognosis and overall survival (Okubo et al., 1999). Cancer cell lines lose the original characteristics of their tumour source due to their growth as a clonal population in defined culture media without influencing factors from the tumour microenvironment (Wilding and Bodmer, 2014); this means that cell lines often have an altered

expression profile and behave differently to the original tissue. This factor most likely explains why CAL27 cells exhibit high WFDC2 expression when SCC of the tongue tissue sections are devoid of WFDC2 protein when analysed by IHC. Similarly, it may explain why two different cell lines derived from the same subtype of tumour (CAL27 and H357) express such different levels of WFDC2.

Cytokines are frequently implicated in the development of cancer as inappropriate cytokine expression can lead to chronic inflammation which can subsequently induce malignant transformation of cells (Hussain and Harris, 2007). Cytokines have also been implicated in the progression of cancer, whereby certain inflammatory mediators appear to mediate pro-tumour effects (Dranoff, 2004). When CAL27 and H357 were exposed to TNF-α and IL-1β, there was no significant effect on WFDC2 expression. This result supports the cytokine exposure assay described in Chapter 4 using HuSL cells and may suggest that WFDC2 is not significantly upregulated in response to inflammation; however, it is possible that CAL27 cells cannot exhibit any further upregulation due to having reached a maximal possible expression of WFDC2. TNF-α and IL-1β are known to mediate pro-cancer effects that are involved in processes such as metastasis and angiogenesis, therefore, if WFDC2 expression is not upregulated in response to cytokines, this suggests that the protein is not involved in such tumorigenic effects (Kim et al., 2009; Orosz et al., 1995; Voronov et al., 2003).

Interestingly, both H357 and CAL27 cells exhibited a significant downregulation of WFDC2 in response to IFN-γ. This result did not appear to be caused by IFN-γ-induced cell death, as the cellular B2M expression levels, which approximately correlate with cell number, were not decreased compared to untreated cells (see Appendix for CT values of B2M). IFN-γ is more of a mediator of immune responses compared to TNF-α and IL-1β (which are pro-inflammatory) and is involved in key immune processes such as activation of NK cells and macrophages and induction of antigen presentation via the major histocompatibility complex (Boehm et al., 1997). In cancer, IFN-γ has been shown to elicit anti-tumour activity by being pro-apoptotic, anti-proliferative and by enabling tumour surveillance due to its capacity for

stimulating antigen presentation (Dunn et al., 2004; Miller et al., 2009). Indeed, IFN-γ knockout mice are reported to be more susceptible to developing carcinogen-induced tumours (Osawa et al., 2006).

The correlation of high serum WFDC2 levels with poor prognosis in ovarian cancer patients, as well as reports that show increased proliferative and invasive capacities of WFDC2 overexpressing clones (reviewed in Section 1.4), suggest that WFDC2 is pro-tumorigenic. It is possible that IFN-γ downregulates WFDC2 in cancer (and subsequently in cancer cell lines) as an anti-tumour response.

It is also worthy of noting that a down-regulation of SLPI in response to IFN-γ has been observed in macrophages (Jin et al., 1997). The authors suggest that this is due to SLPI being upregulated in response to LPS where it can suppress the LPS-induced activation of NF-κB and its downstream pro-inflammatory pathways, meaning that a downregulation of SLPI by IFN-γ reinstates the defensive responses of macrophages. This is supported by literature showing that Slpi knockout mice are more susceptible to septic shock, suggesting that Slpi is involved in dampening excessive inflammation (Nakamura et al., 2003). It seems unlikely that WFDC2 would be downregulated for the same reason as SLPI as the evidence described in Chapter 4 suggests that the proteins do not elicit the same functions. It is more likely that WFDC2 plays a role in tumorigenesis and is thus downregulated in cancer cell lines by IFN-γ in an attempt to suppress tumour growth and invasion.

The main aim of the work described in this chapter was to identify the role of WFDC2 in influencing cancer cell behaviour. To do so, a cancer cell line with high endogenous WFDC2 expression, namely CAL27, was selected as a target for CRISPR gene editing and then used for downstream functional assays. Although siRNA knockdown WFDC2 clones have previously been described in the literature, WFDC2 CRISPR knockout cancer cells have not yet been used to study cancer cell behaviour. CRISPR has the potential to induce complete gene silencing, whereas a knockdown will only partially silence expression.

The CRISPR sgRNAs were designed to recruit the Cas9 nuclease to exon 2 of WFDC2, which would thus silence three out of five WFDC2 splice variants, including the most prevalent isoform, V0. Exon 2 encodes the N-terminal WFDC domain and also includes the site of N-glycosylation (Asn-44). The presence of a deletion in colony 11 was visualised on a gel where it exhibited a smaller size than the control colonies and was subsequently analysed by sequencing.

The three PCR products derived from colony 11 gDNA showed that all sequences lacked the 3'AG splice site of intron 1. The 3'AG is a requirement for splicing, therefore its deletion prevents the excision of intron 1 which subsequently excludes exon 2 from the resultant mature mRNA. The deletion of residues within exon 2 in PCR fragment 1 (see Fig 5.11 and 5.12) meant that a premature stop codon was also introduced, further supporting the notion that no WFDC2 isoforms containing exon 2 could form. As a result, the synthesis of mature mRNA encoding protein isoforms V0, V1 and V4 was unfeasible but successful splicing could still occur between exon 1 and 4 to form isoform V2, and exon 4 and 5 to produce isoform V3.

To further determine whether the CRISPR editing had been successful in silencing WFDC2, RT-PCR was carried out using cDNA derived from the CRISPR cells and controls. The primers spanned the whole gene transcript (targeting exon 1 and 5), meaning that they could amplify V0, as well as V1 and V2. The results showed that in the two controls, V0 was by far the most abundantly expressed isoform (450 bp). A faint second band was also visible at ~300 bp which most likely represents V2 (306 bp). Interestingly, the CRISPR PCR product showed a single intense band at ~300 bp, with no other visible bands at 450 bp or 579 bp, suggesting that the CRISPR cells were no longer able to express V0 or V1 and instead directed all pre-mRNA into synthesising V2. When primers specific to V2 were utilised, the CRISPR cells expressed more V2 than the controls, suggesting that V2 expression is indeed exacerbated due to the silencing of exon 2-containing isoforms.

When specific primers designed to target V0 were used for RT-PCR the results showed that, although V0 expression was significantly more abundant in WT and control cells, some V0

expression was also detectable in the CRISPR cells. This result suggests that a mixed population of cells remained within the CRISPR colony despite every effort to expand the colony from a single mutant cell. Nonetheless, the low band intensity shows that successfully edited cells (which lack expression of V0, V1 and V4) predominate in the mixed population. This hypothesis is also supported by RT-qPCR data which shows a significant reduction of V1 expression in CRISPR cells compared to the controls, with CRISPR cells expressing 1% of the WFDC2 expression found in controls.

When WFDC2 expression was analysed at the protein level, the results were conclusive. Western blotting showed that when 45 µg of total protein was analysed, CRISPR cells secreted no detectable WFDC2 in their conditioned media, whereas the negative control cells continued to express protein. Similarly, when the concentration was analysed by ELISA, the CRISPR cells had no detectable WFDC2 in their conditioned media, whereas the controls had an abundance of WFDC2. The absence of protein is particularly interesting when compared to the expression of WFDC2 in the different WT cancer cell lines (Fig 5.6), as even the lowest expressing cell lines still secreted detectable levels of protein (the lowest concentration being 452 pM, secreted by NCI-H322 cells). Although the RT-PCR analysis suggested that some V0 expression may remain in the CRISPR cells, protein analysis appeared to show that the mRNA detected was not translated into protein. For this reason, the cells were subsequently used for downstream functional analyses. Behavioural assays were used to compare the growth characteristics of the CRISPR and control cells, with the aim of determining the role of WFDC2 in influencing cancer cell behaviour.

When the metabolic activity and proliferative capacity of the cells was assayed using MTS reagent and cell counting, both the negative control and CRISPR cells exhibited reduced proliferation compared to WT. This is most likely due to both groups having undergone transfection, FACS sorting and single cell expansion, all of which are harsh treatments that may affect cell viability. This evidence suggests that, contrary to the data published in the

literature, WFDC2 does not increase cell proliferation and thus it is unlikely to propagate tumour growth directly.

The effect of WFDC2 on cell adhesion was tested against collagen I and fibronectin, two main components of the ECM, and no significant difference in adhesion was identified between the CRISPR cells and controls. Given that adhesion molecules, e.g. integrins, are key to influencing migration and invasion, if WFDC2 was capable of inducing changes to cell motility its effects would be unlikely to be mediated by adhesive mechanisms. The disparity in the number of adhered WT cells versus the negative control and CRISPR cells most likely arises from the reduced viability of the latter groups.

The CRISPR cells were further analysed for changes in cell motility. The WFDC2-knockout cells appeared to show a reduced capacity for migration, although the difference compared to the WT control was not statistically significant. Interestingly, when analysed for changes in invasive capacity, the CRISPR edited cells showed a statistically significant reduction in invasion compared to the control cells; a mean reduction of 13-17%. Given the aggressive and infiltrative phenotype of CAL27 cells, the fact that WFDC2-silencing has an inhibitory effect on their ability to migrate and invade may suggest that the gene exhibits pro-metastatic activity. It is feasible that if a less invasive cell line was used for the CRISPR studies, the effect on cell motility may have been more pronounced.

The lack of extensive behavioural changes identified from the CRIPSR cell analysis was surprising, especially given the WFDC2-mediated effects reported in the literature using siRNA knockdown clones, which express considerably more WFDC2 compared to the CRISPR cells engineered in this study. The siRNA clones are reported to exhibit significantly reduced proliferation, migration and invasion rates (Guo et al., 2015; Lu et al., 2012; Zhu et al., 2016); indeed, the siRNA knockdowns induced ~50-60% reduction in invasion compared to controls (Lu et al., 2012; Guo et al., 2015). While the CRISPR cells in this study did show changes in migratory and invasive activities, the key difference between the CRISPR cells and the siRNA clones described in the literature is their disparate proliferation rates.

When considered more closely, some of the assays reported in the literature show only moderate differences between clones with modified WFDC2 expression and their controls. For example, Zhu et al (2016) uses both WFDC2-overexpressing and WFDC2-knockdown clones to analyse changes in viability using an MTT assay. Although there is a statistically significant difference in absorbance readings between groups (overexpressing cells have a higher proliferation rate than the mock control and knockdown cells have a lower proliferation rate), the error bars overlap and mean OD readings are close together, suggesting that the result is not biologically significant. Li et al (2012) describes significantly higher proliferation in WFDC2-overexpressing cells compared to mock cells, yet their MTS assay results are plotted as 'a ratio of DNA content' and no error bars are shown. The unconventional expression of this data perhaps calls its reliability into question. Data described in Gao et al (2011) and Lu et al (2012) support the findings in this study, as they describe no significant changes in cell viability/proliferation of WFDC2-overexpressing cells when assessed against controls.

Discrepancies in the literature are also notable in xenograft studies. WFDC2-overexpressing ovarian cancer cells have been used to induce tumour development via subcutaneous injection in two separate reports; however, the resultant tumours were significantly larger than controls when isolated and measured in the work of Zhu et al (2016), whereas Gao et al (2011) describe the opposite, with the cells giving rise to significantly smaller tumours than controls. Both studies used the same ovarian cancer cell line to synthesise overexpression clones.

Further investigations will be required to more resolutely determine the role of WFDC2 in influencing cancer cell behaviour. The evidence discussed in this study suggests that by knocking out the expression of WFDC2, the cells have reduced invasive and migratory capacities. To further examine this hypothesis, it would be of interest to additionally edit the exon 2-knockout CAL27 cell line to cause silencing of exon 4. This would silence all 5 WFDC2 splice variants and could subsequently be used as a complete WFDC2-knockout cell line.

The fact that the CRISPR cells produced in this study still produce V2 mRNA means that it is entirely probable that some functional V2 isoforms were expressed by the CRISPR cells. The

hypothesised upregulation of V2 (shown in Fig 5.15) may have had an effect on the CRISPR phenotype, meaning that, although the cells could not secrete V0 protein, they expressed a compensatory level of V2 – an isoform expressing a WFDC domain. Unfortunately, this theory cannot be tested as, thus far, no antibody has been raised against V2. If the function of the C-terminal and N-terminal domain are the same, then the C-terminal domain in V2 may act in a compensatory fashion when V0 is silenced. In conclusion, the effects on cell behaviour mediated by the V0-knockout were moderate, with cells appearing to show some reduction in invasion and migration. The limited behavioural changes may be due to the compensatory activity of V2. It should also be noted that in all of the WFDC2-knockdown models described in the literature, no reference is made to the fact that WFDC2 can be alternatively spliced and therefore the prospect of four isoforms continuing to be produced at normal, or indeed elevated, levels is not discussed.

Chapter 6

Analysis of the Wfdc2-knockout mouse phenotype

6.1. Introduction

Multiple *in vitro* and xenograft studies have been described in the literature with the overall aim of elucidating the function of WFDC2 (Gao et al., 2011; Li et al., 2013; Lu et al., 2012; Zhu et al., 2016), yet little has been gleaned from these studies regarding the role of WFDC2 in healthy tissues. One research tool yet to be utilised is the generation of Wfdc2-knockout mice. The analysis described in this chapter will compare the histopathological phenotype of tissues derived from heterozygous knockout and WT adult mice. The phenotype of homozygous knockout mouse embryos will also be studied to determine whether Wfdc2 plays a role in embryonic development. These studies will fill a gap in the literature, providing a novel perspective to the discussion surrounding the function of Wfdc2 in mice.

6.2. Aims and objectives

The aim of the work described in this chapter was to compare the histological phenotype of heterozygous Wfdc2-knockout and wild-type (WT) mouse tissues. Organs from WT and heterozygote adult mice were studied, firstly by analysing their tissue architecture by H&E staining, and secondly by assessing the extent of Wfdc2 expression within the tissues to determine whether the mutant phenotype resulted in a concomitant reduction in Wfdc2 protein production. Analysis of Wfdc2 staining in WT mice was important as no IHC data is available in the literature regarding murine Wfdc2 expression. The expression of Slpi, a closely related WFDC protein, was also quantified to determine whether its expression was upregulated in heterozygotes to compensate for any reduction in Wfdc2 expression. To assess the phenotype of homozygote Wfdc2-knockout mice, embryos were analysed by micro-computed

tomography (micro-CT) at the MRC Harwell Institute to identify any distinct structural defects in homozygotes, to establish a biological function for Wfdc2 in development.

6.3. Materials and methods

To access the relevant materials and methods utilised in this chapter, please refer to the sections below:

- Knockout mice tissues were processed as per the methods in Section 2.11.1 to 2.11.3.

- H&E staining was carried out as described in Section 2.11.5.

- Immunohistochemistry was performed by Jessica-Leigh Tallis (a Sheffield Undergraduate Research Experience scheme student, supervised by Hannah Armes) using the method described in Section 2.11.6.

- The antibodies used are summarised in Table 2.13 and positive staining was quantified using the method outlined in Section 2.11.8.

- Micro-CT analysis of homozygous embryos was performed by researchers at the MRC Harwell Institute.

6.4. Wfdc2-knockout mice

Wfdc2-knockout mice were established at the MRC Harwell Institute and were generated via CRISPR-induced mutation. A 2431 nucleotide region spanning intron 2 to intron 4 was deleted, thus removing two essential exons (namely exons 3 and 4) to induce a null allele. The homozygote knockout animals exhibited a lethal phenotype, whereby the animals died shortly after birth and prior to weaning, while heterozygote animals were viable.

6.5. Analysis of heterozygotes

6.5.1. Tissue architecture

Tissues derived from WT and heterozygous adult mice were processed, sectioned and then stained with haematoxylin and eosin (H&E) to visually analyse the structure of the tissues. Salivary gland, trachea, lung, stomach, gastrointestinal tract, kidney, testis, epididymis and ovary derived from six WT animals (three male, three female) and six heterozygotes (three male, three female) were analysed. The tissues were visually inspected for any apparent architectural differences or anomalies between phenotypes, in addition to looking for any fibrosis or ulceration, lymphocyte infiltration or reductions in specific cell types likely to express Wfdc2 e.g. secretory cells. The airways were studied for any thickening of alveolar/bronchial walls or any reduction in alveoli. Tissues of the gastrointestinal system were analysed for loss of villi or reduction in mucous-secreting cells.

Visual analysis of the WT and heterozygote tissues showed no discernible structural differences between the two phenotypes in any of the tissues studied. Figures 6.1 to 6.3 show representative images of H&E stained tissues derived from WT and heterozygote mice.
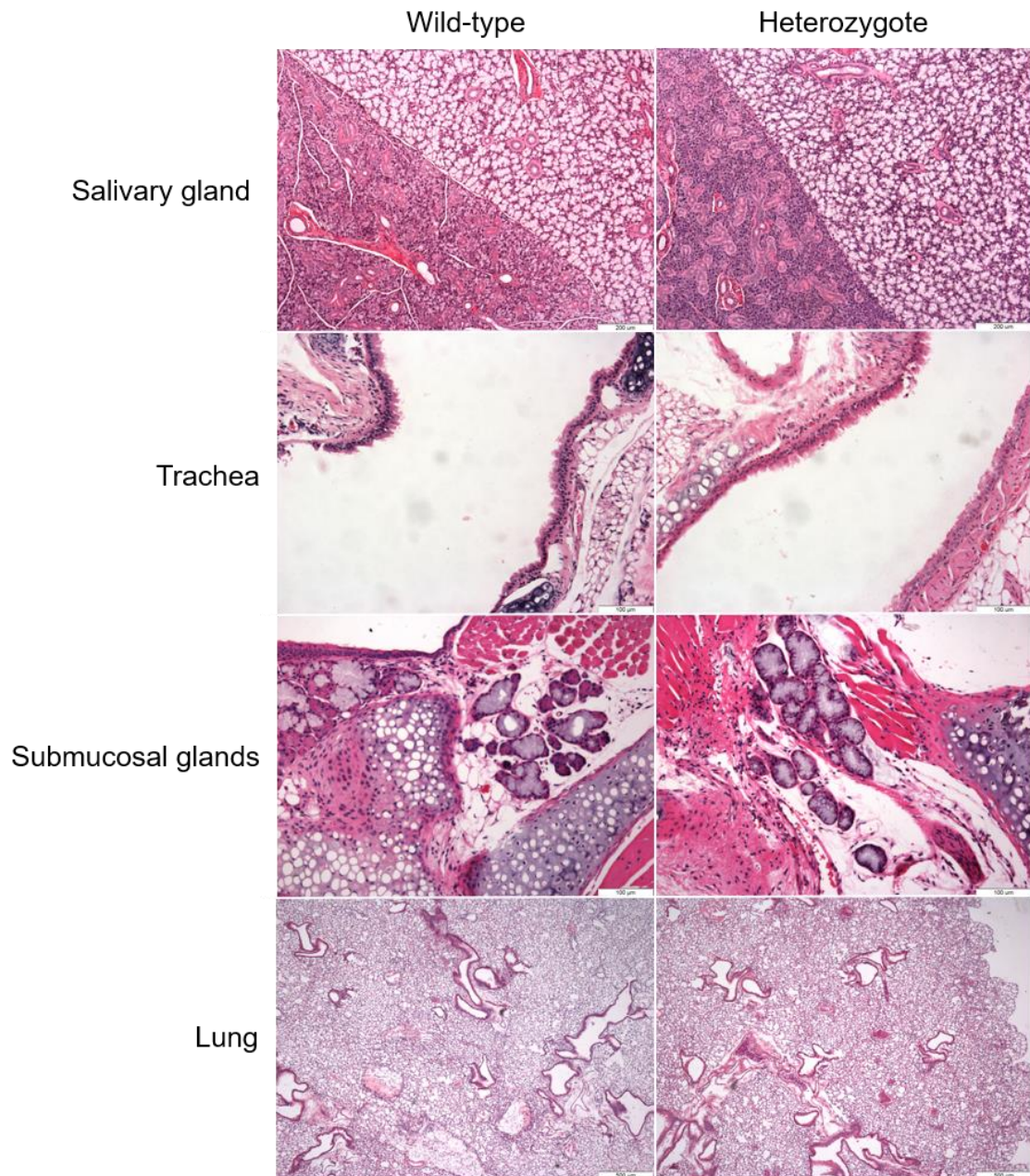
Fig 6.1. H&E stained tissues of murine salivary glands and respiratory tract. Tissue sections show H&E staining in WT (left panel) and heterozygote (right panel) murine salivary gland (10X magnification), trachea (20X), mucosal glands (20X) and lung (4X).
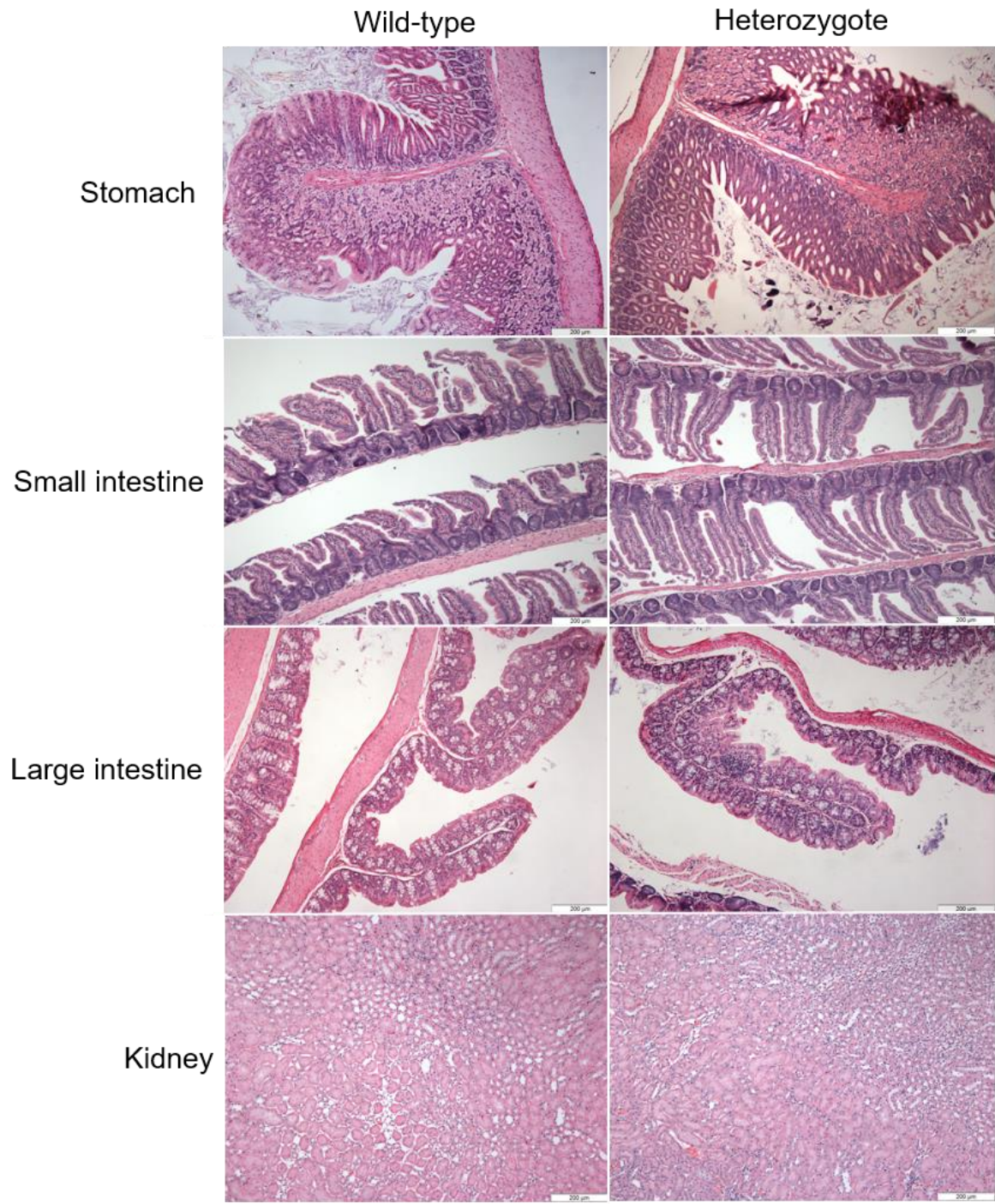
Fig 6.2. H&E stained tissues of murine gastrointestinal tract and kidney. Tissue sections show H&E staining in WT (left panel) and heterozygote (right panel) murine stomach (10X), small intestine (10X), large intestine (10X) and kidney (10X).
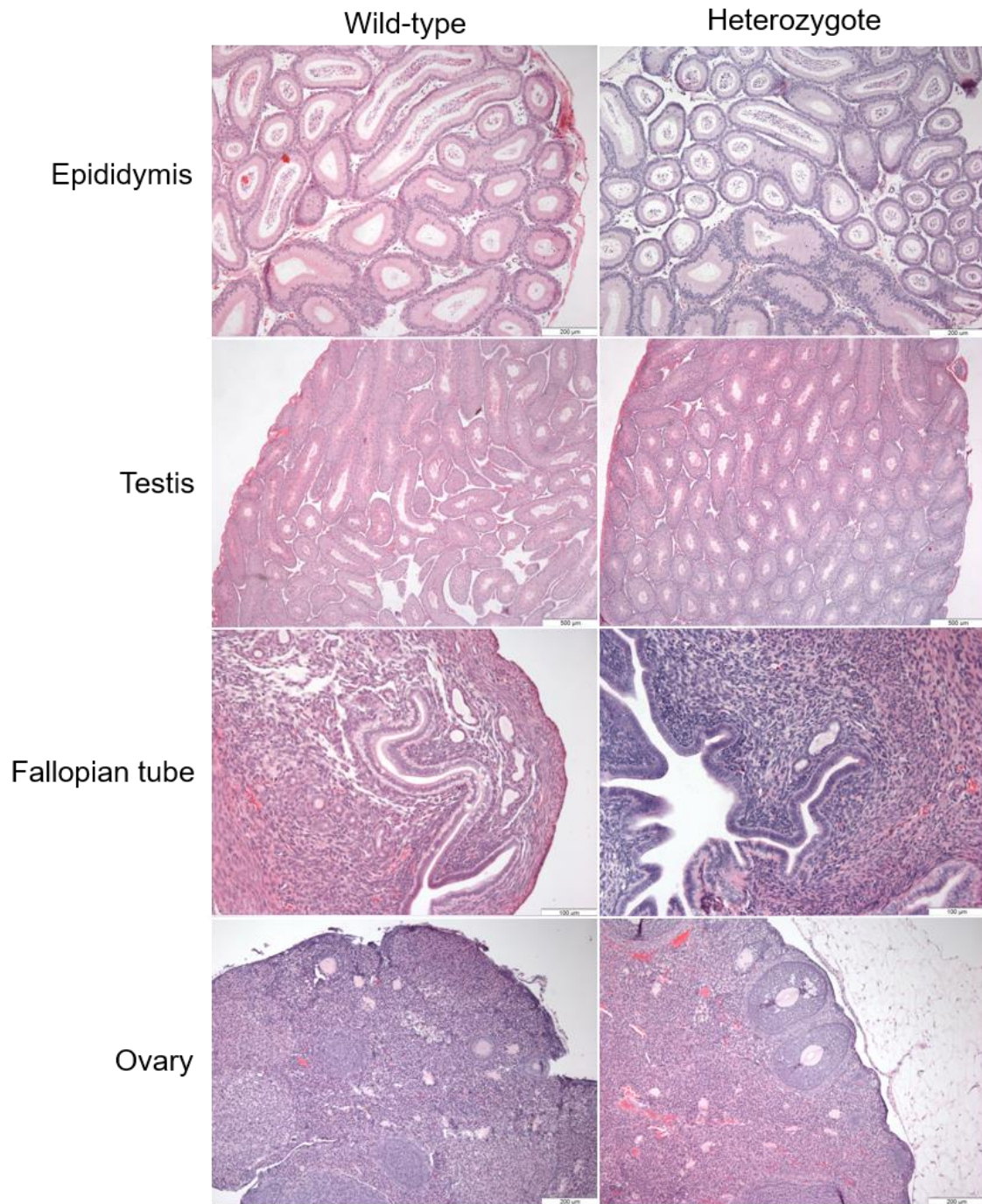
Fig 6.3. H&E stained tissues of murine reproductive tracts. Tissue sections show H&E staining in WT (left panel) and heterozygote (right panel) murine epididymis (10X), testis (4X), fallopian tube (20X) and ovary (10X).

6.5.2. Wfdc2 expression

Immunohistochemistry (IHC) was used to determine the localisation of Wfdc2 protein in numerous murine tissues. A polyclonal antibody raised against murine Wfdc2 was used to detect the protein. Due to the absence of data in the literature regarding murine Wfdc2 protein localisation, WT tissues were carefully analysed for positive Wfdc2 staining and compared to the array data available online (Su et al., 2002). Negative controls are shown in the Appendix.

Analysis of positive staining showed that the localisation of Wfdc2 protein is largely consistent with the gene expression data published on the BioGPS portal (see Fig 1.10 in Section 1.2.4). Salivary gland, trachea, lung, stomach, intestines, kidney, fallopian tube and testis all stained positively for Wfdc2 while epididymis and ovary were largely negative.

Wfdc2 was present in all major salivary glands. In the parotid gland, the protein was localised to the acinar cells and was notably absent in the ducts (Fig 6.4a). Conversely, in the sublingual gland (Fig 6.4b and Fig 6.4d) Wfdc2 was localised almost entirely to the ducts, although some positivity was visible in the serous demilunes. Staining was also predominantly found in the ducts of the submandibular gland (Fig 6.4c and Fig 6.4d). When positivity was quantified using Aperio ImageScope software and presented as a percentage of positively stained pixels, Wfdc2 was shown to be most abundant within the parotid gland, followed by the submandibular gland (Fig 6.5).
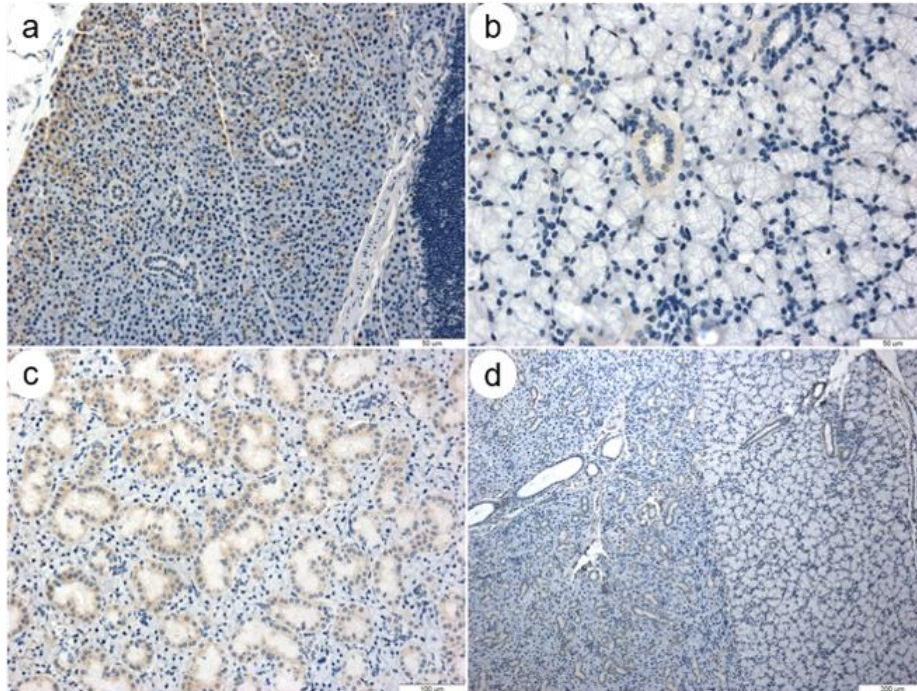
Fig 6.4. Localisation of Wfdc2 in murine salivary glands. Tissue sections show WT murine salivary gland tissue stained for murine Wfdc2 by IHC. Positive staining was identified in parotid (a), sublingual (b, d) and submandibular (c, d) glands. Fig 6.4d shows both sublingual (right) and submandibular gland (left) tissue. Original magnification of the images was 40X (a, b), 20X (c) and 10X (d).
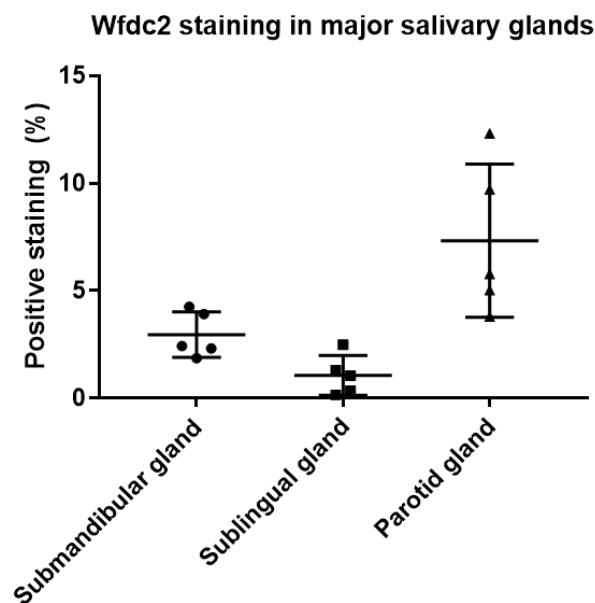
Fig 6.5. Quantification of Wfdc2 staining in murine salivary glands. Quantification of positive Wfdc2 staining of murine salivary gland tissue analysed by IHC. Quantification was based on the percentage of positively stained pixels from total pixels present determined using Aperio ImageScope software. Mean ± SD are shown, with data points representing individual Wfdc2 expression for each animal. Staining of negative controls were subtracted to account for background.

