THE UNIVERSITY
of EDINBURGH

Linear Domain Interactome and Biological Function of Anterior Gradient 2

Melanie Lawrence

A thesis submitted for the degree of

Doctor of Philosophy

Institute of Genetics and Molecular Medicine

School of Medicine and Veterinary Medicine

The University of Edinburgh

February 2013

Contents

Declaration

I hereby declare that the work presented here is my own and has not been submitted in any form for any degree at this or any other university. Any contribution made by other parties is clearly acknowledged.

Laura Melanie Alexandra Lawrence

Acknowledgements

I would like to thank the MRC for funding this work and my PhD.

I would also like to thank my supervisors, Prof. Ted Hupp, Prof. Nicholas Hastie and Dr. Peter Hohenstein as well as all members of the Hupp and Hastie/Hohenstein labs past and present. In particular, I would like to thank Dr. Euan Murray and Dr. Magdalena Maslon for advice relating to proteomics and general discussions relating to Agr2, and Dr Nils Lindstrom for invaluable advice and help relating to kidney organ culture and development. I also thank colleagues at the MRC Human Genetics Unit, without whose help I could not have completed this work; Dr Karamjit Singh Dolt for work relating to the universal expression vector, Rachel Berry for advice on IHC, Dr Laura Lettice for advice on making targeting vectors and recombineering, Prof Bob Hill for always providing feedback, Dr Abdelkader Essafi for discussions on protein biochemistry, Dr Emma Hall and Dr Pleasantine Mill for advice, reagents and cell lines relating to cilia, all the staff in stores, shared services and of course the wash-up ladies.

For providing the IMCD3 cell lines used extensively in this thesis I would like to thank Prof Ian Jackson's group and Dr Helen Dawe.

For advice and antibodies relating to Meckelin I would like to thank Dr Colin Johnson (Leeds).

For help with the membrane localisation assays for Meckelin and AGR2 I

would like to thank Prof Rory Duncan from Heriot Watt University.

I would also like to acknowledge the help, in large or small part, of all the members of the IGMM over the course of this PhD either through advice, through learning from them at the many unit lectures and talks that I have attended, or simply for a chat and a coffee from time to time.

Finally, I would like to thank my parents for fostering in me a curiosity about the world and telling me that I could do anything I wanted to.

The scientist does not study nature because it is useful; he studies it because he delights in it, and he delights in it because it is beautiful. If nature were not beautiful, it would not be worth knowing, and if nature were not worth knowing, life would not be worth living.

- Henri Poincaré

For my husband David, whom I love,

and for my beautiful children, Cameron and Sophie, who light up my life,

and for Margot, on whose strength I drew

Publication arising from this work:

Karamjit Singh Dolt*, Melanie L. Lawrence*, Eve Miller-Hodges, Anna Thornburn, Paul Devenney and Peter Hohenstein (2013)

A Universal Vector for High-Efficiency Multi-Fragment Recombineering of BACs and Knock-In Constructs.

PLoS ONE 8(4): e62054. doi:10.1371/journal.pone.0062054

* Denotes joint first author publication

Abstract

There is an increasing need for more efficient generation of transgenic constructs. Here we present a universal multi-site Gateway vector to use in recombineering reactions. We show its use in the generation of BAC transgenics as well as the generation of endogenous targeting vectors. The modular nature of the vector allows for rapid modification of constructs to generate different versions of the same construct. As such it will help streamline the generation of series of related transgenic models.

Abstract

Abstract

The Anterior Gradient 2 (AGR2) protein has been implicated in a variety of biological systems linked to cancer and metastasis, tamoxifen-induced drug resistance, pro-inflammatory diseases like IBD and asthma, and limb regeneration. The molecular mechanisms by which AGR2 mediates these various phenotypes in disease progression in both cancer and IBD are poorly understood, as is the biological function(s) of AGR2 under non-disease conditions. Here, we use a combination of biochemical techniques, organ culture, cell biology and mouse genetics to investigate the biological significance of AGR2 both in cell lines and *in vivo*. We present data based on phage-peptide interactomics screens suggesting a role for AGR2 in mediating the maturation and trafficking of a class of membrane and secretory proteins, and investigate a putative interaction between AGR2 and one member of this class of proteins. We also describe the construction of a universal vector for use in making a variety of transgenic animals, and then present data showing its use as a promoter reporter, and attempt to investigate the temporal and spatial expression of AGR2 in the developing and adult mouse. Further, we present data describing the localisation pattern of AGR2 in the developing murine kidney using

a combination of organ culture and antibody staining, and suggest a role for AGR2 in the developing kidney based on this data that is in agreement with a chaperone function for membrane and secretory proteins. Together, these data suggest that AGR2 has an intrinsic consensus docking site for a subset of its client proteins, that AGR2 plays a role in protein maturation in ciliated cell types, and provides a novel biological model to dissect the role of AGR2 in ER-trafficking.

List of Figures

List of Tables

Abbreviations

AGR2	Anterior Gradient 2
APS	Ammonium persulphate
BAC	Bacterial artificial chromosome
BSA	Bovine serum albumin
CM	Cap mesenchyme
CoIp	Co-immunoprecipitation
Da	Daltons
DAPI	4',6-diamidino-2-phenylindole (nuclear stain)
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxy(ribo)nucleotide triphosphate
DTT	Dithiothreitol
E12.5	Embryonic day 12.5 (mouse)
EDTA	Ethylene diamine tetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
ER	Endoplasmic reticulum
ERα	Oestrogen receptor alpha
GFP	Green fluorescent protein
HA	Haemaggluttinin (epitope)

HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IBD	Inflammatory bowel disease
IM	Intermediate mesoderm
kb	Kilo base pairs
LB	Luria Bertani medium
MET	Mesenchymal to epithelial transition
MM	Metanephric mesenchyme
MOPS	3-(N-morpholino)propanesulphonic acid
NC	Nephrogenic cord
ND	Nephric duct
NEB	New England Biolabs
OD	Optical density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PIPES	Piperazine-N,N'-bis(2-ethanesulphonic acid)
RFP	Red fluorescent protein
RV	Renal vesicle
SDS	Sodium dodecyl sulphate

SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
siRNA	Small interfering ribonucleic acid
SP	Signal peptide
SRP	Signal recognition particle
TAE	Tris-acetate Ethylene diamine tetraacetic acid buffer
TE	Tris Ethylene diamine tetraacetic acid buffer
TIRF-M	Total internal reflection microscopy
TEMED	N'N'N'-tetramethyl-ethylene-diamine
UB	Ureteric bud

Chapter 1

Introduction

1.1 Investigating Gene Function

In the 200 years since the "father of genetics," Gregor Mendel, first discovered that traits in pea plants could be passed from one generation to another in a pattern now described as Mendelian Inheritance, the science of genetics has yielded many insights into the biology of living beings, and disease. Although Mendel first described genes as traits that could be inherited as discrete entities, the term gene was not used until much later, and the main composition of the gene, DNA, was not discovered until the 1940s [Avery et al., 1944] [Watson and Crick, 1953a] [Watson and Crick, 1953b]. Since then, we have gained insights into the mechanism of genetic inheritance and translation into proteins, and the structure, packaging and complexity of genetic material. More recently, the complex nature of gene expression has become clearer, with layers of regulation of a gene such as alternative splicing [McManus and Graveley, 2011], epigenetic changes to the genetic material [Bird, 2007], and small inter-

fering RNAs being discovered [Fire et al., 1998] [Wianny and Zernicka-Goetz, 2000]. Genetics seeks to determine the biologicial function of a genetic product, the protein or enzyme, and to investigate its regulation spatially and temporally, at the level of the chromosome and post-translationally. Identification of disease genes has been made easier since the advent of high-throughput and next generation sequencing technologies [Pettersson et al., 2009], and investigating the characteristics of disease-linked genes can give insights into gene function. With estimates of between 20,000-25,000 predicted genes having been identified from the Human Genome Project [Lander et al., 2001] [Pennisi, 2003], the *in vivo* function of many genes remains to be characterised. There are a number of approaches that have been employed to investigate gene function. Transcript expression analysis is a powerful tool for providing quantitative information about mRNA levels in a given organism, tissue type or disease tissue [Loven et al., 2012]. Protein expression pattern analysis using antibodies on fixed tissue can provide information about the *in vivo* localisation patterns of a protein either in normal or in disease tissues. The advent of animal transgenics, whereby foreign genetic material is integrated into the genome of a model animal such as mouse, zebrafish or the fruit fly, has opened up possibilities for exploring gene function *in vivo*. New tools in animal transgenics have also been instrumental in elucidating gene function, for example through reporter lines, gene knockouts or knock-ins, and tissue-specific transgenes. The ability to culture cancerous, transformed non-cancerous or non-transformed primary cells has further enriched our ability to investigate gene function, and the discovery of small interfering RNAs and subsequent use in functional assays has revolutionised the field of genetics. The rise of

the proteomics and interactomics fields have provided a more global approach to uncovering gene function, with techniques such as microarray, chromatin immunoprecipitation chips (ChIP on chip), yeast-two-hybrid (Y2H) screens and phage display proteomics identifying novel interactions or gene expression patterns that can help to characterise protein function.

1.1.1 Protein expression pattern analysis using antibody based methods

The protein expression pattern in a particular cell type and/or at particular points in development can hint at a specific function in the developing or adult organism, and co-expression patterns with known or suspected interacting partners can be particularly useful in elucidating biological function [Zhang et al., 2004]. Antibody-based methods using fixed and sectioned tissues or using organ culture followed by tissue fixing, are one method of determining the protein expression pattern of a gene. Normal or diseased tissues such as tumours can be analysed for the expression of particular proteins using fixed (for example, with paraformaldehyde), processed and wax-embedded samples that are cut into sections 4-10μm thick. Often, the efficacy of antibody binding requires antigen retrieval, a process whereby the proteins in a given sectioned tissue sample are treated to reveal their epitopes. The process by which various methods of antigen retrieval work are not clear, but it is thought that breaking the intermolecular cross-linkages that are formed during fixation may be one mechanism [Leong and Leong, 2007]. Organ culture is now a possibility for some embryonic organs such as the mouse kidney, and these organs can be fixed, permeabilised and stained with antibodies to reveal spatial protein ex-

pression patterns *in vivo*. This technique will be discussed further in chapters 2 and 6.

Although useful, there are limitations to antibody-based methods. Fixed tissue can only give information on the protein expression pattern of a gene at a given point in time. In addition, tissue penetration by antibodies can be a problem for fixed whole organs, and background staining in immunohistochemistry can also be a problem. The process of antigen retrieval described above is not always successful in unmasking an epitope made unavailable by fixing techniques. Availability of suitable antibodies for use in these applications can also be problematic, with many antibodies effective for immunofluorescence studies on cell culture monolayer not suitable on fixed and sectioned tissue. Nevertheless, in some cases there is no alternative to antibody-based techniques. For many proteins and antibodies the challenges listed above can be surmounted, and valuable data can be obtained. For example, comparisons between a particular protein expression pattern or level in normal versus cancerous tissue from a particular tissue type can be useful in highlighting proteins involved in onset or progression of cancer [Como et al., 2002] [Armes et al., 2013]. Although protein expression analysis with antibodies tends to be less quantitative than transcript analysis, new advances such as proximity ligation assays, using antibody based detection systems in fixed tissue and cell culture, may prove useful for providing valuable quantitative protein expression data in the future [Recupero et al., 2013] [Camilo et al., 2012].

1.1.2 Promoter reporters and lineage tracing

Another effective way to investigate expression patterns is to use a marker for a given gene's expression in a model animal (Figure 1.1). It is possible to tag a protein *in vivo* with a marker such as green fluorescent protein (GFP), making a fusion protein, either at the endogenous locus (affecting one or both alleles of the gene), or at another locus using BAC transgenics (see section 3.2). Tagged cDNA constructs with a gene specific short promoter, or enhancer with minimal promoter sequences, have also been described, however, these are less reproducible or predictable than endogenous constructs or those based on BAC transgenics. For promoter reporters, the marker is expressed from the gene's own promoter, leading to temporally-accurate and tissue-specific expression of the marker. In the case of a fusion protein at the endogenous locus, the total amount of protein is not affected but the product of at least one of the alleles might not retain its biological function due to the tag, and in the case of a fusion protein based on BAC transgenics and inserted elsewhere in the genome, there may be an excess of protein that may lead to off-target biological effects. It is also possible to design a promoter reporter to simply express the marker without being tagged to the protein, thus removing the possibility that the tag may influence the function of the protein, although this will no longer provide intracellular information. This can be achieved in the same ways as described above, either by insertion of the marker at the endogenous locus or elsewhere in the genome. Integration at the start codon (replacing it) of the endogenous locus is useful as the location of the foreign transgene is known, however, there may be issues of haplo-insufficiency that need to be considered as the endogenous gene will not be expressed.

Figure 1.1: Promoter Reporters

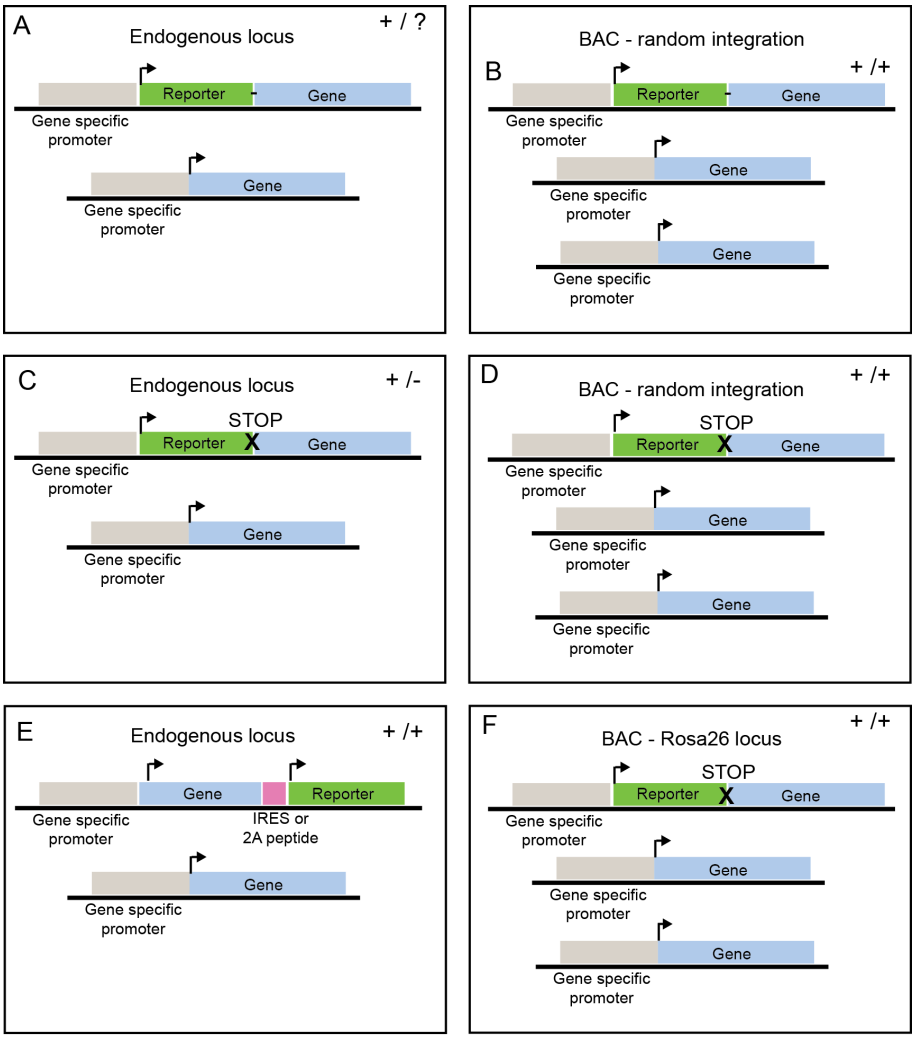


Figure 1.1 Promoter Reporters
Reporter genes are indicated in green, gene of interest in blue, IRES or 2A peptide for restart of transcription in pink. Allele status of the cell is indicated in the top right corner. (A) Reporter-protein fusion promoter reporter at the endogenous locus. Only one allele remains unchanged, the second allele may have disrupted function. (B) Promoter reporter with reporter-protein fusion on randomly-integrated BAC. Endogenous alleles are unaffected but extra gene product or random integration of BAC may lead to off-target effects. (C) Promoter reporter not fused to protein, therefore no intracellular information. Endogenous gene is only expressed from one allele, may cause haplo-insufficiency. (D) Promoter reporter with reporter on randomly-integrated BAC, reporter not fused to protein. No intracellular information and random targeting of BAC may lead to off-target effects. (E) Reporter at 3' end of gene with IRES sequence or 2A peptide to restart transcription of reporter. Reporter not fused to gene, endogenous gene expressed from both alleles. (F) Promoter reporter with reporter on BAC that is targeted to a particular locus such as Rosa26. Reduced chance of off-target effects. Can be fusion or reporter only.

Another option is to insert the marker at the C-terminus of the endogenous gene, using an internal ribosomal entry site (IRES) [Pelletier and Sonenberg, 1988] or 2A peptide from viruses such as enteroviruses [Szymczak et al., 2004] to link the expression of the reporter to the gene of interest. This ensures that the endogenous expression pattern is maintained without losing the allele, thus removing the issue of haploinsufficiency. This method however, also has its disadvantages, since IRES sequences that allow the polymerase to dock back onto the DNA can be unreliable [Hennecke et al., 2001], and 2A peptides are not thought to be reliable for membrane proteins due to extra amino acids left at the N-terminus during the process, although some studies do show efficiency for membrane proteins with modified 2A peptides [Trichas et al., 2008] [Tang et al., 2009] [Kim et al., 2011]. Conversely, insertion at a random location is not likely to affect the native protein, although it is possible for the transgene to integrate randomly into a deleterious genomic location, again leading to off-target effects. Additionally, random integration can also lead to the transgene being located at a region that is sub-optimal for expression such as telomeres, or silenced loci. For this reason, transgenes are sometimes harboured at loci that are known to be "safe" and yet retain widespread tissue expression such as Rosa26 in mouse, which also has been shown to be conducive to a high rate of integration events through homologous recombination [Irion et al., 2007] [Soriano, 1999]. All of these types of transgenic models have advantages and disadvantages as discussed, and these need to be taken into consideration when analysing the data from transgenic models.

Despite the considerations for all of these types of promoter reporters,

they can provide an effective means of investigating the expression pattern of a gene in that the markers are expressed only from the promoter linked to the gene in question. This means that the marker will be expressed not only in the same place and at the same time as the gene of interest, but also at the same levels, providing subtle information that it is not necessarily easily possible to gain using antibodies (although the long half-life of markers such as lacZ and, to some extent GFP, must be kept in mind when analysing data). Furthermore, promoter reporters can be adapted to provide lineage tracing using recombination enzymes such as Cre-recombinase, in which all daughter cells that express a gene of interest will be labelled, even when they no longer express the original gene. This can provide data about the contribution of a gene's expression to the maturation of a particular cell type.

Lineage tracing is a powerful method of following the progeny from a single cell *in vivo*. Essentially, a cell is labelled in such a way that the marker is permanently passed down to all daughter cells, providing information about the number, location and differentiation status of the progeny [Kretzschmar and Watt, 2012]. It has been used as a technique in developmental biology, for example to follow the lineage of stem cells [Snippert et al., 2010], and in cancer to follow the clonal expansion of cancerous cells [Wang et al., 2007]. Fundamental to the concept of lineage tracing is the notion of conditional or inducible genetic systems that can introduce genetic mutations somatically, in a temporally and spatially controllable manner [Nagy, 2000]. The most common systems used to achieve this are the Cre-*loxP* system, the inducible CreERT2 system and the FLP/FRT-recombinase system [Kretzschmar and Watt, 2012]. Cre recombinase is a recombinase enzyme from the P1 bacte-

riophage. It specifically recognises a sequence of DNA known as *loxP* that is 34 base pairs long, and catalyses a recombination event between two loxP sites [Sternberg and Hamilton, 1981] [Hamilton and Abremski, 1984]. The enzyme was subsequently found to be active in eukaryotic cells [Lakso et al., 1992], and was quickly used by ES cell and transgenic technologies as a valuable tool in mediating site-specific changes to the mouse genome [Gu et al., 1993]. The CreERT2 system was designed to provide temporally inducible expression of Cre *in vivo*. In this system, Cre is fused to the estrogen receptor (ER) ligand binding domain, which has been mutated. Upon ligand (tamoxifen) binding, the ER-Cre fusion protein is translocated to the nucleus where the Cre moeity can effect recombination between two loxP sites [Feil et al., 1997]. In this way, recombination to allow expression of a given transgene can be temporally controlled by the administration of tamoxifen. The flippase system works in a similar way to the Cre-*loxP* system, with the enzyme Flp recoginising identical FRT sites on the DNA [Zhu and Sadowski, 1995]. These systems can be employed to label the progeny of a cell by using activation of a recombinase in a founder cell specifically. In a simple example (Figure 1.2), a mouse line with Cre expressed in specific cells only (for example, under the control of a gene-specific or tissue-specific promoter), is then crossed to a separate mouse line in which a marker such as GFP or lacZ is harboured. In the genome at a specific locus, several stop codons and polyA signals flanked by *loxP* sites are placed between the marker cDNA and its promoter, preventing functional expression of the marker. Crossing this line to a Cre-expressing line in which Cre is expressed in a specific cell type or tissue type brings about excision of the *loxP*-flanked stop codon, and therefore expression of the transgenic marker.

Figure 1.2: Lineage Tracing

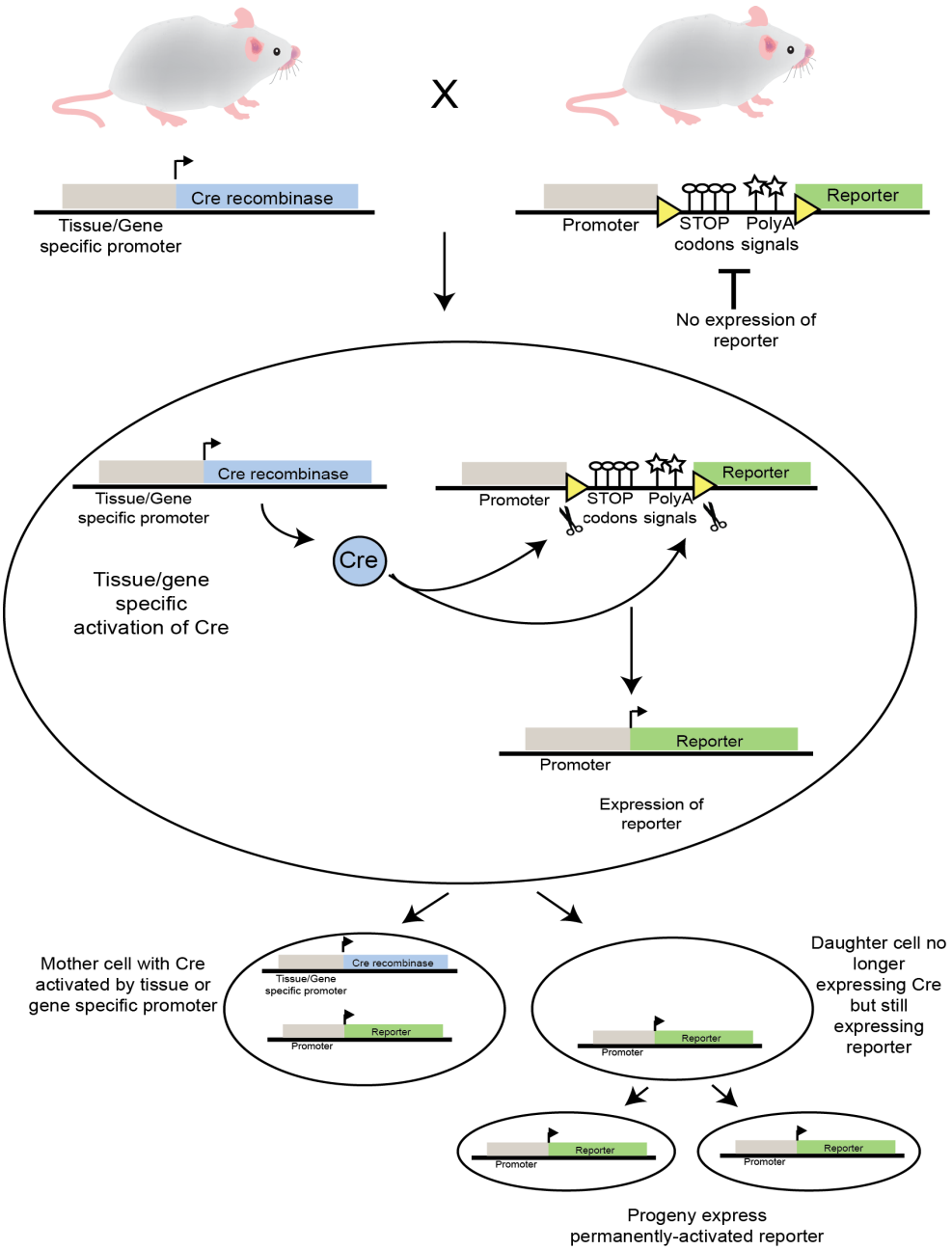


Figure 1.2 Lineage tracing example
(A) A mouse expressing Cre-recombinase under the control of a tissue-specific or gene-specific promoter is crossed to a reporter mouse with multiple stop codons and polyA signals blocking expression of the reporter. (B) Cre is expressed in certain tissues and effects its recombinase activity in those tissues only, removing the stop codons and polyA signals to allow expression of the reporter transgene. (C) Progeny continue to express the reporter transgene even after the Cre-recombinase is no longer being expressed from the tissue or gene-specific promoter. *Mouse image supplied by C Nicol*

In this way, only cells that express Cre or descendants of these cells will be labelled. An elegant example of the CreERT2 system was carried out using intestinal stem-cell specific expression of CreERT2 in a mouse line, crossed with a mouse line harbouring a loxP-flanked Neomycin resistance cassette that effectively stopped transcription of four fluorescent markers. In this model, the choice of which of the four reporters to be activated was random in each activated cell. Expression of Cre-ERT2 only in the intestinal stem cells allowed lineage tracing and clonal analysis of single stem cells, and determination of their fate in the intestinal crypt, leading to a new model of stem cell division and fate [Snippert et al., 2010]. Other more complex lineage tracing systems have also been described for lineage tracing of cancerous cells, such as the MADM system in which interchromosomal recombination can be followed *in vivo* [Zong et al., 2005] [Luo, 2007].

1.1.3 Interactomics

Whilst antibody expression data and animal transgenics can provide information about a gene's biological function, proteomics screens and functional protein-based assays have also added to our toolkit for investing gene function. Using biochemical and imaging techniques such as co-immunprecipitation, enzymatic assays, immunofluoresence and mass spectrometry, many questions relating to protein interactions or activity can be answered. More recently, larger-scale approaches to identifying gene function such as mass-spectrometry based identification of organelle compostion [Ishikawa et al., 2012], transcriptomics screens in healthy versus diseased tissues [Strausberg et al., 2004], phenotypic screens using siRNA-mediated depletion [Boutros and Ahringer,

2008] and discovery of interacting proteins or peptides with yeast-two-hybrid (Y2H) [Rual et al., 2005] [Bruckner et al., 2009] or phage peptide display screens [Murray et al., 2007] have been employed. The growing field of interactomics uses these techniques to identify protein networks, defining the "interactome" of proteins to help give a more global picture of protein function in certain environments and cellular compartments.

Linear Peptide motifs

Both transient and more robust protein-protein interactions have classically been thought to occur through the effect of hydrophobic patches between large globular domains, but many examples of short linear peptides binding to globular domains have now been described [Neduva and Russell, 2006] (Figure 1.3). For example, the C-terminal KDEL ER-retrieval motif has been shown to bind a receptor that mediates retrograde transport and retrieval from the golgi, [Raykhel et al., 2007], and a consensus motif ψ-K-x-D/E (where ψ is a hydrophobic residue) is found on proteins that are sumoylated and is recognised by the SUMO transferase Ubc9 [Teng et al., 2012]. These motifs may lie on the surface of a protein or within a disordered, accessible loop [Dyson and Wright, 2005] and some motifs are specific to regions on proteins that are post-translationally modified such as phosphorylation or glycosylation [Huang et al., 2008]. For proteins capable of binding a linear peptide motif *in vivo*, identifying those motifs is key to defining a group or class of proteins with which they interact, and thus determining their interactomes [Neduva and Russell, 2005].

Bioinformatics advances and the sequencing of the human genome (and

Figure 1.3: Linear Peptides Mediating Interactions

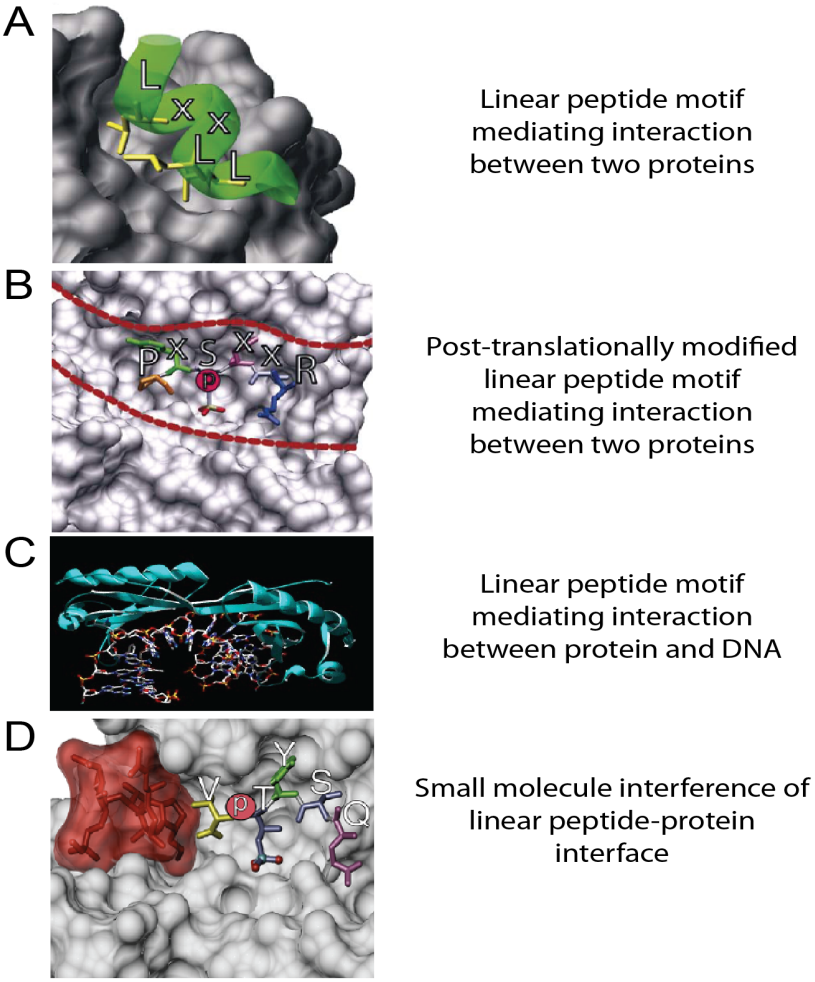


Figure 1.3 Linear peptides mediating interactions
(A) An α-helical peptide from oestrogen co-activator GRIP1 shown in green, bound to human oestrogen receptor (gray), mediated by linear peptide motif LXXLL. (B) Serine-posphorylated linear peptide bound to human 14-3-3 ζ protein (gray), which only recognises the phosphorylated form of the linear peptide motif. (C) Human 14-3-3 protein binds to DNA sequence TATAAAAG specifically through a linear motif consisting of four phenylalanine residues that interdigitate into the minor groove of the DNA, bending it and enhancing the protein-DNA interaction. (D) Small molecules can interfere with peptide-protein interactions; the small molecule toxin fosicoccin stabilises the interaction between a plant 14-3-3 protein (gray) and a peptide ligand from plant H^+ATP-ase by interacting with both protein and ligand [Neduva and Russell, 2006].
Images A, B, and D reproduced from Neduva and Russell, 2006. Image C reproduced from http://www.wiley.com/college/pratt/0471393878/student/structure/dna_binding_proteins/index.html

other species) have allowed genome-scale interaction studies to identify previously-unknown linear motifs [Neduva et al., 2005]. Defining interactomes may be particularly important for "hub" proteins such as p53 [Vogelstein et al., 2000] or the ATM kinase [Matsuoka et al., 2007], that interact with many different proteins either in a common pathway or at a particular cellular location. In addition, identifying peptide-globular domain interactions may be important for disrupting medically-relevant interactions with small molecules, since very few examples of small molecules disrupting large interfaces have been described [He et al., 2005], [Neduva and Russell, 2006]. Screening for linear peptide motifs that provide docking sites for a given protein can be achieved using various techniques, and one technique that has been employed is phage peptide display [Smith, 1985] [Rossenu et al., 1997].

Phage peptide display proteomics

Phage display is a technique that was developed to screen for protein-protein, protein-peptide and protein-DNA interactions [Smith and Petrenko, 1997]. First described in 1985, phage display involves the use of a bacteriophage as a host system for expression of polypeptides on one of the coat proteins on the surface of the phage virion (normally M13 filamentous phage) [Smith and Petrenko, 1997]. In this way, large combinatorial phage peptide libraries with a random peptide expressed on the virion coat protein of each virion can be built up [Noren and Noren, 2001]. The phage library, containing many different presented peptides, is washed over immobilised target protein or DNA in a microtitre plate, and non-binding phage is washed away. The bound phages are then harvested and used to infect bacteria, and the phage DNA can then

be sequenced to identify the unique cDNA, and thus peptide or polypeptide, that binds the target protein or DNA. The technique was first used in the application of phage antibody libraries for making custom antibodies to target proteins [McCafferty et al., 1990] [Hoogenboom et al., 1998], with the immunoglobulin variable region displayed on the virion surface to screen for binding affinity. The process was improved by developing a system of selective enrichment, in which phage virions that bind the target protein are enriched through successive rounds of infection and selection, allowing rare binding peptides to be identified out of a large pool (as much as one in a million) [Crameri and Walter, 1999] (Figure 1.4). Later, the technique has been used for identification of receptor-ligand interactions [Hartley, 2002], and linear peptide-protein interactions [Smothers and Henikoff, 2001]. For example, phage peptide display was able to identify linear motifs that bound to the p53 co-activator p300 (PxxP, PxP, PRLP), and this motif was found to be present in p53 and validated as a p300 docking site [Dornan et al., 2003]. The proline repeat motif has since been shown to be important in transcription factor signalling [Neduva and Russell, 2007].

More recently, some criticism of phage display as a technique for identifying peptide-protein interactions has been put forward, with a suggestion that the technique may be biased toward hydrophobic interactions [Luck and Trave, 2011], and that this bias may exclude important interactions based on linear peptides such as those incorporating electrostatic interactions or Van der Waals interactions. Nevertheless, it is clear that phage peptide display proteomics, when paired with bioinformatics resources and annotated genomes, is a very powerful technique for interrogating protein-protein interaction networks and

Figure 1.4: Phage Peptide Display

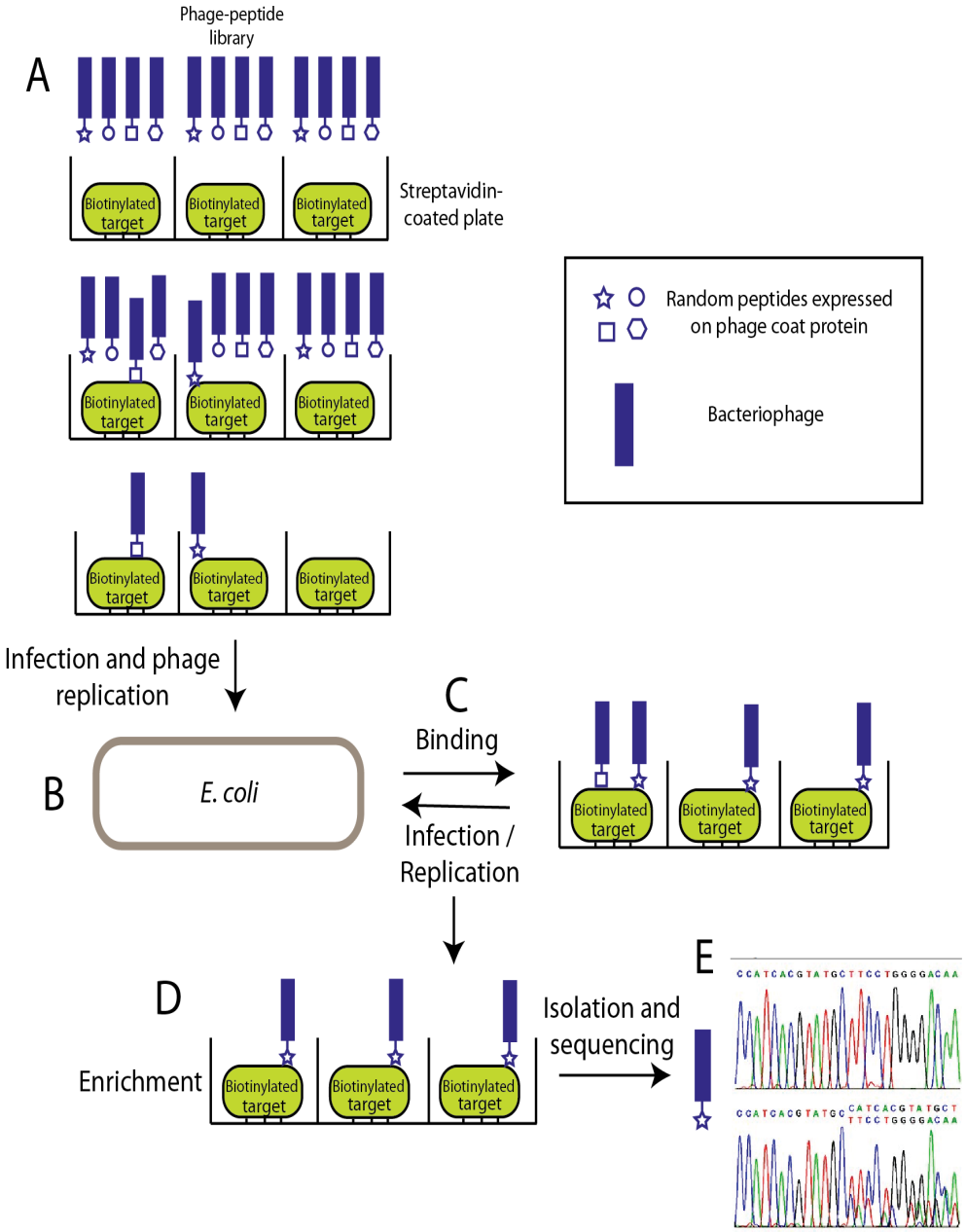


Figure 1.4 Phage Peptide Display
(A) A phage peptide library expressing random peptides on a virion coat protein is washed over immobilised target protein or DNA, and unbound phage virions are washed away. (B) Bound phages are used to infect *E.coli* for replication. (C) Successive rounds of binding, selection and infection lead to enrichment of binding peptides with highest affinity (D). Enriched phage-peptides are isolated and sequenced (E).

providing leads for validation *in vivo*.

1.2 Anterior Gradient 2

1.2.1 AGR2 expression patterns in vertebrates

The Anterior Gradient-2 (AGR2) family was first described in 1998 as a secreted protein in *Xenopus laevis* eggs [Aberger et al., 1998]. The secreted protein in *X. laevis* was found to be involved in anterior specification of the embryo, and expression was reported in the forebrain and mucous-secreting cement gland of the developing *Xenopus* embryo. Expression of AGR2 homologues have also been reported in other vertebrates. In zebrafish (*Danio rerio*), zAgr2 is widely expressed in many tissues including epidermis, oesophagus, olfactory bulbs, pharynx, gills and goblet cells of the intestine, as verified by qRT-PCR and whole-mount in situ hybridisation [Shih et al., 2007] [Chen et al., 2012]. Another study found that an AGR2 homologue was upregulated in infected tissues of Atlantic salmon with amoebic gill disease, and expression in healthy salmon was mainly found in epithelial tissue of the gill, intestine and brain [Morrison et al., 2006]. In the gibel carp (*Carassius auratus gibelio*), an AGR2 homologue was shown to be expressed in the mucus-secreting hatching gland, and later the pharynx, swim bladder and pronephric duct. In the adult carp, AGR2 mainly localised to the kidney epithelia, and was found at lower levels in the intestine, ovary and gills [Xia et al., 2009].

Expression of the AGR2 homologue nAG was also described in the adult salamander in a study investigating newt limb regeneration [Kumar et al.,

2007]. Newt limbs regenerate by de-differentiating limb cell types (skeletal, bone and non-neuronal cells of the nervous system) into earlier stem-cell like precursors, followed by proliferation of these stem-cell like cells mediated by various developmental genes into a mass called the blastema [Stocum, 2007]. Re-differentiation, however, into specified cell types to re-grow a limb is poorly understood. In this study, the newt (salamander) AGR2 homologue *n*AG was found to promote limb regeneration after amputation and to interact directly with the protein Prod-1 in Y2H and co-immunoprecipitation assays. Prod-1 is a GPI-anchored cell-surface protein that is expressed in a proximal-distal gradient in the limb blastema, and is essential for limb regeneration. The proximal-distal gradient of Prod-1 exists in both regenerating and normal limbs, and Prod-1 is regulated by retinoic acid [da Silva et al., 2002]. *n*AG was expressed in the nerve sheath of the amputated limb, and in glands in the blastema of the regenerating limb. Surprisingly, introduction of *n*AG by electroporation directly into the blastema of the amputated limb was able to restore regeneration of distal structures (digits), but without promoting nerve regeneration in the the limb. As a secreted ligand for Prod-1, *n*AG represents a novel factor in the process of limb regeneration in amphibians and may also represent an important factor in regenerative medicine in humans. Human AGR2 has been reported to interact with another GPI-anchored protein, metastasis protein C4.4, which shares some domain architechture with Prod-1 (cysteine-rich domains and Ly6-type domains) [Fletcher et al., 2003], suggesting that there might be a conserved, functional interaction pathway.

In mice, transcript analysis initially revealed expression of the AGR2 homologue (*mAgr2*, *Gob4*) in the intestines, colon and stomach of the adult

mouse [Komiya et al., 1999], and in situ hybridization determined that *m*Agr2 was localised to the mucus-secreting cells (goblet cells) of these tissues. Expression of AGR2 in the goblet cells of the intestine in mice was also reported in a more recent study [Park et al., 2009]. Goblet cells are specialised cells of the intestinal epithelium, and their main function is to secrete mucins. Mouse Agr2 and *Xenopus laevis* Agr2 share significant homology (46%), and both contain a putuative cleavable leader sequence required for targeting proteins to the secretory pathway [Komiya et al., 1999]. This finding in mice, combined with the conserved expression of the AGR2 homologue in *X. laevis* in mucus-secreting cement gland, was the first suggestion that AGR2 might be involved in the secretory pathway.

The Anterior Gradient family in humans consists of two closely related genes, *AGR2* and *AGR3*, located close to each other (60 kilobases apart) on Chromosome 7p21.1 [Petek et al., 2000] (Figure 1.5A) and sharing 71% identity [Fletcher et al., 2003]. Another homologue, *AGR1* (also known as *Erp18/19*, or TXNDC12, is located on Chromosome 1 in humans and is less closely related to *AGR2* and *AGR3* [Rowe et al., 2009]. The human *AGR2* gene spans 8 exons over 11 kilobases (Figure 1.5B), with the start codon in the second exon. The AGR2 protein shares close homology with the oxidoreductase family of protein disulphide isomerases (PDIs) [Persson et al., 2005]. The active motif of the PDI family is CXXC, which is also found in thioredoxins and is therefore also referred to as Tx-domain motifs [Ferrari and Soling, 1999]. Although both AGR2 and AGR3 share homology with thioredoxin, in humans they lack the disulphide active site, with a serine in place of the second cysteine (Figure 1.5C).

Figure 1.5: Human Anterior Gradient Homologues

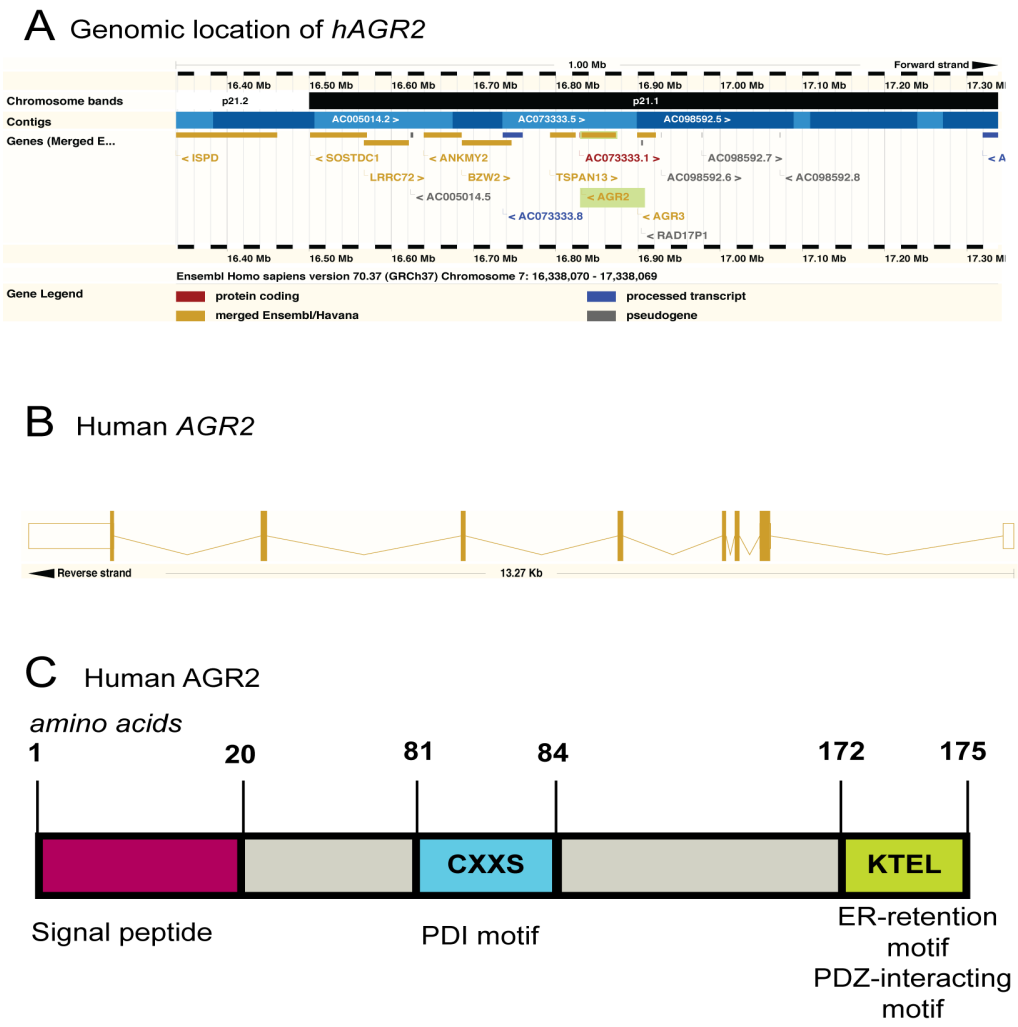


Figure 1.5 Human *AGR2*
(A) Human *AGR2* is located at Chromosome 7p21.1, 60kb from homologue *AGR3* and spanning approximately 11kb. (B) *hAGR2* is transcribed from the reverse strand and consists of 8 exons. The transcript is 1778bp and the start site is in exon 2. (C) The gene encodes a full length protein of 175 amino acids. The protein contains a putative cleavable signal peptide at the N-terminus (residues 1-20) that targets the nascent polypeptide to the secretory pathway, a thioredoxin-like PDI motif (residues 81-84) and an ER-retention sequence at the C-terminus (residues 172-175). The protein also contains a putative PDZ-interacting motif at the C-terminal, X-T/S-X-Φ, where Φ is a hydrophobic amino acid and X is any amino acid.
A, B reproduced from Ensembl (www.ensembl.org)

This CXXS motif has been identified in another ERp protein, albeit with lower PDI activity than members containing the CXXC motif [Anelli et al., 2002]. Sequence alignment of AGR2 homologues from *Xenopus laevis* AG-1 and AG-2, *Xenopus laevis* Agr2, Pufferfish (*Tetraodon nigris*) agr2, Atlantic salmon (*Salmo salar*) agr2, Atlantic salmon Agr3, Zebrafish (*Danio rerio*) Agr2, Zebrafish Agr1, Mouse (*Mus musculus*) Agr2, Human (*Homo sapiens*) AGR2, Mouse Agr3, Human AGR3 and Human AGR1 shows that the least conserved regions of these genes is in the N-terminal region (Figure 1.6A, B). Phylogenetic analysis of Agr homologues in fish, amphibians, reptiles, birds and mammals determined that there were three subfamilies of Agrs; Ag1, Agr2 and Agr3 [Ivanova et al., 2013]. Agr2 and Agr3 subfamily members are present only in amphibians and higher vertebrates, and are located on the same syntenic fragment, whilst the Ag1 subfamily members are located on a different syntenic fragment and are only present in fishes and amphibians. These Ag1 subfamily members are lost in reptiles, birds and mammals, suggesting that loss of this subfamily member may correspond to loss of regenerative ability associated with fish and amphibians [Ivanova et al., 2013]. It has been suggested that human Agr2 and Agr3 are the result of tandem duplication [Fletcher et al., 2003], and this may also be the case for mouse anterior gradient 2 family genes.

1.2.2 Animal Models and Biology of AGR2

The human Anterior Gradient 2 gene (*hAGR2*, *HAG2*, *GOB4*) encodes a 19,878 Da protein that is regulated by the proteasome [Pohler et al., 2004], and which has been proposed as a member of the PDI family of molecular

Figure 1.6: Human Anterior Gradient 2

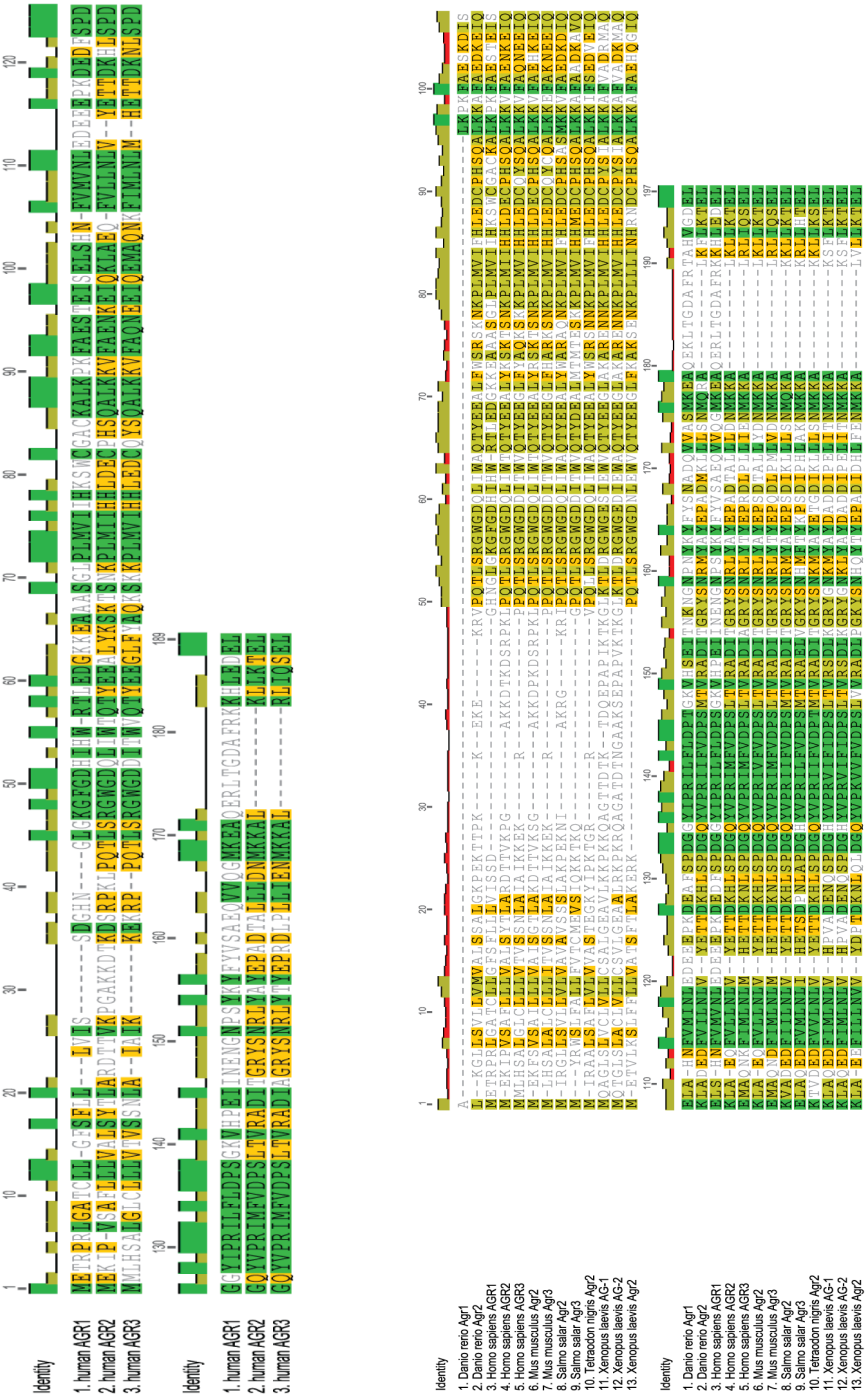


Figure 1.6 Alignment of Anterior Gradient homologues

(A) Alignment of human Anterior Gradient homonlogues (B) Multiple alignment of various anterior gradient homologues from Human (*Homo sapiens*), Mouse (*Mus musculus*), Altlantic Salmon (*Salmo salar*), Pufferfish (*Tetraodon nigris*), Zebrafish (*Danio rerio*) and African clawed frog (*Xenopus laevis*). *Figures produced in Geneious* (www.geneious.com)

chaperones [Persson et al., 2005]. Recent work has shown that it is capable of dimerising via a cysteine bridge allowing it to mediate ER-stress responses [Ryu et al., 2012], and it has also been shown to function as a PDI *in vivo*, interacting with target protein MUC2 in the ER in part through this cysteine [Park et al., 2009]. *h*AGR2 is 175 amino acids long and contains a putative N-terminal cleavable leader sequence (also known as a signal peptide) [Pohler et al., 2004] that targets proteins to the ER membrane and secretory pathway (Figure 1.6C). A predicted cleavage site would yield a mature, processed form, after cleavage of the N-terminal residues by signal peptidase in the ER lumen. Biochemical studies into the localisation of these two forms of AGR2 using transfected full-length or mature forms have suggested differential localisation [Fourtouna et al., 2009], with the full length protein localising predominantly to the ER and the shorter, mature protein lacking the N-terminal leader sequence localising primarily to the nucleus. While the mature protein does contain a potential nuclear localisation signal, the presence of such a consensus site has not been shown experimentally. The leader sequence targets nascent proteins to the ER membrane and is then cleaved, with membrane domains remaining in the ER membrane or, in the case of soluble proteins, localising to the ER lumen (discussed further in section 1.3). Co-translational cleavage by signal peptidase is thought to be essential for correct folding of the nascent protein [Vitale and Denecke, 1999], and it is presumed that the transfected, full-length form would be targeted, upon translation by the ribosome, to the ER membrane by the signal peptide and be processed into a soluble, mature form resident in the ER lumen. The transfected form of mature AGR2 lacks this leader sequence and would therefore not be targeted to the ER membrane,

but would rather remain soluble within the cell and outside the ER lumen. It is not clear whether the differential localisation of transfected, mature and transfected, full length AGR2 is due simply to the lack of the leader sequence in the mature form, however, studies have shown that AGR2 can function as a p53 inhibitor in the cytoplasm, promoting nuclear exclusion [Pohler et al., 2004], that endogenous AGR2 localises to the nucleus as well as the ER in fractionation assays [Fourtouna et al., 2009], and that it forms a complex with reptin, a predominantly nuclear protein [Maslon et al., 2010], suggesting distinct nuclear functions for AGR2, or possibly an unidentified isoform lacking the N-terminal leader sequence and allowing a pool of cytoplasmic and nuclear AGR2 to exist.

In addition to the N-terminal leader sequence, AGR2 also shares a C-terminal ER-retrieval sequence with other members of the molecular chaperone families. The classical 4-amino acid motif consists of H/KDEL, and has been shown to act as an ER retrieval sequence to return ER-resident proteins once they have left the ER for the secretory pathway [Munro and Pelham, 1987]. Early studies determined that the motif is recognised by the golgi-resident KDEL receptor in mammalian cells, which binds to the protein and triggers retrograde transport back to the ER lumen [Yamamoto et al., 2003]. Variations on the sequence have been identified [Persson et al., 2005], including KTEL, the terminal 4-amino acid sequence at the C-terminus of AGR2, and differing affinities for the KDEL receptor are associated with these variations [Raykhel et al., 2007] [Alanen et al., 2011]. Other mechanisms exist for ER-retention of proteins; ER-membrane resident proteins often have a di-lysine motif in their C-terminal region [Teasdale and Jackson, 1996], and ER-membrane or

luminal associated chaperone proteins such as BiP (Grp78) are retained in the ER in large complexes mediated by their membrane or luminal domains [Nilsson and Warren, 1994]. Although the H/KDEL motif has been shown to function as an ER-retention motif, examples of proteins with this motif at their C-terminal that have functions and localisations outside the ER have been described [Inohara et al., 1989] [Akasofu et al., 1989]. It is not known if this is a result of the error rate of the KDEL motif as an ER-retrieval sequence, or whether escape from retrieval is mediated by an unknown factor [Vitale and Denecke, 1999]. More recently, Dong et al. have shown that deletion of the KTEL motif at the C-terminus of AGR2 in human cancer cells led to secretion of AGR2, and that it is essential for the induction of *AREG* (amphiregulin, an EGF-receptor ligand) by AGR2 [Dong et al., 2011] in esophageal and lung adenocarcinomas. However, achieving ER-retention using the classical KDEL motif did not rescue function [Gupta et al., 2012a], suggesting that the C-terminal motif in AGR2 confers specificity for the function of AGR2 and is not simply an ER-retrieval sequence.

Other studies have found secreted extracellular AGR2 in the serum of lung adenocarcinoma patients [Chung et al., 2011], the serum of prostate cancer patients [Kani et al., 2013] and at the extracellular cell surface of circulating pancratic cancer cells [Dumartin et al., 2011]. In this last study, AGR2 was also found to increase the level of secretion of the precursor of cathepsin D, a secreted lysosomal protein. Furthermore, immunohistochemical analysis of AGR2 expression in prostate cancer tissues demonstrated a strong expression pattern in secretory epithelial cells [Zhang et al., 2005], and in the chicken oviduct, enriched with secretory cells secreting proteins such as ovalbumin,

AGR-2 was identified as being highly expressed [Kim et al., 2007], further supporting a role for AGR2 in the secretory pathway. Another study found that AGR2 localised to the ER lumen, and was found to be bound to cargo proteins in the secretory pathway as well as nascent polypeptide chains as they translocated through the ER membrane [Higa et al., 2011]. In this study AGR2 was found to be involved in maintenance of ER homeostasis, and that its expression was controlled by the unfolded protein response (UPR), implicating AGR2 not only in folding and trafficking of membrane or secreted proteins, but also in the ER stress response. In a separate study of pancreatic cancer cells, induction of AGR2 regulated expression of ubiquitin-proteasome degradation pathways, ER chaperones such as PDI, and lysosomal proteases such as cathepsins B and D, suggesting association with the ER-stress response in these cells [Dumartin et al., 2011].

The first studies characterising the human AGR2 gene and protein identified similarities with members of the molecular chaperone families, such as the classical PDI/thioredoxin structural fold and strong ER association [Persson et al., 2005] [Higa et al., 2011]. This has provided some suggestions as to the biological function of AGR2. Animal models implicate AGR2 in intestinal goblet cell function [Park et al., 2009] and gastric cell differentiation [Gupta et al., 2012b], and it has also been suggested to play a role in other developmental processes such as limb regeneration in amphibians [Kumar et al., 2007]. In the first mouse knockout paper, the authors provide a link with a potential substrate protein, MUC2, demonstrating an interaction between these two proteins involving the single cysteine in the AGR2 PDI-like motif and demonstrating that AGR2 is required for the production of MUC2 [Park

et al., 2009]. MUC2 is a mucin, one of the secreted mucin proteins manufactured by goblet cells in the intestine. Other mucins are present in other tissues as well, and AGR2 has also been shown to be sufficient and required for the production of the cancer-associated mucin MUC1 in pancreatic ductal adenocarcinoma cells [Norris et al., 2012]. Furthermore, a zebrafish knockdown of AGR2 using morpholinos confirmed a role for AGR2 in goblet cell mucin production, as seen by a reduction in alcian blue staining of mucins (unpublished, personal communication, E Patton), and a more recent study has found that zebrafish Agr2 is essential for mucin production and terminal differentiation of intestinal goblet cells [Chen et al., 2012]. Although the authors of the first mouse knockout paper did not report any other noticeable phenotypes, a more recent AGR2 mouse knockout study has implicated AGR2 in gastric cell differentiation [Gupta et al., 2012a]. In this paper, the authors identify AGR2 as essential for differentiation of several stomach cell types including chief cells, parietal cells, enteroendocrine cells and mucus-secreting pit cells. Loss of Agr2 in these mice resulted in an increase of stem-cell like stomach precursors expressing Sox9, and a decrease in terminally differentiated cells, suggesting a role for Agr2 in mediating the stem-cell - differentiated cell switch required for terminal differentiation. There may be an important link between the findings of this study and the nAG (newt homolog) involvement in newt limb regeneration discussed in section 1.2.1, where cells in the amputated limb revert to a more stem-cell like state before differentiating into the appropriate cell types in an nAG-dependent manner.

1.2.3 Disease association of *h*AGR2

AGR2 in cancer

Cancer type	Mechanism of Involvement	Source
Breast adenocarcinoma	Upregulation/Overexpression	[Duran et al., 2008] [Fletcher et al., 2003]
	Associated with ERBB2^{+} tumours	[Duran et al., 2008]
	Poor prognosis	[Barraclough et al., 2009] [Hrstka et al., 2010]
	Metastasis	[Liu et al., 2005] [Barraclough et al., 2009] [Fletcher et al., 2003]
Mucoid Ovarian	Upregulation	[Park et al., 2011]
	Overexpression	[Gray et al., 2012]
	High plasma concentration	[Edgell et al., 2010]
Other Ovarian	Poor prognosis	[Darb-Esfahani et al., 2012]
	Upregulation/Overexpression	[Armes et al., 2013]
Prostate	Overexpression in blood serum	[Kani et al., 2013]
	Upregulation/Overexpression	[Kristiansen et al., 2005] [Bu et al., 2011]
	Metastasis	[Hu et al., 2012]
	Poor prognosis	[Zhang et al., 2007]
Pancreatic	Upregulation/Overexpression	[Barry et al., 2012] [Ramachandran et al., 2008]
	Secreted in pre-malignant neoplastic juice	[Chen et al., 2010] [Makawita et al., 2011]
	Promotes initiation/progression of pancreatic epithelial neoplasia	[Norris et al., 2012]
Gastric	Metastasis	[Lee et al., 2011]
Colorectal	Poor prognosis	[Valladares-Ayerbes et al., 2012]
Pre-malignant Barrett's epithelial	Overexpression	[Pohler et al., 2004]
Lung	Overexpression	[Chung et al., 2012] [Fritzsche et al., 2007]

Table 1.1: *AGR2* in Cancer

Although the biological function of AGR2 has not been fully elucidated, links with disease, and particularly cancer, have been well documented. AGR2 and its close homologue AGR3 have been shown to be involved in both onset and disease progression of cancer. AGR2 has in particular been highlighted in numerous screens for genes differentially expressed or regulated in various

Resistance	Involvement of AGR2	Source
Tamoxifen	Induced by tamoxifen, overexpressed in tamoxifen-resistant tumours	[Hrstka et al., 2010]
Cisplatin	Mediates resistance in xenograft models	[Gray et al., 2012]
Doxorubicin	Overexpressed in doxorubicin-resistant tumours	[Hrstka et al., 2010]
Docetaxel	Reduced expression in docetaxel-resistant tumours	[Zhao et al., 2009]
Anti-estrogen therapy	AGR2 and Cathepsin D mRNA upregulated in resistant cell line	[Huber et al., 2004]

Table 1.2: *AGR2* in Cancer Drug Resistance

types of cancer including breast [Duran et al., 2008] [Fletcher et al., 2003], pancreatic [Barry et al., 2012], prostate [Kani et al., 2013] [Kristiansen et al., 2005] [Bu et al., 2011], and lung [Chung et al., 2012] [Fritzsche et al., 2007]. AGR2 is a survival factor that has been shown to contribute to inhibition of the tumour suppressor p53 [Pohler et al., 2004] and to promote nuclear exclusion of p53 in cultured cells subjected to DNA damage [Fourtouna et al., 2009]. In a screen of premalignant lesions from Barrett's epithelium, a pre-neoplastic oesophageal disease caused by damage from bile reflux, AGR2 was found to be strikingly overexpressed [Pohler et al., 2004], suggesting that it contributes to cellular transformation. AGR2 has also been shown to be co-expressed with the estrogen receptor in breast cancer cell lines [Thompson and Weigel, 1998], and the gene has been shown to be estrogen responsive, containing up to four putative estrogen response elements (EREs) in its promoter region [Kim et al., 2007] [Hrstka et al., 2010]. In fact, AGR2 is overexpressed in many hormone-responsive cancers. In breast carcinomas, high AGR2 expression was associated with ERBB2 positive tumours [Duran et al., 2008], a member of the epidermal growth factor receptor family involved in pathogenesis of aggressive breast cancers. In breast carcinomas it is also associated with poor

prognosis [Barraclough et al., 2009] [Hrstka et al., 2010], and with a malignant metastatic phenotype [Barraclough et al., 2010]. In ovarian cancers, AGR2 is associated particularly with mucinous tumours [Park et al., 2011] [Gray et al., 2012], but also with endometrioid tumours [Armes et al., 2013] and serous tumours [Darb-Esfahani et al., 2012]. A high blood plasma concentration of AGR2 has also been associated with both serous and non-serous ovarian cancers [Edgell et al., 2010]. Increased levels of AGR2 are also associated with many prostate carcinomas [Kristiansen et al., 2005] [Kovalev et al., 2006]. AGR2 transcripts have been detected in the urinary sediments of both healthy and cancerous patients, with increased levels in the urinary sediments of patients suffering from prostate cancer [Bu et al., 2011]. AGR2 has also been detected at higher levels in the serum of prostate cancer patients, leading some to propose AGR2 as a blood-based biomoarker for prostate cancer [Kani et al., 2013]. Knockdown of AGR2 in a prostate cancer cell line induced cellular sensescence, linking AGR2 not only to horomone-related tumours but also metastasis in prostate cancer [Hu et al., 2012]. In addition, AGR2 is associated with poor survival in prostate cancer patients [Zhang et al., 2007].

AGR2 has also been studied in non-hormone related cancers. Elevated levels of AGR2 have been found in the pancreatic juices of pre-malignant neoplasia of the pancreas [Chen et al., 2010] [Makawita et al., 2011]. It has been found to promote initiation and progression of pancreatic epithelial neoplasia [Norris et al., 2012], and was found to be highly differentially expressed in pancreatic cancers (14-fold increase of mRNA levels) compared to normal pancreatic tissue in one study [Ramachandran et al., 2008]. AGR2 is implicated in gastric cancer cells that have a high metastatic potential [Lee et al.,

2011], and more recently has also been implicated in colorectal cancer and is associated with poor prognosis [Valladares-Ayerbes et al., 2012]. Furthermore, AGR2 was found to be present at higher levels in the serum of lung adenocarcinoma patients [Chung et al., 2011], and in one study to be elevated in 94% of lung adenocarcinomas [Chung et al., 2012] athough in this last study AGR2 was not found to be associated with worse prognosis. Other groups have also found elevated expression of AGR2 in non-small cell lung adenocarcinomas [Fritzsche et al., 2007] [Pizzi et al., 2012], and again the elevated expression was not found to be prognostic of poor outcome. In addition, AGR2 has been implicated in renal Wilms tumours, where a locus including AGR2 was identified as a as a loss of heterozygosity site associated with some of these tumours [Sossey-Alaoui et al., 2003], although the region contains several other candidate genes as well and this association has not yet been validated.

There is evidence suggesting that AGR2 can mediate cytotoxic drug resistance in tamoxifen and doxorubicin-resistant tumours, that it is induced by tamoxifen [Hengel et al., 2011] [Hrstka et al., 2010], and that both AGR2 and AGR3 can mediate cisplastin resistance in xenograft models [Gray et al., 2012]. AGR2 is overexpressed in tamoxifen and doxorubicin-resistant breast cancers, is associated with poor prognosis in tamoxifen-treated tumours [Hrstka et al., 2010], and has been shown to play a role in Docetaxel resistance (although in this model AGR2 downregulation is associated with resistance) [Zhao et al., 2009]. In a study involving anti-estrogen therapy-resistant cancer cell line T47Dr, AGR2 and cathepsin D were found to be among those genes whose mRNA was upregulated in the resistant cell line compared with the non-resistant control [Huber et al., 2004].

Links between AGR2 and metastasis have also been found, with one study finding that AGR2 interacts with metastasis protein C4.4 [Fletcher et al., 2003]. It has also been shown to mediate metastasis in animal models [Liu et al., 2005], with overexpression of AGR2 in this study shown to induce non-metastatic breast cancer cells to metastasise to the lung. In another study, AGR2 was shown to mediate tumour growth, cellular transformation and cell migration both *in vivo* and *in vitro* [Wang et al., 2008]. Other studies also demonstrate the ability of AGR2 to promote cell proliferation, cell cycle progression and induced cell death [Park et al., 2011] [Vanderlaag et al., 2010]. In pancreatic cancer cells, AGR2 is positively correlated with invasiveness potential of the cells, and the precursor of the secreted lysosomal protease cathepsin D is a functional downstream target of this pro-invasive activity [Dumartin et al., 2011], suggesting a link between the secretory pathway, AGR2 and invasiveness of tumours. This was supported by a metastasis screen comparing pancreatic ductal carcinoma tumour tissue and liver hepatocarcinomas, in which the invasiveness potential of AGR2 was reported [Barry et al., 2012]. Furthermore, in androgen-responsive prostate cell lines, increased AGR2 expression was also found to be correlated with increased migratory and invasiveness potential in these cell lines [Bu et al., 2011].

It is clear that AGR2 is involved in the onset and progression of many cancers, and that it has a role to play in metastasis, resistance and aggressiveness of many of these cancers. It is likely that there are different mechanisms and pathways by which AGR2 mediates these effects from cancer type to cancer type, and possibly from subtype to subtype. Whether AGR2's involvement in most of these cancers is related to its functions as a molecular chaperone, or

whether there are other modes of action, for example, mediating cytoplasmic p53 inhibition in the cytoplasm or nucleus, remains to be shown.

Inflammatory Bowel Disease

As well as the association with cancer, AGR2 has been linked to inflammatory bowel disease (IBD). In 1996, a genome-wide association study identified a susceptibility locus at chromosome 7p21.1, a region of the genome that encodes both AGR2 and AGR3 [Satsangi et al., 1996]. Supporting the findings of the 1996 genome-wide association study, a study of single nucleotide polymorphisms (SNPs) in humans found that SNPs in the 5-prime region of AGR2 were associated with increased susceptibility to ulcerative colitis [Zheng et al., 2006]. The study also found that the AGR2 promoter could be regulated by the goblet-cell specific transcription factors FOXA1 and FOXA2. As described above, recent findings have highlighted the importance of AGR2 in goblet cell function and development and in two murine gene knockouts of Agr2 the mice were found to be susceptible to ulcerative colitis, further cementing the link between AGR2 and IBD [Park et al., 2009] [Zhao et al., 2010]. In the second mouse knockout study AGR2 was found to be important for ER homeostasis [Zhao et al., 2010], and a different ER-stress related gene, XBP1, has also been demonstrated to be a genetic risk factor for inflammatory bowel disease [Adolph et al., 2012]. The intestines of patients suffering from some forms of IBD often display marked ER stress [Bogaert et al., 2011] [Niederreiter and Kaser, 2011], suggesting that dysregulation of the ER stress response associated with low levels of AGR2 may pose a secondary pathway for increased susceptibility to IBD.

Asthma

AGR2 has more recently been implicated in the pathophysiology of asthma. The first study to highlight a link between AGR2 and asthma involved a screen for genes overexpressed upon allergen exposure in a mouse model of asthma [Valentin et al., 2009]. In these mice, genome-wide microarray analysis demonstrated that Agr2, along with other mucous-secretion associated genes, was overexpressed, in mice exposed to an allergen (ovalbumin). A further study using the Agr2 mouse knockout line described previously [Park et al., 2009] demonstrated that AGR2 was induced in asthma following allergen exposure, and that mucin overproduction was observed [Schroeder et al., 2012]. In their model, immature lung mucin Muc5AC forms a complex with AGR2, and the knockout mice exhibit a 50% reduction in Muc5AC and Muc5B, togther with an increase in mucins retained in the ER as shown by alcian blue staining. In a separate study in human lung epithelial cells, MUC5AC was found to be induced by interleukin-13 (IL-13), and a reduction in IL-13 was found to lead to a decrease in both AGR2 and MUC5AC levels [Yu et al., 2012], suggesting that IL-13 may be upstream of AGR2 in a particular pathway. This data further confirms the importance of AGR2 in maturation of mucins, and proposes a novel disease model of mucin dysfunction involving AGR2.

1.3 Secretory pathway and trafficking transmembrane proteins to cellular compartments

AGR2 has been shown to be secreted in several types of cancer as discussed above, and studies on homologues in other species also associate AGR2 with the secretory pathway (such as in the chicken oviduct, as a secreted protein in *Xenopus laevis*, and in the regenerating newt limb blastema). The findings that AGR2 is required for maturation of at least two mucins, a class of secretory proteins that mature and are trafficked through the secretory pathway, further implicate AGR2 in this pathway. The secretory pathway is the principal mechanism by which soluble, secreted proteins are trafficked to the cell plasma membrane for exocytosis. From the endoplasmic reticulum (ER) lumen, soluble proteins are first exported from the ER in vesicles, and these vesicles are then trafficked through the golgi network by successive fusion through the golgi compartments. This process is mediated by many proteins, enzymes and co-factors. Likewise, the principal mechanism by which integral, transmembrane proteins mature and arrive at their resident location is also through the secretory pathway (Figure 1.7). Hydrophobic patches are shielded from the aqueous intracellular environment through association with chaperone proteins and vesicular transport through the cytosol by way of the golgi network. Nascent integral membrane proteins that are being synthesized by ribosomes in the ER must be inserted into the ER membrane. For multipass membrane proteins, insertion into the ER membrane occurs either via signal peptides at the N-terminus, or more commonly, the first alpha-helix in the growing peptide that acts as signal-anchor sequence [Lodish et al., 2000].

Figure 1.7: Targeting Proteins to the ER Membrane and Secretory Pathway

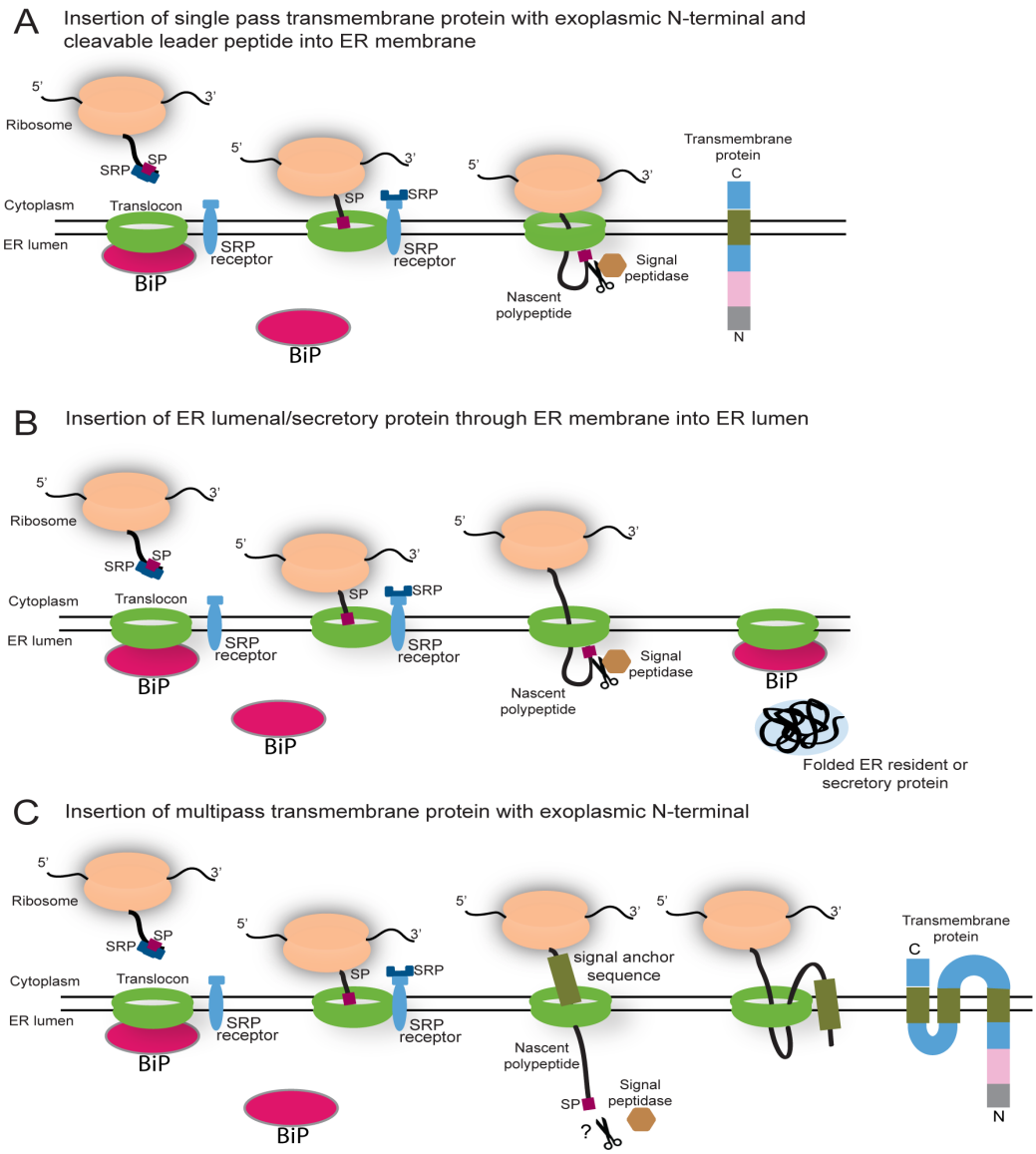


Figure 1.7 Targeting proteins to the ER membrane and secretory pathway
Signal peptide at the N-terminus of nascent polypetide emerging from the ribosome is recognised by signal recognition particle (SRP) and targeted to the translocon resident in the ER membrane. The polypeptide continues to grow until the first transmembrane domain inserts into the ER membrane (A) and the signal peptide is co-translationally cleaved. If no transmembrane domain exists the polypeptide will translocate into the ER lumen (B) where it is folded and becomes resident or packaged into vesicles for export. For many multipass membrane proteins an internal signal anchor sequence similar to the N-terminal signal sequence is recognised by SRP, leading to insertion into the ER membrane of the first alpha-helix. The parameters of the alpha helices determine the topology of the transmembrane protein. A cleavable or non-cleavable N-terminal signal peptide may also exist, targeting the N-terminal to the ER lumen (C).

In the case of the N-terminal signal peptide (also referred to as a leader sequence), a cleavage site allows this hydrophobic, ER-targeting sequence to be removed, and evidence suggests that proper removal of this peptide may be necessary for correct processing and folding of the N-terminal domain of the protein [Gillikin et al., 1997] [Huang et al., 2010]. The signal peptide at the N-terminus is recognised the signal recognition particle (SRP) complex in the ER membrane, which then targets the transmembrane protein to an ER component known as a translocon. The function of the translocon is to allow the hydrophilic portion of the transmembrane protein to pass through the ER membrane into the ER lumen. The lumenal side of the ER membrane maintains its integrity through the action of the BiP chaperone on the lumenal side, which associates with the translocon, moving away when the ribosome that is synthesising the nascent protein associates with the cytoplasmic side of the translocon [Vitale and Denecke, 1999]. Chaperone proteins sometimes assist in the process of transport through the secretory network, aiding in folding and/or assembly of subunits whilst preventing unwanted protein aggregation due to exposure of hydrophobic patches to the aqueous environment [Ellis, 1999] [Ellis, 2006]. In this way, the hydrophobic portions of the nascent polypeptide are shielded from the aqueous environment, whist allowing the hydrophilic portions destined for the cytoplasm or the extracellular environment to pass through to the ER lumen or cytoplasm.

In the first step towards entering the secretory pathway, membrane proteins in the ER membrane or secretory proteins in the ER lumen are packaged into vesicles that are exported to the cis-Golgi network (Figure 1.8).

The cargo-loaded vesicles pass through the trans-Golgi network, shut-

Figure 1.8: The Secretory Pathway

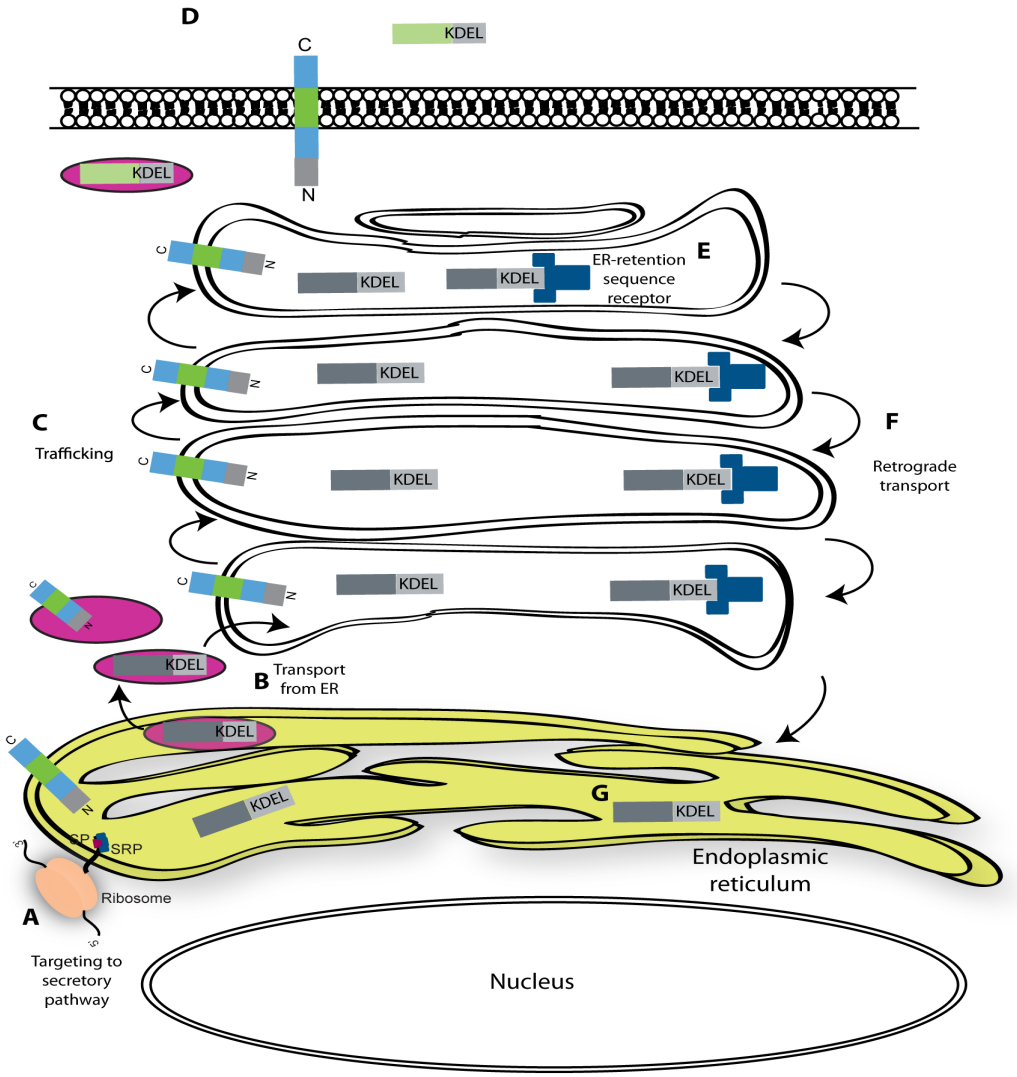


Figure 1.8 Secretory Pathway
Nascent proteins emerging from the ribosome are targeted to the ER membrane by a signal recognition particle (A) where they are integrated into the ER membrane or translocated into the ER lumen. Secretory pathway proteins are then packaged into transport vesicles and exported from the ER to the Golgi network (B). Membrane or secretory vesicles are trafficked through the *cis*-Golgi, medial Golgi and on to the *trans*-Golgi (C), probably through a combination of vesicular transport involving budding and fusing, or through continuous Golgi maturation that progressively moves proteins to the *trans*-Golgi. Transmembrane proteins are inserted into the plasma membrane (D) or secretory vesicles move to the plasma membrane and are released into the extracellular space by fusion with the plasma membrane in a poorly -understood process. Proteins with the KDEL motif are recognised by an ER-retention sequence receptor (E), and trafficked back by retrograde transport (F), where they return to the ER.

tling vesicle-loaded proteins to their destined location at the plasma membrane or extracellular environment. The cargo-loaded secretory vesicles eventually fuse with the plasma membrane, either discharging their cargo into the extracellular environment in the case of soluble secretory proteins, or siting the integral membrane proteins into the plasma membrane with the portion located on the internal side of the vesicle being placed into the extracellular environment [Lodish et al., 2000]. This process is mediated by at least three specialised proteins known as SNARE proteins [Sutton et al., 1998], although the mechanisms of this process are not currently clear, mainly because imaging these processes is extremely difficult due to the rapid nature of the interaction and current imaging resolution limits.

1.4 Aims

Although the link between AGR2 and disease has been well established, little is known about the mechanistic roles of AGR2 in either disease or development. In order to understand how AGR2 contributes to the formation of tumours, mediates metastasis and cytotoxic drug resistance, and how polymorphisms in the gene contribute to IBD, it is important to identify protein networks involving AGR2 (its interactome).

The overall aims of the PhD project are to identify and validate putative interacting proteins for AGR2 from phage peptide display data, and to investigate the biological role of AGR2 in normal tissue and development. These will be achieved using (i) biochemical techniques including co-immunoprecipitation, colocalisation by confocal microscopy and imaging techniques such as total internal reflection microscopy (TIRF-M), (ii) generation and characterisation of antibodies against putative AGR2 binding proteins, (iii) creation of a transgenic mouse line that expresses a GFP-Cre fusion protein from the Agr2 promoter to analyse expression patterns in developing and mature mice and provide a Cre-driver mouse for lineage tracing or cell-type specific co-expression and (iv) characterisation of AGR2 expression in embryonic mouse tissue using antibody-based techniques. Taken together, this *in vivo* and *in vitro* data will begin to define the tissue-specific interactome of AGR2 and help determine novel pathways that mediate normal development as well as onset of disease and disease progression.

Chapter 2

Materials and Methods

2.1 Cell culture

2.1.1 Cell lines and culture conditions

Cell lines were cultured in humidified incubators equipped with CO_2 gas cylinders. Antibiotics were used to maintain cell lines, however, media without antibiotics were used for cells to be transfected. Media were obtained from Invitrogen (GIBCO) and supplemented with fetal calf serum as detailed, and stored at 4°C. Trypsin-EDTA (Invitrogen) was used to suspend cells for splitting and was also stored at 4°C.

The cell lines used are detailed below:

Cell lines

Name	Description	Growth conditions	Media	Max passage number
A375	Human metastatic melanoma	5% CO_2, 37°C	DMEM, 10% FCS, 1% Penicillin-Streptomycin when not transfecting	n+15
(American) MCF7	Human breast adenocarcinoma	5% CO_2, 37°C	DMEM, 10% FCS, 1% Penicillin-Streptomycin when not transfecting	35
(European) MCF7	Human breast adenocarcinoma	5% CO_2, 37°C	DMEM, 10% FCS, 1% Penicillin-Streptomycin when not transfecting	n+15
IMCD3	Mouse SV-40 transformed renal epithelial	5% CO_2, 37°C	DMEM-F12, 10% FCS, 1% Penicillin-Streptomycin when not transfecting	24
IMCD3 (mKate2-Arl13B stable)*	Mouse SV-40 transformed renal epithelial	5% CO_2, 37°C	DMEM-F12, 10% FCS, 1% Penicillin-Streptomycin when not transfecting	24
hTERT-RPE	Human hTERT-immortalised pigmented retinal epithelial	5% CO_2, 37°C	DMEM-F12, 10% FCS, 0.25% sodium bicarbonate, 1% Penicillin-Streptomycin when not transfecting	n+2
NIH-3T3	Mouse immortalised fibroblast	5% CO_2, 37°C	DMEM, 10% FCS, 1% Penicillin-Streptomycin when not transfecting	n+2

Table 2.1: Cell Lines - Description and growth conditions, *kindly provided by Dr E Hall, MRC Human Genetics Unit, Edinburgh

Culture Medium	Description
DMEM:	Dulbecco's Minimal Essential Medium containing sodium pyruvate, L-glutamine, 4500mg/L glucose and phenol red
DMEM-F12:	Dulbecco's Minimal Essential Medium in a 1:1 ratio with Ham's F12 medium containing 15mM HEPES, sodium pyruvate, L-glutamine, 4500mg/L glucose and phenol red
MEM:	Minimal Essential Medium containing L-glutamine and phenol red

2.1.2 Subculturing, storage and recovery of cells

Subculturing

Cell lines were subcultured every 3 days (A375, IMCD3, hTERT-RPE, NIH-3T3), or 5 days (MCF7) in either 10cm dishes or 75cm^2 flasks (Corning) in sterile tissue culture hoods. Cells were washed once in sterile PBS and trypsinised with Trypsin-EDTA (Invitrogen) before incubating at 37°C for 3 min (MCF7, A375 cells) or 5-10 minutes (IMCD3, hTERT-RPE, NIH-3T3 cells). Cells were then resuspended in warm media before splitting into fresh dishes or flasks and incubating as described in Table 2.1 and at a splitting dilution dependent on cell type and experimental requirements. Final volume for subculturing was 10ml (10cm dish) or 25 ml (75cm^2 flask). For experiments in 6 well plates final volume of media before transfection was 2 ml, or 1ml for 12 well plates.

Storage

Cell lines were grown to 95% confluence in 10cm dishes and trypsinised as described above before diluting trypsinised cells in a final volume of 10ml. Cells were pelleted by centrifuging at 1000rpm (Sorvall Legend X1R) for 5 minutes in sterile 15ml falcon tubes and resuspended in fetal calf serum supplemented with 10% DMSO. Cells were aliquoted in 1ml aliquots into cryotubes (Nunc) and placed in gradient freezing containers (Nalgene) for 24 hours at -80°C, before transferring to liquid nitrogen containers for long-term storage.

Recovery

Cells in cryotubes in liquid nitrogen were placed into a water bath at 37°C for 5 minutes to thaw before being transferred into a 25cm^2 flask containing 5ml media (constituents dependent on cell line as described in Table 2.1) at 37°C. Media was replaced after 12-24 hours to remove DMSO and cells were grown to 80-95% confluence before transferring to 75cm^2 flask for culturing as described above.

2.1.3 Transient Transfection

Cells were grown as described above to 80-95% confluence before plating according to cell type and experimental conditions. IMCD3 cells were split 1:4 - 1:6 and incubated for 24 hours at 37°C before transfection. MCF7 cells (American and European) and A375 cells were split 1:3 and incubated for 24 hours at 37°C before transfection. Transfections were carried out with either Attractene Transfection Reagent (Qiagen) or Lipofectamine 2000 (Life Tech-

nologies) as per manufacturers instructions. For MCF7 cells and A375 cells, transfections with Lipofectamine were carried out using Optimem (GIBCO) again as per manufacturers instructions. IMCD3 cells were transfected with Lipofectamine 2000 as follows: Per 6 well plate, 1.6μg DNA combined with 800μl Optimem in one tube, 8μl Lipofectamine combined with 800μl Optimem in another, incubation for 5 minutes at room temperature before combining solutions and incubating for 20 minutes at room temperature. Media in wells was replaced with 2ml fresh media, and transfection complexes added to wells before swirling and incubating at 37°C for 24-48 hours or as stated.

2.2 Microbiological Techniques

All procedures using bacteria were carried out under aseptic condtions. Cloning was carried out in competent DH5α cells or competent Stbl2 DH5α cells. Recombineering was carried out in recombineering-ready SW102 cells (section 2.4.6). Luria-Bertani broth (Table 2.2) was sterilised by autoclaving at 121°C for 20 minutes, selective antibiotics (Table 2.2) were sterilised by filter-sterilising (0.22μm filter). For cultures requiring zeocin as selective antibiotic low-salt LB broth was used containing 0.5% (w/v) NaCl.

2.2.1 Bacterial culture and reagents

For small DNA preparation extraction from bacterial cell culture (minipreps) 5-10 ml LB medium containing the appropriate antibiotic was inoculated with a bacterial colony grown on agar or from glycerol stock and incubated overnight at 37°C with shaking at 225rpm. For larger DNA preparations (maxipreps)

LB broth	Selective Antibiotics
1% Tryptone	Ampicillin (Sigma), 50-100μg/ml
0.5% (w/v) Yeast extract	Kanamycin (Sigma), 10-50μg/ml
1% (w/v) NaCl	Chloramphenicol (Sigma), 25μg/ml
	Spectinomycin (Sigma), 50-100μg/ml
	Zeocin (Invitrogen), 25-50μg/ml *low salt LB

Table 2.2: Bacterial culture medium and antibiotics

5 ml LB medium containing the appropriate antibiotic was inoculated with a bacterial colony grown on agar or from glycerol stock and incubated for 8 hrs or overnight at 37°C with shaking at 225rpm, before transferring 1-2 ml of starter culture to 250ml LB broth containing appropriate antibiotic and incubating at 37°C with shaking at 225rpm for 20-24 hours. Culture flasks were a minimum of 4 times the volume of culture for proper aeration. Glycerol stocks were taken as appropriate by transferring 0.7ml of culture to 0.3ml of glycerol (0.5%) (total glycerol concentration 0.15%) in a sterile cryovial (Nunc) before snap-freezing in liquid nitrogen and storing at -80°C.

Bacterial culture dishes

LB agar
1% Tryptone
0.5% (w/v) Yeast extract
1% (w/v) NaCl
1.5% (w/v) agar powder

LB agar was melted by heating to greater than 60°C and allowed to cool

to lower than 50°C before addition of appropriate antibiotic as described in section 2.2.1. Liquefied agar containing antibiotic was poured into 100mm petri dishes and left in a sterile hood to solidify. Plates were sealed with Parafilm® and stored at 4°C. Prior to use plates were left at room temperature with Parafilm® removed for up to 4 hours.

2.2.2 Preparation of competent *E. coli* cells and transformation

Where competent cells were not bought commercially (Subcloning Efficiency DH5α™Competent Cells, Invitrogen; MAX Efficiency® Stbl2 ™Competent Cells), DH5α cells were made chemically competent as follows:

Competent Cell Buffer 1	Competent Cell Buffer 2
100mM RbCl	10mM MOPS buffer pH 6.8
50mM $MgCl_2$	10mM RbCl
30mM potassium acetate	75mM $CaCl_2$
10mM $CaCl_2$	15% glycerol (v/v)
15% glycerol (v/v)	
pH adjusted to 5.8 with dilute acetic acid	pH adjusted to 6.8 with sodium hydroxide
filter sterilised	filter sterilised

Table 2.3: Buffers for making chemically competent cells

LB broth (5 ml) was inoculated with glycerol stock of DH5α cells and shaken overnight at 37°C. 250 μl of starter culture was used to inoculate 50ml of LB broth in a 250 ml flask and shaken at 37°C until the culture reached an OD (600nm) of 0.4 (log phase of bacterial growth). Cells were pelleted by

centrifuging at 2737xg (4000rpm) (Sorvall Legend X1R) for 15 minutes at 4°C and gently resuspended in ice-cold Competent Cell Buffer 1 (16 ml buffer/50 ml culture) and left on ice for 10 minutes. Cells were pelleted as before and gently resuspended in ice-cold Competent Cell Buffer 2 (2 ml/50 ml culture) and left on ice for a further 10 minutes. Competent cells were aliquoted in the cold room into 50 μl aliquots in pre-chilled tubes and snap-frozen in liquid nitrogen before storage at -80°C. These cells were used for transformation with DNA using the heat shock method (see below).

Transformation of bacterial cells with plasmid DNA

Bacterial cells (commercially bought Subcloning Efficiency DH5α ^{TM}Competent Cells, commercially bought MAX Efficiency® Stbl2 ^{TM}Competent Cells, or DH5α cells made chemically competent as described above) were transformed using the heat shock method. 50-100μl competent cells were thawed on ice and mixed with 0.1-1.0μg of plasmid DNA in microcentrifuge tubes, and incubated on ice for 10 minutes. Competent cells mixed with DNA were placed in a water bath at 42°C for 1-3 minutes and incubated on ice for a further 5 minutes. 800μl LB broth was added to the transformed competent cells and mixture placed at 37°C for 1h with shaking at 225rpm. Cells were pelleted by centrifuging at 3000rpm for 10 minutes and cells were resuspended in 120μl fresh LB before plating onto LB agar plates with appropriate selective antibiotic, and growing at 37°C overnight in an incubator.

2.3 Plasmid and BAC DNA

2.3.1 Purification of plasmid DNA

LB broth was inoculated with DH5α cells transformed with desired plasmid DNA from a colony on an agar plate or from glycerol stock and grown at 37° overnight or as described in section 2.2.1, depending on the size of the DNA plasmid preparation. Bacterial cultures were pelleted by centrifuging at 6000xg in a Sorvall RC6+ Centrifuge (maxiprep) or a Sorvall legend X1R table top centrifuge (miniprep). Plasmid DNA was retrieved and purified from pelleted cells using maxiprep or miniprep kits (Qiagen Plasmid Plus Maxi Kit, or Qiaprep Spin Miniprep Kit) as per manufacturers instructions. DNA was resuspended in distilled, sterile water (sterilised by autoclaving at 121°C for 20 minutes). Plasmid DNA was stored at -20°C, and quantified using a NanoDrop spectrometer.

2.3.2 Propagation and preparation of BAC DNA

An 85 kilobase BAC containing the mouse Agr2 gene and flanking regions (bMQ251g14) was obtained from the 129S7/AB2.2 clone library, shipped in *E. coli* in agar. A toothpick was used to streak the bacteria containing the BAC onto an agar plate containing 25μg/ml chloramphenicol (the selective antibiotic on the BAC backbone) and grown at 37°C overnight. Colonies were picked from the agar plate into 3 ml LB broth containing 25μg/ml chloramphenicol and placed at 37°C overnight with shaking at 225rpm. After making glycerol stock as described above, the remaining 1.6 ml of culture was used to isolate

and purify the BAC DNA.

Commerical buffers from the Qiagen Plasmid Plus Maxi Kit were used for the first stage (Buffer P1 resuspension buffer, Buffer P2 lysis buffer, Buffer P3 neutralisation buffer). For each preparation, 1.6 ml of bacterial culture containing the Agr2 BAC was pelleted in an Eppendorf 5417R microcentrifuge at 5000xg, and the pellet resuspended in 100μl of Buffer P1 before incubating on ice for 5 minutes. 200μl of Buffer P2 was added followed by a further 5 minutes on ice. 200μl of Buffer P3 was added and preparations were again incubated on ice for 5 minutes. Preparations were centrifuged at 10,000xg for 5 minutes, and supernatant containing BAC DNA was transferred to a fresh tube. BAC DNA was precipitated by adding 900μl of ethanol (95%) and incubating on ice for 30 minutes. DNA was pelleted by centrifuging at 10,000xg for 5 minutes, and the glassy pellet was washed with 200μl ethanol (70%). Preparations were centrifuged again at 10,000xg for 5 minutes and pellets were air-dried on the bench for 1 hour. DNA was redissolved in 50μl endotoxin-free TE buffer (Qiagen kit) and DNA concentrations were determined by NanoDrop. Purity of BAC DNA was assessed by NanoDrop (A260/A280 ratio at or greater than 1.8) and by endonuclease test digest with Hin$d$III (20μg BAC DNA, 3μl restriction enzyme, enzyme-specific buffer, total volume 50μl) for 3h at 37°C and then analysing the digest pattern on an agarose gel (0.8%).

2.3.3 Agarose gel electrophoresis for DNA analysis

50XTAE Buffer was diluted in distilled water prior to use to make a 1x solution. Ultrapure agarose (weight as appropriate for percentage of agarose gel) was added to 1x TAE buffer and the solution was boiled in a microwave until

the agarose had dissolved. The solution was allowed to cool until just warm to the touch and ethidium bromide was added (5μl/150 ml). The solution was poured into an electrophoresis gel tank with appropriate comb size and left to solidify. Combs were removed and gel was submerged in 1x TAE buffer. 6x DNA loading buffer was added to samples (1:6 ratio) and samples were loaded into the gel wells. Electrophoresis was carried out at 140V until DNA was adequately separated, followed by visualisation under a u.v. lamp (Biodoc-It Imaging System, UVP) with built-in camera (Fluor Cam 210). For samples requiring recovery of DNA, appropriate bands were excised and recovered using the QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's instructions.

50x TAE (Tris-Acetate EDTA) Running Buffer*	6x DNA Loading Buffer
2M Tris pH 6.8	15% Ficoll 400 (w/v)
0.1M $Na_2EDTA.2H_20$	0.25% Bromophenol blue (w/v)
4% Acetic acid (v/v)	0.25% Xylene cyanol (w/v)
pH adjusted to pH 8.5	
*obtained from Shared Services, MRC HGU	

Table 2.4: TAE and DNA Loading Buffers

2.4 Cloning methods

Primers I

Name	Sequence
MKS3wt-GW-nV5-F	ggggacaagtttgtacaaaaaagcaggctcaatggcggtttggtccctctta
MKS3wt-GW-nV5-R	ggggaccactttgtacaagaaagctgggtactattaaatcaaaaatctttg
MKS3-Y385A-F	gactttcccactcctatatttgctgatgtgtacc
MKS3-Y385A-R	ggtacacatcagcaaatataggagtgggaaagtc
MKS3-P382T-F384Y-F	gactttcccactactatatattatgatgtgtaccttg
MKS3-P382T-F384Y-R	caaggtacacatcataatatatagtagtgggaaagtc
5prime-F-Agr2BAC	cagcctgtggtgtgggccat
5prime-R-Agr2GCIP	ttgctcagggcggactgggt
3prime-F-Agr2GCIP	ttacggtatcgccgctcccg
3prime-R-Agr2BAC	gagaagccactggggtgcatgt
amp-F	gcttatcgatgataagctgtcaaacatgagaattgatccggaacccttaatctta caatttaggtggcact
amp-R	tccgatgcaagtgtgtcgctgtcgacggtgaccctatagtcgagggacctaat atgagtaaacttggtctga
loxPtoAmptestF	gtgccgaggatgacgatgagc
loxPtoAmptestR	ccgtgccggcacgttaacc
5primeMCS/TOP	tcgaagatctcatatggtttaaacacgcgtgctagca
5primeMCS/BOTTOM	tcgatgctagcacgcgtgtttaaaccatatgagatct
3primeMCS/TOP	ggccttaattaaatttaaatggcgcgcc
3primeMCS/BOTTOM	ggccggcgcgccatttaaatttaattaa
Agr2HomArm-5prime/F	cgcgctcgagttctatcttcatctacatatg
Agr2HomArm-5prime/R	gcgcgtcgacggcaaatctgcatgggaaa
Agr2HomArm-3prime/F	gcgcttaattaagagaaattttcagtgtctgc
Agr2HomArm-3prime/R	tataggcgcgccgtcagatgagctgatcg

Table 2.5: Primers I

Primers II

Name	Sequence
eGFPCre F	ggggacaagtttgtacaaaaaagcaggcttaaccatggtgagcaagggcgag
eGFPCre R	ggggacaacttttgtatacaaagttgtctaatcgccatcttccag
mCherry F	ggggacaagtttgtacaaaaaagcaggcttaatggtgagcaagggcgag
mCherry R	ggggacaacttttgtatacaaagttgtctacttgtacagctcgtccatg
IRES F	ggggacaactttgtatacaaaagttgtggttattttccaccatattgccgtc
IRES R	ggggacaactttgtatagaaaagttgggtgtattatcatcgtgtttttt
puroR	ggggacaacttttctatacaaagttgctaccatgaccgagtacaagccc
puroR	ggggacaactttattatacaaagttgttcaggcaccgggcttgcg
pA F	ggggacaactttgtataataaagttgctctgtgccttctagttgccagcc
pA R	ggggaccactttgtacaagaaagctgggtaatagagcccaccgcatcc

Table 2.6: Primers II

2.4.1 Gateway cloning

Gateway cloning is a system from Invitrogen which enables quick and efficient cassette exchange without the need for restriction endonucleases. The principle of Gateway cloning makes use of bacteriophage λ recombinase enzymes to mediate the exchange of DNA cassettes between recognition sites in *E. coli* [Hartley et al., 2000]. Recombination between specific sites is one-directional, so that an attB site on one DNA segment and an attP site on another are recombined to yield new sites, attL and attR. Conversely, recombination between attL and attR sites yield attB and attP sites. The directionality is recognised by two different recombinase enzymes now marketed by Invitrogen, known as BP Clonase and LR Clonase enzymes. DNA fragments, usually from PCR products, are amplified containing attB sites flanking the

DNA sequence to be cloned, and a BP reaction using BP clonase causes recombination into attP sites in a pDONR vector. For selection, the *E.coli* lethal *ccdB* gene is used. Correct recombination of the PCR fragment into the *ccdB* site removes the lethal gene and allows positive clone selection. Once the desired DNA fragment has been successfully recombined into the pDONR vector, the vector, now termed pENTR and with attL and attR sites, is combined with a destination vector containing attB and attP sites (pDEST) in a reverse recombination event using LR Clonase enzyme. Correct clones are selected for with antibiotic.

Like single-fragment Gateway recombination, Multisite Gateway ® recombination can be carried out in two recombinase-mediated steps, however, it allows specific cassette exchange for up to 4 different fragments. Multisite Gateway pDONR vectors contain unique *attP* sites which confer substrate specificity for the BP clonase enzyme. By using the unique sites only the fragments with correct flanking sites can recombine (see Figure 3.1). Thus, these sites are not interchangeable [Cheo et al., 2004]. This method will be discussed further in section 2.4.4 and in Chapter 3.

2.4.2 Cloning of V5-tagged wild type MKS3

The human *Meckelin* (MKS3) DNA sequence (longest known isoform, 995 amino acids) was amplified from pCMV-HA-MKS3 (generously donated by Prof C Johnson, Leeds University) with Gateway-compatible primers MKS3wt-GW-nV5-F and MKS3wt-GW-nV5-R (Table 2.5) using the conditions described below:

Amplification of wild type Meckelin	
Component	**Final Concentration**
10x Platinum Taq buffer	1x
10mM dNTP	0.2mM
50mM MgCl	1.5mM
10μM Forward primer	0.4μM
10μM Reverse primer	0.4μM
Platimum Taq 4U/μl	2U
pCMV-HA-MKS3 (800ng/μl)	800ng
	Total volume 50μl

PCR protocol to amplify wild type Meckelin

step 1:	95°C	2:00 min
step 2:	95°C	30 sec
step 3:	60°C	45 sec
step 4:	72°C	3:00 min
step 5:	go to step 1 34x	
step 6:	72°C	5:00 min
step 7:	4°C	for ever

PCR products were loaded onto a 1% agarose gel and DNA was separated by electrophoresis. The appropriate band corresponding to Meckelin was excised and gel purified using the Gel Extraction Kit from Qiagen described previously and eluted in 30μl. 7μl of the PCR products were incubated with pDONR221 (215ng) and 2μl of BP Clonase enzyme mix at room temperature for 1h. Proteinase K was added to inhibit the enzyme (10 minutes at 37°C), and mixture was transformed into DH5α cells as described above before plating on agar containing kanamycin (50μg/ml). Clones were picked off agar plates and grown to make minipreps of the DNA. These were sequenced and correct pENTR-MKS3 clones identified.

In the second step, pENTR-MKS3 DNA (300ng) was incubated with pcDNA^{TM}3.1/nV5-DEST (100ng) and 2μl of LR Clonase enzyme mix for 1h at room temperature. Proteinase K was added to inhibit the enzyme (10 minutes at 37°C), and the mixture was transformed into DH5α cells as described in section 2.2.2 before plating on agar containing ampicillin (100μg/ml). Clones were picked off agar plates into LB broth containing ampicillin, grown to make minipreps and verified by sequencing.

2.4.3 Site-directed Mutagenesis

Mutants of the wild-type V5-tagged MKS3 vector described above were made using the QuikChange ^{TM}site-directed mutagenesis kit from Stratagene (www.chem.agilent.com). The reason for choosing these mutations will be further discussed in subsequent chapters. Primers used to generate mutations are listed in Table 2.5; MKS3-Y385A-F, MKS3-Y385A-R, MKS3-P382T-F384Y-F, MKS3-P382T-F384Y-R. In the first step, the forward and reverse primers containing desired mutation(s) are annealed to the two denatured strands of parental (vector) wild type DNA. A PCR cycle with a long elongation phase allows the entire vector sequence to be amplified with the desired mutation(s). Repeating the cycles enriches for the mutated DNA. DpnI digest then removes parental DNA by recognising *E coli* DNA methylation on the parental strands. The mixture is then transformed into bacteria and plated on agar containing appropriate antibiotic, and clones are picked off the agar plates into LB broth with antibiotic for growth and DNA miniprep as described previously. Using primers with desired mutations listed above, mutations were amplified as described below:

Amplification of mutant Meckelin isoforms	
Component	**Final Concentration**
10x Stratagene Pfu Turbo buffer	1x
10mM dNTP	0.5mM
10μM Forward primer	0.4μM
10μM Reverse primer	0.4μM
Pfu Turbo Polymerase 2.5U/μl	2.5U
pENTR-MKS3 (50ng/μl)	50ng
	Total volume 50μl

PCR protocol for site-directed mutagenesis

step 1:	95°C	30 sec
step 2:	95°C	30 sec
step 3:	55°C	1 min
step 4:	68°C	7:30 min
step 5:	go to step 1 15x	
step 6:	4°C	for ever

PCR reactions were then DpnI-digested (10U/reaction) for 2h at 37°C before transforming into chemically competent DH5α cells as previously described, and plating onto agar containing kanamycin (50μg/ml). Colonies were picked into LB broth with antibiotic and minipreps were made also as previously described. Correct pENTR clones were identified by sequencing. These pENTR clones were recombined into pcDNA^{TM}3.1/nV5-DEST by Gateway cloning as described in section 2.4.2 and sequence verified.

2.4.4 Construction of universal expression construct pMULTIrec

Fragments for four different cDNA sequences (GFP-Cre, IRES, puromycin, polyA) were amplified by PCR using primers as shown in Table 2.6. Each fragment was amplified with specific multi-site Gateway recombinase sites (*attB*) at the 5' and 3' ends. These fragments were purified by gel extraction as described in section 2.3.3 before recombining into pDONR vectors to generate Multisite-Gateway specific pENTR clones as described above. Clones were sequence verified and are listed in Table 2.7. Using these unique entry clones, an LR reaction was performed to bring all four fragments together, in order, in a single reaction. The conditions for 4-fragment Multisite Gateway ® are summarised below. The reaction mixture was incubated at room temperature for 18h before addition of Proteinase K for 10 minutes at 37°C to inactivate the enzyme. Mixtures were then transformed into MachI *E. coli* cells supplied with Multisite Gateway Pro kit according to manufacturers instructions and plated onto agar plates containing ampicillin (50μg/ml). Colonies were picked and grown for DNA miniprep and analysed by analytical restriction endonuclease digests. The correct clone was termed pcDNA6.2-eGFPCre-IRES-puroR-pA.

Because the antibiotic resistance of both vector pcDNA6.2-eGFPCre-IRES-puroR-pA and our final destination vector loxP-F3-neo-F3-DEST were kanamycin, we needed to shuttle the eGFPCre-IRES-puroR-pA cassette through a series of BP reactions and LR reactions to enable positive selection. The four-fragment cassette in pcDNA6.2 (117.5ng) was incubated with pDONR/Zeo from Invitrogen (100ng) and BP Clonase mix (2μl) for 1h at room tempera-

Multisite Gateway 4-fragment reaction	
Component	**Final Concentration**
pENTR-GFP-Cre	10 fmole
pENTR-IRES	10 fmole
pENTR-puroR	10 fmole
pENTR-pA	10 fmole
pcDNA6.2/V5-pL-DEST	20 fmole
LRClonase II Plus	2μl
	Total volume 50μl

ture in a standard Gateway reaction as described above. Because of the enzyme specificity and location of attB sites the entire 4-cassette block was recombined into pDONR/Zeo, yielding pENTR/Zeo-eGFPCre-IRES-puroR-pA. Reaction mixtures were transformed into DH5α cells and selected on low-salt agar containing zeocin (50μg/ml). Clones were picked into LB broth with zeocin for DNA miniprep, and correct clones were identified by analytical restriction endonuclease digest.

The final LR reaction into the destination vector could now be carried out by performing an LR reaction between the pENTR/Zeo-eGFPCre-IRES-puroR-pA vector (with attL sites now flanking the entire 4-cassette block) and the final destination vector loxP-F3-neo-F3-DEST (with attR sites) with kanamycin resistance. 150ng of pENTR/Zeo-eGFPCre-IRES-puroR-pA was incubated with 150ng loxP-F3-new-F3-DEST and 2μl of LR Clonase II for 2h at room temperature, before addition of Proteinase K for 10 minutes at 37°C to inactivate the enzyme. Mixtures were transformed into DH5α and plated on agar containing kanamycin (50μg/ml). Clones were picked into LB broth with kanamycin and grown for DNA miniprep as described. Correct clones were

identified by analytical restriction endonuclease digest and sequence verified.

Vectors used in construction of pMULTIrec	vector size (bp)
pENTR-eGFPCre/L1-R5	4800
pENTR-IRES/L5-L4	4350
pENTR-puroR/R4-R3	4500
pENTR-pA/L3-L2	4100
pcDNA6.2/V5-pL-DEST	6693
pcDNA6.2-eGFPCre-IRES-puroR-pA	7686
pDONR/Zeo	4291
pENTR/Zeo-eGFPCre-IRES-puroR-pA	5283
loxP-F3-neo-F3-DEST	
GFPCre-IRES-puroR-loxP-F3-neo-F3-DEST	8363
pMULTIrec	8428

Table 2.7: Vectors used in construction of pMULTIrec

Recombineering into a BAC uses homologous arms, regions of 50-400 nucleotides flanking the sequence to be recombineered. In order to make this construct easy to adapt to any BAC, multiple cloning sites were inserted at the 5' and 3' ends of the 4-cassette expression and neomycin resistance cassettes. A multiple cloning site comprising PacI, SwaI, AscI was inserted into the NotI site at the 3' side of the expression cassettes. Oligonucleotides (Table 2.5) corresponding to these sites were annealed by combining 50μl of each oligo (at 100μM) and incubating for 10 minutes at 95°C. Oligos were then allowed to cool gently to room temperature to form double-stranded DNA segments with overhangs for ligation. These were diluted 1:10. The GFPCre-IRES-puroR-loxP-F3-neo-F3-DEST vector was digested with NotI as shown below

and incubated for 1h at 37°C before inactivating the enzyme at 65°C for 10 minutes.

Restriction endonuclease digest of

GFPCre-IRES-puroR-loxP-F3-neo-F3-DEST

Component	**Final Concentration**
GFPCre-IRES-puroR-loxP-F3-neo-F3-DEST	117ng
NotI	1μl
10x NEB Buffer 3	1x
1000x BSA	1x
	Total volume 10μl

The annealed oligonucleotides with 3' overhangs were ligated into the the vector as shown below and incubated overnight at 16°C. Ligation reactions were purified using the QIAquick PCR purification kit (Qiagen) and then redigested with NotI as shown in the table above to remove any self-ligated vectors. After inactivating the enzyme the reactions were transformed into DH5α cells and plated on kanamycin. Correctly ligated clones (in the correct orientation) were identified by sequencing.

Ligation reaction - annealed oligonucleotides

Component	**Final Concentration**
T4 ligase	1.5U
10x T4 Buffer	1x
Annealed oligs	1μ
NotI-digested vector	10ng
	Total volume 20μl

The process was repeated at the 5' end exactly as described for the 3'

multiple cloning site insertion. The XhoI site at the 5' side of the expression cassettes was used for insertion of the new multiple cloning site. Oligonucleotides used for annealing and ligation are listed in Table 2.5. The new multiple cloning site at the 5' end consisted of BglII-NdeI-PmeI-MluI-NheI. Clones with correct insertion of the new multiple cloning site were sequence verified. The insertion of this site, however, caused the entire vector to become unstable and undergo spontaneous rearrangements (the reason for this is not known, although high sequence similiarity with a region of the backbone may have been responsible). For this reason two of the sites were swapped by Dr K Singh-Dolt to give the following 5' multiple cloning site sequence: NdeI-BglII-PmeI-MluI-NheI. We named this universal expression construct, with multiple cloning sites for ligation of BAC homologous arms, pMULTIrec. A list of all the vectors used to make the universal expression construct can be found in Table 2.7.

In addition, various versions of pMULTIrec with different expression cassettes were made, including replacing the eGFPCre cassette with mCherry. The method is explained in Chapter 3. The first of the four expression cassettes was replaced with the *ccdB E. coli* lethal gene by carrying out a Gateway BP reaction with pMULTIrec and pDONR-P1-P5r (this was done by Dr K Singh-Dolt) to make pDEST-IRES-puroR-pA MULTIrec. In parallel, the mCherry sequence was amplified from vector pCAGGs-fucci using the Gateway primers listed in Table 2.6, and the PCR product was placed into pDONR-P1-P5r by a Gateway BP reaction to make pENTR-mCherry-L1-R5. Combining the two vectors pDEST-IRES-puroR-pA-MULTIrec with pENTR-mCherry-L1-R5 in a Gateway LR reaction provided the new expression vector with mCherry in

place of GFP-Cre.

2.4.5 Creation of pMULTIrec-Agr2

Homologous arms corresponding to the 242bp immediately upstream of the mouse *Agr2* start codon and 250bp immediately downstream of the mouse *Agr2* start codon were amplified by PCR using the primers listed in Table 2.5 and using the conditions and PCR protocol shown below.

Amplification of Agr2 3' and 5' homologous arms	
Component	**Final Concentration**
10x Platinum Taq buffer	1x
10mM dNTP	0.2mM
50mM MgCl	1.5mM
10μM Forward primer	0.4μM
10μM Reverse primer	0.4μM
Platimum Taq 4U/μl	2U
Agr2 BAC miniprep	5μg
	Total volume 100μl

PCR protocol to amplify Agr2 homologous arms

step 1:	95°C	2:00 min
step 2:	95°C	1 minute
step 3:	58°C	30 sec
step 4:	72°C	30 sec
step 5:	go to step 1 34x	
step 6:	72°C	5:00 min
step 7:	4°C	for ever

For the 5' homologous arm the pMULTIrec vector was digested with NdeI and SalI before ligating the amplified homologous arm. For the 3' homologous

arm pMULTIrec (with 5' Agr2 homologous arm inserted) was digested with PacI and AscI before ligating the amplified homologous arm. All double digests were carried out using enzymes from NEB (New England Biolabs) and according to standard protocols. These were then ligated as previously described using T4 ligase and buffer, 10ng of digested vector and a range of dilutions of amplified homologous arms. Correct insertion of homologous arms was verified by analytical PCR (see Figure 3.7).

In addition, homologous arms corresponding the 400bp immediately upstream of the mouse *Meckelin* start codon and 400bp immediately downstream of the mouse *Meckelin* start codon were amplified by PCR and cloned into the same restriction sites on pMULTI-rec-mCherry, and verified by analytical PCR, however, this targeting expression construct has not yet been recombineered into a Meckelin BAC to create a Meckelin mCherry promoter reporter.

2.4.6 Recombineering

SW102 cells are bacterial cells with the recombination machinery integrated into their genome. This machinery consists of three temperature sensitive λ Red genes: *exo*, *bet* and *gam*. These genes encode an exonuclease (*exo*), a pairing protein that mediates homologous recombination with complementary DNA (*bet*), and an inhibitor of *E. coli* exonuclease RecBCD to protect the DNA targeting cassette from degradation (*gam*). In this system, the recombineering genes are not expressed when the bacteria are kept at 32°C, but when the bacteria are incubated for 15 minutes at 42°C, the genes are induced and homologous recombination can occur [Warming et al., 2005].

Introducing BAC into SW102 cells by electroporation

SW102 cells were grown in 5 ml LB broth at 32°C overnight. This starter culture was used to inoculate 40 ml of fresh LB broth in a 250 ml flask and grown at 32°C for one hour. Cells were pelleted by centrifuging (in a 50 ml falcon tube) for 10 minutes at 2737xg (4000rpm) in a Sorvall Legend X1R centrifuge. The pellet was washed in 20 ml ice-cold water and centrifuged at 2737xg (4000rpm) for a further 10 minutes, at 4°C. The pellet was resuspended in 1 ml ice-cold water, transferred to a microcentrifuge tube and centrifuged at 10,000xg for 30 seconds at 4°C. This last step was repeated once more and pellet was resuspended in ice-cold water to a final volume of 200μl. This was divided into 50μl aliquots and 100ng of BAC DNA (Agr2, bMQ251g14) was added to each electrocompetent SW102 aliquot, retaining one without DNA as a negative control. DNA-SW102 mixtures were left on ice for 5 minutes and then placed into pre-chilled cuvettes for electroporation (1.8kV, 25μF capacitance, 200ω resistance). After electroporation, cells were placed in 15 ml falcon tubes with 1 ml LB broth and incubated at 32°C for one hour with shaking at 225rpm. Cells were then pelleted by centrifuging at 4000xg and resuspended in 200μl of fresh LB broth before plating onto fresh agar plates containing chloramphenicol (25μg/ml). Plates were incubated at 32°C overnight. Introduction of BAC DNA into the recombineering cells was confirmed by culture, miniprep as described in section 2.3.2 (retaining a portion of the culture for glycerol stock), and analytical digest as described also in section 2.3.2.

Recombineering of target DNA into Agr2 BAC

SW102 cells harbouring the Agr2 BAC (bMQ251g14) as described above were picked from glycerol stock into 10 ml LB broth with chloramphenicol (25μg/ml) and grown overnight at 32°C. 1 ml of overnight culture was then added to 50 ml LB broth with chloramphenicol (25μg/ml) and grown for 1 hour at 32°C. The 50 ml culture was then divided into 2x 25 ml in conical flasks, and one flask was placed at 42°C with swirling for 15 minutes (to induce the Red genes) while the other flask remained at 32°C (uninduced control). After induction both flasks were immediately plunged into an ice-cold water slurry for 10 minutes, with swirling to cool the cultures. Cultures were transferred to two pre-chilled 50 ml falcon tubes and cells pelleted by centrifuging at 2737 x g (4000rpm) for 10 minutes at 4°C. The soft pellet was gently resuspended first in 1 ml ice-cold water and then a further 30 ml ice-cold water was added. Cells were pelleted as before and soft pellet resuspended in 1 ml ice-cold water before transferring to fresh microcentrifuge tubes. Cells were centrifuged at 10,000xg for 30 seconds at 4°C, supernatant removed, and cells resuspended in 1 ml ice-cold water. Cells were pelleted once more at 10,000xg for 30 seconds at 4°C before resuspending in 200μl ice-cold water. On ice, 100ng salt-free targeting cassette DNA (pMULTIrec-Agr2; section 2.4.5) was combined with 50μl electrocompetent induced or uninduced cells (and one induced aliquot without DNA for negative control) and transferred immediately to the pre-chilled microcuvettes before electroporating as described above. After electroporation, 1 ml LB was added and cells were incubated at 32°C for 2h. Cells were then pelleted at 4000xg and resuspended in 200μl fresh LB before plating on agar containing kanamycin (10μg/ml; selective antibiotic for targeting DNA cassette). Correct

clones were identified by picking colonies from the plates, making minipreps as described in section 2.3.2 (retaining a portion of each culture as glycerol stock also as described in section 2.2.1), and analysing by (i) Hin*d*III digest pattern comparison and (ii) PCR analysis. Primers for correct recombineering of the DNA targeting cassette at the correct location (start codon) on the Agr2 BAC are listed in Table 2.5 (5prime-F-Agr2BAC, 5prime-R-Agr2GCIP; 3prime-F-Agr2GCIP; 3prime-R-Agr2BAC). The primers were chosen to amplify 5' and 3' regions at the cassette integration site, and each primer pair was designed to amplify a product that spanned the integration site (from BAC backbone and from DNA targeting cassette). Analytical gels are shown in Chapter 3, Figure 3.7.

Removal of loxP site

A *loxP* site in the backbone of the Agr2 BAC required removal in order not to interfere with the targeting cassette that also contains a *loxP* site (for futher discussion see Chapter 3). Primers were designed (Dr K Singh Dolt) to amplify the ampicillin resistance gene (amp-F, amp-R; Table 2.5) from vector pROSA and with 50bp homologous arms identical to the region directly upstream and downstream of the *loxP* site on the backbone of the Agr2 BAC. The conditions and PCR protocol are listed below. The ampicillin resistance gene was recombineered into the Agr2 BAC-pMULTIrec as described above, replacing the *loxP* sequence. After electroporation cells were plated on agar containing ampicillin (50μg/ml) and kanamycin (10μg/ml) for selection. Correct removal of the *loxP* site and replacement with the ampicillin gene was confirmed by PCR analysis using primers that amplified the region from outside the ampi-

cillin cassette. Amplification of the loxP site yielded a 200bp product whilst amplification of the ampicillin sequence yielded a 1200bp site (see Chapter 3 Figure 3.7 for analytical PCRs).

Amplification of ampicillin resistance cassette	
Component	**Final Concentration**
10x Platinum Taq buffer	1x
10mM dNTP	0.2mM
50mM MgCl	1.5mM
10μM Forward primer	0.4μM
10μM Reverse primer	0.4μM
Platimum Taq 4U/μl	2U
pROSA (100ng/μl)	100ng
	Total volume 50μl

PCR protocol to amplify ampicillin resistance cassette

step 1:	95°C	2:00 min
step 2:	95°C	30 sec
step 3:	56°C	30 sec
step 4:	72°C	1:00 min
step 5:	go to step 1 34x	
step 6:	72°C	5:00 min
step 7:	4°C	for ever

2.5 Protein methods

2.5.1 SDS-PAGE and Immunoblotting

Preparation of cell lysate

Cell lysates were prepared by harvesting cells into an appropriate volume of PBS (generally 1 ml) and centrifuging for 10 minutes at 4000xg. Pellets were resuspended in an appropriate volume of lysis buffer (0.2 - 1.0 ml) and incubated on ice for 30-40 minutes. Lysis buffers used are listed in table 2.8.

Cell Lysis Buffers

Urea Lysis Buffer	NP40 Lysis Buffer
7M Urea	
0.1M DTT	10mMDTT
0.05% TritonX100	0.1%NP40 (Igepal)
25mM NaCl	40mM KCl
20mM HEPES.KOHpH7.6	25mM HEPES.KOHpH7.6
1x protease inhibitor mix (Roche)	1x protease inhibitor mix (Roche)

Table 2.8: Cell Lysis Buffers

After incubation in lysis buffer, lysate was centrifuged at 13,000rpm in a table top centrifuge for 30 minutes at 4°C to pellet cellular debris. The supernatant was collected and either used for analysis or assay, or snap-frozen in liquid nitrogen before storing at -80°C. Prior to analysis, protein content of the lysates was quantified either by Bradford assay or using NanoDrop spectroscopic quantification.

Bradford assays were carried out using Bradford Reagent from Biorad.

The reagent was diluted 1:5 in distilled water and filtered (0.25μm filter). 200μl of diluted, filtered Bradford reagent was added in transparent 96-well plates to 2μl of sample lysate or standard preparations of bovine serum albumin (BSA) in a range of concentrations. There is a linear relationship between protein concentration and the absorbance of a sample with Bradford reagent at 595nm. Samples were mixed well and absorbance of each sample was measured with a spectrophotometer at 595nm (Powerwave XS Microplate spectrophotometer, Biotek). Concentrations were determined by comparison to a standard curve provided by the BSA samples.

Polyacrylamide gel preparation

Pre-cast polyacrylamide gels were used for one assay only (co-immunoprecipitation of AGR2 and MKS3 in MCF7 (American) cells, Figure 4.8. For all other experiments polyacrylamide gels were prepared in the laboratory. For analysis of proteins of high molecular weight 10% polyacrylamide gels were used, whilst for lower molecular weight proteins 12-15% gels were used (Table 2.9). Separating gels were either poured between 1.0mm-gap glass plates or between 1.5mm-gap glass plates dependent on the volume required for loading and allowed to polymerise at room temperature, and stacking gels were added afterwards together with either 10-well or 15 well combs dependent on assay and number of samples. After polymerisation, combs were removed and glass plates containing gels were placed into tanks (Biorad) filled with SDS-Running buffer (Table 2.10). SDS sample buffer was added to the samples to be loaded onto the gel and boiled at 95°C for 3-5 minutes before loading onto the gels. 5μl of pre-stained protein standards were loaded as marker (Fermentas, broad range

prestained protein marker), and proteins were separated by electrophoresis at 140V for 45-90 minutes.

10% Separating Gel	12% Separating Gel	15% Separating Gel	Stacking Gel
10% acrylamide	12% acrylamide	15% acrylamide	5% acrylamide
375mM Tris pH 8.8	375mM Tris pH 8.8	375mM Tris pH 8.8	125mM Tris pH 6.8
0.1% SDS (w/v)	0.1% SDS (w/v)	0.1% SDS (w/v)	0.1% SDS (w/v)
0.1% APS (w/v)	0.1% APS (w/v)	0.1% APS	0.1% APS (w/v)
0.04%TEMED (v/v)	0.04%TEMED (v/v)	0.04%TEMED (v/v)	0.1%TEMED (w/v)

Table 2.9: Composition of Polyacrylamide Gels

SDS-Running Buffer	5x SDS Sample Buffer
192mM Glycine	250mM Tris pH 6.8
25mM Tris	50% glycerol (v/v)
0.1% SDS (w/v)	5% SDS (w/v)
	0.025% bromophenol blue
	200mM DTT (added prior to use)

Table 2.10: SDS-Running Buffer and Sample Buffer

Immunoblotting

After separating proteins by electrophoresis as described above, proteins were transferred to Hybond-C nitrocellulose membranes (Amersham) in tanks filled with transfer buffer (tanks and equipment, Biorad). Briefly, nitrocellulose membranes were soaked in methanol for 1 minute, washed for 10 minutes in distilled water, and soaked in transfer buffer to equilibrate for at least 10 minutes. Gels were removed from glass plates and placed over the pre-soaked

nitrocellulose membrane, and this was placed between 2x 2 card papers also soaked in transfer buffer (Table 2.12). This was then placed in tank holders filled with 1x transfer buffer and transfer was carried out at 300mA for 1 hour on ice, or 30mA overnight. Transferred membranes were removed and rinsed in TBS buffer once, before staining with either ink or Ponceau stain to confirm adequate transfer of protein. Membranes were then washed on an orbital shaker in TBS-Tween-20 (0.1%) (Table 2.12) unless otherwise stated (3x5 minutes) and blocked for 1h in TBS-Tween 0.1% with 5% milk (dried skimmed, Marvel) again unless otherwise stated (Table 2.12). Primary antibodies were diluted in blocking buffer as required (dilution dependent on antibody, Table 2.12), and incubated overnight at 4°C. Blots were then washed as above (3x 5 minutes) and incubated with appropriate secondary antibody diluted in blocking buffer for 1h at room temperature. Blots were washed again (3x 5 minutes) in wash buffer. Blots were then visualised by ECL (enhanced chemiluminescence) using the buffers listed below mixed prior to use in a 1:1 ratio and poured onto the membranes for 1 minute. ECL solution was removed using tissue paper and blots were exposed to film (Kodak, Hyperfilm) to detect chemiluminescence.

List of Primary Antibodies for Protein Biochemistry

Name	Raised in	Dilution	Company	Product number
Anti-AGR2 polyclonal	Rabbit	1:500-1:1000	Moravian Biotech	K47
Anti-AGR2 monoclonal	Mouse	1:1000	Abnova	H00010551-M03
Anti-β-tubulin	Rabbit	1:1000	Cell Signalling Technologies	2128P
Anti-Fibrillarin	Rabbit	1:1000	Cell Signalling Technologies	2639P
Anti-GFP monoclonal	Mouse	1:1000	Abcam	ab1218
Anti-HA monoclonal	Mouse	1:1000	Abcam	ab18181
Anti-MKS3 polyclonal	Rabbit	1:1000	Abcam	ab76786
Anti-MKS3 polyclonal	Rabbit	1:500	Santa-Cruz Biotechnology	sc-87298
Anti-MKS3 (C-terminal) polyclonal	Rabbit	1:100	C Johnson, University of Leeds	N/A
Anti-MKS3 (N-terminal) polyclonal	Rabbit	1:100	C Johnson, University of Leeds	N/A
Anti-V5 epitope monoclonal	Mouse	1:1000	Abcam	ab27672

Table 2.11: List of Primary Antibodies for Protein Biochemistry

Transfer Buffer	TBS-Tween20 wash buffer	Blocking Buffer
192mM glycine	1xTBS	1xTBS
25mM Tris	0.1% Tween20 (v/v)	0.1% Tween20 (v/v)
20% Methanol (v/v)		5% milk powder (w/v)

Table 2.12: Immunoblotting Buffers

2.5.2 Co-immunoprecipitation

Transfection and lysis

Human A375 metastatic melanoma cells grown in 10cm dishes were transfected using Attractene as described in section 2.1.3 with HA-tagged MKS3 and either GFP-AGR2 or GFP vector, before incubating for 24h. Transfected cells were

ECL solution I	ECL solution II
100mM Tris pH 8.5	100mM Tris pH 8.5
2.5mM Luminol	0.02% (v/v) hydrogen peroxide
0.4mM coumaric acid	

ECL solution for chemiluminescence - mixed 1:1

subjected to u.v. at 50J/cm2, or not treated as indicated and incubated at 37°C for a further 24h. Human MCF7 (American) breast adenocarcinoma cells grown in 10cm dishes were transfected with HA-MKS3, and GFP-tagged full length AGR2, RFP-tagged mature AGR2 or GFP vector, before incubating for 48h.

Cells were harvested into 1ml PBS and pelleted by centrifuging at 4000xg for 10 minutes at 4°C. Pellets were resuspended in 1 ml NP40 lysis buffer and incubated on ice for 30 minutes before centrifuging at 13,000xg for 30 minutes at 4°C. Lysates were transferred to fresh tubes. After removing an aliquot as a pre-immunoprecipitation sample, lysates were first pre-cleared to reduce unspecific binding. 100μl of sephadex CL 4B resin was washed 3x in ice-cold lysis buffer (1 ml, 5000xg, 4°C), and added to 1 ml of lysate. Mixtures were incubated for 40 minutes on a rotator at 4°C before centrifuging for 2 minutes at 13,000xg at 4°C. Supernatants were transferred to fresh tubes and protein concentrations were determined by Bradford assay.

Pre-cleared lysates (200-800μg) were then incubated with anti-GFP antibody (2μg, abcam, 1218) on a rotator overnight at 4°C to pull down GFP or RFP-tagged AGR2. Protein G-Sepharose beads were washed 3x in ice-cold lysis buffer (1 ml, 5000xg, 4°C), and antibody-antigen complexes were

then incubated with 20μl of Protein G-Sepharose beads for one hour at room temperature. Beads with bound antibody-antigen complexes were collected by centrifuging at 5000xg for 2 minutes at 4°C. Unbound proteins were removed by washing 4x in 0.5 ml ice-cold lysis buffer (5000xg, 2 minutes, 4°C). Antibody-antigen-binding protein complexes were eluted from the beads by adding 50μl sample buffer with 0.2M DTT and boiling at 95°C for 3 minutes before centrifuging at 5000xg for 5 minutes at 4°C. Immunoprecipitated proteins were analysed by denaturing SDS-PAGE followed by immunoblotting. Where immunoprecipitated fractions were not immediately loaded onto a gel they were snap-frozen in liquid nitrogen and stored at -80°C.

2.5.3 Subcellular Fractionation

Cells were fractionated into subcellular components using the ProteoExtract ® Subcellular Proteome Extraction Kit (Calbiochem). MCF7 cells were grown in 6-well plates (Corning) and transfected as described in section 2.1.3, and incubated at 37°C for 24h. Cells were harvested into 1 ml PBS and centrifuged at 4000xg for 10 minutes at 4°C. Pelleted cells were then fractionated according to manufacturer's instructions to provide fractionation of the cells into cytosolic, membrane and organelle, nuclear and cytoskeletal protein fractions. Protease inhibitors (provided with the kit) were added to each fractionation buffer immediately before use. Fractionated lysates were snap-frozen and stored at -80°C before analysing by SDS-PAGE and immunoblotting.

2.5.4 RNA interference - AGR2

MCF7 (American) cells were grown in 6-well plates or 10 cm dishes (Corning) and transfected with siRNA, control siRNA, or were not transfected. The siRNA to AGR2 used was siGenome Smart Pool M-003626-00-0005, and control siRNA used was from the siGENOME Non-Targeting siRNA Pool (Dharmacon/ThermoScientificBio). Transfections were carried out using Dharmafect transfection reagent according to manufacturer's instructions. siRNA was diluted to 20μM and 5μl of this was used per well for transfection in a total volume of 2 ml of media (final concentration 50nM). Cells were incubated for 72h at 37°C before analysing by SDS-PAGE and immunoblotting or by immunofluorescence.

2.6 Immunofluorescence

Cells were plated on coverslips (generally at 5x10^3 - 1x10^4 cells per well) in the bottom of 12-well dishes (Corning) and incubated for 24h at 37°C. If transfection was required, adherent cells were transfected with 0.8μg/well of DNA using Attractene transfection reagent as described in section 2.1.3. Cells were then incubated for a further 24h at 37°C. Cells were fixed by first rinsing in cold PBS, followed by fixing in Standard Formaldehyde Buffer (Table 2.13) for 10 minutes at room temperature. Cells were then washed 3x 5 minutes in PBS-TritonX100 (0.1%) before blocking in PBS-TritonX100 (0.1%) with 2% BSA (w/v) for 1 hour at room temperature. Cells were incubated overnight at 4°C with primary antibodies as described in Table 2.14 and in the Results sections. Coverslips with fixed cells were then washed 3x5 minutes in PBS-TritonX100

Fixing and Permeabilisation Buffers for Immunofluorescence

	Standard Formaldehyde Fixation Buffer	
	3.7% Formaldehyde 10mM EGTA pH8.0 100mM Pipes pH6.8 1mM $MgCl_2$ TritonX100 0.2% (v/v) 10 minutes room temp	
Formaldehyde Buffer fixation, methanol permeabilisation	**Formaldehyde Buffer fixation, Glycine-Saponin Permeabilisation**	**Formaldehyde Buffer fixation, High TritonX100 permeabilisation**
Fix in Standard Formaldehyde Buffer Additional incubation in ice-cold Methanol for 2 min at -20°C	Fix in Standard Formaldehyde Buffer 15 min Wash 3x100mM glycine-PBS Permeabilise 100mM glycine, 5%(v/v) saponin for 20 min room temp	Fix in Standard Formaldehyde Buffer Wash 2xPBS Permeabilise 0.5% (v/v) Triton X100, BSA 1% (w/v) for 15 min room temp

Table 2.13: Fixing and Permeabilisation Buffers for Immunofluorescence

(0.1%), and incubated with appropriate secondary antibodies (Table 2.15) for 30 minutes at room temperature in the dark. Cells were then washed again 3x5 minutes and either mounted onto glass slides with mounting medium containing DAPI nuclear stain (Vectashield with DAPI hard-set mounting fluid), or were incubated for 15 minutes at 37°C in the dark with 50μl TOPRO-3 nuclear stain diluted to 1:1000 in PBS. For TOPRO-3 treated cells, these were washed again 3x5 minutes in PBS-TritonX100 (0.1%) before mounting on glass slides (Superfrost) with mounting fluid for fluorescent analysis without DAPI (DakoCytomation). Cells were analysed on a Leica SP-5 confocal microscope or on a Nikon A1R confocal microscope with NIS elements imaging software.

List of Antibodies for Immunofluorescence

Name	Raised in	Dilution	Company	Product number
Anti-Acetylated α-tubulin	Mouse	1:1000	Sigma	Clone 6-11B-1 T6793
Anti-γ-tubulin	Mouse	1:1000	Sigma	Clone GTU-88, T6557
Anti-AGR2 polyclonal	Rabbit	1:500-1:1000	Moravian Biotech	K47
Anti-AGR2 monoclonal	Mouse	1:1000	Abnova	H00010551-M01
Anti-HA monoclonal	Mouse	1:1000	Abcam	ab18181
Anti-MKS3 (C-terminal) polyclonal	Rabbit	1:100	C Johnson, Leeds University	N/A
Anti-V5 epitope monoclonal	Mouse	1:1000	Abcam	27672

Table 2.14: List of Antibodies for Immnofluorescence

List of Secondary Antibodies for Immunofluorescence

Name	Dilution
Alexa Fluorophore conjugated goat anti-mouse 488	1:200-1:1000
Alexa Fluorophore conjugated donkey anti-rabbit 488	1:200-1:1000
Alexa Fluorophore conjugated goat anti-mouse 594	1:200-1:1000
Alexa Fluorophore conjugated donkey anti-rabbit 594	1:200-1:1000

Table 2.15: List of Secondary Antibodies for Immunofluorescence

2.6.1 mKate2-Arl13B IMCD3 immunofluorescence assays

IMCD3(mKate2-Arl13B) cells were plated at a density of 10^4 per coverslip and incubated overnight at 37°C on coverslips in a 12 well plate. Cells were fixed in Standard Formaldehyde Buffer (Table 2.13), washed and blocked as described above, before incubation with anti-AGR2 mouse monoclonal antibody (Abnova) or Anti-AGR2 rabbit polyclonal antibody (Moravian Biotech) overnight at 4°C. Coverslips were then washed and incubated with Alexa-fluorophore conjugated antibody (either goat anti-mouse 488 or donkey anti-rabbit 488) at a dilution of 1:1000, washed again and mounted using mounting fluid containing DAPI (Vectashield with DAPI hard-set mounting fluid). Mounted coverslips were analysed by confocal microscopy.

2.6.2 Immunofluorescence for characterising custom MKS3 antibodies

IMCD3 cells were plated at a density of 10^4 per coverslip and incubated overnight at 37°C on coverslips in a 12-well plate. Cells were fixed and permeabilised as per each of the four fixation protocols (Table 2.13), and then blocked in PBS containing TritonX100 (0.1%) and 2% bovine serum albumin (BSA) for one hour. Cells were then incubated overnight at 4°C with mouse anti-acetylated α tubulin and one of the four antibody preparations (fractions) as described in section 5.3 and Figure 5.4. Cells were then washed according to the various wash protocols, and incubated in secondary antibody (Alexa fluorophore conjugated donkey anti-rabbit 488, Alexa fluorophore conjugated goat anti-mouse 594) at a dilution of 1:1000 for 1 hour. Cells were then washed and

coverslips were mounted using DAPI-containing mounting fluid (Vectashield with DAPI hard-set mounting fluid).

2.6.3 Total Internal Reflection Microscopy

We used TIRF-microscopy to image MKS3 at or just beneath the plasma membrane. MCF7 (American) cells were co-transfected as described in section 2.3.1 (Lipofectamine) with V5-tagged MKS3 wild type, V5-tagged MKS3^{PTPIFA} mutant, V5-tagged MKS3^{PTTIYY} double mutant, or V5-empty vector (pcDNA-nV5-3.1/DEST), all together with pAcGFPN1 empty vector to mark transfected cells, and incubated overnight on coverslips. Cells were then fixed in Standard Formaldehyde Buffer (Table 2.13), washed and blocked as described above, before incubating with mouse monoclonal anti-V5 (abcam 27672) and rabbit anti-Ecadherin primary antibodies overnight at 4°C. Cells were washed again as described above before adding fluorophore-conjugated secondary antibodies (Alexa647 anti-mouse, Alexa568 anti-rabbit) at a dilution of 1:1000 for 1h at room temperature. Coverslips were washed 3x5 minutes and then fixed and stained cells were placed in a chamber filled with buffer (PBS) and analysed by Total Internal Reflection Fluorescence Microscopy by Prof R Duncan, Heriot Watt University (Olympus IX81 fully motorised microscope stand, equipped with a PLAN APO 150X TIRFM objective, a 4-line motorised TIRF combiner and 405, 491, 561 and 633 nm lasers). Ecadherin was used as a positive control to ensure that the cells being visualised were still adherent, and expressed GFP was used to locate and define the boundaries of transfected cells being analysed.