In the airways, intense Wfdc2 staining was found in the tracheal epithelium (Fig 6.6a and Fig 6.6b). Submucosal glands in the upper airways were also positive, with staining largely being localised to serous cells within the glands (Fig 6.6b). Positive staining continued down the trachea and into the large airways, with staining visible in the epithelium of bronchi and bronchioles (Fig 6.6c, Fig 6.6d). Epithelia of the alveoli were largely negative for Wfdc2, although some individual cells within the alveolar epithelium were Wfdc2-positive (Fig 6.6d).
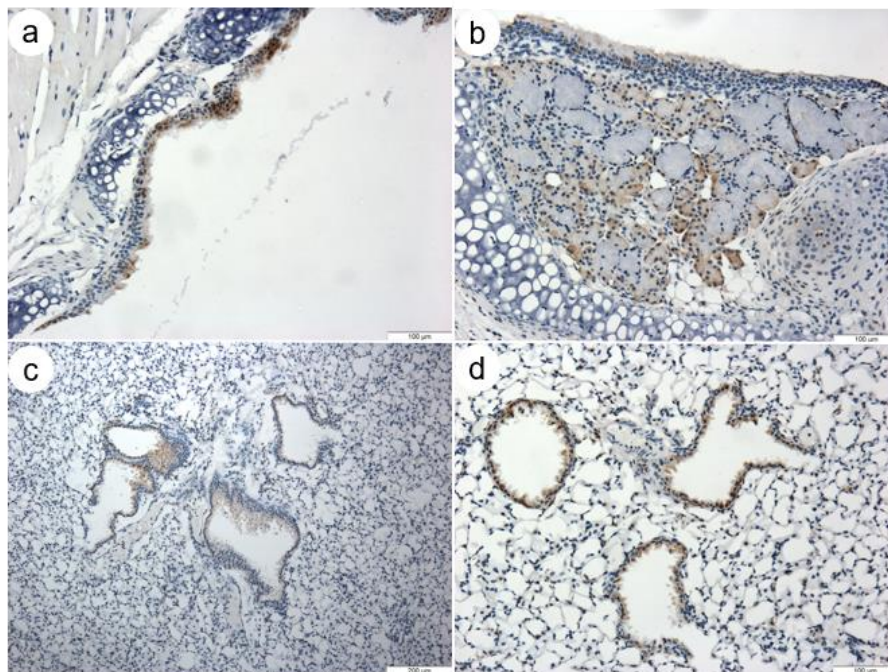
Fig 6.6. Localisation of Wfdc2 in murine airways. Tissue sections show WT murine trachea and lung tissue stained for murine Wfdc2 by IHC. Positive staining was identified in the trachea (a, b), submucosal glands (b) and bronchioles (c, d). Alveoli were largely negative for Wfdc2 (c, d). Original magnification of the images was 20X (a, b, d) and 10X (c).

When analysed more closely, positive Wfdc2 staining within the tracheal epithelium appears to co-localise with non-ciliated cells (Fig 6.7). The cilia project outwards into the lumen of the trachea while Wfdc2-expressing non-ciliated cells tend to be shorter.
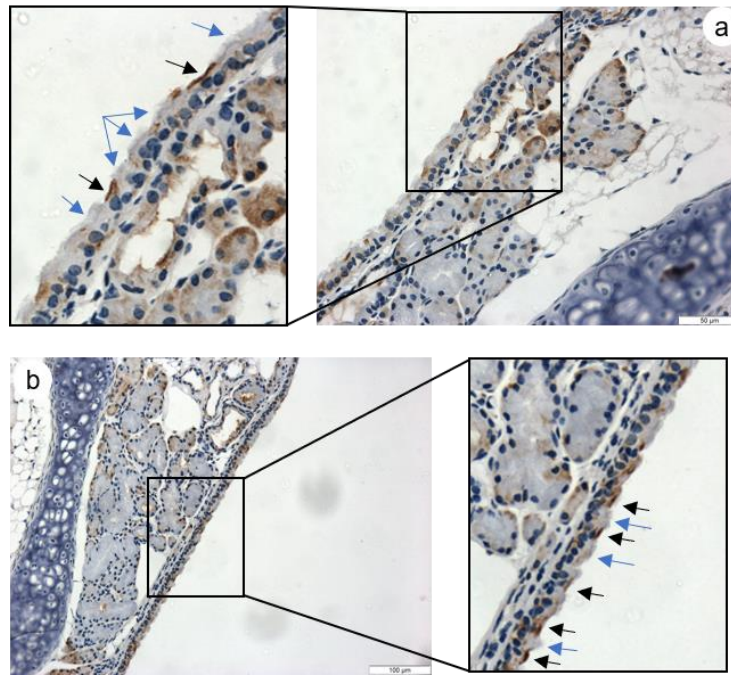
Fig 6.7. Localisation of Wfdc2 in the tracheal epithelium. Tissue sections show the WT murine tracheal epithelium (a, b) which has been stained for murine Wfdc2 by IHC. Positive staining was identified in non-ciliated cells (marked with a black arrow) while ciliated cells appear to be negative for Wfdc2 (marked with blue arrow). Original magnification of the images was 40X (a) and 20X (b).

Wfdc2 staining was abundant throughout the digestive tract. In the stomach, positive staining was seen within the gastric pits of the mucosa which open into the stomach lumen (Fig 6.8a, traverse section visible in Fig 6.8b). When moving laterally through the mucosa, from the lumen towards the submucosa, the centre of the mucosal tissue was largely negative for Wfdc2 staining; however, at the base of the mucosa Wfdc2 appeared to localise to chief cells (Fig 6.8a, Fig 6.8c).

In the small intestine, the brush border of enterocytes of the villi were moderately stained, while the internal lamina propria remained negative for Wfdc2 (Fig 6.8d-f). Interestingly, individual goblet cells scattered between the columnar enterocytes were intensely stained (as highlighted in the expanded area in Fig 6.8e). Crypts present at the base of the villi also appeared to exhibit moderate staining, again displaying marked staining of goblet cells (Fig 6.8d-f).

Wfdc2 staining was abundant within the mucosa of the large intestine, with positive staining visible throughout the tubular glands and crypts of the mucosa (Fig 6.8g-h). Goblet cells were intensely stained, a factor consistent with the small intestine. Positivity appeared to be most abundant at the base of the mucosal glands (Fig 6.8h). Traverse sections through the mucosa reiterate the abundance of positive mucous cell staining (Fig 6.8i).
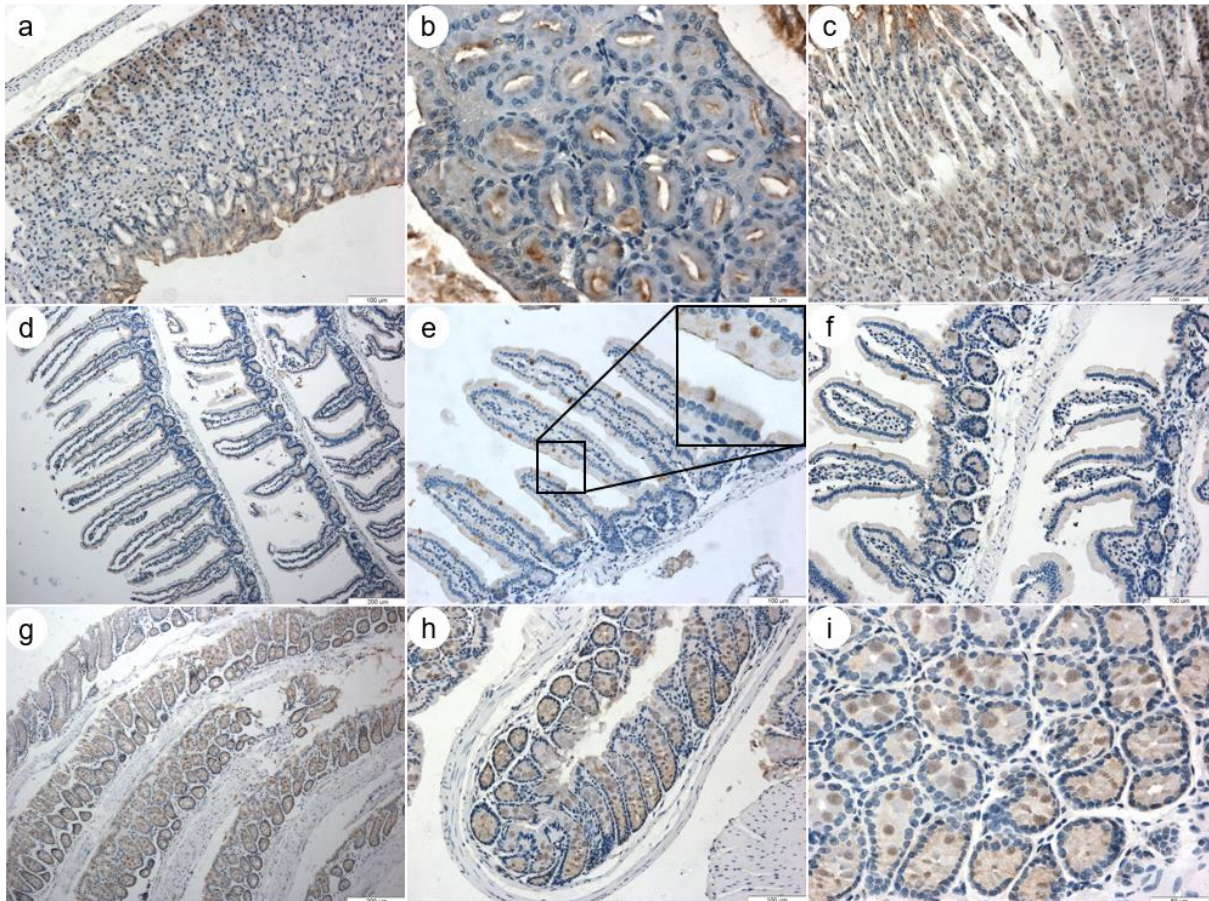
Fig 6.8. Localisation of Wfdc2 in murine gastrointestinal tract. Tissue sections show WT murine gastrointestinal tissue stained for murine Wfdc2 by IHC. Positive staining was identified in the stomach mucosa (a, b, c), small intestine (d, e, f) and large intestine (g, h, i). A magnified section of a villus in the small intestine is shown (e) so that individual goblet cells are visible. Original magnification of the images was 40X (b, h, i), 20X (a, c, e, f) and 10X (d, g).

Wfdc2 was moderately expressed within the kidney. Staining was visible throughout the convoluted tubules of the cortex (Fig 6.9a) and could be found within the lumen of the tubules (see magnified section in Fig 6.9c). The glomeruli were negative (Fig 6.9a) and the collecting tubules of the medullary ray were only weakly positive (Fig 6.9b).
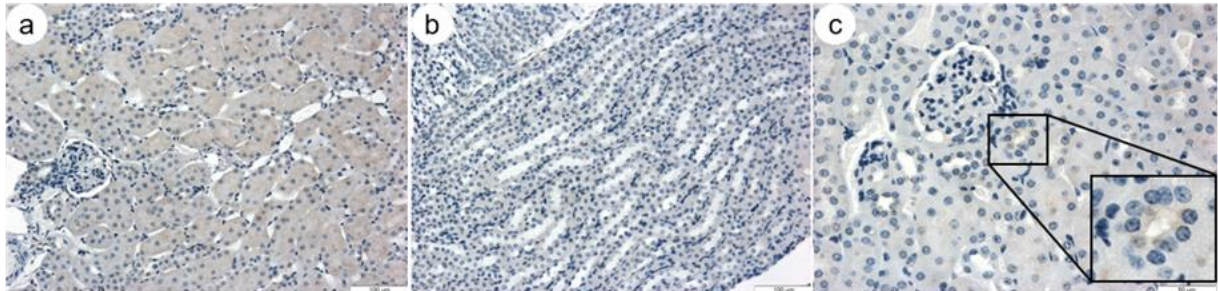
Fig 6.9. Localisation of Wfdc2 in murine kidney. Tissue sections show WT renal tissue stained for murine Wfdc2 by IHC. Positive staining was identified in the renal cortex (a, c) and moderate staining found in the medullary ray (b). Glomeruli were negative for Wfdc2 (a, c). A section of the renal cortex is magnified (c) to show the lumen of a convoluted tubule. Original magnification of the images was 40X (c) and 20X (a, b).

In the male reproductive tract, Wfdc2 expression was variable. In the testis, positive staining was abundant within the germ cells, although the late stage germ cells found in the lumen appeared to stain more intensely than the early stage germ cells seen at the basal layer (Fig 6.10a-b). The interstitial tissue, seminiferous epithelium and Sertoli cells were all negative for Wfdc2. Similarly, the epididymis appeared to lack Wfdc2 staining, although there was some faint positivity visible in the brush border of the columnar epithelium (Fig 6.10c-d).

In the female reproductive tract, Wfdc2 staining was also mostly negative. In the ovary there was some minor staining visible within follicles of different stages, but no staining was seen in the remainder of the tissue (Fig 6.11a-b). Conversely, significant staining was found in the columnar epithelium of the fallopian tube (Fig 6.11c).
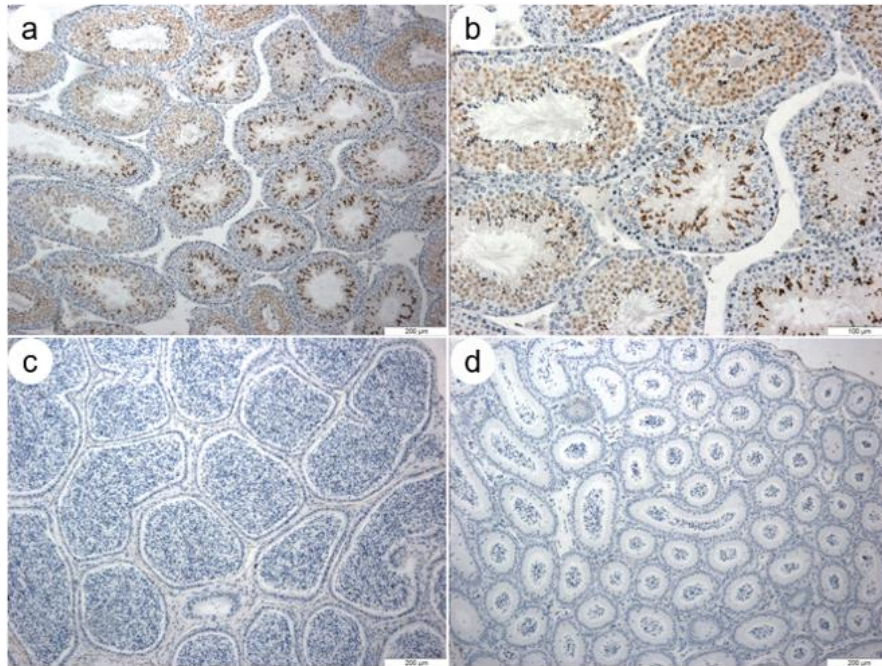
Fig 6.10. Localisation of Wfdc2 in murine male reproductive tract. Tissue sections show WT reproductive tract tissue stained for murine Wfdc2 by IHC. Positive staining was identified in the testis (a, b). The epididymis was largely negative for Wfdc2 (c, d). Original magnification of the images was 20X (b) and 10X (a, c, d).

Fig 6.11. Localisation of Wfdc2 in murine female reproductive tract. Tissue sections show WT reproductive tract tissue stained for murine Wfdc2 by IHC. The ovary was largely negative for Wfdc2 (a, b). Positive staining was identified in the fallopian tube (c). Original magnification of the images was 20X (b, c) and 10X (a).

Wfdc2 staining within mouse gastrointestinal tissues appeared to frequently co-localise with mucous or mucous-producing cells. When tissues of the gastrointestinal tract were stained with alcian blue, a dye which stains acid mucins, a similar staining pattern to that of Wfdc2 emerged. In the small intestine, mucous-secreting goblet cells stained positively for alcian blue within the epithelia of villi and in the crypts at the base of the villi (Fig 6.12a-b). In the large

intestine, positive blue staining was profuse due to the abundance of large goblet cells spanning the crypts of the mucosa (Fig 6.12c-d). This staining pattern correlates with Wfdc2 staining in the intestines (see Fig 6.8).
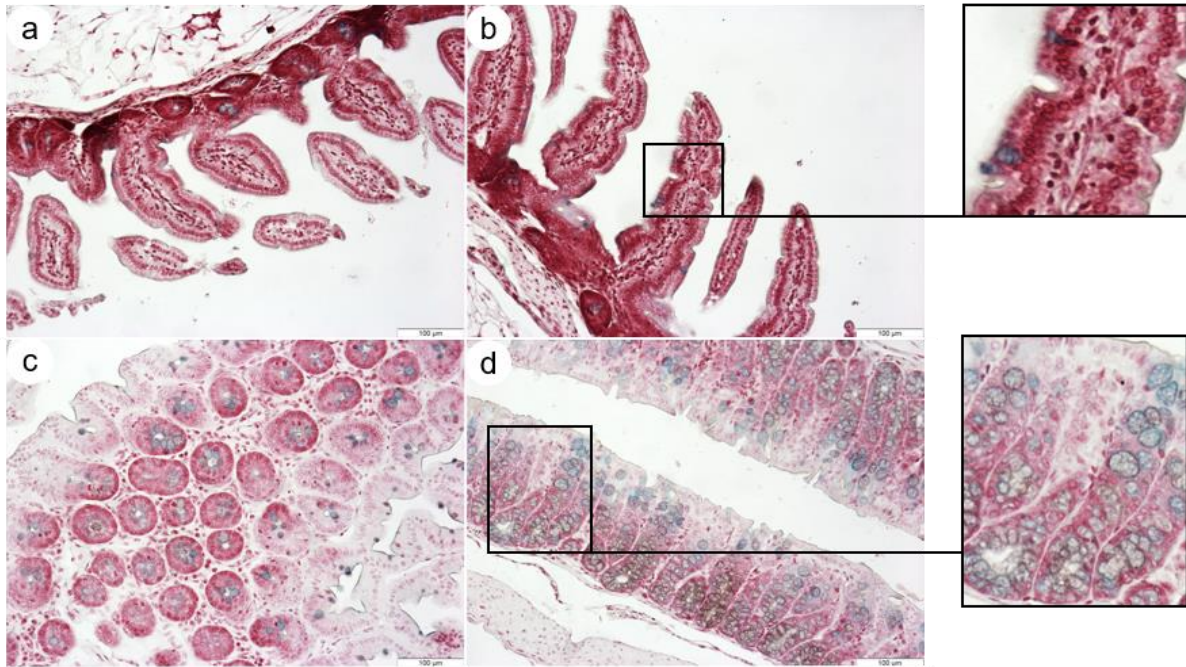
Fig 6.12. Alcian blue staining of murine gastrointestinal tract. Tissue sections show WT murine gastrointestinal tissue stained with alcian blue. Positive staining was identified in the goblet cells of the small intestine (a, b) and large intestine (c, d). Magnified sections of a villus within the small intestine (b) and of a tubular mucosal gland of the large intestine (d) are shown so that individual cells are visible. Original magnification of all images was 20X.

Although a trend of co-localisation between alcian blue and Wfdc2 was observed in tissues of the gastrointestinal tract, a similar trend was not seen in the respiratory tract or oral cavity. For instance, Wfdc2 was found to be present within the submandibular and sublingual glands (see Fig 6.4) but its pattern of expression was distinct to that of salivary mucins (Fig 6.13). Wfdc2 was specifically found in the ducts of the submandibular and sublingual glands, whereas alcian blue staining was notably absent from the ducts, instead positivity was seen in the mucous acini (Fig 6.13a and 6.13c).   When submandibular glands were analysed for mucin 5B (MUC5B) expression using IHC, the protein was abundant within the mucous cells (Fig 16.3d).

In submucosal glands of the upper respiratory tract, both alcian blue and MUC5B localised to mucous within the glands (Fig 6.13e-f). The staining pattern appeared to be distinct to Wfdc2 staining, which showed positivity in smaller serous glands rather than the obviously mucous-filled ducts (Fig 6.6b).
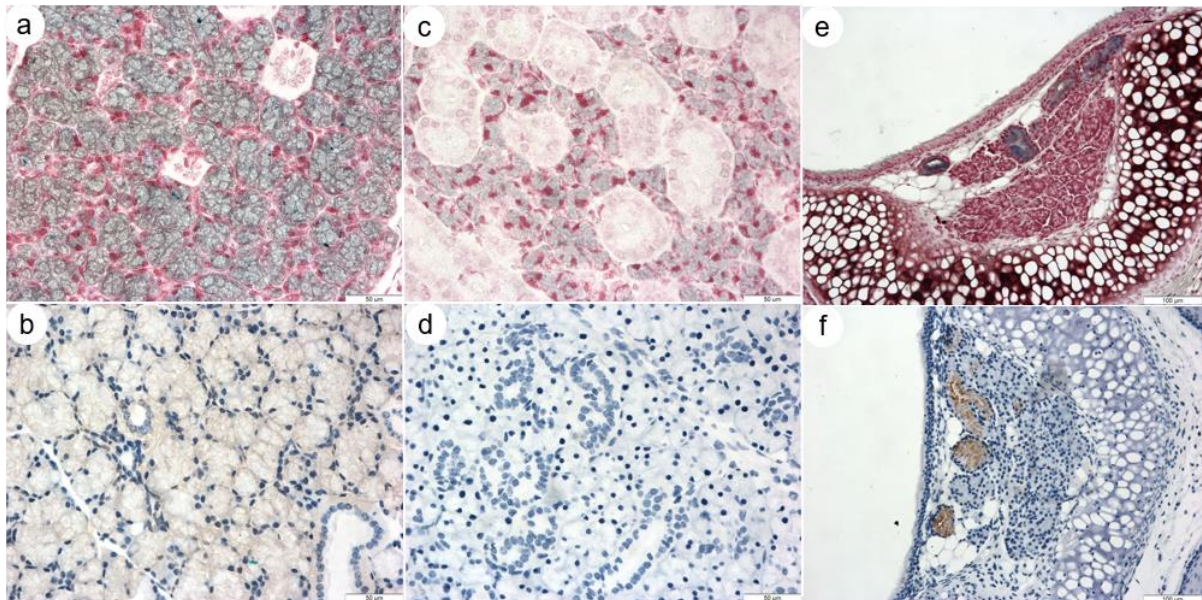
Fig 6.13. Mucous staining of murine salivary glands and submucosal glands. Tissue sections show WT murine salivary glands (a-d) and airway submucosal glands (e-f) stained with alcian blue (a, c, e) and stained for MUC5B (b, d, f). Positive alcian blue staining was identified in the mucous cells of the sublingual gland (a) and submandibular gland (b). Positive MUC5B staining was seen in the sublingual gland (c) but was absent in the submandibular gland (d). Both alcian blue and MUC5B localised to mucous cells of the submucosal glands (e, f). Original magnification of images was 40X (a-d) and 20X (e-f).

Wfdc2 is expressed in non-mucous secreting cells such as club cells, as evidenced by the co-localisation of Wfdc2 with club cell secretory protein (CCSP), which is the predominant secretory protein of club cells (Fig 6.14). Club cells are present within the tracheal and bronchial epithelium.
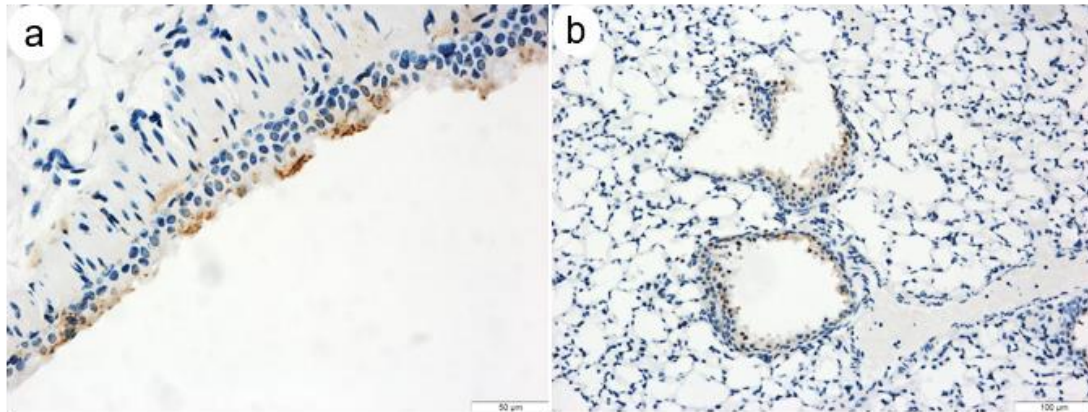
Fig 6.14. Localisation of CCSP in murine airways. Tissue sections show WT murine tracheal and lung tissue stained for murine CCSP by IHC. Positive staining was identified in the trachea (a) and bronchi (b). Original magnification of the images was 40X (a) and 20X (b).

6.5.3. Slpi expression

Given the frequent parallels drawn between the human orthologues of WFDC2 and SLPI, the expression profile of murine Slpi was studied via IHC to determine whether it co-localises with Wfdc2. The results showed that Slpi and Wfdc2 have a similar yet distinct pattern of expression.

In the respiratory tract, Slpi staining was present within the tracheal and bronchial epithelium (Fig 6.15a-b) and was largely absent in the peripheral lung. Likewise, in the oral cavity, Slpi staining was similar to Wfdc2 in the major salivary glands, with staining in the acinar cells of the parotid gland and in the ducts and serous demilunes of the sublingual glands (Fig 6.15d-e). In contrast to Wfdc2, the submandibular gland stained positively for Slpi in the acini as well as the ducts (Fig 6.15f).

Slpi staining within the stomach presented a similar pattern of staining to Wfdc2, with positivity at the surface of the mucosa within the gastric pits, in addition to staining within chief cells at the base of the gastric glands (Fig 6.15g). Interestingly, in the intestines, Slpi localisation was distinct to Wfdc2, largely due to its notable absence in goblet cells. In the small intestine staining was present within the brush border of the columnar epithelium of the villi and within

the intestinal crypts, while goblet cells within the crypts and villi were negative (Fig 6.15h). Minor Slpi staining was found in the large intestine, particularly at the mucosal surface epithelium. The crypts were largely negative for Slpi and goblet cells were negative (Fig 6.15i-j).

In the kidney, weak Slpi staining was visible throughout the convoluted tubules of the cortex and the glomeruli were negative (Fig 6.15k). In the male reproductive tract, Slpi was abundant in the brush border of the columnar epithelium of the epididymis. Staining was strongest at the initial segment and caput of the epididymis but weaker towards the cauda (Fig 6.15l-m). Slpi staining was absent in the testis (Fig 6.15n) and in the female reproductive tract, Slpi was also largely absent within the ovary and the fallopian tube (Fig 6.15o).
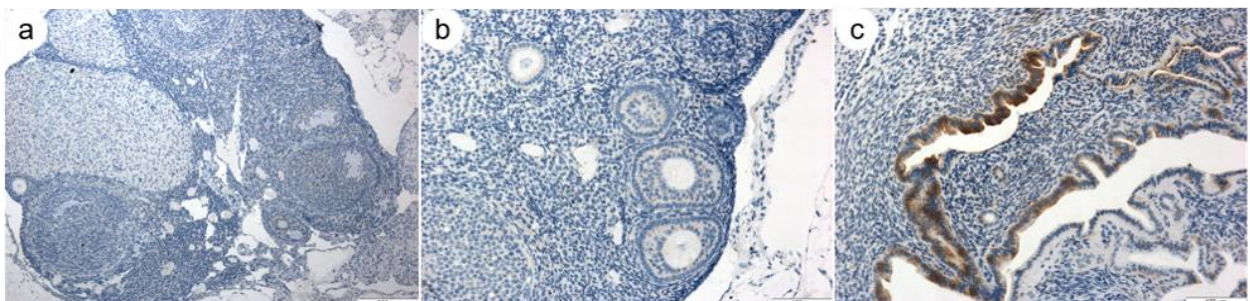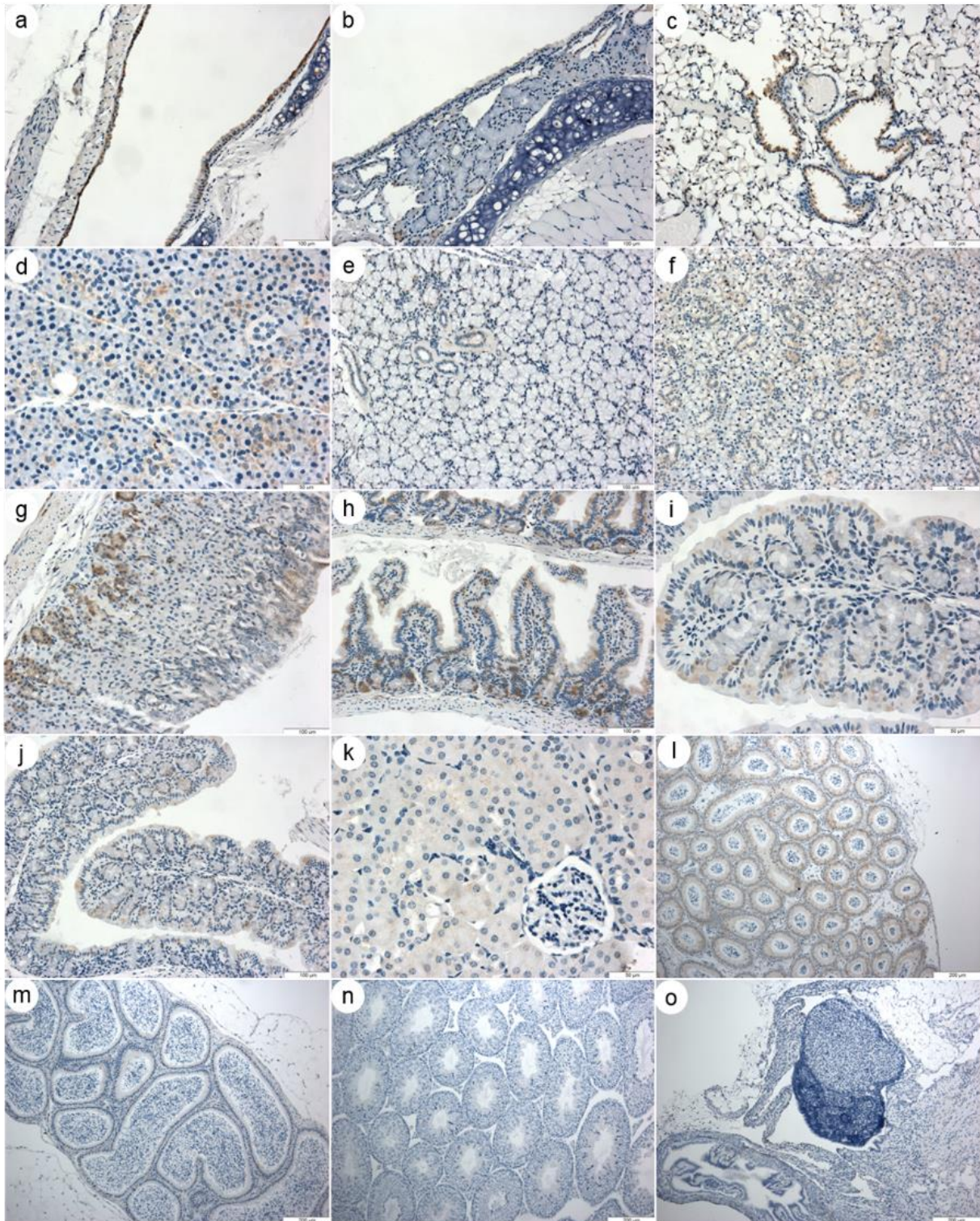
Fig 6.15. Localisation of Slpi in murine tissues. Tissue sections show WT tissues stained for murine Slpi by IHC. Positive staining was identified in the trachea (a, b), bronchi (c), parotid gland (d), sublingual gland (e), submandibular gland (f), stomach (g), small intestine (h), large intestine (i, j), kidney (k) and epididymis (l, m). Testis (n) and ovary (o) were negative for Slpi. Original magnification of the images was 40X (d, i, k), 20X (a, b, c, e, f, g, h, j) and 10X (l, m, n, o).

6.6.   Comparison of heterozygote and wild-type protein expression

To determine whether the heterozygous knockout mice, which by definition have a single copy of the Wfdc2 gene and thus theoretically express a 50% reduction in gene expression, resultantly produce less Wfdc2 protein, the heterozygous and WT tissues were analysed by IHC. Tissue sections from each animal were stained for Wfdc2 protein using a polyclonal antibody and then positive DAB staining was quantified using Aperio ImageScope software. Positive staining was presented as a percentage of positively stained pixels.

Quantification of positive staining demonstrated that there was no significant difference in Wfdc2 protein production between phenotypes (Fig 6.16), except for the large intestine which showed significantly higher staining in tissue derived from heterozygote animals. The percentage of positive staining was extensively variable between animals, regardless of phenotype. Higher mean percentages of positive staining were frequently seen in heterozygote tissues, including the submandibular gland, intestines and male and female reproductive tissues, although this difference was not significant.
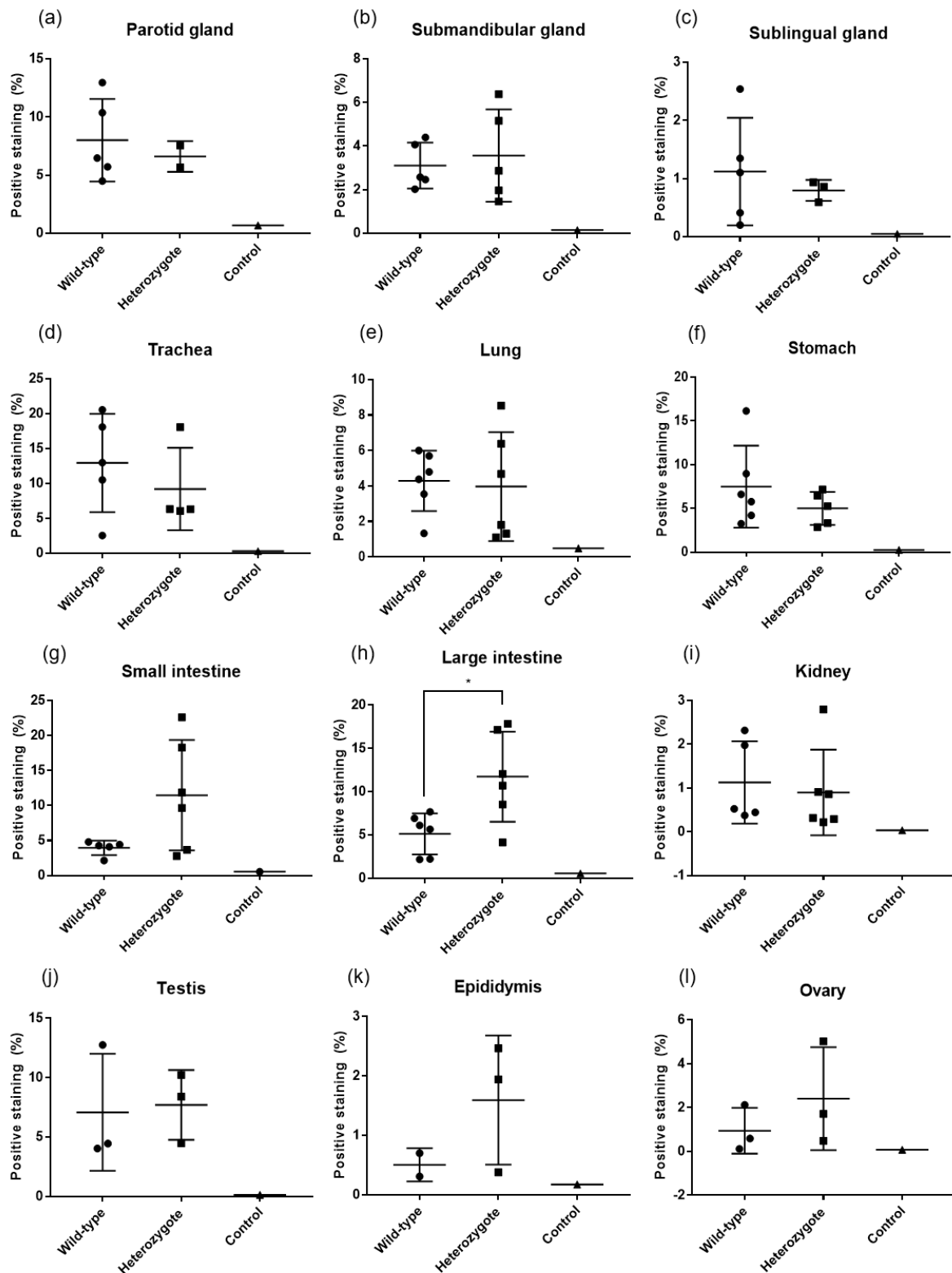
Fig 6.16. Quantification of positive Wfdc2 immunostaining in WT and heterozygote mouse tissues. Positive staining was quantified and presented as a percentage of positively stained pixels per tissue. Graphs show positive staining in WT and heterozygotes for each tissue analysed (a-l) and individual data points represent each animal. The control shows background when tissues were incubated with secondary antibody only. Data represents the mean ± SD; * P ≤ 0.05.

197

The expression of Slpi was also analysed to determine whether heterozygote animals produce more Slpi protein than WT animals to compensate for the effect of Wfdc2 silencing. IHC was carried out using a polyclonal antibody raised against Slpi and positive DAB staining was quantified using Aperio ImageScope as previously described. As testis, ovary and kidney tissues from both phenotypes were largely negative for Slpi, this staining was not quantified.

Quantification of positive Slpi staining demonstrated that there was no significant difference in Slpi protein production between tissues from WT and heterozygous animals (Fig 6.17). In the parotid gland staining was significantly higher in WT animals than heterozygotes, however the number of parotid tissue samples derived from heterozygotes was low (n=2) so the results are most likely due to the low sample number. Extensive variability in the percentage of positivity was visible between animals regardless of phenotype, particularly in the submandibular gland, trachea, intestines and epididymis.
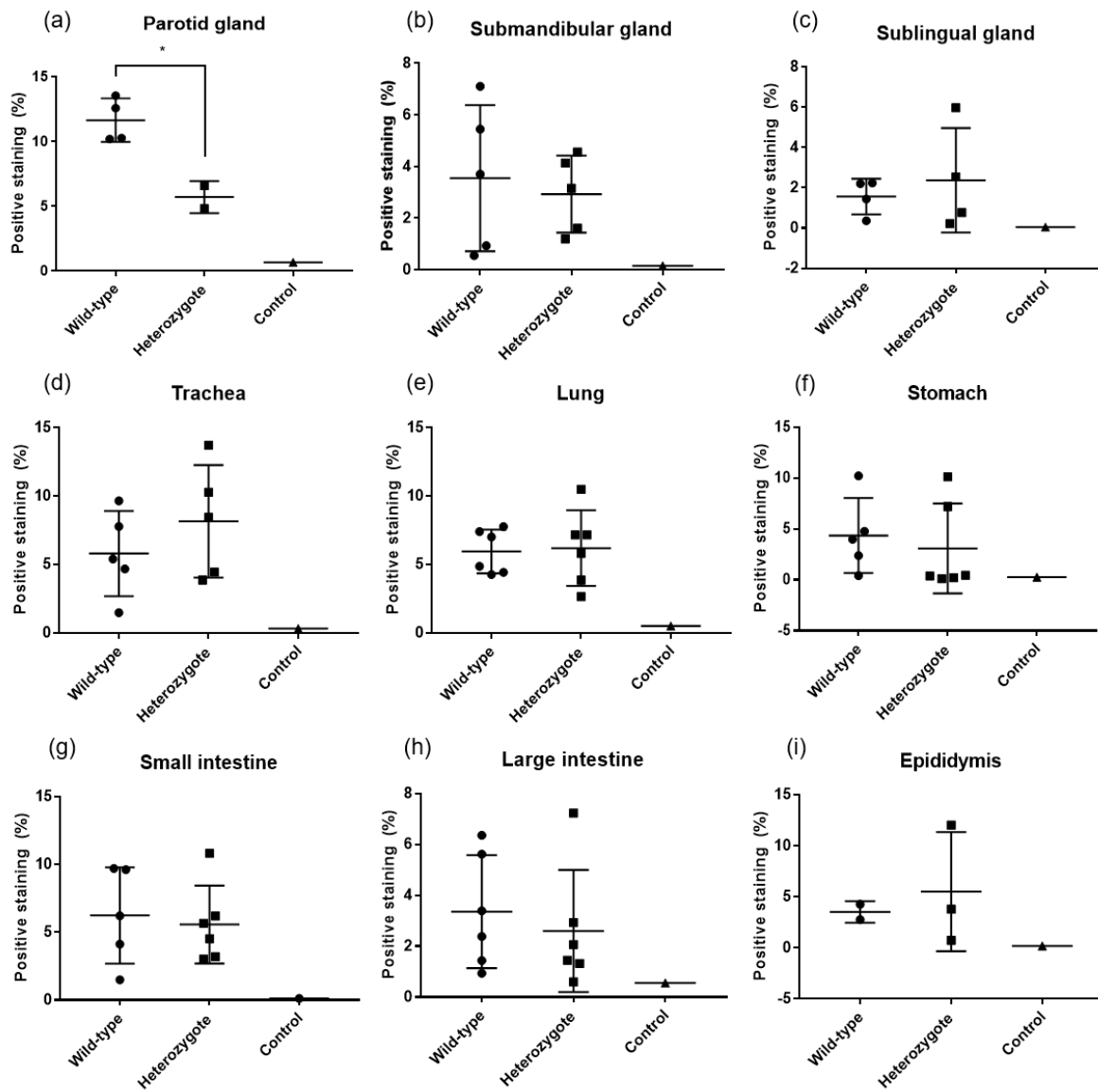
Fig 6.17. Quantification of positive Slpi immunostaining in WT and heterozygote mouse tissues. Positive Slpi staining was quantified in WT and heterozygotes for each tissue analysed (a-i). Individual data points represent each animal. The control shows background when tissues were incubated with secondary antibody only. Data represents the mean ± SD; * P ≤ 0.05.

To determine whether heterozygote animals displayed a greater propensity for fibrosis, tissues were analysed using Picrosirius red – an anionic dye which associates with collagen networks. Representative images of Picrosirius staining are shown in Fig 6.18. The Picrosirius red staining of lung sections from WT and heterozygotes was compared and quantified. A slight increase in the amount of positive staining in the heterozygote was identified, although the difference was not statistically significant (Fig 6.19a). When kidneys were subjected to the

same analysis, the results also showed no significant difference in positively stained pixels between the two phenotypes (Fig 6.19b).
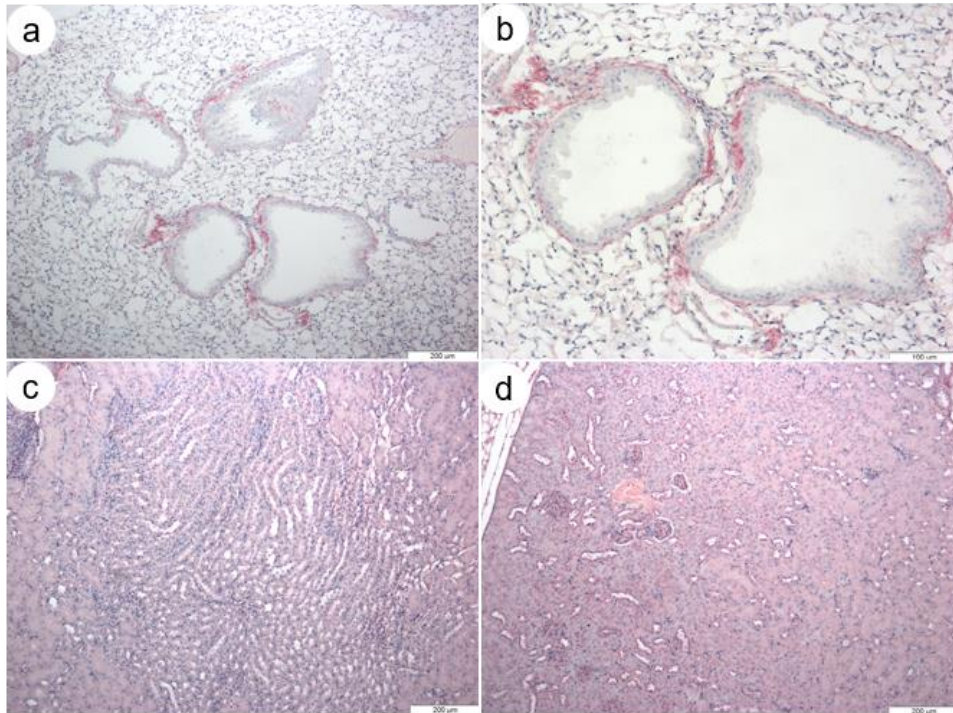
Fig 6.18. Picrosirius staining of murine lung and kidney tissue. Tissue sections show WT murine lung (a, b) and kidney (c, d) tissue stained with Picrosirius stain. Staining was visible around the bronchi but was absent in the alveoli (a, b). Staining was diffuse throughout the kidneys (c, d). Original magnification of images was 10X (a, c, d) and 20X (b).
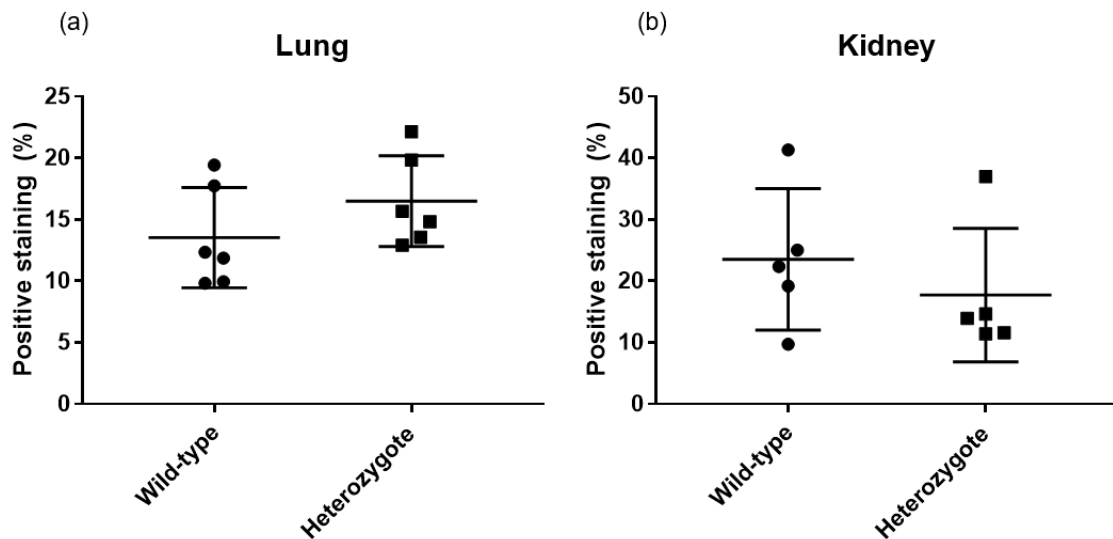
Fig 6.19. Quantification of Picrosirius red staining in WT and heterozygote mouse lungs and kidneys. Positive Picrosirius staining was quantified in WT and heterozygote lungs (a) and kidneys (b). Individual data points represent each animal. Data represents the mean ± SD.

6.7.   Analysis of homozygous embryos

Homozygous Wfdc2-knockout animals were viable at embryonic day E18.5 but the phenotype was lethal by weaning. The embryonic phenotype was studied at the MRC Harwell Institute and all results described in Section 6.7 are those communicated from researchers at MRC Harwell.

Embryos at E14.5 and E18.5 were analysed by micro computed tomography (micro-CT) to give 3D assessments of the embryos. Relative embryo size was measured by deriving embryo volume and crown-to-rump length from the micro-CT images.

When WT (N=2) and homozygote (N=4) embryos were compared, there was no difference in the volume or length of the embryos at E14.5 or E18.5 (Fig 6.20). The homozygote embryo volumes showed a greater spread of data at both embryonic days. Although there was no difference in the size of the embryos between phenotypes, at E18.5 two homozygote animals did exhibit a 20% larger volume than the mean of the WT embryos (Fig 6.21).
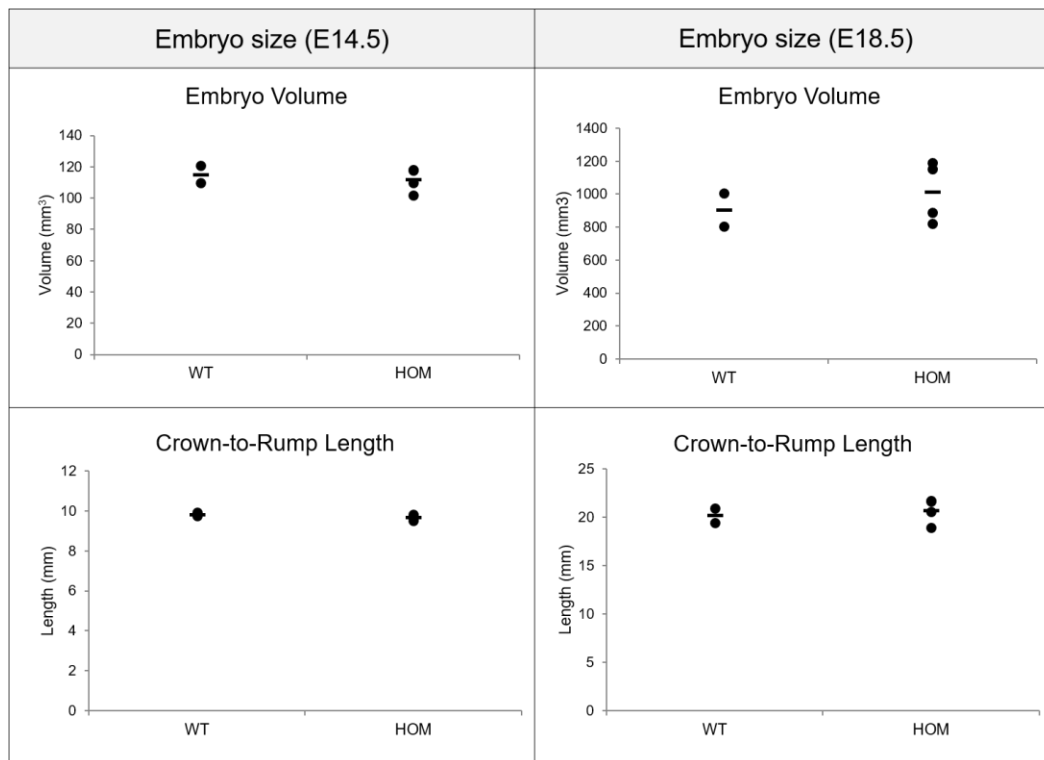
Fig 6.20. Assessment of WT and homozygote embryo size. Embryo size was analysed at E14.5 and E18.5 by measuring embryo volume (mm^3) and crown-to-rump length (mm). Individual data points represent each embryo and the mean is shown.
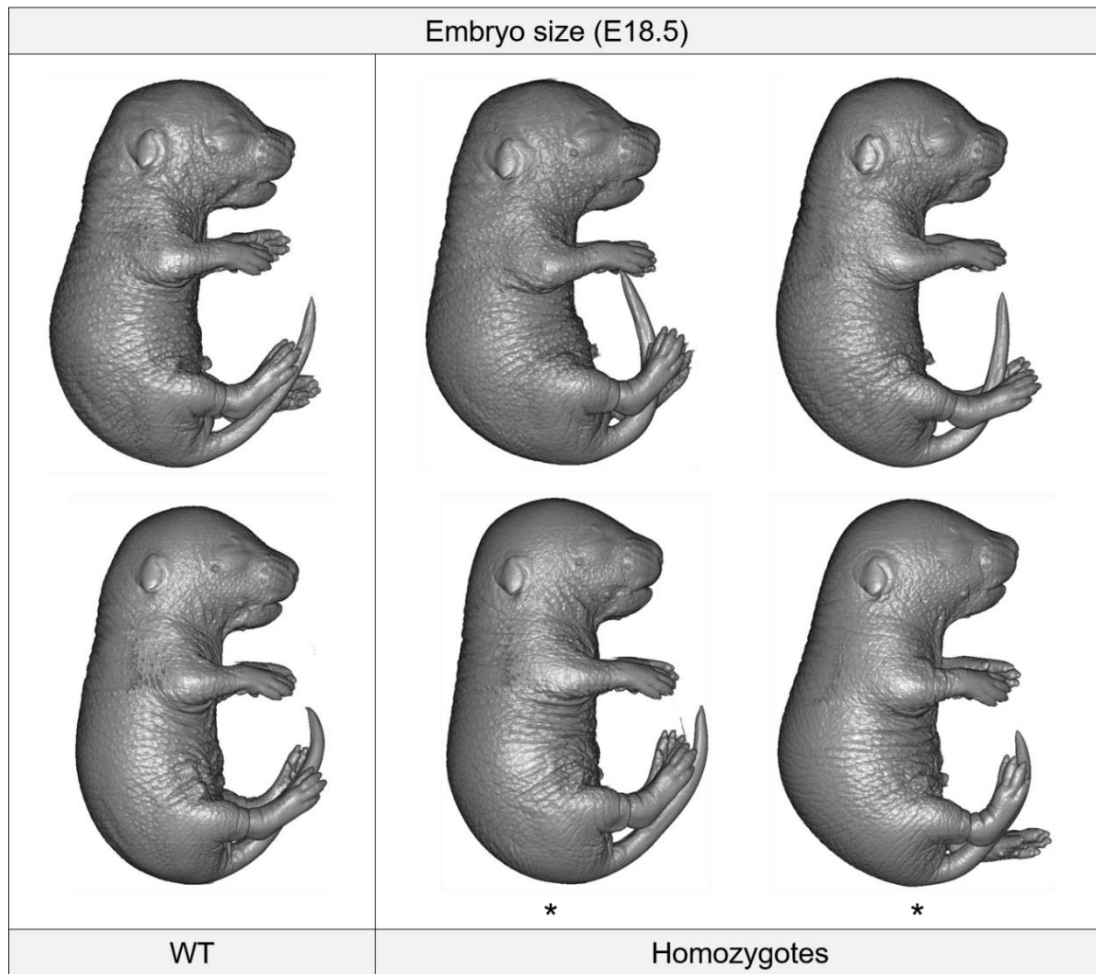
Fig 6.21. Micro-CT images of WT and homozygote embryo size. Representative images of WT and homozygote embryo size at E18.5 when analysed by micro-CT. Homozygote individuals that were 20% larger than the mean WT embryo size are marked with an asterisk (*).

Tissues/organs visualised by micro-CT were manually annotated and any morphological anomalies in the homozygote embryos were stated as a ratio (embryos with anomalies/total embryos analysed). At E14.5 and E18.5, the trachea and main bronchi were consistently found to exhibit structural differences compared to the WT animals in all four homozygotes analysed at both developmental time points. Other tissues including the fusion soft palate – nasal septum (E14.5), testis (E18.5) and ventricles (E18.5) had anomalies that were found in one out of four homozygotes. Two out of four animals also exhibited oedema at E18.5. Full details of the tissues and organs analysed and the number of annotations given are shown in Table 6.1.

3D assessment	Annotation of homozygotes E14.5	E18.5	Heart and lungs	Annotation of homozygotes E14.5	E18.5
Digits	0/4	0/4	Vertebra	-	0/4
Tail	0/4	0/4	Oesophagus	0/4	0/4
Pinna	0/4	0/4	Trachea	4/4	4/4
Oedema	2/4	0/4	Main bronchus	4/4	4/4
Vibrissa	0/4	0/4	Lung: cranial lobe	0/4	0/4
Size	0/4	2/4	Lung: middle lobes	0/4	0/4
Spinal shape	0/4	0/4	Lung: caudal lobe	0/4	0/4
Limbs	-	0/4	Lung: accessory lobe	0/4	0/4
Brain and CNS	E14.5	E18.5	Thymus	0/4	0/4
Midbrain roof	0/4	0/4	Vena cava	0/4	0/4
Midbrain	0/4	0/4	Subclavian artery	0/4	0/4
Cerebellum	0/4	0/4	Ductus arteriosus	0/4	-
Cerebral aqueduct	0/4	-	Aorta	0/4	0/4
Lateral ventricles	0/4	0/4	Atria	0/4	0/4
CP in lateral ventricles	0/4	0/4	Ventricles	0/4	1/4
CP in 4th ventricle	0/4	0/4	Tricuspid valve	0/4	0/4
Median sulcus	0/4	0/4	Mitral valve	0/4	0/4
4th ventricle	0/4	1/4	Interventricular septum	0/4	1/4
3rd ventricle	0/4	0/4	Pericardial cavity	0/4	-
Cortex	0/4	-	Pericardium	0/4	-
Olfactory lobe	0/4	0/4	Sternum primordium	0/4	0/4
Trigeminal nerves	0/4	-	First rib	0/4	0/4
Trigeminal ganglia	-	0/4	Pulmonary trunk	-	0/4
Rathke's pouch	0/4	-	**Liver and gut**	E14.5	E18.5
Pituitary gland	-	0/4	Diaphragm	0/4	0/4
Emerging cervical spinal cord	0/4	0/4	Peritoneal cavity	0/4	0/4
Spinal cord: central canal	0/4	-	Liver: R/L lobes	0/4	0/4
Dorsal root ganglia	0/4	0/4	Liver: caudate lobe	0/4	0/4
Face	E14.5	E18.5	Intralobular space	0/4	0/4
Cochlea	0/4	0/4	Stomach: pyloric antrim	-	0/4
Nasal septum	0/4	0/4	Stomach: fundus	-	0/4
Olfactory epithelia	0/4	-	Stomach: lumen	0/4	0/4
Nasal cavity	0/4	0/4	Hermiated gut	0/4	-
Intra-retinal space	-	0/4	Small intestine	-	0/4
Jacobson's organ	0/4	-	**Kidney and gonads**	E14.5	E18.5
Baso/sphenoid bone	0/4	0/4	Adrenals	0/3	0/4
Hyaloid cavity	0/4	0/4	Kidney	0/3	1/4
Lens	0/4	0/4	Bladder	0/3	0/4
Eyelids	0/4	0/4	Ovaries	0/3	0/4
Jugular lymph sacs	0/4	-	Testis	0/3	1/4
Eustachian tube	0/4	0/4	Spleen	-	0/4
Palatal shelf	1/4	0/4			
Fusion soft palate - nasal septum	1/4	-			
Epiglottis	0/4	0/4			
Tongue	0/4	0/4			
Submandibular gland	0/4	0/4			
Upper incisors	0/4	-			
Lower molar teeth	-	0/4			
Upper molar teeth	-	0/4			
Brown fat	-	0/4			
Thyroid	-	1/4			

Table 6.1. Analysis of morphological defects in homozygote embryos. Homozygote embryos imaged using micro-CT at E14.5 and E18.5 were manually annotated and compared to WT. Any morphological defects in tissues or organs were catalogued.

The micro-CT analysis of homozygote animals listed that all four homozygote embryos at embryonic day E14.5 and 18.5 exhibited structural defects in the trachea and main bronchus. When the micro-CT images from E14.5 were studied, the tracheal lumen in all homozygotes was significantly constricted or completely absent, while WT animals had a visible space down the centre of their tracheas (Fig 6.22 and 6.23). Similarly, the main bronchi of the homozygotes appeared to lack a lumen, while the WT embryos had a clear luminal space within each bronchus (Fig 6.24 and 6.25).
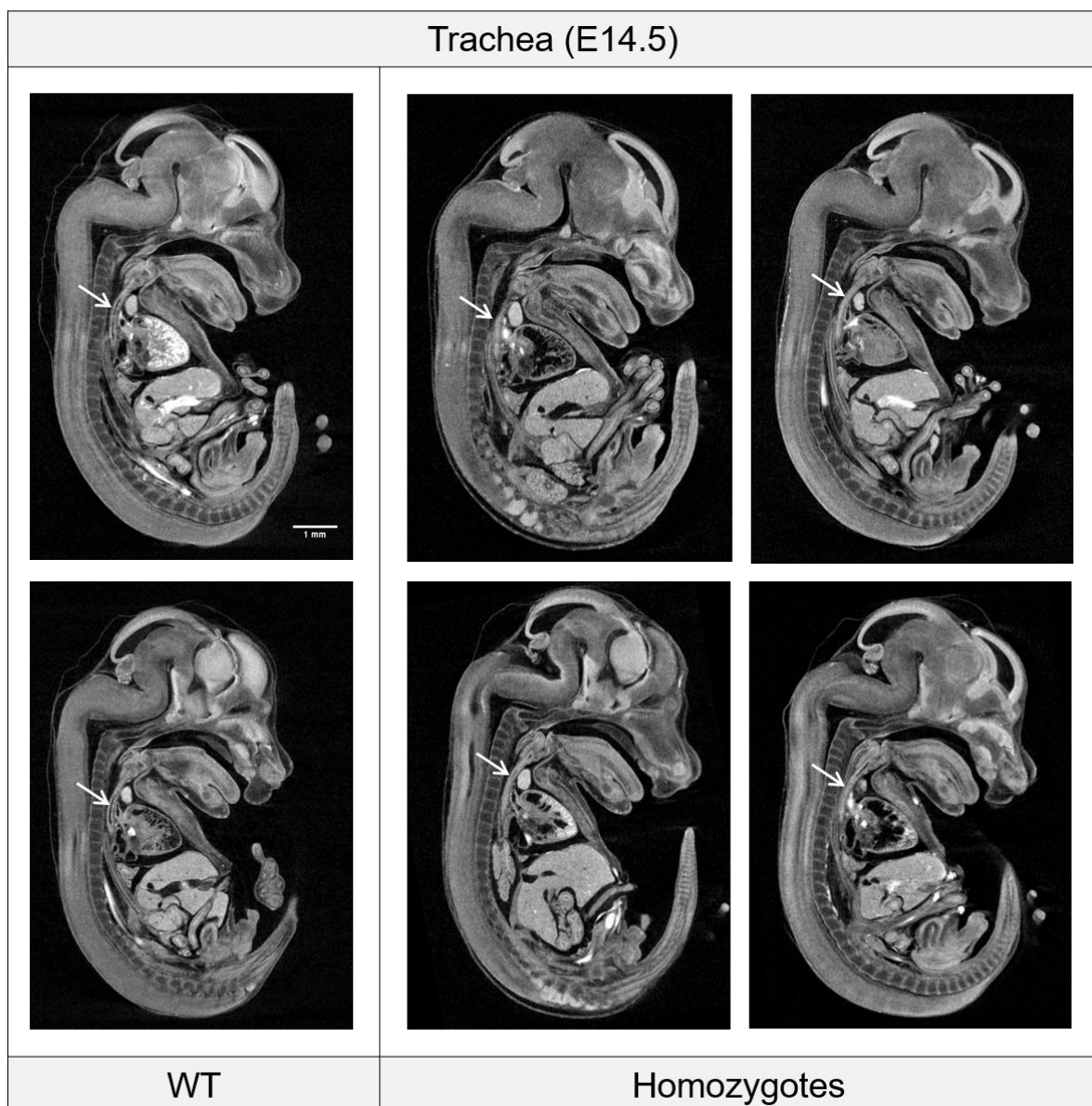
Fig 6.22. Micro-CT images of WT and homozygote tracheas (E14.5). Representative images of WT (N=2) and homozygote (N=4) embryo tracheas at E14.5, analysed by micro-CT. Tracheas are indicated with a white arrow.
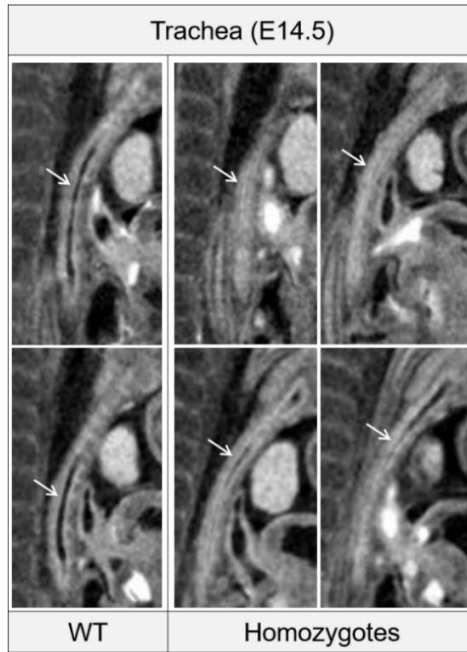
Fig 6.23. Expanded view of micro-CT images of tracheas (E14.5). Micro-CT images shown in Fig 6.22 are expanded 200X to visualise the trachea (indicated with an arrow).
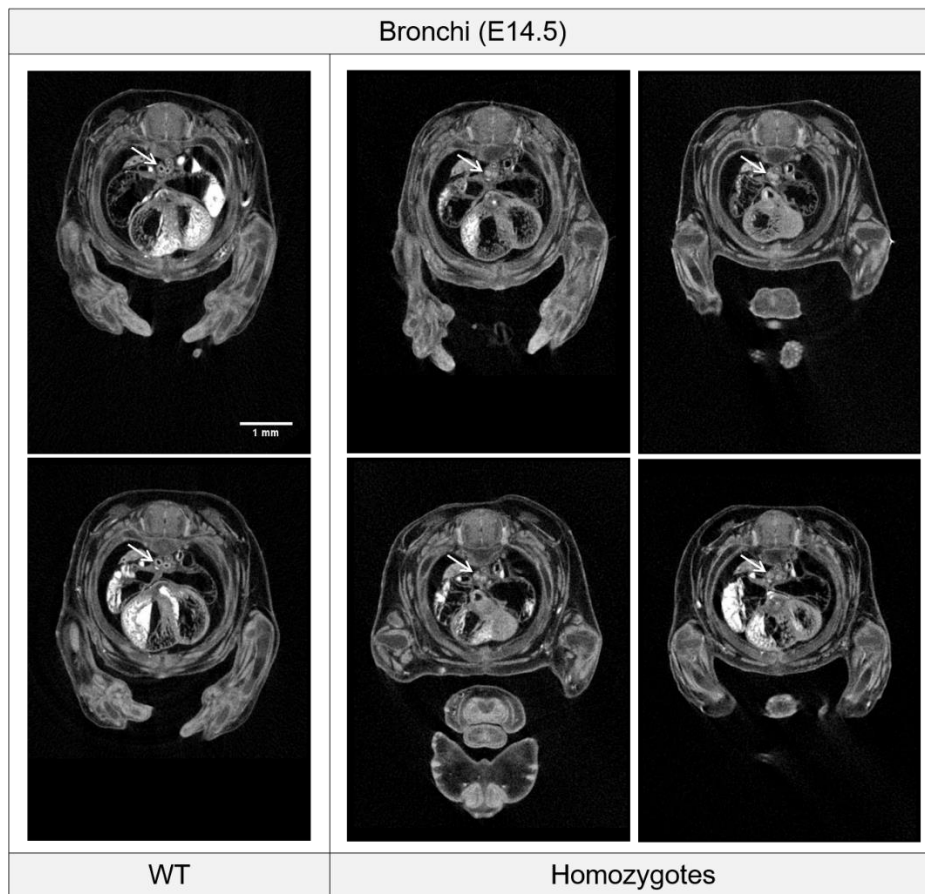
Fig 6.24. Micro-CT images of WT and homozygote main bronchi (E14.5). Representative images of WT (N=2) and homozygote (N=4) embryo main bronchi at E14.5, analysed by micro-CT. Bronchi are indicated with a white arrow.
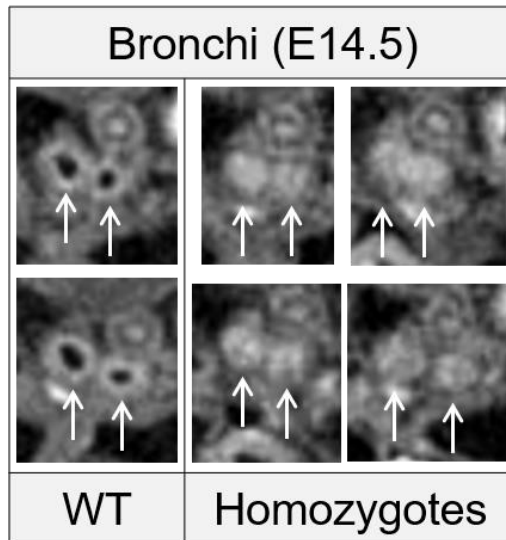
Fig 6.25. Expanded view of micro-CT images of main bronchi (E14.5). Micro-CT images shown in Fig 6.24 are expanded 250X to visualise the bronchi (indicated with arrows).