2.6.4 Proximity Ligation Assay

MCF7 (American) cells were seeded onto coverslips in 12-well plates at $5x10^4$ cells per well and incubated at 37°C for 24h. Cells were transfected with GFP-tagged DNA aptamers or HA-tagged reptin (from Dr M Maslon, University of Edinburgh) as described in section 4.2 (Figure 4.5), using Attractene transfection reagent according to manufacturer's instructions. HA-tagged reptin was used as a positive control for the assay as a published, known binding parter for AGR2 and with a known nuclear subcellular localisation pattern. Cells were fixed in Standard Formaldehyde Buffer (Table 2.13), washed and blocked as described above in section 2.6 before incubating with primary antibodies overnight at 4°C (anti-GFP mouse monoclonal, Abcam, 1:1000; anti-AGR2 rabbit polyclonal, Moravian Biotech, 1:1000). After incubation overnight with primary antibodies, cells were processed according to manufacturer's instructions using the Duolink PLS In Situ kit (PLA probe anti-rabbit MINUS, PLA probe anti-mouse PLUS, Detection Reagents Red). Cells were imaged using an automated microscope (Olympus BX61), with 60 individual fields imaged per coverslip and with 7 images per field moving down through the cell from the top of the cell to the coverslip. These "stacks" were combined to provide one merged image per field. Images were analysed by an automated program (Ariol) that counted the number of fluorophores per field and number of cells counted. Microsoft Windows Excel was used to analyse the data.

2.7 Embryonic kidney organ culture

2.7.1 Embryonic kidney organ culture and immunofluorescence

Plugged CD1 female mice were euthanised during gestation, at embryonic day E12.5 (12.5 days post fertilisation) and embryonic kidneys were microdissected from the embryos using sterile needles and forceps under a light microscope. Tail ends or limb buds were kept for genotyping (section 2.9). Kidneys were placed onto permeable culture membranes in 6-well plates (Costar Transwell, Corning, 24mm diameter inserts, 0.4μm pore size) with 1.5 ml MEM media in the base of the wells such that membrane is not submerged. Kidneys were grown for between 24h and 96h at 37°C in 5% CO_2 before fixing and staining.

After culturing for 24-96h, kidneys were removed from culture (still adherent to the porous membrane) and fixed in 4% paraformaldehyde (PFA) in 5 ml PBS (in bijou tubes) for 30 minutes at room temperature. Kidneys adhered to membranes were then washed (5x30 minutes, 5 ml PBS) and incubated with primary antibodies (Table 2.16) diluted appropriately in 0.6 ml PBS overnight at 4°C. Kidneys were then washed (5x30 minutes, 5 ml PBS), and incubated with secondary antibodies (Table 2.17) at 1:400 overnight at 4°C (protected from the light with foil). Kidneys were washed in PBS (5x15 minutes) and mounted on glass slides (Superfrost), using glass coverslips as spacers, with mounting fluid containing DAPI (Vectashield with DAPI mounting fluid - note NOT hard-set). Fixed, stained and mounted kidneys were analysed by confocal microscopy with a X60 oil immersion lens.

List of Antibodies for Kidney Organ Culture Immunofluorescence

Name	Raised in	Dilution	Company	Product number
Anti-AGR2 polyclonal	Rabbit	1:500	Moravian Biotech	K47
Anti-Acetylated α-tubulin	Mouse	1:500	Sigma	Clone 6-11B-1, T6793
Anti-γ-tubulin	Mouse	1:500	Sigma	Glone GTU-88, T6557
Anti-JAG1 polyclonal	goat	1:100	RandD Systems	AF599
Anti-Ecadherin monoclonal	mouse	1:100	BD Transduction Laboratories	610182
Anti-ZO-1	rat	1:20	DSHB (Developmental Structures Hybridoma Bank)	TBC
Anti-NCAM monoclonal	mouse	1:200	Sigma	9672
Anti-podocalyxin polyclonal	mouse	1:200	RandD Systems	AF1556

Table 2.16: List of Antibodies for Kidney Organ Culture Immunofluorescence

2.7.2 Embryonic kidney organ culture for analysis of transgenic line Agr2PR(GFPCre)

Agr2PR(GFPCre) x CD1 embryonic kidneys were microdissected as above at embryonic day E12.5 and placed into culture on membranes, also as described above. Kidneys were cultured for 24h before fixing in 4% paraformaldehyde for 30 minutes at 4°C. Kidneys, still adherent to membranes, were washed in PBS (5x15 minutes) and mounted on glass slides (Superfrost) as described above using mounting fluid containing DAPI (Vectashield with DAPI mounting fluid). Fixed, stained and mounted kidneys were analysed by confocal microscopy with a x60 oil immersion lens (Nikon A1R as described above).

List of Secondary Antibodies for Organ Culture Immunofluorescence

Name	Dilution
Alexa Fluorophore conjugated goat anti-mouse 488	1:200-1:1000
Alexa Fluorophore conjugated donkey anti-rabbit 488	1:200-1:1000
Alexa Fluorophore conjugated donkey anti-rabbit 594	1:200-1:1000
Alexa Fluorophore conjugated donkey anti-goat 488	1:200-1:1000
Alexa Fluorophore conjugated donkey anti-rat 488	1:200-1:1000

Table 2.17: List of Secondary Antibodies for Organ Culture Immunofluorescence

2.8 Immunohistochemistry

Agr2PR(GFPCre) pups were euthanised at 8 weeks of age by cervical dislocation (performed by technician in transgenic facility) and intestines were dissected and made into a gut roll. Intestines were placed into ice-cold PBS before fixing overnight at 4°C in 4% paraformaldehyde-PBS. Organs were washed 2x30 minutes in cold PBS on an orbital shaker at 4°C, and then processed according to Table 2.18.

Organs were removed from liquid paraffin and embedded in paraffin blocks before cutting into 6μm sections using a microtome (Leica RM2235) and placing on glass slides (Superfrost). Sections were baked at 50°C for 1h before placing in xylene overnight. Sections were then washed 3x10 minutes in 100% ethanol, 2x10 minutes in 96% ethanol, 1x10 minutes in 70% ethanol and 3x10 minutes in water. Antigen retrieval was carried out using TEG buffer (10mM Tris, 1mM EGTA, pH9.0) by boiling in the microwave for 10 minutes and sections were then left to cool at room temperature for 1h. Sections were then washed 1x30 minutes in ammonium chloride buffer (50mM in PBS, pH7.4). Sections were further washed 3x10minutes in PBS with gelatine

Protocol for Paraffin-Embedding of Mouse Intestines

Solution	Time (min)	Temperature (°C)
25% ethanol	15	25
50% ethanol	15	25
75% ethanol	15	25
	Placed into TissueTek VIP 5Jr IHC machine	
70% ethanol	40	35
85% ethanol	40	35
95% ethanol	40	35
100% ethanol	40	35
100% ethanol	40	35
Xylene	40	35
Xylene	40	35
Paraffin	40	58
Paraffin	40	58
Paraffin	40	58
Paraffin	40	58

Table 2.18: Protocol for Paraffin-Embedding of Organs

(0.2%), saponin (0.05%) and BSA 1%, and then incubated with primary antibody (rabbit polyclonal anti-GFP, Abcam 6556) at a dilution of 1:500 or 1:1000 in PBS-TritonX100 (0.3%) with BSA (0.1%) overnight at 4°C. Sections were then washed 3x10 minutes in PBS with gelatine (0.2%), Saponin (0.05%) and BSA 0.1%) before incubating with secondary antibodies (Alexa 488 donkey anti-rabbit, 1:200 in PBS-TritonX100 (0.3%) with BSA (0.1%) for 90 minutes at room temperature (in the dark). Sections were then washed 3x in PBS and dried gently before adding mounting fluid (Vectashield with DAPI) and covering with a coverslip. Nail varnish was used to adhere coverslip to glass

slide. Immunostained sections were analysed for GFP expression on a Zeiss axioplan 2 fluorescence microscope equipped with a Hamamatsu ORCA AG camera.

2.9 Creation of transgenic animals

2.9.1 Microinjection of oocytes

All oocyte microinjections were carried out by Paul Devenney and Emma Murdoch, transgenic unit, MRC Human Genetics Unit, Western General Hospital. DNA for microinjection was prepared as follows: The Agr2 BAC with expression cassette integrated at Agr2 start codon as described in sections 2.4.5 and 2.4.6 was linearised with PI-SceI restriction endonuclease digest (single restriction site in backbone of BAC) with the protocol shown below.

Component	Final concentration
BAC DNA	2μg
PI-SceI restriction endonuclease (NEB)	5U
PI-SceI Buffer (NEB)	1x
BSA (NEB)	1x
	to final volume 30μl
Incubation 37°C 12h, Enzyme denatured at 65°C for 20 min	

Table 2.19: Linearising Agr2 BAC for microinjection

DNA was ethanol precipitated by adding 3μl of 3M NaOAc to the 30μl sample, and then adding 75μl of ice-cold ethanol before incubating at -20°C for 20 minutes. DNA was pelleted by centrifuging at 10,000xg for 15 minutes

and pellet was washed with 1 ml of 70% ethanol. DNA was again pelleted in the same way and pellet was allowed to air-dry. DNA was resuspended in 30μl of microinjection buffer at 37°C for 5 minutes with gentle shaking. Concentrations were determined by NanoDrop and diluted to 0.5ng/μl in microinjection buffer prior to microinjection.

Component	Final concentration
Tris-HCl pH 7.5	10mM
EDTA pH8.0	0.1mM
NaCl	100mM
Spermine (Sigma S-1141)	30μM
Spermidine (Sigma S-2501)	70μM
distilled, sterile H_2O	to appropriate volume

Table 2.20: DNA Microinjection Buffer

2.9.2 Genotyping

Transgenic mice were genotyped using earclip (for pups) or tail or limb bud (for embryos). DNA was extracted by incubating tissue in 75μl sodium hydroxide (0.1M) at 95°C for 30 minutes followed by neutralising by addition of 75μl Tris-HCl to a final concentration of 20mM. Genotyping was carried out by PCR for the Cre trangene using the components and protocol below. Primers for Cre and for internal control Fabpi are also listed below. All genotyping was carried out by myself unless otherwise stated.

Genotyping PCR

Amplification of Cre transgene	
Component	**Final Concentration**
10x Platinum Taq buffer	1x
10mM dNTP	0.2mM
50mM MgCl	2mM
10μM Forward primer	0.4μM
10μM Reverse primer	0.4μM
Platimum Taq 4U/μl	2U
Template DNA from tissue	5μl
	Total volume 25μl

PCR protocol to amplify Cre-recombinase gene

step 1:	95°C	2:00 min
step 2:	95°C	30 sec
step 3:	56°C	30 sec
step 4:	72°C	1:00 min
step 5:	go to step 1 34x	
step 6:	72°C	5:00 min
step 7:	4°C	for ever

Genotyping Primers

Name	Sequence
Cre/F	GCATTACCGGTCGATGCAACGAGTGATGAG
Cre/R	GAGTGAACGAACCTGGTCGAAATCAGTGCG

Table 2.21: Genotyping Primers

Chapter 3

In Vivo Analysis of Agr2 expression

3.1 Introduction

The involvement of Agr2 in diseases such as cancer and inflammatory bowel disease has been well established, however, the mechanisms by which it promotes disease are still unclear. It has been implicated in limb regeneration [Kumar et al., 2007], at least in amphibians, suggesting a biological role that involves maintenance of cell growth, and yet other studies implicate Agr2 in terminal intestinal goblet cell differentiation [Chen et al., 2012] and terminal differentiation of gastric epithelial cells in a mouse knockout model [Gupta et al., 2012b]. Insights have been gained into the biological function of the protein with the publication of this most recent mouse knockout, however, more subtle phenotypes may not have yet come to light, and the mechanism by which Agr2 effects the balance between the stem cell and differentiated cell populations

are yet to be discovered.

Establishing the expression pattern of a gene through developmental stages and adult tissues can be useful for establishing its biological role [Alberts et al., 2002]. There are numerous ways to investigate the expression pattern of a gene, including antibody-based methods using fixed and sectioned tissues or using organ culture followed by tissue fixing as described previously. However, antibody-based techniques can have drawbacks such as antibody specificity and compatibility with tissue processing, and do not give real-time information possible with other types of expression analyses. Alternative methods for determining the expression pattern of a gene in mice include the use of transgenic animals and promoter reporters which, coupled with lineage tracing, may also provide information about the contribution of a gene to the differentiation or maintenance of a particular cell type (discussed in sections 1.1.2 and 3.4.4). In order to fully investigate the global expression pattern of Agr2 in both the developing mouse embryo and in adult tissues, and to investigate the contribution of Agr2-expressing cells to particuar cell types, we decided to make a transgenic promoter reporter for the murine *Agr2* gene, expressing GFP-Cre from the *Agr2* promoter using BAC transgenics, and with random insertion in the mouse genome. In addition, expressing Cre recombinase from an ectopic *Agr2* promoter would provide tissue-specific Cre-drivers for tissue-specfic expression of other transgenes following genetic crosses, making the mouse line useful for the AGR2 field in general.

Construction of targeting vectors for use in model animal transgenesis can be lengthy and complex. Rather than making a construct that would only be useful for our present study into the expression patterns of Agr2, we de-

cided to create a universal, modular expression vector that could be used with other projects and genes. In this chapter we describe a universal and highly adaptable construct for gene expression from a BAC or endogenous promoter, using a BAC cloning platform and multi-site Gateway technology [Sasaki et al., 2004]. As expression is driven by the promoter available on the BAC, this construct is readily adaptable for the development of transgenic animals used for a variety of analyses such as gene expression during development, tissue-specific knockdown and tissue-specific inducible gene expression. The modular nature of the vector makes it suitable for efficiently making variant BAC constructs and as such is adaptable to high-throughput generation of BAC transgenes or targeting vectors using recombineering [Poser et al., 2008] [Skarnes et al., 2011]. We demonstrate its effectiveness *in vivo* as a promoter reporter using the known expression pattern of *Six2* as proof-of-concept, and then discuss attempts to use the universal construct as a means of determining the expression pattern and for lineage tracing of Agr2.

3.2 Construction of the Universal Transgenic Expression Vector

The generation of highly adaptable constructs for use in transgenic projects is key to streamlining the production of transgenic lines. In addition, large genomic contructs such as Bacterial Artificial Chromosomes (BACs) have been shown to provide high predictability for expression of transgenic constructs [Giraldo and Montoliu, 2001]. Commercially available BAC libraries contain a collection of large genomic constructs from a particular genome,

such as mouse, and each BAC includes all the genomic components of that region including introns, exons, and upstream or downstream regions. In some cases they may include the promoter region of a given gene and, depending on the size of the BAC, may also include regulatory elements essential for the expression of the gene. Using a technique termed *recombineering*, BAC-based vectors for the creation of transgenic animals can be readily made in *E. Coli* [Copeland et al., 2001]. Recombineering is a technique that harnesses the power of lambda phage recombination enzymes in *E. Coli* to facilitate homologous recombination at specific sites. Strains of *E. Coli* have been engineered with temperature-sensitive recombination genes from the lambda phage, in which they are repressed at 32°C. When the bacteria are placed at 42°C for a brief period, the genes become derepressed, and the recombineering machinery within the bacteria becomes active. A recombineering-ready bacterial strain harbouring a BAC of interest can be electroporated with a DNA vector that contains sites of homology (homologous arms) with the BAC, and it will be recombined into the corresponding region of homology. In this way, DNA containing a transgene can efficiently and effectively be *recombineered* into a BAC containing the gene of interest. For the above-mentioned reasons we decided to base our universal and adaptable expression vector on BAC transgenics.

In order to make this construct adaptable we used the Gateway cloning platform from Invitrogen which enables quick and efficient cassette exchange [Hartley et al., 2000]. Like single-fragment Gateway recombination, Gateway multisite recombination can be carried out in two recombinase-mediated steps, however, it allows specific cassette exchange for up to 4 different fragments (Figure 3.1) [Cheo et al., 2004]. To begin with, we set out to create

Figure 3.1: Universal Expression Construct Cloning Process

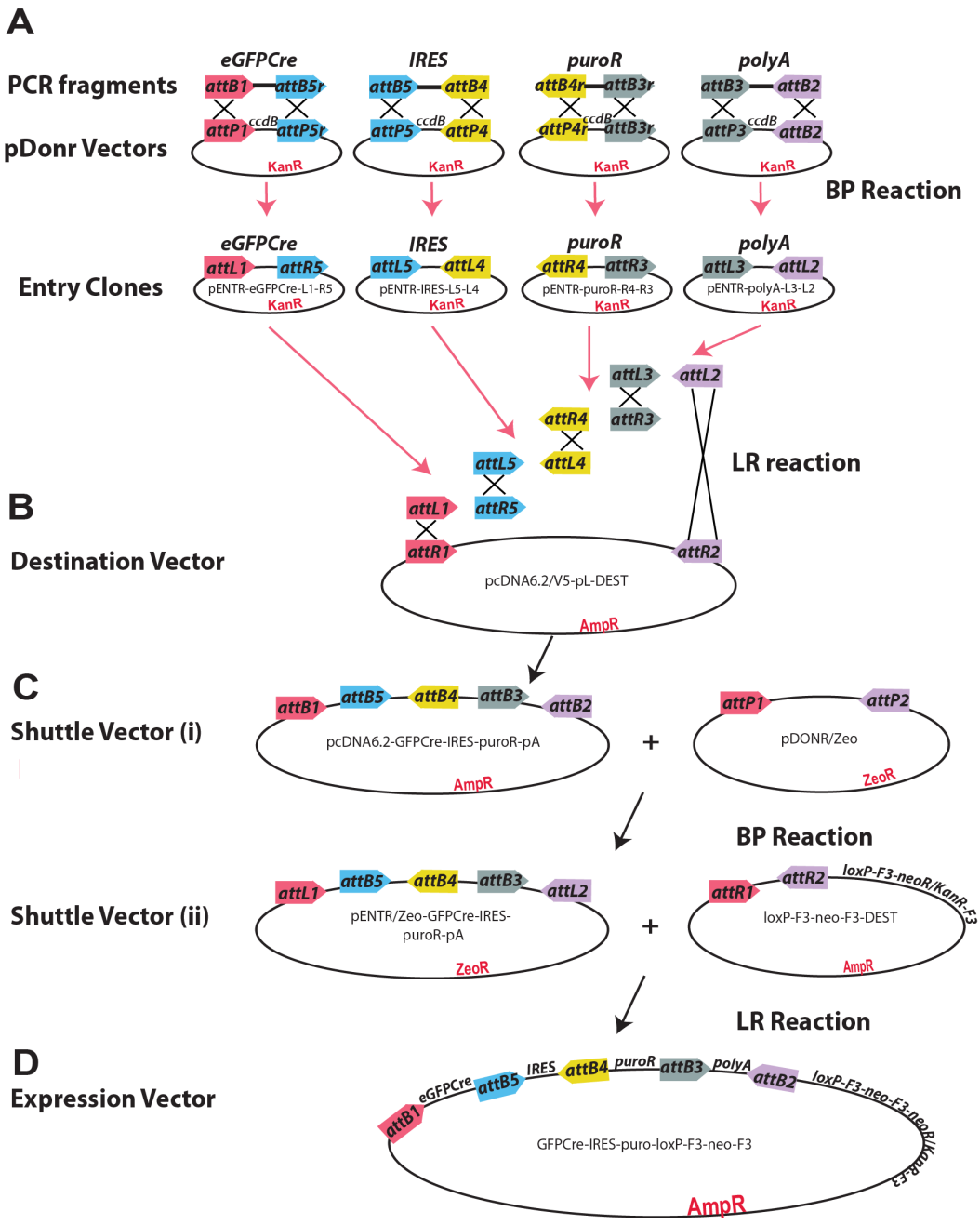


Figure 3.1 - Universal Expression Construct Cloning Process - Gateway entry clones were made for each of four cassettes (A) and then recombined in order, in one reaction using Multisite Gateway, into a destination vector (B), shuttle vector (i). Using a reverse Gateway reaction (C) the entire four-cassette block was placed into a second destination vector, shuttle vector (ii). A final LR reaction recombined the entire four-cassette block into the expression vector (D).

a 4-cassette expression vector containing eGFP-Cre, IRES, puromycin resistance and polyA cassettes. Each fragment was amplified by PCR with specific multi-site Gateway recombinase sites (*attB*) at the 5' and 3' ends (Figure 3.1A, primers table 2.5). The priniciple of Gateway recombinase cloning uses the *E.coli* lethal *ccdB* gene for positive selection; correct recombination of the PCR fragment removes the lethal gene and allows positive clone selection. A "BP reaction" (recombination by the BP clonase enzyme between *attB* sites and *attP* sites on pDONR vectors) is carried out to generate entry clones with kanamycin resistance. Multisite Gateway pDONR vectors contain unique *attP* sites which confer substrate specificity for the BP clonase enzyme. Thus, these sites are not interchangeable. Using these unique entry clones, an "LR reaction" (recombination by the LR clonase enzyme between *attL* and *attR* sites) was performed to bring all four fragments together, in order, in a single reaction (Figure 3.1B). The destination vector, pcDNA6.2/V5-pL-DEST, is a simple Gateway-compatible vector with ampicillin resistance for positive selection. Since we wanted to create an adaptable construct for use in transgenics, a final expression vector that included a removable neomycin/kanamycin resistance cassette for ES cell and bacterial selection respectively was required. The expression vector chosen was a Gateway-compatible loxP-F3-neo-F3-DEST with such a cassette, flanked by F3 sites for removal (Figure 3.1C). The bacterial kanamycin resistance gene in this vector prevented us from performing an LR reaction with the entry vectors directly into this expression vector. Therefore, we first placed the multi-site Gateway cassettes into the first vector described above (Shuttle Vector A; Figure 3.1B), and then performed a BP reaction between this vector, containing the multi-site cassettes, and a

Gateway-compatible vector, pDONR/Zeo, containing zeocin resistance (Figure 3.1C). Due to the substrate specificity of the attB sites in multisite Gateway, the entire 4-cassette block was recombined into pDONR/Zeo, resulting in pENTR/Zeo-GFPCre-IRES-PuroA-pA Shuttle Vector B (Figure 3.1C). An LR reaction with the final destination vector containing the neomycin/kanamycin cassette was performed, again recombining the entire 4-cassette block into the expression vector in one step (Figure 3.1D).

In this way, we created an expression vector for expression of four cDNAs, each flanked by substrate-specific *attB* recombinase sites for specific, efficient cassette exchange. One, two, three or even four cassettes can be replaced or removed either individually, in pairs or in a block (Figure 3.2A). However, we wanted to make this expression construct readily adaptable to a BAC transgenics platform using recombineering. Recombineering into a BAC uses homologous arms, regions of 50-400 nucleotides flanking the sequence to be recombineered. In order to make this construct easy to adapt to any BAC, we inserted multiple cloning sites at the 5' and 3' ends of the expression cassettes and neomycin resistance cassette (Figure 3.2A). We now had a highly adaptable construct comprising readily-exchangeable expression cassettes mediated by two-step Gateway recombinase activity, flexible multiple cloning sites for insertion of any homologous arms, and removable neomycin/kanamycin resistance under the control of bacterial (EM7) and mammalian (PGK) promoters for bacterial and ES cell selection (Figure 3.2A). We named this universal expression construct pMULTIrec. A list of all the vectors used to make the universal expression construct can be found in Table 2.7.

In order to replace a cassette, first a PCR product of the new cDNA

Figure 3.2: Universal Expression Construct schematic and cassette exchange process

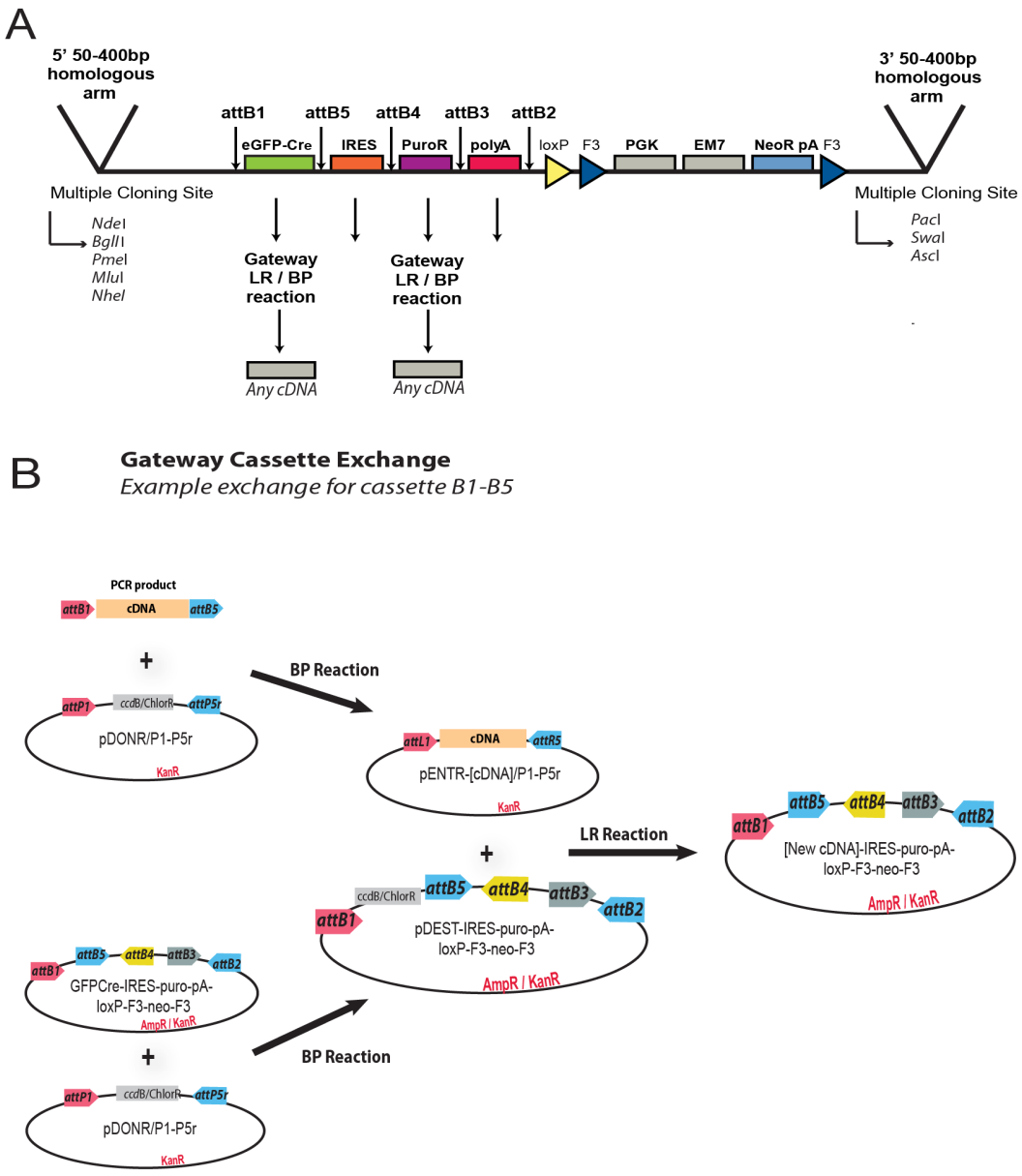


Figure 3.2 Universal Expression Construct schematic and cassette exhange process - (A) The universal expression construct contains four exchangable expression cassettes. Unique Multisite Gateway recombination sites flanking each cassette allow exchange of one, two, three or four cassettes at a time. At the 3' end the construct contains a Neomycin selectable marker under the control of PGK and EM7 promoters for mammalian and bacterial expression. The sites are flanked by F3 sites for removal in ES cells. The entire construct is flanked by multiple cloning sites for insertion of homologous arms for recombineering. (B) Each expression cassette or block of cassettes can be exchanged in two steps once an entry clone has been generated.

is amplified with the specific *attB* sites. Together with the corresponding pDONR vector (Figure 3.1A and Figure 3.2B), a BP reaction provides the new entry clone. In parallel, a BP reaction is performed between the existing expression vector (in this case eGFPCre-IRES-puroR-pA-loxP-F3-neo-F3) and the pDONR vector corresponding to the entry vector (Figure 3.2B). The result is a destination vector with one Gateway cassette replaced by the *ccdB* gene and flanked by *attR/attL* sites. A final LR reaction between the new entry vector containing the specific *attL/R* sites, and the new destination vector containing an "empty" cassette provides the vector for use in transgenics with one or more cassettes exchanged (Figure 3.2B).

To demonstrate the interchangeable nature of this construct, we exchanged the first cassette with other cDNAs to create new expression vectors useful for transgenic analysis. Using the method outlined above, we simply and efficiently created variations on the promoter reporter, including an mCherry expression vector (Figure 3.3A). Correct replacement of the eGFP-Cre cassette with mCherry cDNA was verified by sequencing (not shown) and restriction endonuclease digest pattern (Figure 3.3B). Other transgenic expression vectors easily adapted from this universal construct include an mKate2 promoter reporter (Karamjit Singh Dolt) and a promoter-driven eGFP-Cre-ERT2 expression vector (Eve Miller-Hodges), which have been used to create BAC mini-targeting vectors and as the basis for endogenous knock-ins. Insertion of the eGFP-Cre fusion protein allows tissue-specific Cre driver mice for generating tissue-specific conditional mouse strains and GFP-Cre-ERT2 provides inducible activation of Cre.

Although we have created these versions of the universal construct, any

Figure 3.3: Vectors and mCherry expression construct verification

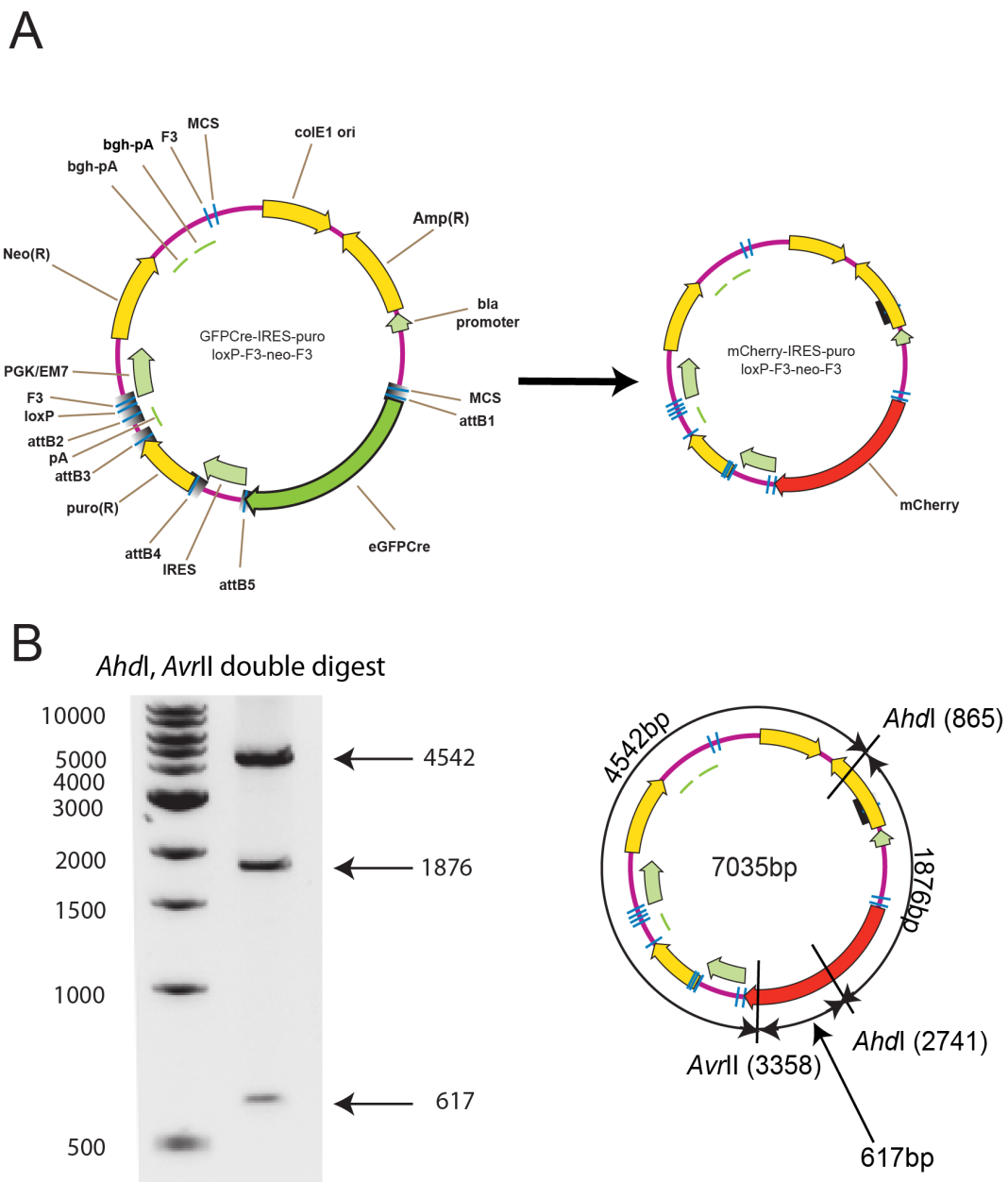


Figure 3.3 - Vectors and mCherry expression construct verification - (A) Vector maps of unviersal expression construct GFP-Cre variant (GFPCre-IRES-puro-loxP-F3-neo-F3; GCiP) and mCherry variant (mCherry-IRES-puro-loxP-F3-neo-F3; mCiP). mCherry variant has GFP-Cre cassette replaced with mCherry only. (B) Restriction digest analysis of mCherry variant to confirm correct cassette exchange.

reporter or other cDNA cassette can in two steps be integrated into any of the four exchangeable cassettes. Using multi-site Gateway technology combined with recombineering, the universal nature of the construct makes it a useful tool for high-throughput genetic analysis, and fast and easy development of transgenic constructs.

3.3 Testing the universal expression vector system *in vivo* using a known gene expression pattern

In order to test our system *in vivo* a gene with a known expression pattern was required. Six2 expression in the developing kidney provided this known expression pattern, as it has previously been shown by lineage tracing (using different targeting constructs for the same gene) to be expressed in the nephrogenic progenitor cells [Kobayashi et al., 2008]. The developing kidney involves extension and branching of an epithelial structure termed the ureteric bud into the metanephric mesenchyme. Communication between these two cell types leads to a process known as mesenchymal-to-epithelial transition (MET) in which the mesenchymal cells become epithelialised and form the basis of the developing nephrons (the filtering units in the adult kidney). The nephrogenic progenitor cells, or cap mesenchyme, are the mesenchymal cells that will undergo MET and become nephrons [Costantini and Kopan, 2010]. The expression of *Six2* in the developing murine kidney was chosen as a way to test our expression vector in part because the developing kidney provides

a useful medium for analysis. Kidneys can be dissected out of mouse embryos as early as E10.5-E11.5 and grown in culture, where they continue to grow and develop for up to six days [Saxen and Lehtonen, 1987]. The kidneys are transparent and are therefore ideal for visualising expressed fluorescent markers without the need for sectioning, and can be fixed and mounted whole for confocal imaging or other types of microscopy. Furthermore, the ability to grow the kidneys in culture makes them applicable for live imaging studies.

3.3.1 Random integration of BAC construct

The making of the transgenic *Six2* promoter reporter mice and subsequent analysis was carried out by Dr Karamjit Singh Dolt in Prof Hastie's lab at the IGMM under the supervision of Dr Peter Hohenstein. Briefly, homologous arms to the *Six2* gene were inserted at the 5' and 3' multiple cloning sites of the eGFPCre-IRES-puromycin-pA vector (pMULTIrec) using *Nde*I and *Sal*I (5' end) and *Pac*I and *Asc*I (3' end). This was then targeted by recombineering into a BAC containing the *Six2* locus (Figure 3.4A), replacing the start codon in exon 1. This recombineered BAC was linearised and microinjected into oocytes, and transgenic mice were generated. Analysis of the transgenic embryos (Figure 3.4B) and developing, cultured kidneys (Figure 3.4C) demonstrated GFP expression as expected. Cultured kidneys were fixed and stained with E-cadherin to highlight structures within the developing kidney, and analysed by confocal microscopy. GFP positive cells were found in the cap mesenchyme cells surrounding the ureteric bud (Figure 3.4D) as has been previously described [Kobayashi et al., 2008]. Thus, the universal expression vector was effectively able to ectopically express inserts via BAC transgenics.

Figure 3.4: Ectopic expression of Six2 promoter reporter made from universal expression construct

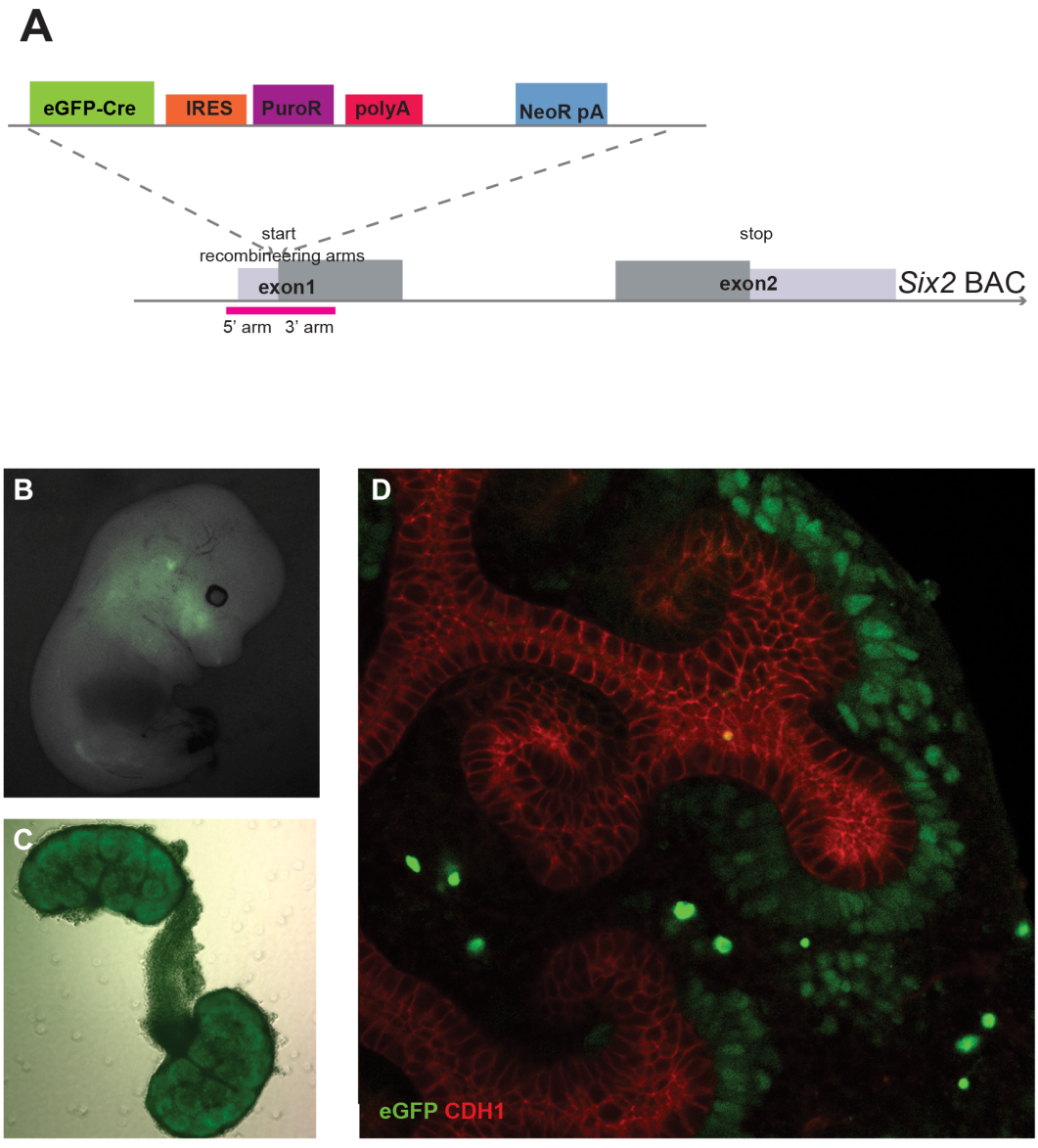


Figure 3.4 Ectopic expression of *Six2* promoter reporter made from universal expression construct - The universal expression construct containing a GFP-Cre-IRES-PuroA-polyA expression cassette (A) was recombineered using homologous arms into the start site of the *Six2* gene on a BAC containing the entire *Six2* gene (Dr K S Dolt). The expression construct was microinjected into oocytes and genomic DNA from the embryos were screened for the transgene using PCR. Positive embryos were analysed for expression of GFP (B) in the whole embryo, (C) in the embryonic kidneys and (D) in the cap mesenchyme of the developing embryonic kidney (analysis by Dr K S Dolt).

3.3.2 Endogenous targeting of BAC construct

The expression vector can be used for ectopic expression from the BAC in the model animal genome, or for expression from an endogenous locus, by first retrieving a fragment of the BAC into a normal targeting vector, followed by transfection and selection in ES cells. In order to test whether the universal construct would allow expression from endogenous loci, a targeting construct with large homologous arms from the *Six2*-GCiP BAC was retrieved into a plasmid backbone by recombineering. A thymidine kinase (TK) negative selection gene to eliminate non-homologous recombination events was then recombineered into the plasmid backbone (Figure 3.5). The *Six2*-GCiP BAC already contains a neomycin positive selection marker for ES cell selection as part of the universal construct, thus allowing both positive and negative selection to screen for correct targeting of the construct at the endogenous locus in ES cells. The vector was electroporated into mouse ES cells, and was integrated at the *Six2* endogenous locus by homologous recombination. Dr Dolt used long-range PCR and Southern blotting to identify correctly targeted clones (Figure 3.6A), which were then used to generate $Six2^{+/GCiP}$ knock-in mice. These mice were healthy and viable as has been previously shown in other knock-in studies of this gene [Self et al., 2006], and the GFP expression patterns were identical to the randomly-integrated Six2-GCiP BAC mentioned above.

The first variant of the universal expression construct, GCiP (with stop codon), was created for use as a promoter reporter that could also provide lineage tracing or tissue-specific Cre-driver mice for generating tissue-specific conditional mouse strains. Expression of Cre recombinase fused to eGFP provides the possibility of lineage tracing by crossing the transgenic line with a

Figure 3.5: Method for retrieval of targeting vector with eGFP-Cre (GCiP) construct as example

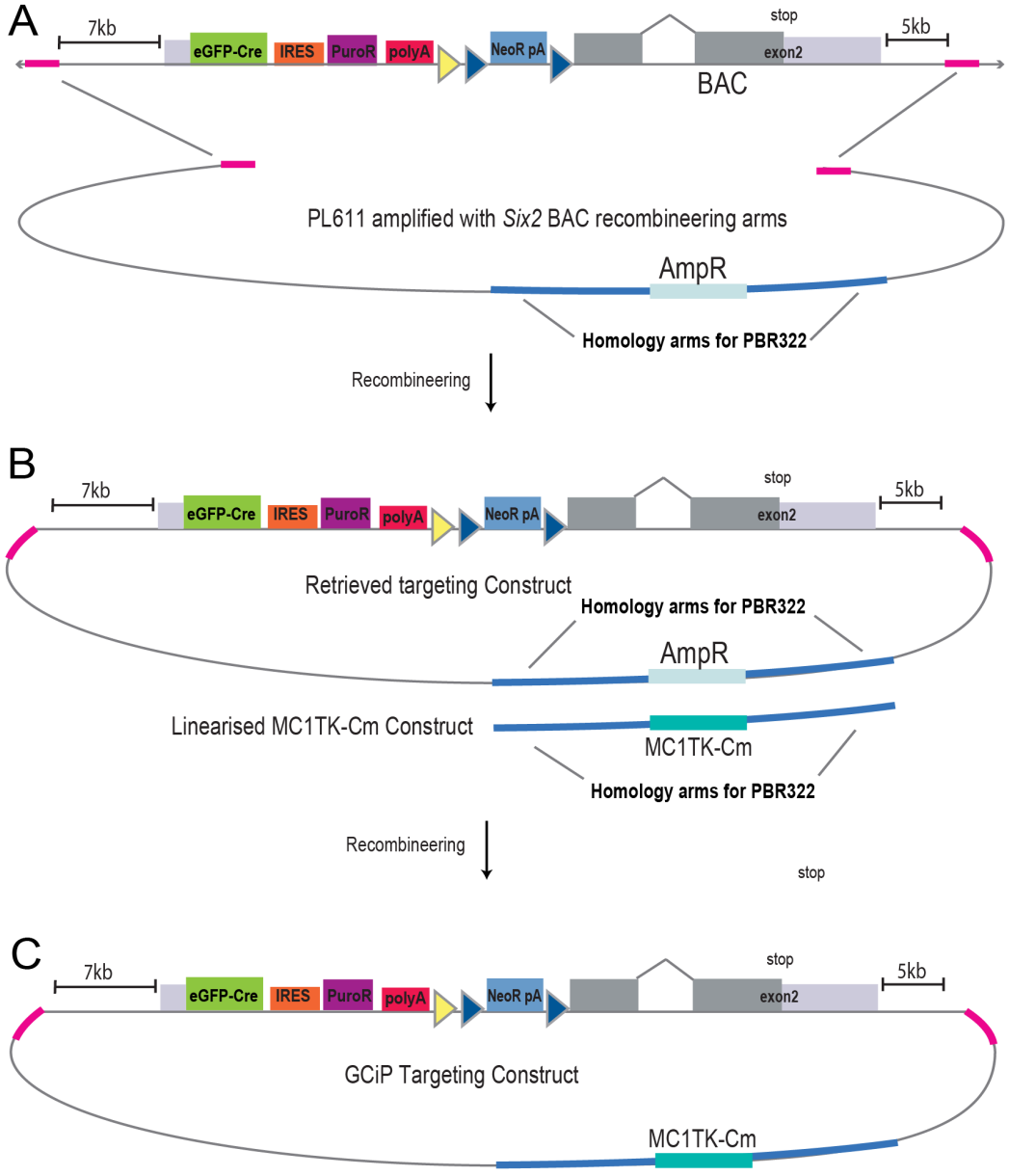
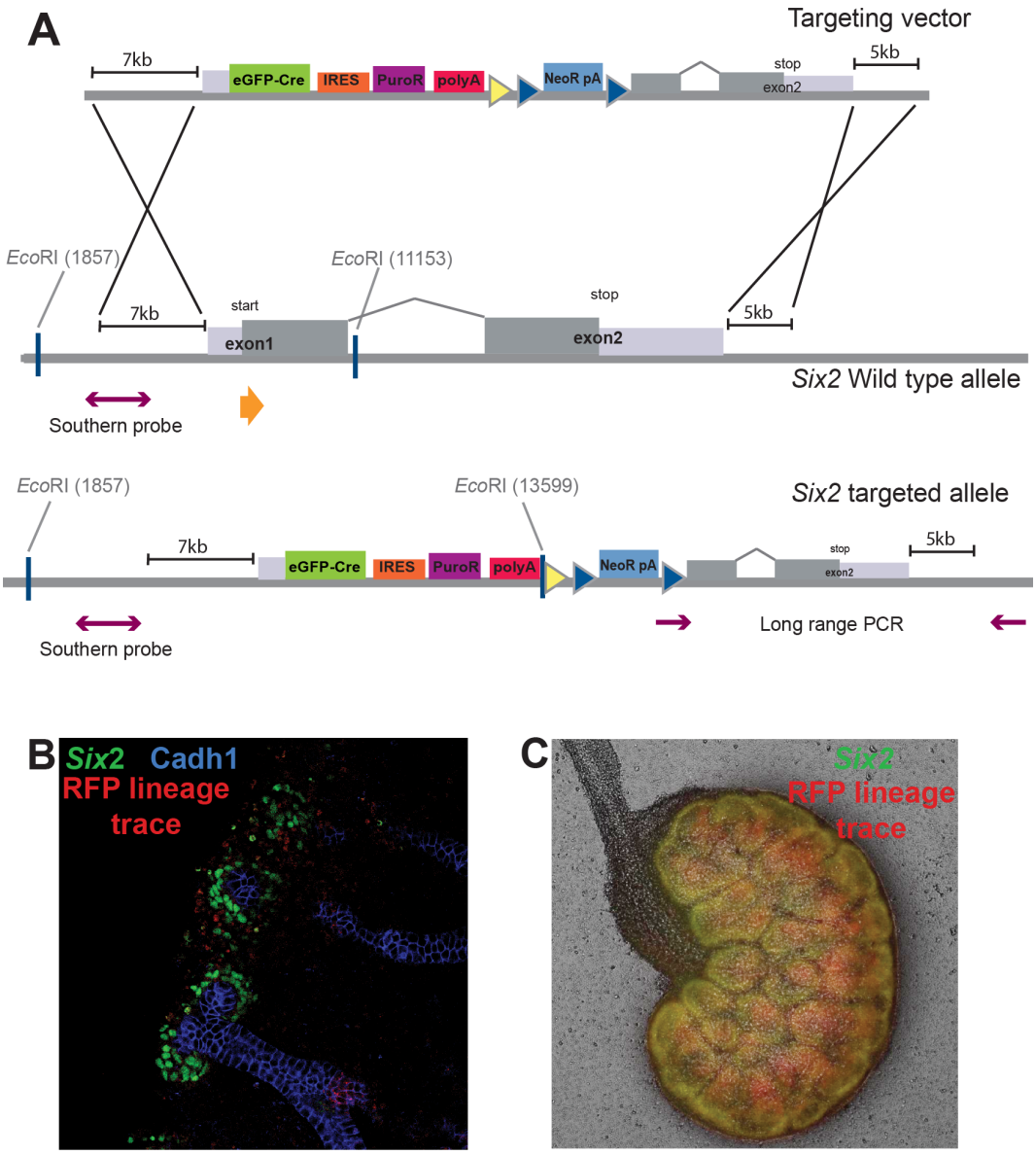


Figure 3.5 Method for retrieval of targeting vector with eGFP-Cre (GCiP) construct as example - (A) Plasmid PL611 containing ampicillin resistance gene was amplified by PCR with 5' and 3' homologous arms to the *Six2* BAC upstream and downstream of the *Six2* gene, and with 5' and 3' homologous arms to plamsid PBR322 upstream and downstream of the MC1TK-Cm cDNA on that plasmid. (B) The BAC region between the homologous arms was recombineered into plasmid PL611 generating a retrieved targeting construct. (C) The MC1TK-Cm gene was integrated into the targeting construct by recombineering, replacing the ampicillin resistance gene and generating the GCiP *Six2* endogenous targeting construct. Generation of retrieval construct by Dr K S Dolt.

Cre-activated reporter. In order to test the Cre moiety of this variant of the universal construct, the *Six2*$^{+/GCiP}$ mouse was crossed to a *Rosa26*^{tdRFP} Cre reporter mouse [Luche et al., 2007]. Expression of Cre drives removal of the stop codons and polyA signals that prevent expression of the tdRFP marker, leading to constitutive RFP expression in all cells of this lineage (see section 3.4.4 and 1.1.2 for further discussion of Cre-driven lineage tracing). Dr Dolt used time-lapse imaging of the brightfield, GFP and tdRFP signals to follow the dynamics of the lineage trace. As before, cap mesenchyme cells surrounding the ureteric bud expressed GFP, and these cells rapidly became RFP-positive while remaining GFP-positive as the Cre moiety activated expression of the tdRFP reporter allele (Figure 3.6B). As the cells continued to develop and progressed through MET they lost GFP expression, and post-MET nephrons on these kidneys were GFP-negative whilst retaining the expression of tdRFP (Figure 3.6C). The timelapse video can be accessed in Supplementary Data of the publication listed on page xiii.

The data from the *Six2*$^{+/GCiP}$ knock-in construct described above confirms that the universal recombineering vector can be used for the generation of endogenous targeting vectors and the generation of transgenic BAC models as well as randomly integrated BAC expression vectors.

Figure 3.6: Endogenous targeting of Universal Expression Construct with Six2 as proof of concept



Fluorescence images from Karamjit Singh-Dolt, Hastie lab, IGMM

Figure 3.6 Endogenous targeting of universal expression construct with *Six2* as proof of concept - (A) GCiP Six2 endogenous targeting vector was recombined in ES cells into Six2 locus by homologous recombination. Correct recombination at locus was verified by long-range PCR and Southern blot. Genotyped transgenic mice were screened for transmission of transgene, and then founder males were crossed to a Cre-driven Rosa26tdRFP reporter for lineage tracing. Embryonic kidneys from the cross were analysed for GFP expression in the cap mesenchyme and subsequent RFP expression in nephrogenic progenitor cells (B) and in the whole kidney by live cell imaging (C).

3.4 Creating the *Agr2* promoter reporter

3.4.1 Adapting the universal construct as a promoter reporter for *Agr2*

As mentioned, we wanted to create a promoter reporter for the *Agr2* gene that did not affect the endogenous loci, and which would express Cre recombinase as part of the reporter to enable lineage tracing. The strategy for ectopic expression of eGFP-Cre from an Agr2 BAC promoter is summarised in Figure 3.7A. The mouse *Agr2* gene spans eight exons (7 coding) over around 11 kilobases on chromosome 12 in the mouse genome, and encoding a protein of 175 amino acids (Figure 1.6). As in human *AGR2* (*hAGR2*), the first coding exon in mouse *Agr2* is exon 2, which contains the start codon. We therefore chose homologous arms of 250bp upstream and downstream of the start codon in exon 2 to target our universal construct into the BAC by recombineering. We chose the 85 kilobase *Agr2* BAC from the 129S7/AB2.2 clone library which encompasses the entire Agr2 gene as well as approximately 30 kilobases upstream and 43 kilobases downstream of the gene, but does not include the closely neighbouring *Agr3* gene.

The entire *Agr2* BAC was electroporated into SW102 bacteria, and selected for with chloramphenicol, the in-built antibiotic resistance gene in the BAC backbone. SW102 is a strain of recombineering-ready *E. Coli* bacteria as described in the introduction to this chapter. Once the BAC was successfully placed into these bacteria, the pMULTIrec vector with *Agr2* homologous arms was electroporated in and the bacteria were placed at 42°C for 15 minutes

Figure 3.7: Agr2 GFP-Cre promoter reporter

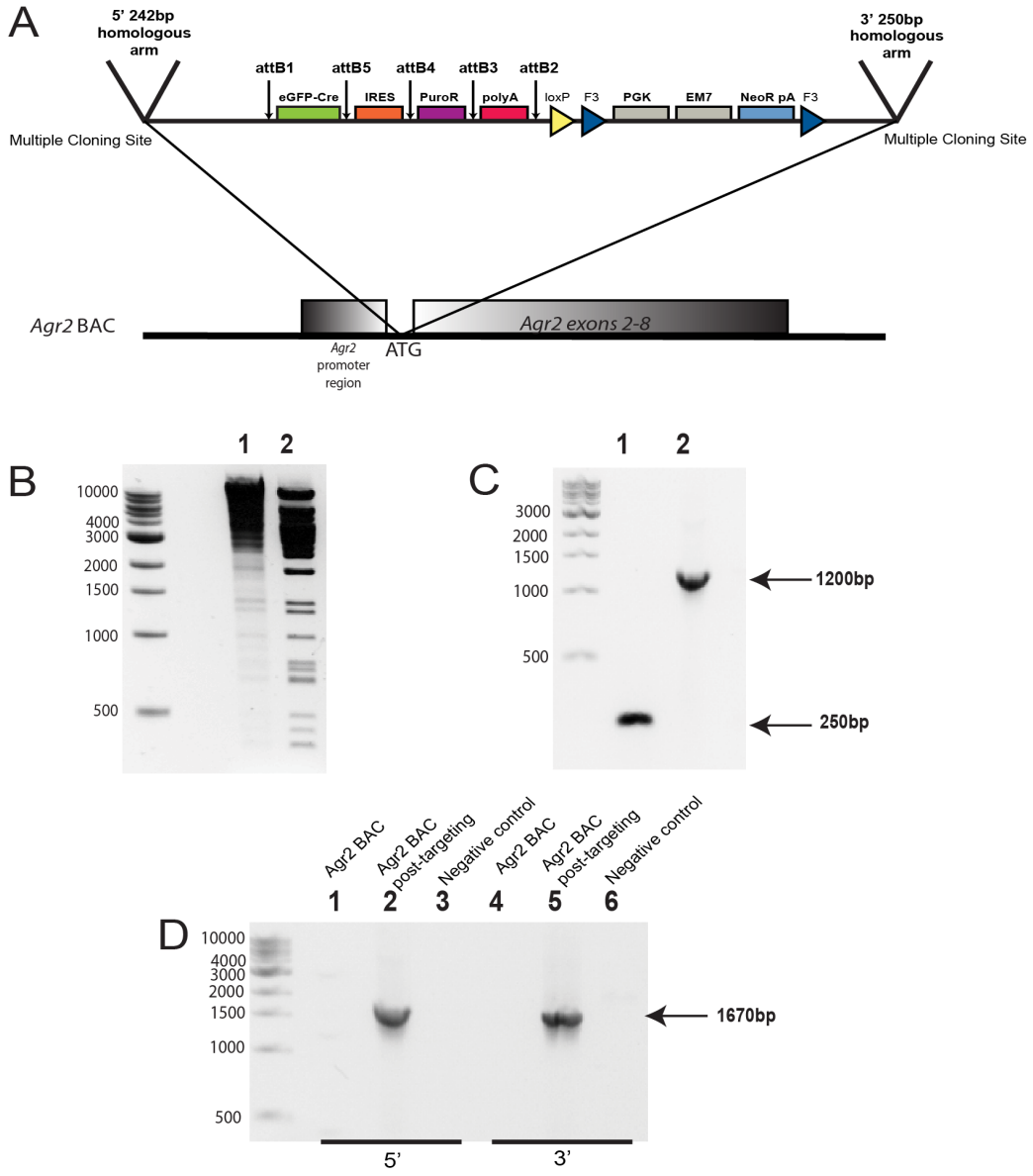


Figure 3.7 Agr2 GFP-Cre promoter reporter - 250bp homologous arms to the 5' and 3' region of the Agr2 gene start codon were cloned into the universal expression construct upstream and downstream of the expression cassettes as shown (A). The construct was targeted by recombineering to the start codon of the Agr2 gene on a BAC containing the entire Agr2 gene. Correct targeting to the BAC was confirmed by restriction endonuclease digest of the entire BAC either before targeting (B-1) or after targeting (B-2), and by amplification of a 1670bp PCR product containing both BAC and construct sequences at the 5' end of the trargeted construct (D-1,2,3), and the 3' end of the construct (D-4,5,6). The loxP site in the backbone of the BAC was replaced with an ampicillin resistance cassette by recombineering and replacement verified by amplification of the region (with loxP 250bp, with replacement by amp 1200bp) either before recombineering (C-1) or after recombineering (C-2)

to activate the recombineering genes. Successfully recombineered clones were selected for on kanamycin and chloramphenicol and then screened for correct integration events (i.e. at the start codon of the *Agr2* gene). Clones were *Hin*dIII digested to reveal a classic BAC pattern, and comparison of pre-and post-recombineered clones shows a shift in some bands of the digest pattern (Figure 3.7B) as would be expected with the integration of a piece of DNA into the BAC, but with most bands remaining the same height indicating that no unexpected large changes had occured.

Most commercial BACs have a single loxP site built into their backbone. Close inspection of the universal vector reveals another, single loxP site within the transgenic region (just upstream of the F3 site). This is a problem, since expression of the Cre-recombinase enzyme would lead to excision of the region between these loxP sites, removing a large portion of the universal vector (including the section containing the PGK and EM7 promoters and neomycin/kanamycin resistance genes). For this reason, we needed to remove the loxP site in the backbone of the BAC by replacing it with another selectable marker, ampicillin. Again, we used recombineering to swap the loxP site DNA with our own cassette containing the ampicillin resistance gene, and then screened for successful recombinants using PCR. Primers (Materials and Methods Table 2.5) were designed to the flanking regions of the loxP site. Before recombineering the amplicon size was 250bp, and after recombineering it was 1200bp (Figure 3.7C). Finally, to ensure that the integration of the universal vector GCiP was at the correct locus in the BAC, primers were designed (Materials and Methods section 2.4.6, Table 2.5) for the 5' and 3' ends of the recombination region, with one primer for each from outside the recom-

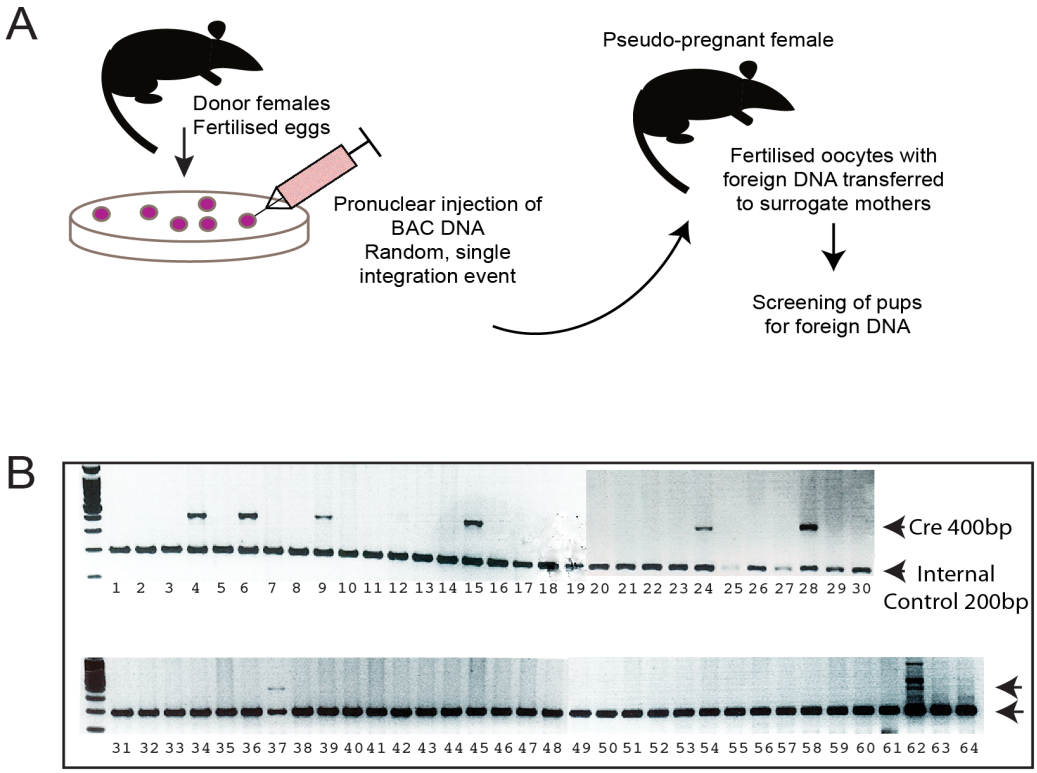
bineered vector region, and one from within. Thus, only correctly integrated clones should be able to amplify the correct size amplicon by PCR (Figure 3.7D). The clone chosen for use in making transgenic mice was tested in each of the three ways described, and Figures 3.7A,B,C were all made from analysis of the same clone.

3.4.2 Creation of the *Agr2* transgenic line "AGR2PR(GFPCre)"

The recombineered *Agr2* BAC clone with start codon in exon 2 replaced with the GCiP vector was ultra-purified using the large genomic construct maxiprep kit from Qiagen, and then used for microinjection into fertilised mouse oocytes (CBA/c57bl/6J F1 x CBA/c57bl/6J F1) at a concentration of 0.5 ng/μl. All microinjections were carried out by Paul Devenney from the MRC Human Genetics Unit. These fertilised oocytes were then transferred to pseudo-pregnant females as surrogates (CD1 mice) (summarised in Figure 3.8A). The pups were weaned and earclipped at around 2 weeks of age, and earclips were used for genotyping. In total, from 2 micro-injection sessions, there were 64 pups born to 4 surrogates. Genotyping was carried out using PCR on genomic DNA from the earclips, with primers (see Materials and Methods) designed to amplify the *Cre* recombinase transgene (400bp), together with an internal control for ubiquitous housekeeping gene *Fabp1* (200bp). Of the 64 pups we identified 8 founder mice; 3 male and 5 female (Figure 3.8B).

The table in Figure 3.8C summarises the breeding of these founders. To begin with, we tested the male founders. Males 6, 9 and 15 were set up with female CD1 mice as breeding partners, and offspring were tested for germline transmission of the transgene using the same genotyping PCR as for

Figure 3.8: Creation of Agr2 promoter reporter line AGR2PR(GFP-Cre)



C

Founder	Sex	Germline Transmission	Average Litter Size (xCD1)	Average Transmission Rate/Litter (%)
4	F	N/A	N/A	N/A
6	M	Y	15.5	30.0
9	M	Y	2.7	12.5
15	M	N	16.5	0
24	F	N/A	N/A	N/A
28	F	N/A	N/A	N/A
37	F	N/A	N/A	N/A
62	F	Deceased	N/A	N/A

Figure 3.8 Figure 3.8 Creation of Agr2 promoter reporter line AGR2PR(GFP-Cre) - (A) Agr2 BAC DNA containing the expression cassette at the start codon was introduced into fertilised oocytes from donor females by pronuclear injection, and injected oocytes were transferred to pseudo-pregnant females. (B) Genomic DNA from pups was screened by PCR analysis for presence of the foreign transgene Cre from the eGFP-Cre expression cassette (400bp) with internal control Fabpi for PCR efficiency (200bp). (C) Three founder males and 5 founder females were identified from 64 live pups. These were tested for fertility with CD1 females and their litters were analysed for germline transmission rate. Founder male 6 was selected for transgenic analyses of Agr2-driven GFP-Cre expression.

the founders. CD1 mice are often used especially for embryo analysis as they tend to produce large litter sizes compared to other strains such as C57BL/6. Male 15 produced average sized litters when bred with a CD1 strain female, however, none of his offspring yielded transmission of the *Cre* recombinase transgene across 3 litters with 3 different CD1 females (see Figure 3.8C). We thus concluded that the integration event happened after the germline lineage was established in this oocyte, making this male unsuitable for our purposes. Breeding of male 9 did produce transgenic offspring, however, the litter sizes were extremely small (Figure 3.8C), with only 2 or 3 pups per litter. It may be that the recombineered BAC construct integrated at a site involved in fertility, or it is possible that this male had a fertility issue unrelated to the transgenic addition but either way, it made this male less than ideal for transgenic analysis. Founder male 6 was also bred with CD1 females, resulting in average litter sizes and germline transmission to his offspring. We therefore decided to use this male for our analyses.

3.4.3 Analysis of transgenic mice for eGFP-Cre expression

Male 6 was set up with two CD1 females for analysis of embryos and pups. We first wanted to establish that our BAC transgenic vector was expressing eGFP-Cre as expected. AGR2 is known to be expressed at clearly detectable levels in the gut of both humans [Thompson and Weigel, 1998] and mice, and specifically in the goblet cells of the mouse intestine [Park et al., 2009], so we looked for expression of GFP in this location to confirm that the promoter reporter was faithfully expressing as it should. One female was permitted to

litter down and produced 13 pups, three of which were clearly transgenic based on PCR-based genotyping (Figure 3.9A, lower panel). The three positive pups and one negative were euthanised at 6 weeks of age, and the entire intestine removed and fixed. These were then treated and embedded in paraffin blocks for sectioning. The sections were processed for immunohistochemical staining (Materials and Methods section 2.8) and incubated with anti-GFP antibody overnight (Abcam 6556 rabbit polyclonal), and then incubated with Alexa-488 donkey anti-rabbit secondary antibody. Sections were then mounted in DAPI-containing mounting fluid and analysed both by widefield and confocal fluorescence microscopy. Some fluorescence was observed at the base and along the length of the villi (Figure 3.9A right panel), in accordance with the known location of goblet cells [Johnson et al., 2012], however, the signal was very weak and not clear. In contrast, no signal was seen in a negative control from the same litter. In parallel, we decided to analyse the embryonic kidney for expression of GFP in transgenic embryos. The reason for this will be discussed further in subsequent chapters, but briefly, antibody staining and biochemical data implied we should find expression of our transgene at the apical side of the ureteric bud and in the cap mesenchyme of early (E12.5) embryonic mouse kidneys. Male 6 was set up with a CD1 female mouse and the pregnant female was euthanised at E12.5. The embryos were dissected out and tails were kept for genotyping, whilst the kidneys were removed and placed into culture overnight.

In total, there were 13 embryos from this female, and genotyping confirmed that 5 were definitely transgenic (Figure 3.9B, lower panel). After 24 hours in culture, the kidneys from both positive and negative embryos

Figure 3.9: Analysis of Agr2 GFP-Cre promoter reporter for GFP expression in adult gut and cultured embryonic kidneys

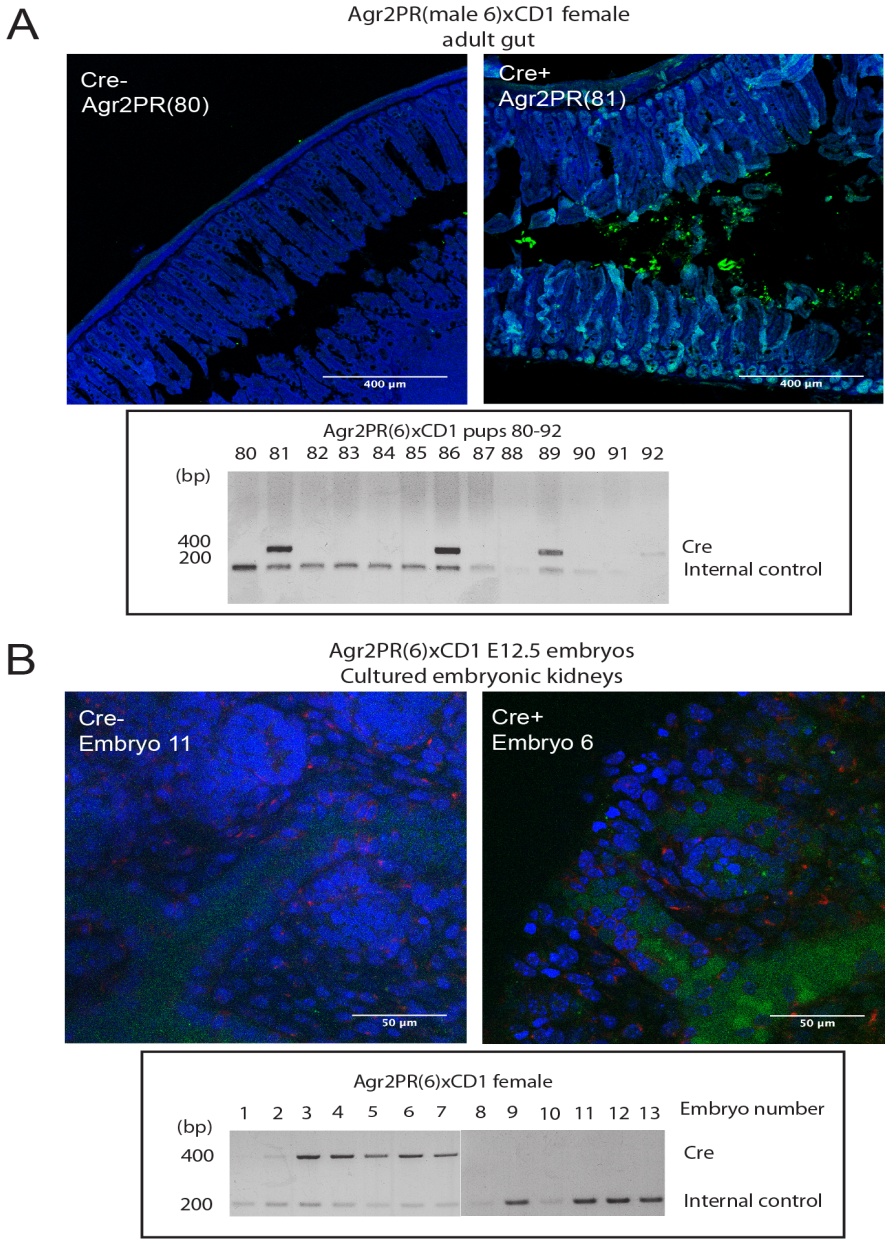
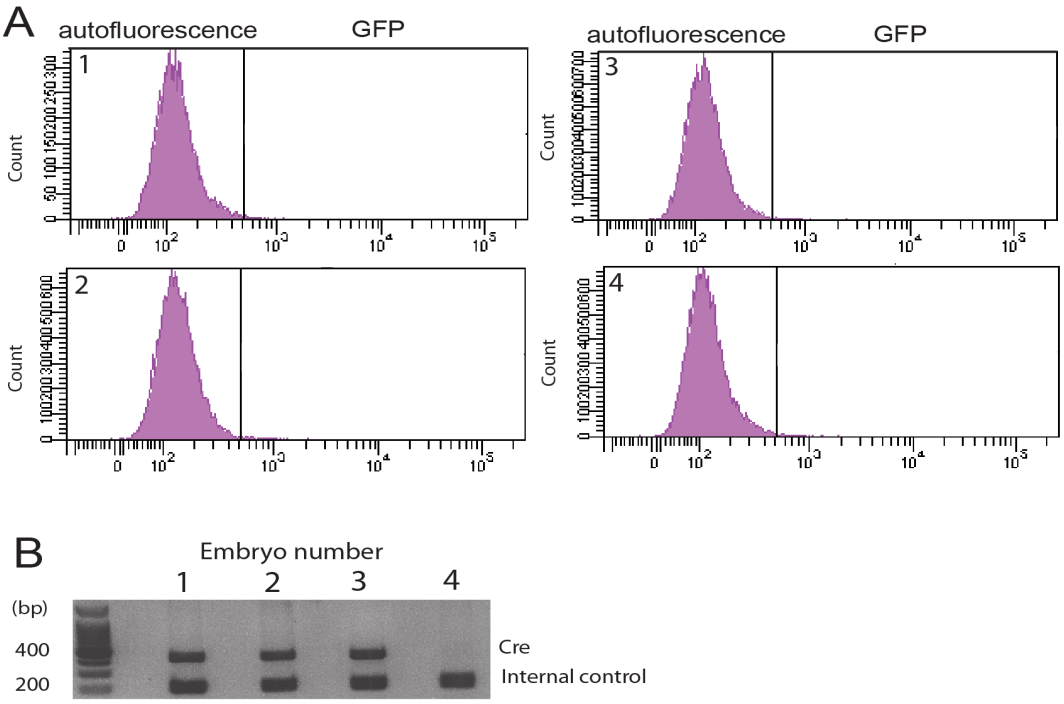


Figure 3.9 Analysis of Agr2PR(GFP-Cre) transgenic promoter reporter line for GFP expression in adult gut and cultured embryonic kidneys - A litter from Agr2PR(GFP-Cre) was earclipped and genotyped for presence of Cre (A, bottom panel). Intestines from pups positive for Cre and one negative were paraffin-embedded and sectioned. Sections were analysed for expression of GFP by IHC using anti-GFP polyclonal antibody (Sigma) (A). Embryos were removed at E12.5 from a CD1 cross with Agr2PR(GFP-Cre) founder male 6, genotyped from tail tissue, and kidneys were microdissected and placed in culture for 24h. Kidneys from embryos positive for Cre and one negative (B, bottom panel) were analysed for native GFP expression by confocal microscopy (B).

were fixed in 4% paraformaldehyde (PFA), and mounted with DAPI-containing mounting fluid onto glass slides for analysis by confocal microscopy. Representative images taken from positive embryos (Figure 3.9B right panel) again showed a weak GFP signal, in the region roughly corresponding to the ureteric bud, and diffuse throughout the cell as expected. However, in order to obtain detectable fluorescence the gain needed to be increased to higher than normal levels. Nevertheless, co-excitation of the specimen at a different wavelength (594 Texas Red) did not produce a red signal indicating that the emission signal captured from the 488 excitation is not simply autofluorescence from the tissue, and the negative control did not produce any detectable green signal when excited at the same levels (Figure 3.9B, left panel), suggesting that the GFP-Cre may be expressing correctly from the BAC expression vector although at very low levels.