By E18.5, a visible tracheal lumen had developed in the homozygote animals; however, the lumen appeared to be narrower in the homozygotes compared to the WT (Fig 6.26 and 6.27). When measured, the trachea diameter was found to be 37% narrower in homozygotes compared to WT after being normalised to embryo size (crown-to-rump length) suggesting that the homozygote tracheas were significantly constricted (Fig 6.28). The main bronchi also showed a similar trend: by E18.5, the homozygotes had developed luminal spaces in the bronchi but the lumina appeared to be narrowed in comparison to WT (Fig 6.29 and 6.30). The diameter of the bronchial lumen was not measured but visual analysis appeared to suggest that constriction had occurred in the homozygote bronchi.
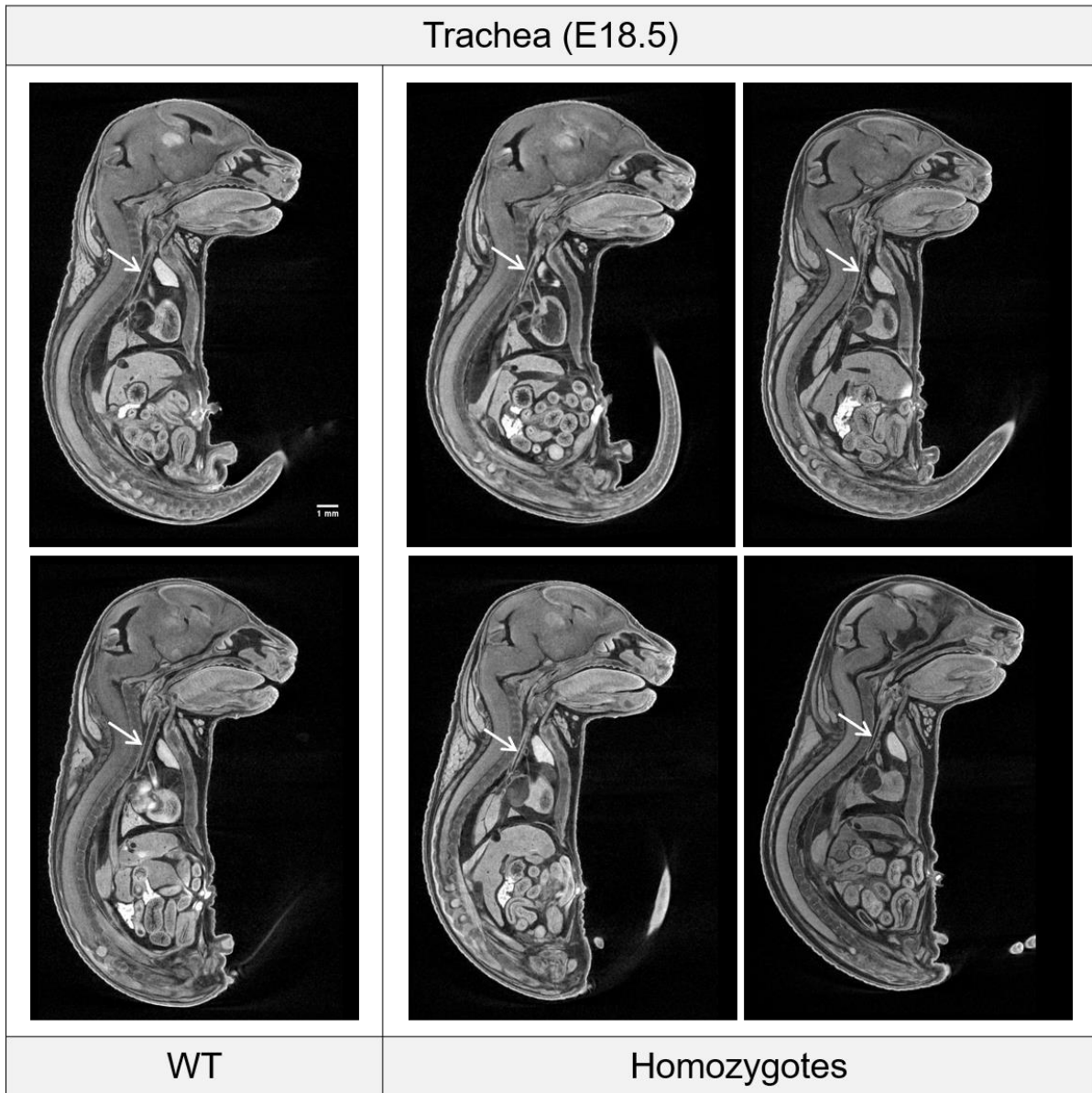
Fig 6.26. Micro-CT images of WT and homozygote tracheas (E18.5). Representative images of WT (N=2) and homozygote (N=4) embryo tracheas at E18.5, analysed by micro-CT. Tracheas are indicated with a white arrow.
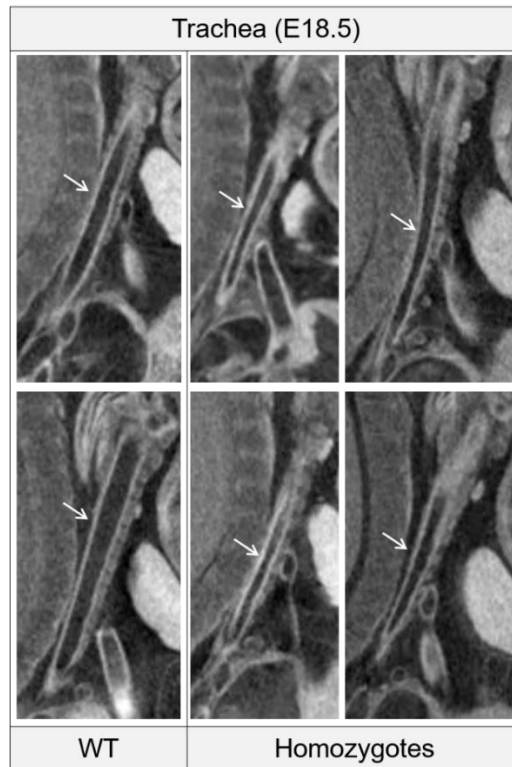
Fig 6.27. Expanded view of micro-CT images of tracheas (E18.5). Micro-CT images shown in Fig 6.26 are expanded 200X to visualise the trachea (indicated with an arrow).
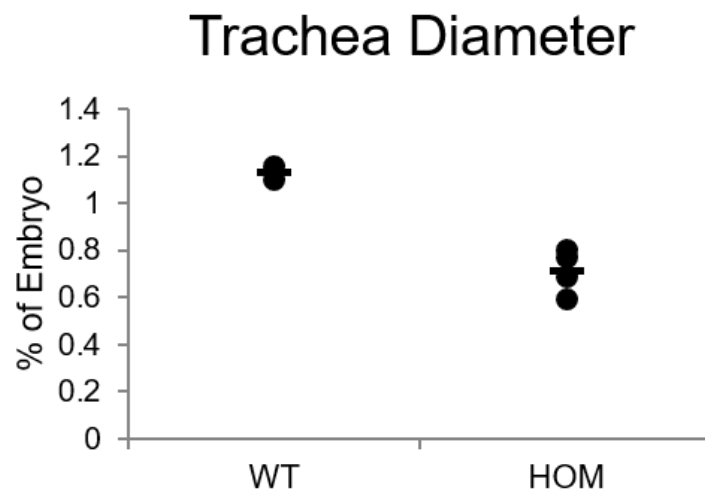
Fig 6.28. Assessment of WT and homozygote trachea diameter. The trachea diameter of WT (N=2) and homozygote (N=4) animals at E18.5 were measured from micro-CT images. Data is expressed as a percentage of the embryo (normalised to embryo size). Individual data points represent each embryo and the mean is shown.
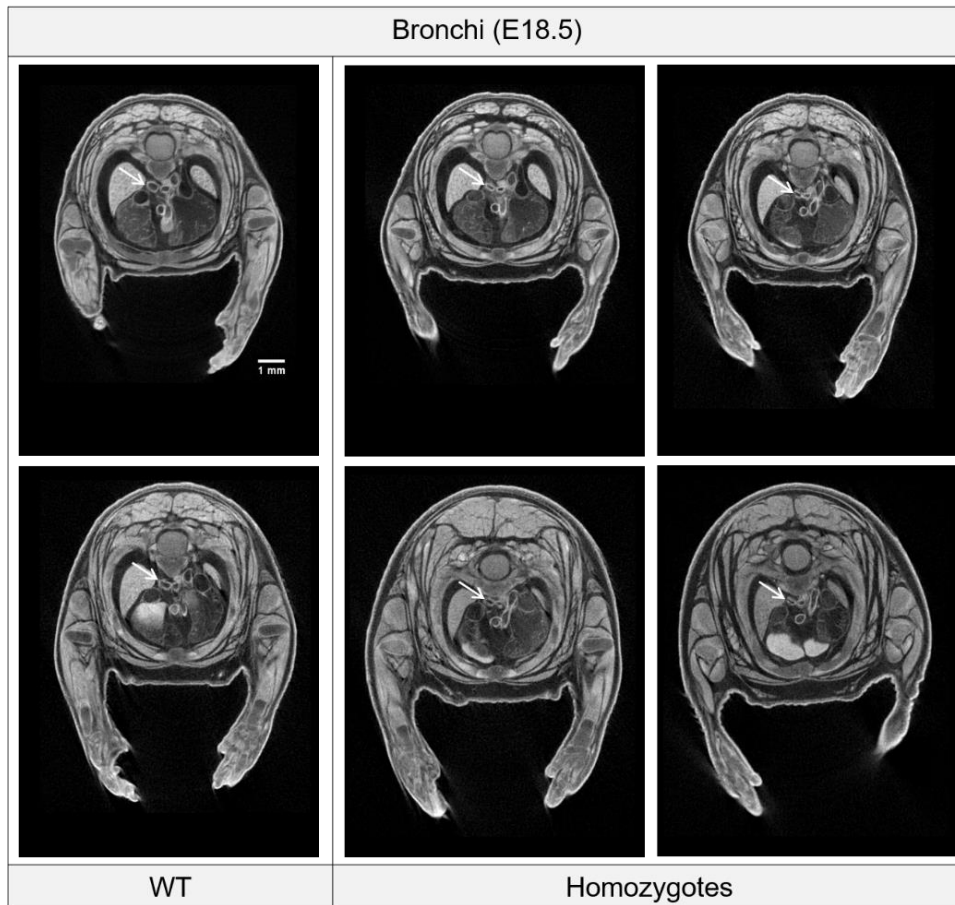
Fig 6.29. Micro-CT images of WT and homozygote main bronchi (E18.5). Representative images of WT (N=2) and homozygote (N=4) embryo main bronchi at E18.5, analysed by micro-CT. Bronchi are indicated with a white arrow.
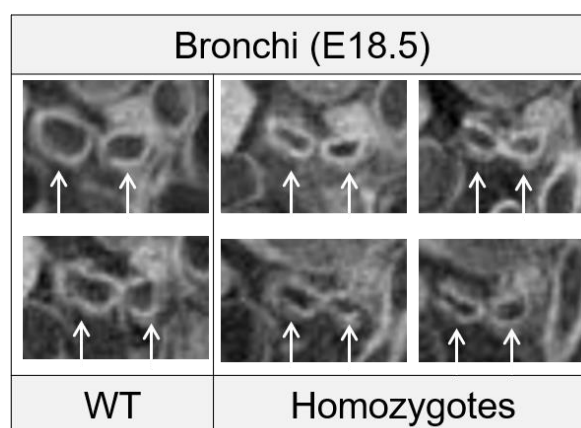
Fig 6.30. Expanded view of micro-CT images of main bronchi (E18.5). Micro-CT images shown in Fig 6.29 are expanded 250X to visualise the bronchi (indicated with arrows).

6.8. Discussion

Although the function of WFDC2 has previously been analysed using *in vitro* and xenograft studies, the results are conflicted and its biological role is yet to be firmly established. One investigative method that has not yet been employed is the use of knockout mice models. The recent establishment of Wfdc2-knockout mice at the MRC Harwell Institute provided a novel tool for investigating Wfdc2 function. The work described in this chapter aimed to use WT animals to establish the expression profile of Wfdc2 in mice, in addition to using knockout animals to identify any phenotypic changes that resulted from Wfdc2 silencing. The overall aim was to further elucidate the function of Wfdc2 in health and disease.

Histological analysis of WT mouse tissues was important for furthering the understanding of Wfdc2 expression; thus far, no IHC data has been published in the literature establishing the localisation of Wfdc2 in mice. IHC carried out during this study showed that positive Wfdc2 expression largely correlated with the high throughput sequencing data available in the literature (see Fig 1.10); for instance, according to BioGPS, the lung, large intestine, salivary gland and stomach all express Wfdc2 (Su et al., 2002). The level of gene expression described by Su et al (2002) appears to correlate with the amount of positive IHC staining identified in this study, for instance, array data shows that in the gastrointestinal system Wfdc2 is most abundant in the large intestine, followed by the stomach, with the small intestine expressing the least Wfdc2. This fits the same trend as that visualised by IHC.

The IHC analysis also demonstrated that murine Wfdc2 exhibits a similar staining pattern to its human counterpart, with both orthologues positively staining the bronchi, salivary glands, large intestine, fallopian tube, kidney and testis (Uhlén et al., 2015). WFDC2 expression tends to be localised to ductal/secretory cell types in human and mouse; for instance, in the salivary glands WFDC2 orthologues are localised to the ducts, in the large airways their expression is limited to the respiratory epithelium and underlying submucosal glands, while the large intestine shows WFDC2 positivity in the tubular glands and their associated goblet cells (Bingle et al., 2006; Uhlén et al., 2015).

The staining pattern of Wfdc2 in the airways suggests that it is most commonly the product of serous cells. For instance, Wfdc2 expression in the airways closely correlates with that of club cells (see Fig 6.6), which are non-ciliated, serous cells found exclusively in the bronchiolar epithelium (Rokicki et al., 2016). Although Wfdc2 co-localises with club cells in the bronchiolar epithelium it is also abundant in the large airways where club cells are largely absent, suggesting that it is expressed by other subtypes of epithelial cell. Positive staining was also seen in what appear to be serous acini of submucosal glands (Fig 6.6). This expression pattern is similar to other antimicrobial factors such as lysozyme, β-defensins, lactotransferrin and SLPI, which tend to be produced by serous cells of the respiratory tract e.g. club cells/ciliated cells of the epithelium and serous-type submucosal glands (Bingle et al., 2006; Bowes et al., 1981; Dohrman et al., 1994; Grubor et al., 2006; Uhlén et al., 2015). To further support the theory that Wfdc2 is expressed by serous cells, airway tissue was stained for markers of mucous, such as alcian blue and MUC5B (see Fig 6.13e-f). The staining pattern for mucins was distinct to Wfdc2, with no visible co-localisation of Wfdc2 with alcian blue or MUC5B.

In the oral cavity, Wfdc2 also appears to be expressed by serous cells rather than mucous cells. For example, positive staining was identified in serous acini of the parotid gland, whereas mucous acini of the submandibular and sublingual glands were devoid of Wfdc2 as shown by the distinct localisation to MUC5B and alcian blue. In the mixed glands, Wfdc2 is found exclusively in ducts and serous demilunes. The ductal expression of Wfdc2 suggests that the protein is supplemented into the viscous saliva produced by mixed glands as it is excreted from ducts rather than being initially exuded into saliva from acinar cells. This ductal staining is also seen in other host defence proteins such as lysozyme, SLPI and lactotransferrin, but, unlike Wfdc2, this is often coupled with positive expression in acini of mixed glands (Bingle et al., 2006; Korsrud and Brandtzaeg, 1982; Uhlén et al., 2015).

Curiously, the converse is true in the gastrointestinal tract, where Wfdc2 appears to exhibit mucous-associated secretion. In the small intestine, anti-microbial peptides e.g. α-defensins and lysozyme are secreted from Paneth cells at the base of intestinal crypts (Cunliffe et al.,

2001; Erlandsen et al., 1974), therefore, given the similar staining patterns of such factors in the airways and the shared putative defence functions, Wfdc2 would perhaps be expected to be co-expressed at the same site. Instead, Wfdc2 is almost exclusively linked to mucous cells, namely in goblet cells of the small and large intestine and mucous cells of the gastric pits of the stomach. Intestinal goblet cells are specialised for mucin production (largely of MUC2), carrying out a lubricating/protective function and serving to facilitate stool transit (Johansson et al., 2009). The mucous cells at the surface of the stomach mucosa secrete high levels of MUC5AC (Rodríguez-Piñeiro et al., 2013). Wfdc2 expression does not appear to correlate with expression of specific mucins, in as much that its expression is most abundant in the stomach and large intestine whereas MUC2 is most abundant in the large intestine, and least expressed in the stomach and MUC5AC shows the converse (Rodríguez-Piñeiro et al., 2013).

The presence of Wfdc2 in chief cells is also atypical for anti-microbial proteins. Chief cells are present at the base of gastric glands, where they secrete pepsin (a proteolytic enzyme). The role of Wfdc2 in these cells is unknown, however the expression has been shown to be upregulated 30-fold during the metaplastic transition of chief cells to spasmolytic polypeptide-expressing metaplasia (SPEM) cells (Weis et al., 2014). In fact, Wfdc2 has been proposed to act as a novel marker of SPEM cells (Nozaki et al., 2008; Weis et al., 2014).

To summarise, the IHC analysis of murine tissues shows that Wfdc2 is a secretory protein that is expressed by serous cells of the oral cavity and airways where it may contribute to the biological function of mucous secretions such as saliva and airway surface lining fluid. It is absent at sites where mucous is lacking, such as the alveoli of the lung. Its expression in mucous-secreting cells of the stomach and intestines, as well as the glandular epithelium of the fallopian tubes, also links Wfdc2 expression to secretory mucosal surfaces. These sites of expression suggest that Wfdc2 is involved in contributing to the composition and function of mucosal surface lining fluids, a theory which continues to propagate the view that Wfdc2 is involved in host defence in some way. However, it is possible that secretion at these sites may play a different and, as yet, unidentified role.

When the histology of WT tissues was compared to those of heterozygote animals no structural defects or significant differences in tissue architecture were identifiable. This was most likely owing to the lack of change in Wfdc2 protein expression between WT and heterozygotes: when positive staining was quantified, there appeared to be no significant reduction in Wfdc2 protein in heterozygotes. In fact, many heterozygote tissues showed a higher mean percentage of positive staining compared to WT, including the large intestine where the difference was statistically significant.

The 'normal' level of Wfdc2 protein production in heterozygotes suggests that the single functional Wfdc2 allele undergoes a compensatory upregulation in expression in response to the knockout event. Such genetic adaptive responses have been reported in the literature; for instance, DNA lesions have been shown to induce reorganisation of chromatin whereby epigenetic changes make compensatory genes more accessible and result in increased expression (Ziv et al., 2006). It has been proposed that dosage compensation systems, that were originally thought to be sex-chromosome-specific, play an important role in buffering the harmful phenotypes resulting from genetic alterations in autosomes (El-Brolosy and Stainier, 2017; Valsecchi et al., 2018). Monoallelic expression of Wfdc2 in heterozygotes most likely results in dosage compensation due to the fact that Wfdc2 is essential for embryonic development, as determined by the lethal phenotype observed in homozygotes.

Genetic knockouts may also induce upregulation of paralogous genes with redundant functions; therefore, the heterozygous tissues were assessed to determine whether they also exhibited a compensatory increase in Slpi protein. Quantification of Slpi showed that there was no significant difference in protein levels between phenotypes. This may be due to two factors. Firstly, it may result from the dosage compensation of Wfdc2 in heterozygotes, meaning that the animals did not exhibit any loss of Wfdc2 function that could be rectified by an increase in Slpi. Secondly, it may be due to the fact that, despite their shared possession of WFDC domains, Slpi and Wfdc2 elicit entirely different biological functions. A lack of functional redundancy would explain why Slpi is not upregulated in heterozygotes.

Heterozygotes and WT animals were additionally analysed for any observable differences in collagen networks due to the fact that WFDC2 has been implicated in human fibrotic diseases such as cystic fibrosis and renal fibrosis (see Section 1.5): in fact, it is hypothesized to be involved in the process of collagen turnover in renal fibrosis (LeBleu et al., 2013). When lung and kidney heterozygous tissues were studied using Picrosirius red stain, there was no significant difference in the percentage of positive staining. Similarly, no changes in collagen thickness or accumulation were observed between phenotypes. Again, it is difficult to speculate the significance of this due to the 'normal' Wfdc2 production in the heterozygotes and the potential absence of a mutant phenotype.

Although the heterozygous tissues elucidated little about the function of Wfdc2, the homozygous knockouts provided compelling insights into its role in development. Most interestingly, the phenotype was lethal: homozygote animals died shortly after birth at some point prior to weaning. Lethality after birth is suggestive of a respiratory phenotype as, prior to birth, gas exchange occurs via the placenta, meaning that successful respiration after birth will only occur if the airways have developed correctly during embryogenesis. In corroboration with this theory, the trachea and main bronchi of homozygote embryos at embryonic day 14.5 and 18.5 were visibly constricted compared to WT. When measured at E18.5, the trachea diameters were found to be significantly narrower in homozygotes.

An understanding of the genetic factors involved in mammalian embryonic airway development has steadily progressed over the last 20 years and transgenic mouse models have been a valuable resource to comprehend the complex interplay of regulatory signals involved. Many factors have been implicated in the process of airway organogenesis due to the revealing phenotypes of transgenic mice. For instance, Wnt5a homozygous mutants exhibit significantly truncated trachea and over-branching of the distal airways (Li et al., 2002). Traf4 mutants have a constriction in the upper trachea at the junction between the trachea and pharynx (Shiels et al., 2000) and Hoxa-5 homozygous mutants experience tracheal occlusion as a result of narrowing of the tracheal lumen, as well as having reduced lung

branching (Aubin et al., 1997). The phenotype of Wfdc2 homozygous mutant embryos suggests that Wfdc2 is also involved in development of the trachea and main bronchi, however its exact function is unclear.

A number of different tissue components are involved in the multistage process of tracheal morphogenesis, including cartilage rings, smooth muscle and pseudostratified columnar epithelial cells. In the mouse embryo, tracheal and lung development is initiated at E9. At E11.5, mesenchymal cells begin to differentiate into smooth muscle and by E14.5 chondrogenesis is underway. From E14.5 onwards the epithelium changes to form a pseudostratified epithelium and at E18.5 the trachea is considered mature. Careful analysis of this developmental process performed by Kishimoto et al (2018) has provided important insights into the factors involved in driving tracheal growth and expansion. Although tracheal elongation is largely facilitated by a specific smooth muscle cell alignment (mediated by *Wnt5a* signalling), luminal expansion is regulated by two distinct factors: cartilage formation and epithelial organisation (Kishimoto et al., 2018).

Between E10.5 to E14.5 the external diameter of the murine trachea is unchanged, but from E14.5 onwards it continually expands. One of the factors affecting tracheal diameter is the cartilage, which begins to develop from E14.5 forming C-shaped rings around the trachea to prevent it from collapsing (Kishimoto et al., 2018). To determine the role of cartilage in luminal growth, Sox9 transgenic mice were used; *Sox9* mutants lack cartilage progenitor cells and therefore their trachea develop without cartilage. When the lumen of *Sox9* mutants were measured, the animals showed a 50% reduction in the total luminal surface area, while its length was not significantly altered (Kishimoto et al., 2018). Although this finding may, at a glance, suggest that Wfdc2 influences chondrogenesis, on further inspection the phenotypes of the two mutants appear distinct. Sox9 mutants show a collapsed trachea at E18.5 with no visible lumen and the external diameter of the trachea remains unchanged between E14.5 and E18.5 (Kishimoto et al., 2018; Turcatel et al., 2013). Wfdc2 mutants, however, have a visible lumen along the entire length of the trachea which is narrowed compared to the WT

(Fig 6.23 and 6.27) and, in contrast to Sox9, the external diameter of Wfdc2 mutant tracheas between E14.5 and E18.5 does appear to be increased (although this has not been quantified from the micro-CT images). The fact that there is no evidence in humans or mice that Wfdc2 is expressed in cartilage or that its expression in tissues correlates with a proximity to cartilage furthers the notion that Wfdc2 does not affect chondrogenesis.

The second important factor to influence tracheal diameter is the luminal epithelium. Although it may be supposed that epithelial cells can only instigate luminal growth by proliferating, they actually perform a more complex role in tracheal expansion. Mesenchymal and epithelial cells proliferate rapidly from E12.5 to E14.5 (at which the proliferation rate peaks) and the degree of cell growth is consistent with the fold-change in luminal surface area. However, from E14.5 onwards, the cell proliferation rate decreases while the luminal surface continues to grow. By analysing the lumen using confocal microscopy, Kishimoto et al (2018) determined that the additional factor affecting luminal diameter was the 'reshaping' of the epithelium. The group showed that the proportion of luminal cells with an exposed apical surface increased over time and that, once the cells were present at the lumen, their apical surfaces expanded. The morphological changes at the lumen surface were quantified to demonstrate the extent of the rearrangement: between E14.5 to E16.5, 60% of tracheal cells had an apical surface exposed and 40% were contained at the basal side; by E18.5 the number of apical surface cells had increased to 80% due to further migration/expansion from the basal side to the apical surface. As a result of the remodelling of the epithelium, the tracheal lumen increased in surface area independently of proliferation (Kishimoto et al., 2018).

The Wfdc2 mutant phenotype showed differences in luminal expansion at E14.5 and E18.5 compared to WT and these differences could be attributed to failed epithelial organisation or reduced proliferation. The fact that the homozygotes showed a tracheal phenotype at E14.5 may suggest that Wfdc2 is involved in influencing epithelial proliferation, as tracheal growth at this stage is largely due to cell number expansion (Kishimoto et al., 2018); however, by E14.5 60% of tracheal epithelial cells had acquired an apical surface so it is also possible that without

Wfdc2 expression, the epithelium had failed to enlarge due to an absence of apical surface enlargement and emergence. At E18.5, the mutant trachea had expanded somewhat, but still exhibited a 37% reduced lumen diameter compared to WT. The expansion that occurred between E14.5 and E18.5 independently of Wfdc2 may be due to continued Sox9-mediated cartilage growth. The differences in luminal diameter between WT and Wfdc2 knockout animals could be attributable to either a reduction in epithelial cell proliferation or due to failed epithelial reorganisation. These theories will be discussed in more detail in Chapter 7.

The change in diameter of the main bronchi is probably attributable to the same Wfdc2-mediated developmental limitations as the trachea, as the bronchi are continuous with the trachea and are structurally similar i.e. are both are reinforced with cartilage and lined with a pseudostratified columnar epithelium.

Given the well documented expression of Wfdc2 in the epithelium of murine and human tissues including the large airways, it is feasible that its developmental role would impact epithelial cell behaviour rather than that of chondrocytes or smooth muscle cells. Nonetheless, extensive analysis is required to fully understand how Wfdc2 affects tracheal and bronchial development. The continued and abundant expression of Wfdc2 in tissues of the adult airway suggests that the protein continues to play a functional role outside of development.

In addition to the trachea and bronchi, other tissues/organs were listed to exhibit structural defects in the Wfdc2 mutants compared to WT. The annotations with regard to the palatal shelf, nasal septum, thyroid, ventricles, interventricular septum, kidney and testis all occurred in 1/4 homozygotes, and were only found in one of the two timepoints (E14.5 or E18.5). The inconsistency in the annotations mean it is difficult to draw substantial conclusions from the data or attribute the changes to an actual phenotypic change resulting from Wfdc2 silencing, therefore they will not be discussed in more detail.

Chapter 7

Final discussion

7.1. Discussion

The overarching aim of the work described in this thesis was to determine the function of WFDC2, with the primary objective of understanding whether it is involved in tumorigenesis. A multifaceted approach was adopted using recombinant WFDC2 protein, a WFDC2-knockout oral cancer-derived cell line and Wfdc2-knockout mice as tools to elucidate its function.

Although WFDC2 is hypothesised to be involved in mucosal defence, the recombinant human and murine WFDC2 orthologues synthesised in this study were found to exhibit no antibacterial effects against the four species of bacteria tested, nor anti-protease activity against trypsin or MMPs. Although these results contradicted a widely held hypothesis, when considered more closely, the potential reasoning behind the negative data became apparent.