We decided to take another approach to analyse the tissue, using flow cytometry. Again, male 6 was set up with a female CD1 mouse, and embryos were harvested at E12.5. Kidneys were dissected out and treated to make them into a single-cell suspension, whilst tails were taken from each corresponding embryo for genotyping (Figure 3.10B). In total, three positive transgenic embryos and one negative embryo were analysed. Flow cytometry was carried out by E Freyer, a technician at the MRC Human Genetics Unit. The parameters for exclusion of autofluorescent signal were set using the negative control, and dead or dying cells were also exlcuded from analysis. Flow cytometric analysis on the four embryonic kidneys concluded that there was no difference between the negative (embryo 4) and the positives (embryos 1-3), confirmed by the number of GFP-postive cells counted as a percentage of the total number of cells analysed (Figure 3.10 A, C).

Figure 3.10: Flow cytometry analysis of Agr2 promoter reporter (AGR2PR(GFP-Cre) X CD1 wt) cultured embryonic kidneys



C

Embryo number	Genotype	GFP + cells/Total cells counted (%)
1	Cre+	0.3
2	Cre+	0.4
3	Cre+	0.4
4	Cre-	0.4

Figure 3.10 Flow cytometry analysis of Agr2 promoter reporter (AGR2PR(GFP-Cre) X CD1 wt) cultured embryonic kidneys - Embryos were removed at E12.5 from a CD1 cross with Agr2PR(GFP-Cre) founder male 6, genotyped from tail tissue (B), and kidneys were microdissected. Kidneys positive for Cre and one negative were dissociated into single-cell suspension and analysed for GFP by flow cytometry (A) using the negative embryo for thresholding. There was no difference in the percentage of live GFP positive cells between the negative kidney and the three positives (C).

3.4.4 Cross to Rosa26YFP Cre reporter mouse for further expression analysis

The data above suggested that if there was expression of the GFP-Cre transgene in the Agr2 promoter reporter mouse line from male 6, Agr2PR(GFPCre)(6), it is very weak expression. Crossing to a Cre-driven reporter mouse strain should activate constitutive expression of a reporter (such as YFP, or tdRFP as was used in the Six2 promoter reporter analysis, above). A Cre-driven reporter mouse strain is one that contains an ectopic expression cassette for a reporter such as YFP or lacZ. The reporter expression cassette is not expressed under normal circumstances as there are several stop codons and polyA signals between its promoter and the start of the reporter sequence. The stop codons and polyA signals are flanked by loxP sites (sequences of DNA that are recognised by the Cre-recombinase enzyme), and expression of the Cre-recombinase enzyme causes excision of the stop codon, allowing constitutive expression of the reporter. Thus, only in the presence of Cre-recombinase is it possible for the reporter to be expressed. One such Cre-driven reporter mouse strain that we had access to is the Rosa26YFP Cre reporter mouse, in which a YFP expression cassette is harboured in the Rosa26 locus (Figure 3.11). As previously mentioned, the Rosa26 locus in the mouse is known to be a safe locus for harbouring ectopic expression constructs [Friedrich and Soriano, 1991] [Zambrowicz et al., 1997], and does not lead to off-target effects that could interfere with normal gene function. Stop codons and several polyA signals between the endogenous *Rosa26* promoter and the YFP expression cassette prevent expression of YFP until it is activated by excision of the stop codon and polyAs, allowing the full YFP cDNA to be transcribed.

Figure 3.11: Part I - Analysis of Agr2PR(GFP-Cre) X YFP reporter line Rosa26YFP

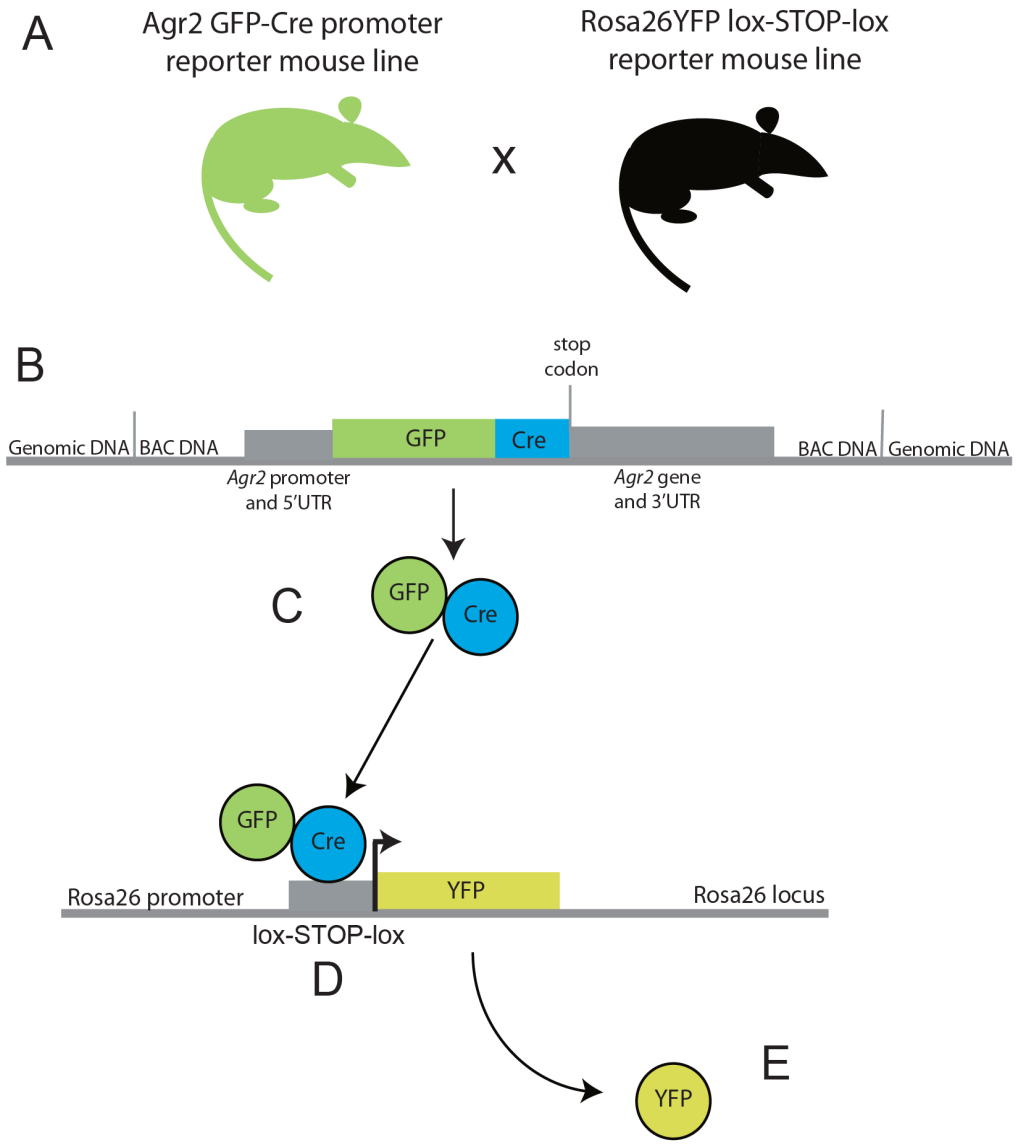


Figure 3.11 Part I - Analysis of Agr2PR(GFP-Cre) X YFP reporter line Rosa26YFP Agr2PR(GFP-Cre) male founder 6 was crossed to a Rosa26YFP homozygous female (A). The GFP-Cre recombinase fusion protein (C) is expressed under the control of the Agr2 promoter on the Agr2 BAC (B). Presence of the Cre-recombinase enzyme in cells expressing Agr2 causes permanent excision of the STOP cassette between the YFP cassette and the Rosa26 promoter (D), and the YFP protein can be expressed (E). The YFP signal will be present in all daughter cells.

By crossing the Agr2PR(GFPCre)(6) mouse to a homozygous female Rosa26YFP Cre reporter mouse (Figure 3.11A), any mice transgenic for the Agr2 promoter reporter construct (Figure 3.11B) will be able to activate expression of YFP (Figure 3.11C,D) but only in the tissues expressing Cre recombinase and crucially, only in cells derived from the original Agr2-expressing cells with the genomic YFP activated. In this way, expression of Cre should be enough to drive excision of the STOP codons and polyA signals that prevent expression of the YFP signal, and harbouring of the YFP expression cassette in the Rosa26 locus ensures that the genomic envirnonment or location is unlikely to affect the level of expression of the marker gene.

Agr2PR(GFPCre) founder male 6 was crossed to a Rosa26YFP Cre reporter mouse for analysis of embryos. The cross produced 6 embryos for analysis, and genotyping (Figure 3.12D) for presence of the Cre transgene confirmed that there was one positive embryo and 5 negative. Images were taken using a fluorescence microscope, first of the whole embryo (Figure 3.12A), and then of the whole embryo dissected rostro-caudally to reveal the inner tissues (Figure 3.12B brightfield microscopy, Figure 3.12C fluorescence microscopy). There was no fluorescence seen in any of the embryos, and no observable difference between the negative embryos (1-5) and the positive embryo (6). Kidneys were dissected out and placed into culture for 24 hours. Kidneys were then analysed for YFP expression using a time-lapse microscope for live-cell imaging with the help of Dr Nils Lindstrőm (Institute of Genetics and Molecular Medicine) (Figure 3.13A brightfield, 3.13B fluorescence microscopy). Again, there was no observable difference seen between the negative embryonic kidneys (1-5) and the positive kidney (6). For comparison, a *Six2* promoter reporter trans-

Figure 3.12: Part II - Analysis of Agr2PR(GFP-Cre) X YFP reporter line Rosa26YFP - E12.5 whole embryos

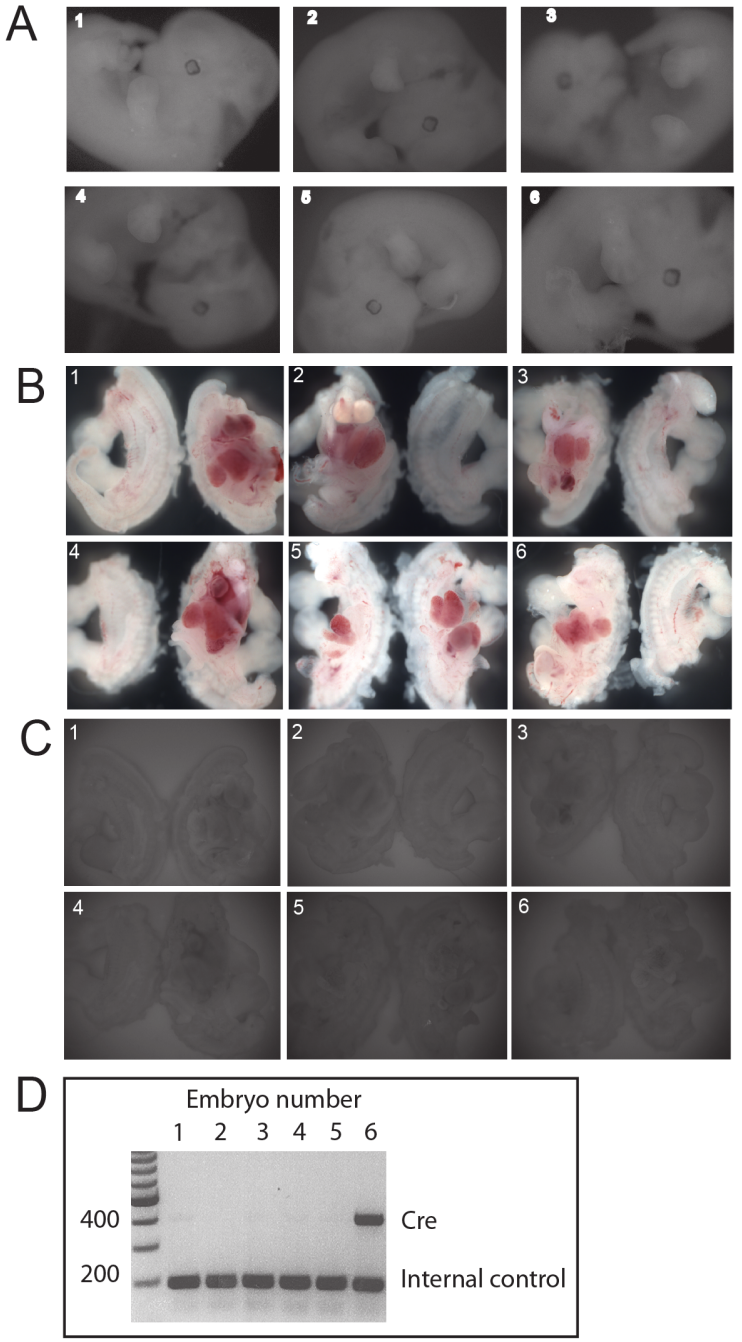


Figure 3.12 Part II - Analysis of Agr2PR(GFP-Cre) X YFP reporter line Rosa26YFP - E12.5 whole embryos
Agr2PR(GFP-Cre) male founder 6 was crossed to a Rosa26YFP homozygous female and E12.5 embryos were harvested, genotyped for presence of Cre (D), and analysed for expression of YFP whole (A) or dissected rostro-caudally (B, brightfield) (C, fluorescence).

genic kidney from the knock-in described previously in this chapter is shown in Figure 3.13D (top panel brightfield, lower panel fluorescence microscopy). This image was taken by Dr Nils Lindstrőm using the same microscope as was used for the kidneys from Agr2PR(GFPCre) x Rosa26YFP Cre reporter cross in Figure 3.13.

Due to the disappointing result from the YFP cross embryos, we decided to test adult tissue for expression of the transgene by western blot and FACS analysis. As previously mentioned, the mature gut is a known location for AGR2 expression. We decided to test both the adult gut and adult kidney for expression of GFP. The anti-GFP monoclonal antibody from Abcam (1218) is predicted to cross-react with YFP protein, and so we should be able to detect both GFP and YFP variants if they are being expressed in the transgenic mice. Agr2PR(GFPCre) founder male 6 was crossed to a new female Rosa26YFP Cre reporter mouse, and the female was permitted to litter down, producing a litter of 6 pups. Pups were earclipped and genotyped (Figure 3.14A), again with one positive transgenic mouse and 5 negative mice. The positive pup (pup 3) and one negative pup (pup 1) were euthanised at 8 weeks of age and the entire intestine and both kidneys were dissected out. A 3cm portion of the upper colon from each mouse was removed for FACS analysis, and the rest of the intestine and both kidneys were placed in 0.32M sucrose overnight for subsequent protein extraction. After saturation of the tissue in sucrose buffer, the tissues were homogenised using a rotor-stator homogeniser, and protein lysate was isolated by spinning the suspension at high speed. The lysates were analysed by SDS-PAGE and western blot. Adult gut and kidney lysates from the positive transgenic littermate (pup 3) were analysed for expression of AGR2

Figure 3.13: Part III - Analysis of Agr2PR(GFP-Cre) X YFP reporter line Rosa26YFP - E12.5 embryonic cultured kidneys

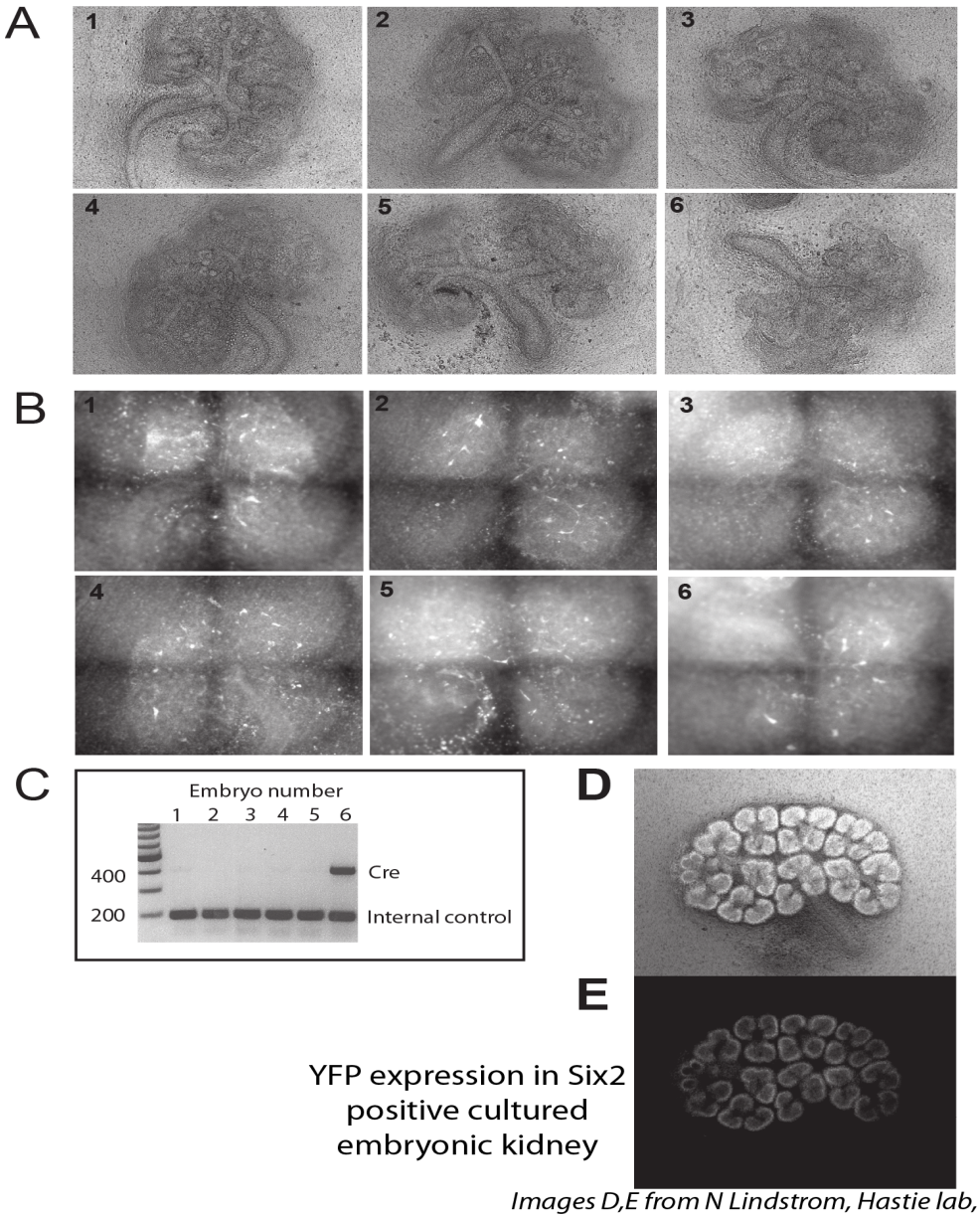


Figure 3.13 Part III - Analysis of Agr2PR(GFP-Cre) X YFP reporter line Rosa26YFP - E12.5 embryonic cultured kidneys
Agr2PR(GFP-Cre) male founder 6 was crossed to a Rosa26YFP homozygous female and E12.5 embryos were harvested and genotyped for presence of Cre (C). Kidneys were micro-dissected and analysed for expression of YFP (A, brightfield) (B, fluorescence). For comparison, E12.5 embryonic kidneys from the *Six2*$^{+GCiP}$ (section 3.3.2) crossed to a Rosa26YFP female are shown in brightfield imaging (D) and fluorescence imaging (E).

(Figure 3.14B) using anti-AGR2 polyclonal antibody K47 (Moravian Biotech). As expected, AGR2 protein expression was observed in the adult murine gut by western blot, as well as the murine kidney. It is not possible to compare protein levels using this method of protein extraction since the concentration of protein in the lysate cannot be accurately be determined due to the presence of other factors such as blood serum. In this case we merely wished to confirm that AGR2 protein was in fact being expressed in these tissues, before looking for expression of a transgenic protein under the control of the Agr2 promoter. Having confirmed the presence of AGR2 protein in both the adult gut and kidney of the mouse, we tested the lysates from kidney and gut of both the positive and negative littermates for presence of GFP-Cre or YFP. GFP and YFP differ in only one amino acid and are thus both predicted to migrate at around 28 KDa, whereas GFP-Cre is predicted to migrate at approximately 60KDa. We used a positive control for the blot from cell lysate that had been transfected with eGFP empty vector (Clonetech pAcGFPN1), and loaded the maximum lysate possible from the positive and negative littermate tissues. Figure 3.14B (farthest right lane) shows the AGR2 levels detectable from this amount of lysate. Blots were probed for GFP/GFP-Cre or YFP using abcam anti-GFP as described above, and a clear band at 28KDa was observed in the control lane. There were no bands detected at the correct heights (28KDa, 60KDa) in the lanes containing lysate from tissues from the positive littermate, and the only bands detectable were observed in both positive and negative littermates, at around 32 or 33 KDa (Figure 3.14C).

In parallel, the intestines dissected for FACS analysis were dissociated into single-cell suspension. These were analysed for GFP/YFP expression by

Figure 3.14: Part IV - Analysis of Agr2PR(GFP-Cre) X YFP reporter line Rosa26YFP - adult gut and kidney

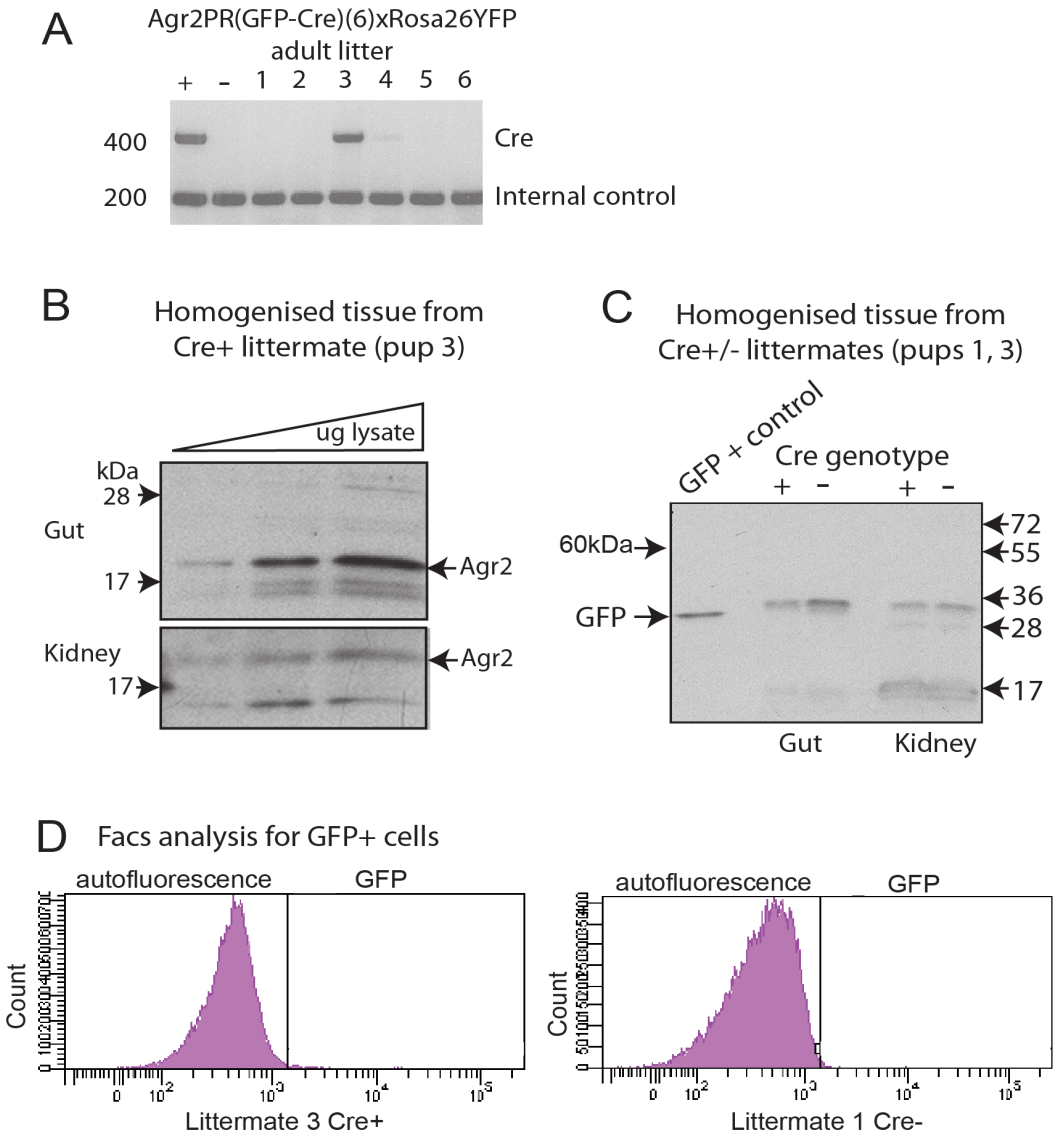


Figure 3.14 Part IV - Analysis of Agr2PR(GFP-Cre) X YFP reporter line Rosa26YFP – adult gut and kidney
Agr2PR(GFP-Cre) male founder 6 was crossed to a Rosa26YFP homozygous female and pups were earclipped and genotyped for presence of Cre (A) (*A Thornburn*). Gut and kidneys were dissected from one positive pup and one negative pup (pups 3 and 1) at 6 weeks of age, and tissue was homogenised in sucrose buffer. Gut and kidney homogenate was analysed by western blot for presence of AGR2 (B), and GFP-Cre (C). After dissection, some gut tissue from pups 1 (negative) and 3 (positive) was dissociated into single-cell suspension and analysed by flow cytometry, and no YFP positive cells were detected for either sample.

flow cytometry (Elisabeth Freyer, MRC Human Genetics Unit). Again, the parameters for exclusion of autofluorescence or dead cells were set according to the negative littermate, and cell clumps were also excluded. There was almost no GFP expression detectable, and no observable difference in GFP/YFP expression between the gut cells from the positive and the negative pups (Figure 3.14D). Taken together, we concluded from this data that there was very little expression, if any, of the GFP-Cre transgene driven by the ectopic BAC Agr2 promoter at least in the case of founder male 6. Of the female founders, one died as a result of husbandry error, and the remaining four females were set up for breeding after male 6 was extensively tested as described above, but all failed to produce any litters and were eventually culled.

3.5 Discussion

There are several possible reasons that we were unable to detect expression of the GFP-Cre transgene in these mice. Since we directly microinjected (Paul Devenney) the BAC expression construct into oocytes rather than endogenous targeting via ES cells, the integration event into the mouse genome is random. The mice were viable and seemed healthy, but in the case of male founder 9, there was a problem with fertility. The possibility that the transgene had integrated at a locus required for male fertility has already been discussed. In the case of male 6, although the integration event was not deleterious to the founder male or his offspring (that we are aware of), it is possible that the BAC vector integrated at a genomic location that is suboptimal for expression. For example the transgene may have been inserted into a "closed" region of the

genome that is silenced, and therefore not accessible to initiation of transcription machinery. Alternatively, the BAC we chose from the BAC library may not have contained all the components required for the promoter to activate transcription. For example, an enhancer region may be absent from the BAC, which can be anything from a few kilobases away to 1 megabase away [Lettice et al., 2002].

Rather than try to establish exactly where the transgene had integrated by plasmid retrieval and sequencing, we felt it was more productive to start with a new micro-injection session for random integration and also to target the endogenous locus in ES cells as described for the *Six2* knock-in. Work on this is ongoing, and will form the basis of future expression studies *in vivo*. The universal vector itself, however, has been used as a highly adaptable and flexible cloning and recombineering system for several in-house projects, and has been made widely available to other research groups via Addgene.

In the following chapter, another approach to investigating the biological function of Agr2 will be described, using phage-peptide proteomics as the basis for identifying biologically relevant interacting partners of AGR2.

Chapter 4

Identification of MKS3 as a putative novel AGR2 interacting protein

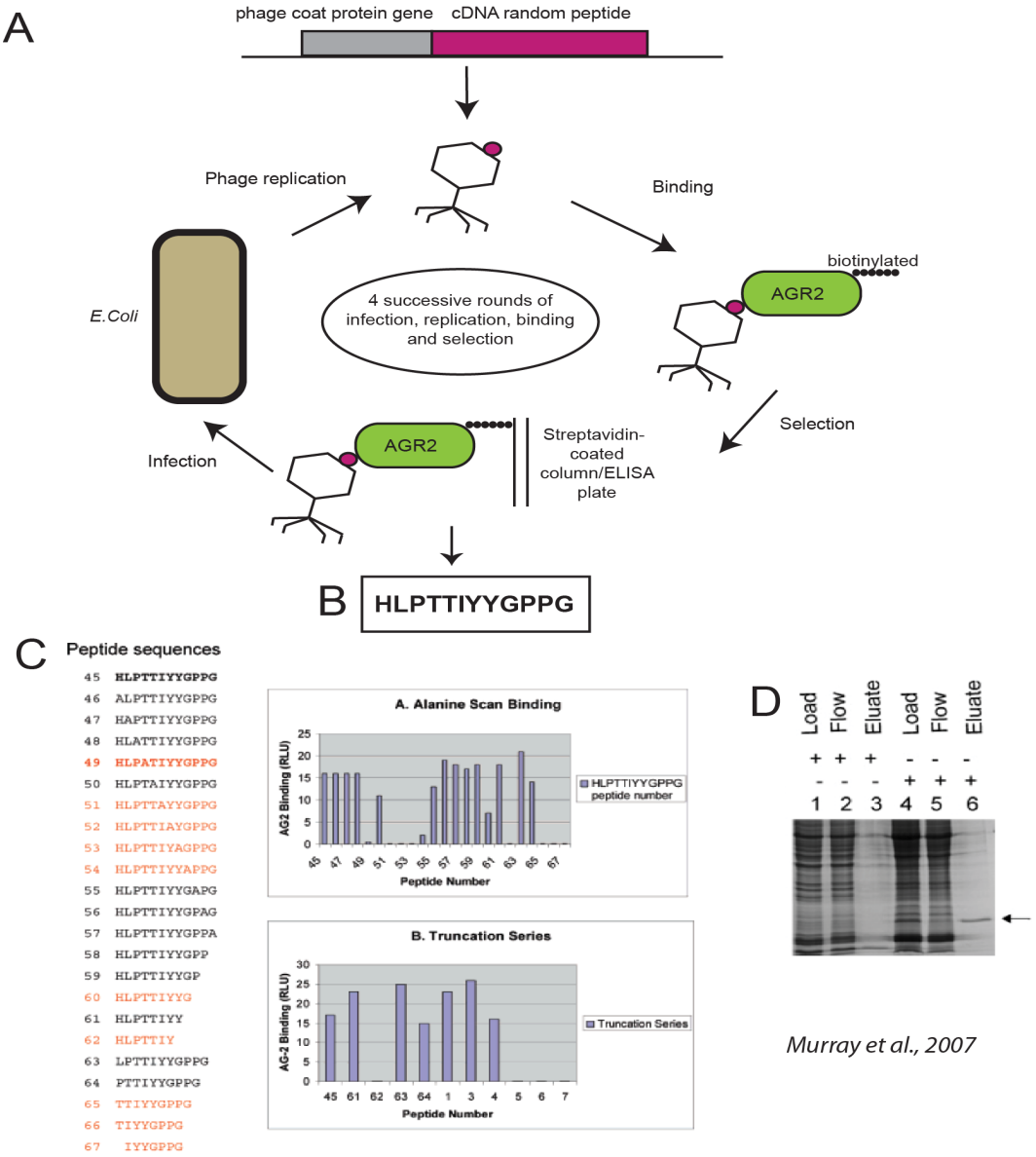
4.1 Introduction

As previously discussed, the biological function of Agr2 is not fully understood. It has been shown to be involved in a variety of cancers when upregulated or overexpressed [Liu et al., 2005] [Maresh et al., 2010] [Vanderlaag et al., 2010] [Zhang et al., 2005], and to be involved in mucin production [Park et al., 2009], contributing to inflammatory bowel disease when levels are reduced [Zheng et al., 2006] . Studies into the association of Agr2 with cancer point to a nuclear function and interaction with p53 [Pohler et al., 2004] [Maslon et al., 2010], yet the interaction with MUC2 protein through a disulphide bond [Park et al., 2009] implicates a role for Agr2 outside the nucleus. Agr2 has

a putative N-terminal leader sequence with cleavage site that destines AGR2 for the secretory pathway, and previous work has shown that the transfected "mature" form of Agr2 lacking the N-terminal leader sequence localises to the nucleus and cytoplasm when transfected into cells [Fourtouna et al., 2009]. Evidence points to dual or multi-faceted roles for Agr2 *in vivo*, and insights into its biological functions could help shed light onto the mechanisms by which Agr2 contributes to disease.

Identifying the interacting partners of a protein can aid in elucidating its biological function(s). Once an interaction is validated, its function and mechanisms of interaction can be investigated. Short linear peptide motifs on the surface of proteins or within disordered loops are known to provide binding sites for globular proteins [Neduva and Russell, 2005] [Neduva and Russell, 2006]. These linear peptide motifs can be an important tool in proteomics and the study of interactomes. Screening for linear peptide motifs that provide docking sites for a given protein can be achieved using various techniques, including phage peptide display [Smith, 1985] [Rossenu et al., 1997], as discussed in section 1.1.3. Briefely, phage peptide display involves the expression of a peptide on the major coat protein of a phage virion (normally M15 or filamentous phage fd [Smith and Petrenko, 1997]). Large combinatorial phage peptide display libraries, in which a single random peptide is expressed on the surface of each virion can be built up, and this library can be used to screen for potential binding peptides for a given protein or DNA. Using this approach (Figure 4.1A), a peptide was identified that bound AGR2 (Figure 4.1B), and was validated by various methods including ELISA based deletion and alanine scan (Figure 4.1C) [Murray et al., 2007].

Figure 4.1: Phage display identifies minimal peptide that binds to AGR2 in vitro



Murray et al., 2007

Figure 4.1 Phage display identifies minimal peptide that binds to AGR2 in vitro
(A) Phage peptide display: random cDNA peptide fused to bacteriophage coat protein is expressed on the surface of the phage virion. Some peptides bind biotinylated AGR2 bound immobilised on straptavidin plate and non-bound phage virion peptides are washed away. Bound virions are recovered and used to infect E.Coli, where binding phages are concentrated. Successive rounds of binding, selection, infection and replication enrich for peptides with high binding affinity. (B) AGR2 binding peptide HLPTTIYYGPPG was identified and the minimal binding peptide identified by alanine scan and truncation series analysis (C). Minimal binding peptide PTTIYY was used to precipitate AGR2 to near homogeneity from Barrett's oesophagus tissue crude lysate. (*Murray et al., 2007)*

Further analysis of the identified binding peptide revealed a degree of degeneracy in the sequence [Fourtouna et al., 2009], with certain amino acids or families of amino acids tolerated in substitution (discussed further, below). These data were used as the basis for database mining in the human genome in an attempt to uncover possible AGR2 binding proteins.

This chapter will focus on the interaction between AGR2 and the linear peptide identified by phage display, and validation of a potential interaction between AGR2 and a protein containing the linear peptide motif.

4.2 Linear peptide as basis for identification of AGR2 interactome

The AGR2-binding linear peptide HLPTTIYYGPPG was identified using a combinatorial phage peptide library as described above. The essential binding residues PTTIYY were determined using a series of deletion and alanine mutants (Figure 4.1 C) and the dissociation constant (Kd) was determined to be in the region of 30μM. AGR2 was previously shown to be highly overexpressed in Barrett's epithelium [Pohler et al., 2004], a pre-neoplastic condition that can lead to adenocarcinoma of the oesophagus. In order to test the ability of the linear peptide PTTIYY to bind native AGR2, cell lysates from Barrett's epithelia were washed over a streptavidin column containing biotinylated peptide. The peptide was able to purify native AGR2 from Barrett's epithelium to near homogeneity (Figure 4.1D) [Murray et al., 2007]. In contrast, AGR2 was not purified from normal squamous oesophageal epithelia under the same conditions, and the peptide did not pull down any other protein in this assay,

demonstrating specificity.

4.2.1 Expression of AGR2 in cancer cell lines as tools for investigation

Using this data, the ability of the peptide to bind native AGR2 in cell culture was evaluated. To investigate this, the cell lines to be used in this and subsequent experiments were characterised. AGR2 has been shown to be over-expressed in a variety of cancers, including several hormone-related cancers such as breast cancer, some ovarian cancer subtypes and prostate cancer, as well as other hormone-independent cancers [Brychtova et al., 2011]. To study the biology of AGR2 in cell cuture, we acquired a panel of cancer cells that express varying levels of AGR2. MCF7s are a human estrogen receptor alpha (ERα)-positive breast adenocarcinoma cell line that have been used extensively worldwide for research purposes [Soule et al., 1973] [Lacroix and Leclercq, 2004]. These cells originally had a karyotype of 85 chromosomes, however the MCF7s in use today have a karyotype of 69 chromosomes, and studies have shown divergence within MCF7 cell lines from different sources [Villalobos et al., 1995] [Liscovitch and Ravid, 2007], highlighting that not all cell lines of a given type are necessarily identical. We acquired two MCF7 lines from different sources ("American" and "European"), and these were tested for expression of AGR2 both at the mRNA level (not shown) and by Western blot using two different antibodies to AGR2; polyclonal anti-AGR2 K47 from Moravian Biotech (Figure 4.2A), and anti-AGR2 monoclonal from Abnova (Figure 4.2B).

Figure 4.2: Expression of the pro-oncogenic protein AGR2 in two human MCF7 breast cancer cell lines and human melanoma cell line A375

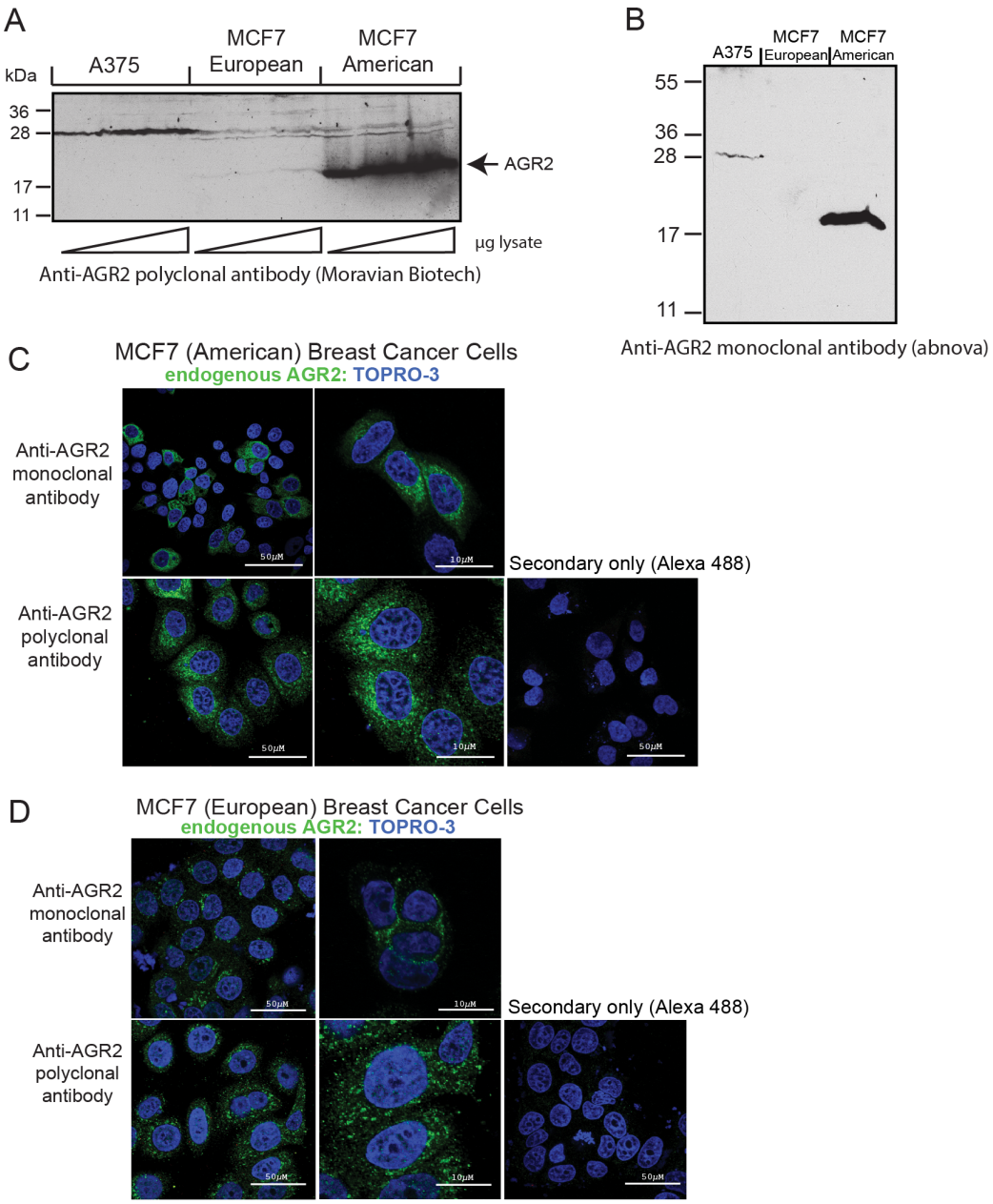


Figure 4.2 Expression of the pro-oncogenic protein AGR2 in two human MCF7 breast cancer cell lines and human melanoma cell line A375
Expression of endogenous AGR2 in MCF7 (European) and MCF7 (American) human breast cancer cell lines, and in human melanoma cell line A375 analysed by western blot using (A) rabbit polyclonal anti-AGR2 from Moravian Biotech or (B) mouse monoclonal anti-AGR2 from Abnova. Expression of endogenous AGR2 by immunofluorescence using polyclonal and monoclonal antibodies in American MCF7 breast cancer cell line (C) or European breast cancer cell line (D).

The results demonstrate that one MCF7 cell line (American) expresses AGR2 at very high levels, and in comparison, a second MCF7 breast cancer cell line (European) expresses much lower levels of the protein. A third cell line, human malignant melanoma cell line A375, was also tested and found to be absent for AGR2 expression at both the mRNA (not shown) and protein levels, making this a useful tool for analyses. The expression of AGR2 was also analysed in both MCF7 cells lines by immunofluorescence using both the anti-AGR2 polyclonal and anti-AGR2 monoclonal antibodies (Figure 4.2C, D). The localisation of endogenous AGR2 is somewhat different in the two MCF7 cells lines. AGR2 localised to both the endoplasmic reticulum (ER) and cytoplasm with a small amount in the nucleus in American MCF7 cells. Although co-localisation with ER markers was not shown in these figures, the ER association of AGR2 has been reported previously using co-localisation with ER markers by immunofluorescence [Fourtouna et al., 2009]. In addition, the distribution pattern of AGR2 mainly in the ER and with some localisation to the cytoplasm and nucleus is also in agreement with previous data [Fourtouna et al., 2009]. In contrast, the localisation of AGR2 in European MCF7 cells appears less diffusely distributed, and is also at lower levels, in agreement with the protein blot described above. In order to evaluate the specificity of the anti-AGR2 polyclonal antibody by immunofluorescence, high-Agr2-expressing MCF7 cells were transfected with siRNA to AGR2 and analysed by immunofluorescence and western blot after 72h incubation, using anti-AGR2 polyclonal antibody K47. AGR2 was knocked down in this cell line as seen by western blot and by immunofluorescence, compared to control siRNA or untransfected controls, validating the polyclonal antibody for use by immunofluorescence (Figure 4.3).

Figure 4.3: Knockdown of AGR2 using AGR2-specific siRNA in MCF7 (American) breast cancer cells

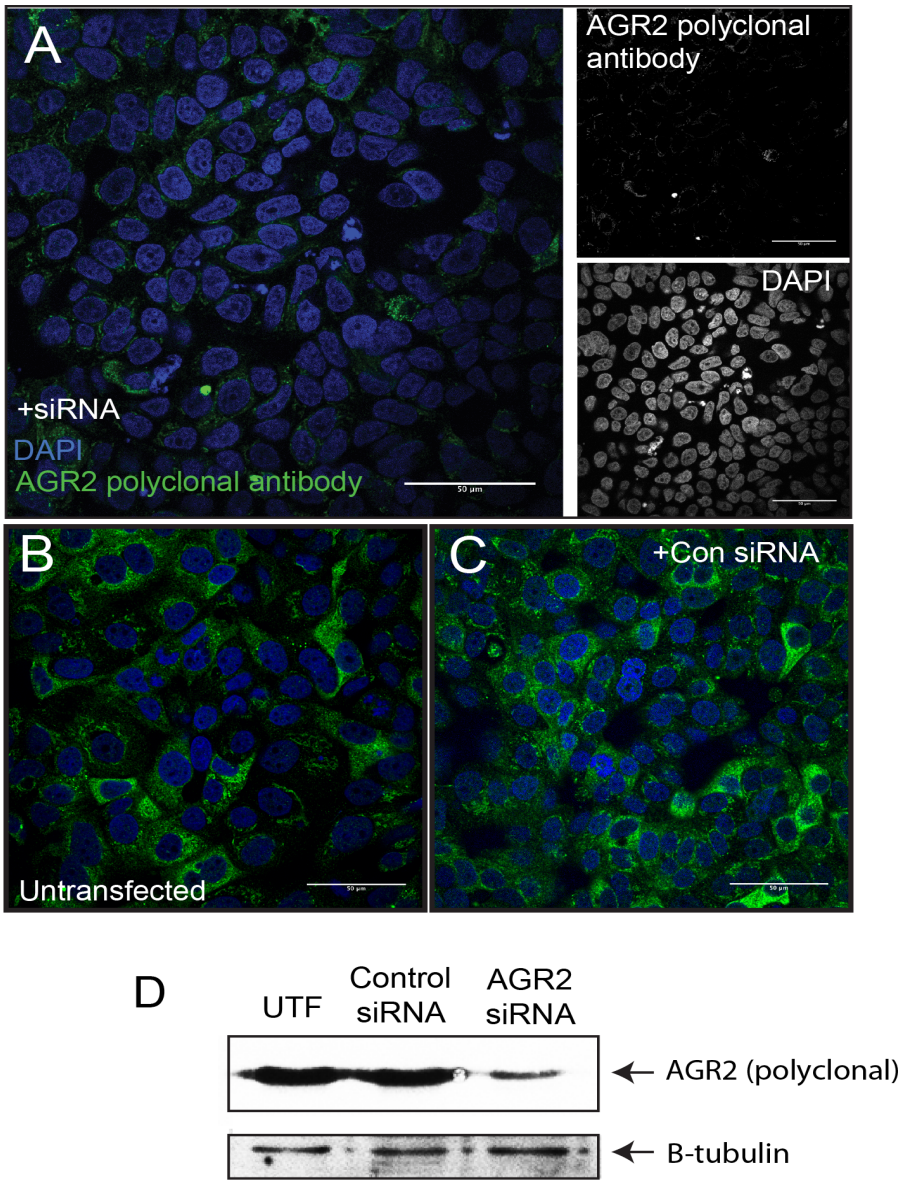


Figure 4.3 Knockdown of AGR2 using AGR2-specific siRNA in MCF7 (American) breast cancer cells
MCF7 (American) breast cancer cells were treated with either siRNA to AGR2 (A), control siRNA (C), or not transfected (B), and analysed by immunofluorescence using rabbit polyclonal anti-AGR2 antibody (Moravian Biotech). Partial knockdown of AGR2 was confirmed by western blot using the same antibody (D).

4.2.2 Linear Peptide Motif Interaction with AGR2

In order to evaluate the ability of the hexapeptide PTTIYY to bind native AGR2 in cell culture, MCF7 (American, high AGR2-expressing cells) were tranfected with GFP-aptamers (GFP-hexapeptide fusion) as shown in Figure 4.4A. Three cloned GFP-tagged aptamers (gifted by Argyro Fourtouna, Hupp lab) as shown were tested for their ability to bind AGR2; GFP-PTTIYY representing the minimal hexapeptide, GFP-PTTIYA mutant, and GFP-STOP-PTTIYY in which a stop codon was inserted before the PTTIYY generating native GFP expression. The PTTIYA mutant peptide had previously been shown to reduce binding to AGR2 *in vitro* by ELISA [Murray et al., 2007]. The constructs were transiently transfected into MCF7 (American) cells and incubated for 24 hours before fixing and analysis by immunofluorescence. A polyclonal antibody to AGR2 (Moravian Biotech, K47) was used to visualise endogenous AGR2 in these cells, and colocalisation to expressed GFP was evaluated by confocal microscopy.

In MCF7 (American) cells transfected with the GFP-stop codon-PTTIYY aptamer, AGR2 was seen to localise primarily to the endoplasmic reticulum (ER) as previously reported [Fourtouna et al., 2009] whilst expressed GFP was seen to accumulate in the nucleus as is commonly seen with untagged, expressed GFP (Figure 4.4). Cells transfected with GFP-PTTIYA aptamer were similarly seen to accumulate expressed GFP in the nucleus, with AGR2 localising to the ER as expected. In contrast, in cells transfected with the GFP-PTTIYY aptamer, GFP was seen to partially co-localise with AGR2 in the ER, with less GFP accumulation in the nucleus.

Figure 4.4: Partial co-localisation of GFP-tagged peptides with AGR2 in MCF7 breast cancer cells

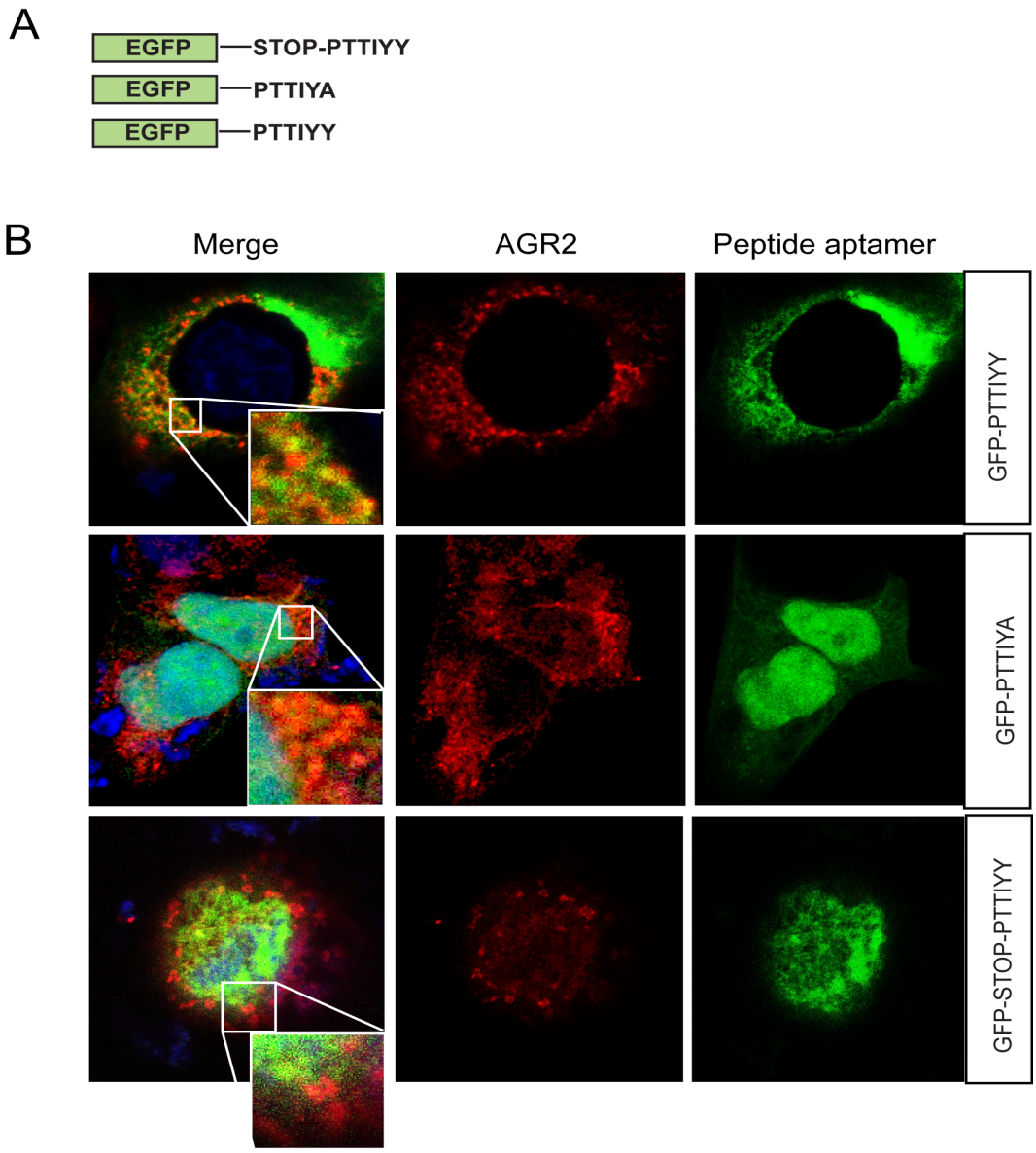
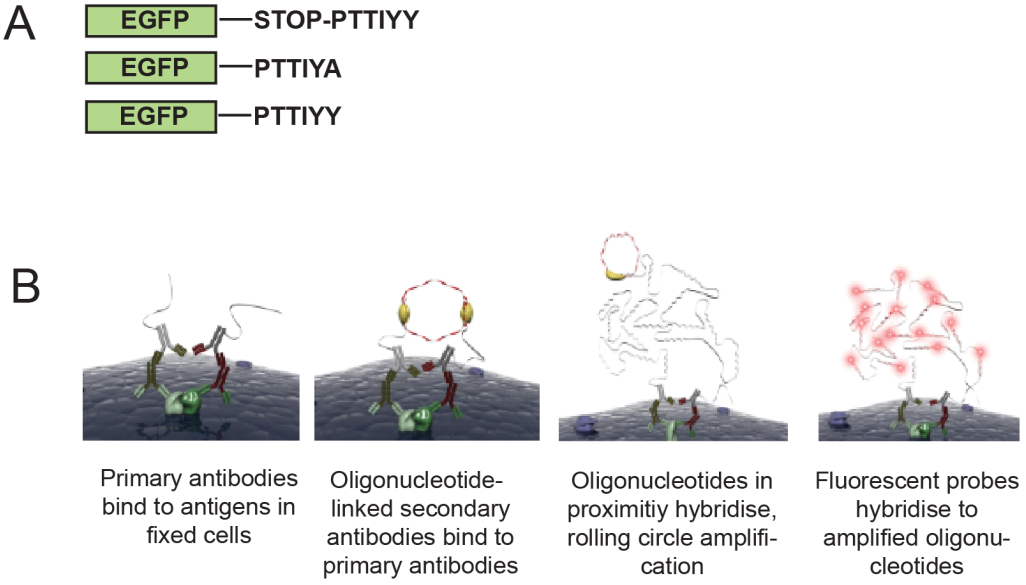


Figure 4.4 Partial co-localisation of GFP-tagged peptides with AGR2 in MCF7 breast cancer cells
GFP-tagged peptide aptamers (A) were transfected into MCF7 (American) cells and analysed by confocal microscopy (B). Endogenous AGR2 was stained with rabbit polyclonal anti-AGR2 (red) and analysed for co-localisation with expressed GFP (green).

This differential distribution suggests that AGR2 and the hexapeptide are bound in these cells either directly or in a complex, retaining expressed GFP in the ER. Whilst these data suggest that the AGR2-peptide interaction identified *in vitro* also holds true in cell culture, colocalisation does not in itself demonstrate a direct interaction, nor does it allow accurate quantification of the interaction compared to controls. A technique that has been developed recently by outside groups (in collaboration with the Edinburgh Cancer Research Centre) that can help identify direct protein-protein interactions is known as proximity ligation assay. This technique involves binding of primary antibodies to the proteins of interest, followed by ligation of oligonucleotides (fused to secondary antibodies) that are within very close proximity (less than 40nm), coupled with polymerase rolling circle amplification and fluorescent probe hybridisation [Gullberg et al., 2003] [Jarvius et al., 2006] (Figure 4.5B). This results in a binary signal from a particular fluorophore that should only be visible when the initial ligation event has occurred (ie when the two proteins of interest are in close enough proximity).

MCF7 (American) cells were transfected with the GFP-PTTIYY aptamer, GFP-PTTIYA aptamer, or GFP-STOP-PTTIYY aptamer, and incubated for 24 hours. In addition, cells were transfected with HA-tagged Reptin, a known AGR2-binding protein [Maslon et al., 2010] as a positive control. After fixing and permeabilising, fixed cells were incubated with monoclonal anti-GFP (Abcam 1218, raised in mouse) and polyclonal anti-AGR2 (Moravian Biotech, K47, raised in rabbit). Oligonucleotide-fused secondary antibodies from O-link Bioscience against either mouse or rabbit were bound to the primary antibodies and then amplified by rolling circle polymerisation.

Figure 4.5: Proximity Ligation Assay describes varying degree of co-localisation of peptides with AGR2



Reproduced from Olink Bioscience product literature

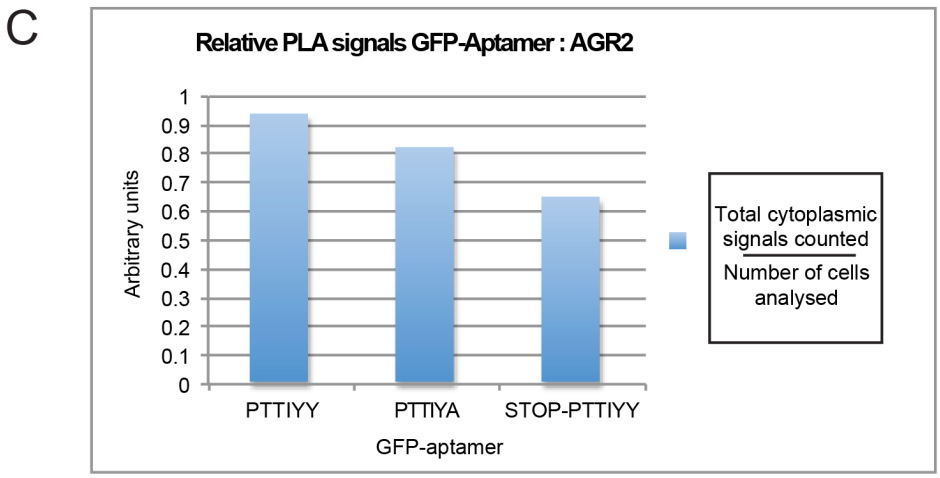


Figure 4.5 Proximity Ligation Assay describes varying degree of co-localisation of GFP-tagged peptides with AGR2
American MCF7 breast cancer cells were transfected with GFP-tagged peptide aptamers (A) and analysed by proximity ligation assay (B). Sixty (60) fields comprising 7 z-stacks were imaged for each condition, and the raw data analysed using automated software (Dr Bartlett group, proprietary). Data are shown as total number of cytomplasmic signals in Texas Red emission range per condition, normalised to the number of cells counted per condition. Experiment was not repeated, therefore no error bars are shown.

A red (DsRed) fluorophore-conjugated probe for the amplified oligonucleotide was hybridised and the signals analysed by fluorescence microscopy. Seven z-stacks were taken for each of 60 individual fields of sight for each condition, and each set of seven stacks was merged to form one image giving a total of 60 images per condition. This was done to ensure that signals were not missed due to being out of the focal plane, giving a clear picture of number of signals throughout the depth of the cell. The signals were analysed using a proprietary software (Ariol) designed in part by a lab at CRUK (Dr John Bartlett's group), in which the software is taught to define cell boundaries, and count fluorescent signals of a given size per cell and per field of vision.

The results were disappointing, in that there appeared to be a high level of noise in the system. The Proximity Ligation Assay technique is becoming more common as a tool to study protein-protein interaction and localisation, however, inherent in the system is a reliance on specificity of the primary antibodies due to the quantitative nature of the assay. A single proximity event can be amplified to produce a fluorescent signal making the assay more sensitive than standard immunofluorescence. Unspecific binding or insufficient washing can cause noise in the system, resulting in poor or ambiguous results. Cells transfected with HA-tagged Reptin and incubated with anti-HA and anti-AGR2 primary antibodies did yield nuclear DsRed fluorophores as expected although these were not quantified. In the case of the inital GFP-aptamer - AGR2 interaction assay, there did appear to be a difference in the number of cytoplasmic signals between the PTTIYY aptamer and GFP alone, with the PTTIYA mutant at an intermediate level between the two (Figure 4.5C). However, a high level of signal was seen even with the GFP-STOP-PTTIYY

aptamer. This suggests that there may be unspecific binding between GFP alone and AGR2, or that the GFP antibody or anti-AGR2 antibody is not specific enough for this assay. Attempts to optimise the AGR2 antibody or GFP antibody to reduce background signal were unsuccessful and therefore we did not repeat the assay using this technique. Nevertheless, the initial proximity ligation assay supports the findings from the immunofluorescence data and biochemical techniques that the PTTIYY aptamer binds AGR2.

4.3 Identification and validation of linear peptide-containing human protein Meckelin as possible AGR2-binding protein

Phage-peptide display as a technique was useful in identifying the AGR2-binding hexapeptide PTTIYY as discussed above, but the next question was whether this information could now be used to identify AGR2-binding proteins in the human proteome. The hexapeptide, or linear motif, was shown *in vitro* to be highly specific (Figure 4.1) for binding to AGR2, and the published Kd for the interaction between the hexapeptide and AGR2 is in the μM range. This is in agreement with the peptide ligand-protein dissociation constants of other molecular chaperones that must be able to bind and release their substrates under physiological conditions [Feifel et al., 1998], unlike ligands such as pharmacological drugs that often have dissociation constants in the nanomolar range. We used the results of the alanine-scan and substitutional mutation scan of the linear peptide motif published by Murray et al. (2007) and

Fourtouna et al. (2009) to probe the human genome and search for potential AGR2 interacting proteins.

4.3.1 Database mining the human proteome for linear peptide motif

The original phage display peptide was a 12-mer with the sequence HLPT-TIYYGPPG. By truncating the 12-mer from the N-terminal it was shown that the presence of the first position of the hexapeptide was essential for binding to AGR2, however, this position was labile in its identity, and the alanine scan showed that mutation of the proline to an alanine did not significantly affect binding to AGR2 [Murray et al., 2007]. Similarly, truncation of the 12-mer from the C-terminal indicated that the final tyrosine residue in position 6 of the hexapeptide was essential for binding, reducing the core binding motif to a hexapeptide. A mutational scan of the linear peptide motif determined which residue substitutions could be tolerated to maintain binding affinity to AGR2 [Fourtouna et al., 2009]. A competition assay determined which residues could be substituted in each of the six positions of the hexapeptide, resulting in a linear peptide motif with the sequence X-[TS]-[AGNQPLMVITS]-I-[YF]-[YF] (Figure 4.6 A). Positions 2, 4, 5 and 6 in the hexapeptide were found to be relatively fixed, tolerating only moderate changes to their identity. Threonine and serine at position 2 are similar in that they are both polar amino acids bearing an alcohol group, and tyrosine and phenylalanine are also very similar in that they are both aromatic amino acids bearing a large circular benzene ring and share similar structure, indicating a possible spacing requirement for maximal binding of the peptide to the AGR2 protein.

Figure 4.6: Database mining of human genome for proteins containing linear doman motif

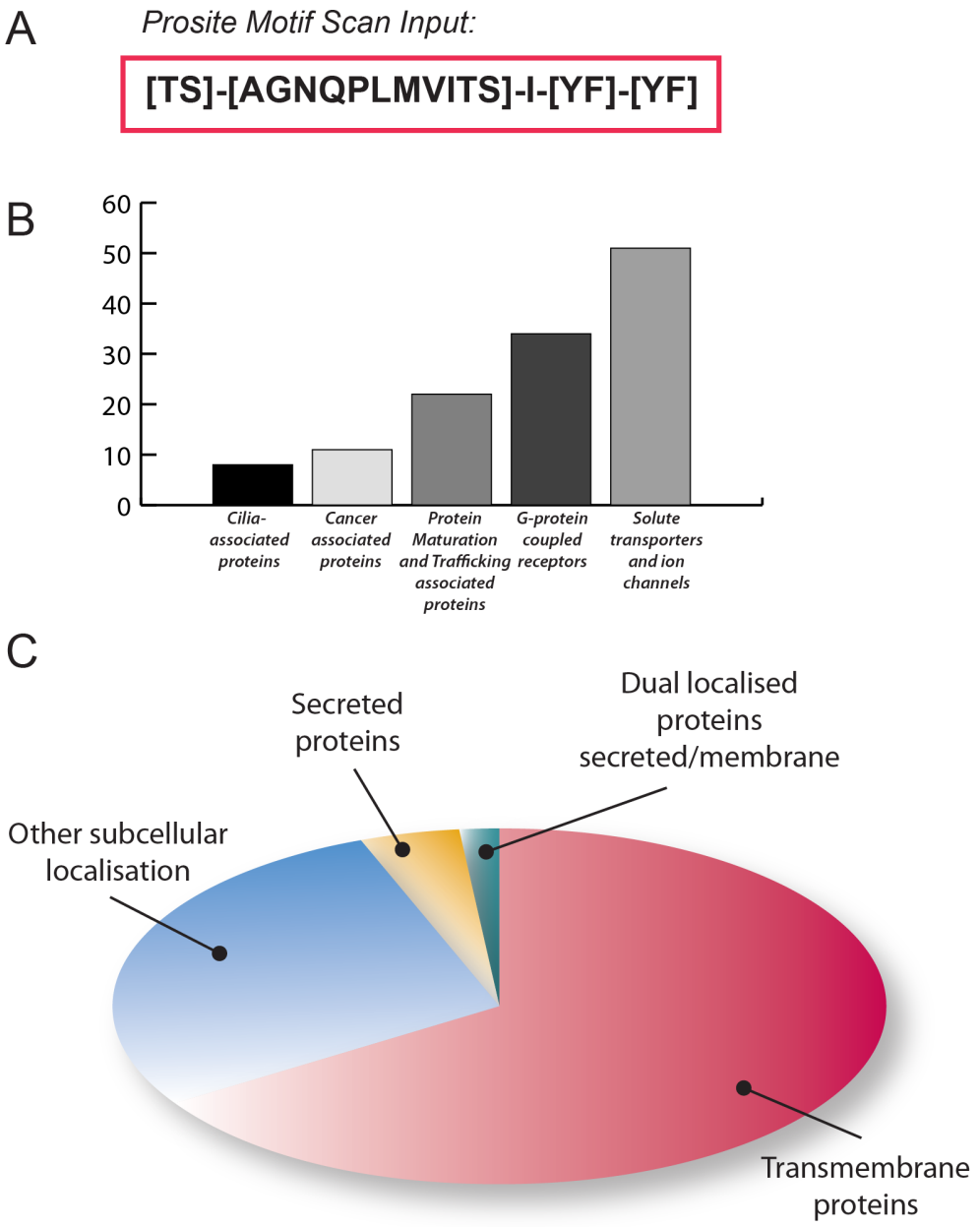


Figure 4.6 Database mining of human genome for proteins containing linear doman motif - The result of a substitution scan of the linear peptide PTTIYY for binding to AGR2 was used to probe the human genome for proteins containing the linear peptide binding motif (A). Resulting hits were scored for function (B) and subcellular localisation (C).

The linear peptide and its possible substitutions represented a potential candidate for an *in vivo* binding motif for AGR2. To look for proteins in the human proteome that contain this linear peptide motif, we used a search engine from the Prosite website called Motifscan. This database search resulted in 225 hits when multiple isoforms of the same gene were excluded (Appendix I). This list of hits was analysed for both functional and structural/localisation trends. Among the hits were cilia-associated proteins, protein-folding and transport-associated proteins, glycosylases, solute transporters and ion channels, and cancer-associated proteins. Interestingly, a large proportion (74 percent) of the hits were secreted or transmembrane proteins, suggesting a role for AGR2 in maturation or chaperoning of this class of protein. In fact, both the 2009 AGR2 mouse knockout paper [Park et al., 2009] and another more recent paper [Norris et al., 2012] suggest a role for AGR2 as an ER chaperone for MUC2 and MUC1 respectively, as well as another mucin family member, MUC5AC [Schroeder et al., 2012] with both of these being secreted proteins, in agreement with the large number of transmembrane and secreted proteins found in the database search. Furthermore, a large number of all the hits from the search were large, cysteine-rich proteins that undergo extensive post-translational modifications, again in agreement with the characteristics of MUC2, MUC1 and MUC5AC.

Using a web-based program (ToppFun, http://toppgene.cchmc.org/ enrichment.jsp), we input all the gene names that were identified by the Prosite Motifscan program. Programs for probing enrichment patterns of a given gene set have been described previously [Subramanian et al., 2005], and this gene set enrichment program is based on the same concept but with more options available. The program assesses a user-defined set of genes for enrinchment of var-

ious parameters compared to the genome as a whole. These include GO (gene ontology) terms, co-expression analysis, protein class, or disease associations. The output provides a measure of the enrichment of the gene set compared to the genome for a particular parameter. The measure is given as a P-value that represents the statistical significance of the parameter's over-representation in the gene set compared to the genome. The results are displayed from highest over-representation for whichever parameter has been chosen to lowest, with "top" scorers topping the output list. The results can help give a picture of the biological function or cellular localisation of a gene set, or provide global information about gene expression clusters or networks. A complete list of all the hits can be found in Appendix I, and some sample spreadsheets from the input of our gene set into the ToppFun program are in Appendix II.

When biological function is used as a parameter for comparison, the top 8 scores are all related to transport across the plasma membrane. The top score is transmembrane transport with a P-value of 3.49^{e-14}. In fact, 17 of the top 20 scores are related to transport, the lowest of these still with a P-value of 2.97^{e-05}. When a different parameter, Molecular Function, is analysed, the top scores are neurotransmitter symporter/transporter activity (P=1.44^{e-21} and P=1.26^{e-19} respectively). Also of note for this category, is G-protein coupled receptor activity with a P-value of 4.03^{e-13}. Using Cellular Component as a parameter, the top score for enrichment is proteins integral to plasma membrane (P=1.48^{e-11}) and proteins found at the apical surface of the plasma membrane (P=2.46^{e-06}). These high levels of enrichment further support a role for AGR2 in the secretory pathway. Another cellular compartment highly enriched included the cilium (including axoneme), with a P-value of

4.86^{e-3}. In addition, when the Prosite Motifscan list was analysed for known gene expression patterns from Gudmap (a consortium of transcript and microarray data from mouse tissue) [Harding et al., 2011], kidney tissue types were enriched including renal podocytes (P=2.78^{e-03}), renal proximal tubules (P=2.19^{e-03}), renal cap mesenchyme (P=2.89^{e-02}) and renal S-shaped bodies (early nephrons) (P=1.89^{e-03}). The possible significance of this will be discussed further in Chapters 6 and 7.