WFDC2 is largely hypothesised to exhibit antimicrobial and anti-protease activity due to its homology to SLPI and elafin. However, aside from their shared possession of WFDC domains, there are a number of key differences between the proteins. Firstly, SLPI and elafin are cationic at physiological pH (pI 9.11 and 9.14, respectively), meaning it is probable that they exhibit their moderate antimicrobial effects via membrane disruption (as discussed in Section 1.3.2), whereas human and murine WFDC2 are negatively charged at physiological pH (pI 4.69 and 5.21, respectively) and would be unable to induce membrane effects in the same way (pI values determined using ExPASy tool; Section 2.12). Secondly, WFDC2 is abundantly glycosylated whereas SLPI and elafin are devoid of glycan modifications; AMPs characteristically lack post-translational modifications (Wiesner and Vilcinskas, 2010). Thirdly, as extensively discussed in Sections 1.1.4 and 1.3.5, WFDC2 has an alternative cysteine

spacing to the conserved alignment found in SLPI and elafin that facilitates protease inhibition; without this cysteine spacing, the would-be active site of WFDC2 is, in all probability, too small to permit binding of a protease ligand.

Perhaps the most compelling evidence to indicate that WFDC2 and SLPI exhibit different functions are the phenotypes of the resultant knockout mice which differ greatly in their severity. Slpi-knockout mice develop normally to adulthood and the homozygote phenotype only becomes apparent following experimental treatments; for instance, the mice are more susceptible to *Mycobacteria* infection (Nishimura et al., 2008) and experience exacerbated disease when infected with *Leishmania* (McCartney-Francis et al., 2014). They also have a heightened susceptibility to septic shock following exposure to LPS (Nakamura et al., 2003). In stark contrast, we found that Wfdc2-knockout mice exhibited a lethal phenotype, where the pups died shortly after birth, prior to weaning. The premature death of the mice appeared to result from respiratory defects as the lumen of the trachea and bronchi were constricted, thus respiration was obstructed.

The phenotype of the Wfdc2 knockout mice analysed in this study led to the hypothesis that Wfdc2 was involved in the process of tubulogenesis of the trachea and bronchi, most likely by regulating the reorganisation of the luminal epithelium to enable expansion of its apical surface (as discussed in Section 6.8). A protein involved in luminal development in *Drosophila melanogaster* known as Crumbs (Crb) presents an interesting example of an embryonic regulator of epithelial organisation. Crb localises to the epithelial plasma membrane where it mediates epithelial cell polarity and apical membrane growth (Pellikka et al., 2002; Wodarz et al., 1995). Uncontrolled expression of Crb results in expansion of the luminal apical membrane of the trachea during late stage *Drosophila* embryogenesis (Laprise et al., 2010). Similarly, overexpression of Crb causes salivary gland lumina to expand while loss of Crb function leads to defects in salivary gland formation, including constriction of the lumen, as a result of changes in apical membrane growth (Myat and Andrew, 2002).

It is possible, given the phenotypic effects described in this study, that Wfdc2 plays a similar role to Crb in reorganising the luminal epithelial surface of mammalian glands and tubes. Although the *Drosophila* Crb gene provides a viable example of how luminal defects such as those found in Wfdc2-knockout mice may arise, the evolutionary distance between mice and fruitflies limits its relevance somewhat.

A Crb orthologue that exists in mammals, Crumbs 3 (CRB3), appears to perform similar activities to its orthologue in *Drosophila*, for instance, *in vitro* it has been reported to be involved in regulating epithelial apicobasal polarity (Szymaniak et al., 2015). Interestingly, when Crb3-deficient mice are analysed, they exhibit a number of phenotypic defects that are similar to those of Wfdc2-knockout mice. The Crb3-homozygote mice die perinatally as a result of respiratory distress, while the heterozygotes are viable and show no obvious defects. When homozygote embryos at E18.5 are analysed, lung defects including 'atrophy of luminal spaces', and 'airways filled with ectopic debris' are evident. The kidneys also exhibit defects in polarisation which cause the formation of cysts and the accumulation of fibrotic tissue. Additionally, the intestinal villi are irregular and frequently fused together (Charrier et al., 2015; Whiteman et al., 2014).

The Crb3 knockout mice are relevant to this study as they demonstrate how proteins which regulate epithelial development can cause defects in lung organogenesis when their expression is silenced, as well as defects that extend far beyond the airways to sites such as the kidney and intestines. Although the analysis of Wfdc2-knockout mice reported in this study only showed defects in the large airway lumen, the level of detail that could be visualised using micro-CT imaging was limited. It is entirely possible that other architectural defects are present in the embryos, and these defects will only be discernible upon further histopathological inspection. If other sites do exhibit phenotypic changes, it is most probable that they will be found in ducts and tubules that are positive for Wfdc2 in adult mice, such as salivary glands, intestinal crypts, the fallopian tube, renal tubules and seminiferous tubules of the testis.

Clearly Wfdc2 is important for development of the airways, yet its role in adult tissues remains unclear. The fact that tracheal and bronchial WFDC2 expression continues into adulthood suggests that it continues to serve a purpose in these tissues. Indeed, expression of genes involved in embryonic development are often sustained in adult tissues where they frequently play roles in maintenance and repair. For instance, Sox9, a key protein required for tracheal development (see Section 6.8), is expressed in a subpopulation of adult pulmonary basal cells where it is reported to be involved in regeneration of damaged lung tissue in response to injury signals (Ma et al., 2018). Crb3, however, appears to be involved in maintenance of contact inhibition and regulation of tight/adherens junctions, as determined by *in vitro* analysis of immortalised mammalian cell lines (Karp et al., 2008; Mao et al., 2017). It is possible that Wfdc2 regulates epithelial surface dimensions and morphology in adulthood in a similar way to its hypothesised role in the embryo.

With this in mind, it is wholly understandable how dysregulation of proteins that are involved in maintaining the organisation of epithelia could lead to tumour formation. If such processes become uncontrolled, they can promote activities such as cell migration and metastasis. There are many examples of this, for instance, Twist (a transcription factor that controls embryonic epithelial-mesenchymal transition; EMT) has been shown to play an essential role in promoting metastasis in breast cancer (Yang et al., 2004). Loss of polarity, loss of cell-to-cell contacts and increased cell migration are characteristic of EMT, however these processes are also key to tumorigenesis. As a result, proteins involved in embryonic EMT are frequently implicated in cancer. Other cell polarity proteins, such as Par3 and Crb3 are also implicated in tumorigenesis. Loss of Par3, a scaffolding protein localised to tight junctions, can enhance cell invasion and metastasis due to downstream effects on actin remodelling and cell-ECM interactions (McCaffrey et al., 2012). Similarly, loss of Crb3 expression can have detrimental effects on cell polarity and cell-cell junctions resulting in increased cell migration (Karp et al., 2008).

If WFDC2 plays a role in regulating epithelial organisation, this explains why the protein is so frequently upregulated in diseases characterised by epithelial remodelling and destruction, such as cystic and renal fibrosis. It would also explain how the frequent upregulation of WFDC2 that occurs in cancer could lead to tumour-enhancing effects.

This theory corroborates the results presented in Chapter 5, whereby the CRISPR gene edited cells were shown to exhibit significantly reduced cell invasion as a result of WFDC2 silencing. If WFDC2 is indeed involved in promoting cell migration and metastasis it would explain why WFDC2 serum levels frequently correlate with poor prognosis and increasing disease stage (see Section 1.4.1). Although heightened WFDC2 expression occurs in ovarian cancer as a result of the frequent amplification of the WFDC2-encoding locus, 20q13 (Section 1.4.3), it is also possible that WFDC2 is selectively upregulated by the cancer cells as a result of its pro-tumour effects.

A multitude of evidence has shown that WFDC2 expression becomes amplified in cancer, yet there is no data in the literature describing loss-of-function WFDC2 mutations. By analysing exome/genome sequencing datasets such as The Genome Aggregation Database (gnomAD), the frequency of single nucleotide variants (SNVs) in the general population can be assessed. The data from gnomAD suggests that WFDC2 has a strong selective pressure against loss-of-function variants, as shown by the absence of any homozygous loss-of-function mutations in the population analysed. This supports the phenotype of the homozygous knockout mice, suggesting that total silencing of WFDC2 expression is not viable in humans either.

Two missense homozygous mutations are recorded, but neither of them target the conserved cysteine residues, therefore it is possible that the mutations do not drastically alter the secondary structure of the protein enough to affect its function, or that the mutations only affect one WFDC domain lobe and therefore the other domain remains functional. The phenotype of the heterozygous knockout mice appear to show that a single functional copy of Wfdc2 is sufficient to enable normal development to adulthood, therefore it is feasible that the various missense/nonsense heterozygous SNVs reported in gnomAD are not deleterious to the host,

as a single allele is able to preserve a healthy phenotype. It is also possible that the different WFDC2 splice variants exist in order to maintain expression of the gene if V0 expression becomes disrupted.

Given that the RT-PCR analysis described in Section 5.7.2 suggested that the CRISPR cells directed their transcriptional machinery towards synthesis of isoform V2 as a consequence of V0 silencing (Fig 5.14 and 5.15), it is possible that isoforms V1-V4 are functionally redundant to V0. With this in mind, it is likely that the downstream effect of silencing all five WFDC2 isoforms would induce much greater changes to cell behaviour than identified in this study.

Limitations of the study and future work

The assessment of antimicrobial activity performed in this study used a small panel of bacteria which were chosen due to their capacity to colonise sites proven to secrete WFDC2. In the interest of further clarity, it may be useful to expand the range of bacteria tested. Bacteria strains that have been reported to correlate with elevated WFDC2 levels would be interesting strains to analyse, including *Gardnerella vaginalis* (Orfanelli et al., 2014) and colonisers of cystic fibrosis lung such as *Pseudomonas aeruginosa* (Nagy et al., 2016). Generation of a dose-response curve may serve as a more appropriate means of resolutely showing whether WFDC2 can inhibit bacterial growth, although it is important to avoid concentrations that cannot occur naturally. Further analysis of WFDC2 levels in human secretions including epithelial secretions of the trachea, intestine and male and female reproductive tracts would ensure that the biological concentration of WFDC2 is well established. In the interest of consistency, it may also be more reliable to assess the concentration of WFDC2 in endogenous samples using the recombinant protein derived from this study as the standard when quantifying the WFDC2 concentration by ELISA.

One of the difficulties in assessing the activity of recombinant WFDC2 is that its biological activity is unknown; as a result, no WFDC2 control exists to show that the recombinant protein produced in this study was active under the defined test conditions. SLPI is a good control for

anti-protease activity as it has well-documented inhibitory activity against trypsin and elastase; this information only serves to add to the perplexity regarding the unusual elastase inhibition assay results (Section 4.4). Contrary to its potent anti-protease activity, SLPI only has moderate antibacterial activity and at a concentration of 1 µM it is unable to significantly reduce the number of colony forming units in any of the assays referenced in Table 1.1. This would therefore make it a poor candidate to use as a control in the bacterial assays described in Section 4.5.

The studies analysing the effect of cytokine exposure on cellular WFDC2 expression described in Sections 4.6.1. and 5.6. have only limited value, as they lack an appropriate positive control (such as IL-8) to show that the cytokines are actively capable of inducing downstream changes to gene expression. Additionally, the lack of change in WFDC2 expression shown by CAL27 cells following exposure to cytokines may be due to the cells reaching a limit in the amount of WFDC2 they are capable of secreting – CAL27 cells endogenously express very high levels of WFDC2 without the presence of stimuli.

Within the scope of this study CRISPR edited cancer cells were generated, with the aim of silencing expression of the predominant WFDC2 splice variant, V0. Although V0 protein production was successfully abolished via CRISPR editing, the expression of splice variant V2 appeared to become upregulated in place of V0 in the CRISPR cells, which may therefore have reinstated some WFDC2 function as a result of genetic redundancy. In the interest of clarity, it would be valuable to perform further gene editing to silence the remaining splice variants in the CAL27 cells and then repeat the cell behaviour assays. This would determine whether a complete knockout of all WFDC2 isoforms induces a more significant effect upon the cell phenotype and, in turn, would provide evidence as to the redundancy of V0 and V2 isoforms. It would also certainly be of interest to further analyse the differences between the CRISPR knockout and control cells by comparing their transcriptome using RNA-seq.

Finally, although the micro-CT analysis of homozygous Wfdc2 knockout mice was illuminating in demonstrating that the large airways of the mice were constricted, unfortunately, only limited

conclusions could be drawn from the data. It would be extremely interesting to histologically analyse the E14.5 and E18.5 embryos. Firstly, H&E staining could be used to show the cellular morphology of the tracheal and bronchial lumen, providing information as to how the tissue architecture was malformed. Secondly, immunohistochemistry could be used to visualise the presence of adherens junctions (via E-cadherin staining) and tight junctions (via ZO-1) to study any changes to the apical surface of the lumen. It would also be interesting to analyse whether any changes in tissue architecture occurred outside of the airways, for instance in the major salivary glands, large and small intestine, kidney and reproductive tracts, given the abundance of Wfdc2 identified at these sites in adult mice.

The potential studies that could be carried out using the knockout mouse embryos are extensive. One of the most crucial investigations would be to perform RNA-seq to compare the transcriptome of different tissues between knockout and WT embryos, to try to elucidate pathways affected by Wfdc2 silencing and to further determine its role in development. It would also be very interesting to isolate tracheal epithelial cells from conditional mutant mice and culture them *in vitro* to observe differences in growth characteristics and morphology between the WT and mutant cell populations.

7.2. Final conclusions and clinical relevance

The work described in this thesis suggests that WFDC2 is not simply an innocent bystander that becomes upregulated during tumorigenesis as a result of amplification of the 20q13 locus. Instead, WFDC2 appears to promote tumour progression by promoting cancer cell invasion, which could explain the correlation between WFDC2 levels and disease severity. Further rigorous investigations will be required to determine the exact mechanisms through which WFDC2 is able to induce cell invasion; however, it is possible that the protein may provide an interesting therapeutic target for inhibition of metastasis in the future.

Another factor worthy of further study is the WFDC2-encoding locus, 20q13, which is reported to undergo frequent amplifications. Assuming that this genetic event induces an upregulation

of WFDC2 expression (thus promoting its hypothesised pro-malignant effects), then it may provide an interesting marker to identify patients at risk of developing an ovarian malignancy. It would certainly be of clinical interest to study the frequency of 20q13 locus amplifications more closely in ovarian cancer patients, as the locus has the potential to act as a useful screening tool to identify both sporadic and familial cases of disease. Patients identified to be at higher cancer risk due to amplification of 20q13 would benefit from routine tests to give the greatest likelihood of an early cancer diagnosis.

Finally, evidence presented within this study has shown that WFDC2 splice variants have the potential to undergo compensatory upregulation when V0 expression is downregulated. This compensatory expression suggests that the isoforms exhibit redundant functions. With this in mind, it is possible that WFDC2 serum levels may be incorrectly quantified in the clinic via the use of the FDA-approved ELISA kit, which is only able to detect V0. If the V1-V4 splice variants are overexpressed in cancer patients without a concomitant upregulation in V0, the serum levels of a patient may be identified as being normal and the affected individual would go undiagnosed.

To conclude, this study has shown that WFDC2 is not capable of antimicrobial or anti-protease activity as is frequently hypothesised. Instead WFDC2 appears to play an, as yet, unidentified role in the development of the trachea and bronchi. Its effects on tissue morphology extend beyond the confines of embryonic development, as WFDC2 appears to affect cancer cell behaviour by promoting invasion. This suggests that WFDC2 is involved in the process of tumorigenesis, and as such, its overexpression in cancer is indicative of poor prognosis.

Bibliography

Abbinante-Nissen, J.M., Simpson, L.G., and Leikauf, G.D. (1993). Neutrophil elastase increases secretory leukocyte protease inhibitor transcript levels in airway epithelial cells. Am J Physiol *265*, L286-292.

Abbink, K., Zusterzeel, P.L., Geurts-Moespot, A.J., Herwaarden, A.E.V., Pijnenborg, J.M., Sweep, F.C., and Massuger, L.F. (2018). HE4 is superior to CA125 in the detection of recurrent disease in high-risk endometrial cancer patients. Tumour Biol *40*, 1010428318757103.

Aida, S., Shimazaki, H., Sato, K., Sato, M., Deguchi, H., Ozeki, Y., and Tamai, S. (2004). Prognostic analysis of pulmonary adenocarcinoma subclassification with special consideration of papillary and bronchioloalveolar types. Histopathology *45*, 468-476.

Akram, K.M., Moyo, N.A., Leeming, G.H., Bingle, L., Jasim, S., Hussain, S., Schorlemmer, A., Kipar, A., Digard, P., Tripp, R.A.*, et al.* (2018). An innate defense peptide BPIFA1/SPLUNC1 restricts influenza A virus infection. Mucosal Immunol *11*, 1008.

Alkemade, J.A., Molhuizen, H.O., Ponec, M., Kempenaar, J.A., Zeeuwen, P.L., de Jongh, G.J., van Vlijmen-Willems, I.M., van Erp, P.E., van de Kerkhof, P.C., and Schalkwijk, J. (1994). SKALP/elafin is an inducible proteinase inhibitor in human epidermal keratinocytes. J Cell Sci *107 (Pt 8)*, 2335-2342.

Anastasi, E., Granato, T., Falzarano, R., Storelli, P., Ticino, A., Frati, L., Panici, P.B., and Porpora, M.G. (2013). The use of HE4, CA125 and CA72-4 biomarkers for differential diagnosis between ovarian endometrioma and epithelial ovarian cancer. J Ovarian Res *6*, 44.

Ao, A., Hao, J., Hopkins, C.R., and Hong, C.C. (2012). DMH1, a novel BMP small molecule inhibitor, increases cardiomyocyte progenitors and promotes cardiac differentiation in mouse embryonic stem cells. PLoS One *7*, e41627.

Argüeso, P., Spurr-Michaud, S., Russo, C.L., Tisdale, A., and Gipson, I.K. (2003). MUC16 mucin is expressed by the human ocular surface epithelia and carries the H185 carbohydrate epitope. Invest Ophthalmol Vis Sci *44*, 2487-2495.

Aubin, J., Lemieux, M., Tremblay, M., Bérard, J., and Jeannotte, L. (1997). Early postnatal lethality in Hoxa-5 mutant mice is attributable to respiratory tract defects. Dev Biol *192*, 432-445.

Bandiera, E., Romani, C., Specchia, C., Zanotti, L., Galli, C., Ruggeri, G., Tognon, G., Bignotti, E., Tassi, R.A., Odicino, F.*, et al.* (2011). Serum human epididymis protein 4 and risk for ovarian malignancy algorithm as new diagnostic and prognostic tools for epithelial ovarian cancer management. Cancer Epidemiol Biomarkers Prev *20*, 2496-2506.

Baranger, K., Zani, M.L., Chandenier, J., Dallet-Choisy, S., and Moreau, T. (2008). The antibacterial and antifungal properties of trappin-2 (pre-elafin) do not depend on its protease inhibitory function. FEBS J *275*, 2008-2020.

Bast, R.C., Klug, T.L., St John, E., Jenison, E., Niloff, J.M., Lazarus, H., Berkowitz, R.S., Leavitt, T., Griffiths, C.T., Parker, L.*, et al.* (1983). A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. N Engl J Med *309*, 883-887.

Bellemare, A., Vernoux, N., Morin, S., Gagné, S.M., and Bourbonnais, Y. (2010). Structural and antimicrobial properties of human pre-elafin/trappin-2 and derived peptides against Pseudomonas aeruginosa. BMC Microbiol *10*, 253.

Betsuyaku, T., Takeyabu, K., Tanino, M., and Nishimura, M. (2002). Role of secretory leukocyte protease inhibitor in the development of subclinical emphysema. Eur Respir J *19*, 1051-1057.

Bingle, C.D., Araujo, B., Wallace, W.A., Hirani, N., and Bingle, L. (2013). What is top of the charts? BPIFB1/LPLUNC1 localises to the bronchiolised epithelium in the honeycomb cysts in UIP. Thorax *68*, 1167-1168.

Bingle, C.D., and Vyakarnam, A. (2008). Novel innate immune functions of the whey acidic protein family. Trends in immunology *29*, 444-453.

Bingle, L., Barnes, F.A., Lunn, H., Musa, M., Webster, S., Douglas, C.W., Cross, S.S., High, A.S., and Bingle, C.D. (2009). Characterisation and expression of SPLUNC2, the human orthologue of rodent parotid secretory protein. Histochem Cell Biol *132*, 339-349.

Bingle, L., Cross, S.S., High, A.S., Wallace, W.A., Rassl, D., Yuan, G., Hellstrom, I., Campos, M.A., and Bingle, C.D. (2006). WFDC2 (HE4): a potential role in the innate immunity of the oral cavity and respiratory tract and the development of adenocarcinomas of the lung. Respir Res *7*, 61.

Bingle, L., Singleton, V., and Bingle, C.D. (2002). The putative ovarian tumour marker gene HE4 (WFDC2), is expressed in normal tissues and undergoes complex alternative splicing to yield multiple protein isoforms. Oncogene *21*, 2768-2773.

Bingle, L., Tetley, T.D., and Bingle, C.D. (2001). Cytokine-mediated induction of the human elafin gene in pulmonary epithelial cells is regulated by nuclear factor-kappaB. Am J Respir Cell Mol Biol *25*, 84-91.

Bingle, L., Wilson, K., Musa, M., Araujo, B., Rassl, D., Wallace, W.A., LeClair, E.E., Mauad, T., Zhou, Z., Mall, M.A.*, et al.* (2012). BPIFB1 (LPLUNC1) is upregulated in cystic fibrosis lung disease. Histochem Cell Biol *138*, 749-758.

Birse, K.M., Burgener, A., Westmacott, G.R., McCorrister, S., Novak, R.M., and Ball, T.B. (2013). Unbiased proteomics analysis demonstrates significant variability in mucosal immune factor expression depending on the site and method of collection. PLoS One *8*, e79505.

Boehm, U., Klamp, T., Groot, M., and Howard, J.C. (1997). Cellular responses to interferon-gamma. Annu Rev Immunol *15*, 749-795.

Boudier, C., and Bieth, J.G. (1994). Oxidized mucus proteinase inhibitor: a fairly potent neutrophil elastase inhibitor. Biochem J *303 (Pt 1)*, 61-68.

Bowes, D., Clark, A.E., and Corrin, B. (1981). Ultrastructural localisation of lactoferrin and glycoprotein in human bronchial glands. Thorax *36*, 108-115.

Braicu, E.I., Fotopoulou, C., Van Gorp, T., Richter, R., Chekerov, R., Hall, C., Butz, H., Castillo-Tong, D.C., Mahner, S., Zeillinger, R.*, et al.* (2013). Preoperative HE4 expression in plasma predicts surgical outcome in primary ovarian cancer patients: results from the OVCAD study. Gynecol Oncol *128*, 245-251.

Brennan, D.J., Hackethal, A., Metcalf, A.M., Coward, J., Ferguson, K., Oehler, M.K., Quinn, M.A., Janda, M., Leung, Y., Freemantle, M.*, et al.* (2014). Serum HE4 as a prognostic marker in endometrial cancer--a population based study. Gynecol Oncol *132*, 159-165.

Cancer Research UK (2018a). Lung cancer statistics (https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer).

Cancer Research UK (2018b). Ovarian cancer statistics (https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer).

Cancer Research UK (2018c). Uterine cancer statistics (https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer).

Capriglione, S., Plotti, F., Miranda, A., Ricciardi, R., Scaletta, G., Aloisi, A., Guzzo, F., Montera, R., and Angioli, R. (2015). Utility of tumor marker HE4 as prognostic factor in endometrial cancer: a single-center controlled study. Tumour Biol *36*, 4151-4156.

Chan, K.K., Chen, C.A., Nam, J.H., Ochiai, K., Wilailak, S., Choon, A.T., Sabaratnam, S., Hebbar, S., Sickan, J., Schodin, B.A.*, et al.* (2013). The use of HE4 in the prediction of ovarian cancer in Asian women with a pelvic mass. Gynecol Oncol *128*, 239-244.

Chapman, B.S., Thayer, R.M., Vincent, K.A., and Haigwood, N.L. (1991). Effect of intron A from human cytomegalovirus (Towne) immediate-early gene on heterologous expression in mammalian cells. Nucleic Acids Res *19*, 3979-3986.

Charrier, L.E., Loie, E., and Laprise, P. (2015). Mouse Crumbs3 sustains epithelial tissue morphogenesis in vivo. Sci Rep *5*, 17699.

Chen, R., Wang, L., Liu, S., Chen, X., Hu, Y., Liu, H., Zhang, H., Jiang, Y., Wang, Q., Ye, D.*, et al.* (2017). Bcl-3 is a novel biomarker of renal fibrosis in chronic kidney disease. Oncotarget *8*, 97206-97216.

Chen, W.T., Gao, X., Han, X.D., Zheng, H., Guo, L., and Lu, R.Q. (2014). HE4 as a serum biomarker for ROMA prediction and prognosis of epithelial ovarian cancer. Asian Pac J Cancer Prev *15*, 101-105.

Chhikara, N., Saraswat, M., Tomar, A.K., Dey, S., Singh, S., and Yadav, S. (2012). Human epididymis protein-4 (HE-4): a novel cross-class protease inhibitor. PLoS One *7*, e47672.

Childers, N.K., Li, F., Dasanayake, A.P., Li, Y., Kirk, K., and Michalek, S.M. (2006). Immune response in humans to a nasal boost with Streptococcus mutans antigens. Oral Microbiol Immunol *21*, 309-313.

Choi, S.I., Jang, M.A., Jeon, B.R., Shin, H.B., Lee, Y.K., and Lee, Y.W. (2017). Clinical Usefulness of Human Epididymis Protein 4 in Lung Cancer. Ann Lab Med *37*, 526-530.

Clauss, A., Lilja, H., and Lundwall, A. (2002). A locus on human chromosome 20 contains several genes expressing protease inhibitor domains with homology to whey acidic protein. Biochem J *368*, 233-242.

Clauss, A., Ng, V., Liu, J., Piao, H., Russo, M., Vena, N., Sheng, Q., Hirsch, M.S., Bonome, T., Matulonis, U.*, et al.* (2010). Overexpression of elafin in ovarian carcinoma is driven by genomic gains and activation of the nuclear factor kappaB pathway and is associated with poor overall survival. Neoplasia *12*, 161-172.

Consortium., C.S.a.A. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. Nature *437*, 69-87.

Cross, B.W., and Ruhl, S. (2018). Glycan recognition at the saliva - oral microbiome interface. Cell Immunol.

Cunliffe, R.N., Rose, F.R., Keyte, J., Abberley, L., Chan, W.C., and Mahida, Y.R. (2001). Human defensin 5 is stored in precursor form in normal Paneth cells and is expressed by some villous epithelial cells and by metaplastic Paneth cells in the colon in inflammatory bowel disease. Gut *48*, 176-185.

Davies, J.R., Kirkham, S., Svitacheva, N., Thornton, D.J., and Carlstedt, I. (2007). MUC16 is produced in tracheal surface epithelium and submucosal glands and is present in secretions from normal human airway and cultured bronchial epithelial cells. Int J Biochem Cell Biol *39*, 1943-1954.

Denny, P., Hagen, F.K., Hardt, M., Liao, L., Yan, W., Arellanno, M., Bassilian, S., Bedi, G.S., Boontheung, P., Cociorva, D*., et al.* (2008). The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions. J Proteome Res *7*, 1994-2006.

Dewan, R., Dewan, A., Hare, S., Bhardwaj, M., and Mehrotra, K. (2017). Diagnostic Performance of Serum Human Epididymis Protein 4 in Endometrial Carcinoma: A Pilot Study. J Clin Diagn Res *11*, XC01-XC05.

Dohrman, A., Tsuda, T., Escudier, E., Cardone, M., Jany, B., Gum, J., Kim, Y., and Basbaum, C. (1994). Distribution of lysozyme and mucin (MUC2 and MUC3) mRNA in human bronchus. Exp Lung Res *20*, 367-380.

Drannik, A.G., Nag, K., Yao, X.D., Henrick, B.M., Jain, S., Ball, T.B., Plummer, F.A., Wachihi, C., Kimani, J., and Rosenthal, K.L. (2012). Anti-HIV-1 activity of elafin is more potent than its precursor's, trappin-2, in genital epithelial cells. J Virol *86*, 4599-4610.

Dranoff, G. (2004). Cytokines in cancer pathogenesis and cancer therapy. Nat Rev Cancer *4*, 11-22.

Drapkin, R., von Horsten, H.H., Lin, Y., Mok, S.C., Crum, C.P., Welch, W.R., and Hecht, J.L. (2005). Human epididymis protein 4 (HE4) is a secreted glycoprotein that is overexpressed by serous and endometrioid ovarian carcinomas. Cancer Res *65*, 2162-2169.

Drenth, J., Low, B.W., Richardson, J.S., and Wright, C.S. (1980). The toxin-agglutinin fold. A new group of small protein structures organized around a four-disulfide core. Journal of biological chemistry *255*, 2652-2655.

Dumont, J., Euwart, D., Mei, B., Estes, S., and Kshirsagar, R. (2016). Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. Crit Rev Biotechnol *36*, 1110-1122.

Dunn, G.P., Old, L.J., and Schreiber, R.D. (2004). The immunobiology of cancer immunosurveillance and immunoediting. Immunity *21*, 137-148.

Díaz-Padilla, I., Razak, A.R., Minig, L., Bernardini, M.Q., and María Del Campo, J. (2012). Prognostic and predictive value of CA-125 in the primary treatment of epithelial ovarian cancer: potentials and pitfalls. Clin Transl Oncol *14*, 15-20.

Eade, C.R., Diaz, C., Chen, S., Cole, A.L., and Cole, A.M. (2015). HIV-Enhancing Factors Are Secreted by Reproductive Epithelia upon Inoculation with Bacterial Vaginosis-Associated Bacteria. Protein Pept Lett *22*, 672-680.

Einhauer, A., and Jungbauer, A. (2001). The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. J Biochem Biophys Methods *49*, 455-465.

Eisenberg, S.P., Hale, K.K., Heimdal, P., and Thompson, R.C. (1990). Location of the protease-inhibitory region of secretory leukocyte protease inhibitor. J Biol Chem *265*, 7976-7981.

El-Brolosy, M.A., and Stainier, D.Y.R. (2017). Genetic compensation: A phenomenon in search of mechanisms. PLoS Genet *13*, e1006780.

El-Naggar, A.K., Chan, J.K.C., Grandis, J.R., Takata, T., and Slootweg, P.J. (2017). WHO Classification of Head and Neck Tumours.

Ellerbrock, K., Pera, I., Hartung, S., and Ivell, R. (1994). Gene expression in the dog epididymis: a model for human epididymal function. Int J Androl *17*, 314-323.

Ericson, T., and Rundegren, J. (1983). Characterization of a salivary agglutinin reacting with a serotype c strain of Streptococcus mutans. Eur J Biochem *133*, 255-261.

Erlandsen, S.L., Parsons, J.A., and Taylor, T.D. (1974). Ultrastructural immunocytochemical localization of lysozyme in the Paneth cells of man. J Histochem Cytochem *22*, 401-413.

Erlanger, B.F., Kokowsky, N., and Cohen, W. (1961). The preparation and properties of two new chromogenic substrates of trypsin. Arch Biochem Biophys *95*, 271-278.

Escudero, J.M., Auge, J.M., Filella, X., Torne, A., Pahisa, J., and Molina, R. (2011). Comparison of serum human epididymis protein 4 with cancer antigen 125 as a tumor marker in patients with malignant and nonmalignant diseases. Clin Chem *57*, 1534-1544.

Fang, X., Yang, L., Wang, W., Song, T., Lee, C.S., DeVoe, D.L., and Balgley, B.M. (2007). Comparison of electrokinetics-based multidimensional separations coupled with electrospray ionization-tandem mass spectrometry for characterization of human salivary proteins. Anal Chem *79*, 5785-5792.

Fernie-King, B.A., Seilly, D.J., Davies, A., and Lachmann, P.J. (2002). Streptococcal inhibitor of complement inhibits two additional components of the mucosal innate immune system: secretory leukocyte proteinase inhibitor and lysozyme. Infect Immun *70*, 4908-4916.

Ferreira, Z., Hurle, B., Andrés, A.M., Kretzschmar, W.W., Mullikin, J.C., Cherukuri, P.F., Cruz, P., Gonder, M.K., Stone, A.C., Tishkoff, S.*, et al.* (2013a). Sequence diversity of Pan troglodytes subspecies and the impact of WFDC6 selective constraints in reproductive immunity. Genome Biol Evol *5*, 2512-2523.

Ferreira, Z., Seixas, S., Andrés, A.M., Kretzschmar, W.W., Mullikin, J.C., Cherukuri, P.F., Cruz, P., Swanson, W.J., Clark, A.G., Green, E.D.*, et al.* (2013b). Reproduction and immunity-driven natural selection in the human WFDC locus. Mol Biol Evol *30*, 938-950.

Franken, C., Meijer, C.J., and Dijkman, J.H. (1989). Tissue distribution of antileukoprotease and lysozyme in humans. J Histochem Cytochem *37*, 493-498.

Franzke, C.W., Baici, A., Bartels, J., Christophers, E., and Wiedow, O. (1996). Antileukoprotease inhibits stratum corneum chymotryptic enzyme. Evidence for a regulative function in desquamation. J Biol Chem *271*, 21886-21890.

Frey, A.M., Satur, M.J., Phansopa, C., Parker, J.L., Bradshaw, D., Pratten, J., and Stafford, G.P. (2018). Evidence for a carbohydrate-binding module (CBM) of. Biochem J *475*, 1159-1176.

Galgano, M.T., Hampton, G.M., and Frierson, H.F. (2006). Comprehensive analysis of HE4 expression in normal and malignant human tissues. Mod Pathol *19*, 847-853.

Gao, L., Cheng, H.Y., Dong, L., Ye, X., Liu, Y.N., Chang, X.H., Cheng, Y.X., Chen, J., Ma, R.Q., and Cui, H. (2011). The role of HE4 in ovarian cancer: inhibiting tumour cell proliferation and metastasis. J Int Med Res *39*, 1645-1660.

Gauthier, F., Fryksmark, U., Ohlsson, K., and Bieth, J.G. (1982). Kinetics of the inhibition of leukocyte elastase by the bronchial inhibitor. Biochim Biophys Acta *700*, 178-183.

Georgakopoulos, P., Mehmood, S., Akalin, A., and Shroyer, K.R. (2012). Immunohistochemical localization of HE4 in benign, borderline, and malignant lesions of the ovary. Int J Gynecol Pathol *31*, 517-523.

Gomez, S.A., Argüelles, C.L., Guerrieri, D., Tateosian, N.L., Amiano, N.O., Slimovich, R., Maffia, P.C., Abbate, E., Musella, R.M., Garcia, V.E.*, et al.* (2009). Secretory leukocyte

protease inhibitor: a secreted pattern recognition receptor for mycobacteria. Am J Respir Crit Care Med *179*, 247-253.

Grubor, B., Meyerholz, D.K., and Ackermann, M.R. (2006). Collectins and cationic antimicrobial peptides of the respiratory epithelia. Vet Pathol *43*, 595-612.

Grütter, M.G., Fendrich, G., Huber, R., and Bode, W. (1988). The 2.5 A X-ray crystal structure of the acid-stable proteinase inhibitor from human mucous secretions analysed in its complex with bovine alpha-chymotrypsin. The EMBO journal *7*, 345-351.

Guo, Y.D., Wang, J.H., Lu, H., Li, X.N., Song, W.W., Zhang, X.D., and Zhang, W.M. (2015). The human epididymis protein 4 acts as a prognostic factor and promotes progression of gastric cancer. Tumour Biol *36*, 2457-2464.

Gurung, S., Werkmeister, J.A., and Gargett, C.E. (2015). Inhibition of Transforming Growth Factor-β Receptor signaling promotes culture expansion of undifferentiated human Endometrial Mesenchymal Stem/stromal Cells. Sci Rep *5*, 15042.

Hellström, I., Raycraft, J., Hayden-Ledbetter, M., Ledbetter, J.A., Schummer, M., McIntosh, M., Drescher, C., Urban, N., and Hellström, K.E. (2003). The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. Cancer Res *63*, 3695-3700.

Helmig, R., Uldbjerg, N., and Ohlsson, K. (1995). Secretory leukocyte protease inhibitor in the cervical mucus and in the fetal membranes. Eur J Obstet Gynecol Reprod Biol *59*, 95-101.

Hennighausen, L.G., and Sippel, A.E. (1982). Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. Nucleic Acids Res *10*, 2677-2684.

Hertlein, L., Stieber, P., Kirschenhofer, A., Krocker, K., Nagel, D., Lenhard, M., and Burges, A. (2012). Human epididymis protein 4 (HE4) in benign and malignant diseases. Clin Chem Lab Med *50*, 2181-2188.

Hiemstra, P.S., Maassen, R.J., Stolk, J., Heinzel-Wieland, R., Steffens, G.J., and Dijkman, J.H. (1996). Antibacterial activity of antileukoprotease. Infect Immun *64*, 4520-4524.

Hough, C.D., Sherman-Baust, C.A., Pizer, E.S., Montz, F.J., Im, D.D., Rosenshein, N.B., Cho, K.R., Riggins, G.J., and Morin, P.J. (2000). Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer. Cancer Res *60*, 6281-6287.

Hua, L., Liu, Y., Zhen, S., Wan, D., Cao, J., and Gao, X. (2014). Expression and biochemical characterization of recombinant human epididymis protein 4. Protein Expr Purif *102*, 52-62.

Hurle, B., Swanson, W., Green, E.D., and Program, N.C.S. (2007). Comparative sequence analyses reveal rapid and divergent evolutionary changes of the WFDC locus in the primate lineage. Genome Res *17*, 276-286.

Hussain, S.P., and Harris, C.C. (2007). Inflammation and cancer: an ancient link with novel potentials. Int J Cancer *121*, 2373-2380.

Ismail, R.S., Baldwin, R.L., Fang, J., Browning, D., Karlan, B.Y., Gasson, J.C., and Chang, D.D. (2000). Differential gene expression between normal and tumor-derived ovarian epithelial cells. Cancer Res *60*, 6744-6749.

Israeli, O., Goldring-Aviram, A., Rienstein, S., Ben-Baruch, G., Korach, J., Goldman, B., and Friedman, E. (2005). In silico chromosomal clustering of genes displaying altered expression patterns in ovarian cancer. Cancer Genet Cytogenet *160*, 35-42.

Iwabuchi, H., Sakamoto, M., Sakunaga, H., Ma, Y.Y., Carcangiu, M.L., Pinkel, D., Yang-Feng, T.L., and Gray, J.W. (1995). Genetic analysis of benign, low-grade, and high-grade ovarian tumors. Cancer Res *55*, 6172-6180.

Iwahori, K., Suzuki, H., Kishi, Y., Fujii, Y., Uehara, R., Okamoto, N., Kobayashi, M., Hirashima, T., Kawase, I., and Naka, T. (2012). Serum HE4 as a diagnostic and prognostic marker for lung cancer. Tumour Biol *33*, 1141-1149.

Jacob, F., Meier, M., Caduff, R., Goldstein, D., Pochechueva, T., Hacker, N., Fink, D., and Heinzelmann-Schwarz, V. (2011). No benefit from combining HE4 and CA125 as ovarian tumor markers in a clinical setting. Gynecol Oncol *121*, 487-491.

Jiang, L., Ji, N., Zhou, Y., Li, J., Liu, X., Wang, Z., Chen, Q., and Zeng, X. (2009). CAL 27 is an oral adenosquamous carcinoma cell line. Oral Oncol *45*, e204-207.

Jiang, S.W., Chen, H., Dowdy, S., Fu, A., Attewell, J., Kalogera, E., Drapkin, R., Podratz, K., Broaddus, R., and Li, J. (2013). HE4 transcription- and splice variants-specific expression in endometrial cancer and correlation with patient survival. Int J Mol Sci *14*, 22655-22677.

Jiang, Y., Wang, C., Lv, B., Ma, G., and Wang, L. (2014). Expression level of serum human epididymis 4 and its prognostic significance in human non-small cell lung cancer. Int J Clin Exp Med *7*, 5568-5572.

Jin, F., Nathan, C.F., Radzioch, D., and Ding, A. (1998). Lipopolysaccharide-related stimuli induce expression of the secretory leukocyte protease inhibitor, a macrophage-derived lipopolysaccharide inhibitor. Infect Immun *66*, 2447-2452.

Jin, F.Y., Nathan, C., Radzioch, D., and Ding, A. (1997). Secretory leukocyte protease inhibitor: a macrophage product induced by and antagonistic to bacterial lipopolysaccharide. Cell *88*, 417-426.

Johansson, M.E., Thomsson, K.A., and Hansson, G.C. (2009). Proteomic analyses of the two mucus layers of the colon barrier reveal that their main component, the Muc2 mucin, is strongly bound to the Fcgbp protein. J Proteome Res *8*, 3549-3557.

Jordan, M., Köhne, C., and Wurm, F.M. (1998). Calcium-phosphate mediated DNA transfer into HEK-293 cells in suspension: control of physicochemical parameters allows transfection in stirred media. Transfection and protein expression in mammalian cells. Cytotechnology *26*, 39-47.

Kamei, M., Yamashita, S., Tokuishi, K., Hashioto, T., Moroga, T., Suehiro, S., Ono, K., Miyawaki, M., Takeno, S., Yamamoto, S.*, et al.* (2010). HE4 expression can be associated with lymph node metastases and disease-free survival in breast cancer. Anticancer Res *30*, 4779-4783.

Kappler-Hanno, K., and Kirchhoff, C. (2003). Rodent epididymal cDNAs identified by sequence homology to human and canine counterparts. Asian J Androl *5*, 277-286.

Karlsen, M.A., Sandhu, N., Høgdall, C., Christensen, I.J., Nedergaard, L., Lundvall, L., Engelholm, S.A., Pedersen, A.T., Hartwell, D., Lydolph, M.*, et al.* (2012). Evaluation of HE4, CA125, risk of ovarian malignancy algorithm (ROMA) and risk of malignancy index (RMI) as diagnostic tools of epithelial ovarian cancer in patients with a pelvic mass. Gynecol Oncol *127*, 379-383.

Karp, C.M., Tan, T.T., Mathew, R., Nelson, D., Mukherjee, C., Degenhardt, K., Karantza-Wadsworth, V., and White, E. (2008). Role of the polarity determinant crumbs in suppressing mammalian epithelial tumor progression. Cancer Res *68*, 4105-4115.

Kesimer, M., Kirkham, S., Pickles, R.J., Henderson, A.G., Alexis, N.E., Demaria, G., Knight, D., Thornton, D.J., and Sheehan, J.K. (2009). Tracheobronchial air-liquid interface cell culture: a model for innate mucosal defense of the upper airways? Am J Physiol Lung Cell Mol Physiol *296*, L92-L100.

Kiechle, M., Jacobsen, A., Schwarz-Boeger, U., Hedderich, J., Pfisterer, J., and Arnold, N. (2001). Comparative genomic hybridization detects genetic imbalances in primary ovarian carcinomas as correlated with grade of differentiation. Cancer *91*, 534-540.

Kim, S., Takahashi, H., Lin, W.W., Descargues, P., Grivennikov, S., Kim, Y., Luo, J.L., and Karin, M. (2009). Carcinoma-produced factors activate myeloid cells through TLR2 to stimulate metastasis. Nature *457*, 102-106.

King, A.E., Critchley, H.O., Sallenave, J.M., and Kelly, R.W. (2003). Elafin in human endometrium: an antiprotease and antimicrobial molecule expressed during menstruation. J Clin Endocrinol Metab *88*, 4426-4431.

Kirchhoff, C., Habben, I., Ivell, R., and Krull, N. (1991). A major human epididymis-specific cDNA encodes a protein with sequence homology to extracellular proteinase inhibitors. Biology of reproduction *45*, 350-357.

Kirchhoff, C., Osterhoff, C., Habben, I., Ivell, R., and Kirchloff, C. (1990). Cloning and analysis of mRNAs expressed specifically in the human epididymis. Int J Androl *13*, 155-167.

Kishimoto, K., Tamura, M., Nishita, M., Minami, Y., Yamaoka, A., Abe, T., Shigeta, M., and Morimoto, M. (2018). Synchronized mesenchymal cell polarization and differentiation shape the formation of the murine trachea and esophagus. Nat Commun *9*, 2816.

Korsrud, F.R., and Brandtzaeg, P. (1982). Characterization of epithelial elements in human major salivary glands by functional markers: localization of amylase, lactoferrin, lysozyme, secretory component, and secretory immunoglobulins by paired immunofluorescence staining. J Histochem Cytochem *30*, 657-666.

Kramps, J.A., van Twisk, C., Appelhans, H., Meckelein, B., Nikiforov, T., and Dijkman, J.H. (1990). Proteinase inhibitory activities of antileukoprotease are represented by its second COOH-terminal domain. Biochim Biophys Acta *1038*, 178-185.

Kulpa, J., Wójcik, E., Reinfuss, M., and Kołodziejski, L. (2002). Carcinoembryonic antigen, squamous cell carcinoma antigen, CYFRA 21-1, and neuron-specific enolase in squamous cell lung cancer patients. Clin Chem *48*, 1931-1937.

Lamy, P.J., Plassot, C., and Pujol, J.L. (2015). Serum HE4: An Independent Prognostic Factor in Non-Small Cell Lung Cancer. PLoS One *10*, e0128836.

Lan, W.G., Hao, Y.Z., Xu, D.H., Wang, P., Zhou, Y.L., and Ma, L.B. (2016). Serum human epididymis protein 4 is associated with the treatment response of concurrent chemoradiotherapy and prognosis in patients with locally advanced non-small cell lung cancer. Clin Transl Oncol *18*, 375-380.

Laprise, P., Paul, S.M., Boulanger, J., Robbins, R.M., Beitel, G.J., and Tepass, U. (2010). Epithelial polarity proteins regulate Drosophila tracheal tube size in parallel to the luminal matrix pathway. Curr Biol *20*, 55-61.

Larramendy, M.L., Lushnikova, T., Björkqvist, A.M., Wistuba, I.I., Virmani, A.K., Shivapurkar, N., Gazdar, A.F., and Knuutila, S. (2000). Comparative genomic hybridization reveals complex genetic changes in primary breast cancer tumors and their cell lines. Cancer Genet Cytogenet *119*, 132-138.

LeBleu, V.S., Teng, Y., O'Connell, J.T., Charytan, D., Müller, G.A., Müller, C.A., Sugimoto, H., and Kalluri, R. (2013). Identification of human epididymis protein-4 as a fibroblast-derived mediator of fibrosis. Nat Med *19*, 227-231.

Lechner, J.F., Haugen, A., McClendon, I.A., and Pettis, E.W. (1982). Clonal growth of normal adult human bronchial epithelial cells in a serum-free medium. In Vitro *18*, 633-642.

Lee, S.K., Lee, S.S., Hirose, S., Park, S.C., Chi, J.G., Chung, S.I., and Mori, M. (2002). Elafin expression in human fetal and adult submandibular glands. Histochem Cell Biol *117*, 423-430.

Lee, W.L., and Downey, G.P. (2001). Leukocyte elastase: physiological functions and role in acute lung injury. Am J Respir Crit Care Med *164*, 896-904.

Li, C., Xiao, J., Hormi, K., Borok, Z., and Minoo, P. (2002). Wnt5a participates in distal lung morphogenesis. Dev Biol *248*, 68-81.

Li, J., Chen, H., Mariani, A., Chen, D., Klatt, E., Podratz, K., Drapkin, R., Broaddus, R., Dowdy, S., and Jiang, S.W. (2013). HE4 (WFDC2) Promotes Tumor Growth in Endometrial Cancer Cell Lines. Int J Mol Sci *14*, 6026-6043.

Lin, A.L., Johnson, D.A., Stephan, K.T., and Yeh, C.K. (2004). Salivary secretory leukocyte protease inhibitor increases in HIV infection. J Oral Pathol Med *33*, 410-416.

Liu, L., Teng, J., Zhang, L., Cong, P., Yao, Y., Sun, G., Liu, Z., Yu, T., and Liu, M. (2017). The Combination of the Tumor Markers Suggests the Histological Diagnosis of Lung Cancer. Biomed Res Int *2017*, 2013989.

Liu, W., Yang, J., Chi, P.D., Zheng, X., Dai, S.Q., Chen, H., Xu, B.L., and Liu, W.L. (2013). Evaluating the clinical significance of serum HE4 levels in lung cancer and pulmonary tuberculosis. Int J Tuberc Lung Dis *17*, 1346-1353.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Lou, E., Johnson, M., Sima, C., Gonzalez-Espinoza, R., Fleisher, M., Kris, M.G., and Azzoli, C.G. (2014). Serum biomarkers for assessing histology and outcomes in patients with metastatic lung cancer. Cancer Biomark *14*, 207-214.

Lu, Q., Chen, H., Senkowski, C., Wang, J., Wang, X., Brower, S., Glasgow, W., Byck, D., Jiang, S.W., and Li, J. (2016). Recombinant HE4 protein promotes proliferation of pancreatic and endometrial cancer cell lines. Oncol Rep *35*, 163-170.

Lu, R., Sun, X., Xiao, R., Zhou, L., Gao, X., and Guo, L. (2012). Human epididymis protein 4 (HE4) plays a key role in ovarian cancer cell adhesion and motility. Biochem Biophys Res Commun *419*, 274-280.

Lundwall, Å., and Clauss, A. (2011). Genes encoding WFDC- and Kunitz-type protease inhibitor domains: are they related? Biochem Soc Trans *39*, 1398-1402.

Lunn, H.L. (2013). Purification and Functional Analysis of BPIFA2.

Ma, Q., Ma, Y., Dai, X., Ren, T., Fu, Y., Liu, W., Han, Y., Wu, Y., Cheng, Y., Zhang, T.*, et al.* (2018). Regeneration of functional alveoli by adult human SOX9. Protein Cell *9*, 267-282.

Mao, X., Li, P., Wang, Y., Liang, Z., Liu, J., Li, J., Jiang, Y., Bao, G., Li, L., Zhu, B.*, et al.* (2017). CRB3 regulates contact inhibition by activating the Hippo pathway in mammary epithelial cells. Cell Death Dis *8*, e2546.

Masuda, K., Kamimura, T., Watanabe, K., Suga, T., Kanesaki, M., Takeuchi, A., Imaizumi, A., and Suzuki, Y. (1995). Pharmacological activity of the C-terminal and N-terminal domains of secretory leukoprotease inhibitor in vitro. Br J Pharmacol *115*, 883-888.

McCaffrey, L.M., Montalbano, J., Mihai, C., and Macara, I.G. (2012). Loss of the Par3 polarity protein promotes breast tumorigenesis and metastasis. Cancer Cell *22*, 601-614.

McCartney-Francis, N., Jin, W., Belkaid, Y., McGrady, G., and Wahl, S.M. (2014). Aberrant host defense against Leishmania major in the absence of SLPI. J Leukoc Biol *96*, 917-929.

Medeiros, L.R., Rosa, D.D., da Rosa, M.I., and Bozzetti, M.C. (2009). Accuracy of CA 125 in the diagnosis of ovarian tumors: a quantitative systematic review. Eur J Obstet Gynecol Reprod Biol *142*, 99-105.

Meyer-Hoffert, U., Wichmann, N., Schwichtenberg, L., White, P.C., and Wiedow, O. (2003). Supernatants of Pseudomonas aeruginosa induce the Pseudomonas-specific antibiotic elafin in human keratinocytes. Exp Dermatol *12*, 418-425.

Mihaila, A., and Tremblay, G.M. (2001). Human alveolar macrophages express elafin and secretory leukocyte protease inhibitor. Z Naturforsch C *56*, 291-297.

Miller, C.H., Maher, S.G., and Young, H.A. (2009). Clinical Use of Interferon-gamma. Ann N Y Acad Sci *1182*, 69-79.

Miyoshi, T., Satoh, Y., Okumura, S., Nakagawa, K., Shirakusa, T., Tsuchiya, E., and Ishikawa, Y. (2003). Early-stage lung adenocarcinomas with a micropapillary pattern, a distinct pathologic marker for a significantly poor prognosis. Am J Surg Pathol *27*, 101-109.

Molina, R., Escudero, J.M., Augé, J.M., Filella, X., Foj, L., Torné, A., Lejarcegui, J., and Pahisa, J. (2011). HE4 a novel tumour marker for ovarian cancer: comparison with CA 125 and ROMA algorithm in patients with gynaecological diseases. Tumour Biol *32*, 1087-1095.

Moore, R.G., Brown, A.K., Miller, M.C., Badgwell, D., Lu, Z., Allard, W.J., Granai, C.O., Bast, R.C., and Lu, K. (2008). Utility of a novel serum tumor biomarker HE4 in patients with endometrioid adenocarcinoma of the uterus. Gynecol Oncol *110*, 196-201.

Moore, R.G., Hill, E.K., Horan, T., Yano, N., Kim, K., MacLaughlan, S., Lambert-Messerlian, G., Tseng, Y.D., Padbury, J.F., Miller, M.C.*, et al.* (2014). HE4 (WFDC2) gene overexpression promotes ovarian tumor growth. Sci Rep *4*, 3574.

Moore, R.G., Jabre-Raughley, M., Brown, A.K., Robison, K.M., Miller, M.C., Allard, W.J., Kurman, R.J., Bast, R.C., and Skates, S.J. (2010). Comparison of a novel multiple marker assay vs the Risk of Malignancy Index for the prediction of epithelial ovarian cancer in patients with a pelvic mass. Am J Obstet Gynecol *203*, 228.e221-226.

Moore, R.G., McMeekin, D.S., Brown, A.K., DiSilvestro, P., Miller, M.C., Allard, W.J., Gajewski, W., Kurman, R., Bast, R.C., and Skates, S.J. (2009). A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. Gynecol Oncol *112*, 40-46.

Moore, R.G., Miller, M.C., Disilvestro, P., Landrum, L.M., Gajewski, W., Ball, J.J., and Skates, S.J. (2011). Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass. Obstet Gynecol *118*, 280-288.

Moremen, K.W., Tiemeyer, M., and Nairn, A.V. (2012). Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol *13*, 448-462.

Mulay, A., Akram, K.M., Williams, D., Armes, H., Russell, C., Hood, D., Armstrong, S., Stewart, J.P., Brown, S.D., Bingle, L.*, et al.* (2016). An in vitro model of murine middle ear epithelium. Dis Model Mech *9*, 1405-1417.

Myat, M.M., and Andrew, D.J. (2002). Epithelial tube morphology is determined by the polarized growth and delivery of apical membrane. Cell *111*, 879-891.

Nagy, B., Fila, L., Clarke, L.A., Gönczy, F., Bede, O., Nagy, D., Újhelyi, R., Szabó, Á., Anghelyi, A., Major, M.*, et al.* (2016). Human epididymis protein 4 (HE4): a novel serum inflammatory biomarker in cystic fibrosis. Chest.

Nagy, B., Krasznai, Z.T., Balla, H., Csobán, M., Antal-Szalmás, P., Hernádi, Z., and Kappelmayer, J. (2012). Elevated human epididymis protein 4 concentrations in chronic kidney disease. Ann Clin Biochem *49*, 377-380.

Nakamura, A., Mori, Y., Hagiwara, K., Suzuki, T., Sakakibara, T., Kikuchi, T., Igarashi, T., Ebina, M., Abe, T., Miyazaki, J*., et al.* (2003). Increased susceptibility to LPS-induced endotoxin shock in secretory leukoprotease inhibitor (SLPI)-deficient mice. J Exp Med *197*, 669-674.

Nemajerova, A., Kramer, D., Siller, S.S., Herr, C., Shomroni, O., Pena, T., Gallinas Suazo, C., Glaser, K., Wildung, M., Steffen, H*., et al.* (2016). TAp73 is a central transcriptional regulator of airway multiciliogenesis. Genes Dev *30*, 1300-1312.

Nishimura, J., Saiga, H., Sato, S., Okuyama, M., Kayama, H., Kuwata, H., Matsumoto, S., Nishida, T., Sawa, Y., Akira, S*., et al.* (2008). Potent antimycobacterial activity of mouse secretory leukocyte protease inhibitor. J Immunol *180*, 4032-4039.

Nishimura, T. (2008). Total number of genome alterations in sporadic gastrointestinal cancer inferred from pooled analyses in the literature. Tumour Biol *29*, 343-350.

Nozaki, K., Ogawa, M., Williams, J.A., Lafleur, B.J., Ng, V., Drapkin, R.I., Mills, J.C., Konieczny, S.F., Nomura, S., and Goldenring, J.R. (2008). A molecular signature of gastric metaplasia arising in response to acute parietal cell loss. Gastroenterology *134*, 511-522.

Nustad, K., Bast, R.C., Brien, T.J., Nilsson, O., Seguin, P., Suresh, M.R., Saga, T., Nozawa, S., Børmer, O.P., de Bruijn, H.W*., et al.* (1996). Specificity and affinity of 26 monoclonal antibodies against the CA 125 antigen: first report from the ISOBM TD-1 workshop. International Society for Oncodevelopmental Biology and Medicine. Tumour Biol *17*, 196-219.

Ohlsson, K., Bjartell, A., and Lilja, H. (1995). Secretory leucocyte protease inhibitor in the male genital tract: PSA-induced proteolytic processing in human semen and tissue localization. J Androl *16*, 64-74.

Ohlsson, S., Ljungkrantz, I., Ohlsson, K., Segelmark, M., and Wieslander, J. (2001). Novel distribution of the secretory leucocyte proteinase inhibitor in kidney. Mediators Inflamm *10*, 347-350.

Okubo, K., Mark, E.J., Flieder, D., Wain, J.C., Wright, C.D., Moncure, A.C., Grillo, H.C., and Mathisen, D.J. (1999). Bronchoalveolar carcinoma: clinical, radiologic, and pathologic factors and survival. J Thorac Cardiovasc Surg *118*, 702-709.

Ono, K., Tanaka, T., Tsunoda, T., Kitahara, O., Kihara, C., Okamoto, A., Ochiai, K., Takagi, T., and Nakamura, Y. (2000). Identification by cDNA microarray of genes involved in ovarian carcinogenesis. Cancer Res *60*, 5007-5011.

Orfanelli, T., Jayaram, A., Doulaveris, G., Forney, L.J., Ledger, W.J., and Witkin, S.S. (2014). Human epididymis protein 4 and secretory leukocyte protease inhibitor in vaginal fluid: relation to vaginal components and bacterial composition. Reprod Sci *21*, 538-542.

Orosz, P., Krüger, A., Hubbe, M., Rüschoff, J., Von Hoegen, P., and Männel, D.N. (1995). Promotion of experimental liver metastasis by tumor necrosis factor. Int J Cancer *60*, 867-871.

Ortiz-Muñoz, B., Aznar-Oroval, E., García García, A., Covisa Peris, A., Perez Ballestero, P., Sanchez Yepes, M., Garcia Lozano, T., Illueca Ballester, C., and García Garcia, E. (2014). HE4, Ca125 and ROMA algorithm for differential diagnosis between benign gynaecological diseases and ovarian cancer. Tumour Biol *35*, 7249-7258.

Osawa, E., Nakajima, A., Fujisawa, T., Kawamura, Y.I., Toyama-Sorimachi, N., Nakagama, H., and Dohi, T. (2006). Predominant T helper type 2-inflammatory responses promote murine colon cancers. Int J Cancer *118*, 2232-2236.

Payne, J.E., Dubois, A.V., Ingram, R.J., Weldon, S., Taggart, C.C., Elborn, J.S., and Tunney, M.M. (2017). Activity of innate antimicrobial peptides and ivacaftor against clinical cystic fibrosis respiratory pathogens. Int J Antimicrob Agents *50*, 427-435.

Pellikka, M., Tanentzapf, G., Pinto, M., Smith, C., McGlade, C.J., Ready, D.F., and Tepass, U. (2002). Crumbs, the Drosophila homologue of human CRB1/RP12, is essential for photoreceptor morphogenesis. Nature *416*, 143-149.

Pera, I., Ivell, R., and Kirchhoff, C. (1994). Regional variation of specific gene expression in the dog epididymis as revealed by in-situ transcript hybridization. Int J Androl *17*, 324-330.

Peters-Hall, J.R., Brown, K.J., Pillai, D.K., Tomney, A., Garvin, L.M., Wu, X., and Rose, M.C. (2015). Quantitative proteomics reveals an altered cystic fibrosis in vitro bronchial epithelial secretome. Am J Respir Cell Mol Biol *53*, 22-32.

Petersen, I., Hidalgo, A., Petersen, S., Schlüns, K., Schewe, C., Pacyna-Gengelbach, M., Goeze, A., Krebber, B., Knösel, T., Kaufmann, O*., et al.* (2000). Chromosomal imbalances in brain metastases of solid tumors. Brain Pathol *10*, 395-401.

Pfundt, R., van Ruissen, F., van Vlijmen-Willems, I.M., Alkemade, H.A., Zeeuwen, P.L., Jap, P.H., Dijkman, H., Fransen, J., Croes, H., van Erp, P.E*., et al.* (1996). Constitutive and inducible expression of SKALP/elafin provides anti-elastase defense in human epithelia. J Clin Invest *98*, 1389-1399.

Pilch, B., and Mann, M. (2006). Large-scale and high-confidence proteomic analysis of human seminal plasma. Genome Biol *7*, R40.

Piletz, J.E., Heinlen, M., and Ganschow, R.E. (1981). Biochemical characterization of a novel whey protein from murine milk. J Biol Chem *256*, 11509-11516.

Powell, J.L., Hill, K.A., Shiro, B.C., Diehl, S.J., and Gajewski, W.H. (2005). Preoperative serum CA-125 levels in treating endometrial cancer. J Reprod Med *50*, 585-590.

Radicioni, G., Cao, R., Carpenter, J., Ford, A.A., Wang, T.T., Li, Y., and Kesimer, M. (2016). The innate immune properties of airway mucosal surfaces are regulated by dynamic interactions between mucins and interacting proteins: the mucin interactome. Mucosal Immunol *9*, 1442-1454.

Raghu, G., Richeldi, L., Jagerschmidt, A., Martin, V., Subramaniam, A., Ozoux, M.L., Esperet, C.A., and Soubrane, C. (2018). Idiopathic Pulmonary Fibrosis: Prospective, Case-Controlled Study of Natural History and Circulating Biomarkers. Chest *154*, 1359-1370.

Ramachandran, P., Boontheung, P., Xie, Y., Sondej, M., Wong, D.T., and Loo, J.A. (2006). Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry. J Proteome Res *5*, 1493-1503.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308.

Ranganathan, S., Simpson, K.J., Shaw, D.C., and Nicholas, K.R. (1999). The whey acidic protein family: a new signature motif and three-dimensional structure by comparative modeling. Journal of molecular graphics and modelling *17*, 106-106.

Ray, S., Fanti, J.A., Macedo, D.P., and Larsen, M. (2014). LIM kinase regulation of cytoskeletal dynamics is required for salivary gland branching morphogenesis. Mol Biol Cell *25*, 2393-2407.

Reid, P.T., Marsden, M.E., Cunningham, G.A., Haslett, C., and Sallenave, J.M. (1999). Human neutrophil elastase regulates the expression and secretion of elafin (elastase-specific inhibitor) in type II alveolar epithelial cells. FEBS Lett *457*, 33-37.

Rodríguez-Piñeiro, A.M., Bergström, J.H., Ermund, A., Gustafsson, J.K., Schütte, A., Johansson, M.E., and Hansson, G.C. (2013). Studies of mucus in mouse stomach, small intestine, and colon. II. Gastrointestinal mucus proteome reveals Muc2 and Muc5ac accompanied by a set of core proteins. Am J Physiol Gastrointest Liver Physiol *305*, G348-356.

Rokicki, W., Rokicki, M., Wojtacha, J., and Dżeljijli, A. (2016). The role and importance of club cells (Clara cells) in the pathogenesis of some respiratory diseases. Kardiochir Torakochirurgia Pol *13*, 26-30.

Ruggeri, G., Bandiera, E., Zanotti, L., Belloli, S., Ravaggi, A., Romani, C., Bignotti, E., Tassi, R.A., Tognon, G., Galli, C.*, et al.* (2011). HE4 and epithelial ovarian cancer: comparison and clinical evaluation of two immunoassays and a combination algorithm. Clin Chim Acta *412*, 1447-1453.

Saitoh, H., Masuda, T., Shimura, S., Fushimi, T., and Shirato, K. (2001). Secretion and gene expression of secretory leukocyte protease inhibitor by human airway submucosal glands. Am J Physiol Lung Cell Mol Physiol *280*, L79-87.

Sallenave, J.M. (2000). The role of secretory leukocyte proteinase inhibitor and elafin (elastase-specific inhibitor/skin-derived antileukoprotease) as alarm antiproteinases in inflammatory lung disease. Respir Res *1*, 87-92.

Sallenave, J.M. (2010). Secretory leukocyte protease inhibitor and elafin/trappin-2: versatile mucosal antimicrobials and regulators of immunity. Am J Respir Cell Mol Biol *42*, 635-643.

Sallenave, J.M., Marsden, M.D., and Ryle, A.P. (1992). Isolation of elafin and elastase-specific inhibitor (ESI) from bronchial secretions. Evidence of sequence homology and immunological cross-reactivity. Biol Chem Hoppe Seyler *373*, 27-33.

Sallenave, J.M., and Ryle, A.P. (1991). Purification and characterization of elastase-specific inhibitor. Sequence homology with mucus proteinase inhibitor. Biol Chem Hoppe Seyler *372*, 13-21.

Sallenave, J.M., Shulmann, J., Crossley, J., Jordana, M., and Gauldie, J. (1994). Regulation of secretory leukocyte proteinase inhibitor (SLPI) and elastase-specific inhibitor (ESI/elafin) in human airway epithelial cells by cytokines and neutrophilic enzymes. Am J Respir Cell Mol Biol *11*, 733-741.