The high over-representation of the neurotransmitter transporter activity and G-protein coupled receptors (compared to other membrane transporters or receptors) suggests that the AGR2-binding linear peptide motif identified by phage peptide display proteomics is a valid motif and not simply due to a bias inherent in the phage display process. Furthermore, Cathepsin D was identified in the list of hits, a lysosomal protease whose secretion has previously been shown to be promoted by AGR2 in pancreatic cancer cells [Dumartin et al., 2011], as was a member of the mucin family, MUC16 (also known as CA125), a glycoprotein that is over-expressed in mucinous ovarian cancers, promotes tumour invasiveness, and is associated with poor prognosis [Streppel et al., 2012]. AGR2 has previously been shown to interact with other members of the mucin family, and to be essential for their maturation and secretion [Park et al., 2009] [Norris et al., 2012]. In addition, MUC16 has been shown to bind mesothelin, promoting cell adhesion and metastasis [Rump et al., 2004], and mesothelin is another protein identified in the Prosite Motifscan list as containing the AGR2-binding linear peptide motif. As mentioned in the Introduction, over-expression of AGR2 is highly associated with onset and negative prognosis of mucinous ovarian cancers, and the presence of a putative linear peptide

binding motif in both MUC16 and mesothelin suggests a possible functional pathway involving AGR2.

Of the numerous proteins identified within the human proteome containing the linear peptide motif, we decided to focus on validating a few in particular in order to test our system of using phage-peptide interactomics to identify interactions *in vivo*. One of these was a human transmembrane protein called MKS3, or TMEM67 (*Meckelin*). We chose MKS3 as it fit the criteria of the largest group identified by the Prosite motifscan data, namely, it is a large (995 amino acid), glycosylated [Smith et al., 2006], 3-transmembrane protein (predicted) with numerous cysteines (26 in the N-terminal region) [Flicek et al., 2012] (Figure 4.7). The protein is evolutionarily conserved, with homologues in *C. elegans* and *D. melanogaster*, but no known paralogues [Dawe et al., 2007b]. The predicted N-terminal structure of the protein has putative topological similarities to the Frizzled family of receptors [Dawe et al., 2007b], a family that has been implicated in Wnt ligand signalling and in differentiation and development [Bhanot et al., 1996]. It also harbors an N-terminal signal peptide that may target it to the secretory/membrane pathways [Smith et al., 2006]. It has been shown to be involved in centriole migration that is essential for initiation of ciliation in mammalian cells [Dawe et al., 2007b] as well as mitosis, and *Meckelin* is a gene that, when mutated in humans causes ciliopathies. Numerous different mutations have been described leading to varying levels of severity of disease. In the case of Meckel-Gruber syndrome neonatal or *in utero* death occurs [Chen, 2007] [Gunay-Aygun et al., 2009], however other mutations in Meckelin lead to less severe phenotypes [Baala et al., 2007].

MKS3 wild type protein localises to the ER and Golgi, plasma membrane

Figure 4.7: Human Meckelin (TMEM67, MKS3, meckelin) - Gene and protein domains

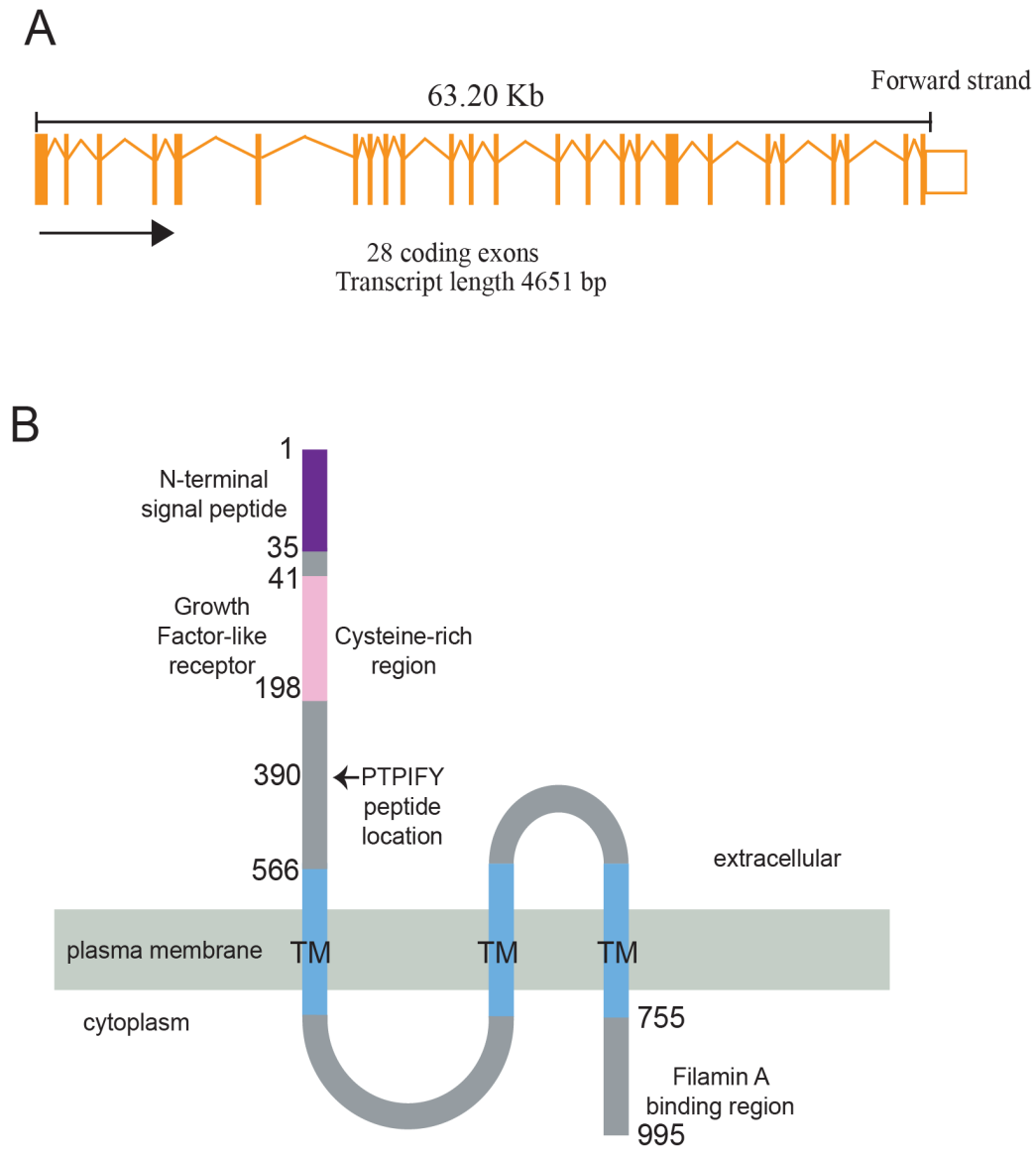


Figure 4.7 Human Meckelin (TMEM67, MKS3, meckelin) - Gene and protein domains (A) Gene structure of human Meckelin. (B) Protein domains of human MKS3 (not to scale, transmembrane domains predicted).

Adapted from *Adams et al. 2012* and Ensembl (www.ensembl.org)

and primary cilium in ciliated cell lines [Dawe et al., 2007b] as well as to basolateral actin cables [Dawe et al., 2009] [Adams et al., 2012]. In addition, it is seen to localise to the mitotic spindle in dividing cells (personal communication, C Johnson, Leeds). In ciliated cell lines depleted of MKS3 protein by siRNA, migration of the centrioles to the plasma membrane is blocked, leading to loss of cilia [Dawe et al., 2007b]. Furthermore, epithelial branching morphogenesis was also severely impaired, indicating a role for *Meckelin* in epithelial morphogenesis, although the mechanisms for this are unclear. In contrast, in a spontaneous Meckelin mouse knockout a different phenotype is observed, with abnormally long cilia seen rather than an absence of cilia [Cook et al., 2009]. The primary cilia in mammalian cells act as sensors that play an important role in signal transduction, differentiation [Satir and Christensen, 2008], and cell cycle [Plotnikova et al., 2009] and are required for the proper development of many organs. In addition, cilia are easily visualised in many cell lines using immunofluorescence microscopy, and thus can provide a readout for phenotypic changes. One of the symptoms of ciliopathies caused by mutations in *Meckelin* is polycystic kidneys. The ToppFun suite analysis of the putative AGR2-binding linear peptide motif identified that cilia-associated genes and genes expressed in the developing kidney were strongly over-represented. Taken together, we felt that *Meckelin* represented an ideal candidate for demonstrating proof-of-concept that phage-display proteomics could be used to identify relevant biological interactions *in vivo*, and if correct, could provide insight into a biological role for *Agr2*.

4.3.2 Validating the AGR2 - MKS3 Interaction

As a starting point, we decided to test a putative interaction between MKS3 and AGR2 using co-immunoprecipitation (CoIp). CoIp involves capturing a target protein using an antibody against that protein, and then binding the heavy chain of the antibody indirectly to a solid phase (e.g. Sepharose beads covalently linked to Protein A or Protein G which bind the heavy chains of most antibodies). The captured protein may remain bound to another protein or even complex of proteins, and when the unbound protein pool is washed away, the remaining bound protein(s) can be identified using standard biochemical techniques such as western blot or mass spectrometry. It is a technique that is widely used and can be a powerful means of identifying physiologically relevant protein interactions. As a technique, however, it does have its limitations, as it will only identify protein-protein interactions that are stable enough to withstand the process. As a result, more transient interactions may be overlooked, and the system may be prone to inconsistency. Nevertheless, overexpressed, HA-tagged MKS3 (Figure 4.8A) was found to co-immunoprecipitate with GFP-tagged AGR2 in both MCF7 (American) breast cancer cells and in human metastatic melanoma cell line A375 (Figure 4.8B,C).

A375 cells were co-transfected with either C-terminally tagged GFP-AGR2 or GFP vector (Clonetech pAcGFPN$_1$), and wild type HA-tagged MKS3 (Figure 4.8B). AGR2 has been previously shown to bind and inhibit p53 [Pohler et al., 2004], so we decided to investigate whether induction of p53 with u.v. light had any effect on Meckelin binding to AGR2. Cells transfected with GFP-AGR2 were subjected to u.v. at 50J/cm2, or not treated, as shown in Figure 4.8B, and then incubated overnight.

Figure 4.8: Overexpressed HA-tagged Meckelin co-immunoprecipitates with GFP-tagged AGR2 in human breast cancer cell line MC7 and human melanoma cell line A375

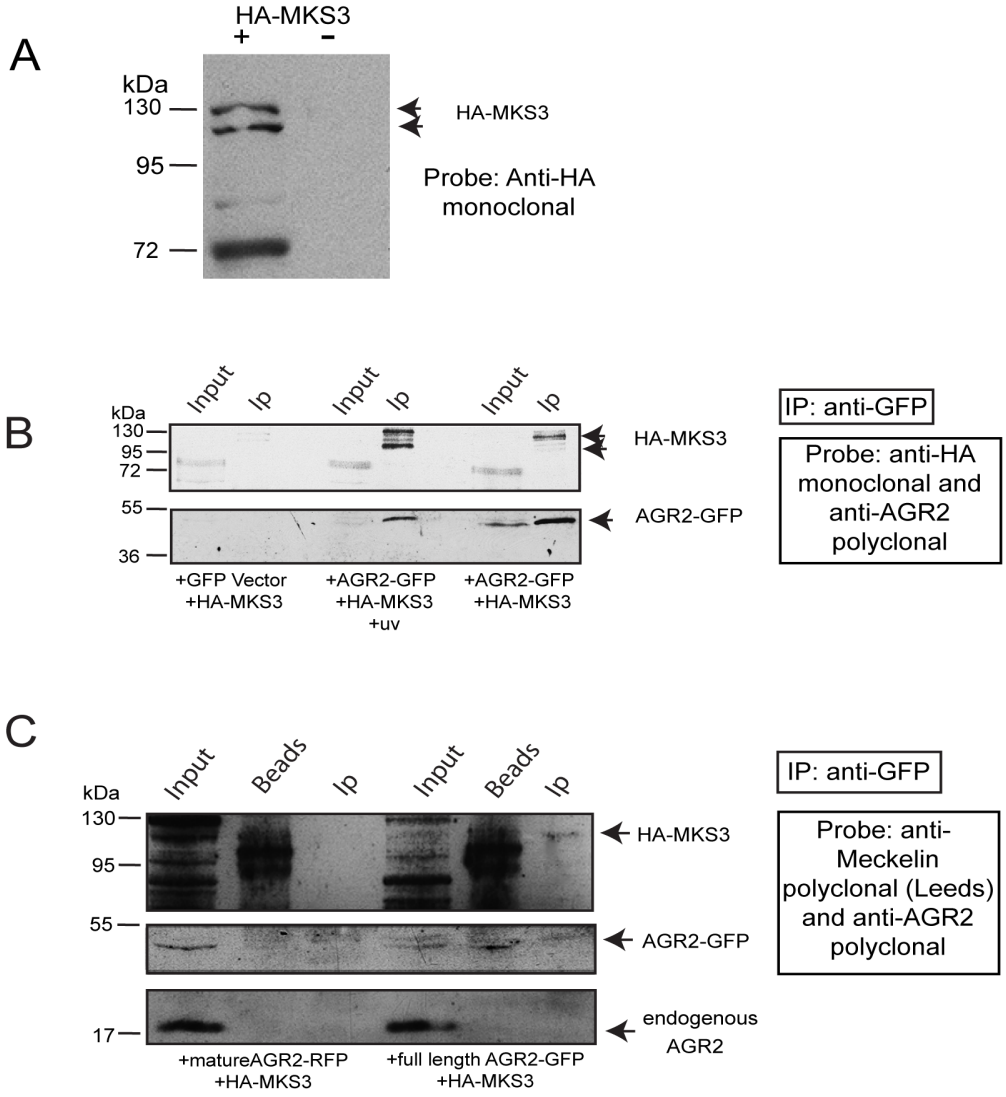


Figure 4.8 Overexpressed HA-tagged Meckelin co-immunoprecipitates with GFP-tagged AGR2 in human breast cancer cell line MC7 and human melanoma cell line A375

MCF7 (American) human breast cancer cells were transfected with HA-tagged MKS3 and lysates were analysed for expression of HA-MKS3 by western blot (A). A375 human melanoma cells (B) or American MCF7 breast cancer cells (C) were transfected with HA-tagged MKS3 and GFP- or RFP-tagged full length or mature AGR2 as indicated, and treated with uv (B) or not (C), before immunoprecipitating also as indicated. Eluted antigen-antibody complexes were analysed by western blot and probed with anti-HA mouse monoclonal antibody (B) or rabbit polyclonal anti-Meckelin N-terminal antibody (C).

After lysis, GFP-AGR2 was immunoprecipitated from the lysate with anti-GFP monoclonal antibody (abcam 1218), and antibody/antigen complexes were pulled down on Protein G-Sepharose beads and analysed by western blot. The blot was probed using anti-AGR2 polyclonal antibody K47 (described above) to confirm pulldown of GFP-AGR2 (50KDa), and was then reprobed using anti-HA monoclonal antibody (Abcam 18181). Full length MKS3 is predicted to migrate at 110KDa [Flicek et al., 2012], however, both untagged and HA-tagged Meckelin has also been described as migrating at 120-130KDa [Dawe et al., 2007b] [Dawe et al., 2009], possibly indicating some form of post-translational modification. The significance of this potential modification has not been reported. A western blot of HA-tagged Meckelin indicates that both the 110 and 130KDa forms can be seen in transfected cells (Figure 4.8A). In addition, lower bands were clearly observed, and similar bands observed in this thesis using anti-V5-tag and anti-MKS3 antibodies (Figures 4.12 and 5.1), as well as anecdotal reports from other groups that observe similar bands using different antibodies, suggest that these may represent processed or alternative isoforms that have not yet been characterised.

Overexpressed HA-tagged MKS3 co-immunoprecipitated with GFP-tagged AGR2 in A375 cells, with the 110KDa band appearing in u.v. treated cells, and the 130KDa band appearing in both treated and untreated cells, but not in cells transfected with GFP-vector control, indicating that neither the GFP itself nor the beads bind MKS3. The significance of the second (predicted) form of MKS3 at 110KDa in u.v. treated cells but not in untreated cells is unclear, and may form part of future work on Meckelin. The lack of band of appropriate size in the input lanes is disappointing, however, it is possible that

transfection levels were low, with bands appearing in the immunoprecipitated lanes due to concentration of the protein of interest (in this case MKS3) during the immunoprecipitation process.

AGR2 contains a putative leader sequence at its N-terminal with cleavage site [Pohler et al., 2004] that would target it to the endoplasmic reticulum (ER) membrane and secretory pathway. Previous work has shown that a mature form of AGR2 lacking the N-terminal leader sequence (mAGR2) localises to the nucleus and cytoplasm [Fourtouna et al., 2009] and further studies have shown that AGR2 can bind reptin and p53, both primarily nuclear proteins [Maslon et al., 2010]. Furthermore, subcellular fractionation in MCF7 (American) cells shows that AGR2 is localised mainly to the membranes and organelles, but with a significant pool in the nucleus [Fourtouna et al., 2009]. For this reason, we tested whether both the full-length (ER and Golgi localised) and mature (nuclear and cytoplasmic) AGR2 could bind MKS3 in cells (Figure 4.8C). In order to test both of these we transfected MCF7 (American) cells with either C-terminally tagged RFP-mature AGR2 or C-terminally tagged GFP-full length AGR2 together with HA-tagged MKS3. Anti-GFP monoclonal antibody was used, as before, to pull down GFP-full length AGR2 or RFP-mature AGR2, and then the antibody-antigen complexes were bound to Sepharose-Protein G beads. Unbound fractions were washed away and the bound antigen-antibody-interacting protein complexes were eluted from the beads as described above. Only full-length AGR2 was seen to co-immunoprecipitate with HA-tagged MKS3 in these cells, and not mature AGR2. This is consistent with the cellular localisation of MKS3 in the ER, Golgi and plasma membrane [Dawe et al., 2007b]. Interestingly, in

these cells only the 110KDa-migrating band is seen despite not being treated with u.v., and possibly indicating a cell-specific difference in dominant MKS3 isoforms. In addition, MKS3 was not seen to co-immunoprecipitate with the mature form of AGR2 (lacking the N-terminal leader sequence) even when the C-terminal ER-localisation motif was mutated to a strong ER-retention sequence previously shown to restrict the mature form of AGR2 to the ER rather than the nucleus (data not shown) [Fourtouna et al., 2009]. This suggests that the N-terminal leader sequence may have an intrinsic quality that is required for binding of AGR2 to MKS3.

We next wanted to investigate whether AGR2 and MKS3 could co-localise in cell lines by immunofluorescence. To do this, we started with MCF7 (American) breast cancer cells since they express high levels of endogenous AGR2. MCF7 cells were transfected with HA-tagged wild type MKS3 and incubated for 24h. Cells were fixed and incubated overnight with anti-AGR2 polyclonal antibody together with anti-HA monoclonal antibody, both as described above. Cells were subsequently incubated with Alexa-fluorophore-conjugated secondary antibodies (anti-mouse 488, anti-rabbit 594) and Topro-3 nuclear stain, and analysed by confocal microscopy. The results show that HA-tagged Meckelin partially colocalises with endogenous AGR2 in MCF7 cells in a perinuclear staining pattern (Figure 4.9A), in agreement with previous work showing that AGR2 colocalises with the ER and Golgi network in these cells [Fourtouna et al., 2009]. Untransfected (Figure 4.9B) and no primary (Figure 4.9C) controls indicate that there is no unspecific binding of the HA antibody or of the secondary antibodies.

Further analysis in A375 cells co-transfected with GFP-tagged (full-

Figure 4.9: Overexpressed HA-Meckelin and endogenous AGR2 partially co-localise in human breast cancer cell line MCF7 (American)

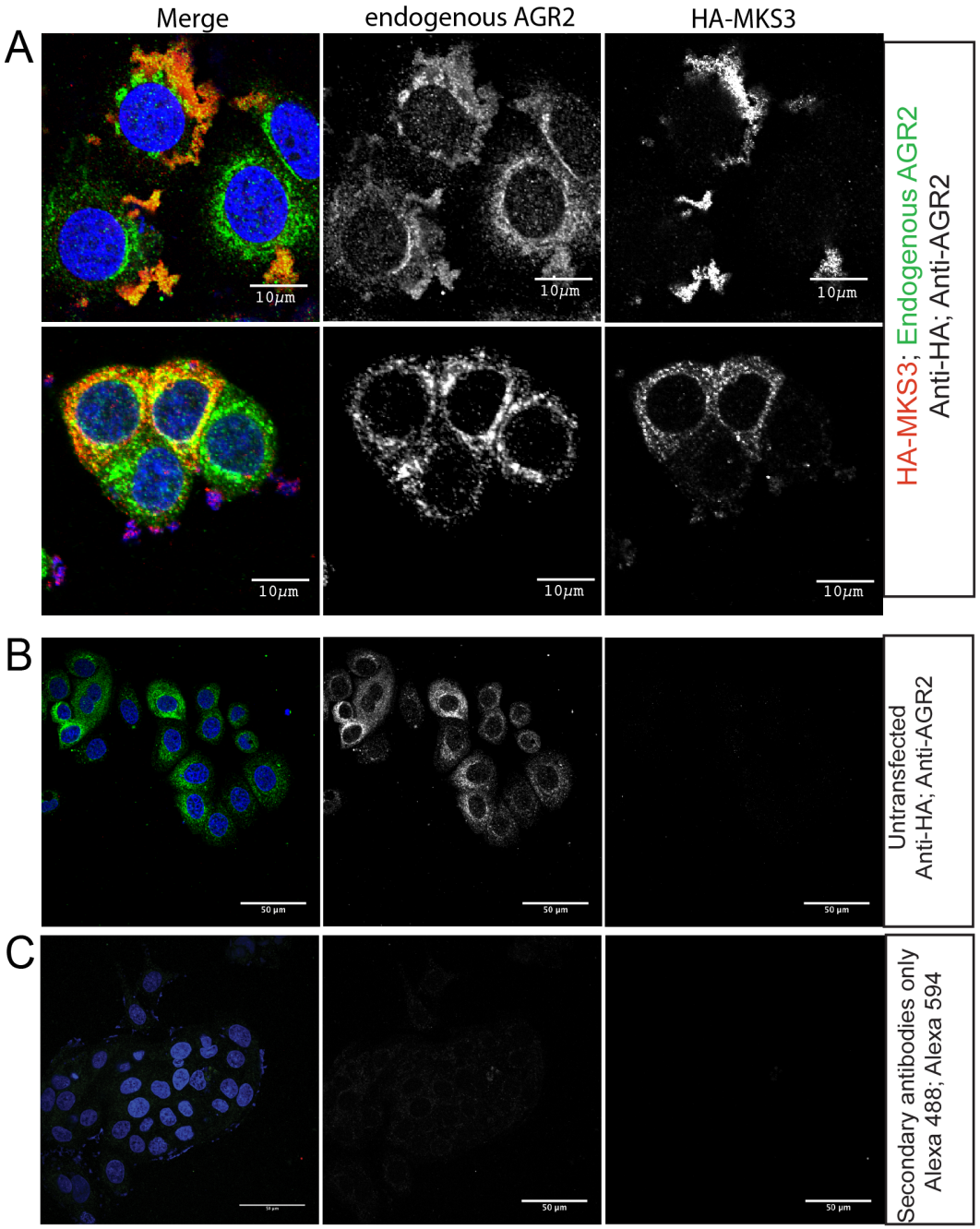


Figure 4.9 Overexpressed HA-Meckelin and endogenous AGR2 partially co-localise in human breast cancer cell line MCF7 (American)
MCF7 (American) human breast cancer cells were either transfected with HA-tagged MKS3 (A, C) or not transfected (B) and incubated with secondary antibodies (A, B) or no secondary antibodies (C), and analysed by confocal microscopy.

length) AGR2 and HA-MKS3 (Figure 4.10A) confirmed that the partial colocalisation of AGR2 and MKS3 is seen in a separate cell type, although in this case using GFP-AGR2 rather than an antibody to the endogenous protein due to the lack of endogenous AGR2 in the human A375 melanoma cell line. Controls for unspecific binding of the anti-HA antibody (Figure 4.10B) and unspecific binding of the secondary antibody (Figure 4.10C) in these cells showed that the antibody binding was specific and that untransfected cells did not express GFP or have detectable auto-fluorescence in the GFP range.

In order to assess the expression and localisation of endogenous MKS3 by immunofluorescence in MCF7 cells, we used a polyclonal antibody to a C-terminal peptide of MKS3 (kindly donated by C Johnson, Leeds). MCF7 cells were either transfected with HA-MKS3, or not transfected, then fixed and incubated overnight with anti-AGR2 monoclonal antibody, and affinity purifed anti-MKS3 C-terminal antibody. Cells transfected with HA-MKS3 demonstrated partial co-localisation with AGR2 in a very defined peri-nuclear localisation as well as more diffuse colocalisation in what is likely to be the ER (Figure 4.11A). This is similar to the localisation of the overexpressed MKS3 and AGR2 as described above. Interestingly, however, endogenous MKS3 was only seen at the spindle poles during anaphase in dividing cells (Figure 4.11B), whilst AGR2 was localised diffusely in the cytoplasm of these cells. MCF7 cells in culture are not normally ciliated [Yuan et al., 2010], and therefore it is possible that the expression of Meckelin in these cells is cell cycle related, and is temporally restricted to migration of the centrioles and initiation of cell division. In contrast, HA-tagged MKS3 was not seen at the spindle poles, suggesting that the tagged, transfected and overexpressed version of MKS3

Figure 4.10: Overexpressed HA-MKS3 and GFP-tagged AGR2 partially co-localise in human melanoma cell line A375

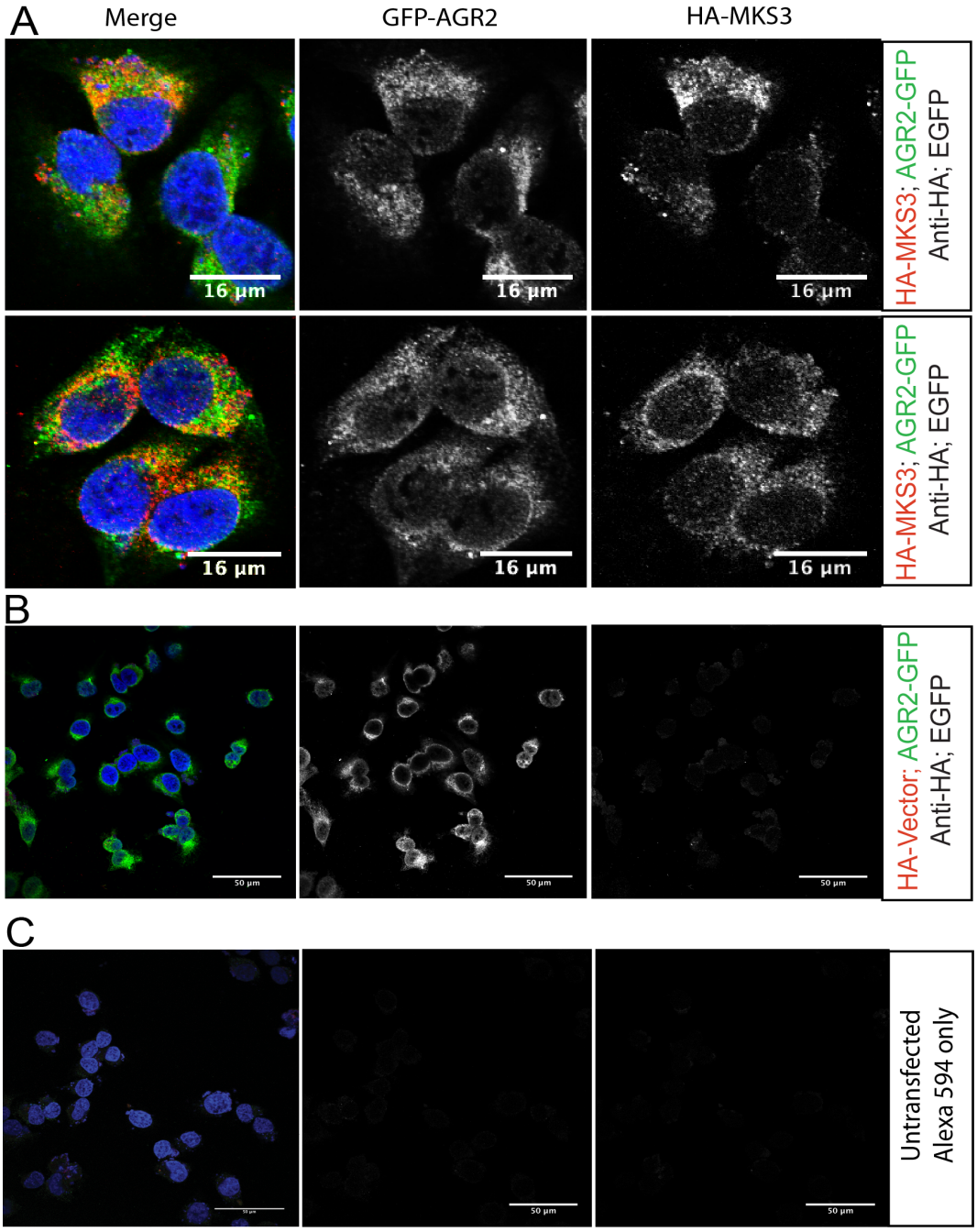


Figure 4.10 Overexpressed HA-MKS3 and GFP-tagged AGR2 partially co-localise in human melanoma cell line A375
A375 human melanoma cells were either transfected with HA-tagged MKS3 and GFP-tagged AGR2 (A), HA empty vector and GFP-tagged AGR2 (B) or not transfected (C) and analysed by confocal microscopy.

Figure 4.11: Expression patterns of overexpressed HA-MKS3 and endogenous MKS3 in MCF7 (American) breast cancer cells

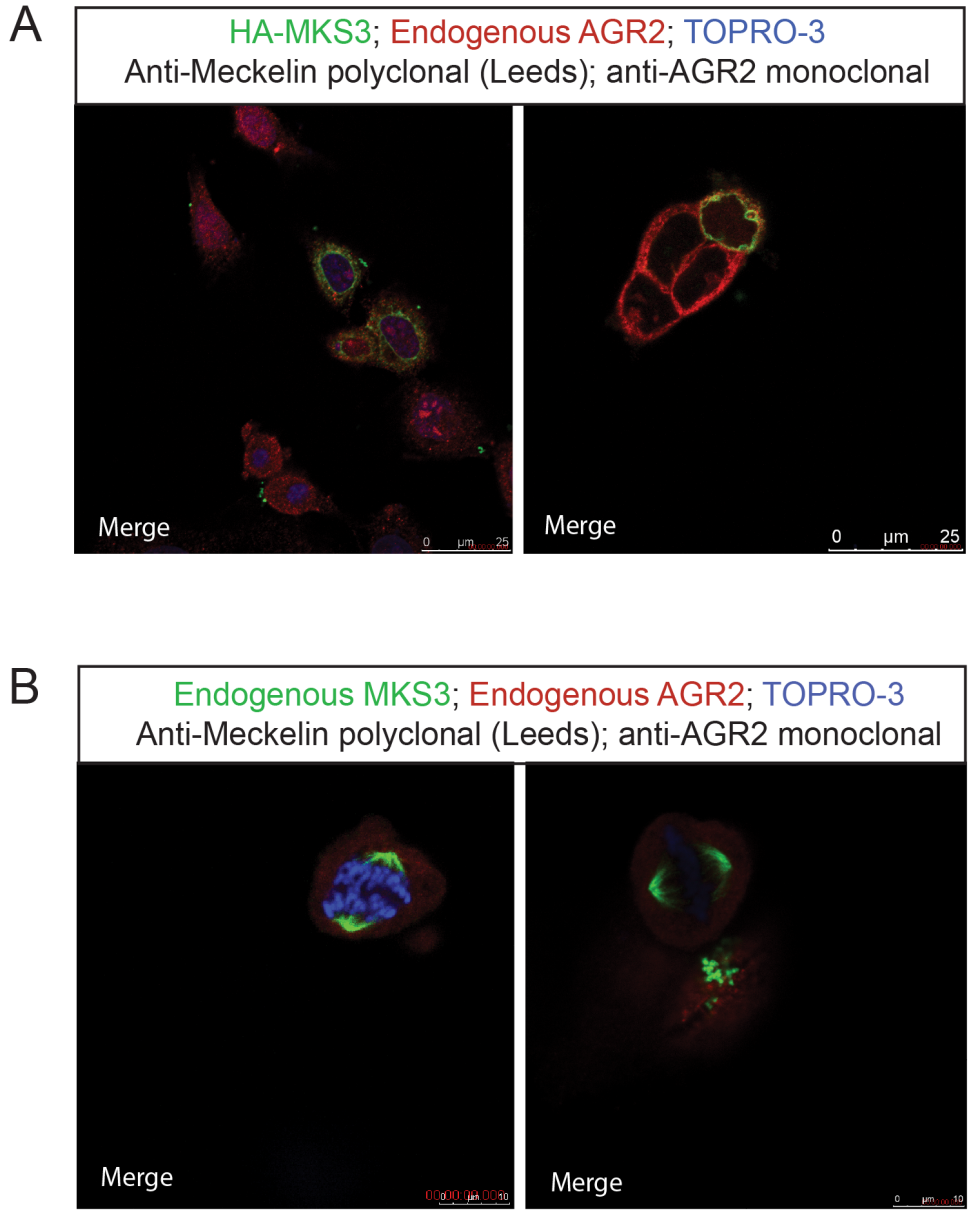


Figure 4.11 Expression patterns of overexpressed HA-MKS3 and endogenous MKS3 in MCF7 (American) breast cancer cells
American MCF7 human breast cancer cells were either transfected with HA-MKS3 (A) or not transfected (B), and analysed by immunoflurescence using rabbit polylconal (affinity purified) anti-MKS3 (C Johnson, Leeds) and mouse monoclonal anti-AGR2 antibodies. Images shown are merged channels only, for comparison between endogenous and transfected HA-MKS3.

may not behave the same way as endogenous MKS3 *in vivo*, or that different isoforms may be localised to different compartments. The differences seen between endogenous MKS3 and overexpressed, HA-tagged MKS3 must be considered when assessing a putative interaction with AGR2 in cell lines.

4.3.3 Assessing the effect of mutations in the linear peptide motif

The putative interaction between AGR2 and MKS3 was identified based on a linear peptide motif as described in section 4.3.1. In order to test whether this was in fact a site of interaction, mutants of MKS3 were cloned with mutations in the linear peptide motif. The wild type sequence of the motif in MKS3 is PTPIFY (note similarity to the linear peptide identified by phage display of PTTIYY). As previously described, mutation of the final residue in the hexapeptide motif to an alanine restricted binding of AGR2 by ELISA, therefore, we made two N-terminally V5-tagged mutants; one double mutant where the linear peptide motif was mutated to the optimal PTTIYY, and one single mutant with the final tyrosine residue mutated to an alanine. MCF7 (American) cells were transfected with wild type or mutant V5-tagged MKS3, or an irrelevant 14kDa V5-tagged protein, and V5-tagged isoforms were tested for expression in mammalian cells (Figure 4.12A). We then looked for differences in co-localisation by immunofluorescence of the mutants compared to the wild type sequence.

We found that mutating the final tyrosine residue in the linear peptide motif of MKS3 restricted co-localisation with endogenous AGR2, although

Figure 4.12: Effect of mutations in the putative AGR2-binding linear domain motif of V5-MKS3 in human breast cancer cell lines - Part I

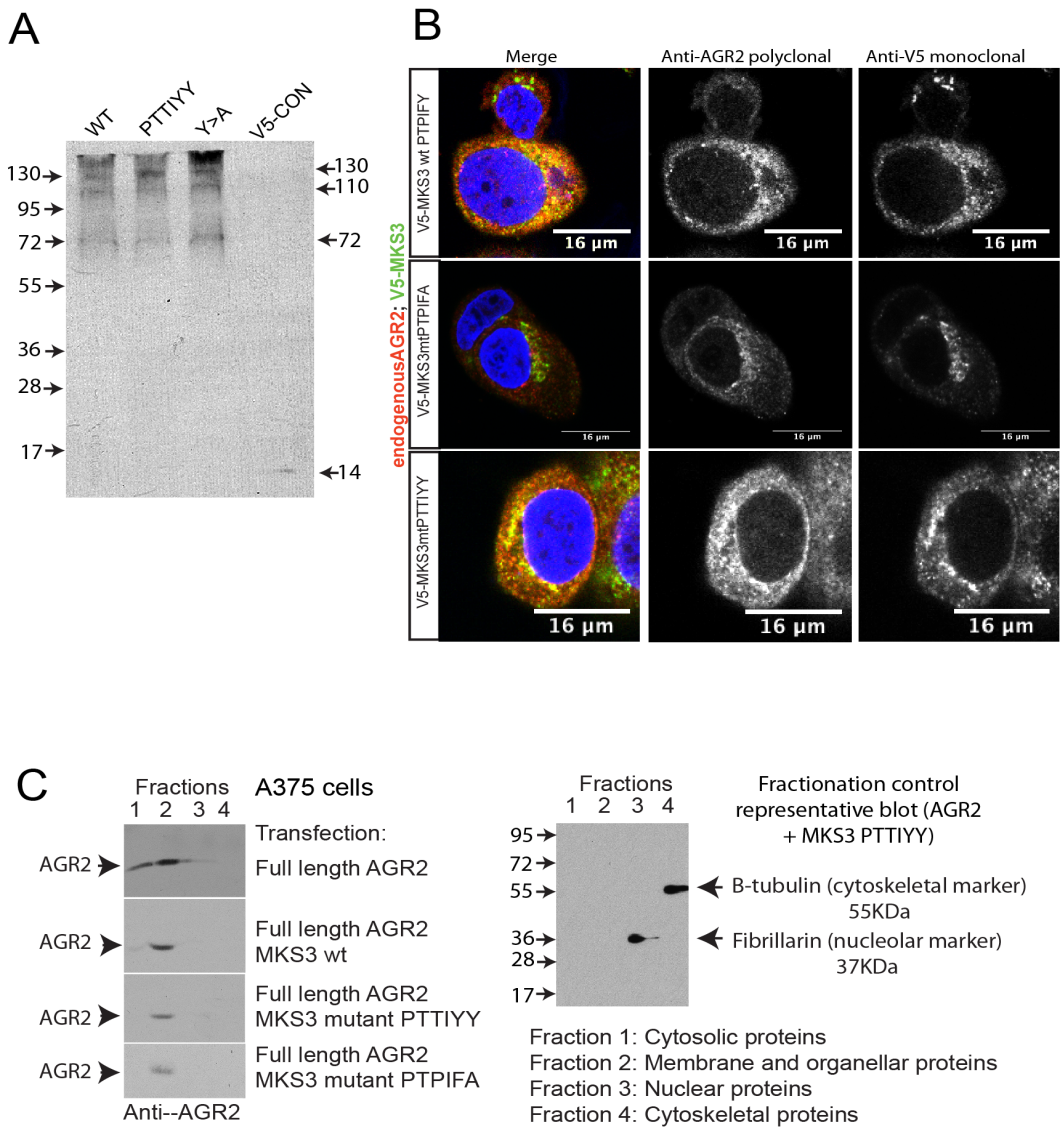


Figure 4.12 Effect of mutations in the putative AGR2-binding linear domain motif of V5-MKS3 in human breast cancer cell lines - Part I
(A) MCF7 (American) breast cancer cells were transfected with N-terminal V5-tagged MKS3 wild type ($MKS3^{PTPIFY}$) or mutants ($MKS3^{PTTIYY}$, $MKS3^{PTPIFA}$), and lysates were analysed for expression of V5-MKS3 by western blot using mouse monoclonal anti-V5 antibody. (B) MCF7 (American) cells were transfected with wild type or mutant V5-MKS3 and analysed for co-localisation with endogenous AGR2 using mouse monoclonal anti-V5 and rabbit polyclonal anti-AGR2 antibodies. A375 human melanoma cells were transfected with full length AGR2 and either wild type or mutant V5-tagged MKS3, and cells were fractionated into cytosolic, membrane, nuclear and cytoskeletal fractions before analysis by western blot (C). Fractionation control for nuclear and cytosolic fractions are shown (C, right panel).

it did not completely abolish it (Figures 4.12B, Figure 4.13). In contrast, mutating the wild type motif to the optimal motif PTTIYY enhanced co-localisation compared to the PTPIFA mutant. This is in agreement with the biochemical findings that mutation of the final tyrosine in position 6 to anything other than tyrosine or phenylalanine restricted the ability of AGR2 to bind the peptide. The differences in colocalisation between AGR2 and wild type or mutant MKS3 was also seen in cell line A375 (data not shown).

We next wanted to assess the effect of transfecting either wild type or mutant MKS3 on endogenous AGR2 localisation in MCF7 cells. Cells were transfected with V5-tagged wild type, PTTIYY mutant or PTPIFA mutant MKS3 and incubated overnight. Cells were fractionated into four fractions; cytosolic proteins, membrane and organellar proteins, nuclear proteins and cytoskeletal proteins. Lysates were analysed by western blot and probed using anti-AGR2 polyclonal antibody. We found that overexpression of any of the V5-tagged versions of MKS3 led to concentration of AGR2 in the membrane and organellar fraction and a reduction of AGR2 in the nuclear and cytoplasmic fractions (Figure 4.12C). This suggests that there may be more than one site of interaction, or that a binding-independent signalling pathway may be responsible for accumulation of AGR2 in the membranes and organelles when one of its target substrates is overexpressed, such as through the unfolded protein response (UPR).

Although the contrast between the PTPIFA mutant and the wild type or double mutant was noted qualitatively, we wanted to assess the differences between the motif mutants and wild type in a quantifiable manner. One of the resident locations for the transmembrane protein MKS3 is the plasma

Figure 4.13: Effect of mutations in the putative AGR2-binding linear domain motif of V5-MKS3 in human breast cancer cell line MCF7 (American): Part II

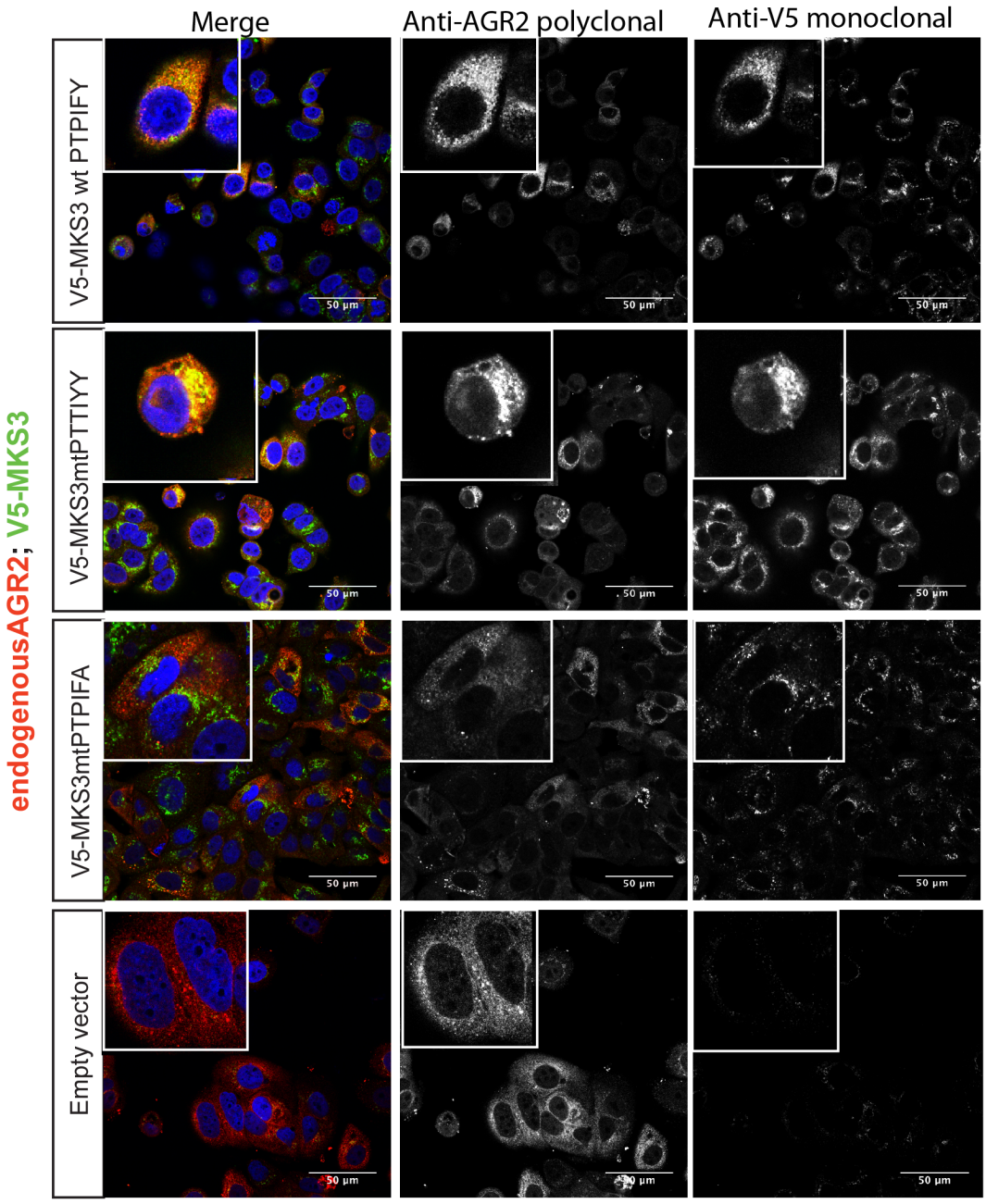


Figure 4.13 Effect of mutations in the putative AGR2-binding linear domain motif of V5-MKS3 in human breast cancer cell line MCF7 (American): Part II
MCF7(American) cells were transfected with wild type (V5-MKS3^{PTPIFY}) or mutant (V5-MKS3^{PTTIYY}, V5-MKS3^{PTPIFA}) MKS3, or with empty V5 vector (nV5pDEST3.1) and analysed by immunofluorescence for co-localisation with endogenous AGR2 using rabbit polyclonal anti-AGR2 and mouse monoclonal anti-V5 antibodies.

membrane [Dawe et al., 2007b]. We hypothesized that a reduction in the affinity of binding to AGR2 caused by a mutation in the linear peptide motif PTPIFY might alter the dynamics of localisation of MKS3 at this cellular compartment.

The mechanism by which integral, transmembrane proteins mature and arrive at their resident location is through the secretory pathway. Hydrophobic patches are shielded from the aqueous intracellular environment through vesicular transport, and by association with chaperone proteins [Lodish et al., 2000]. Integral membrane proteins first insert into the ER membrane co-translationally, before moving into the trans-Golgi network and secretory pathway. Secretory, or ER-lumen bound proteins must pass through the ER membrane into the ER lumen, and this includes resident chaperone proteins such as BiP (Grp78) [Vitale and Denecke, 1999] (Figure 4.14A). For many transmembrane proteins this process in the ER is mediated by an N-terminal leader sequence (signal peptide), although most multi-spanning transmembrane proteins also insert into the ER membrane thanks to the secondary structure of their transmembrane domains (termed signal anchor sequence) that resembles a signal peptide but is not cleaved [Lodish et al., 2000]. Both AGR2 and MKS3 contain putative N-terminal signal peptides (Figure 4.14B) [Petersen et al., 2011] with putative cleavage sites, although N-terminally tagged HA-MKS3 does not lose either the HA-tag or correct localisation [Dawe et al., 2007b], indicating that the signal peptide is not cleaved at least in the tagged form. Once integral or trans-membrane proteins are trafficked to the plasma membrane through the trans-Golgi network, the cargo vesicles fuse with the plasma membrane in a poorly understood process, inserting the protein into

Figure 4.14: Both AGR2 and MKS3 contain putative signal peptides at their N-terminus

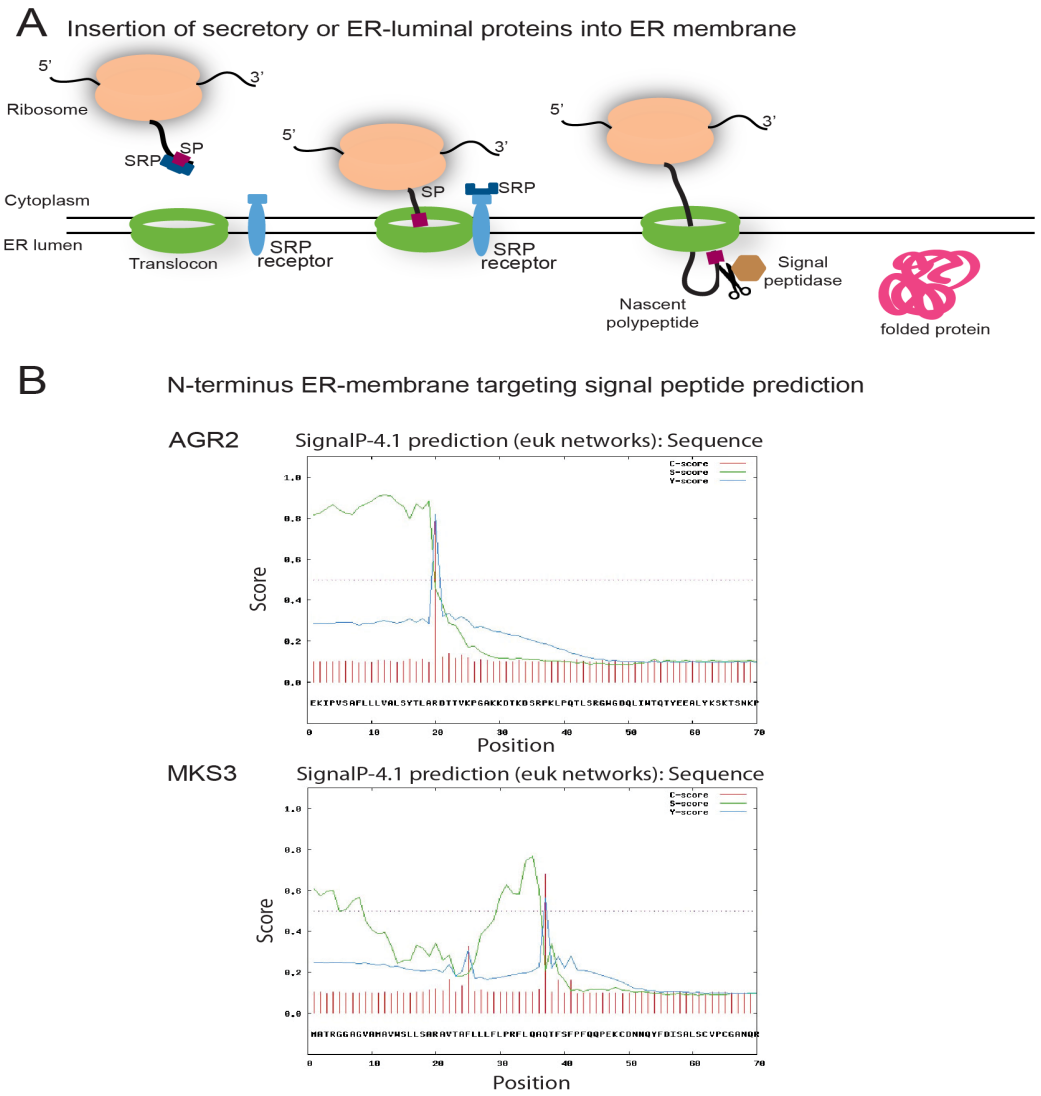


Figure 4.14 Both AGR2 and MKS3 contain putative signal peptides at their N-terminus (A) Secretory and luminal endoplasmic reticulum (ER) proteins are translocated across the ER membrane. A signal peptide at the N-terminus of secretory or ER luminal proteins targets the nascent polypeptide emerging from the ribosome to the ER membrane by interacting with a signal recognition particle (SRP), which binds the SRP receptor. The SRP receptor links the nascent polypeptide to the translocon, a channel across the ER membrane that allows hydrophilic peptides to cross the membrane. The signal peptide is usually cleaved by signal peptidase in the ER lumen, and the protein is folded inside the lumen of the ER. (B) Both AGR2 and MKS3 contain putative signal peptides at their N-terminus (B) Reproduced from *Petersen et al., 2011*

the membrane in the correct orientation [Lodish et al., 2000].

In order to look for changes to the localisation of MKS3 at the plasma membrane, we needed a technique that allowed us to visualise secretory or membrane protein-loaded vesicles at high resolution, at the plasma membrane or just beneath it. Visualising vesicles at the plasma membrane by microscopy can be challenging due to fluroescence signal from outside the field of interest, and the small depth of the area being investigated. A technique termed Total Internal Reflection Fluorecence Microscopy (TIRF-M) is a very powerful way of imaging the precise location of the plasma membrane and any associated vesicles [Loder et al., 2013] [Mattheyses et al., 2010]. The principle of TIRF microscopy (Figure 4.15) uses the difference in the refractive indices of two media, in the case of TIRF-M, the glass coverslip on which cells are grown and aqueous medium above it containing the specimen. To achieve total internal reflection, the refractive index of the specimen must be lower than the coverslip. By shining a laser beam at the sample through a prism, the reflective quality of the laser beam as it passes through the glass and then the aqueous medium causes an evanescent wave to be emitted at the interface between the glass and the aqueous medium before it is reflected away from the specimen. This means that only the fluorophores within the evanescent wave are excited, producing images clear of background noise and allowing fluorescence imaging down to single-molecule resolution when combined with other imaging techniques such as Fluorescence resonance energy transfer (FRET) [Holden et al., 2010]. The excitation wave (produced by the laser beam as it hits the differing refractive indeces of the interface) decays exponentially as the distance increases away from the coverslip.

Figure 4.15: Total Internal Reflection Microscopy

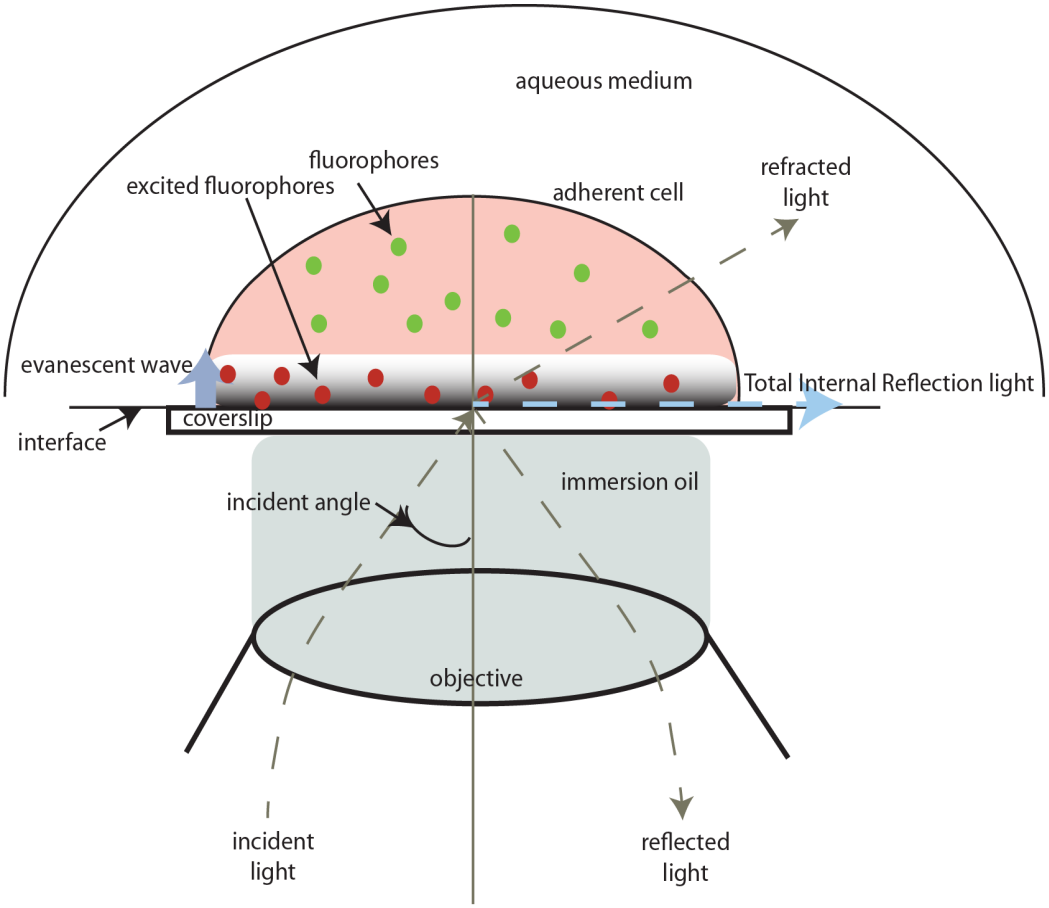


Figure 4.15 Total Internal Reflection Microscopy
Total internal reflection microscopy (TIRF-M) is an imaging technique that makes use of the difference in refractive index between two media (aqueous solution and glass) to image a very narrow range at high resolution. The laser light is directed at a particular angle (greater than a critical angle) at the interface of the coverslip and the aqueous solution, and the mismatch between the refractive indices of the two media creates total internal reflection of the light, and a residual evanescent wave, or electromagnetic field, that extends slightly into the specimen. An evanescent wave is formed at the boundary between two media, and has the same properties as the light that created it, decaying exponentially as the distance increases from the interface (in practice, around 100nm). Fluorophores are excited by the evanescent wave in the the same way as they would be by the laser. Restriction of the excitation field to such a narrow range reduces the interference from other fluorophores or cell autofluorescence outside the imaging field, increasing resolution.

Image adapted from: www.ncl.ac.uk/bioimaging/techniques/tirfm/

The depth can be altered by changing the angle of the laser beam, producing a larger or smaller volume for the excitation wave. In principle, the highest resolution is restricted to a depth of approximately 100nm.

We used TIRF-microscopy to image MKS3 at or just beneath the plasma membrane. TIRF microscopy and quantification analysis was carried out with the help of Prof R Duncan, Heriot Watt University. MCF7 (American) cells were co-transfected with GFP empty vector (pAcGFPN1) and either V5-tagged MKS3 wild type (human 995 residue isoform), V5-tagged MKS3 (PTPIFA) mutant, or V5-tagged MKS3 (PTTIYY) double mutant and incubated overnight on coverslips. Cells were then fixed, permeabilised and incubated with mouse monoclonal anti-V5 (Abcam 27672) and rabbit anti-E-cadherin primary antibody, before adding fluorophore-conjugated secondary antibodies. Coverslips with fixed and stained cells were then placed in a chamber filled with aqueous solution (PBS) and analysed by Total Internal Reflection Fluorescence Microscopy. E-cadherin was used as a positive control to ensure that the cell membrane was being accurately visualised and to control for detachment of the cells from the coverslip, and expressed GFP was used to locate and define the boundaries of transfected cells (data not shown). V5-vector and secondary antibody-only controls were also imaged for unspecific primary or secondary antibody binding (previously shown in these MCF7 cells in Figure 4.13 and Figure 4.2C), and only a negligible signal was detected (data not shown).

V5-tagged MKS3 was clearly visualised in transport vesicles at or just beneath the plasma membrane. TIRF microscopy allows visualisation of the adherent plasma membrane as described above, suggesting that MKS3 may

localise to the basolateral plasma membrane (although MCF7 cells do not normally thought to be polarised). MKS3 has not been described at the basolateral membrane previously, and therefore this may represent a novel cellular localisation for the protein. A representative image of vesicles of wild type V5-MKS3 can be seen in the large image, with representative images of vesicles for each of the V5-tagged MKS3 forms shown in the inset boxes, as labelled (Figure 4.16A). Of note is that the V5-MKS3-containing vesicles seen at the plasma membrane for the PTPIFA mutant are not absent (although the signals were of much lower intensity, gain on representative image shown here is increased to allow visualisation). When the data were examined for number of vesicles (or vesicle-sized signals) at the plasma membrane, the number of signals seen for the PTPIFA mutant was increased compared to the wild type or the PTTIYY double mutant (vesicle counting was automated) as shown in Figure 4.16B. Closer inspection of the vesicles seen in Figure 4.16A reveals that the "vesicles" are much larger than in either the PTTIYY or wild type-transfected forms of V5-MKS3, and that they appear as diffuse, aggregated bundles rather than the clear, discrete vesicles seen for the PTTIYY or wild type. Widefield images for each of the isoforms were also taken immediately after acquiring the TIRF-M images and then deconvolved to produce 3D images of the cell (Figure 4.16C). The distribution pattern of V5-MKS3 (PTPIFA) in these cells is clearly different than wild type MKS3, with evidence of a diffuse localisation pattern compared to the distinct vesicular and ER/Golgi-like localisation pattern of the wild type. This suggests that the PTPIFA V5-MKS3 mutant, whilst still able to reach at least one of its target destinations, may be improperly folded leading to aggregates of misfolded protein.

Figure 4.16: Differential localisation of MKS3 wt and linear-peptide mutant in MCF7 breast cancer cells by TIRF-M (Total internal reflection microscopy)

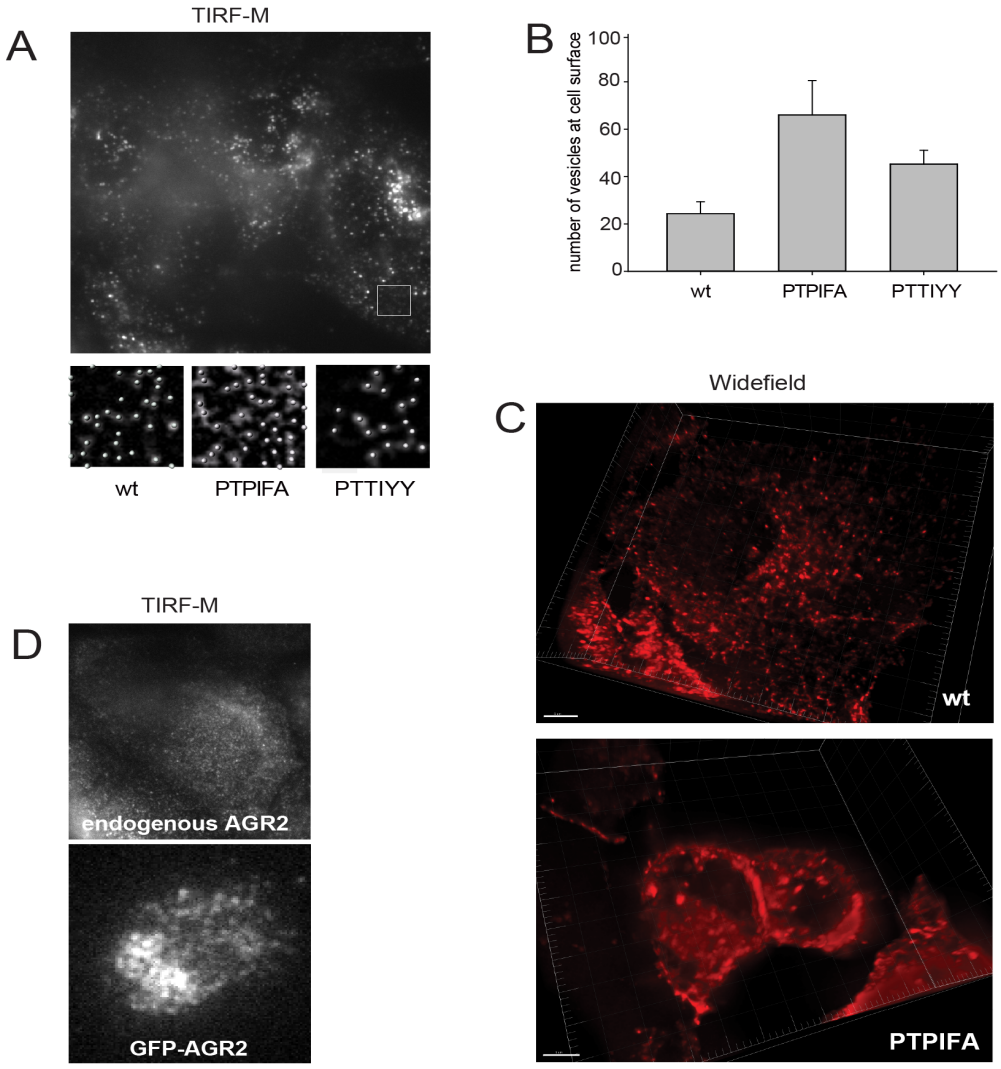


Figure 4.16 Differential localisation of MKS3 wt and linear-peptide mutant in MCF7 breast cancer cells by TIRF-M (Total internal reflection microscopy)
American MCF7 breast cancer cells were transfected with empty GFP vector ($pAcN_1GFP$) and V5-tagged MKS3 wild type or mutants (V5-MKS3^{PTTIYY}, V5-MKS3^{PTPIFA}), and fixed and antibody-stained cells (anti-V5 mouse monoclonal) were analysed by Total Internal Reflection Microscopy (TIRF-M). Cells were localised by widefield GFP imaging and adherence to the coverslip was verified with a membrane marker by TIRF-M (not shown). Adherent, GFP-positive cells were analysed for plasma membrane localisation of V5-MKS3 wild type or mutants (A). Number of vesicles at the membrane containing V5-MKS3 were analysed using a custom automated software (Heriot Watt University, ProfR Duncan) (B). Cells analysed with TIRF-M were also imaged in widefield and deconvolution provided a 3D image (C). Cells were also either transfected with GFP-tagged AGR2 or not transfected and analysed for localisation of AGR2 at the plasma membrane with anti-GFP mouse monoclonal antibody (D, bottom panel) or anti-AGR2 mouse monoclonal antibody (D, top panel). *TIRF microscopy and analysis provided by Prof R Duncan, Heriot Watt University*

This is supported by the migration pattern of V5-MKS3 from transfected lysates (Figure 4.12), which shows more aggregated protein at the top of the wells in the PTPIFA mutant compared to the wild type or the PTTIYY mutant. What is more difficult to assess, even using TIRF-M, is whether or not the PTPIFA mutant is able to insert into the plasma membrane - this would require correct insertion of the nascent protein into the ER membrane as described above.

We also wanted to assess whether AGR2 was localised in vesicles at or beneath the plamsa membrane, in a similar compartment to V5-MKS3 as seen by TIRF-M. Cells were transfected with C-terminally tagged GFP-AGR2 or GFP vector only and incubated overnight. Cells were then fixed and permeabilised as described above and visualised using TIRF-M also as described. GFP-tagged AGR2 was seen to localise to the plasma membrane in vesicles (Figure 4.16D, bottom panel). Similarly, endogenous AGR2 was also seen to localise to vesicles at the plasma membrane. In this case, fixed and stained MCF7 (American) cells were incubated with primary antibody (anti-AGR2 monoclonal) and Alexa secondary antibody to allow visualisation of the endogenous protein. Although it is encouraging that AGR2 is seen at the plasma membrane in secretory vesicles in a similar way to MKS3, it is not possible to colocalise the signals using TIRF-M, and therefore no specific conclusions can be drawn about whether AGR2 and MKS3 are directly binding in this cellular compartment.

4.4 Discussion

In this chapter evidence has been presented to support a possible interaction between a linear peptide motif-containing protein, identified through phage-display proteomics, and its binding partner. Specifically, we have shown that AGR2 partially co-localises with a linear peptide *in vivo* (supporting previous data showing that the linear peptide and AGR2 bind *in vitro*), and that this linear peptide is found in a number of membrane and secretory proteins in the human genome. In addition, we have shown that AGR2 and overexpressed MKS3 can be co-immunoprecipitated from cell lines, and that overexpression of MKS3 alters the balance of subcellular localisation of AGR2, concentrating it into the membranes and organelles. We have also shown that mutating the linear peptide motif, identified from the phage display proteomics, in MKS3, affects both the degree of colocalisation of MKS3 and AGR2, and the cellular localisation of MKS3 as seen by TIRF-microscopy. The evidence presented in this chapter does not conclusively demonstrate an interaction between MKS3 and AGR2, however, it does support a possible interaction between these two proteins. Further indirect evidence of a possible interaction between AGR2 and MKS3 will be presented in Chapter 6, and future work will be required to show a functional interaction between the two proteins. This will include siRNA-mediated depletion of AGR2 to investigate the effect on MKS3 localisation by TIRF-M (although this would only support a functional interaction rather than direct binding). FRET studies to evaluate the binding of AGR2 to MKS3 wild type and mutants at the plasma membrane is also planned, as is depletion of AGR2 in epithelial, ciliated cell lines such as IMCD3 cells (discussed later), and analysis of cilia phenotype. The idea of AGR2 as a molecular chaperone has

been suggested before [Park et al., 2009], although this has not been proven. These data lend support to the model of full-length AGR2 as a chaperone protein resident in the ER and involved in the processing and maturation of large, cysteine-rich membrane or secretory proteins.

The discovery of a putative novel interacting partner for the disease protein AGR2 is potentially important both for helping us to understand the function of AGR2 and for providing a model system for investigating AGR2-linked pathways. The discovery is potentially all the more important in that it provides validation that phage-display proteomics could be harnessed to identify interacting partners and networks on a larger scale. Furthermore, it may add to the growing field of interactomics in which protein-protein interactions are increasingly seen to be mediated by linear peptides binding to larger globular domains.

In the process of investigating the interaction between AGR2 and MKS3 we discovered that commercially available anti-Meckelin antibodies did not perform well, at least when used in our assays. Indeed, it is for this reason that much of the work described in this chapter was carried out using tagged, overexpressed Meckelin. We felt that having a reliable antibody made would be useful for our studies, and therefore designed a strategy to have a new polyclonal antibody made by a commercial group. The next chapter describes this strategy and the subsequent characterisation of two new anti-Meckelin polyclonal antibodies.

Chapter 5

Design and Characterisation of Novel Anti-MK3 Antibody

5.1 Introduction

Antibodies are small multi-subunit protein complexes that are produced by both higher and lower vertebrates [Marchalonis and Edelman, 1965] as part of the immune system. They are also known as immunoglobulins, and are Y-shaped glycoproteins produced by B-cells [Janeway et al., 2001]. Their basic structure is composed of four polypeptide chains, with two heavy chains and two identical light chains held together by disulphide bonds [Tonegawa, 1983]. In mammals, there are five different types of heavy chain, and these define the class of antibody (IgA, IgD, IgE, IgG, IgM). IgG is the main class of antibodies responsible for detecting and binding to invading pathogens [Burry, 2010]. The variable region on the tips of the "Y" shape represents the antigen-binding domain [Hilschmann and Craig, 1965]. This region is made by somatic

recombination events in B-cells (cells of the immune system in vertebrates), also called *VJD recombination* [Tonegawa, 1983], and the variable region may recognise and bind a peptide from the target protein, or several disjointed sequences from a folded protein forming a single epitope [Burry, 2010]. Randomly produced antibodies made by VJD recombination in the B cells of the bone marrow circulate in the body at low concentrations. When a pathogen such as a bacterium or virus is present at high enough levels in the body, an antibody containing a variable region that specifically recognises a part of that antigen (the epitope) binds the antigen and begins a chain of events that may include clonal expansion, where the B-cells begin to manufacture high levels of that particular antibody containing that specific variable region. In this way the circulating concentration (or titre) of the antibody is increased [Janeway et al., 2001]. Vaccination works in this way, exposing the body to a particular foreign antigen to increase the circulating titre of its corresponding antibody so that the body is pre-primed for defense in the event of exposure.

In the last few decades, it has become possible to harness the ability of vertebrates to produce specific antibodies in response to antigen exposure, in order to make both polyclonal and monoclonal antibodies as a tool for use in research [Kennett, 1979] [Kohler and Milstein, 1975]. When an antigen is circulating, more than one variable region antibody may bind several different epitopes on the same antigen. If these are purified as a single pool, the result is a "polyclonal" antibody, one preparation that contains more than one antibody to the same antigen. If, however, each antibody that binds is individually purified such that each preparation contains antibodies all with the same variable region, then this is known as a "monoclonal" antibody. Monoclonal and

polyclonal antibodies are useful tools in research in that they bind an epitope strongly. Numerous downstream applications have been designed including their use in immunofluorescence, immunohistochemistry and biochemical techniques such as co-immunoprecipitation. In order to make an antibody towards a particular antigen, usually a protein, either the entire protein or a peptide from that protein can be used. The latter is especially useful when purifying a protein in its entirety is not possible. A purified protein or a polypeptide chain from the protein is used to immunise animals, often rabbits or mice. The protein or polypeptide is injected into the rabbit or mouse and this is repeated two, three or four times over a period of time to increase the exposure of the animal to the antigen and increase the antibody titre, much the same way as repeated vaccination. In the case of making a polyclonal antibody, serum from the immunised animal is tested for immune response and may then be affinity purified to remove any other contaminating immunoglobulins (Ig) that are present in the serum of the animal [Harlow and Lane, 1988].

We decided to make a polyclonal antibody against murine MKS3 (meckelin) for use in our research. When commercially available antibodies as well as two custom antibodies (Prof C Johnson, Leeds University), were tested by western blot in various cell lines, we found that all tested demonstrated poor specificity in our hands (Figure 5.1A,B,C). MCF7 (American) cells (Figure 5.1A,B) or a panel of cell lines including ciliated cell lines (Figure 5.1C) were transfected with HA-tagged MKS3 or not transfected, and then lysed in urea lysis buffer. Lysates were loaded onto a 10% reducing polyacrylamide gel and separated by SDS-PAGE before being transferred to nitrocellulose membrane and probed with various antibodies. The anti-MKS3 N-terminal polyclonal

Figure 5.1: Specificity and sensitivity of commercial anti-MKS3 antibodies

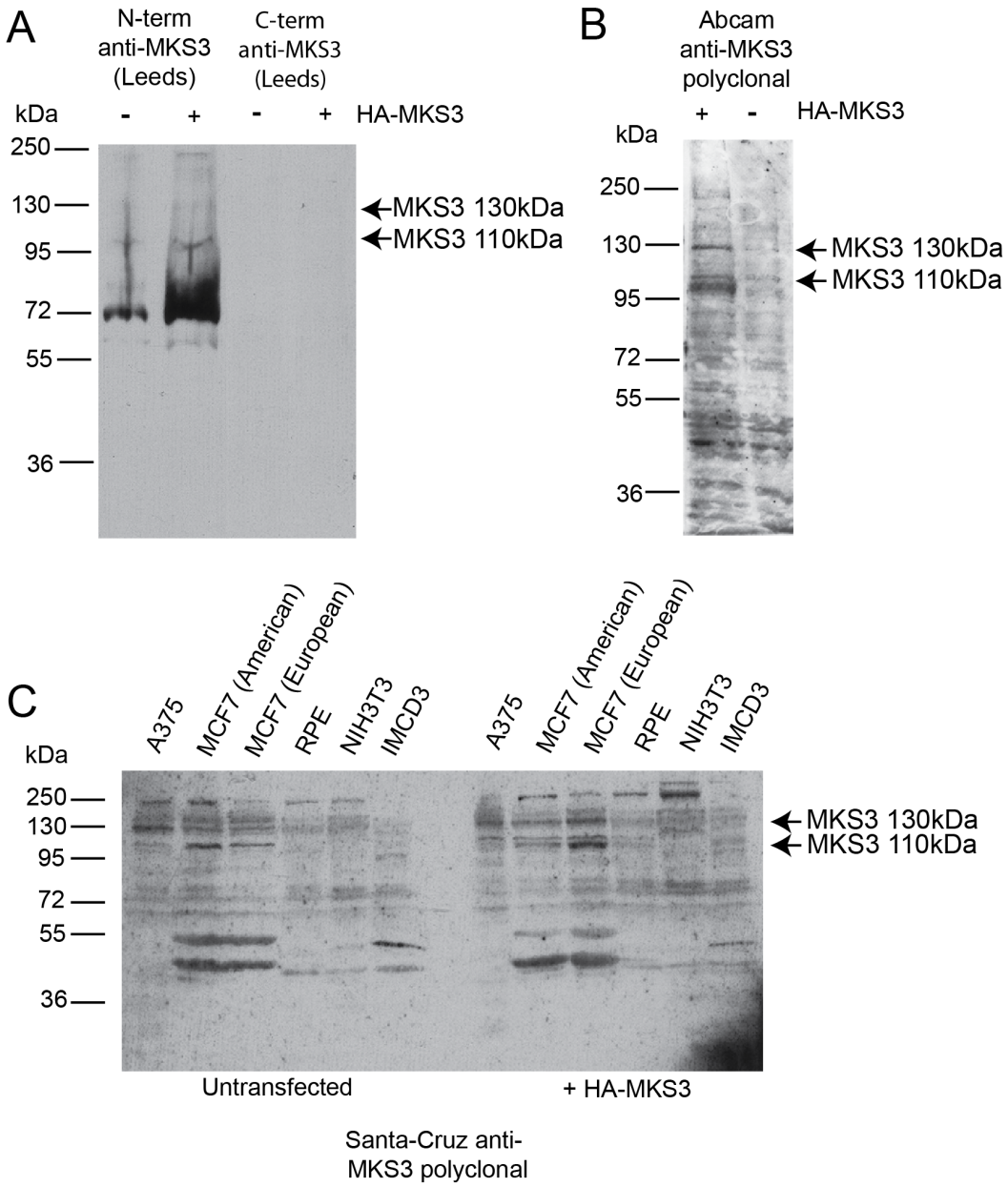


Figure 5.1 Specificity and sensitivity of commercial anti-MKS3 antibodies
MCF7 (American) cells were transfected with HA-tagged MKS3 or not transfected and lysates were analysed by western blot using (A) custom rabbit polyclonal anti-MKS3 antibody to the N-terminus or C-terminus at dilutions of 1:100 (C Johnson, Leeds), or (B) commercial rabbit polyclonal anti-MKS3 at a dilution of 1:1000 (abcam). (C) A panel of different cell lines as shown were transfected with HA-tagged MKS3 or not transfected and lysates were analysed by western blot using a commercial rabbit polyclonal anti-MKS3 at a dilution of 1:500 (Santa Cruz).

antibody from C Johnson did identify a band of appropriate size (Figure 5.1A), however, a stronger band was also detected at 72KDa. It is possible that this band corresponds to a processed form of meckelin, however, published and predicted band sizes by western blot for MKS3 are 110-130KDa in size as described in chapter 4. The affinity-purified C-terminal antibody did not detect any band at all by western blot (although see Figure 4.11 for an example of its use in immunofluorescence studies). Many antibodies are suitable for one application but not others, such as is the case for this C-terminal antibody, and this may be due to differences in pH associated with different techniques, or accessibility of an epitope after denaturing the target protein or fixation with cross-linking agents. Both the commercial antibodies tested were similarly lacking in specificity (Abcam and Santa Cruz), and whilst it is possible to detect bands at appropriate sizes, the numerous other bands did not make these antibodies useful in our assays.

5.2 Strategy for design of polyclonal antibody to MKS3 (meckelin)

The mouse *Meckelin* gene spans 29 exons across approximately 50 kilobases. It encodes several transcripts, two of which are predicted to be protein coding. The shortest of these encodes a 995 amino acid protein from 28 exons, and the longer one encodes a predicted 1061 amino acid protein from 29 exons, with an extra exon at the 5' end and the protein differing only at its N-terminal (www.ensembl.org Mouse). The significance of the longer protein in mouse has not been reported, and the longest human transcript for *Meckelin* encodes

a 995 residue protein (www.ensembl.org Human). This human protein shares close homology to the mouse 995 residue protein (85% identity, 93% similarity) and for this reason we decided to make antibodies to this isoform of MKS3. Mouse and human *Meckelin* have three to seven predicted transmembrane domains and contain 26 cysteines in the predicted extracellular and membrane spanning domains, as well a signal peptide at the far N-terminal domain that may target it to the ER and secretory pathway (Figure 5.2A). As mentioned above, antibodies can be made to an entire protein, or to polypeptides from the N-terminal, C-terminal or intra-protein regions. In our earlier investigations into the interaction between MKS3 and AGR2 we attempted to purify MKS3 however, it was not possible to purify the protein in its entirety for technical reasons despite repeated attempts. Large, cysteine-rich eukaryotic membrane proteins are notoriously difficult to purify in bacterial cells due to the lack of specialised folding machinery, and the reducing conditions of the cytoplasm [Schwarz et al., 2008]. This can sometimes be overcome in insect cells, however, we were ultimately unsuccessful in expressing MKS3 in either bacterial or insect cells (in fact, the overexpression of MKS3 appeared toxic to *E. Coli* as the cells did not grow well; data not shown). In order to overcome this, we decided to use peptides from the N-terminal and far C-terminal of MKS3 as antigens for immunisation and antibody production, as shown in Figure 5.2A. The N-terminal peptide chosen was outwith the predicted signal peptide region, encompassing residues 52-62 in the 995 amino acid isoform of MKS3. The C-terminal peptide we used for antibody production comprised the final 10 residues of the protein, amino acids 986-995. These residues were chosen in part for their location (more likely to be accessible in the native protein),

Figure 5.2: Design Strategy for production of antibodies to mouse MKS3

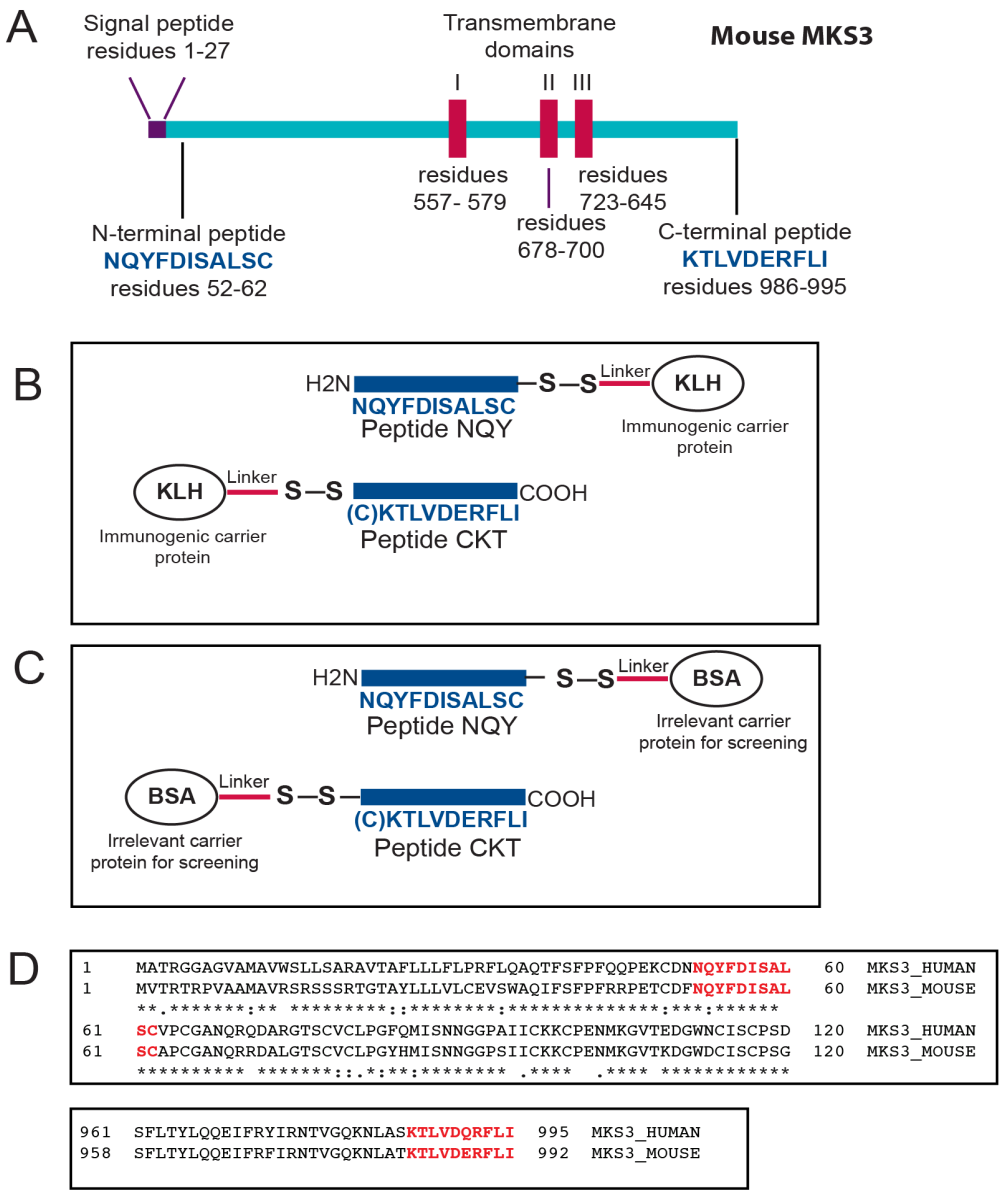


Figure 5.2 Design Strategy for production of antibodies to mouse MKS3
(A) Putative membrane domains of the multi-pass transmembrane protein MKS3 and position of the N-terminal and C-terminal peptides used for antibody generation. (B) Peptides used as antigens for antibody production with immunogenic carrier KLH covalently linked to the peptides. (C) Peptides used as antigens covalently linked to irrelevant protein bovine serum albumin (BSA) for adsoprtion to membranes or microtitre plates used for screening immunogenic response. (D) Comparison of the mouse and human sequences of the antigenic N-terminal and C-terminal peptides from MKS3 used for antibody production. *Alignment reproduced from Uniprot*

and also in part due to their homology with the human protein (Figure 5.2D), in order to increase the chance of cross-reactivity of any antibodies to both human and mouse meckelin. In the case of the N-terminal peptide, a cysteine residue was already present at the C-terminal end of the peptide, whereas for the C-terminal peptide, a cysteine was added to the N-terminal end of the peptide. The sulfhydryl group on these cysteines were used to link an immunogenic carrier via a disulphide bond.

Short peptide antigens that are used for immunisation are often linked to another moiety to increase the immunogenicity of the antigen - in other words, to increase the level of immune response. Keyhole limpet hemocyanin, or KLH, is a copper-containing polypeptide from the species *Megathura crenulata* of the monotypic genus Megathura, that can be linked by a disulphide bond to a peptide antigen to increase immunogenicity [Helling et al., 1994]. It can be used as a vaccine carrier, and is highly immunogenic due to its large size and complex structure. It is composed of two subunits of 350KDa and 390KDa each, but forms aggregates of 0.5 - 8 million KDa [Coligan et al., 1996]. The KLH protein was linked to each of our MKS3 peptides as shown in Figure 5.2B, and this was used for immunisation of rabbits by a commercial company, Moravian Biotech. In order to aid in screening the serum of the rabbits for immune response to the immunised peptides, our peptides were also linked to Bovine Serum Alubumin (BSA) (Figure 5.2C). BSA can be used as an immunogenic carrier in its own right, however it is often used for screening purposes. Short peptides do not adsorb as well as larger proteins to microtitre plates, however, using KLH-linked peptide in our case was not appropriate for screening serum since antibodies to the carrier (or *hapten*) will also be present

in the serum of the immunised animal. BSA can be used as an irrelevant carrier to aid the adsorption of peptides to a plate, but without binding to antibodies present in the serum [Coligan et al., 1996]. In this way, specific immune responses to the short polypeptides rather than to the KLH carrier can be discerned.

5.3 Serum testing and affinity purification

After three rounds of immunisation to increase the titre of any antibodies to the peptides, serum was tested by Moravian Biotech for immune response to the peptides using a dot-blot (Figure 5.3A). This involves spotting peptide onto a membrane and letting it adsorb, and then testing the serum from immunised animals for specific binding to the membrane. Serum from reactive animals was isolated and used for further initial screening (Figure 5.3B,C). MCF7 (American) cells were transfected with HA-tagged MKS3 and lysed in urea lysis buffer. Lysates were analysed by western blot and serum from an animal that had been immunised with either the C-terminal or N-terminal peptides linked to KLH were used to probe the membranes. The potential antibodies were designated CKT (for the C-terminal antibody) or NQY (for the N-terminal antibody), and these were tested at dilutions of 1:100 or 1:500. As controls, blots were also probed using the pre-immune sera (serum from the same animal, before immunisation with the peptide), and also with commercial anti-MKS3 (Abcam). In the case of both CKT (Figure 5.3B) and NQY (Figure 5.3C) antibody-containing sera there was evidence of bands at the approximate height expected that were not present in the pre-immune sera.

Figure 5.3: Immune response to injected peptides and serum testing

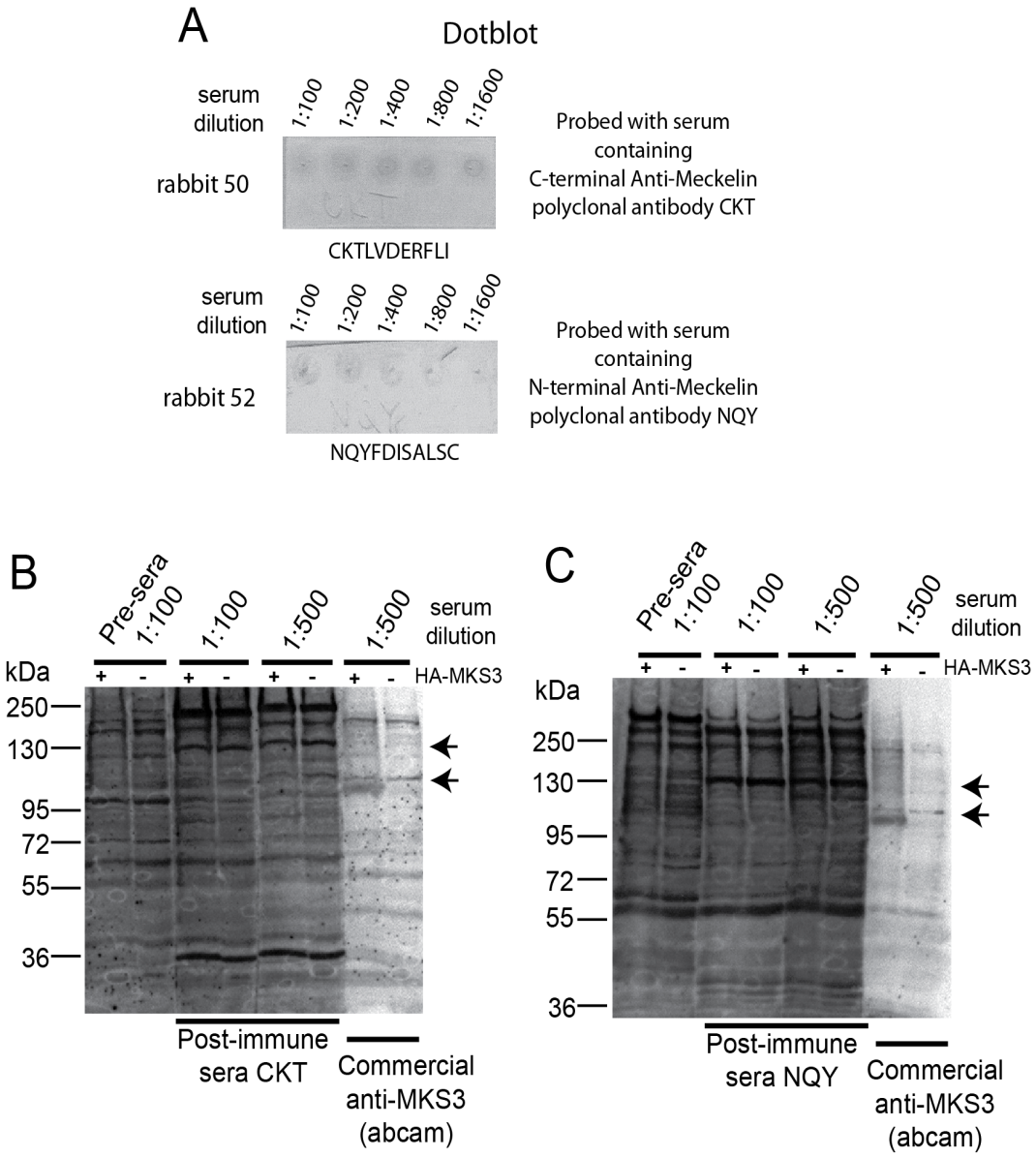


Figure 5.3 Immune response to injected peptides and serum testing
BSA-linked peptides were spotted onto membranes and probed with serum from immunised animals containing C-terminal or N-terminal anti-MKS3 antibody. Dotblots shown are for the serum from two animals (50, 52) with immunogenic response to peptides (A). MCF7 (American) cells were transfected with HA-MKS3 or not transfected and lysates were analysed by western blot. Membranes were cut and probed using serum containing pre-immune serum, anti-MKS3 CKT (C-terminal) (B) or anti-MKS3 NQY (N-terminal) (C), or commercial anti-MKS3 (abcam).

As there were many other bands also present, we decided to have the antibodies affinity purified from the sera to increase specificity to MKS3.

Affinity purification is a process whereby a protein is concentrated from a preparation by taking advantage of the binding properties to another molecule, often another protein. In the case of antibody purification, the antigen may be immobilised onto a column such as resin using a variety of immobilisation techniques, and the serum containing the specific antibody washed through the column. Any antibodies specific to the antigen are bound and remain in the column, whilst all other proteins and antibodies present in the serum do not bind and wash through. The bound antibodies are then eluted in fractions, for example by decreasing the pH until a range is found that inhibits binding of the antibody to the antigen, allowing the antibody-antigen complex to dissociate. Affinity purification was carried out in this way by Moravian Biotech. The peptide antigens were covalently coupled to a CDI-activated resin via the free amine group at the N-terminal (Reactigel, Pierce). Serum from the immunised and selected rabbits was run through the column, allowing the antibody to the peptide to bind the column. Elution of the IgG (antibody) was carried out by reducing the pH (0.1M Glycine, pH 2.2) and preparation was then neutralised with 1M Tris buffer (pH 8.8). Fractions were analysed by SDS-PAGE for immunoglobulin heavy and light chains, and two fractions were chosen for further analysis based on the results (Figure 5.4A); fractions 1.5 and 2.4 for antibody CKT, and fractions 1.4 and 2.3 for antibody NQY.

Figure 5.4: Optimisation of affinity purifed polyclonal antibodies to mouse MKS3 by western blot - Part I

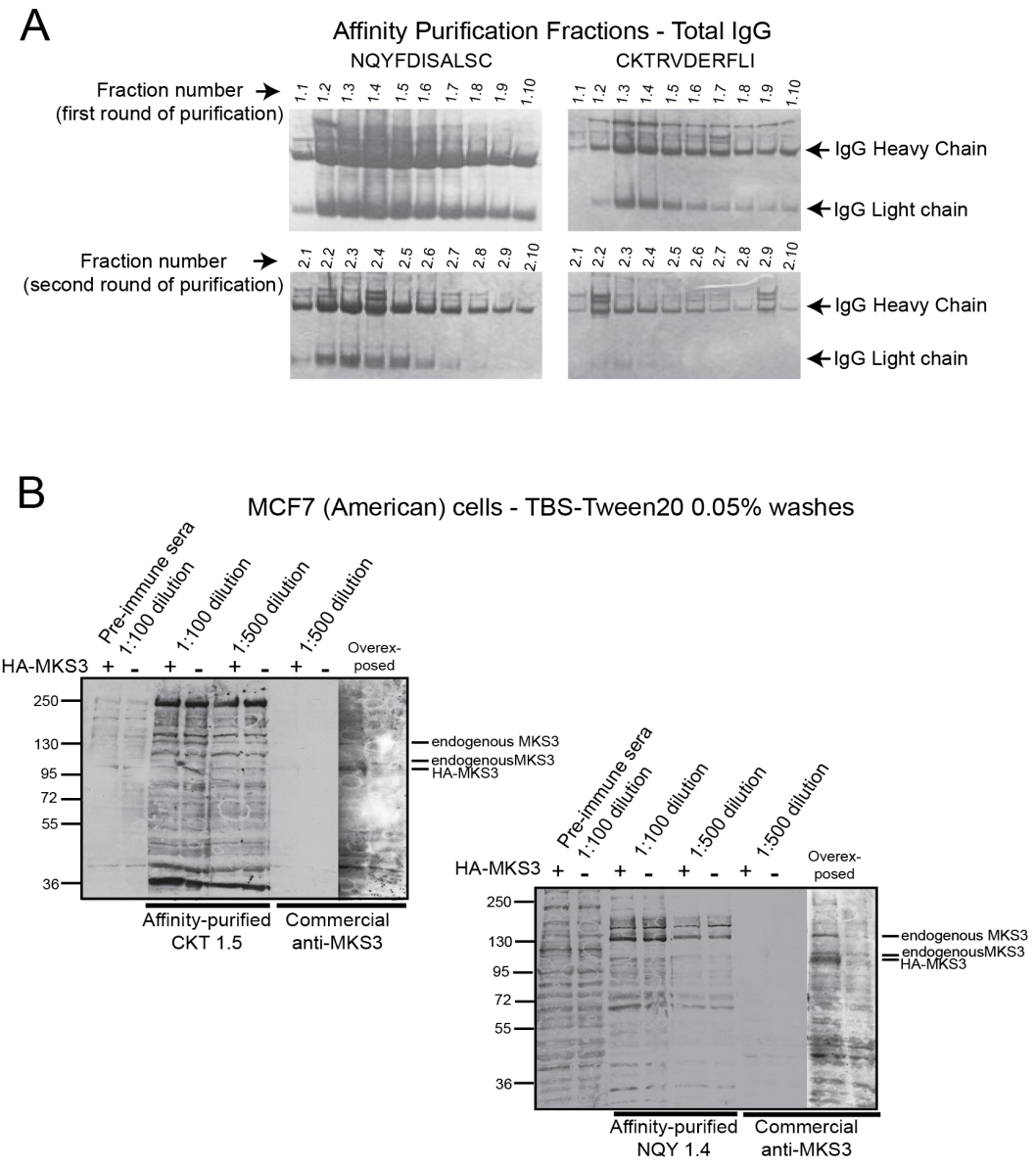


Figure 5.4 Optimisation of affinity purifed polyclonal antibodies to mouse MKS3 by western blot - Part I
(A) Serum from immunised rabbits 50 and 52 was affinity purified using CDI-activated resin (Reactigel, Pierce), eluted at low pH, and fractions were analysed by SDS-PAGE for presence of immunoglobulin heavy and light chains. (B) MCF7 (American) cells were transfected with HA-MKS3 or not transfected and lysates were analysed by western blot. Membranes were cut and probed with pre-immune sera, affinity-purified anti-MKS3 CKT antibody (left panel) or affinity-purified anti-MKS3 NQY antibody (right panel), or commercial anti-MKS3 (abcam).

5.4 Optimisation of affinity purified antibodies by western blot

Initially, the affinity purified antibodies for the C-terminal (CKT) and N-terminal (NQY) regions of MKS3 were assayed by western blot at low dilutions (1:100-1:500) and using TBS-Tween20 for the washes (Figure 5.4B). 2% BSA in TBS was used for blocking and for diluting antibody. These conditions allowed both antibodies to detect bands at the appropriate height (110-130KDa) and matching bands detected with the commercial anti-MKS3 antibody (Abcam) that were not present in the pre-immune sera. Due to the non-specific bands present, optimisation of these conditions was required, and this included varying dilutions (1:100 - 1:5000), blocking conditions (milk or BSA, 2% - 5%), wash conditions (TBS or PBS), and detergents used for washing (TritonX100 or Tween20, 0.05% - 0.1%). For both antibodies, the optimal conditions were found to be as follows: antibody dilution of 1:1000 in 2% milk and PBS washes with 0.1% TritonX100. Conditions were tested for lysates from both MCF7 (American) and ciliated cell line IMCD3 as summarised in Figure 5.5 A (antibody CKT, fractions 1.5 and 2.4), and Figure 5.5 B (antibody NQY, fractions 1.4 and 2.3). In particular, antibody CKT fraction 1.5 was found to detect MKS3 in both MCF7 (American) cells and IMCD3s well with little unspecific reactivity. The identity of the larger bands detected by this antibody is not known, however, these bands are detected by both antibodies and in both cell types, despite their distinct epitopes, perhaps suggesting a modified form of MKS3 or aggregates due to boiling that causes many membrane proteins to aggregate.

Figure 5.5: Optimisation of affinity purified polyclonal antibodies to mouse MKS3 by western blot - Part II

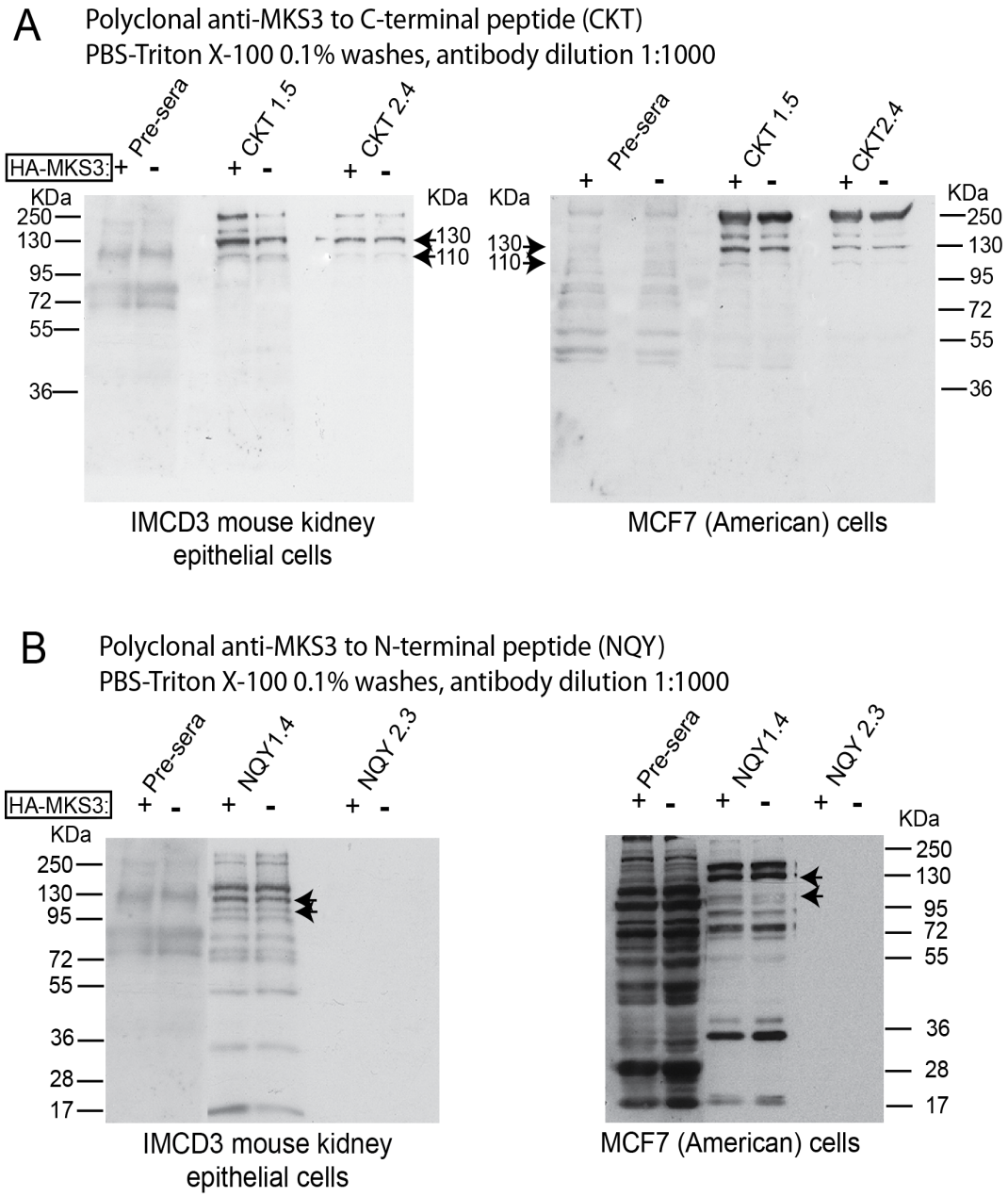


Figure 5.5 Optimisation of affinity purified polyclonal antibodies to mouse MKS3 by western blot - Part II
MCF7 (American) breast cancer cells or IMCD3 kidney epithelial cells were transfected with HA-MKS3 or not transfected and lysates were analysed by westernblot. Membranes were cut and probed with pre-immune sera, affinity-purified MKS3 CKT antibody (A) or affinity-purified MKS3 NQY antibody (B). Shown are blots using optimised wash and antibody dilution conditions. Arrows refer to expected sizes of MKS3 of 110KDa and 130KDa.

In summary, the antibody produced to the C-terminal of meckelin was optimised and found to be appropriate for use in assays involving detection by western blot. The antibody produced to the N-terminal of meckelin was also optimised, and one fraction, from the affinity purification column, fraction 1.4, was found to be suitable for analysis by western blot although with some degree of background detection present. The other N-terminal fraction tested, fraction 2.3, was not seen to detect MKS3 (or any other protein) by western blot, making it unsuitable for use in these assays.

5.5 Optimisation and assay of anti-MKS3 antibodies for immunofluoresence

We next decided to characterise the anti-MKS3 antibodies to the N-terminal and C-terminal of meckelin for use in immunofluorescence studies. As discussed in chapter 4, known intracellular locations for meckelin include the plasma membrane, ER and Golgi, and the primary cilium [Dawe et al., 2007b] [Dawe et al., 2009]. In particular, the ciliary basal body [Adams et al., 2012] and transition zone [Williams et al., 2011] have been described as a known location for meckelin. This provided us with a specific and readily-identifiable intracellular location to test the specificity of our antibodies by immunofluorescence. Primary cilia are long projections from the cellular plasma membrane with a variety of functions including cell signalling and development [Goetz et al., 2009] [Berbari et al., 2009], regulation of cell cycle [Alaiwi et al., 2009], and mechano-sensory roles [Praetorius and Spring, 2001]. All mammalian cells are believed to be able to form cilia, and in the case of primary cilia, only one cilia

exists per cell at any one time under normal circumstances. One ciliated cell line frequently used for cilia studies is a cell line derived from the mouse kidney known as IMCD3 cells (inner medullary collecting duct cells) [Rauchman et al., 1993]. These cells, when plated at high confluence, project long primary cilia that are easily visualised using antibodies. We used this cell line to visualise primary cilia and look for co-localisation of MKS3 at the cilium or basal body to confirm that the antibodies were suitable for immunofluoresence assays.

Each of the two fractions for each antibody assayed by western blot (antibody CKT fractions 1.5 and 2.4, antibody NQY fractions 1.4 and 2.3) were tested for suitability by immunofluorescence, using four different fixation and permeabilisation protocols and three different andtibody dilutions. The results are summarised in Figure 5.6, with one asterisk referring to some incidences of co-localisation of the antibody at the basal body, three asterisks referring to most or all ciliary basal bodies visualised co-localising with the antibody, and no asterisk depicting no co-localisation. All protocols included fixing in 3-4% paraformaldehyde (PFA), either with or without methanol. Permeabilisation was with either TritonX100 0.1% (fixation protocol 1), Triton X100 0.5% (fixation protocol 4), methanol and TritonX100 0.5% (fixation protocol 2) or glycine and saponin (fixation protocol 3) (see Materials and Methods Table 2.13). Anti-acetylated alpha tubulin allows detection of the cilia axoneme (the projection from the cell body), whereas anti-gamma tubulin identifies the basal body of the cilia. The purpose of this screen was to identify MKS3 in one of the known cellular locations as a means of validating the polyclonal antibodies for use in immunofluorescence assays. For this reason, only clear visualisation of the ciliary axoneme was required, with the basal body located

directly adjacent to the proximal end of the axoneme [Marshall, 2008].

All fixation protocols were found to be compatible with the mouse anti-acetylated alpha tubulin antibody used (Figures 5.7, 5.8, 5.9, 5.10). Affinity purified antibody CKT (fraction 1.5) to the C-terminal peptide of meckelin was found to clearly co-localise with the base of the cilia but only using fixation protocols 1 and 2, which involved fixation in PFA with or without methanol and followed by permeabilisation in TritonX100 (0.1%) (Figure 5.7A,D,G,J). By contrast, fix protocols 3 (PFA fixation followed by glycine/saponin permeabilisation) and 4 (PFA fixation followed by permeabilisation in high concentration of TritonX100 0.5%) did not allow detection of MKS3 using this antibody (Figure 5.8 A,D,G J). Interestingly, another fraction from the purification column of the same antibody (CKT, fraction 2.4) did not detect meckelin using any of the fixing protocols (data not shown), indicating that this fraction of the affinity purified serum is not suitable for immunofluorescence studies.

Affinity purified antibody NQY (fraction 2.3) to the N-terminal peptide of meckelin was also found to co-localise with the base of cilia, although this time only with fixation protocol 4 (PFA fixation followed by permeabilisation in 0.5% TritonX100) (Figure 5.9 G,J). Fixation protocols 1, 2 and 3 as described above did not allow detection of meckelin at the base of the ciliary axoneme (Figure 5.9 A, D and Figure 5.10 A, D, G, J). In addition, no antibody signal at the basal body was detected for affinity purified fraction 1.4 of the antibody to the N-terminal of meckelin using any of these fixation protocols or at any of the antibody dilutions assayed (data not shown).

We had shown that both antibodies detected a protein at the base of the ciliary body, however, there are numerous proteins at the base of the

Figure 5.6: Optimisation screening of affinity purified polyclonal antibodies to mouse MKS3 for immunofluorescence

Fixation		CKT 1.5	CKT 2.4	NQY 1.4	NQY 2.3
PFA + 0.1% Triton	1:100	-	-	*	-
	1:500	*	-	-	-
	1:1000	***	-	-	-
PFA + Methanol	1:100	-	-	-	-
	1:500	*	*	-	-
	1:1000	**	*	-	-
PFA + Permeabilisation Glycine/Saponin	1:100	-	-	-	-
	1:500	-	-	-	-
	1:1000	-	-	-	-
PFA + Permeabilisation 0.5% Triton	1:100	-	-	-	-
	1:500	-	-	-	*
	1:1000	-	-	-	**

Figure 5.6 Optimisation screening of affinity purified polyclonal antibodies to mouse MKS3 for immunofluorescence
IMCD3 mouse kidney epithelial cells were plated at high confluence and fixed with one of four fixation and permeabilisation conditions as indicated. Fixed cells were incubated with ciliary axoneme marker (mouse monoclonal anti-acetylated alpha-tubulin) and either affinity-purified anti-MKS3 CKT antibody fractions 1.5 or 2.4, or affinity-purified anti-MKS3 NQY antibody fractions 1.4 or 2.3, at antibody dilutions also as indicated. Cells were analysed for localisation of affinity-purified anti-MKS3 antibodies at the base or axoneme of the cilia, and scored for number of localising signals per field (*=low, **=moderate, ***=high/all visible cilia) or no localising signals to the cilia (-)

Figure 5.7: Affinity purified anti-MKS3 polyclonal antibody CKT (Fraction 1.5) localises to the ciliary basal body using PFA or PFA-methanol fixation

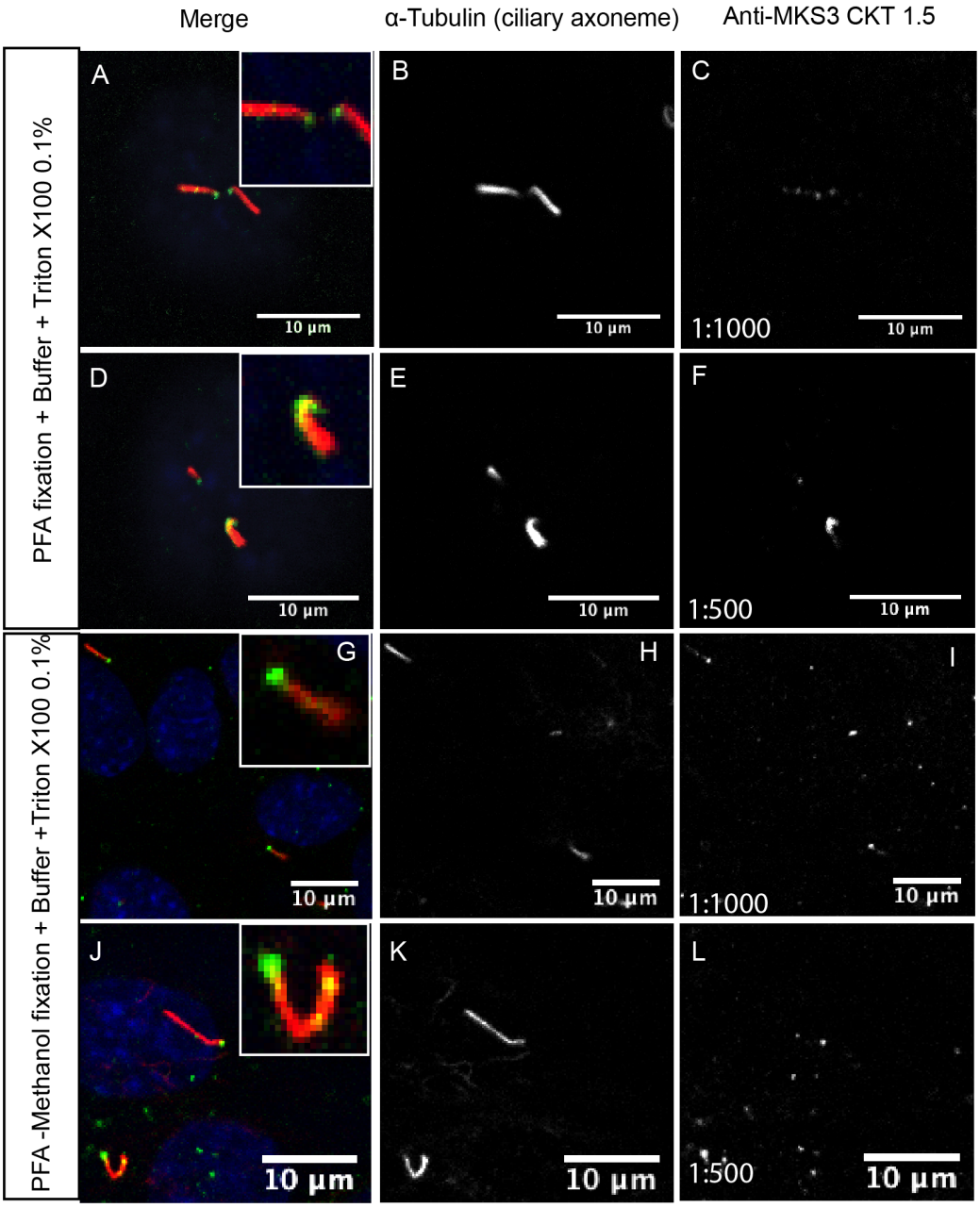


Figure 5.7 Affinity purified anti-MKS3 polyclonal antibody CKT (Fraction 1.5) localises to the ciliary basal body using PFA or PFA-methanol fixation
IMCD3 kidney epithelial cells were fixed and permeabilised with paraformaldehyde (PFA) alone and TritonX100 0.1% (A-F) or PFA/methanol and TritonX100 0.1% (G-L). Cells were incbuated with ciliary axoneme marker anti-acetylated alpha tubulin and affinity-purified anti-MKS3 CKT (fraction1.5) antibodies at antibody dilutions as indicated. Cells were assayed for localisation of anti-MKS3 antibody at the base or axoneme of the cilia.

Figure 5.8: Affinity purified anti-MKS3 polyclonal antibody CKT (Fraction 1.5) does not localise to the ciliary base or axoneme using Glycine-Saponin or 0.5% Triton permeabilisation

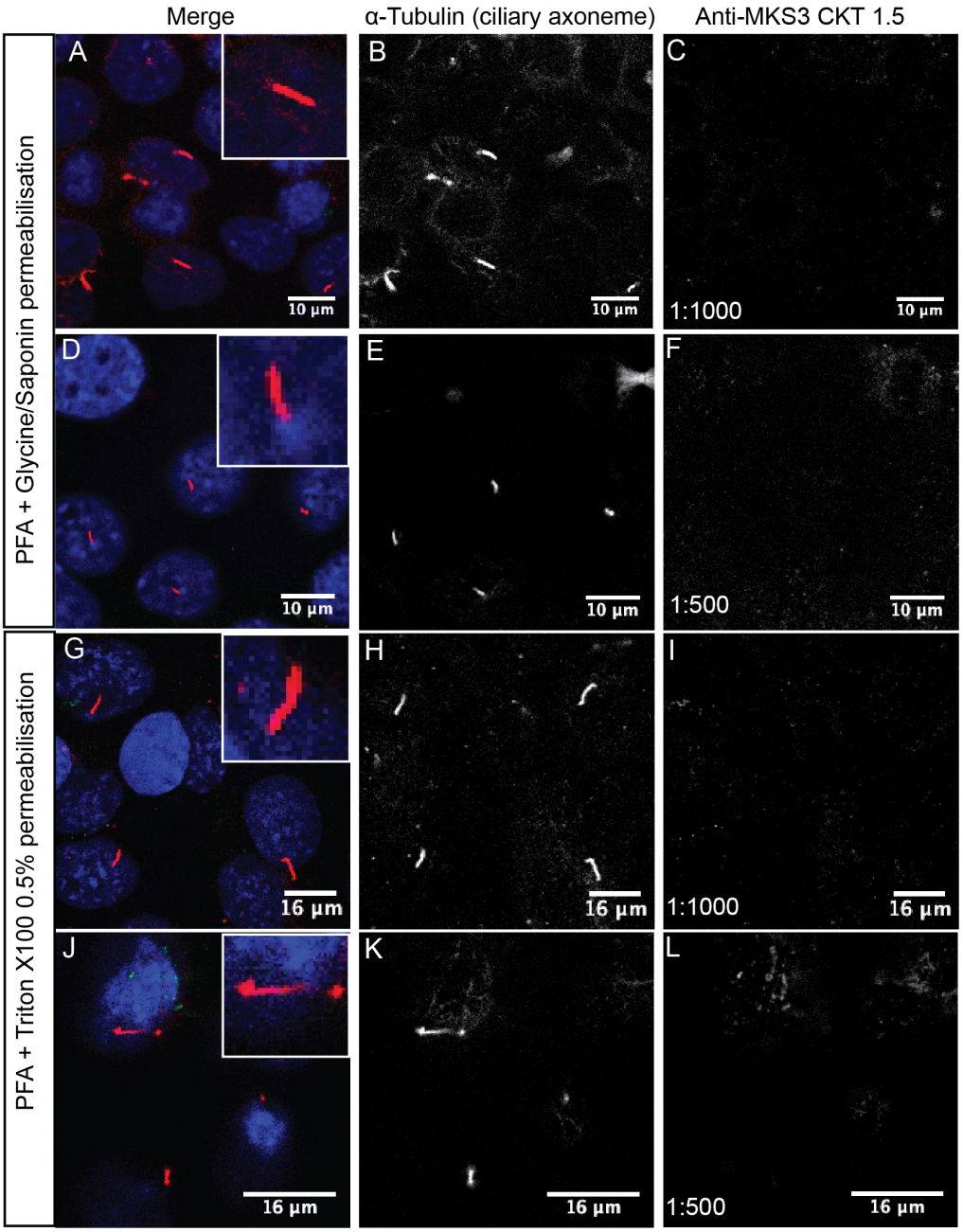


Figure 5.8 Affinity purified anti-MKS3 polyclonal antibody CKT (Fraction 1.5) does not localise to the ciliary base or axoneme using Glycine-Saponin or 0.5% Triton permeabilisation
IMCD3 kidney epithelial cells were fixed and permeabilised with paraformaldehyde (PFA) and Glycine/Saponin (A-F), or PFA alone and TritonX100 0.5% (G-L). Cells were incbuated with ciliary axoneme marker anti-acetylated alpha tubulin and affinity-purified anti-MKS3 CKT (fraction 1.5) antibodies at antibody dilutions as indicated. Cells were assayed for localisation of anti-MKS3 antibody at the base or axoneme of the cilia.

Figure 5.9: Affinity purified anti-MKS3 polyclonal antibody NQY (Fraction 2.3) localises to the ciliary base or axoneme using PFA fixation with 0.5% Triton permebilisation, but not with Glycine-Saponin permeabilisation

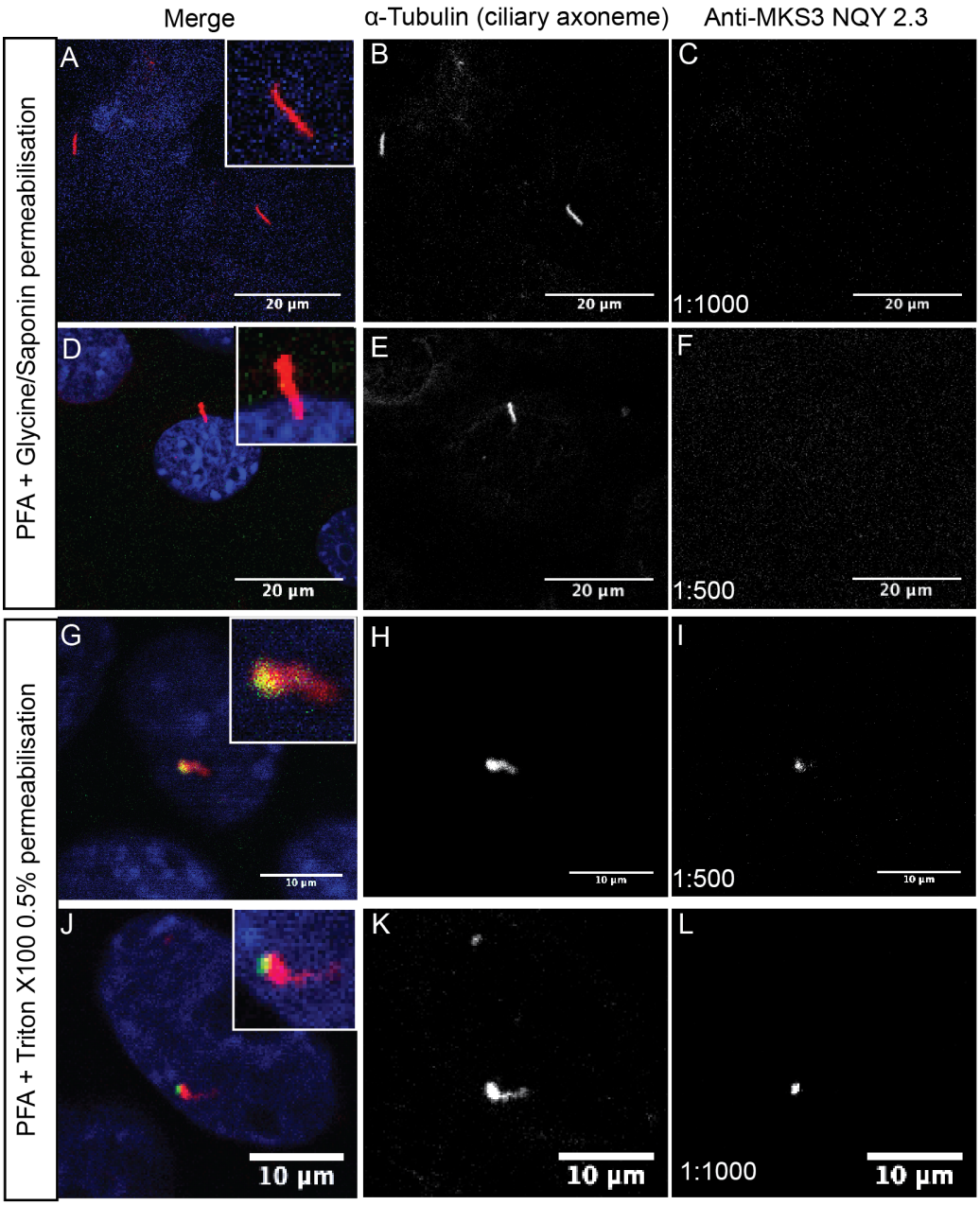


Figure 5.9 Affinity purified anti-MKS3 polyclonal antibody NQY (Fraction 2.3) localises to the ciliary base or axoneme using PFA fixation with 0.5% Triton permebilisation, but not with Glycine-Saponin permeabilisation
IMCD3 kidney epithelial cells were fixed and permeabilised with paraformaldehyde (PFA) and Glycine/Saponin (A-F), or PFA alone and TritonX100 0.5% (G-L). Cells were incbuated with ciliary axoneme marker anti-acetylated alpha tubulin and affinity-purified anti-MKS3 NQY (fraction 2.3) antibodies at antibody dilutions as indicated. Cells were assayed for localisation of anti-MKS3 antibody at the base or axoneme of the cilia.

Figure 5.10: Affinity purified anti-MKS3 polyclonal antibody NQY (Fraction 2.3) does not localise to the ciliary base or axoneme using PFA or PFA-methanol fixation

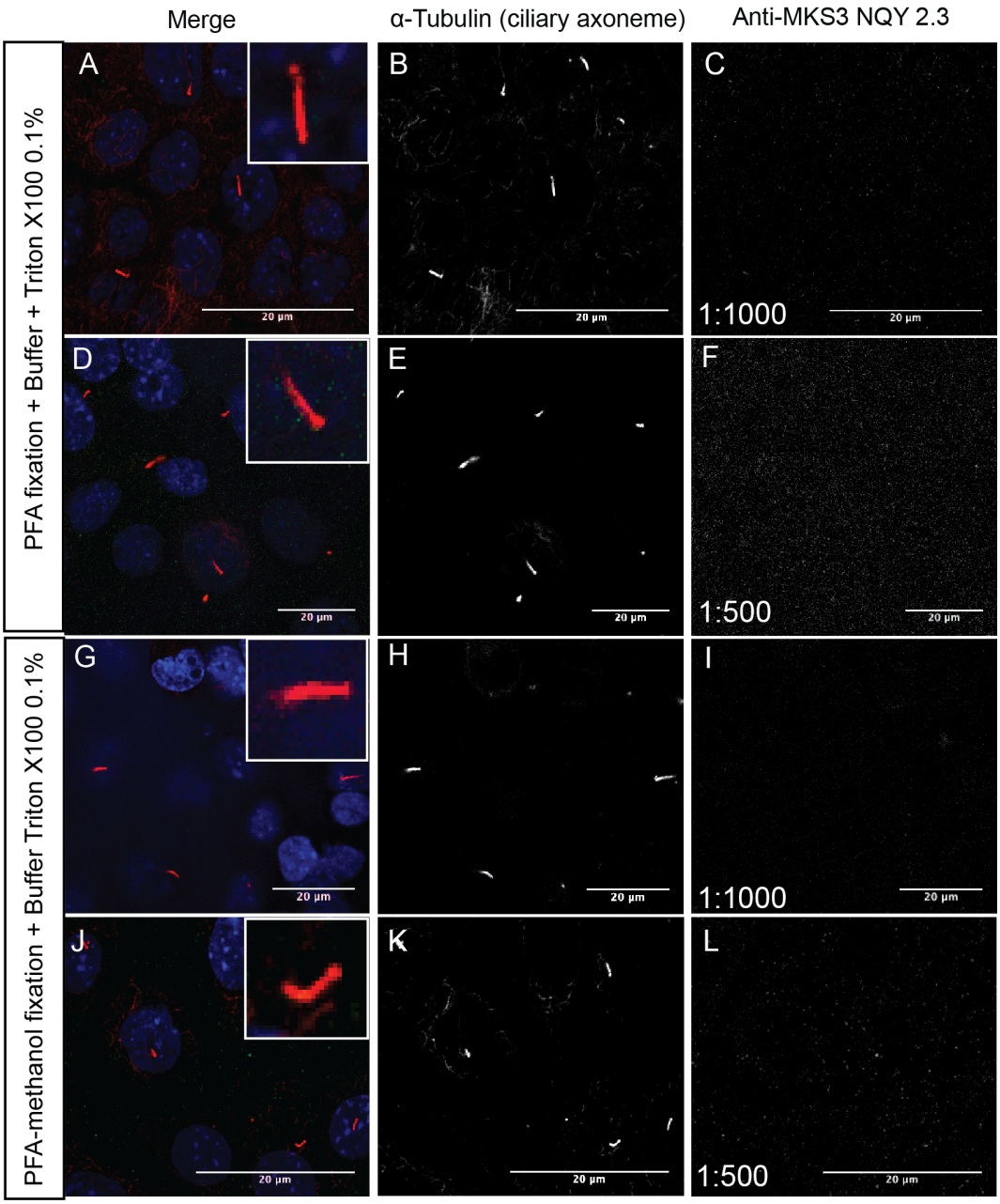


Figure 5.10 Affinity purified anti-MKS3 polyclonal antibody NQY (Fraction 2.3) does not localise to the ciliary base or axoneme using PFA or PFA-methanol fixation
IMCD3 kidney epithelial cells were fixed and permeabilised with paraformaldehyde (PFA) alone and TritonX100 0.1% (A-F) or PFA/methanol and TritonX100 0.1% (G-L). Cells were incbuated with ciliary axoneme marker anti-acetylated alpha tubulin and affinity-purified anti-MKS3 NQY (fraction 2.3) antibodies at antibody dilutions as indicated. Cells were assayed for localisation of anti-MKS3 antibody at the base or axoneme of the cilia.

cilia [Marshall, 2008], and the centrosomes which also localise to the base of the cilia are known to be “sticky" to some rabbit polyclonal antibodies (personal communication, P Mill). Nevertheless, detection by these antibodies of the appropriate band size when assayed by western blot does indicate that they are capable of detecting meckelin. In order to further verify that the antibodies were able to detect meckelin by immunofluorescence, we transfected IMCD3 cells with HA-tagged Meckelin, overexpressing the HA-tagged protein. Cells were fixed using PFA (3.7%) and permeabilised with Triton X100 (0.1%), and blocked in PBS containing TritonX100 and 2% BSA. Transfected cells were then incubated overnight at 4°C with anti-HA antibody raised in mouse (as described previously, abcam) and affinity-purified anti-MKS3 antibody CKT (fraction 1.5) raised in rabbit at a dilution of 1 in 1000. Cells were washed and incubated with secondary antibodies (Alexa fluorophore-conjugated donkey anti-rabbit 488, Alexa fluorophore-conjugated goat anti-mouse 594) before mounting with mounting fluid containing DAPI. Analysis showed that there was co-localisation of the two antibody signals in the peri-nuclear region of these IMCD3 cells (Figure 5.11A), as seen in previous immunofluoresence assays in a different cell types (Chapter 4, Figure 4.9, 4.10). Untransfected controls confirmed that the anti-HA antibody was specific (Figure 5.11B), and secondary antibody only controls also demonstrated specificity of the Alexa secondaries in this cell type (Figure 5.11 C).

Figure 5.11: Affinity purified anti-MKS3 antibody CKT partially co-localises with anti-HA antibody in MCF7 cells transfected with HA-tagged Meckelin

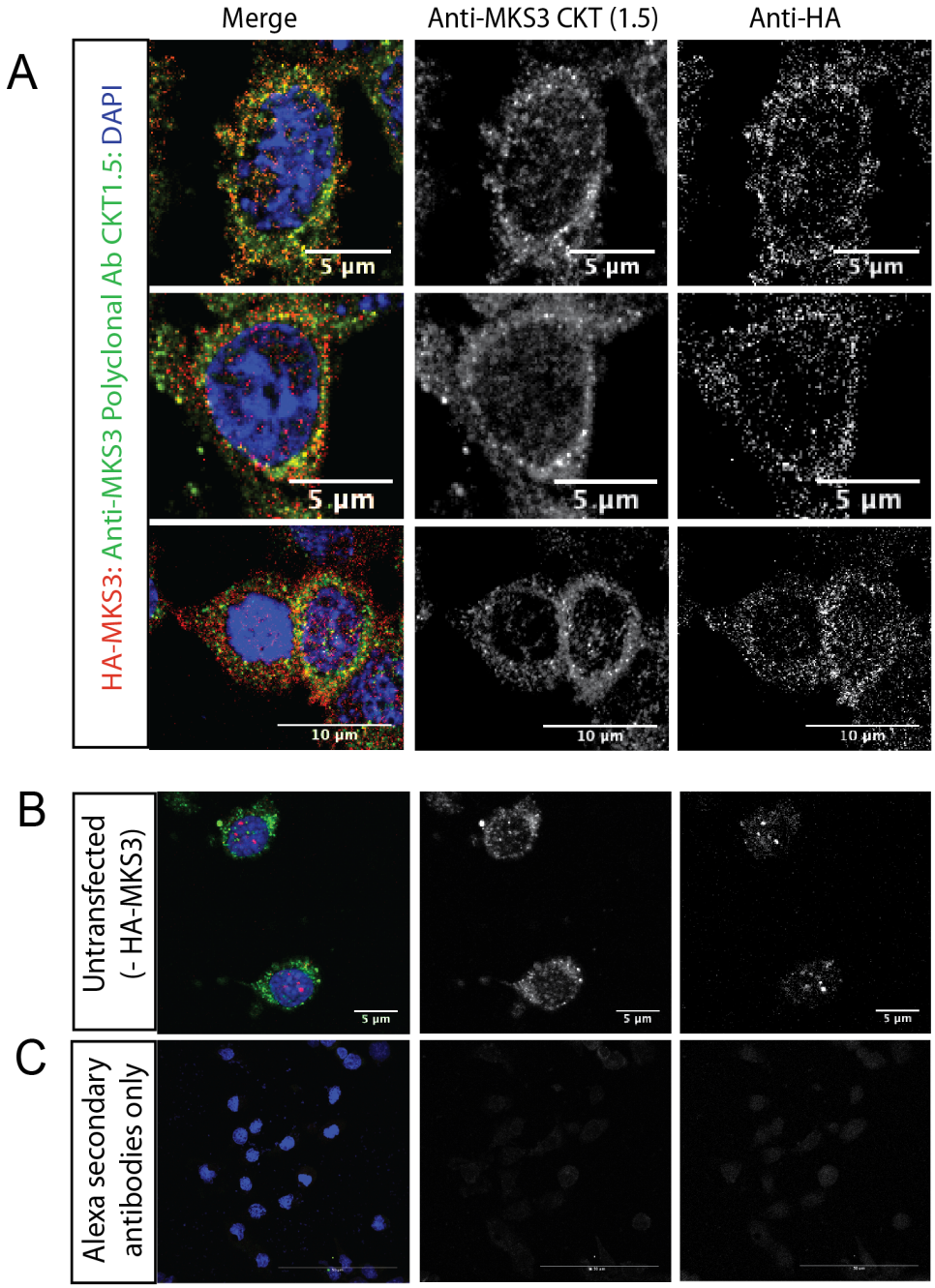


Figure 5.11 Affinity purified anti-MKS3 antibody CKT partially co-localises with anti-HA antibody in MCF7 cells transfected with HA-tagged Meckelin
MCF7 (American) cells were transfected with HA-tagged MKS3 (A, C) or not transfected (B), fixed and analysed by confocal microscopy using mouse monoclonal anti-HA antibody and rabbit polyclonal affinity purified anti-MKS3 antibody (fraction 1.5).

5.6 Discussion

One way of testing the validity of an antibody is to remove the antigen either with gene knockout animal tissue or using siRNA to knock down the gene. In order to verify the antibody for immunofluorescence by loss of signal, the gene must be fully or at least very substantially depleted by siRNA. In the case of *Meckelin*, the gene has been shown to be essential for the formation of cilia in ciliated cell lines such as IMCD3s, making our assay for antibody detection at the cilia body or base redundant when Meckelin is depleted. For this reason, we used transfected HA-tagged Meckelin, overexpressing the HA-tagged protein to verify that the C-terminal antibody, affinity purified fraction 1.5, was detecting Meckelin and not some other basal-body or transition zone-associated protein. As previously discussed, Meckelin knockout animals retain the formation of cilia, although they are abnormal in length; this could potentially serve as a mechanism for further verifying the antibody specificity although the antibodies have not to date been tested for use on tissue using immunohistochemistry.

Differences exist between the two antibodies, and even between affinity-purified fractions of each antibody. For example, a later fraction from the purification column of the antibody to the C-terminal (CKT fraction 2.4) did not detect MKS3 whereas an earlier fraction did. This may be due to differences in the concentration or purity of the fractions collected. Both of these fractions were able to detect MKS3 by western blot, suggesting a possible difference in the affinity of the interaction when binding to fully denatured (western blot) or fixed proteins in immunofluorescence assays.

It is noteworthy that fraction 2.3 of the N-terminal antibody did not detect MKS3 by western blot, although it did detect MKS3 at the base of the cilia using fixation protocol 4 (PFA fixation with 0.5% TritonX100 permeabilisation). The number of antibody signals seen to co-localise with the cilia were much fewer than that seen with the C-terminal antibody, perhaps reflecting a reduced affinity of the interaction. Both antibodies were made using peptides, however, the C-terminal peptide is at the very far end of the protein, whereas the N-terminal peptide is further into the protein, partly to avoid the signal peptide. This might reflect a degree of accessibility of the antigen in the native protein. Furthermore, a naked peptide may not carry the same conformation as a polypeptide within the context of a native protein.

In this chapter, we demonstrated a need for a reliable anti-MKS3 antibody for use in our assays, and described a strategy for producing our own antibodies to the C-terminal and N-terminal of the protein. We then characterised some of the resulting affinity-purified fractions for use in two common assays, detection by western blot and immunofluorescence. We used a known cellular location for MKS3 that is convenient and easy to identify as a means of verifying its usefulness in immunofluorescence assays, and validated its specificity using a transfected, tagged version of the same protein. We have demonstrated that the two antibodies are compatible with differing fixation and permeabilisation protocols, and characterised which conditions are suitable for each, thus giving options for compatibility with other antibodies for co-immunofluorescence assays.

Chapter 6

Agr2 Expression in a Ciliated Cell Line and in cultured Embryonic Kidneys

6.1 Introduction

The usefulness of analysis of a gene's expression pattern was discussed in Chapter 3, where we made an expression vector for ectopic expression of a reporter driven by the *Agr2* promoter. Although we were not to date able to produce detectable reporter gene expression in mice from this Agr2 line, the benefit of gene expression pattern analysis for investigating the biological function of *AGR2* is still valid. Endogenous expression of a reporter from a gene's own promoter *in vivo* is an excellent and reliable way to analyse expression patterns. However, expression analysis using antibodies can also be very useful in determining the expression patterns of a gene. A promoter reporter such as

the one we made for AGR2 will determine which cells are expressing the gene of interest, however, the reporter is not tagged to the protein and remains diffuse in the cytoplasm or nucleus of the cell. Since antibodies bind to proteins, expression analysis using antibodies may provide intracellular information as well as more global expression data.

As discussed in Chapter 4, a novel putative interaction between AGR2 and MKS3 was discovered as a result of a phage-peptide interactomics screen. Biochemical techniques were employed to help validate the potential interaction, however, we wished to investigate this interaction *in vivo*. As mentioned previously, *Meckelin* (MKS3) is required for ciliogenesis and localises to the ER, Golgi network, plasma membrane and cilia in mammalian cells [Dawe et al., 2007b] [Dawe et al., 2009] [Adams et al., 2012]. If AGR2 and MKS3 were interacting *in vivo*, we hypothesised that AGR2 might also be localised to the cilia, a highly specialised organelle found in nearly all mammalian cells [Berbari et al., 2009]. We decided to test whether AGR2 localised to this cellular compartment using both monoclonal and polyclonal antibodies to AGR2. In addition, we decided to identify the expression pattern of AGR2 in the developing embryonic kidney. Disease-causing mutations in Meckelin cause polycystic kidneys, among other phenotypes, brought about by abnormal or absent primary cilia during embryogenesis and postnatally [Chen, 2007]. Given the possible interaction between AGR2 and MKS3 described in Chapter 4, we were interested in the expression pattern of Agr2 in the developing kidney and any overlap with ciliated cells expressing Meckelin. Furthermore, a consortium that provides expression data for the mouse genito-urinary tract based on *in situ* hybridisation techniques, immunohistochemistry, transgenic reporters and microarray data

(GUDMAP), has suggested that Agr2 is expressed in the embryonic kidney in specific compartments [Harding et al., 2011] [Davies et al., 2012].

Apart from being able to be cultured for up to 6 days with relative ease once dissected from the embryo, the mouse embryonic kidney continues to develop the specialised structures found in the mammalian kidney for some time [Srinivas et al., 1999] [Giuliani et al., 2008]. Thus, kidneys dissected at E12.5 will continue to undergo branching morphogenesis of the ureteric bud, and go on to develop nephrons. These nephrons progress through several stages and form comma shaped bodies, S-shaped bodies and early nephrons that contain proximal, medial and distal tubules, as well as glomeruli and associated podocytes. Many antibodies (though by no means all) penetrate with depth into the embryonic kidney once fixed, as do commonly used nuclear stains such as DAPI allowing distinct structures to be imaged [Barak and Boyle, 2011].

6.2 AGR2 localisation at the base of the cilium

Cilia are a specialised organelle that are thought to be potentially found on all mammalian cells. They are tubulin-based projections from the plasma membrane, that grow and recede in a cell-cycle dependent manner. Initiation of ciliogenesis for primary cilia is brought about by rearrangement of the centrioles and docking at the apical cell surface to form a basal body from which the cilium elongates [Dawe et al., 2007a]. Meckelin (MKS3) together with MKS1 were previously shown to be essential for faithful rearrangement of the centrosomes and the initiation of ciliogenesis for primary cilia [Dawe et al.,

2007b]. Primary cilia differ in structure and function from motile cilia [Satir and Christensen, 2008], and unlike motile cilia, only one primary cilia is found per cell under normal conditions. Models to study cilia and the genetics of ciliogenesis include cell lines that readily produce cilia under the appropriate conditions. IMCD3 cells are an SV40-transformed cell line derived from the inner medullary collecting duct (terminal tubule) of a mouse kidney [Rauchman et al., 1993]. IMCD3 cells are ciliated when grown at high confluence, and put out long primary cilia that are easily visualised using antibodies to a number of ciliary proteins. Antibodies against acetylated α-tubulin (marker of stabilised tubulin in the ciliary axoneme) and γ tubulin (ciliary basal body) are commonly used to visualise the cilium by immunofluorescence as seen in Chapter 5. Another commonly used cilia marker is Arl13B, a member of the Ras superfamily of small GTPases that localises to the primary cilia axoneme in diverse organs including kidney, retina and cerebellum [Cantagrel et al., 2008]. In order to avoid possible issues with antibody staining or compatibility, we used IMCD3 cells that had been adapted to stably express fluorescently-tagged Arl13B (generously donated by E Hall, IGMM, Prof I Jackson group). These IMCD3 cells express mKate2-Arl13B, a fluorescent marker in the far red spectrum (excitation/emission maxima 588nm and 633nm respectively). This fluorescent marker is a monomeric protein with high pH stability and photostability, and also exhibits low cytotoxicity [Shcherbo et al., 2007], making it a useful tag for a variety of cell-based assays.

6.2.1 Ciliary localisation of Agr2 in IMCD3 kidney epithelial cells

IMCD3(mKate2-Arl13B) cells were plated at high confluence and incubated overnight at 37°C on coverslips. Cells were fixed and incubated with anti-AGR2 mouse monoclonal antibody (Abnova). Fixed, stained and mounted coverslips were analysed by confocal microscopy. In cells stained with Agr2 monoclonal antibody, cilia projections were readily visualised in the far red spectrum, and Agr2 was seen to localise to the base of the cilium in each cell as well as in punctuated regions along the cilium axoneme (Figure 6.1 A-I). Cells incubated with anti-AGR2 polyclonal antibody also demonstrated localisation of AGR2 at the cilium base, and also along the cilium (data not shown). This localisation is in agreement with known intracellular localisations for MKS3 at the ciliary basal body, transition zone and ciliary axoneme [Dawe et al., 2007b], [Garcia-Gonzalo et al., 2011] [Williams et al., 2011], further lending support to a possible interaction between the two proteins.

Whilst encouraging that Agr2 localised to the same intracellular compartment as MKS3 in cell-based assays, we wished to investigate whether Agr2 localised to the cilia and basal body *in vivo*. As mentioned above, murine embryonic kidneys can be dissected from embryos from embryonic day 11.5-13.5 (E11.5-E13.5) and cultured for up to 6 days, and many antibodies can penetrate into the organ after fixation allowing imaging of specialised structures including cilia. This provided a model system to investigate AGR2 localisation at the cilium *in vivo*.