Sallenave, J.M., Si Tahar, M., Cox, G., Chignard, M., and Gauldie, J. (1997). Secretory leukocyte proteinase inhibitor is a major leukocyte elastase inhibitor in human neutrophils. J Leukoc Biol *61*, 695-702.

Savelieva, E., Belair, C.D., Newton, M.A., DeVries, S., Gray, J.W., Waldman, F., and Reznikoff, C.A. (1997). 20q gain associates with immortalization: 20q13.2 amplification correlates with genome instability in human papillomavirus 16 E7 transformed human uroepithelial cells. Oncogene *14*, 551-560.

Schalkwijk, J., Wiedow, O., and Hirose, S. (1999). The trappin gene family: proteins defined by an N-terminal transglutaminase substrate domain and a C-terminal four-disulphide core. Biochem J *340 (Pt 3)*, 569-577.

Schaner, M.E., Ross, D.T., Ciaravino, G., Sorlie, T., Troyanskaya, O., Diehn, M., Wang, Y.C., Duran, G.E., Sikic, T.L., Caldeira, S.*, et al.* (2003). Gene expression patterns in ovarian carcinomas. Mol Biol Cell *14*, 4376-4386.

Schummer, M., Ng, W.V., Bumgarner, R.E., Nelson, P.S., Schummer, B., Bednarski, D.W., Hassell, L., Baldwin, R.L., Karlan, B.Y., and Hood, L. (1999). Comparative hybridization of an

array of 21,500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas. Gene *238*, 375-385.

Schäfer, B., von Horsten, H.H., Dacheux, J.L., Holtz, W., and Kirchhoff, C. (2003). Cloning and characterization of boar epididymal secretory proteins by homology to the human. Reprod Domest Anim *38*, 111-118.

Scotto, L., Narayan, G., Nandula, S.V., Arias-Pulido, H., Subramaniyam, S., Schneider, A., Kaufmann, A.M., Wright, J.D., Pothuri, B., Mansukhani, M.*, et al.* (2008). Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. Genes Chromosomes Cancer *47*, 755-765.

Shiels, H., Li, X., Schumacker, P.T., Maltepe, E., Padrid, P.A., Sperling, A., Thompson, C.B., and Lindsten, T. (2000). TRAF4 deficiency leads to tracheal malformation with resulting alterations in air flow to the lungs. Am J Pathol *157*, 679-688.

Shugars, D.C. (1999). Endogenous mucosal antiviral factors of the oral cavity. J Infect Dis *179 Suppl 3*, S431-435.

Si-Tahar, M., Merlin, D., Sitaraman, S., and Madara, J.L. (2000). Constitutive and regulated secretion of secretory leukocyte proteinase inhibitor by human intestinal epithelial cells. Gastroenterology *118*, 1061-1071.

Simpson, A.J., Maxwell, A.I., Govan, J.R., Haslett, C., and Sallenave, J.M. (1999). Elafin (elastase-specific inhibitor) has anti-microbial activity against gram-positive and gram-negative respiratory pathogens. FEBS Lett *452*, 309-313.

Sivadasan, P., Gupta, M.K., Sathe, G.J., Balakrishnan, L., Palit, P., Gowda, H., Suresh, A., Kuriakose, M.A., and Sirdeshmukh, R. (2015). Human salivary proteome--a resource of potential biomarkers for oral cancer. J Proteomics *127*, 89-95.

Smith, C.E., and Johnson, D.A. (1985). Human bronchial leucocyte proteinase inhibitor. Rapid isolation and kinetic analysis with human leucocyte proteinases. Biochem J *225*, 463-472.

Smith, D.J., and Taubman, M.A. (1996). Experimental immunization of rats with a Streptococcus mutans 59-kilodalton glucan-binding protein protects against dental caries. Infect Immun *64*, 3069-3073.

Sonoda, G., Palazzo, J., du Manoir, S., Godwin, A.K., Feder, M., Yakushiji, M., and Testa, J.R. (1997). Comparative genomic hybridization detects frequent overrepresentation of chromosomal material from 3q26, 8q24, and 20q13 in human ovarian carcinomas. Genes Chromosomes Cancer *20*, 320-328.

Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A.*, et al.* (2002). Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A *99*, 4465-4470.

Suzuki, Y., Furukawa, M., Abe, J., Kashiwagi, M., and Hirose, S. (2000). Localization of porcine trappin-2 (SKALP/elafin) in trachea and large intestine by in situ hybridization and immunohistochemistry. Histochem Cell Biol *114*, 15-20.

Szymaniak, A.D., Mahoney, J.E., Cardoso, W.V., and Varelas, X. (2015). Crumbs3-Mediated Polarity Directs Airway Epithelial Cell Fate through the Hippo Pathway Effector Yap. Dev Cell *34*, 283-296.

Tabach, Y., Kogan-Sakin, I., Buganim, Y., Solomon, H., Goldfinger, N., Hovland, R., Ke, X.S., Oyan, A.M., Kalland, K.H., Rotter, V.*, et al.* (2011). Amplification of the 20q chromosomal arm occurs early in tumorigenic transformation and may initiate cancer. PLoS One *6*, e14632.

Tanner, M.M., Grenman, S., Koul, A., Johannsson, O., Meltzer, P., Pejovic, T., Borg, A., and Isola, J.J. (2000). Frequent amplification of chromosomal region 20q12-q13 in ovarian cancer. Clin Cancer Res *6*, 1833-1839.

Tas, F., Aydiner, A., Topuz, E., Yasasever, V., Karadeniz, A., and Saip, P. (2000). Utility of the serum tumor markers: CYFRA 21.1, carcinoembryonic antigen (CEA), and squamous cell carcinoma antigen (SCC) in squamous cell lung cancer. J Exp Clin Cancer Res *19*, 477-481.

Thibault, B., Castells, M., Delord, J.P., and Couderc, B. (2014). Ovarian cancer microenvironment: implications for cancer dissemination and chemoresistance acquisition. Cancer Metastasis Rev *33*, 17-39.

Thompson, R.C., and Ohlsson, K. (1986). Isolation, properties, and complete amino acid sequence of human secretory leukocyte protease inhibitor, a potent inhibitor of leukocyte elastase. Proc Natl Acad Sci U S A *83*, 6692-6696.

Tomazic, P.V., Birner-Gruenberger, R., Leitner, A., Obrist, B., Spoerk, S., and Lang-Loidolt, D. (2014). Nasal mucus proteomic changes reflect altered immune responses and epithelial permeability in patients with allergic rhinitis. J Allergy Clin Immunol *133*, 741-750.

Tomee, J.F., Hiemstra, P.S., Heinzel-Wieland, R., and Kauffman, H.F. (1997). Antileukoprotease: an endogenous protein in the innate mucosal defense against fungi. J Infect Dis *176*, 740-747.

Tominaga, E., Tsuda, H., Arao, T., Nishimura, S., Takano, M., Kataoka, F., Nomura, H., Hirasawa, A., Aoki, D., and Nishio, K. (2010). Amplification of GNAS may be an independent, qualitative, and reproducible biomarker to predict progression-free survival in epithelial ovarian cancer. Gynecol Oncol *118*, 160-166.

Tsunemi, M., Matsuura, Y., Sakakibara, S., and Katsube, Y. (1996). Crystal structure of an elastase-specific inhibitor elafin complexed with porcine pancreatic elastase determined at 1.9 A resolution. Biochemistry *35*, 11570-11576.

Turcatel, G., Rubin, N., Menke, D.B., Martin, G., Shi, W., and Warburton, D. (2013). Lung mesenchymal expression of Sox9 plays a critical role in tracheal development. BMC Biol *11*, 117.

Uchida, K., Oga, A., Mano, T., Nagatsuka, H., Ueyama, Y., and Sasaki, K. (2010). Screening for DNA copy number aberrations in mucinous adenocarcinoma arising from the minor salivary gland: two case reports. Cancer Genet Cytogenet *203*, 324-327.

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F.*, et al.* (2017). A pathology atlas of the human cancer transcriptome. Science *357*.

Uhlenbruck, F., Sinowatz, F., Amselgruber, W., Kirchhoff, C., and Ivell, R. (1993). Tissue-specific gene expression as an indicator of epididymis-specific functional status in the boar, bull and stallion. Int J Androl *16*, 53-61.

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A.*, et al.* (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419.

Valsecchi, C.I.K., Basilicata, M.F., Semplicio, G., Georgiev, P., Gutierrez, N.M., and Akhtar, A. (2018). Facultative dosage compensation of developmental genes on autosomes in Drosophila and mouse embryonic stem cells. Nat Commun *9*, 3626.

Van Gorp, T., Cadron, I., Despierre, E., Daemen, A., Leunen, K., Amant, F., Timmerman, D., De Moor, B., and Vergote, I. (2011). HE4 and CA125 as a diagnostic test in ovarian cancer: prospective validation of the Risk of Ovarian Malignancy Algorithm. Br J Cancer *104*, 863-870.

van Wetering, S., van der Linden, A.C., van Sterkenburg, M.A., Rabe, K.F., Schalkwijk, J., and Hiemstra, P.S. (2000). Regulation of secretory leukocyte proteinase inhibitor (SLPI) production by human bronchial epithelial cells: increase of cell-associated SLPI by neutrophil elastase. J Investig Med *48*, 359-366.

Varki, A. (1993). Biological roles of oligosaccharides: all of the theories are correct. Glycobiology *3*, 97-130.

Voronov, E., Shouval, D.S., Krelin, Y., Cagnano, E., Benharroch, D., Iwakura, Y., Dinarello, C.A., and Apte, R.N. (2003). IL-1 is required for tumor invasiveness and angiogenesis. Proc Natl Acad Sci U S A *100*, 2645-2650.

Wahl, S.M., McNeely, T.B., Janoff, E.N., Shugars, D., Worley, P., Tucker, C., and Orenstein, J.M. (1997). Secretory leukocyte protease inhibitor (SLPI) in mucosal fluids inhibits HIV-I. Oral Dis *3 Suppl 1*, S64-69.

Walter, M., Plotnick, M., and Schechter, N.M. (1996). Inhibition of human mast cell chymase by secretory leukocyte proteinase inhibitor: enhancement of the interaction by heparin. Arch Biochem Biophys *327*, 81-88.

Wan, J., Wang, Y., Cai, G., Liang, J., Yue, C., Wang, F., Song, J., Wang, J., Liu, M., Luo, J.*, et al.* (2016). Elevated serum concentrations of HE4 as a novel biomarker of disease severity and renal fibrosis in kidney disease. Oncotarget *7*, 67748-67759.

Wang, H., Zhu, L., Gao, J., Hu, Z., and Lin, B. (2015). Promotive role of recombinant HE4 protein in proliferation and carboplatin resistance in ovarian cancer cells. Oncol Rep *33*, 403-412.

Wang, X., Fan, Y., Wang, J., Wang, H., and Liu, W. (2014). Evaluating the expression and diagnostic value of human epididymis protein 4 (HE4) in small cell lung cancer. Tumour Biol *35*, 6847-6853.

Wang, Y., Han, C., Teng, F., Bai, Z., Tian, W., and Xue, F. (2017). Predictive value of serum HE4 and CA125 concentrations for lymphatic metastasis of endometrial cancer. Int J Gynaecol Obstet *136*, 58-63.

Wang, Z., Widgren, E.E., Richardson, R.T., and O'Rand, M.G. (2007). Characterization of an eppin protein complex from human semen and spermatozoa. Biol Reprod *77*, 476-484.

Watt, A.P., Sharp, J.A., Lefevre, C., and Nicholas, K.R. (2012). WFDC2 is differentially expressed in the mammary gland of the tammar wallaby and provides immune protection to the mammary gland and the developing pouch young. Dev Comp Immunol *36*, 584-590.

Weis, V.G., and Goldenring, J.R. (2009). Current understanding of SPEM and its standing in the preneoplastic process. Gastric Cancer *12*, 189-197.

Weis, V.G., Petersen, C.P., Mills, J.C., Tuma, P.L., Whitehead, R.H., and Goldenring, J.R. (2014). Establishment of novel in vitro mouse chief cell and SPEM cultures identifies MAL2 as a marker of metaplasia in the stomach. Am J Physiol Gastrointest Liver Physiol *307*, G777-792.

Welsh, J.B., Zarrinkar, P.P., Sapinoso, L.M., Kern, S.G., Behling, C.A., Monk, B.J., Lockhart, D.J., Burger, R.A., and Hampton, G.M. (2001). Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Proc Natl Acad Sci U S A *98*, 1176-1181.

Westin, U., Fryksmark, U., Polling, A., and Ohlsson, K. (1994). Localisation of secretory leucocyte proteinase inhibitor mRNA in nasal mucosa. Acta Otolaryngol *114*, 199-202.

Westin, U., Polling, A., Ljungkrantz, I., and Ohlsson, K. (1999). Identification of SLPI (secretory leukocyte protease inhibitor) in human mast cells using immunohistochemistry and in situ hybridisation. Biol Chem *380*, 489-493.

Whiteman, E.L., Fan, S., Harder, J.L., Walton, K.D., Liu, C.J., Soofi, A., Fogg, V.C., Hershenson, M.B., Dressler, G.R., Deutsch, G.H.*, et al.* (2014). Crumbs3 is essential for proper epithelial development and viability. Mol Cell Biol *34*, 43-56.

Wiedow, O., Harder, J., Bartels, J., Streit, V., and Christophers, E. (1998). Antileukoprotease in human skin: an antibiotic peptide constitutively produced by keratinocytes. Biochem Biophys Res Commun *248*, 904-909.

Wiedow, O., Lüademann, J., and Utecht, B. (1991). Elafin is a potent inhibitor of proteinase 3. Biochem Biophys Res Commun *174*, 6-10.

Wiesner, J., and Vilcinskas, A. (2010). Antimicrobial peptides: the ancient arm of the human immune system. Virulence *1*, 440-464.

Wilding, J.L., and Bodmer, W.F. (2014). Cancer cell lines for drug discovery and development. Cancer Res *74*, 2377-2384.

Wilkinson, T.S., Dhaliwal, K., Hamilton, T.W., Lipka, A.F., Farrell, L., Davidson, D.J., Duffin, R., Morris, A.C., Haslett, C., Govan, J.R.*, et al.* (2009). Trappin-2 promotes early clearance of Pseudomonas aeruginosa through CD14-dependent macrophage activation and neutrophil recruitment. Am J Pathol *174*, 1338-1346.

Wodarz, A., Hinz, U., Engelbert, M., and Knust, E. (1995). Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of Drosophila. Cell *82*, 67-76.

Wu, J., Kobayashi, M., Sousa, E.A., Liu, W., Cai, J., Goldman, S.J., Dorner, A.J., Projan, S.J., Kavuru, M.S., Qiu, Y.*, et al.* (2005). Differential proteomic analysis of bronchoalveolar lavage fluid in asthmatics following segmental antigen challenge. Mol Cell Proteomics *4*, 1251-1264.

Wyszyńska-Pawelec, G., Koryczan, P., Zapała, J., Gontarz, M., Opach, M., Kuśnierz, P., Kosowski, B., and Adamek, D. (2013). Minor salivary gland mucinous adenocarcinoma of buccal mucosa - case report and review of the literature. Pol J Pathol *64*, 312-316.

Xu, W.D., Wang, L.F., Miao, S.Y., Zhao, M., Fan, H.Y., Zong, S.D., Wu, Y.W., Shi, X.Q., and Koide, S.S. (1996). Identification of a rabbit epididymal protein gene. Arch Androl *37*, 135-141.

Yamashita, S., Tokuishi, K., Moroga, T., Yamamoto, S., Ohbo, K., Miyahara, S., Yoshida, Y., Yanagisawa, J., Hamatake, D., Hiratsuka, M.*, et al.* (2012). Serum level of HE4 is closely associated with pulmonary adenocarcinoma progression. Tumour Biol *33*, 2365-2370.

Yan, W., Apweiler, R., Balgley, B.M., Boontheung, P., Bundy, J.L., Cargile, B.J., Cole, S., Fang, X., Gonzalez-Begne, M., Griffin, T.J.*, et al.* (2009). Systematic comparison of the human saliva and plasma proteomes. Proteomics Clin Appl *3*, 116-134.

Yang, J., Mani, S.A., Donaher, J.L., Ramaswamy, S., Itzykson, R.A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R.A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell *117*, 927-939.

Yenugu, S., Richardson, R.T., Sivashanmugam, P., Wang, Z., O'rand, M.G., French, F.S., and Hall, S.H. (2004). Antimicrobial activity of human EPPIN, an androgen-regulated, sperm-bound protein with a whey acidic protein motif. Biol Reprod *71*, 1484-1490.

Yılmaz, S.A., Altınkaya, S., Kerimoglu, Ö., Tazegül Pekin, A., Akyürek, F., Ilhan, T.T., Benzer, N., Unlu, A., Yuksel, H., and Celik, C. (2017). The role of human epididymis secretory protein

E4 in patients with endometrial cancer and premalignant endometrial lesions. J Obstet Gynaecol *37*, 58-63.

Yin, B.W., and Lloyd, K.O. (2001). Molecular cloning of the CA125 ovarian cancer antigen: identification as a new mucin, MUC16. J Biol Chem *276*, 27371-27375.

Ying, Q.L., Kemme, M., and Simon, S.R. (1994). Functions of the N-terminal domain of secretory leukoprotease inhibitor. Biochemistry *33*, 5445-5450.

Ying, Q.L., and Simon, S.R. (1993). Kinetics of the inhibition of human leukocyte elastase by elafin, a 6-kilodalton elastase-specific inhibitor from human skin. Biochemistry *32*, 1866-1874.

Yuan, T., and Li, Y. (2017). Human Epididymis Protein 4 as a Potential Biomarker of Chronic Kidney Disease in Female Patients With Normal Ovarian Function. Lab Med *48*, 238-243.

Zanotti, L., Bignotti, E., Calza, S., Bandiera, E., Ruggeri, G., Galli, C., Tognon, G., Ragnoli, M., Romani, C., Tassi, R.A.*, et al.* (2012). Human epididymis protein 4 as a serum marker for diagnosis of endometrial carcinoma and prediction of clinical outcome. Clin Chem Lab Med *50*, 2189-2198.

Zeeuwen, P.L., Hendriks, W., de Jong, W.W., and Schalkwijk, J. (1997). Identification and sequence analysis of two new members of the SKALP/elafin and SPAI-2 gene family. Biochemical properties of the transglutaminase substrate motif and suggestions for a new nomenclature. J Biol Chem *272*, 20471-20478.

Zhang, L., Liu, D., Li, L., Pu, D., Zhou, P., Jing, Y., Yu, H., Wang, Y., Zhu, Y., He, Y.*, et al.* (2017). The important role of circulating CYFRA21-1 in metastasis diagnosis and prognostic value compared with carcinoembryonic antigen and neuron-specific enolase in lung cancer patients. BMC Cancer *17*, 96.

Zhang, M., Yuan, L., Yao, F., Cao, P., Rong, J., Zhang, B., and Su, J. (2018). Human epididymis protein 4 concentration is not associated with liver fibrosis and cirrhosis in a case control study. Clin Chim Acta *484*, 213-217.

Zhang, Y., Qiao, C., Li, L., Zhao, X., and Li, Y. (2014). Serum HE4 is more suitable as a biomarker than CA125 in Chinese women with benign gynecologic disorders. Afr Health Sci *14*, 913-918.

Zhou, L., Zhao, S.Z., Koh, S.K., Chen, L., Vaz, C., Tanavde, V., Li, X.R., and Beuerman, R.W. (2012). In-depth analysis of the human tear proteome. J Proteomics *75*, 3877-3885.

Zhu, H., Lam, D.C., Han, K.C., Tin, V.P., Suen, W.S., Wang, E., Lam, W.K., Cai, W.W., Chung, L.P., and Wong, M.P. (2007). High resolution analysis of genomic aberrations by metaphase and array comparative genomic hybridization identifies candidate tumour genes in lung cancer cell lines. Cancer Lett *245*, 303-314.

Zhu, L., Zhuang, H., Wang, H., Tan, M., Schwab, C.L., Deng, L., Gao, J., Hao, Y., Li, X., Gao, S.*, et al.* (2016). Overexpression of HE4 (human epididymis protein 4) enhances proliferation, invasion and metastasis of ovarian cancer. Oncotarget *7*, 729-744.

Ziv, Y., Bielopolski, D., Galanty, Y., Lukas, C., Taya, Y., Schultz, D.C., Lukas, J., Bekker-Jensen, S., Bartek, J., and Shiloh, Y. (2006). Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. Nat Cell Biol *8*, 870-876.

Appendices

Supplementary methods and reagents

DNA agarose gel electrophoresis

For a 1% gel, 1 g of agarose was added to 100 ml 1X Tris-acetate-EDTA (TAE), heated and supplemented with 0.2 µg/ml ethidium bromide. Colourless DNA samples were mixed with 1 µl of loading dye prior to loading. PCR Ranger 100 bp DNA ladder (Norgen Biotek) and GeneRuler 1 kb DNA ladder (ThermoFisher Scientific) were used as reference ladders. TAE buffer was made as 50X stock solution (242 g of Tris base, 57.1 ml of glacial acetic acid, 100 ml of 0.5 M pH 8.0 EDTA, made to 1 L with H_2O). The stock solution was diluted in water to give a 1X TAE working solution.

Polyacrylamide gel electrophoresis and Western blotting

Lower Tris buffer for SDS-PAGE

Lower Tris buffer was made with 18.17 g of Tris (ThermoFisher Scientific), 0.4 g of SDS pellets (ThermoFisher Scientific) and made to 100 ml with H_2O. The pH of the solution was adjusted to pH 8.8.

Upper Tris buffer for SDS-PAGE

Upper Tris buffer was made with 6.06 g of Tris (ThermoFisher Scientific), 0.4 g of SDS (ThermoFisher Scientific) and made up to 100 ml with H_2O. The pH of the solution was adjusted to pH 6.8.

SDS-PAGE 12% acrylamide gel

To make the resolving gel for a 1.5 mm 12% gel, 3 ml of 40% acrylamide (ThermoFisher Scientific) and 2.5 ml of Lower Tris were added to 4.3 ml of H_2O. Prior to pouring between the

glass plates, 350 µl of 10% ammonium persulfate (APS; Acros Organics) and 5 µl of TEMED (AppliChem) were added to the gel solution to induce crosslinking. The gel was protected from drying out by layering isopropanol on top of the resolving gel which was removed prior to addition of the stacking gel.

For the stacking gel, 2.1 ml of Upper Tris and 975 µl of 40% acrylamide were added to 4.725 ml of H_2O. Prior to pouring, 100 µl of 10% APS and 17 µl of TEMED were added to the solution. A comb was added between the plates to form lanes for protein loading.

2X SDS loading buffer

Samples for analysis by Western blotting were lysed in 2X SDS lysis buffer, made by adding 1 protease inhibitor tablet (Roche) to 1 ml of 1 M DTT (ThermoFisher Scientific), 1 ml of 20% SDS (ThermoFisher Scientific), 2 ml of glycerol (ThermoFisher Scientific), 1.25 ml of 0.5 M pH 6.8 Tris-HCl (ThermoFisher Scientific), 0.2 ml of 0.02% bromophenol blue (Sigma-Aldrich) and 4.55 ml of H_2O.

SDS-PAGE running buffer

SDS-PAGE running buffer was made as a 10X stock solution using 10 g of SDS pellets (ThermoFisher Scientific), 188 g of glycine (ThermoFisher Scientific) and 30.2 g of Tris base (ThermoFisher Scientific) made up to 1 L with H_2O. This stock solution was diluted in water to give a 1X working solution.

Semi-dry transfer buffer

Semi-dry transfer buffer was made using 5.8 g of Tris base, 2.9 g of glycine, 0.37 g of SDS and 200 ml of methanol, made up to 1 L with H_2O.

Lower Tris buffer for native PAGE

Lower Tris buffer was made by adding 54.5 g of Tris base (ThermoFisher Scientific) to 300 ml of H_2O, with the solution adjusted to pH 8.8.

Upper Tris buffer for native PAGE

Upper Tris buffer was made by adding 6 g of Tris (ThermoFisher Scientific) to 100 ml of H_2O with the solution adjusted to pH 6.8.

Native PAGE 10% acrylamide gel

To make the resolving gel for a 1.5 mm 10% native gel, 3.33 ml of 40% acrylamide (ThermoFisher Scientific) and 2.5 ml of Lower Tris were added to 4.17 ml of H_2O. Prior to pouring between the plates, 50 µl of 10% APS (Acros Organics) and 5 µl of TEMED (AppliChem) were added to the gel solution to cause crosslinking. To make the stacking gel, 2.5 ml of Upper Tris and 3.33 µl of 40% acrylamide were added to 4.17 ml of H_2O. Prior to pouring, 50 µl of 10% APS and 10 µl of TEMED were added to the solution and a comb was added between the plates to form lanes for protein loading.

Native PAGE sample buffer

Native PAGE sample buffer was made using 4.8 ml of H_2O, 1 ml of 0.5 M pH 6.8 Tris-HCl (ThermoFisher Scientific), 2 ml of glycerol (ThermoFisher Scientific) and 0.2 ml of 0.5% (w/v) bromophenol blue (Sigma-Aldrich).

Native PAGE running buffer

Native PAGE running buffer was made with 30.3 g of Tris base (ThermoFisher Scientific) and 144 g of glycine (ThermoFisher Scientific), made up to 1 L with H_2O. The pH was adjusted to pH 8.3.

10X TBS

A 10X Tris buffered saline (TBS) solution was made with 80 g of NaCl (ThermoFisher Scientific), 2 g of KCl (BDH Chemicals) and 30 g of Tris base (ThermoFisher Scientific), made up to 1 L with H_2O. The pH was adjusted to pH 7.6. This stock solution was diluted to a 1X working solution.

1X TBS-Tween

TBS-Tween wash buffer was made by adding 1 ml of Tween-20 (VWR Chemicals) to 1 L of 1X TBS.

Zymography

5X non-reducing sample buffer

Non-reducing sample buffer was made with 10 g of SDS, 50 ml of glycerol, 25 mg of bromophenol blue and 4.91 g of Tris, made up to 250 ml with H_2O.

7.5% acrylamide gel with gelatin

To make the resolving gel for a 1.0 mm 7.5% gel, 1.5 ml of 40% acrylamide (ThermoFisher Scientific), 2 ml of 1.5 M Tris (pH 8.8), 40 µl of 20% SDS and 2 ml of gelatin (4 mg/ml) were added to 2.5 ml of H_2O. Prior to pouring between the glass plates, 80 µl of 10% APS (Acros Organics) and 10 µl of TEMED (AppliChem) were added to the gel solution to induce crosslinking. The gel was protected from drying out by layering isopropanol on top of the resolving gel which was removed prior to addition of the stacking gel.

For the stacking gel, 1.25 ml of 0.5 M Tris (pH 6.8), 0.5 ml of 40% acrylamide and 25 µl of 20% SDS were added to 3.245 ml of H_2O. Prior to pouring, 50 µl of 10% APS and 10 µl of TEMED were added to the solution and a comb was added between the plates to form lanes for protein loading.

Zymography washing buffer

Washing buffer was made with 6.25 ml of Triton-X 100, 12.5 ml of 1 M Tris HCl (pH 7.5), 625 µl of 2 M $CaCl_2$ and 2.5 µl of 0.1 M $ZnCl_2$, made up to 250 ml with H_2O.

Zymography incubation buffer

Incubation buffer was made with 2.5 ml of Triton-X 100, 12.5 ml of 1 M Tris HCl (pH 7.5), 625 µl of 2 M $CaCl_2$ and 2.5 µl of 0.1 M $ZnCl_2$, made up to 250 ml with H_2O.

Additional results

Glycosylation site prediction

Online NetNGlyc 1.0 and NetOGlyc 4.0 Server prediction results for human and mouse WFDC2 are shown in the figures below. The positions of predicted N- or O-glycans are indicated.
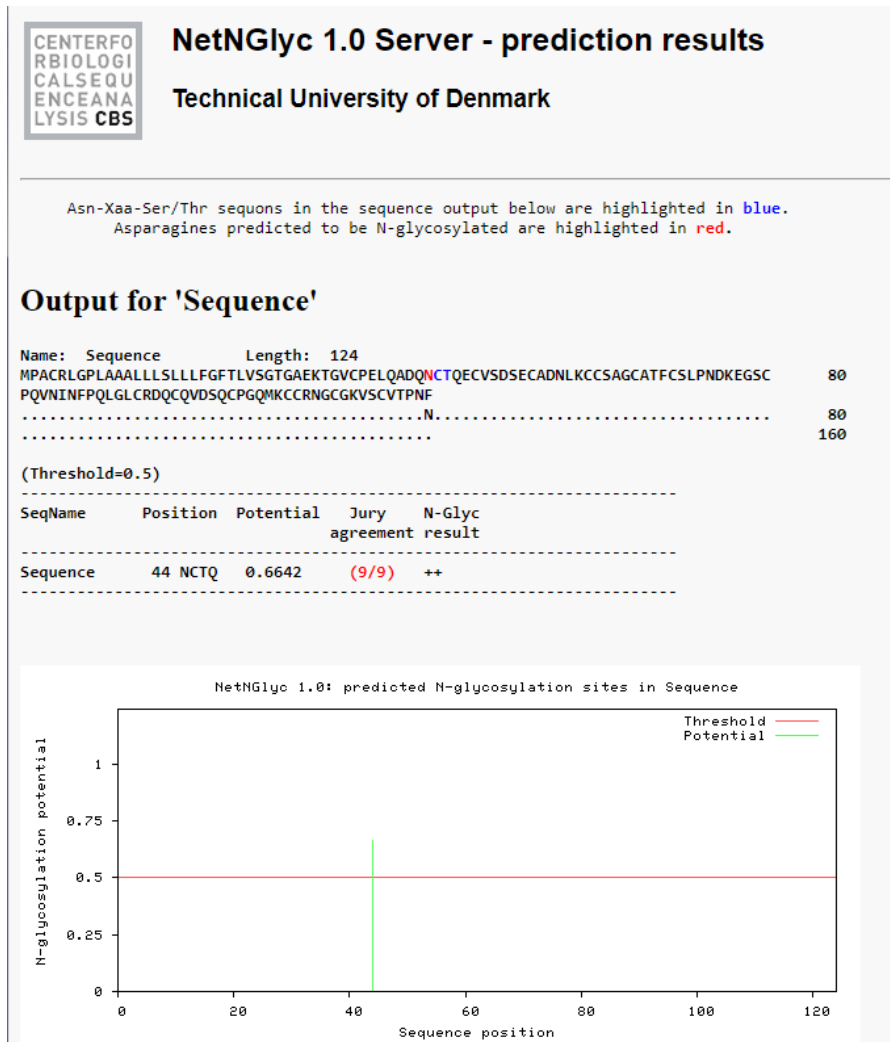
NetNGlyc 1.0 Server prediction results for human WFDC2.

NetOGlyc 4.0 Server - prediction results

Technical University of Denmark

```
##gff-version 2
##source-version NetOGlyc 4.0.0.13
##date 18-7-30
##Type Protein
#seqname     source  feature start   end      score   strand  frame   comment
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  16      16     0.0144401        .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  23      23     0.0728325        .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  26      26     0.272241         .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  28      28     0.1344  .        .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  33      33     0.548403         .       .       #POSITIVE
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  46      46     0.430131         .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  51      51     0.474784         .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  53      53     0.239209         .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  63      63     0.341761         .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  68      68     0.133138         .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  71      71     0.410181         .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  79      79     0.290445         .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  101     101    0.308183         .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  118     118    0.296661         .       .
SEQUENCE     netOGlyc-4.0.0.13       CARBOHYD  121     121    0.144912         .       .
```

NetOGlyc 4.0 Server prediction results for human WFDC2.