Figure 6.1: AGR2 localises to the ciliary base and segments of the axoneme in IMCD3 kidney epithelial cells stably expressing mKate-Arl13B

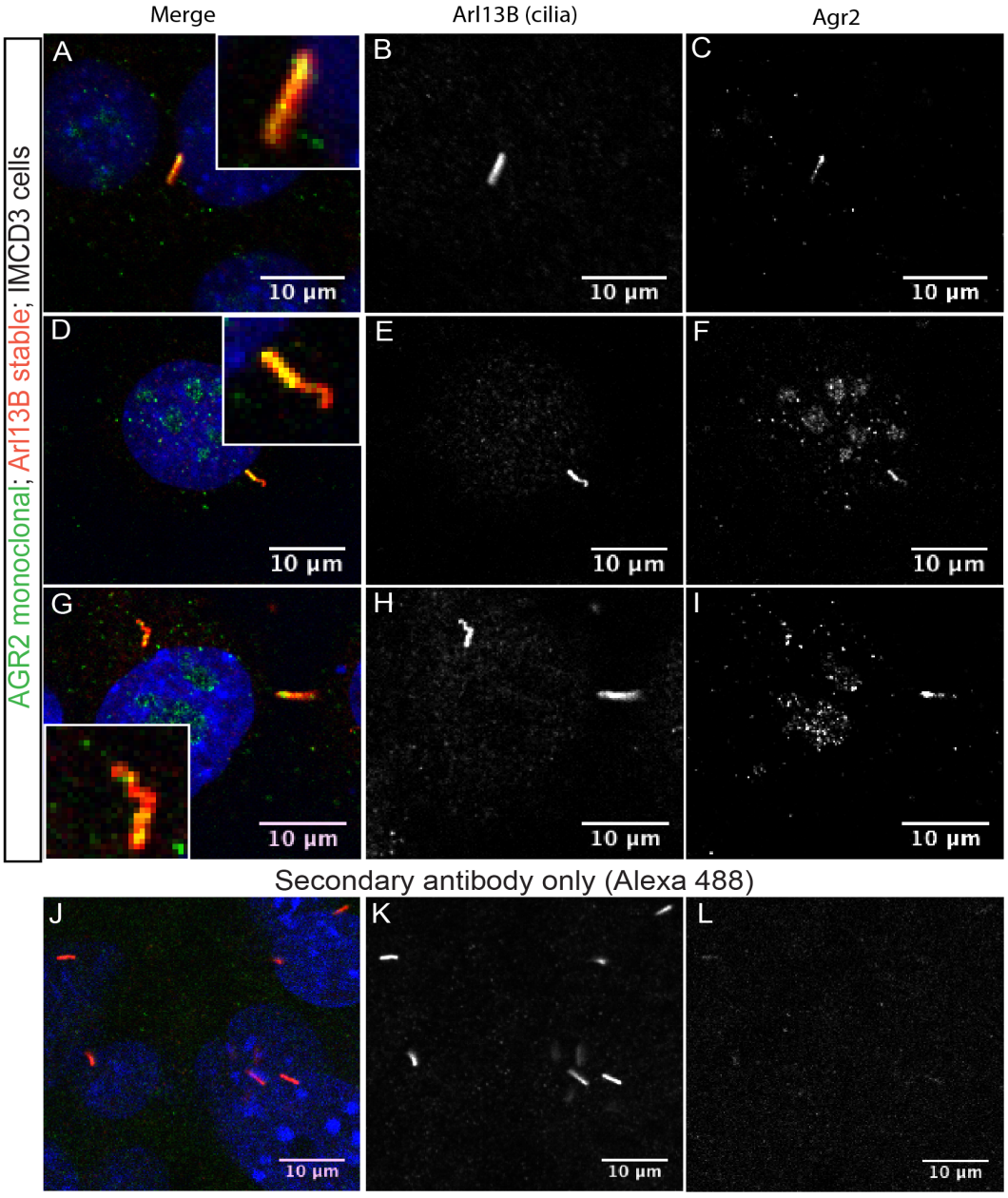


Figure 6.1 AGR2 localises to the ciliary base and segments of the axoneme in IMCD3 kidney epithelial cells stably expressing mKate2-Arl13B
IMCD3 mouse kidney epithelial cells stably expressing mKate2-Arl13B fusion protein were fixed and incubated with anti-AGR2 mouse monoclonal antibody (Abnova) and either secondary antibody (A-I) or no secondary antibody (J-L), and analysed for localisation of Agr2 to the cilia base or axoneme.
mKate2-Arl13B stable IMCD3 cell line was created by E Hall, IGMM, Prof I Jackson lab

6.2.2 Ciliary localisation of Agr2 in cultured mouse embryonic kidney

Embryonic mouse kidneys were dissected from wild type CD1 mouse embryos at embryonic day 12.5 (E12.5) and cultured for 48h on porous membranes according to published protocols [Barak and Boyle, 2011]. Cilia were stained with mouse monoclonal anti-aceylated α tubulin (Sigma, T6793) and Agr2 was visualised with anti-AGR2 rabbit polyclonal antibody (Moravian Biotech). Fixed, stained and mounted kidneys were analysed by confocal microscopy with a 60x oil immersion lens. Cilia were clearly visible on many cells in the developing embryonic kidney on both mesenchymal and epithelial cells, and Agr2 was seen to localise at and around the base of the cilium (Figure 6.2). Clustering of proteins at the cilia base has been reported for proteins known to localise to the axoneme, basal body or transition zone such as IFT proteins [Sedmak and Wolfrum, 2010] [Wang et al., 2009]. This clustering was not seen in the IMCD3 cells described above, however, mKate2-Arl13B stable IMCD3 cells have been seen to project very elongated cilia, indicating that the biology of these cells and cilia may be altered due to the tagged Arl13B protein (personal communication, Dr P Mill, Dr E Hall, Jackson group, MRC Human Genetics Unit). Nevertheless, clear localisation of Agr2 at the base of the cilia both in cell culture and in embryonic renal organ culture supports a role for Agr2 in the biology of cilia.

Figure 6.2: AGR2 protein localises to the base of the cilia in cultured embryonic mouse kidneys

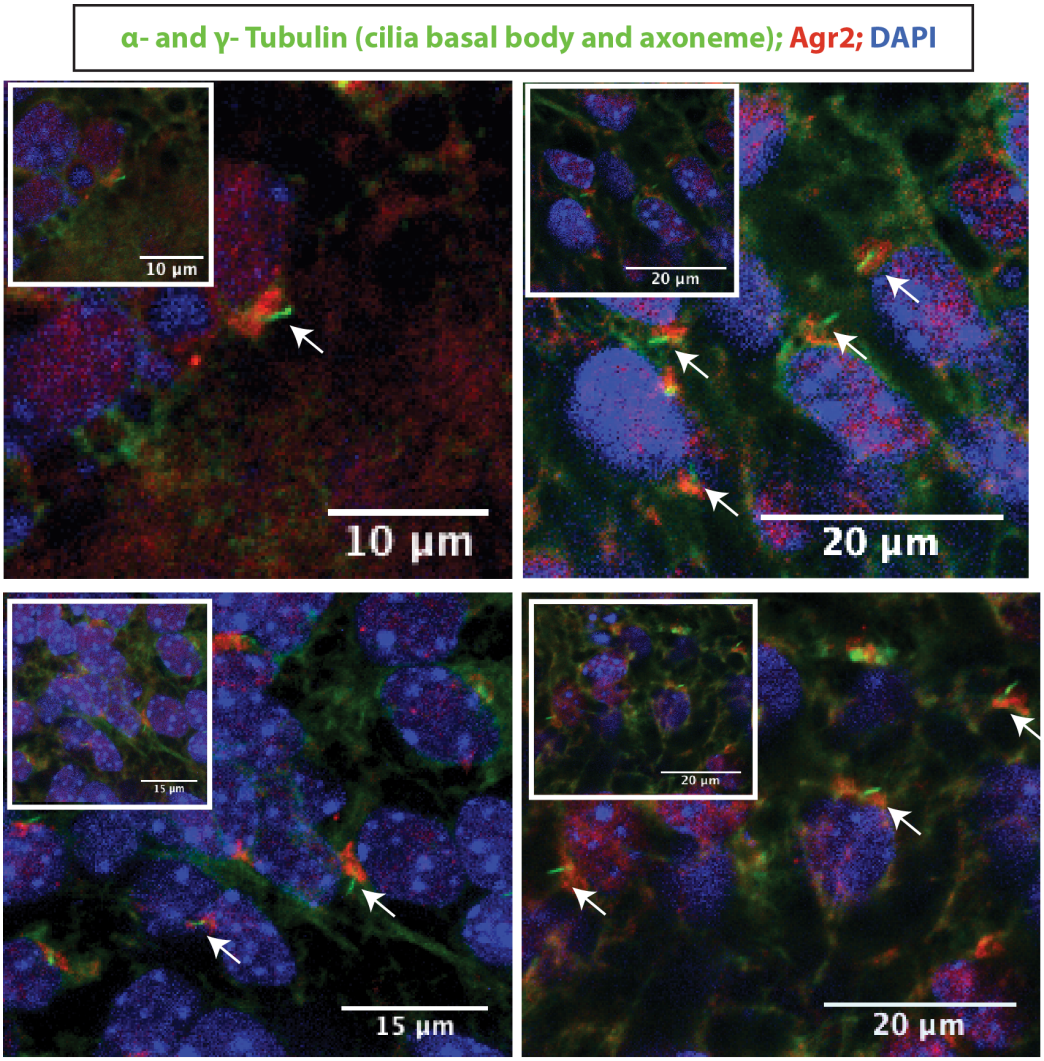


Figure 6.2AGR2 protein localises to the base of the cilia in cultured embryonic mouse kidneys
Embryonic mouse kidneys were dissected from embryos at E12.5 and cultured for 24 hours before fixing and staining with mouse monoclonal anti-acetylated tubulin, mouse monoclonal anti-gamma tubulin and rabbit polyclonal anti-AGR2 antibodies. Fixed, stained and mounted kidneys were analysed by confocal microscopy. Agr2 (green) was analysed for localisation to the cilia (green).

6.3 AGR2 expression patterns in developing embryonic kidney

6.3.1 Embryonic kidney development in mice and associated markers of renal structures *in vivo*

Organogenesis of the embryonic kidney is a dynamic process that progresses in a very specific temporal and spatial manner. In the mouse embryo, cells from the intermediate mesoderm (IM) give rise to the embryonic kidney starting from about embryonic day 9.5 (E9.5) [Saxen and Lehtonen, 1987]. Dorsal intermediate mesoderm cells come together to form the nephric duct (ND), also known as the Wolffian duct, whilst ventral IM mesenchymal cells remain undifferentiated. This population of cells is called the nephrogenic cord (NC) [Costantini and Kopan, 2010] [Saxen and Lehtonen, 1987]. Later, the IM, now divided into nephric duct (ND) and nephrogenic cord (NC), aligns along a rostro-caudal axis of the embryo (along the length of the body), forming an epithelialised tubule in the ND and a group of cells at the caudal end called the metanephric mesenchyme (MM) (Figure 6.3) [Costantini and Kopan, 2010]. At around embryonic day 10.5 (E10.5), the metanephric mesenchyme (MM) signals to the nephric duct epithelia to evaginate, forming the ureteric bud (UB). This then branches repeatedly in response to bi-directional signalling between the ureteric bud tips and metanephric mesenchyme, an example of epithelial branching morphogenesis [Costantini, 2006] [Costantini and Shakya, 2006] [Dressler, 2006]. These branched structures will contribute to the future collecting ducts of the mature kidney.

Figure 6.3: Mammalian Embryonic Kidney Development

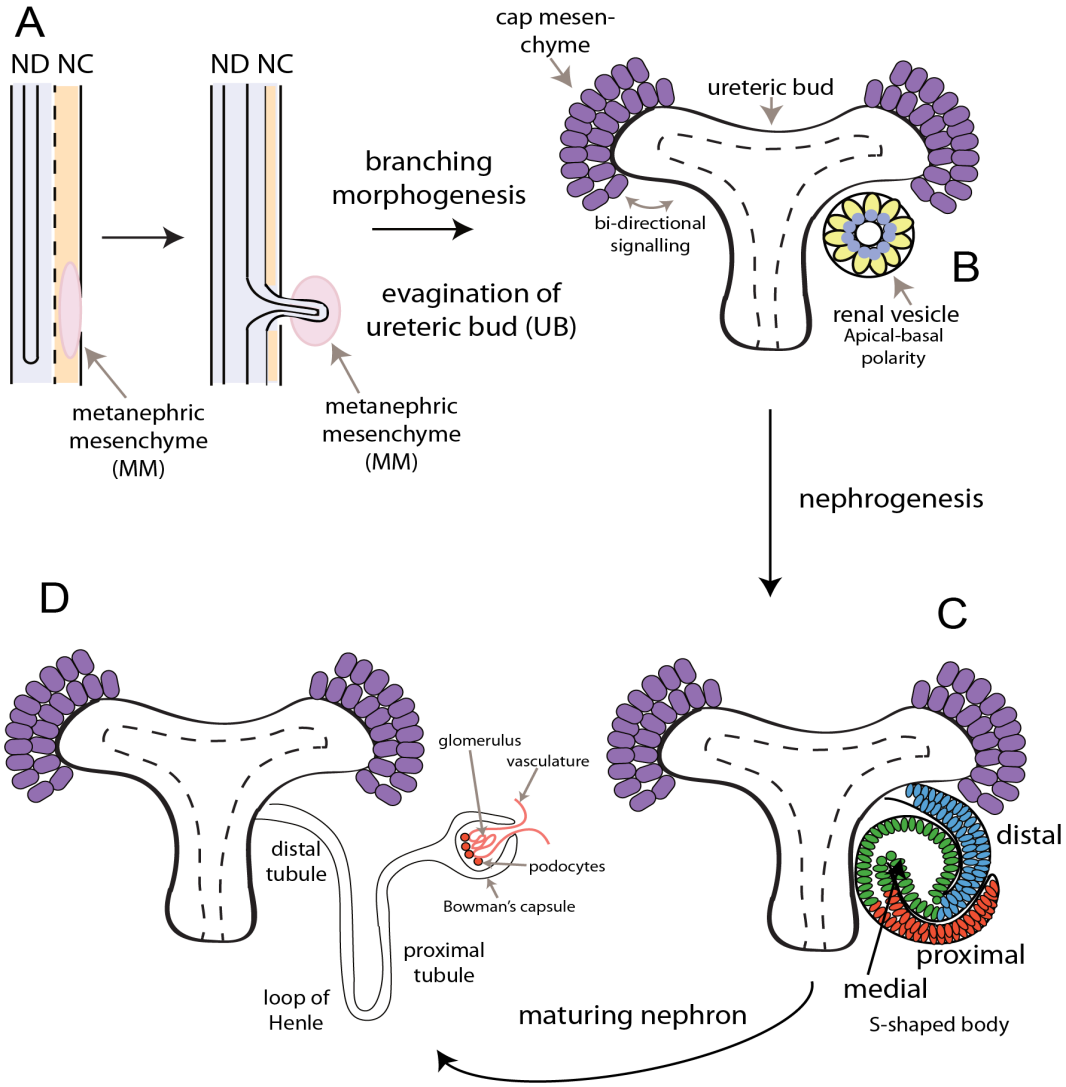


Figure 6.3 Mammalian embryonic kidney development
(A) The intermediate mesoderm develops into the nephric duct (ND) and nephrogenic cord (NC). A primitive tubule forms along the rostro-caudal axis of the ND and cells coalesce to form the metanephric mesenchyme (MM) at the caudal end of the nephrogenic cord. Signals from the MM lead to evagination of the ND to form the ureteric bud, followed by repeated branching morphogenesis of the ureteric bud. (B) Bi-directional signalling between the mesenchymal cells surrounding the tips of the ureteric buds and the ureteric buds themselves leads to epithelialisation of the mesenchyme (MET) and the formation of renal vesicles, the earliest nephrons. (C) The nephrons elongate and form comma-shaped and then S-shaped bodies that are segmented into distal, medial and proximal sections, and the distal ends fuse to the ureteric buds while precursor cells at the proximal end develop into glomeruli. (D) The maturing nephrons form into the conneting segment, distal tubule, loop of Henle, and proximal tubule, and end in the glomerulus comprising the vasculature, mesangial cells, podocytes and Bowman's capsule. *Figure adapted from Costantini and Kopan, 2010; and Dr N Lindstrom*

Meanwhile, cells from the MM form a cap around the ureteric bud tips, leading to the formation of the cap mesenchyme (CM) cells [Mugford et al., 2009]. These cap mesenchyme cells contain the progenitor cells that have the potential to form nephrons [Boyle et al., 2008] [Kobayashi et al., 2008]. Bi-directional signalling between the ureteric buds and cap mesenchyme leads to maintenance of the stem cell population, and signals from the ureteric bud induce cap mesenchymal cells to undergo mesenchymal-to-epithelial transition (MET) and generate nephron epithelia [Saxen and Lehtonen, 1987]. The initial epithelial nephrogenic structure to be formed is the renal vesicle (RV), and these epithelial cyst-like structures appear beneath the ureteric bud tips [Saxen and Lehtonen, 1987]. Cells at the distal end of the renal vesicle (at the end closest to the ureteric bud) fuse with the UB to form a connecting segment [Georgas et al., 2009], whilst the renal vesicle continues to mature, elongating to form a "comma-shaped body", and then an "S-shaped body". These bodies continue to lengthen and become more complex in a poorly-understood process, eventually developing into the embryonic nephron and glomerulus. The precursors of the mature nephron are segmented into proximal, medial and distal tubules, and podocyte precursors are present at the most proximal part of the developing nephron. The glomerulus forms at the proximal end of the nephron, and is the main filtering unit of the mature kidney comprising a capillary network enveloped in podocytes and enclosed in a capsule (Bowman's capsule) [Saxen and Lehtonen, 1987]. The entire process is summarised in Figure 6.3.

There are numerous antibodies that can be used to selectively mark the specialised structures found in the developing kidney. Among these, are antibodies against markers of the CM and early nephrons such as NCAM (Neu-

ral cell adhesion molecule), antibodies against markers of podocytes such as podocalyxin, and antibodies against the Notch ligand, Jagged1, which selectively marks the proximal/medial tubule of the developing nephron. Conversely, antibodies against E-cadherin (CDH1) can be used as a specific marker of the distal cell population in developing nephrons as well as other epithelial structures such as the ureteric bud. The antibodies used are summarised in Materials and Methods Table 2.16. Using these antibodies in combination with an anti-AGR2 antibody, expression patterns of AGR2 in the developing embryonic mouse kidney were analysed.

6.3.2 AGR2 expression in embryonic mouse kidney

We wanted to investigate the expression of Agr2 in different segments of the embryonic mouse kidneys, using a previously-validated polyclonal antibody to AGR2. As mentioned above, MKS3 is a cilia-associated protein that is essential for ciliogenesis, and localises to the primary cilium in mammalian cells. Primary cilia are found in many structures of the developing mammalian kidney, including at the apical side of some renal epithelial cells [Mirna Saraga-Babic et al., 2012], and in the case of the ureteric bud this corresponds to the lumen. Embryonic kidneys were dissected from embryos at E12.5 and cultured for 24 hours according to established protocols as described above. Kidneys were fixed as described in section 2.7 and incubated with anti-Agr2 rabbit polyclonal antibody (Moravian Biotech) overnight also as described in section 2.7. Fixed, stained and mounted kidneys were analysed using confocal microscopy.

In the early (E12.5 plus one day) cultured embryonic kidney, Agr2 lo-

calised to the apical side of the ureteric bud, in agreement with reported localisation of primary cilia in mammals [Trapp et al., 2008]. Agr2 was also seen to localise to the cap mesenchyme (CM) in a polarised manner in these early developing kidneys. This is in agreement with GUDMAP data showing expression of Agr2 in the CM. In addition, the staining pattern of Agr2 in the CM was also apical, with the apical side facing away from the ureteric bud basolateral membrane (Figure 6.4 A,B). The metanephric mesenchyme has previously been shown to have primary cilia, as seen in Figure 6.2 and in human embryonic kidneys [Mirna Saraga-Babic et al., 2012].

In order to confirm that the localisation of AGR2 in the ureteric bud was apical, we used a marker of apical-basal polarity that is only present at the apical side of epithelial cells. ZO-1 (Zona Occludens-1) protein binds occludins directly and is a marker of tight junctions and apical (luminal) polarity [Fanning et al., 1998]. Kidneys were harvested at E12.5 and cultured for 24h before fixing, washing and staining. Kidneys were incubated overnight with anti-Agr2 rabbit polyclonal antibody (Moravian Biotech) and anti-ZO-1 rat polyclonal antibody (DSHB). Kidneys were analysed by confocal microscopy. ZO-1 expression, used to mark the apical side of the ureteric bud, indicated that polarised Agr2 staining was detected at the luminal, apical side of the ureteric buds as shown in Figure 6.5.

Having observed that Agr2 was localised to the CM and UB in early stage developing embryonic kidneys, we wished to assess the expression pattern of Agr2 in nephron development. NCAM is a marker of CM and early nephrons [Klein et al., 1988] but not of the ureteric bud. Kidneys were dissected at E12.5 and cultured as described above.

Figure 6.4: Agr2 protein localises to the apical side of the ureteric bud and to focal points in the mesenchyme of cultured E12.5 developing mouse kidney

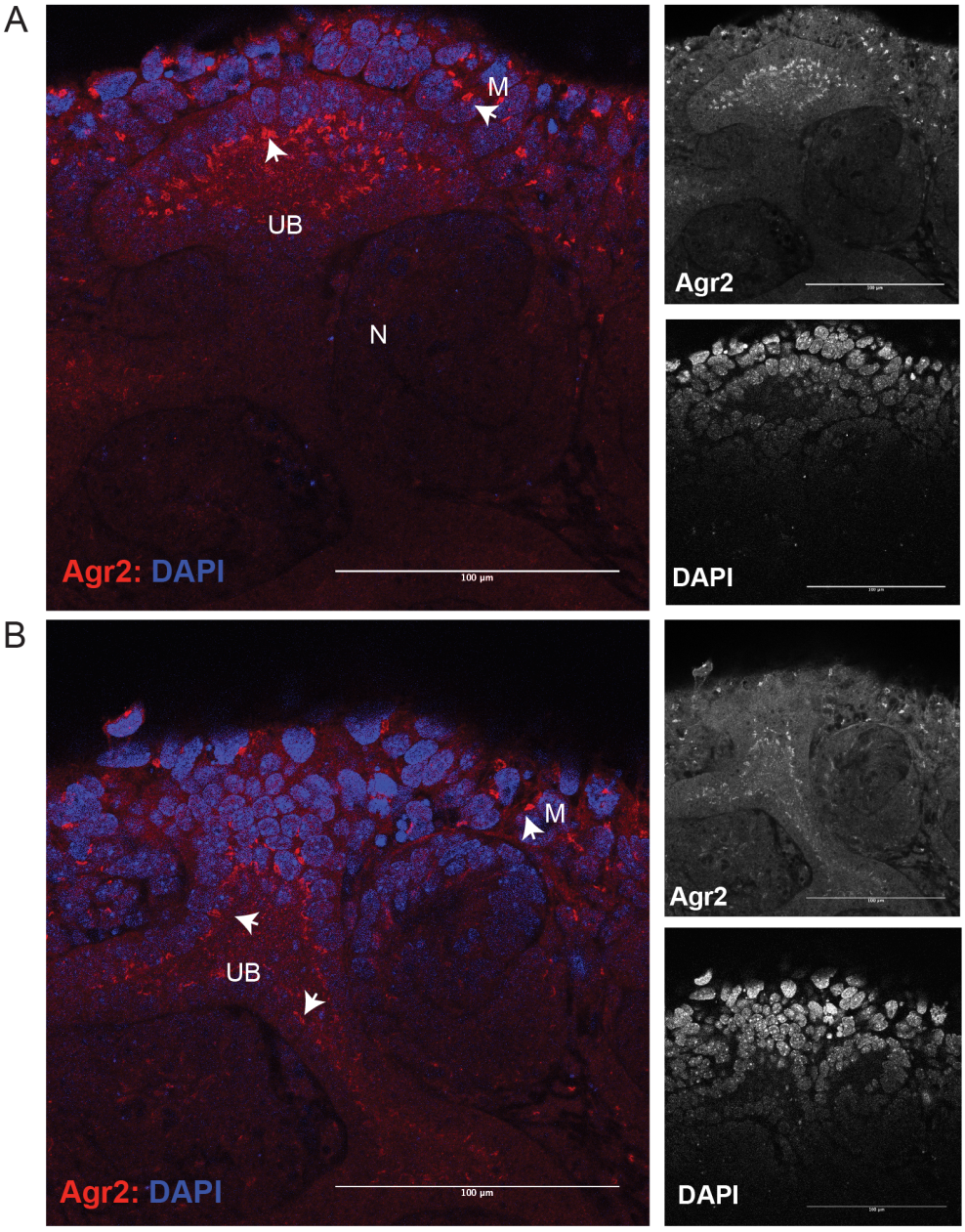


Figure 6.4 AGR2 protein localises to the apical side of the ureteric bud and to focal points in the mesenchyme of cultured E12.5 developing mouse kidney
Embryonic mouse kidneys were dissected from embryos at E12.5 and cultured for 24 hours before fixing and staining with rabbit polyclonal anti-AGR2 antibody. Fixed, stained and mounted kidneys were analysed by confocal microscopy (A, B). A, B show confocal Z planes from the same sample.

Figure 6.5: AGR2 localises to the apical side of the ureteric bud with apical marker ZO-1

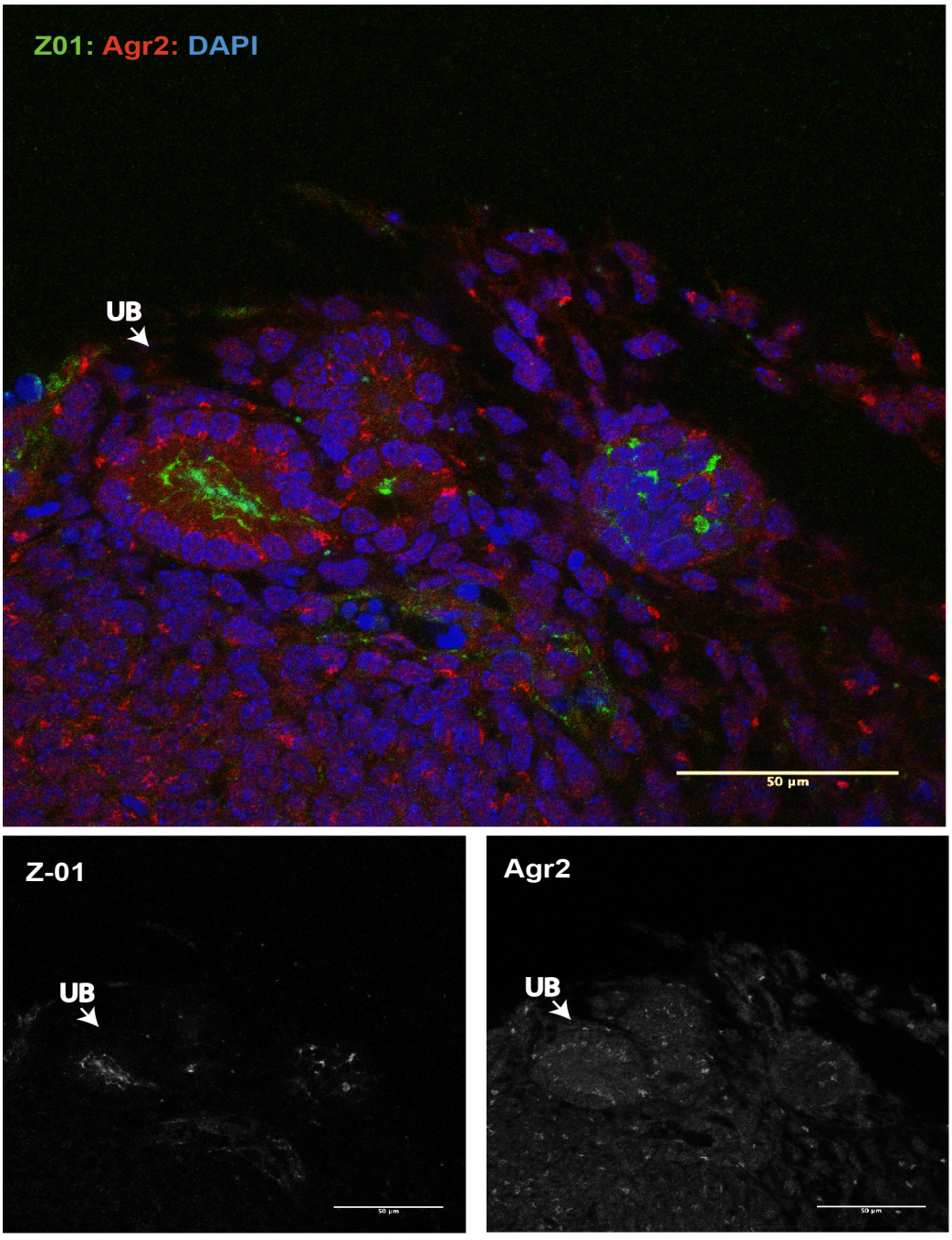


Figure 6.5 AGR2 localises to the apical side of the ureteric bud with apical marker ZO-1 Embryonic mouse kidneys were dissected from embryos at E12.5 and cultured for 24 hours before fixing and staining with rat polyclonal anti-ZO-1 and rabbit polyclonal anti-AGR2 antibodies. Fixed, stained and mounted kidneys were analysed for localisation of Agr2 (red) and ZO-1 (green) by confocal microscopy.

Kidneys were fixed and incubated with anti-Agr2 rabbit polyclonal antibody (Moravian Biotech) and anti-NCAM mouse monoclonal antibody (Sigma), and fixed, stained and mounted embryonic kidneys were analysed as described above. Agr2 was co-expressed with NCAM in early stage nephrons and renal vesicles (Figure 6.6). We also wished to assess the expression pattern of Agr2 in later stages of nephron development. After the renal vesicle stage, in which the nascent nephron is already divided into proximal and distal regions, the developing nephron in mammalian kidneys becomes segmented into the proximal, medial and distal tubules, with the distal-most segment connecting to the ureteric bud. One of the components that helps to create proximal-distal identity is the family of Notch signalling proteins [Costantini and Kopan, 2010].

The Notch signalling pathway is highly conserved and has been shown to control differentiation, growth, and patterning processes [Benedito et al., 2009]. Although studies have revealed that Notch2 is most important for this process, both Notch1 and Notch2 are involved in patterning proximal fate in the nephron [Cheng et al., 2007]. Jagged1 (JAG1) is a ligand for the Notch1 receptor, and as such has been used effectively as a marker of medial/proximal tubule in nephrons. Conversely, E-cadherin (CDH1) is a calcium-dependent cell adhesion protein that has been shown to be expressed mainly in the distal tubules of developing kidneys, and only at very low levels in the proximal tubule [Prozialeck et al., 2004]. Initially, kidneys were dissected at E12.5 and cultured for four to five days before fixing and staining as described above.

To begin with, kidneys were incubated with anti-AGR2 rabbit polyclonal antibody (Moravian Biotech) and mouse monoclonal anti-acetylated α-tubulin (Sigma, T6793). As well as labelling the ciliary axoneme, anti-acetylated α-

Figure 6.6: Agr2 is co-expressed with a marker of CM, renal vesicles and early nephrons, NCAM, in cultured E12.5 developing mouse kidney

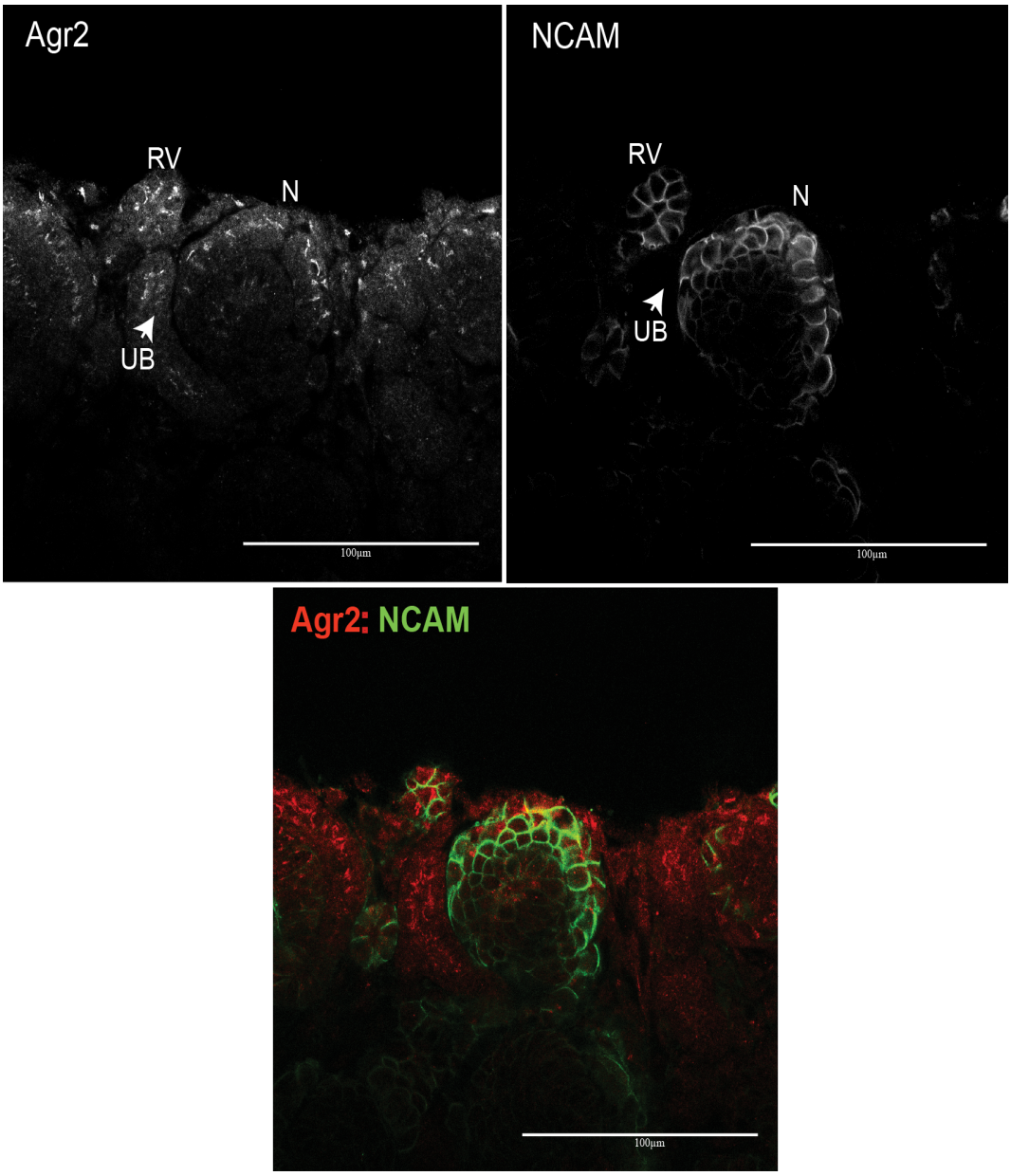


Figure 6.6 Agr2 is co-expressed with a marker of cap mesenchyme, renal vesicles and early nephrons, NCAM, in cultured E12.5 developing mouse kidney
Embryonic mouse kidneys were dissected from embryos at E12.5 and cultured for 24 hours before fixing and staining with mouse monoclonal anti-NCAM and rabbit polyclonal anti-AGR2 antibodies. Fixed, stained and mounted kidneys were analysed by confocal microscopy. Agr2 (red) was analysed for co-expression with a marker of cap mesenchyme, renal vesicles and early nephrons.

tubulin also labels acetylated tubulin filaments within the cell, helping to visualise various specialised structures (Figure 6.7). After incubation with primary antibodies, kidneys were incubated with secondary antibodies and mounted using DAPI-containing mounting fluid. Kidneys dissected at E12.5 and cultured for several days continue to develop, and at this stage segmented nephrons and glomeruli with podocytes can be observed. Agr2 staining was again clearly seen in the ureteric bud and metenephric mesenchyme as previously described. In addition, Agr2 was expressed in the proximal tubule in these kidneys, but was absent in the centre of the glomeruli (Figure 6.7). In order to further probe the segement-specificity of Agr2 expression in the nephron, we also stained E12.5 kidneys that had been cultured for 4 days with markers of proximal and distal tubules. Kidneys were incubated with anti-AGR2 rabbit polyclonal antibody, anti-JAG1(Jagged1) goat polyclonal antibody and anti-CDH1(E-cadherin) monoclonal antibody. Fixed, stained and mounted kidneys were analysed by confocal microscopy. Figure 6.8 shows images from separate stacks through a single nephron, with distal tubule (DT), medial tubule (MT) and proximal tubules (PT) labelled. Agr2 is seen to be expressed in the proximal and medial tubule, with much lower expression in the distal tubule, and expression of Agr2 increases the more proximal the tubule becomes. The proximal tubule is also the site of primary cilia in the embryonic kidney in humans [Mirna Saraga-Babic et al., 2012], and MKS3 has also been shown to be expressed in the proximal renal tubules of early human embryos [Dawe et al., 2007b]. In addition, Agr2 staining in the proximal section of the nephron appeared to stain the visceral and parietal epithelia, although it is not possible to state this definitively without appropriate markers.

Figure 6.7: AGR2 localises to the apical side of the ureteric bud and proximal tubule of the developing nephron but not the centre of the glomerulus in cultured embryonic kidneys

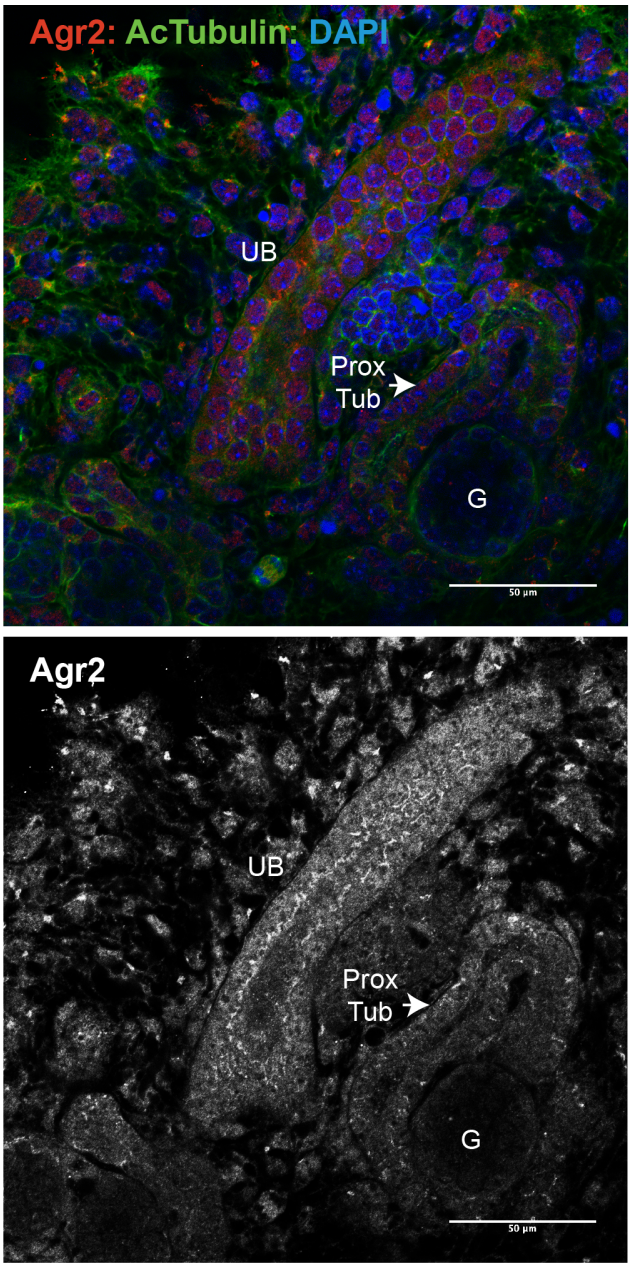


Figure 6.7 AGR2 localises to the apical side of the ureteric bud and proximal tubule of the developing nephron but not the centre of the glomerulus in cultured embryonic kidneys Embryonic mouse kidneys were dissected from embryos at E12.5 and cultured for 72 hours before fixing and staining with mouse monoclonal anti-acetylated alpha tubulin and rabbit polyclonal anti-AGR2 antibodies. Fixed, stained and mounted kidneys were analysed by confocal microscopy for localisation of Agr2 in maturing embryonic kidneys (red).

Figure 6.8: AGR2 is expressed in a gradient in the proximal and medial tubule of the developing mouse embryonic kidney

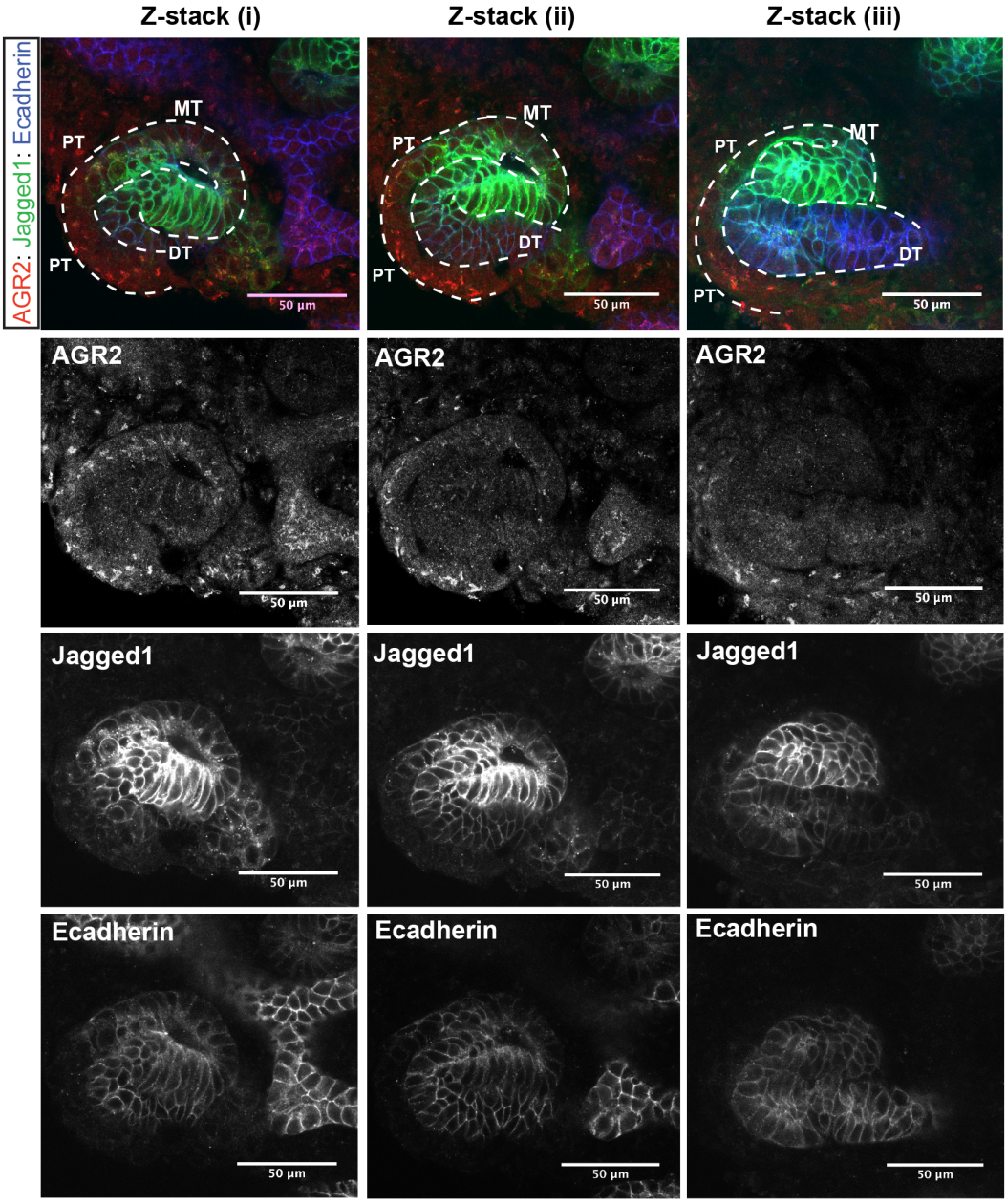


Figure 6.8 AGR2 is expressed in a gradient in the proximal and medial tubule of the developing mouse embryonic kidney
Embryonic mouse kidneys were dissected from embryos at E12.5 and cultured for 96 hours before fixing and staining with goat polyclonal anti-JAG1, mouse monoclonal anti-CDH1 and rabbit polyclonal anti-AGR2 antibodies. Fixed, stained and mounted kidneys were analysed by confocal microscopy for localisation of Agr2 in maturing embryonic kidneys (red). Distal structures and ureteric bud are marked with anti-E-cadherin (blue) and medial tubule is marked with anti-Jagged1 (green).

This would be in agreement with expression data from GUDMAP which also reports expression of Agr2 in the visceral epithelia of the embryonic kidney.

The observation that Agr2 was expressed in segments of the nephron but appeared to stain glomeruli to a lesser degree (Figure 6.7), together with GUDMAP expression data showing high levels of AGR2 in podocytes, led us to investigate the expression pattern of Agr2 in the glomeruli using a podocyte-specific marker. We dissected kidneys from E12.5 embryos and cultured them for 4 days. Kidneys were fixed and incubated with anti-AGR2 rabbit polyclonal antibody and anti-podocalyxin goat polyclonal antibody. Podocytes are polarised epithelial cells with a specialised structure that wrap around the glomerular capillaries. Adjacent podocytes interdigitate like entwined fingers creating filtration slits that are bridged by specialized cell junctions known as slit diaphragms [Dryer and Reiser, 2010]. They are rich in cell surface proteins that restrict the uptake of proteins in the bloodstream whilst allowing smaller molecules and ions to pass [Rodewald and Karnovsky, 1974]. One such abundant cell-surface protein found in podocytes is podocalyxin, making it a useful podocyte marker. Fixed, stained and mounted kidneys were analysed as described above. Agr2 was seen to be expressed in the ureteric buds (UB) as previously observed, and also in the surface of the podocytes (P) (Figure 6.9). By contrast, Agr2 expression appeared absent inside the glomeruli, containing the vasculature (G), in agreement with previous observations (Figure 6.7). Again, this supports the possibility that AGR2 and MKS3 may be interacting and localised to the cilia, since podocytes, but not the rest of the glomeruli, have been shown to have primary cilia in human embryos [Mirna Saraga-Babic et al., 2012]. The specific expression of Agr2 in podocytes is noteworthy, since

Figure 6.9: Agr2 localises to the podocytes and parietal epithelium but not to the centre of the glomerulus in cultured developing mouse embryonic kidneys

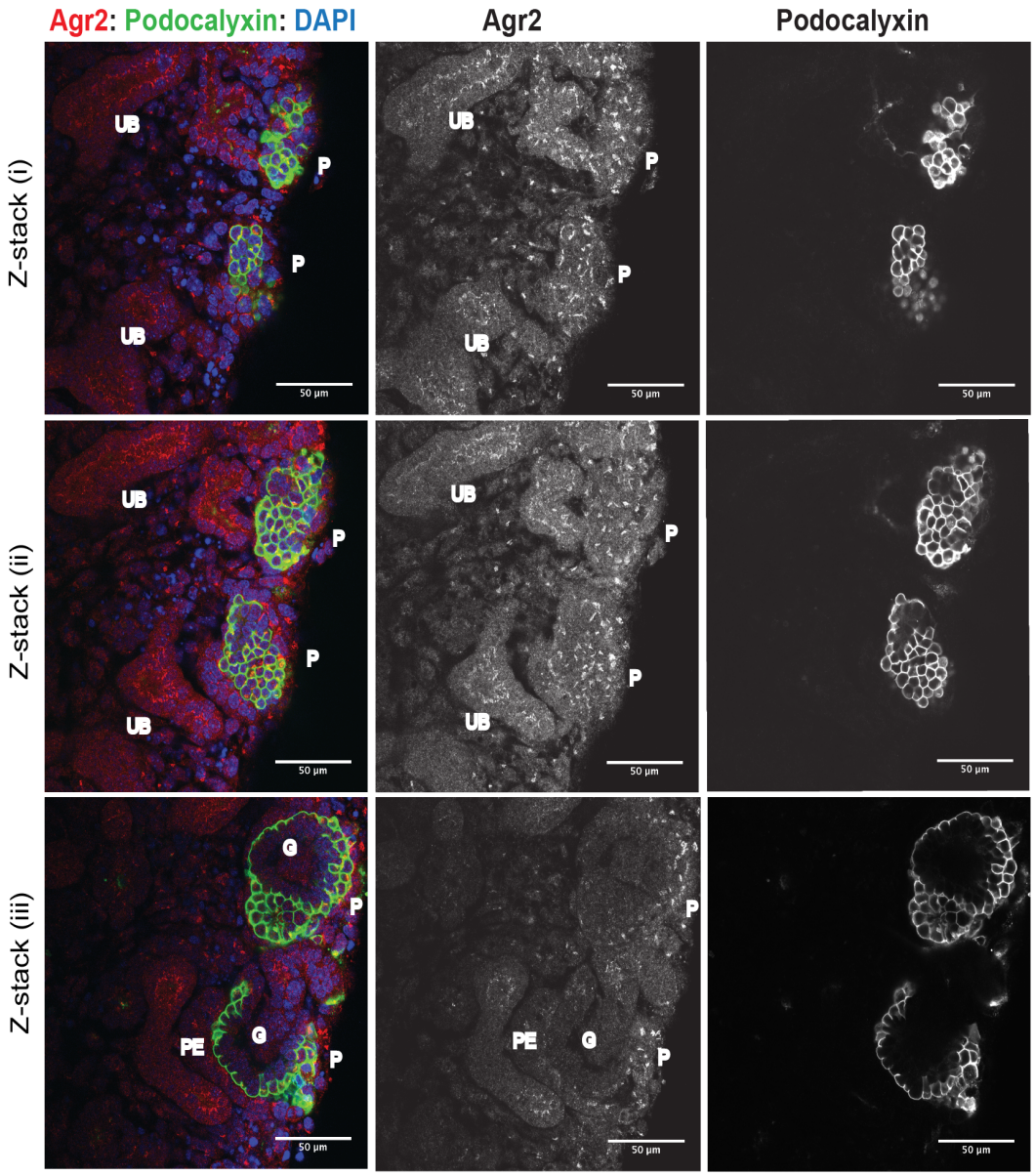


Figure 6.9 Agr2 localises to the podocytes and parietal epithelium but not to the centre of the glomerulus in cultured developing mouse embryonic kidneys
Embryonic mouse kidneys were dissected from embryos at E12.5 and cultured for 96 hours before fixing and staining with goat polyclonal anti-podocalyxin and rabbit polyclonal anti-AGR2 antibodies. Fixed, stained and mounted kidneys were analysed by confocal microscopy for localisation of Agr2 (red) in maturing embryonic kidneys and podocytes (green).

podocytes are rich in transmembrane proteins such as solute channels, membrane proteins [Greka and Mundel, 2012] and some active protein-transport channels [Akilesh et al., 2008]. If Agr2 is indeed, as the phage-display interactomics screen would suggest, a chaperone for a family of large, modified, transmembrane proteins, then this would agree with expression of Agr2 in the podocytes of the embryonic kidney.

6.4 Discussion

Analysis of Agr2 expression first in a murine ciliated cell line and then in cultured embryonic mouse kidneys have revealed a localisation at the primary cilium and in and around the base of the cilium. Further expression analysis using a panel of specific markers to specialised renal structures have revealed a pattern of expression that is in agreement with reported expression data from Gudmap, and with published locations of primary cilia and MKS3 in the kidney. Specifically, Agr2 is expressed in the proximal tubule, ureteric bud, podocytes and metanephric mesenchyme in murine embryonic kidneys, but not in the glomeruli or distal tubules (Figure 6.10). This is also in agreement with tissue expression of the linear-peptide containing gene set, in which the expression of many of the genes is enriched in the proximal tubule, ureteric bud, podocytes and cap mesenchyme.

Primary cilia and MKS3 are known to be involved in epithelial branching morphogenesis [McDermott et al., 2010] [Dawe et al., 2007b]. In developing kidneys, signals between the metanephric mesenchyme and ureteric bud lead to repeated branching of the ureteric buds [Davies and Bard, 1998], and it

Figure 6.10: Model of Agr2 expression in segments of the developing embryonic kidney

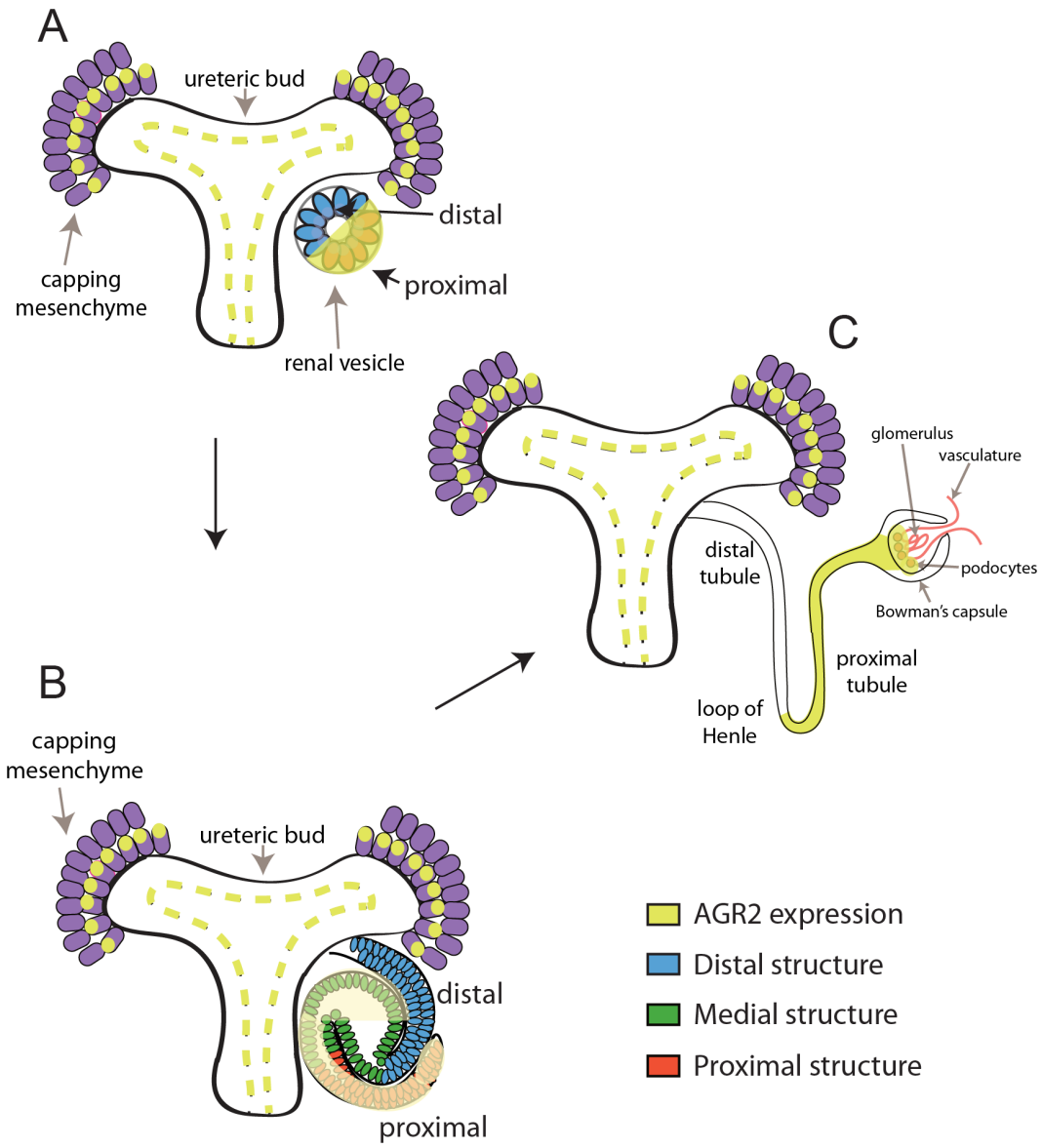


Figure 6.10 Model of Agr2 expression in segments of the developing embryonic kidney (A) Model of Agr2 expression in apical part of the cap mesenchyme, apical side of ureteric bud and renal vesicles in early embryonic kidneys. (B) Model of Agr2 expression in cap mesenchyme, ureteric bud and in a gradient in proximal tubule of maturing embryonic kidneys. (C) Model of Agr2 expression in capping mesenchyme, ureteric bud, proximal tubule and podocytes of later stage developing embryonic kidneys.

is of note that both these structures have primary cilia and thus presumably MKS3, and that both cell types also express Agr2. In addition, the observation that Agr2 is expressed in a polarised manner in the mesenchyme, and particularly in the cap mesenchyme, of early embryonic kidneys is of interest. Recent studies point to the cap mesenchyme as the likely provenance of Wilms tumours [Pode-Shakked et al., 2013] [Metsuyanim et al., 2009], and a study pointed to the region encompassing the Agr2 gene as the site of loss-of-heterozygosity linked to Wilms tumours [Sossey-Alaoui et al., 2003]. Polarisation of the metanephric mesenchyme in the developing kidney has not been reported previously, and polarity is generally thought to be a marker of epithelial cells [Bryant and Mostov, 2008]. Early polarisation of cells in response to epithelialisation signals may include asymmetrical rearrangement of cytoskeletal and membrane-trafficking systems [Goldstein and Macara, 2007] [Bryant and Mostov, 2008]. If so, the metanephric mesenchyme at E12.5 in mouse embryonic kidneys may be displaying early signs of polarisation, perhaps the first step in epithelialisation of the CM as it is induced to form nephron precursors. At first glance, the most recent murine Agr2 knockout mouse does not appear to support an essential role for Agr2 in early polarisation of the mesenchyme and subsequent epithelialisation. However, mice with reduced MET could still potentially be viable. As a putative chaperone protein involved in trafficking and folding of membrane and secretory proteins, Agr2 might play a role, even if some redundancy exists, in the early stage of mesenchymal to epithelial transition, as the cytoskeleton and membrane-trafficking systems redistribute in a polarised manner. Thus, Agr2 may represent an early marker of polarisation in mesenchymal cells.

Chapter 7

Discussion

The importance of AGR2 as a factor in disease onset and progression has become clearer during the last few years. In the last year alone some twenty published articles have highlighted the role of AGR2 in both hormonal and non-hormonal cancers, as well as in the pathology of inflammatory bowel disease and asthma. AGR2 was also shown in studies to be associated with poor prognosis and drug resistance in hormone-related cancers, and was more recently shown to be responsive to ERα (estrogen-receptor alpha) and also induced by tamoxifen [Hrstka et al., 2010]. In addition, several studies have begun to investigate the biology of AGR2, with the most recent mouse gene knock-out implicating AGR2 in terminal differentiation of gastric cell types [Gupta et al., 2012b], although no functional mechanism has yet been established. We wanted to investigate the biology of AGR2 using a combination of interactomics to identify functionally-relevant binding partners, and mouse genetics to determine the expression pattern of AGR2 both in development and in adult mice. In addition, we wanted to create an AGR2-promoter driven Cre

recombinase mouse for use in functional genetics mouse models.

7.1 Linear Domain Interactome of AGR2

AGR2 can form a homodimer, and dimerisation of AGR2 has been shown to be involved in mediating the ER stress response [Ryu et al., 2012]. Using phage-peptide display proteomics, we identified a core peptide that bound AGR2 with high specificity (Figure 4.1) [Murray et al., 2007]. Recent work has identified the structure of the dimer and the dimer interface [Patel et al., 2013], and has identified a region that is thought to stabilise the homodimer (Figure 7.1). In addition, unpublished work from collaborators has identified the likely peptide binding site, corresponding to amino acids 131-135 (Figure 7.1). This region is highly conserved (see Figure 1.6), indicating the likely importance of binding to this interface for protein function. It is thought that binding to the peptide interface on AGR2 stabilises and regulates the formation of the homodimer drawing a possible link between binding to this site and the ER stress response (unpublished, personal communication, T Hupp). Further work identified the linear motif and allowable substitutions that continued to bind AGR2 [Fourtouna et al., 2009], and using this motif we searched the human genome for potential AGR2 binding partners. Many potential interactors were identified, and we analysed the characteristics of this set of genes using a web-based informatics approach. Several classes of proteins with specialised functions or cellular compartments were identified, and more, a pattern in the hits was identified that pointed toward a biological role for AGR2 as a chaperone for a subset of client proteins in the secretory pathway.

Figure 7.1: AGR2 Dimer Structure with Linear Peptide Binding Region

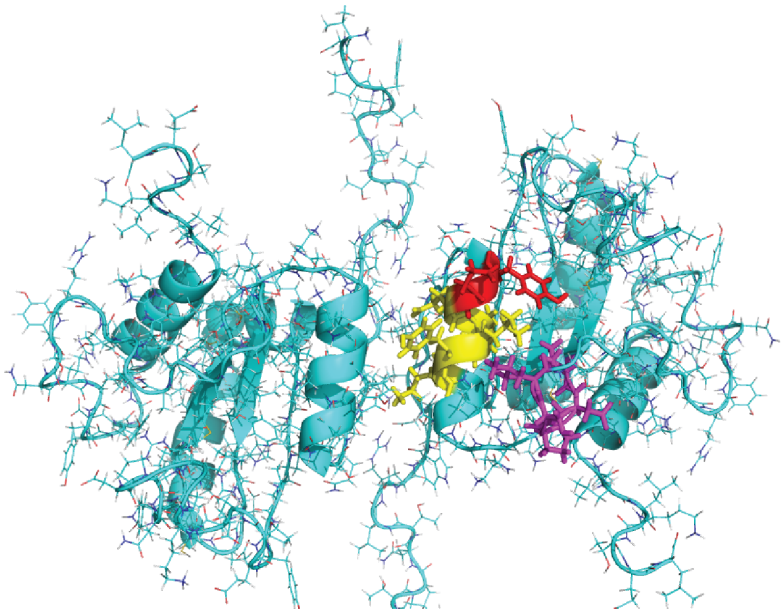


Figure 7.1 AGR2 dimer structure with linear peptide binding region
The AGR2 homodimer binding interface (yellow) with dimer stabilising region (red) and linear peptide binding region (magenta). It is thought that docking to the linear peptide binding site promotes stability of the dimer. *Reproduced from T Gray, Hupp lab, IGMM, using PDB structure 2LNS [Patel et al, 2012].*

Since beginning this project, several studies have also identified a role for AGR2 as a molecular chaperone [Park et al., 2009] [Ryu et al., 2012] [Higa et al., 2011], and also as a secreted protein that increases the levels of another secreted protein, Cathepsin D [Dumartin et al., 2011]. These and other studies, together with the data from the database mining, strongly suggest a biological role for AGR2 in the secretory pathway (Figure 7.2). Many of the genes identified by the phage-display / database mining analysis have functions associated with cell growth, metastasis, and ER-stress responses. As mentioned in the Introduction, however, there are likely to be other cellular localisations and thus functions for AGR2 in the cytoplasm and nucleus of some cell types, indicating that the contribution of AGR2 to cell proliferation, survival, metastasis and drug resistance may be complex and mutli-faceted. By far the most common characteristic of the proteins identified by the phage-display / database mining analysis was membrane association. Agr2 contains a putative PDZ-interacting domain [Xia et al., 2009] in addition to the thioredoxin-like PDI motif (CXXS), the ER retention motif (KDEL) and the N-terminal signal sequence. PDZ proteins mediate the clustering of ion channels through their C-terminal domains [Doyle et al., 1996], by interacting with a short linear peptide on target proteins. This is in agreement with the large number of ion and solute channels identified by the linear motif-based database mining. Most of the membrane or transmembrane proteins identified are known to be post-translationally modified (in particular, glycosylation), and most contained a large number of cysteines. Enrichment analysis for this gene set also showed that many of these genes are expressed in the developing kidney.

We therefore chose Meckelin (MKS3), a glycosylated, transmembrane,

Figure 7.2: Model of AGR2 as a PDI family chaperone

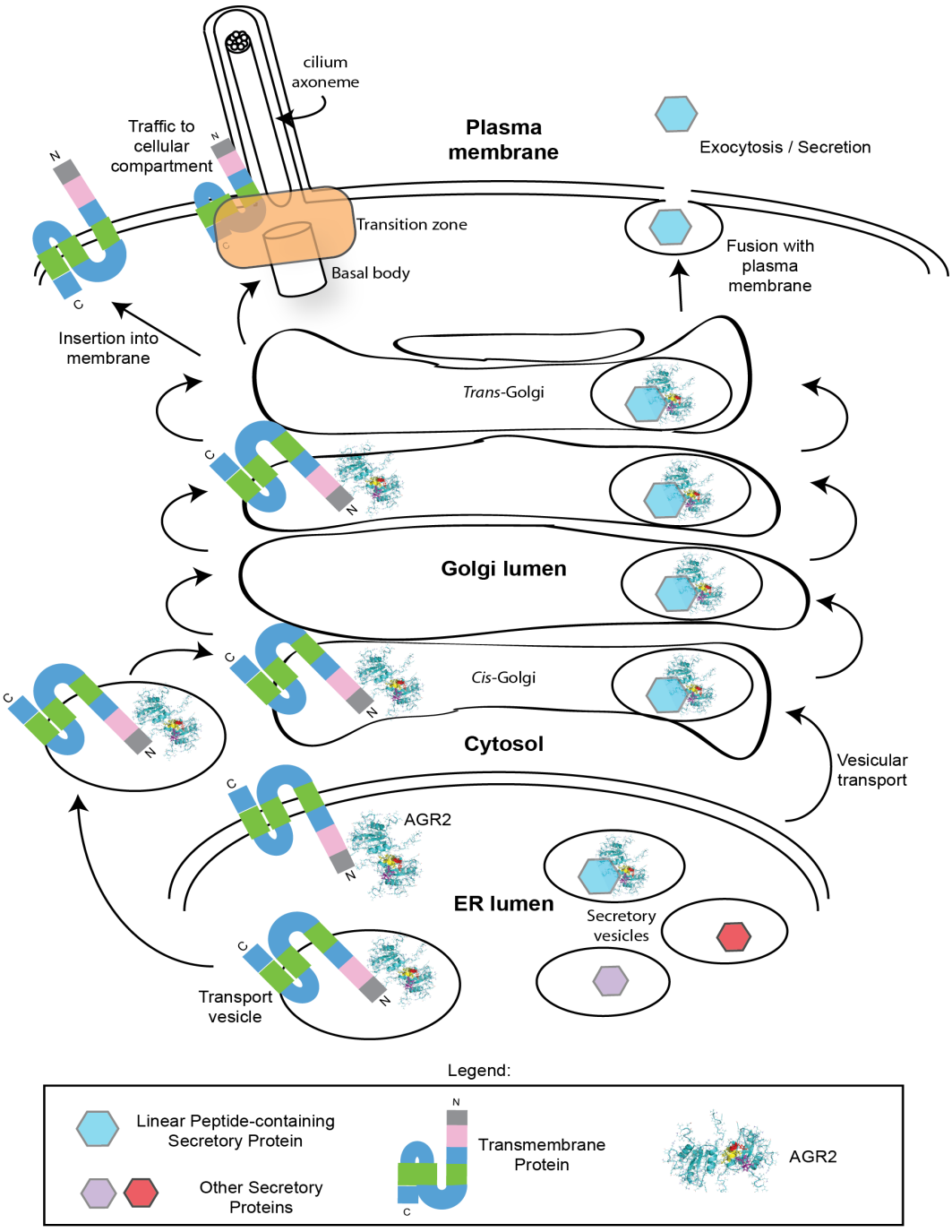


Figure 7.2 Model of AGR2 as PDI chaperone for class of membrane proteins
AGR2 binds to target substrates in the ER lumen and functions as a protein disulphide isomerase (PDI), helping to fold correctly and traffic a subset of transmembrane and secretory proteins as they move through the secretory pathway and on to their resident subcellular locations.

cysteine-rich disease protein, mutations in which cause polycystic kidneys among other defects, to validate as a client protein for AGR2.

Using a number of biochemical approaches, we determined that MKS3 and AGR2 associate in cell lines, when MKS3 is overexpressed. Mutating the linear peptide motif changed the localisation of MKS3 as determined by widefield and TIRF microscopy, with the peptide "loss-of-function" mutated form accumulating in aggregates in the endoplasmic reticulum (ER) and at the plasma membrane. In addition, AGR2 localised to the cilium (and other compartments) in IMCD3 renal epithelial cells, and also localised to the cilium in the developing embryonic mouse kidney. These data support an interaction, either direct or indirect, between AGR2 and MKS3, although further work is being carried out to analyse and verify the interaction *in vivo* and using endogenous MKS3 in cell lines. This task will be aided by the design, optimisation and characterisation of a new anti-MKS3 antibody as described in Chapter 5.

If AGR2 is a member of the PDI chaperone family, helping to isomerise and fold a subset of client proteins into their mature forms while they are trafficked through the secretory pathway, then we would expect a phenotype in mice when Agr2 is functionally deleted. The first published AGR2 mouse knockout paper [Park et al., 2009] did not report major phenotypes such as embryonic lethality or perinatal lethality. In fact, the mice seemed viable and were only susceptible to inflammatory bowel disease in response to dextran sulphate sodium (DSS). In contrast, the second [Zhao et al., 2010] and third [Gupta et al., 2012b] AGR2 mouse knockout papers identified more severe phenotypes both in the gut and in the stomach. The second mouse knockout paper consisted of two knockout lines, one germline and one tamoxifen-inducible, both

with the deletion of exons 2, 3 and 4. They identifed a severe intestinal phenotype upon loss of Agr2, independent of administration of DSS in both lines. They also described a marked increase in Paneth cells, increase in the expression of Paneth cell and stem-cell marker gene Sox9, and a loss of Mucin 2. The loss of Mucin 2 did not include loss of the goblet cell lineage, however, and the inducible mouse line showed that expansion of the Paneth cell population preceded loss of Mucin 2 and intestinal inflammation, suggesting independent pathways. Interestingly, the acute loss of AGR2 in the inducible mouse line caused severe ER stress pathway induction, including increases in apoptosis as monitored by caspase-3 activity as well as decreases in proliferation. The same was not seen in the Agr2 germline knockout mouse line, and the authors hypothesise that this is caused by a gradual upregulation in alternate chaperone / PDI activity that is not possible with the acute deletion of AGR2 seen in the inducible knockout (redundancy). This may explain how AGR2 could function as a chaperone protein for many different proteins including MKS3, (as suggested by the linear motif / database mining analysis, the homology to other PDIs and thioredoxin, and published chaperone-like activity), and yet not prove embryonic or peri-natal lethal when deleted in the germline, instead causing phenotypes that are only related to client proteins for which no alternate PDI is available.

The third mouse knockout paper also removed exons 2, 3 and 4, yet the reported phenotype in these mice was even more severe. The females were infertile and thus necessitated heterozygous breeding to produce Agr2 null mice. These mice were viable at birth, however, by 12 weeks of age most were euthanised due to enlarged stomachs and intestines. Analysis of the stomachs

of these mice showed drastic gastric phenotypes including loss of gastric cell lineages and proliferation of stem-cell like populations, suggesting again that Agr2 may be involved in differentiation of certain cell types, which could possibly include differentiation of nephron progenitors. Although these knockout papers do not report kidney defects, the pronounced intestinal and gastric phenotypes observed may have precluded in-depth analysis of the kidney, however, it is not possible to comment without further analysis of these knockouts.

Although the first reported Agr2 knockout mouse did not appear to have a strong phenotype, this study deleted exon 2 alone to create their germline null mice, and only reported using antibody analysis on tissue by western blot to confirm that the animals were $Agr2^{-}/^{-}$ (null). This commercial antibody was made from a peptide found in exon 2, and would therefore not detect a potential N-terminally truncated protein. In addition, a putative alternate start site that is in frame with the endogenous protein exists near the start of exon 3 of the Agr2 gene. This means that a putative N-terminally truncated protein would retain the original sequence of the Agr2 protein from the new start site onwards. Inspection of this sequence shows that the PDI CXXS motif would be retained, as would the peptide binding interface, the ER-retention sequence at the C-terminus, and the PDI-like loop motif, although the N-terminal leader sequence would be removed. To test this, tissue from this knockout line was obtained, however, the tissue integrity was damaged during transport and thus it was not possible to determine whether an N-terminally truncated protein has been created in this mouse line. It is tempting to hypothesise that such a protein would mimic the mature form of Agr2 (without the leader sequence), although it would not likely be targeted to the ER lumen or secretory pathway,

and the cytoplasmic / nuclear pools that bind p53 may be enriched. If so, it would explain why the first mouse knockout study did not report increased apoptosis in their mice in response to DSS treatment.

Deletion of Meckelin in mice leads to embryonic death or perinatal lethality [Cook et al., 2009], and in humans, truncating mutations also lead to early lethality [Iannicelli et al., 2010]. Some mutations in Meckelin have been identified that lead to a less severe syndrome, however all of these are missense mutations and never truncating mutations [Iannicelli et al., 2010]. In addition to the possibility of redundancy of PDIs in chaperoning membrane proteins that are being trafficked through the secretory pathway, it is possible that the absence of Agr2 in mice does not lead to a complete loss of Meckelin, but rather to a misfolded or partially folded isoform. The TIRF microscopy data support this assertion, with a mutation in the putative linear peptide domain leading to accumulation of MKS3 aggregates. Just as a missense mutation may lead to a less severe phenotype in humans, deletion of Agr2 in mice may lead to a reduced pool of functional Meckelin rather than an absence of the protein, or a pool of partially folded Meckelin retaining some biological function. If true, this could rationalise the findings of an interaction between MKS3 and AGR2 and viability of the Agr2 knockout mouse.

7.2 *In vivo* Expression Pattern of AGR2 in Embryonic Murine Kidney

The expression pattern of Agr2 in the developing kidney was investigated using an antibody-based approach. The putative interaction between Agr2

and MKS3 together with the gene set enrichment data mentioned above, and Gudmap array data showing expression of Agr2 in the developing kidney, suggested that this organ would be rich in AGR2. Using a characterised antibody to AGR2 we analysed the cell-specific localisation of Agr2 in the developing embryonic kidney, using organ culture. The observed expression pattern of Agr2 (in a polarised manner in the cap mesenchyme, in the lumen (apical side) of the ureteric bud, in the apical side of the podocytes and at the proximal end of the developing nephron in a proximo-distal gradient), supports a role for Agr2 in chaperoning membrane proteins. All these renal structures, and in particular the apical side, are rich in ion and solute transporters as well as many signalling proteins such as G-protein coupled receptors. Furthermore, the lumen (apical side) of the ureteric bud, the metanephric mesenchyme and the podocytes have all been shown to have primary cilia, supporting a role for AGR2 in the biology of primary cilia, although this will need to be verified with functional genetic studies *in vivo* or in cell culture.