NetNGlyc 1.0 Server - prediction results
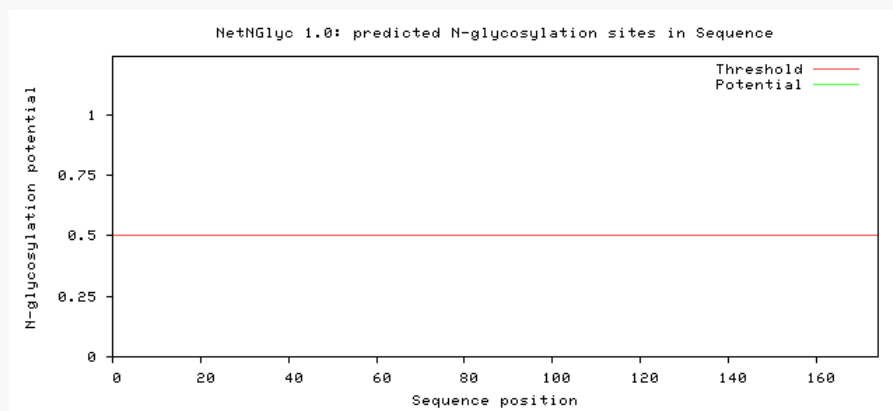
Technical University of Denmark

```
    Asn-Xaa-Ser/Thr sequons in the sequence output below are highlighted in blue.
        Asparagines predicted to be N-glycosylated are highlighted in red.
```

Output for 'Sequence'

```
Name:  Sequence          Length:  174
MPACRLCLLAAGLLLGLLLFTPISATGTDAEKPGECPQLEPITDCVLECTLDKDCADNRKCCQAGCSSVCSKPNGPSEGE          80
LSGTDTKLSETGTTTQSAGLDHTTKPPGGQVSTKPPAVTREGLGVREKQGTCPSVDIPKLGLCEDQCQVDSQCSGNMKCC          160
RNGCGKMACTTPKF
.................................................................................          80
.................................................................................          160
..............                                                                             240

(Threshold=0.5)

No sites predicted in this sequence.
```


NetNGlyc 1.0: predicted N-glycosylation sites in Sequence

NetNGlyc 1.0 Server prediction results for murine Wfdc2.

NetOGlyc 4.0 Server - prediction results

Technical University of Denmark

```
##gff-version 2
##source-version NetOGlyc 4.0.0.13
##date 18-7-31
##Type Protein
#seqname      source  feature start   end     score   strand  frame   comment
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         21     21      0.0430753       .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         24     24      0.105576        .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         26     26      0.111052        .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         28     28      0.148216        .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         43     43      0.328966        .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         50     50      0.213537        .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         67     67      0.325094        .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         68     68      0.27453 .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         71     71      0.339912        .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         77     77      0.769041        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         82     82      0.646716        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         84     84      0.663157        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         86     86      0.685465        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         89     89      0.929123        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         91     91      0.897975        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         93     93      0.945194        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         94     94      0.973035        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         95     95      0.980142        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         97     97      0.967526        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         103    103     0.958659        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         104    104     0.937756        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         112    112     0.981708        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         113    113     0.92998 .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         119    119     0.882016        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         131    131     0.668276        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         134    134     0.884647        .       .       #POSITIVE
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         151    151     0.164441        .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         154    154     0.208249        .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         170    170     0.163734        .       .
SEQUENCE      netOGlyc-4.0.0.13       CARBOHYD         171    171     0.0901897       .       .
```

NetOGlyc 4.0 Server prediction results for murine Wfdc2.

Bacterial count at OD$_{600}$ of 1.0

The number of bacterial cells per cm^3 in cell suspensions at an OD$_{600}$ of 1.0 was counted using a Helber bacterial 1 cell counter (Thoma). The results shown in the table below represent the mean ± SD.

Bacterial strain	Mean number of cells per cm^3 ± SD
E. coli	$1.95 \times 10^9 \pm 7.50 \times 10^7$
S. gordonii	$3.70 \times 10^9 \pm 2.75 \times 10^8$
S. mutans	$6.52 \times 10^9 \pm 6.38 \times 10^8$
S. pneumoniae	$2.70 \times 10^9 \pm 5.00 \times 10^7$

Bacterial count at OD$_{600}$ of 1.0. Mean ± SD cells per cm^3 at OD$_{600}$ of 1.0 are summarised for each species (N=3).

Optimisation of agglutination assay

To determine the appropriate number of cells to be used in the agglutination assay, different volumes of bacterial suspension (at an OD_{600} of 1.0) were pelleted, washed in PBS and then labelled with FITC. Excess FITC was removed by repeated washes and then bacteria were resuspended in 1 ml of PBS which acted as a non-agglutinating negative control. One hundred microlitres of suspension was added to each well and incubated overnight in a U-bottomed 96-well plate. When analysed the following day via a UV box imaging system, wells that had visible 'pellets' of bacteria were identified as having an appropriate number of cells seeded for downstream experiments.

The results showed that the following volumes of bacterial suspension should be used for downstream agglutination assays: 0.5 ml of *E. coli*, 0.9 ml of *S. gordonii,* 0.9 ml of *S. mutans* and 0.7 ml of *S. pneumoniae.*
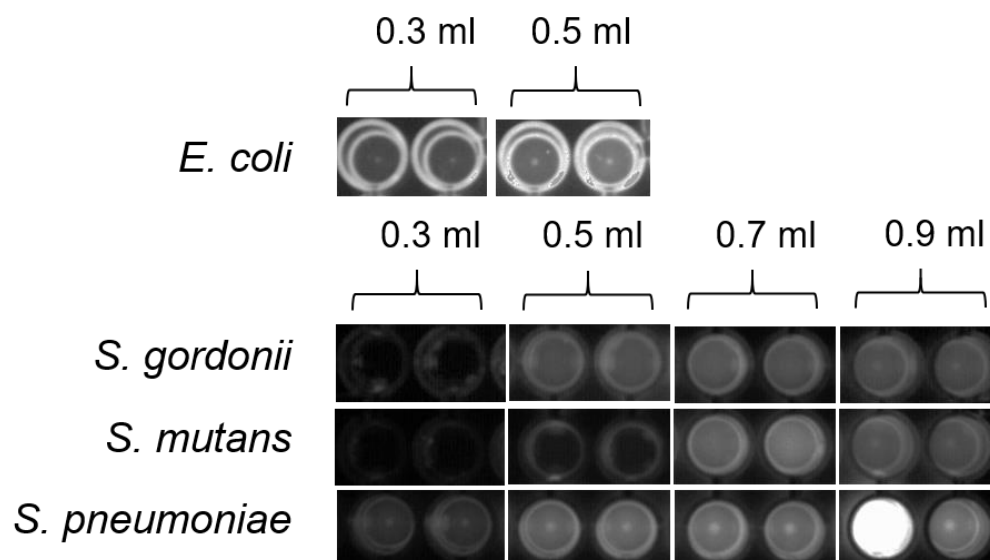
Optimisation of the volume of bacterial suspension required for agglutination assays. Different volumes of bacterial suspension (at OD_{600} of 1.0) were tested to determine the optimal volume required to give downstream visible 'pellets' of bacteria in wells where no agglutination had occurred. Each tested volume of cells was pelleted, resuspended in 1 ml of PBS and stained with FITC. The FITC-stained bacteria were resuspended in 1 ml of PBS and 100 µl were loaded into wells of a U-bottomed 96-well plate (n=2).

**Expression of WFDC2 in salivary gland organoids**

Organoids were developed by Zulaiha Rahman (School of Clinical Dentistry, University of Sheffield) from biopsy samples of normal human sublingual gland tissue. Cells were cultured in a basement membrane matrix (i.e. Matrigel) in an enriched media at an air-liquid interface (ALI). Signalling pathway inhibitors were added to the enriched media to study their effect on cell differentiation within the organoid. A summary of the inhibitors used and their functions are described in the table below.

Inhibitor	Function	Reference
SR7826	LIM kinase inhibitor, inhibits cofilin phosphorylation, inhibits cell migration and invasion	(Ray et al., 2014)
A8301	Transforming growth factor beta (TGFβ) type I receptor superfamily activin-like kinase (ALK5) inhibitor, inhibits epithelial to mesenchymal transition by TGFβ signaling	(Gurung et al., 2015)
DMH1	Bone morphogenic protein (BMP) inhibitor	(Ao et al., 2012)

A summary of the inhibitors used for differentiation studies of salivary organoids.

WFDC2 expression in salivary gland organoids was analysed using RT-qPCR. RNA derived from organoids, kindly donated by Zulaiha Rahman, was used to synthesise cDNA as described in Section 2.2.2 and the cDNA was used for qPCR analysis as described in Section 2.2.5.

HuSL cells have very low endogenous WFDC2 expression (see Fig 4.13) which is potentially due to the lack of highly differentiated cells such as ductular cells and mucous-secreting cells. The HuSL cells were exposed to retinoic acid (RA; 50 nM) for 24 hours and the response in WFDC2 expression was measured via RT-qPCR. RA is a factor important for differentiation of mucous cells (Lechner et al., 1982) so it was hypothesised that exposure to RA would induce an upregulation in WFDC2 expression. The results show that a moderate upregulation in WFDC2 expression did occur in the HuSL cells but this change was not statistically significant due to the variation between biological repeats.
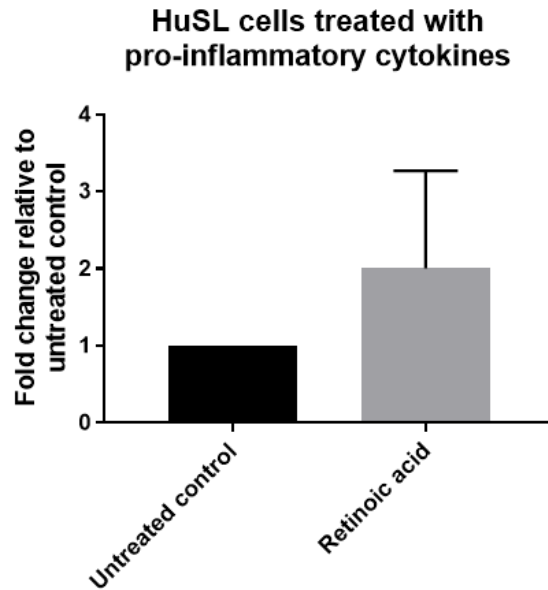
**HuSL cells treated with pro-inflammatory cytokines**

RT-qPCR analysis of WFDC2 expression in HuSL cells following exposure to retinoic acid. WFDC2 expression in human primary sublingual gland (HuSL) cells was tested by RT-qPCR following exposure to retinoic acid. WFDC2 expression is represented as the fold change in expression compared to the untreated HuSL cell controls. Data represents the mean ± SD (N=3; n=3).

HuSL cells can be grown in complex three-dimensional culture systems using extracellular matrices such as Matrigel to form salivary gland organoids. The organoids can develop branches, positively express markers for acinar cells and secrete active amylase, and thus provide a novel *in vitro* system for studying salivary glands and salivary gland diseases.

To determine whether organoids grown from HuSL cells expressed more WFDC2 than the cells grown at a monolayer, RT-qPCR was used to quantify gene expression from these tissues.

The RT-qPCR results show that although WFDC2 expression can be upregulated up to ~150 fold in organoids compared to monolayer cultures, there is a huge amount of variation in WFDC2 expression in the organoids, meaning that the difference between monolayer and organoid culture systems is not significant. The variation of WFDC2 expression in organoids may be due to the variation in formation of structures such as ducts, where WFDC2 would be most abundant.

WFDC2 expression in HuSL cells in different culture conditions
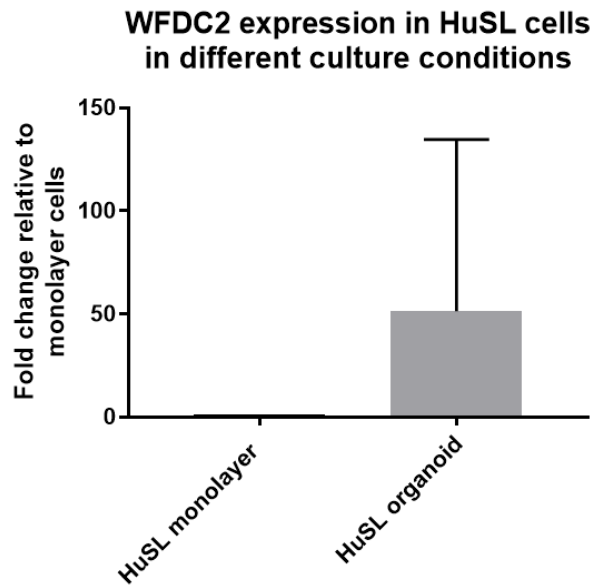
RT-qPCR analysis of WFDC2 expression in HuSL monolayers and organoids. HuSL cells grown as a monolayer and HuSL cells grown in complex 3D culture to form an organoid were analysed by RT-qPCR to quantify WFDC2 expression, which is represented as the fold change in expression compared to the monolayer cells. Data represents the mean ± SD (N=3; n=3).

During development of the organoid culture system, various signalling pathway inhibitors were included in the culture medium in order to prevent spontaneous differentiation of progenitor cells and control the differentiation process. Inhibitors were also used to influence the morphology of the organoids by limiting the progenitor population of cells' capacity for invasion and migration.

To establish whether WFDC2 expression within the HuSL organoids was affected when the tissues were exposed to different inhibitors, WFDC2 expression in the organoids was quantified using RT-qPCR.

The results show that there was no significant difference in WFDC2 expression between organoids that were exposed to signalling pathway inhibitors compared to untreated HuSL cells cultured as a monolayer.

Organoids treated with SR7826, a LIM-kinase inhibitor known to supress cell migration and invasion, showed extensive upregulation in WFDC2 expression in some organoids, however

others exhibited a less drastic upregulation. This broad variation in expression means that the changes in WFDC2 expression were not significant between SR7826-treated organoids compared with untreated organoids or monolayer HuSL cells. The effect of SR7826 on WFDC2 expression is interesting as it implicates the gene in processes such as migration and invasion: if LIM-kinase is inhibited, cytoskeletal factors that mediate invasion/migration will be inhibited, but it is possible that other factors that are independent of the LIM-kinase pathway could be upregulated with the aim of restoring the inhibited function. WFDC2 has been implicated in promoting invasion in cancer cells (see Section 5.8.5).
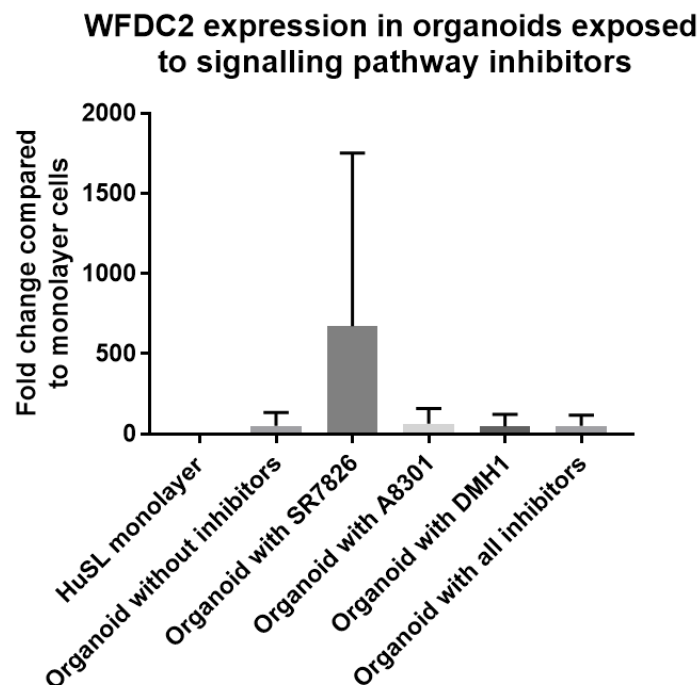
RT-qPCR analysis of WFDC2 expression in organoids treated with signalling pathway inhibitors. Organoids exposed to different signalling pathway inhibitors were analysed by RT-qPCR to quantify WFDC2 expression, which is represented as the fold change in expression compared to untreated HuSL cells grown as a monolayer. Data represents the mean ± SD (N=3; n=3).
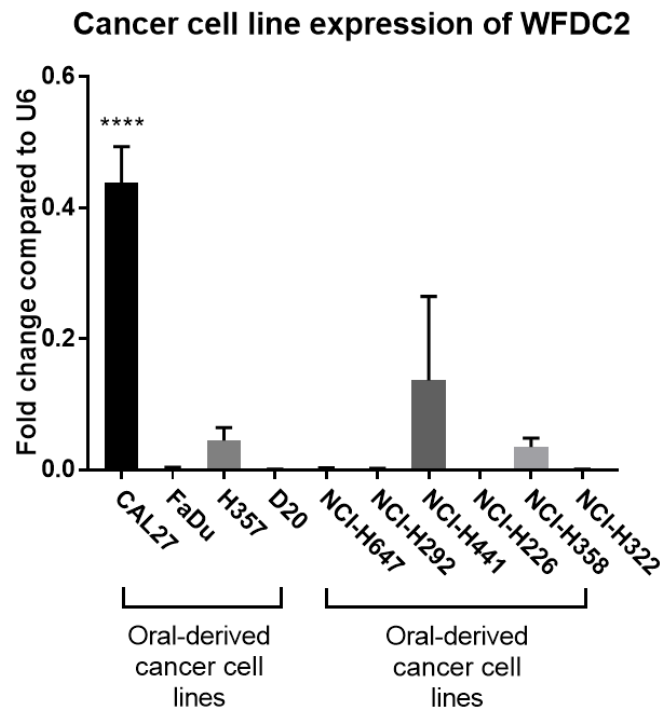
RT-qPCR analysis of WFDC2 in cancer cell lines
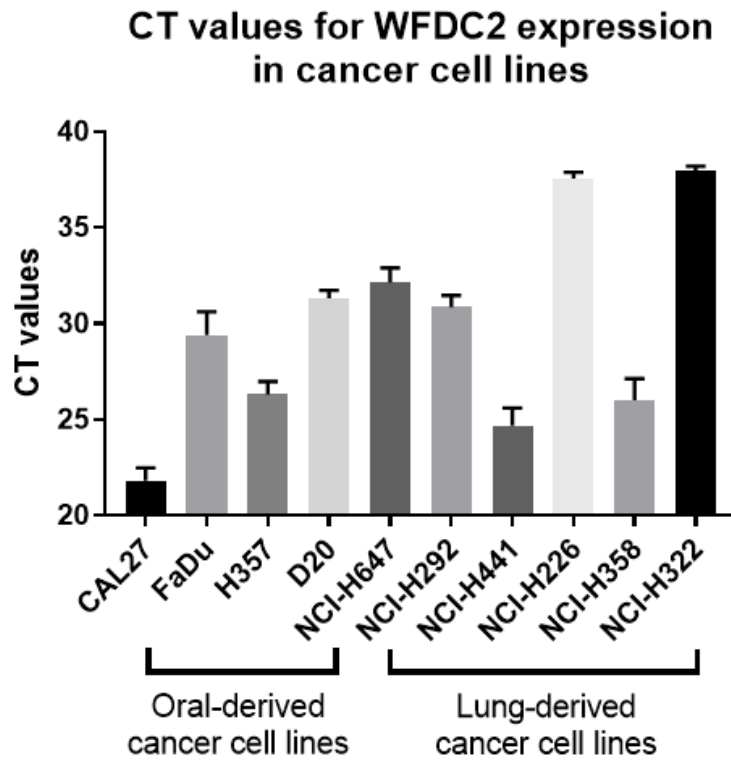
CT values from RT-qPCR quantification of WFDC2 gene expression. WFDC2 V1 expression in 4 oral-derived and 6 lung-derived cancer cell lines was tested using RT-qPCR. The CT values for WFDC2 expression are shown. Data represents the mean ± SD (N=3; n=3).

RT-qPCR analysis of V0 expression in cancer cell lines

WFDC2 variant 0 expression was measured in ten cancer cell lines by RT-qPCR using Sybr Green and the same primers used for RT-PCR (see Table 2.2). U6 was used as the endogenous control (F primer: 5' CTCGCTTCGGCAGCACA and R primer: AACGTTCACGAATTTGCGT) . The results show equivalent levels of expression between cell lines to those derived when analysing variant 1 expression using Taqman primers against B2M endogenous control. Taqman primers were used for all RT-qPCR experiments shown in results chapters as Taqman primers are considered to be more specific with results being reproducible and more accurately quantified.

RT-qPCR quantification of WFDC2 gene expression in 10 cancer cell lines. WFDC2 expression in 4 oral-derived and 6 lung-derived cancer cell lines was tested using RT-qPCR. WFDC2 expression is represented as the fold changed in expression compared to the U6 endogenous control. Data represents the mean ± SD (N=3; n=3) **** $P \leq 0.0001$.

CT values for B2M expression in cancer cell lines treated with cytokines
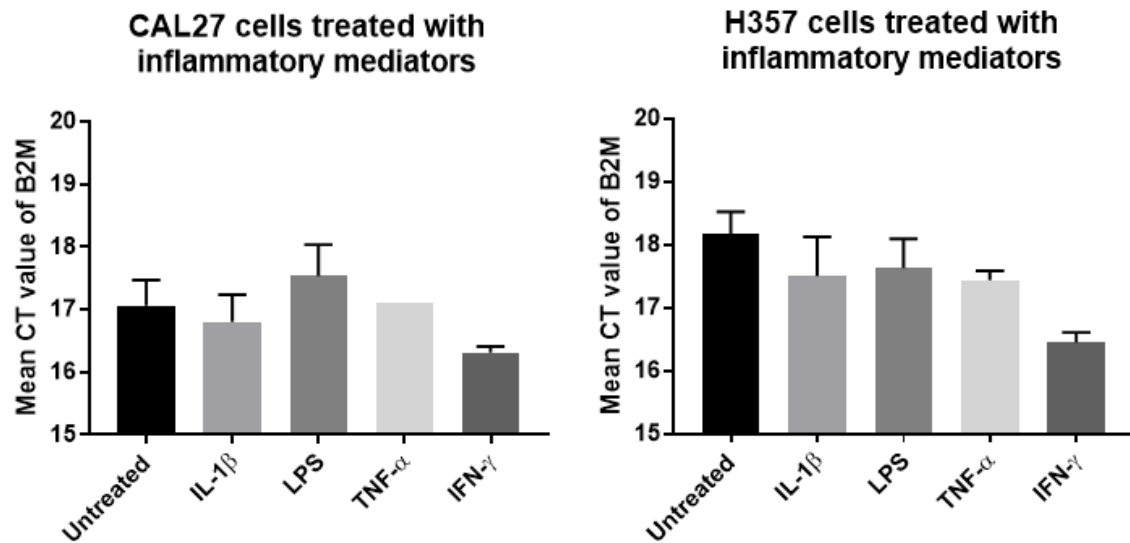
Fig 5.7. CT values from RT-qPCR quantification of WFDC2 gene expression in cells exposed to cytokines. WFDC2 expression in CAL27 and H357 cells was tested by RT-qPCR following exposure to inflammatory mediators: LPS (50 µg/ml), TNF-α (20 ng/ml), IL-1ß (20 ng/ml) and IFN-γ (5 ng/ml). The CT values of the B2M endogenous control are shown. The results suggest that the reduction in WFDC2 expression identified following exposure to IFN-γ did not result from a reduction in cell viability, which would be indicated by an increase in the CT value of the B2M endogenous control. Results show the mean ± SD (N=3; n=3).

Clonal isolation analysis of CAL27 cells

To determine whether CAL27 cells are capable of being used for clonal isolation i.e. are not attachment dependent, the cells were serially diluted within a 96-well plate and grown for 7 days. The cells were fixed and stained which enabled easy identification of single cell colonies. The results showed that CAL27 cells were capable of growing without attachment to neighbouring cells and thus could be used for clonal selection of CRISPR gene edited cells.
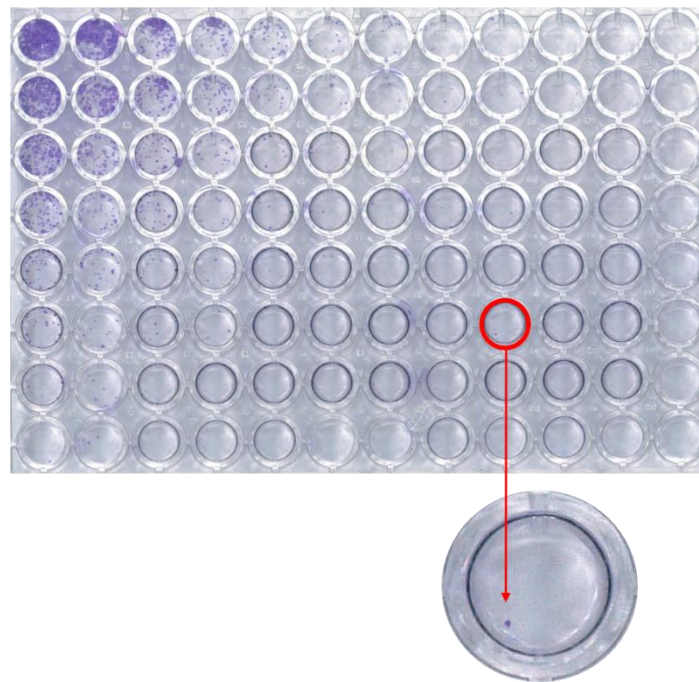

Clonal isolation test of CAL27 cells. CAL27 cells were serially diluted in a 96-well plate and their capacity to grow without attachment to neighbouring cells was tested. Cells were fixed, stained and imaged. The well highlighted shows a small colony of cells derived from a single cell.

Negative control slides for immunohistochemistry
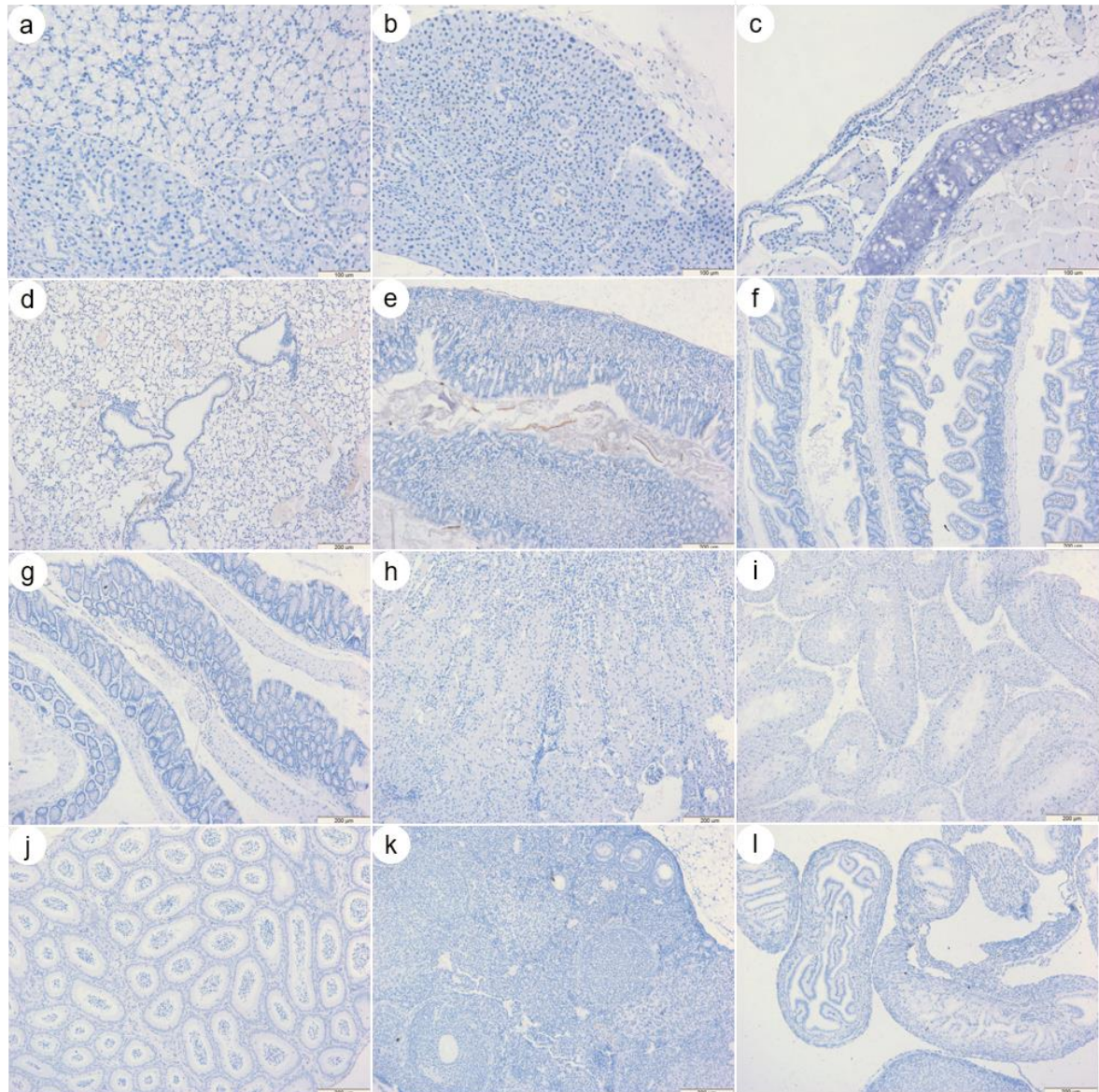
Control immunohistochemistry slides of murine tissues. Tissue sections show WT tissues stained with anti-rabbit-HRP secondary antibody only. Images show representative staining of control submandibular/sublingual gland (a), parotid gland (b), trachea and submucosal glands (c), lung (d), stomach (e), small intestine (f), large intestine (g), kidney (h), testis (i) epididymis (j), ovary (k) and fallopian tube (l). Minor non-specific staining was visible in the parotid gland (b), red blood cells in the lung (d), the chyme of the stomach (e) and the lamina propria of villi of the small intestine (f). Original magnification of the images was 10X (a, d, e, f, g, h, i, j, k, l) and 20X (b, c).

Aperio ImageScope quantification of positive immunohistochemistry staining
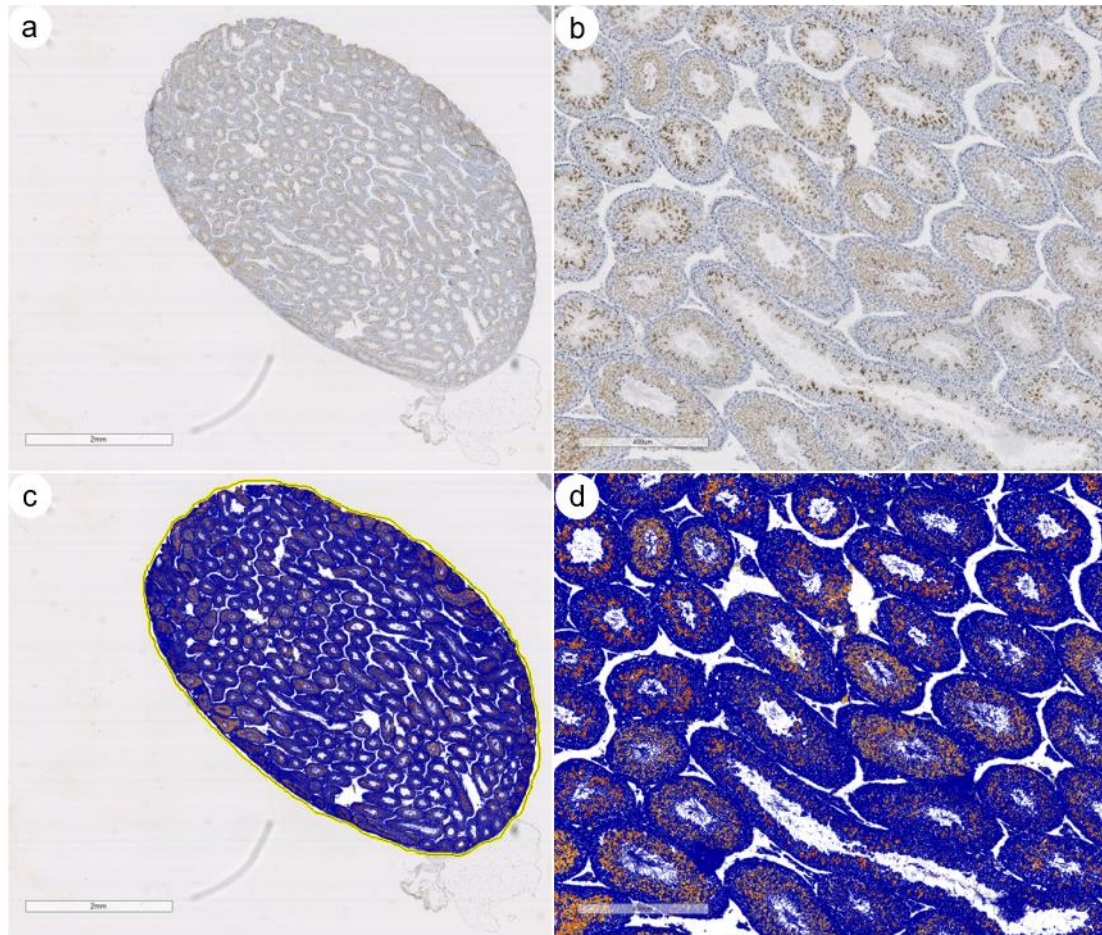
Representative image of immunohistochemistry staining quantification using ImageScope. Scanned slides were analysed for positive DAB staining using Aperio ImageScope software (Leica Biosystems). Tissues were highlighted using the drawing tool and analysed using the Positive Pixel Count v9 algorithm. Positively stained pixels (orange/brown colour) were counted and used to determine percentage of positivity from all pixels. Negatively stained pixels are blue.