7.3 Agr2 GFP-Cre driver mouse line

A modular, adaptable expression vector was created for use with BAC transgenics, as described in Chapter 3. This vector was modified via the 5' and 3' multiple cloning sites to be compatible with the murine Agr2 BAC, and recombineering was used to replace the start codon on the Agr2 BAC with the expression cassettes as detailed in Figure 3.7. The resulting GFP-Cre fusion protein-puroR-pA-Neo/Kan expression cassettes integrated near the start of exon 2 (start codon) on the Agr2 BAC was introduced into the mouse genome

by oocyte microinjection. BAC transgenics have several advantages in that they are usually integrated as a single copy due to their large size, and contain much of the upstream and downstream sequences together with intronic sequences that may be important for the expression of a gene. However, we chose to integrate the modified BAC randomly rather than targeting to a particular locus or to the endogenous locus (a knock-in). It is likely that the low or possibly absent expression of GFP seen in the mice and embryos analysed from this transgenic line is due to sub-optimal localisation of the BAC in the genome, at a region that is silenced or inaccessible to transcription machinery.

The plethora of recent publications on AGR2 and first mouse knockout papers highlight the growing community of researchers investigating AGR2 from several different perspectives - cancer research, pathology of GI disorders, asthma, ER homeostasis, and differentiation cell types. There is therefore an increasing need for tools to study AGR2 *in vivo*, and a Cre-driver mouse would be valuable for such purposes. In addition, a thorough characterisation of AGR2 expression patterns in the tissues of both embryonic and adult mice would provide important background data for researchers looking to investigate the mechanisms of AGR2, and looking to design treatments for AGR2-mediated drug resistance. Retrieval of the AGR2 targeting cassette into a targeting vector as described in Figure 3.5 and creation of an Agr2 GFP-Cre knock-in mouse is a current consideration for future work.

7.4 Ongoing work arising from this project and future outlook

In addition to the research described in this thesis, ongoing work has been undertaken. A GFP-miRNA construct to knockdown Agr2 in mouse cells has been cloned, and testing is underway with the intention to knock down Agr2 in murine IMCD3 cells. This would allow us to carry out phenotypic analysis of cilia both in morphology and in number. In addition, a series of TALENs pairs to cut in the first coding exon of Agr2 have been generated. These are currently being used to generate a frameshift mutation in IMCD3 cells to create a mouse knockout cell line. The TALENs pairs may also be used to generate an in-house Agr2 mouse knockout line, although work on this has not yet begun, and these TALENs may also be used to make the generation of the promoter reporter model at the endogenous locus more efficient.

As well as the pMULITrec with GFP-Cre expression cassette described in Chapter 3, an mCherry version was also constructed (Figure 3.3). Homologous arms to the start codon of the Meckelin gene have been cloned into the multiple cloning sites at the 5' and 3' ends of the mCherry construct, and work has begun to recombine this construct into a Meckelin BAC. If successful, this would provide a Meckelin promoter reporter for analysis of Meckelin protein expression *in vivo*, and a mouse line that could be crossed to an Agr2 promoter reporter such as a GFP-Cre endogenous knock-in for co-expression analysis.

The list of potential interacting partners for AGR2 identified by database mining with the linear peptide motif provides a unique avenue to probe the interactome of AGR2. In this project one "hit" from the list was chosen for

in-depth validation. Further work remains to be carried out including analysis of cilia phenotype and localisation patterns after knockdown or knockout of Agr2 in IMCD3 cells or AGR2 in human MCF7 cells. The new anti-MKS3 antibodies should also be used with AGR2 antibodies to analyse endogenous co-localisation and co-immunoprecipitation of the two proteins. The discovery and further validation of an interaction between the disease-associated MKS3 and AGR2 might provide a functional assay for AGR2 activity and be important in research into the pathophysiology of Meckelin-associated syndromes. However, the entire putative client set for AGR2 as a PDI chaperone is potentially just as useful, and screening the client set for changes in response to reduced or absent AGR2 could give insights into the specificity or level of redundancy of AGR2. Probing this client set further may thus provide a novel biological model to dissect the role of AGR2 in ER-trafficking, and validation of the linear peptide motif as an important part of its action as a chaperone protein could provide a new target for disrupting these interactions with small molecules in the future.

Chapter 8

References

[Aberger et al., 1998] Aberger, F., Weidinger, G., Grunz, H. and Richter, K. (1998). Anterior specification of embryonic ectoderm: the role of the Xenopus cement gland-specific gene XAG-2. Mech Dev *72*, 115–130.

[Adams et al., 2012] Adams, M., Simms, R. J., Abdelhamed, Z., Dawe, H. R., Szymanska, K., Logan, C. V., Wheway, G., Pitt, E., Gull, K., Knowles, M. A., Blair, E., Cross, S. H., Sayer, J. A. and Johnson, C. A. (2012). A meckelin-filamin A interaction mediates ciliogenesis. Hum Mol Genet *21*, 1272–1286.

[Adolph et al., 2012] Adolph, T.-E., Niederreiter, L., Blumberg, R. S. and Kaser, A. (2012). Endoplasmic reticulum stress and inflammation. Dig Dis *30*, 341–346.

[Akasofu et al., 1989] Akasofu, H., Yamauchi, D., Mitsuhashi, W. and Minamikawa, T. (1989). Nucleotide sequence of cDNA for sulfhydryl-

endopeptidase (SH-EP) from cotyledons of germinating Vigna mungo seeds. Nucleic Acids Res *17*, 6733.

[Akilesh et al., 2008] Akilesh, S., Huber, T. B., Wu, H., Wang, G., Hartleben, B., Kopp, J. B., Miner, J. H., Roopenian, D. C., Unanue, E. R. and Shaw, A. S. (2008). Podocytes use FcRn to clear IgG from the glomerular basement membrane. Proc Natl Acad Sci U S A *105*, 967–972.

[Alaiwi et al., 2009] Alaiwi, W. A. A., Lo, S. T. and Nauli, S. M. (2009). Primary cilia: highly sophisticated biological sensors. Sensors (Basel) *9*, 7003–7020.

[Alanen et al., 2011] Alanen, H. I., Raykhel, I. B., Luukas, M. J., Salo, K. E. H. and Ruddock, L. W. (2011). Beyond KDEL: the role of positions 5 and 6 in determining ER localization. J Mol Biol *409*, 291–297.

[Alberts et al., 2002] Alberts, B., Johnson, A. and Lewis, J., eds (2002). Molecular Biology of the Cell. 4th edition. Garland ScienceDisruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly.

[Anelli et al., 2002] Anelli, T., Alessio, M., Mezghrani, A., Simmen, T., Talamo, F., Bachi, A. and Sitia, R. (2002). ERp44, a novel endoplasmic reticulum folding assistant of the thioredoxin family. EMBO J *21*, 835–844.

[Armes et al., 2013] Armes, J. E., Davies, C. M., Wallace, S., Taheri, T., Perrin, L. C. and Autelitano, D. J. (2013). AGR2 expression in ovarian tumours: a potential biomarker for endometrioid and mucinous differentiation. Pathology *45*, 49–54.

[Avery et al., 1944] Avery, O. T., Macleod, C. M. and McCarty, M. (1944). STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. J Exp Med *79*, 137–158.

[Baala et al., 2007] Baala, L., Romano, S., Khaddour, R., Saunier, S., Smith, U. M., Audollent, S., Ozilou, C., Faivre, L., Laurent, N., Foliguet, B., Munnich, A., Lyonnet, S., Salomon, R., Encha-Razavi, F., Gubler, M.-C., Boddaert, N., de Lonlay, P., Johnson, C. A., Vekemans, M., Antignac, C. and Attie-Bitach, T. (2007). The Meckel-Gruber syndrome gene, MKS3, is mutated in Joubert syndrome. Am J Hum Genet *80*, 186–194.

[Barak and Boyle, 2011] Barak, H. and Boyle, S. C. (2011). Organ culture and immunostaining of mouse embryonic kidneys. Cold Spring Harb Protoc *2011*, pdb.prot5558.

[Barraclough et al., 2009] Barraclough, D. L., Platt-Higgins, A., de Silva Rudland, S., Barraclough, R., Winstanley, J., West, C. R. and Rudland, P. S. (2009). The metastasis-associated anterior gradient 2 protein is correlated with poor survival of breast cancer patients. Am J Pathol *175*, 1848–1857.

[Barraclough et al., 2010] Barraclough, D. L., Sewart, S., Rudland, P. S., Shoker, B. S., Sibson, D. R., Barraclough, R. and Davies, M. P. A. (2010). Microarray analysis of suppression subtracted hybridisation libraries identifies genes associated with breast cancer progression. Cell Oncol *32*, 87–99.

[Barry et al., 2012] Barry, S., Chelala, C., Lines, K., Sunamura, M., Wang, A., Marelli-Berg, F. M., Brennan, C., Lemoine, N. R. and Crnogorac-Jurcevic, T. (2012). S100P is a metastasis-associated gene that facilitates transendothelial migration of pancreatic cancer cells. Clin Exp Metastasis *1*, 1–5.

[Benedito et al., 2009] Benedito, R., Roca, C., SŽrensen, I., Adams, S., Gossler, A., Fruttiger, M. and Adams, R. H. (2009). The notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. Cell *137*, 1124–1135.

[Berbari et al., 2009] Berbari, N. F., O'Connor, A. K., Haycraft, C. J. and Yoder, B. K. (2009). The primary cilium as a complex signaling center. Curr Biol *19*, R526–R535.

[Bhanot et al., 1996] Bhanot, P., Brink, M., Samos, C. H., Hsieh, J. C., Wang, Y., Macke, J. P., Andrew, D., Nathans, J. and Nusse, R. (1996). A new member of the frizzled family from Drosophila functions as a Wingless receptor. Nature *382*, 225–230.

[Bird, 2007] Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396–398.

[Bogaert et al., 2011] Bogaert, S., Vos, M. D., Olievier, K., Peeters, H., Elewaut, D., Lambrecht, B., Pouliot, P. and Laukens, D. (2011). Involvement of endoplasmic reticulum stress in inflammatory bowel disease: a different implication for colonic and ileal disease? PLoS One *6*, e25589.

[Boutros and Ahringer, 2008] Boutros, M. and Ahringer, J. (2008). The art and design of genetic screens: RNA interference. Nat Rev Genet *9*, 554–566.

[Boyle et al., 2008] Boyle, S., Misfeldt, A., Chandler, K. J., Deal, K. K., Southard-Smith, E. M., Mortlock, D. P., Baldwin, H. S. and de Caestecker, M. (2008). Fate mapping using Cited1-CreERT2 mice demonstrates that the cap mesenchyme contains self-renewing progenitor cells and gives rise exclusively to nephronic epithelia. Dev Biol *313*, 234–245.

[Bruckner et al., 2009] Bruckner, A., Polge, C., Lentze, N., Auerbach, D. and Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci *10*, 2763–2788.

[Bryant and Mostov, 2008] Bryant, D. M. and Mostov, K. E. (2008). From cells to organs: building polarized tissue. Nat Rev Mol Cell Biol *9*, 887–901.

[Brychtova et al., 2011] Brychtova, V., Vojtesek, B. and Hrstka, R. (2011). Anterior gradient 2: a novel player in tumor cell biology. Cancer Lett *304*, 1–7.

[Bu et al., 2011] Bu, H., Bormann, S., SchŁfer, G., Horninger, W., Massoner, P., Neeb, A., Lakshmanan, V.-K., Maddalo, D., Nestl, A., S§ltmann, H., Cato, A. C. B. and Klocker, H. (2011). The anterior gradient 2 (AGR2) gene is overexpressed in prostate cancer and may be useful as a urine sediment marker for prostate cancer detection. Prostate *71*, 575–587.

[Burry, 2010] Burry, R. W. (2010). Immunocytochemistry. Springer.

[Camilo et al., 2012] Camilo, V., Barros, R., Sousa, S., MagalhŃes, A. M., Lopes, T., Santos, A. M., Pereira, T., Figueiredo, C., David, L. and Almeida,

R. (2012). Helicobacter pylori and the BMP pathway regulate CDX2 and SOX2 expression in gastric cells. Carcinogenesis *33*, 1985–1992.

[Cantagrel et al., 2008] Cantagrel, V., Silhavy, J. L., Bielas, S. L., Swistun, D., Marsh, S. E., Bertrand, J. Y., Audollent, S., AttiŐ-Bitach, T., Holden, K. R., Dobyns, W. B., Traver, D., Al-Gazali, L., Ali, B. R., Lindner, T. H., Caspary, T., Otto, E. A., Hildebrandt, F., Glass, I. A., Logan, C. V., Johnson, C. A., Bennett, C., Brancati, F., Group, I. J. S. R. D. S., Valente, E. M., Woods, C. G. and Gleeson, J. G. (2008). Mutations in the cilia gene ARL13B lead to the classical form of Joubert syndrome. Am J Hum Genet *83*, 170–179.

[Chen, 2007] Chen, C.-P. (2007). Meckel syndrome: genetics, perinatal findings, and differential diagnosis. Taiwan J Obstet Gynecol *46*, 9–14.

[Chen et al., 2010] Chen, R., Pan, S., Duan, X., Nelson, B. H., Sahota, R. A., de Rham, S., Kozarek, R. A., McIntosh, M. and Brentnall, T. A. (2010). Elevated level of anterior gradient-2 in pancreatic juice from patients with pre-malignant pancreatic neoplasia. Mol Cancer *9*, 149.

[Chen et al., 2012] Chen, Y.-C., Lu, Y.-F., Li, I.-C. and Hwang, S.-P. L. (2012). Zebrafish Agr2 is required for terminal differentiation of intestinal goblet cells. PLoS One *7*, e34408.

[Cheng et al., 2007] Cheng, H.-T., Kim, M., Valerius, M. T., Surendran, K., Schuster-Gossler, K., Gossler, A., McMahon, A. P. and Kopan, R. (2007). Notch2, but not Notch1, is required for proximal fate acquisition in the mammalian nephron. Development *134*, 801–811.

[Cheo et al., 2004] Cheo, D. L., Titus, S. A., Byrd, D. R. N., Hartley, J. L., Temple, G. F. and Brasch, M. A. (2004). Concerted assembly and cloning of multiple DNA segments using in vitro site-specific recombination: functional analysis of multi-segment expression clones. Genome Res *14*, 2111–2120.

[Chung et al., 2012] Chung, K., Nishiyama, N., Wanibuchi, H., Yamano, S., Hanada, S., Wei, M., Suehiro, S. and Kakehashi, A. (2012). AGR2 as a potential biomarker of human lung adenocarcinoma. Osaka City Med J *58*, 13–24.

[Chung et al., 2011] Chung, K., Nishiyama, N., Yamano, S., Komatsu, H., Hanada, S., Wei, M., Wanibuchi, H., Suehiro, S. and Kakehashi, A. (2011). Serum AGR2 as an early diagnostic and postoperative prognostic biomarker of human lung adenocarcinoma. Cancer Biomark *10*, 101–107.

[Coligan et al., 1996] Coligan, J., Kruisbeek, A., Margulies, D., Shevach, E. and Strober, W., eds (1996). Current Protocols in Immunology. John Wiley & Sons; New York; pp. 9.0.1Ð9.8.15.

[Como et al., 2002] Como, C. J. D., Urist, M. J., Babayan, I., Drobnjak, M., Hedvat, C. V., Teruya-Feldstein, J., Pohar, K., Hoos, A. and Cordon-Cardo, C. (2002). p63 expression profiles in human normal and tumor tissues. Clin Cancer Res *8*, 494–501.

[Cook et al., 2009] Cook, S. A., Collin, G. B., Bronson, R. T., Naggert, J. K., Liu, D. P., Akeson, E. C. and Davisson, M. T. (2009). A mouse model for Meckel syndrome type 3. J Am Soc Nephrol *20*, 753–764.

[Copeland et al., 2001] Copeland, N. G., Jenkins, N. A. and Court, D. L. (2001). Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet *2*, 769–779.

[Costantini, 2006] Costantini, F. (2006). Renal branching morphogenesis: concepts, questions, and recent advances. Differentiation *74*, 402–421.

[Costantini and Kopan, 2010] Costantini, F. and Kopan, R. (2010). Patterning a Complex Organ: Branching Morphogenesis and Nephron Segmentation in Kidney Development. Dev Cell *18*, 698–712.

[Costantini and Shakya, 2006] Costantini, F. and Shakya, R. (2006). GDNF/Ret signaling and the development of the kidney. Bioessays *28*, 117–127.

[Crameri and Walter, 1999] Crameri, R. and Walter, G. (1999). Selective enrichment and high-throughput screening of phage surface-displayed cDNA libraries from complex allergenic systems. Comb Chem High Throughput Screen *2*, 63–72.

[da Silva et al., 2002] da Silva, S. M., Gates, P. B. and Brockes, J. P. (2002). The newt ortholog of CD59 is implicated in proximodistal identity during amphibian limb regeneration. Dev Cell *3*, 547–555.

[Darb-Esfahani et al., 2012] Darb-Esfahani, S., Fritzsche, F., Kristiansen, G., Weichert, W., Sehouli, J., Braicu, I., Dietel, M. and Denkert, C. (2012). Anterior gradient protein 2 (AGR2) is an independent prognostic factor in ovarian high-grade serous carcinoma. Virchows Arch *461*, 109–116.

[Davies and Bard, 1998] Davies, J. A. and Bard, J. B. (1998). The development of the kidney. Curr Top Dev Biol *39*, 245–301.

[Davies et al., 2012] Davies, J. A., Little, M. H., Aronow, B., Armstrong, J., Brennan, J., Lloyd-MacGilp, S., Armit, C., Harding, S., Piu, X., Roochun, Y., Haggarty, B., Houghton, D., Davidson, D. and Baldock, R. (2012). Access and use of the GUDMAP database of genitourinary development. Methods Mol Biol *886*, 185–201.

[Dawe et al., 2009] Dawe, H. R., Adams, M., Wheway, G., Szymanska, K., Logan, C. V., Noegel, A. A., Gull, K. and Johnson, C. A. (2009). Nesprin-2 interacts with meckelin and mediates ciliogenesis via remodelling of the actin cytoskeleton. J Cell Sci *122*, 2716–2726.

[Dawe et al., 2007a] Dawe, H. R., Farr, H. and Gull, K. (2007a). Centriole/basal body morphogenesis and migration during ciliogenesis in animal cells. J Cell Sci *120*, 7–15.

[Dawe et al., 2007b] Dawe, H. R., Smith, U. M., Cullinane, A. R., Gerrelli, D., Cox, P., Badano, J. L., Blair-Reid, S., Sriram, N., Katsanis, N., Attie-Bitach, T., Afford, S. C., Copp, A. J., Kelly, D. A., Gull, K. and Johnson, C. A. (2007b). The Meckel-Gruber Syndrome proteins MKS1 and meckelin interact and are required for primary cilium formation. Hum Mol Genet *16*, 173–186.

[Dong et al., 2011] Dong, A., Gupta, A., Pai, R. K., Tun, M. and Lowe, A. W. (2011). The human adenocarcinoma-associated gene, AGR2, induces expression of amphiregulin through Hippo pathway co-activator YAP1 activation. J Biol Chem *286*, 18301–18310.

[Dornan et al., 2003] Dornan, D., Shimizu, H., Burch, L., Smith, A. J. and Hupp, T. R. (2003). The proline repeat domain of p53 binds directly to the transcriptional coactivator p300 and allosterically controls DNA-dependent acetylation of p53. Mol Cell Biol *23*, 8846–8861.

[Doyle et al., 1996] Doyle, D. A., Lee, A., Lewis, J., Kim, E., Sheng, M. and MacKinnon, R. (1996). Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell *85*, 1067–1076.

[Dressler, 2006] Dressler, G. R. (2006). The cellular basis of kidney development. Annu Rev Cell Dev Biol *22*, 509–529.

[Dryer and Reiser, 2010] Dryer, S. E. and Reiser, J. (2010). TRPC6 channels and their binding partners in podocytes: role in glomerular filtration and pathophysiology. Am J Physiol Renal Physiol *299*, F689–F701.

[Dumartin et al., 2011] Dumartin, L., Whiteman, H. J., Weeks, M. E., Hariharan, D., Dmitrovic, B., Iacobuzio-Donahue, C. A., Brentnall, T. A., Bronner, M. P., Feakins, R. M., Timms, J. F., Brennan, C., Lemoine, N. R. and Crnogorac-Jurcevic, T. (2011). AGR2 is a novel surface antigen that promotes the dissemination of pancreatic cancer cells through regulation of cathepsins B and D. Cancer Res *71*, 7091–7102.

[Duran et al., 2008] Duran, M. C., Vega, F., Moreno-Bueno, G., Artiga, M. J., Sanchez, L., Palacios, J., Ridley, A. and Timms, J. F. (2008). Characterisation of tumoral markers correlated with ErbB2 (HER2/Neu) overexpression and metastasis in breast cancer. Proteomics Clin Appl *2*, 1313–1326.

[Dyson and Wright, 2005] Dyson, H. J. and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol *6*, 197–208.

[Edgell et al., 2010] Edgell, T. A., Barraclough, D. L., Rajic, A., Dhulia, J., Lewis, K. J., Armes, J. E., Barraclough, R., Rudland, P. S., Rice, G. E. and Autelitano, D. J. (2010). Increased plasma concentrations of anterior gradient 2 protein are positively associated with ovarian cancer. Clin Sci (Lond) *118*, 717–725.

[Ellis, 1999] Ellis, R. J. (1999). Molecular chaperones: pathways and networks. Curr Biol *9*, R137–R139.

[Ellis, 2006] Ellis, R. J. (2006). Molecular chaperones: assisting assembly in addition to folding. Trends Biochem Sci *31*, 395–401.

[Fanning et al., 1998] Fanning, A. S., Jameson, B. J., Jesaitis, L. A. and Anderson, J. M. (1998). The tight junction protein ZO-1 establishes a link between the transmembrane protein occludin and the actin cytoskeleton. J Biol Chem *273*, 29745–29753.

[Feifel et al., 1998] Feifel, B., SchŽnfeld, H. J. and Christen, P. (1998). D-peptide ligands for the co-chaperone DnaJ. J Biol Chem *273*, 11999–12002.

[Feil et al., 1997] Feil, R., Wagner, J., Metzger, D. and Chambon, P. (1997). Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. Biochem Biophys Res Commun *237*, 752–757.

[Ferrari and Soling, 1999] Ferrari, D. M. and Soling, H. D. (1999). The protein disulphide-isomerase family: unravelling a string of folds. Biochem J *339 (Pt 1)*, 1–10.

[Fire et al., 1998] Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806–811.

[Fletcher et al., 2003] Fletcher, G. C., Patel, S., Tyson, K., Adam, P. J., Schenker, M., Loader, J. A., Daviet, L., Legrain, P., Parekh, R., Harris, A. L. and Terrett, J. A. (2003). hAG-2 and hAG-3, human homologues of genes involved in differentiation, are associated with oestrogen receptor-positive breast tumours and interact with metastasis gene C4.4a and dystroglycan. Br J Cancer *88*, 579–585.

[Flicek et al., 2012] Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L., Gordon, L., Hendrix, M., Hourlier, T., Johnson, N., KŁhŁri, A. K., Keefe, D., Keenan, S., Kinsella, R., Komorowska, M., Koscielny, G., Kulesha, E., Larsson, P., Longden, I., McLaren, W., Muffato, M., Overduin, B., Pignatelli, M., Pritchard, B., Riat, H. S., Ritchie, G. R. S., Ruffier, M., Schuster, M., Sobral, D., Tang, Y. A., Taylor, K., Trevanion, S., Vandrovcova, J., White, S., Wilson, M., Wilder, S. P., Aken, B. L., Birney, E., Cunningham, F., Dunham, I., Durbin, R., FernĞndez-Suarez, X. M., Harrow, J., Herrero, J., Hubbard, T. J. P., Parker, A., Proctor, G., Spudich, G., Vogel, J., Yates, A., Zadissa, A. and Searle, S. M. J. (2012). Ensembl 2012. Nucleic Acids Res *40*, D84–D90.

[Fourtouna et al., 2009] Fourtouna, A., Murray, E., Nicholson, J., Maslon, M. M., Pang, L. Y., Dryden, D. T. and Hupp, T. R. (2009). The Anterior Gradient-2 Pathway as a Model for Developing Peptide-Aptamer Anti-Cancer Drug Leads that Stimulate p53 Function. Curr Chem Biol *3*, 124–137 124–137.

[Friedrich and Soriano, 1991] Friedrich, G. and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev *5*, 1513–1523.

[Fritzsche et al., 2007] Fritzsche, F. R., Dahl, E., Dankof, A., Burkhardt, M., Pahl, S., Petersen, I., Dietel, M. and Kristiansen, G. (2007). Expression of AGR2 in non small cell lung cancer. Histol Histopathol *22*, 703–708.

[Garcia-Gonzalo et al., 2011] Garcia-Gonzalo, F. R., Corbit, K. C., Sirerol-Piquer, M. S., Ramaswami, G., Otto, E. A., Noriega, T. R., Seol, A. D., Robinson, J. F., Bennett, C. L., Josifova, D. J., GarcŠa-Verdugo, J. M., Katsanis, N., Hildebrandt, F. and Reiter, J. F. (2011). A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. Nat Genet *43*, 776–784.

[Georgas et al., 2009] Georgas, K., Rumballe, B., Valerius, M. T., Chiu, H. S., Thiagarajan, R. D., Lesieur, E., Aronow, B. J., Brunskill, E. W., Combes, A. N., Tang, D., Taylor, D., Grimmond, S. M., Potter, S. S., McMahon, A. P. and Little, M. H. (2009). Analysis of early nephron patterning reveals a role for distal RV proliferation in fusion to the ureteric tip via a cap mesenchyme-derived connecting segment. Dev Biol *332*, 273–286.

[Gillikin et al., 1997] Gillikin, J. W., Zhang, F., Coleman, C. E., Bass, H. W., Larkins, B. A. and Boston, R. S. (1997). A defective signal peptide tethers the floury-2 zein to the endoplasmic reticulum membrane. Plant Physiol *114*, 345–352.

[Giraldo and Montoliu, 2001] Giraldo, P. and Montoliu, L. (2001). Size matters: use of YACs, BACs and PACs in transgenic animals. Transgenic Res *10*, 83–103.

[Giuliani et al., 2008] Giuliani, S., Perin, L., Sedrakyan, S., Kokorowski, P., Jin, D. and Filippo, R. D. (2008). Ex vivo whole embryonic kidney culture: a novel method for research in development, regeneration and transplantation. J Urol *179*, 365–370.

[Goetz et al., 2009] Goetz, S. C., Ocbina, P. J. R. and Anderson, K. V. (2009). The primary cilium as a Hedgehog signal transduction machine. Methods Cell Biol *94*, 199–222.

[Goldstein and Macara, 2007] Goldstein, B. and Macara, I. G. (2007). The PAR proteins: fundamental players in animal cell polarization. Dev Cell *13*, 609–622.

[Gray et al., 2012] Gray, T. A., MacLaine, N. J., Michie, C. O., Bouchalova, P., Murray, E., Howie, J., Hrstka, R., Maslon, M. M., Nenutil, R., Vojtesek, B., Langdon, S., Hayward, L., Gourley, C. and Hupp, T. R. (2012). Anterior Gradient-3: a novel biomarker for ovarian cancer that mediates cisplatin resistance in xenograft models. J Immunol Methods *378*, 20–32.

[Greka and Mundel, 2012] Greka, A. and Mundel, P. (2012). Cell biology and pathology of podocytes. Annu Rev Physiol *74*, 299–323.

[Gu et al., 1993] Gu, H., Zou, Y. R. and Rajewsky, K. (1993). Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. Cell *73*, 1155–1164.

[Gullberg et al., 2003] Gullberg, M., Fredriksson, S., Taussig, M., Jarvius, J., Gustafsdottir, S. and Landegren, U. (2003). A sense of closeness: protein detection by proximity ligation. Curr Opin Biotechnol *14*, 82–86.

[Gunay-Aygun et al., 2009] Gunay-Aygun, M., Parisi, M. A., Doherty, D., Tuchman, M., Tsilou, E., Kleiner, D. E., Huizing, M., Turkbey, B., Choyke, P., Guay-Woodford, L., Heller, T., Szymanska, K., Johnson, C. A., Glass, I. and Gahl, W. A. (2009). MKS3-related ciliopathy with features of autosomal recessive polycystic kidney disease, nephronophthisis, and Joubert Syndrome. J Pediatr *155*, 386–92.e1.

[Gupta et al., 2012a] Gupta, A., Dong, A. and Lowe, A. W. (2012a). AGR2 gene function requires a unique endoplasmic reticulum localization motif. J Biol Chem *287*, 4773–4782.

[Gupta et al., 2012b] Gupta, A., Wodziak, D., Tun, M., Bouley, D. M. and Lowe, A. W. (2012b). Loss of Anterior Gradient 2 (Agr2) Expression Results in Hyperplasia and Defective Lineage Maturation in the Murine Stomach. J Biol Chem *287*, 4773–4785.

[Hamilton and Abremski, 1984] Hamilton, D. L. and Abremski, K. (1984). Site-specific recombination by the bacteriophage P1 lox-Cre system. Cre-mediated synapsis of two lox sites. J Mol Biol *178*, 481–486.

[Harding et al., 2011] Harding, S. D., Armit, C., Armstrong, J., Brennan, J., Cheng, Y., Haggarty, B., Houghton, D., Lloyd-MacGilp, S., Pi, X., Roochun, Y., Sharghi, M., Tindal, C., McMahon, A. P., Gottesman, B., Little, M. H., Georgas, K., Aronow, B. J., Potter, S. S., Brunskill, E. W., Southard-Smith, E. M., Mendelsohn, C., Baldock, R. A., Davies, J. A. and Davidson, D. (2011). The GUDMAP database–an online resource for genitourinary research. Development *138*, 2845–2853.

[Harlow and Lane, 1988] Harlow, E. and Lane, D. (1988). Antibodies. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

[Hartley et al., 2000] Hartley, J. L., Temple, G. F. and Brasch, M. A. (2000). DNA cloning using in vitro site-specific recombination. Genome Res *10*, 1788–1795.

[Hartley, 2002] Hartley, O. (2002). The use of phage display in the study of receptors and their ligands. J Recept Signal Transduct Res *22*, 373–392.

[He et al., 2005] He, M. M., Smith, A. S., Oslob, J. D., Flanagan, W. M., Braisted, A. C., Whitty, A., Cancilla, M. T., Wang, J., Lugovskoy, A. A., Yoburn, J. C., Fung, A. D., Farrington, G., Eldredge, J. K., Day, E. S., Cruz, L. A., Cachero, T. G., Miller, S. K., Friedman, J. E., Choong, I. C. and Cunningham, B. C. (2005). Small-molecule inhibition of TNF-alpha. Science *310*, 1022–1025.

[Helling et al., 1994] Helling, F., Shang, A., Calves, M., Zhang, S., Ren, S., Yu, R. K., Oettgen, H. F. and Livingston, P. O. (1994). GD3 vaccines for melanoma: superior immunogenicity of keyhole limpet hemocyanin conjugate vaccines. Cancer Res *54*, 197–203.

[Hengel et al., 2011] Hengel, S. M., Murray, E., Langdon, S., Hayward, L., O'Donoghue, J., Panchaud, A., Hupp, T. and Goodlett, D. R. (2011). Data-independent proteomic screen identifies novel tamoxifen agonist that mediates drug resistance. J Proteome Res *10*, 4567–4578.

[Hennecke et al., 2001] Hennecke, M., Kwissa, M., Metzger, K., Oumard, A., KrŽger, A., Schirmbeck, R., Reimann, J. and Hauser, H. (2001). Composition and arrangement of genes define the strength of IRES-driven translation in bicistronic mRNAs. Nucleic Acids Res *29*, 3327–3334.

[Higa et al., 2011] Higa, A., Mulot, A., Delom, F., Bouchecareilh, M., NguyŘn, D. T., Boismenu, D., Wise, M. J. and Chevet, E. (2011). Role of pro-oncogenic protein disulfide isomerase (PDI) family member anterior gradient 2 (AGR2) in the control of endoplasmic reticulum homeostasis. J Biol Chem *286*, 44855–44868.

[Hilschmann and Craig, 1965] Hilschmann, N. and Craig, L. C. (1965). Amino acid sequence studies with Bence-Jones proteins. Proc Natl Acad Sci U S A *53*, 1403–1409.

[Holden et al., 2010] Holden, S. J., Uphoff, S., Hohlbein, J., Yadin, D., Reste, L. L., Britton, O. J. and Kapanidis, A. N. (2010). Defining the limits of single-molecule FRET resolution in TIRF microscopy. Biophys J *99*, 3102–3111.

[Hoogenboom et al., 1998] Hoogenboom, H. R., de BruŢne, A. P., Hufton, S. E., Hoet, R. M., Arends, J. W. and Roovers, R. C. (1998). Antibody phage display technology and its applications. Immunotechnology *4*, 1–20.

[Hrstka et al., 2010] Hrstka, R., Nenutil, R., Fourtouna, A., Maslon, M. M., Naughton, C., Langdon, S., Murray, E., Larionov, A., Petrakova, K., Muller, P., Dixon, M. J., Hupp, T. R. and Vojtesek, B. (2010). The pro-metastatic protein anterior gradient-2 predicts poor prognosis in tamoxifen-treated breast cancers. Oncogene *29*, 4838–4847.

[Hu et al., 2012] Hu, Z., Gu, Y., Han, B., Zhang, J., Li, Z., Tian, K., Young, C. Y. F. and Yuan, H. (2012). Knockdown of AGR2 induces cellular senescence in prostate cancer cells. Carcinogenesis *33*, 1178–1186.

[Huang et al., 2008] Huang, H., Li, L., Wu, C., Schibli, D., Colwill, K., Ma, S., Li, C., Roy, P., Ho, K., Songyang, Z., Pawson, T., Gao, Y. and Li, S. S.-C. (2008). Defining the specificity space of the human SRC homology 2 domain. Mol Cell Proteomics *7*, 768–784.

[Huang et al., 2010] Huang, Y., Wilkinson, G. F. and Willars, G. B. (2010). Role of the signal peptide in the synthesis and processing of the glucagon-like peptide-1 receptor. Br J Pharmacol *159*, 237–251.

[Huber et al., 2004] Huber, M., Bahr, I., KrŁtzschmar, J. R., Becker, A., M§ller, E.-C., Donner, P., Pohlenz, H.-D., Schneider, M. R. and Sommer, A. (2004). Comparison of proteomic and genomic analyses of the human breast cancer cell line T47D and the antiestrogen-resistant derivative T47D-r. Mol Cell Proteomics *3*, 43–55.

[Iannicelli et al., 2010] Iannicelli, M., Brancati, F., Mougou-Zerelli, S., Mazzotta, A., Thomas, S., Elkhartoufi, N., Travaglini, L., Gomes, C., Ardissino, G. L., Bertini, E., Boltshauser, E., Castorina, P., D'Arrigo, S., Fischetto, R., Leroy, B., Loget, P., BonniŔre, M., Starck, L., Tantau, J., Gentilin, B., Majore, S., Swistun, D., Flori, E., Lalatta, F., Pantaleoni, C., Penzien, J., Grammatico, P., Group, I. J. S., Dallapiccola, B., Gleeson, J. G., Attie-Bitach, T. and Valente, E. M. (2010). Novel TMEM67 mutations and genotype-phenotype correlates in meckelin-related ciliopathies. Hum Mutat *31*, E1319–E1331.

[Inohara et al., 1989] Inohara, N., Shimomura, S., Fukui, T. and Futai, M. (1989). Auxin-binding protein located in the endoplasmic reticulum of maize shoots: molecular cloning and complete primary structure. Proc Natl Acad Sci U S A *86*, 3564–3568.

[Irion et al., 2007] Irion, S., Luche, H., Gadue, P., Fehling, H. J., Kennedy, M. and Keller, G. (2007). Identification and targeting of the ROSA26 locus in human embryonic stem cells. Nat Biotechnol *25*, 1477–1482.

[Ishikawa et al., 2012] Ishikawa, H., Thompson, J., Yates, J. R. and Marshall, W. F. (2012). Proteomic analysis of mammalian primary cilia. Curr Biol *22*, 414–419.

[Ivanova et al., 2013] Ivanova, A. S., Tereshina, M. B., Ermakova, G. V., Belousov, V. V. and Zaraisky, A. G. (2013). Agr genes, missing in amniotes, are involved in the body appendages regeneration in frog tadpoles. Sci Rep *3*, 1279.

[Janeway et al., 2001] Janeway, C. A., Travers, P., Walport, M. and Shlomchik, M. J. (2001). Immunobiology 5th edition. Garland Science.

[Jarvius et al., 2006] Jarvius, J., Melin, J., GŽransson, J., Stenberg, J., Fredriksson, S., Gonzalez-Rey, C., Bertilsson, S. and Nilsson, M. (2006). Digital quantification using amplified single-molecule detection. Nat Methods *3*, 725–727.

[Johnson et al., 2012] Johnson, L. R., Ghishan, F. K., Kaunitz, J. D., Merchant, J. L., Said, H. M. and Wood, J. D., eds (2012). Physiology of the Gastrointestinal Tract, 5th edition. Academic Press.

[Kani et al., 2013] Kani, K., Malihi, P. D., Jiang, Y., Wang, H., Wang, Y., Ruderman, D. L., Agus, D. B., Mallick, P. and Gross, M. E. (2013). Anterior gradient 2 (AGR2): Blood-based biomarker elevated in metastatic prostate cancer associated with the neuroendocrine phenotype. Prostate *73*, 306–315.

[Kennett, 1979] Kennett, R. G. (1979). Monoclonal antibodies. Hybrid myelomas–a revolution in serology and immunogenetics. Am J Hum Genet *31*, 539–547.

[Kim et al., 2011] Kim, J. H., Lee, S.-R., Li, L.-H., Park, H.-J., Park, J.-H., Lee, K. Y., Kim, M.-K., Shin, B. A. and Choi, S.-Y. (2011). High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS One *6*, e18556.

[Kim et al., 2007] Kim, N. S., Shen, Y. N., Kim, T.-Y., Byun, S. J., Jeon, I. S. and Kim, S. H. (2007). Expression of AGR-2 in chicken oviduct during laying period. J Biochem Mol Biol *40*, 212–217.

[Klein et al., 1988] Klein, G., Langegger, M., Goridis, C. and Ekblom, P. (1988). Neural cell adhesion molecules during embryonic induction and development of the kidney. Development *102*, 749–761.

[Kobayashi et al., 2008] Kobayashi, A., Valerius, M. T., Mugford, J. W., Carroll, T. J., Self, M., Oliver, G. and McMahon, A. P. (2008). Six2 defines and regulates a multipotent self-renewing nephron progenitor population throughout mammalian kidney development. Cell Stem Cell *3*, 169–181.

[Kohler and Milstein, 1975] Kohler, G. and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. Nature *256*, 495–497.

[Komiya et al., 1999] Komiya, T., Tanigawa, Y. and Hirohashi, S. (1999). Cloning of the gene gob-4, which is expressed in intestinal goblet cells in mice. Biochim Biophys Acta *1444*, 434–438.

[Kovalev et al., 2006] Kovalev, L. I., Shishkin, S. S., Khasigov, P. Z., Dzeranov, N. K., Kazachenko, A. V., Toropygin, I. I. and Mamykina, S. V. (2006). Identification of AGR2 protein, a novel potential cancer marker, using proteomics technologies. Prikl Biokhim Mikrobiol *42*, 480–484.

[Kretzschmar and Watt, 2012] Kretzschmar, K. and Watt, F. M. (2012). Lineage tracing. Cell *148*, 33–45.

[Kristiansen et al., 2005] Kristiansen, G., Pilarsky, C., Wissmann, C., Kaiser, S., Bruemmendorf, T., Roepcke, S., Dahl, E., Hinzmann, B., Specht, T., Pervan, J., Stephan, C., Loening, S., Dietel, M. and Rosenthal, A. (2005). Expression profiling of microdissected matched prostate cancer samples reveals CD166/MEMD and CD24 as new prognostic markers for patient survival. J Pathol *205*, 359–376.

[Kumar et al., 2007] Kumar, A., Godwin, J. W., Gates, P. B., Garza-Garcia, A. A. and Brockes, J. P. (2007). Molecular basis for the nerve dependence of limb regeneration in an adult vertebrate. Science *318*, 772–777.

[Lacroix and Leclercq, 2004] Lacroix, M. and Leclercq, G. (2004). Relevance of breast cancer cell lines as models for breast tumours: an update. Breast Cancer Res Treat *83*, 249–289.

[Lakso et al., 1992] Lakso, M., Sauer, B., Mosinger, B., Lee, E. J., Manning, R. W., Yu, S. H., Mulder, K. L. and Westphal, H. (1992). Targeted oncogene activation by site-specific recombination in transgenic mice. Proc Natl Acad Sci U S A *89*, 6232–6236.

[Lander et al., 2001] Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R.,

Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., BlŽcker, H., Hornischer,

K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., Szustakowki, J. and Consortium, I. H. G. S. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

[Lee et al., 2011] Lee, D. H., Lee, Y., Ryu, J., Park, S. G., Cho, S., Lee, J.-J., Choi, C. and Park, B. C. (2011). Identification of proteins differentially expressed in gastric cancer cells with high metastatic potential for invasion to lymph nodes. Mol Cells *31*, 563–571.

[Leong and Leong, 2007] Leong, T. Y.-M. and Leong, A. S.-Y. (2007). How does antigen retrieval work? Adv Anat Pathol *14*, 129–131.

[Lettice et al., 2002] Lettice, L. A., Horikoshi, T., Heaney, S. J. H., van Baren, M. J., van der Linde, H. C., Breedveld, G. J., Joosse, M., Akarsu, N.,

Oostra, B. A., Endo, N., Shibata, M., Suzuki, M., Takahashi, E., Shinka, T., Nakahori, Y., Ayusawa, D., Nakabayashi, K., Scherer, S. W., Heutink, P., Hill, R. E. and Noji, S. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc Natl Acad Sci U S A *99*, 7548–7553.

[Liscovitch and Ravid, 2007] Liscovitch, M. and Ravid, D. (2007). A case study in misidentification of cancer cell lines: MCF-7/AdrR cells (re-designated NCI/ADR-RES) are derived from OVCAR-8 human ovarian carcinoma cells. Cancer Lett *245*, 350–352.

[Liu et al., 2005] Liu, D., Rudland, P. S., Sibson, D. R., Platt-Higgins, A. and Barraclough, R. (2005). Human homologue of cement gland protein, a novel metastasis inducer associated with breast carcinomas. Cancer Res *65*, 3796–3805.

[Loder et al., 2013] Loder, M. K., Tsuboi, T. and Rutter, G. A. (2013). Live-cell imaging of vesicle trafficking and divalent metal ions by total internal reflection fluorescence (TIRF) microscopy. Methods Mol Biol *950*, 13–26.

[Lodish et al., 2000] Lodish, H., Berk, A., Zipursky, S. and et al. (2000.). Molecular Cell Biology. 4th edition. New York: W. H. Freeman;.

[Loven et al., 2012] Loven, J., Orlando, D. A., Sigova, A. A., Lin, C. Y., Rahl, P. B., Burge, C. B., Levens, D. L., Lee, T. I. and Young, R. A. (2012). Revisiting global gene expression analysis. Cell *151*, 476–482.

[Luche et al., 2007] Luche, H., Weber, O., Rao, T. N., Blum, C. and Fehling, H. J. (2007). Faithful activation of an extra-bright red fluorescent protein in

"knock-in" Cre-reporter mice ideally suited for lineage tracing studies. Eur J Immunol *37*, 43–53.

[Luck and Trave, 2011] Luck, K. and Trave, G. (2011). Phage display can select over-hydrophobic sequences that may impair prediction of natural domain-peptide interactions. Bioinformatics *27*, 899–902.

[Luo, 2007] Luo, L. (2007). Fly MARCM and mouse MADM: genetic methods of labeling and manipulating single neurons. Brain Res Rev *55*, 220–227.

[Makawita et al., 2011] Makawita, S., Smith, C., Batruch, I., Zheng, Y., R§ckert, F., Gr§tzmann, R., Pilarsky, C., Gallinger, S. and Diamandis, E. P. (2011). Integrated proteomic profiling of cell line conditioned media and pancreatic juice for the identification of pancreatic cancer biomarkers. Mol Cell Proteomics *10*, M111.008599.

[Marchalonis and Edelman, 1965] Marchalonis, J. and Edelman, G. M. (1965). Phylogenetic origins of antibody structure. I. Multichain structure of immunoglobulins in the smooth dogfish (Mustelus canis). J Exp Med *122*, 601–618.

[Maresh et al., 2010] Maresh, E. L., Mah, V., Alavi, M., Horvath, S., Bagryanova, L., Liebeskind, E. S., Knutzen, L. A., Zhou, Y., Chia, D., Liu, A. Y. and Goodglick, L. (2010). Differential expression of anterior gradient gene AGR2 in prostate cancer. BMC Cancer *10*, 680.

[Marshall, 2008] Marshall, W. F. (2008). Basal bodies platforms for building cilia. Curr Top Dev Biol *85*, 1–22.

[Maslon et al., 2010] Maslon, M. M., Hrstka, R., Vojtesek, B. and Hupp, T. R. (2010). A divergent substrate-binding loop within the pro-oncogenic protein anterior gradient-2 forms a docking site for Reptin. J Mol Biol *404*, 418–438.

[Matsuoka et al., 2007] Matsuoka, S., Ballif, B. A., Smogorzewska, A., McDonald, E. R., Hurov, K. E., Luo, J., Bakalarski, C. E., Zhao, Z., Solimini, N., Lerenthal, Y., Shiloh, Y., Gygi, S. P. and Elledge, S. J. (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science *316*, 1160–1166.

[Mattheyses et al., 2010] Mattheyses, A. L., Simon, S. M. and Rappoport, J. Z. (2010). Imaging with total internal reflection fluorescence microscopy for the cell biologist. J Cell Sci *123*, 3621–3628.

[McCafferty et al., 1990] McCafferty, J., Griffiths, A. D., Winter, G. and Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. Nature *348*, 552–554.

[McDermott et al., 2010] McDermott, K. M., Liu, B. Y., Tlsty, T. D. and Pazour, G. J. (2010). Primary cilia regulate branching morphogenesis during mammary gland development. Curr Biol *20*, 731–737.

[McManus and Graveley, 2011] McManus, C. J. and Graveley, B. R. (2011). RNA structure and the mechanisms of alternative splicing. Curr Opin Genet Dev *21*, 373–379.

[Metsuyanim et al., 2009] Metsuyanim, S., Harari-Steinberg, O., Buzhor, E., Omer, D., Pode-Shakked, N., Ben-Hur, H., Halperin, R., Schneider, D. and

Dekel, B. (2009). Expression of stem cell markers in the human fetal kidney. PLoS One *4*, e6709.

[Mirna Saraga-Babic et al., 2012] Mirna Saraga-Babic, M., Katarina Vukojevic, K., Bocina, I., Drnasin, K. and Saraga, M. (2012). Ciliogenesis in normal human kidney development and post-natal life. Pediatr Nephrol *27*, 55–63.

[Morrison et al., 2006] Morrison, R. N., Cooper, G. A., Koop, B. F., Rise, M. L., Bridle, A. R., Adams, M. B. and Nowak, B. F. (2006). Transcriptome profiling the gills of amoebic gill disease (AGD)-affected Atlantic salmon (Salmo salar L.): a role for tumor suppressor p53 in AGD pathogenesis? Physiol Genomics *26*, 15–34.

[Mugford et al., 2009] Mugford, J. W., Yu, J., Kobayashi, A. and McMahon, A. P. (2009). High-resolution gene expression analysis of the developing mouse kidney defines novel cellular compartments within the nephron progenitor population. Dev Biol *333*, 312–323.

[Munro and Pelham, 1987] Munro, S. and Pelham, H. R. (1987). A C-terminal signal prevents secretion of luminal ER proteins. Cell *48*, 899–907.

[Murray et al., 2007] Murray, E., McKenna, E. O., Burch, L. R., Dillon, J., Langridge-Smith, P., Kolch, W., Pitt, A. and Hupp, T. R. (2007). Microarray-formatted clinical biomarker assay development using peptide aptamers to anterior gradient-2. Biochemistry *46*, 13742–13751.

[Nagy, 2000] Nagy, A. (2000). Cre recombinase: the universal reagent for genome tailoring. Genesis *26*, 99–109.

[Neduva et al., 2005] Neduva, V., Linding, R., Su-Angrand, I., Stark, A., de Masi, F., Gibson, T. J., Lewis, J., Serrano, L. and Russell, R. B. (2005). Systematic discovery of new recognition peptides mediating protein interaction networks. PLoS Biol *3*, e405.

[Neduva and Russell, 2005] Neduva, V. and Russell, R. B. (2005). Linear motifs: evolutionary interaction switches. FEBS Lett *579*, 3342–3345.

[Neduva and Russell, 2006] Neduva, V. and Russell, R. B. (2006). Peptides mediating interaction networks: new leads at last. Curr Opin Biotechnol *17*, 465–471.

[Neduva and Russell, 2007] Neduva, V. and Russell, R. B. (2007). Proline-rich regions in transcriptional complexes: heading in many directions. Sci STKE *2007*, pe1.

[Niederreiter and Kaser, 2011] Niederreiter, L. and Kaser, A. (2011). Endoplasmic reticulum stress and inflammatory bowel disease. Acta Gastroenterol Belg *74*, 330–333.

[Nilsson and Warren, 1994] Nilsson, T. and Warren, G. (1994). Retention and retrieval in the endoplasmic reticulum and the Golgi apparatus. Curr Opin Cell Biol *6*, 517–521.

[Noren and Noren, 2001] Noren, K. A. and Noren, C. J. (2001). Construction of high-complexity combinatorial phage display peptide libraries. Methods *23*, 169–178.

[Norris et al., 2012] Norris, A. M., Gore, A., Balboni, A., Young, A., Longnecker, D. S. and Korc, M. (2012). AGR2 is a SMAD4-suppressible gene

that modulates MUC1 levels and promotes the initiation and progression of pancreatic intraepithelial neoplasia. Oncogene *5*, 1–5.

[Park et al., 2011] Park, K., Chung, Y. J., So, H., Kim, K., Park, J., Oh, M., Jo, M., Choi, K., Lee, E.-J., Choi, Y.-L., Song, S. Y., Bae, D.-S., Kim, B.-G. and Lee, J.-H. (2011). AGR2, a mucinous ovarian cancer marker, promotes cell proliferation and migration. Exp Mol Med *43*, 91–100.

[Park et al., 2009] Park, S.-W., Zhen, G., Verhaeghe, C., Nakagami, Y., Nguyenvu, L. T., Barczak, A. J., Killeen, N. and Erle, D. J. (2009). The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. Proc Natl Acad Sci U S A *106*, 6950–6955.

[Patel et al., 2013] Patel, P., Clarke, C., Barraclough, D. L., Jowitt, T. A., Rudland, P. S., Barraclough, R. and Lian, L.-Y. (2013). Metastasis-promoting anterior gradient 2 protein has a dimeric thioredoxin fold structure and a role in cell adhesion. J Mol Biol *425*, 929–943.

[Pelletier and Sonenberg, 1988] Pelletier, J. and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature *334*, 320–325.

[Pennisi, 2003] Pennisi, E. (2003). Human genome. A low number wins the GeneSweep Pool. Science *300*, 1484.

[Persson et al., 2005] Persson, S., Rosenquist, M., Knoblach, B., Khosravi-Far, R., Sommarin, M. and Michalak, M. (2005). Diversity of the protein disulfide isomerase family: identification of breast tumor induced Hag2 and

Hag3 as novel members of the protein family. Mol Phylogenet Evol *36*, 734–740.

[Petek et al., 2000] Petek, E., Windpassinger, C., Egger, H., Kroisel, P. M. and Wagner, K. (2000). Localization of the human anterior gradient-2 gene (AGR2) to chromosome band 7p21.3 by radiation hybrid mapping and fluorescencein situ hybridisation. Cytogenet Cell Genet *89*, 141–142.

[Petersen et al., 2011] Petersen, T., Brunak, S., von Heijne, G. and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods *8*, 785–786.

[Pettersson et al., 2009] Pettersson, E., Lundeberg, J. and Ahmadian, A. (2009). Generations of sequencing technologies. Genomics *93*, 105–111.

[Pizzi et al., 2012] Pizzi, M., Fassan, M., Balistreri, M., Galligioni, A., Rea, F. and Rugge, M. (2012). Anterior gradient 2 overexpression in lung adenocarcinoma. Appl Immunohistochem Mol Morphol *20*, 31–36.

[Plotnikova et al., 2009] Plotnikova, O. V., Pugacheva, E. N. and Golemis, E. A. (2009). Primary cilia and the cell cycle. Methods Cell Biol *94*, 137–160.

[Pode-Shakked et al., 2013] Pode-Shakked, N., Shukrun, R., Mark-Danieli, M., Tsvetkov, P., Bahar, S., Pri-Chen, S., Goldstein, R. S., Rom-Gross, E., Mor, Y., Fridman, E., Meir, K., Simon, A., Magister, M., Kaminski, N., Goldmacher, V. S., Harari-Steinberg, O. and Dekel, B. (2013). The isolation and characterization of renal cancer initiating cells from human Wilms'

tumour xenografts unveils new therapeutic targets. EMBO Mol Med *5*, 18–37.

[Pohler et al., 2004] Pohler, E., Craig, A. L., Cotton, J., Lawrie, L., Dillon, J. F., Ross, P., Kernohan, N. and Hupp, T. R. (2004). The Barrett's antigen anterior gradient-2 silences the p53 transcriptional response to DNA damage. Mol Cell Proteomics *3*, 534–547.

[Poser et al., 2008] Poser, I., Sarov, M., Hutchins, J. R. A., HŐrichŐ, J.-K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A. W., Pelletier, L., Kittler, R., Hua, S., Naumann, R., Augsburg, M., Sykora, M. M., Hofemeister, H., Zhang, Y., Nasmyth, K., White, K. P., Dietzel, S., Mechtler, K., Durbin, R., Stewart, A. F., Peters, J.-M., Buchholz, F. and Hyman, A. A. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat Methods *5*, 409–415.

[Praetorius and Spring, 2001] Praetorius, H. A. and Spring, K. R. (2001). Bending the MDCK cell primary cilium increases intracellular calcium. J Membr Biol *184*, 71–79.

[Prozialeck et al., 2004] Prozialeck, W. C., Lamar, P. C. and Appelt, D. M. (2004). Differential expression of E-cadherin, N-cadherin and beta-catenin in proximal and distal segments of the rat nephron. BMC Physiol *4*, 10.

[Ramachandran et al., 2008] Ramachandran, V., Arumugam, T., Wang, H. and Logsdon, C. D. (2008). Anterior gradient 2 is expressed and secreted during the development of pancreatic cancer and promotes cancer cell survival. Cancer Res *68*, 7811–7818.

[Rauchman et al., 1993] Rauchman, M. I., Nigam, S. K., Delpire, E. and Gullans, S. R. (1993). An osmotically tolerant inner medullary collecting duct cell line from an SV40 transgenic mouse. Am J Physiol *265*, F416–F424.

[Raykhel et al., 2007] Raykhel, I., Alanen, H., Salo, K., Jurvansuu, J., Nguyen, V. D., Latva-Ranta, M. and Ruddock, L. (2007). A molecular specificity code for the three mammalian KDEL receptors. J Cell Biol *179*, 1193–1204.

[Recupero et al., 2013] Recupero, D., Daniele, L., MarchiŸ, C., Molinaro, L., Castellano, I., Cassoni, P., Righi, A., Montemurro, F., Sismondi, P., Biglia, N., Viale, G., Risio, M. and Sapino, A. (2013). Spontaneous and pronase-induced HER2 truncation increases the trastuzumab binding capacity of breast cancer tissues and cell lines. J Pathol *229*, 390–399.

[Rodewald and Karnovsky, 1974] Rodewald, R. and Karnovsky, M. J. (1974). Porous substructure of the glomerular slit diaphragm in the rat and mouse. J Cell Biol *60*, 423–433.

[Rossenu et al., 1997] Rossenu, S., Dewitte, D., Vandekerckhove, J. and Ampe, C. (1997). A phage display technique for a fast, sensitive, and systematic investigation of protein-protein interactions. J Protein Chem *16*, 499–503.

[Rowe et al., 2009] Rowe, M. L., Ruddock, L. W., Kelly, G., Schmidt, J. M., Williamson, R. A. and Howard, M. J. (2009). Solution structure and dynamics of ERp18, a small endoplasmic reticulum resident oxidoreductase . Biochemistry *48*, 4596–4606.

[Rual et al., 2005] Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P. and Vidal, M. (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173–1178.

[Rump et al., 2004] Rump, A., Morikawa, Y., Tanaka, M., Minami, S., Umesaki, N., Takeuchi, M. and Miyajima, A. (2004). Binding of ovarian cancer antigen CA125/MUC16 to mesothelin mediates cell adhesion. J Biol Chem *279*, 9190–9198.

[Ryu et al., 2012] Ryu, J., Park, S. G., Lee, P. Y., Cho, S., Lee, D. H., Kim, G. H., Kim, J.-H. and Park, B. C. (2012). Dimerization of pro-oncogenic protein Anterior Gradient 2 is required for the interaction with BiP/GRP78. Biochem Biophys Res Commun *430*, 610–615.

[Sasaki et al., 2004] Sasaki, Y., Sone, T., Yoshida, S., Yahata, K., Hotta, J., Chesnut, J. D., Honda, T. and Imamoto, F. (2004). Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the Multisite Gateway system. J Biotechnol *107*, 233–243.

[Satir and Christensen, 2008] Satir, P. and Christensen, S. T. (2008). Structure and function of mammalian cilia. Histochem Cell Biol *129*, 687–693.

[Satsangi et al., 1996] Satsangi, J., Parkes, M., Louis, E., Hashimoto, L., Kato, N., Welsh, K., Terwilliger, J. D., Lathrop, G. M., Bell, J. I. and Jewell, D. P. (1996). Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. Nat Genet *14*, 199–202.

[Saxen and Lehtonen, 1987] Saxen, L. and Lehtonen, E. (1987). Embryonic kidney in organ culture. Differentiation *36*, 2–11.

[Schroeder et al., 2012] Schroeder, B. W., Verhaeghe, C., Park, S.-W., Nguyenvu, L. T., Huang, X., Zhen, G. and Erle, D. J. (2012). AGR2 is induced in asthma and promotes allergen-induced mucin overproduction. Am J Respir Cell Mol Biol *47*, 178–185.

[Schwarz et al., 2008] Schwarz, D., DŽtsch, V. and Bernhard, F. (2008). Production of membrane proteins using cell-free expression systems. Proteomics *8*, 3933–3946.

[Sedmak and Wolfrum, 2010] Sedmak, T. and Wolfrum, U. (2010). Intraflagellar transport molecules in ciliary and nonciliary cells of the retina. J Cell Biol *189*, 171–186.

[Self et al., 2006] Self, M., Lagutin, O. V., Bowling, B., Hendrix, J., Cai, Y., Dressler, G. R. and Oliver, G. (2006). Six2 is required for suppression of nephrogenesis and progenitor renewal in the developing kidney. EMBO J *25*, 5214–5228.

[Shcherbo et al., 2007] Shcherbo, D., Merzlyak, E. M., Chepurnykh, T. V., Fradkov, A. F., Ermakova, G. V., Solovieva, E. A., Lukyanov, K. A., Bog-

danova, E. A., Zaraisky, A. G., Lukyanov, S. and Chudakov, D. M. (2007). Bright far-red fluorescent protein for whole-body imaging. Nat Methods *4*, 741–746.

[Shih et al., 2007] Shih, L.-J., Lu, Y.-F., Chen, Y.-H., Lin, C.-C., Chen, J.-A. and Hwang, S.-P. L. (2007). Characterization of the agr2 gene, a homologue of X. laevis anterior gradient 2, from the zebrafish, Danio rerio. Gene Expr Patterns *7*, 452–460.

[Skarnes et al., 2011] Skarnes, W. C., Rosen, B., West, A. P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A. O., Thomas, M., Harrow, J., Cox, T., Jackson, D., Severin, J., Biggs, P., Fu, J., Nefedov, M., de Jong, P. J., Stewart, A. F. and Bradley, A. (2011). A conditional knockout resource for the genome-wide study of mouse gene function. Nature *474*, 337–342.

[Smith, 1985] Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science *228*, 1315–1317.

[Smith and Petrenko, 1997] Smith, G. P. and Petrenko, V. A. (1997). Phage Display. Chem Rev *97*, 391–410.

[Smith et al., 2006] Smith, U. M., Consugar, M., Tee, L. J., McKee, B. M., Maina, E. N., Whelan, S., Morgan, N. V., Goranson, E., Gissen, P., Lilliquist, S., Aligianis, I. A., Ward, C. J., Pasha, S., Punyashthiti, R., Sharif, S. M., Batman, P. A., Bennett, C. P., Woods, C. G., McKeown, C., Bucourt, M., Miller, C. A., Cox, P., Algazali, L., Trembath, R. C., Torres, V. E., Attie-Bitach, T., Kelly, D. A., Maher, E. R., Gattone, V. H., Harris, P. C. and Johnson, C. A. (2006). The transmembrane protein meckelin

(MKS3) is mutated in Meckel-Gruber syndrome and the wpk rat. Nat Genet *38*, 191–196.

[Smothers and Henikoff, 2001] Smothers, J. F. and Henikoff, S. (2001). Predicting in vivo protein peptide interactions with random phage display. Comb Chem High Throughput Screen *4*, 585–591.

[Snippert et al., 2010] Snippert, H. J., van der Flier, L. G., Sato, T., van Es, J. H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A. M., van Rheenen, J., Simons, B. D. and Clevers, H. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell *143*, 134–144.

[Soriano, 1999] Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet *21*, 70–71.

[Sossey-Alaoui et al., 2003] Sossey-Alaoui, K., Vieira, L., David, D., Boavida, M. G. and Cowell, J. K. (2003). Molecular characterization of a 7p15-21 homozygous deletion in a Wilms tumor. Genes Chromosomes Cancer *36*, 1–6.

[Soule et al., 1973] Soule, H. D., Vazguez, J., Long, A., Albert, S. and Brennan, M. (1973). A human cell line from a pleural effusion derived from a breast carcinoma. J Natl Cancer Inst *51*, 1409–1416.

[Srinivas et al., 1999] Srinivas, S., Goldberg, M. R., Watanabe, T., D'Agati, V., al Awqati, Q. and Costantini, F. (1999). Expression of green fluorescent protein in the ureteric bud of transgenic mice: a new tool for the analysis of ureteric bud morphogenesis. Dev Genet *24*, 241–251.

[Sternberg and Hamilton, 1981] Sternberg, N. and Hamilton, D. (1981). Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. J Mol Biol *150*, 467–486.

[Stocum, 2007] Stocum, D. L. (2007). Developmental biology. Acceptable nAGging. Science *318*, 754–755.

[Strausberg et al., 2004] Strausberg, R. L., Simpson, A. J. G., Old, L. J. and Riggins, G. J. (2004). Oncogenomics and the development of new cancer therapies. Nature *429*, 469–474.

[Streppel et al., 2012] Streppel, M. M., Vincent, A., Mukherjee, R., Campbell, N. R., Chen, S.-H., Konstantopoulos, K., Goggins, M. G., Seuningen, I. V., Maitra, A. and Montgomery, E. A. (2012). Mucin 16 (cancer antigen 125) expression in human tissues and cell lines and correlation with clinical outcome in adenocarcinomas of the pancreas, esophagus, stomach, and colon. Hum Pathol *43*, 1755–1763.

[Subramanian et al., 2005] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545–15550.

[Sutton et al., 1998] Sutton, R. B., Fasshauer, D., Jahn, R. and Brunger, A. T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. Nature *395*, 347–353.

[Szymczak et al., 2004] Szymczak, A. L., Workman, C. J., Wang, Y., Vignali, K. M., Dilioglou, S., Vanin, E. F. and Vignali, D. A. A. (2004). Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat Biotechnol *22*, 589–594.

[Tang et al., 2009] Tang, W., Ehrlich, I., Wolff, S. B. E., Michalski, A.-M., WŽlfl, S., Hasan, M. T., L§thi, A. and Sprengel, R. (2009). Faithful expression of multiple proteins via 2A-peptide self-processing: a versatile and reliable method for manipulating brain circuits. J Neurosci *29*, 8621–8629.

[Teasdale and Jackson, 1996] Teasdale, R. D. and Jackson, M. R. (1996). Signal-mediated sorting of membrane proteins between the endoplasmic reticulum and the golgi apparatus. Annu Rev Cell Dev Biol *12*, 27–54.

[Teng et al., 2012] Teng, S., Luo, H. and Wang, L. (2012). Predicting protein sumoylation sites from sequence features. Amino Acids *43*, 447–455.

[Thompson and Weigel, 1998] Thompson, D. A. and Weigel, R. J. (1998). hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines. Biochem Biophys Res Commun *251*, 111–116.

[Tonegawa, 1983] Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575–581.

[Trapp et al., 2008] Trapp, M. L., Galtseva, A., Manning, D. K., Beier, D. R., Rosenblum, N. D. and Quarmby, L. M. (2008). Defects in ciliary localization of Nek8 is associated with cystogenesis. Pediatr Nephrol *23*, 377–387.

[Trichas et al., 2008] Trichas, G., Begbie, J. and Srinivas, S. (2008). Use of the viral 2A peptide for bicistronic expression in transgenic mice. BMC Biol *6*, 40.

[Valentin et al., 2009] Valentin, E. D., Crahay, C., Garbacki, N., Hennuy, B., GuŐders, M., NoŚl, A., Foidart, J.-M., Grooten, J., Colige, A., Piette, J. and Cataldo, D. (2009). New asthma biomarkers: lessons from murine models of acute and chronic asthma. Am J Physiol Lung Cell Mol Physiol *296*, L185–L197.

[Valladares-Ayerbes et al., 2012] Valladares-Ayerbes, M., Blanco-Calvo, M., Reboredo, M., Lorenzo-PatiŰo, M. J., Iglesias-DŠaz, P., Haz, M., DŠaz-Prado, S., Medina, V., Santamarina, I., PŐrtega, S., Figueroa, A. and AntŮn-Aparicio, L. M. (2012). Evaluation of the Adenocarcinoma-Associated Gene AGR2 and the Intestinal Stem Cell Marker LGR5 as Biomarkers in Colorectal Cancer. Int J Mol Sci *13*, 4367–4387.

[Vanderlaag et al., 2010] Vanderlaag, K. E., Hudak, S., Bald, L., Fayadat-Dilman, L., Sathe, M., Grein, J. and Janatpour, M. J. (2010). Anterior gradient-2 plays a critical role in breast cancer cell growth and survival by modulating cyclin D1, estrogen receptor-alpha and survivin. Breast Cancer Res *12*, R32.

[Villalobos et al., 1995] Villalobos, M., Olea, N., Brotons, J. A., Olea-Serrano, M. F., de Almodovar, J. M. R. and Pedraza, V. (1995). The E-screen assay: a comparison of different MCF7 cell stocks. Environ Health Perspect *103*, 844–850.

[Vitale and Denecke, 1999] Vitale and Denecke (1999). The endoplasmic reticulum-gateway of the secretory pathway. Plant Cell *11*, 615–628.

[Vogelstein et al., 2000] Vogelstein, B., Lane, D. and Levine, A. J. (2000). Surfing the p53 network. Nature *408*, 307–310.

[Wang et al., 2007] Wang, W., Warren, M. and Bradley, A. (2007). Induced mitotic recombination of p53 in vivo. Proc Natl Acad Sci U S A *104*, 4501–4505.

[Wang et al., 2009] Wang, Z., Fan, Z.-C., Williamson, S. M. and Qin, H. (2009). Intraflagellar transport (IFT) protein IFT25 is a phosphoprotein component of IFT complex B and physically interacts with IFT27 in Chlamydomonas. PLoS One *4*, e5384.

[Wang et al., 2008] Wang, Z., Hao, Y. and Lowe, A. W. (2008). The adenocarcinoma-associated antigen, AGR2, promotes tumor growth, cell migration, and cellular transformation. Cancer Res *68*, 492–497.

[Warming et al., 2005] Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. and Copeland, N. G. (2005). Simple and highly efficient BAC recombineering using galK selection. Nucleic Acids Res *33*, e36.

[Watson and Crick, 1953a] Watson, J. D. and Crick, F. H. (1953a). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737–738.

[Watson and Crick, 1953b] Watson, J. D. and Crick, F. H. (1953b). Genetical implications of the structure of deoxyribonucleic acid. Nature *171*, 964–967.

[Wianny and Zernicka-Goetz, 2000] Wianny, F. and Zernicka-Goetz, M. (2000). Specific interference with gene function by double-stranded RNA in early mouse development. Nat Cell Biol *2*, 70–75.

[Williams et al., 2011] Williams, C. L., Li, C., Kida, K., Inglis, P. N., Mohan, S., Semenec, L., Bialas, N. J., Stupay, R. M., Chen, N., Blacque, O. E., Yoder, B. K. and Leroux, M. R. (2011). MKS and NPHP modules cooperate to establish basal body/transition zone membrane associations and ciliary gate function during ciliogenesis. J Cell Biol *192*, 1023–1041.

[Xia et al., 2009] Xia, J.-H., Jiang, J., Shi, Y.-H. and Gui, J.-F. (2009). Predominant expression and cellular distribution of fish Agr2 in renal collecting system. Comp Biochem Physiol B Biochem Mol Biol *152*, 397–404.

[Yamamoto et al., 2003] Yamamoto, K., Hamada, H., Shinkai, H., Kohno, Y., Koseki, H. and Aoe, T. (2003). The KDEL receptor modulates the endoplasmic reticulum stress response through mitogen-activated protein kinase signaling cascades. J Biol Chem *278*, 34525–34532.

[Yu et al., 2012] Yu, H., Zhao, J., Lin, L., Zhang, Y., Zhong, F., Liu, Y., Yu, Y., Shen, H., Han, M., He, F. and Yang, P. (2012). Proteomic study explores AGR2 as pro-metastatic protein in HCC. Mol Biosyst *8*, 2710–2718.

[Yuan et al., 2010] Yuan, K., Frolova, N., Xie, Y., Wang, D., Cook, L., Kwon, Y.-J., Steg, A. D., Serra, R. and Frost, A. R. (2010). Primary cilia are decreased in breast cancer: analysis of a collection of human breast cancer cell lines and tissues. J Histochem Cytochem *58*, 857–870.

[Zambrowicz et al., 1997] Zambrowicz, B. P., Imamoto, A., Fiering, S., Herzenberg, L. A., Kerr, W. G. and Soriano, P. (1997). Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. Proc Natl Acad Sci U S A *94*, 3789–3794.

[Zhang et al., 2005] Zhang, J.-S., Gong, A., Cheville, J. C., Smith, D. I. and Young, C. Y. F. (2005). AGR2, an androgen-inducible secretory protein overexpressed in prostate cancer. Genes Chromosomes Cancer *43*, 249–259.

[Zhang et al., 2004] Zhang, W., Morris, Q. D., Chang, R., Shai, O., Bakowski, M. A., Mitsakakis, N., Mohammad, N., Robinson, M. D., Zirngibl, R., Somogyi, E., Laurin, N., Eftekharpour, E., Sat, E., Grigull, J., Pan, Q., Peng, W.-T., Krogan, N., Greenblatt, J., Fehlings, M., van der Kooy, D., Aubin, J., Bruneau, B. G., Rossant, J., Blencowe, B. J., Frey, B. J. and Hughes, T. R. (2004). The functional landscape of mouse gene expression. J Biol *3*, 21.

[Zhang et al., 2007] Zhang, Y., Forootan, S. S., Liu, D., Barraclough, R., Foster, C. S., Rudland, P. S. and Ke, Y. (2007). Increased expression of anterior gradient-2 is significantly associated with poor survival of prostate cancer patients. Prostate Cancer Prostatic Dis *10*, 293–300.

[Zhao et al., 2010] Zhao, F., Edwards, R., Dizon, D., Afrasiabi, K., Mastroianni, J. R., Geyfman, M., Ouellette, A. J., Andersen, B. and Lipkin, S. M. (2010). Disruption of Paneth and goblet cell homeostasis and increased endoplasmic reticulum stress in Agr2-/- mice. Dev Biol *338*, 270–279.

[Zhao et al., 2009] Zhao, L., Lee, B. Y., Brown, D. A., Molloy, M. P., Marx, G. M., Pavlakis, N., Boyer, M. J., Stockler, M. R., Kaplan, W., Breit, S. N., Sutherland, R. L., Henshall, S. M. and Horvath, L. G. (2009). Identification of candidate biomarkers of therapeutic response to docetaxel by proteomic profiling. Cancer Res *69*, 7696–7703.

[Zheng et al., 2006] Zheng, W., Rosenstiel, P., Huse, K., Sina, C., Valentonyte, R., Mah, N., Zeitlmann, L., Grosse, J., Ruf, N., N§rnberg, P., Costello, C. M., Onnie, C., Mathew, C., Platzer, M., Schreiber, S. and Hampe, J. (2006). Evaluation of AGR2 and AGR3 as candidate genes for inflammatory bowel disease. Genes Immun *7*, 11–18.

[Zhu and Sadowski, 1995] Zhu, X. D. and Sadowski, P. D. (1995). Cleavage-dependent ligation by the FLP recombinase. Characterization of a mutant FLP protein with an alteration in a catalytic amino acid. J Biol Chem *270*, 23044–23054.

[Zong et al., 2005] Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. and Luo, L. (2005). Mosaic analysis with double markers in mice. Cell *121*, 479–492.