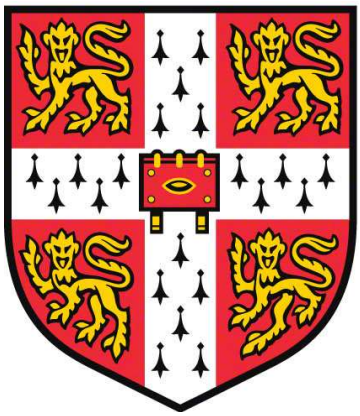
Specific forms of TCF7L2 protein define the Wnt pathway dependent genetic program



Manuela Urbischek

Hughes Hall
Department of Biochemistry
University of Cambridge

This thesis is submitted for the degree of Doctor of Philosophy

January 2021

Preface

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the outlined list of collaborations and contributions. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

It does not exceed the prescribed word limit of 60.000 words for the PhD Degree Committee.

Manuela Urbischek
January 2021

Specific forms of TCF7L2 protein define the Wnt pathway dependent genetic program

by Manuela Urbischek

Summary

The TCF7L2 transcription factor is a critical regulator of Wnt pathway-dependent gene expression in the intestinal epithelia. Mutations in *TCF7L2* have been previously reported in colorectal cancer, however their functional impact on tumour biology has not been investigated.

The present work has determined that expression of the short form of TCF7L2 protein is not a product of alternative splicing; instead, the protein is predominantly expressed as the long form of the protein that is processed to the short form of the protein. I have identified discrete sequence-specific requirements in TCF7L2 by the identification of point mutations that abrogate conversion to the short form. Therefore, TCF7L2 specific processing is the principal mechanism dictating the ratio of long and short forms of the protein and the Wnt pathway-dependant transcriptional program.

The consequence of altered expression ratios of long and short forms of TCF7L2 are biologically important, shown by a specific set of mutations in the *TCF7L2* gene that are selected in some colorectal cancers; the mutations lead to the expression of a truncated version of TCF7L2 that is functionally identical to the processed, short form of the protein. Bioinformatics analyses have identified unique, molecular biomarkers and clinical characteristics of TCF7L2 mutant colorectal cancers that include microsatellite instability, exclusivity of BRAFV600E mutations, better prognosis and differential gene expression. Thus, specific TCF7L2 mutations classify a unique and discrete set of colorectal cancers. I further showed the outcome of TCF7L2 mutations and the altered ratio of long and short versions of the TCF7L2 protein – alteration of the Wnt pathway gene expression program with specific consequences on cell proliferation, clonogenic growth, cell cycle progression and survival.

I also established experimental systems to study TCF7L2 expression levels using organoid models. The technology consisted of organoids cultured in media containing precise activities of essential growth factors, obviating issues endemic to the use of commercially sourced, impure growth factor preparations. I focused on the two essential organoid media growth factors, R-spondin1 and Gremlin-1, that were produced from bacterial expression and optimised for organoid culture.

I further investigated whether R-spondin1 could be further engineered to increase its potency in organoid growth media. My strategy was to introduce mutations that specifically increase its affinity for its cognate receptor LGR5. LGR5 expression is specific to stem cells whereas the closely related LGR4 receptor is expressed by more differentiated cell types in the

intestinal epithelia. My engineering was successful, producing a mutant version of R-spondin1 that displayed a 6-fold increased binding affinity for LGR5 and a significantly reduced ability to bind LGR4.

My work has defined a unique colon cancer subset defined by specific TCF7L2 mutations and developed a system for analysing the consequence of the mutations *in vitro*. Unravelling the recircuitry imposed by mutant TCF7L2 protein will reveal therapeutic vulnerabilities and pave the way for developing treatment options.

Publications relating to this work

Conference poster presentation: Jan. 2020 5th Early Detection Annual Symposium, CRUK Cambridge
“A novel functional variant of TCF7L2 stratifies a unique colon cancer subset”

Journal article
Manuela Urbischek, Helena Rannikmae, Thomas Foets, Katharina Ravn, Marko Hyvönen & Marc de la Roche. Organoid culture media formulated with growth factors of defined cellular activity. Scientific Reports 2019; 9 (6193)

Acknowledgement

First of all, I want to say more than thank you to my supervisor Marc de la Roche. You are a passionate visionary scientist whose enthusiasm was always infectious, inviting for critical and out of the box thinking. Encouraging independent working and an always open door for discussion and exchange of ideas has developed me into the young scientist I am now. But not only on scientific level did you support me, also through my journey becoming a mum and balancing work and motherhood. Thank you for always being accommodating and understanding. Thank you!

Thank you to Andrew Holding, who became a second supervisor to me, analysing data and teaching me how to use R, very patiently and always on call for any little advice I needed. I hope I can implement what I have learned from you in my scientific future.

I am also very grateful for the excellent scientific input from Gerald Evan and Trevor Littlewood. The exceptional scientific environment that was created through your research has massively impacted me on my way through my PhD. Special thank you to you Trevor for your and time and very much appreciated advice.

Marko Hyvönen, thank you for your answers to my biochemical questions as well as your collaboration. I enjoyed our critical discussions loaded with refreshing directness.
I also want to thank Maike de la Roche for her collaboration giving me insights into T-cell biology and for always supporting my projects. Danke Maike!

And of course, what would have been without my colleagues. Helena, thank you for being tidy and organised, it and myself would have been a mess without you. I always enjoyed our chats and laughs. Thomas, despite being a bit “unorganised’ on your lab bench (or mine), you are the heart of the lab and a great scientist! Also, thanks to all the students who came along through the years. It was always fun! I also want to express my gratitude to Siolian Ball, for excellent supervision in the writing period of my thesis.

Luca, Roderik, Cathy, Peter, Ben, Tania, Phil, Lucia, Linsey, Nicole, Daniel: you have created an amazing scientific atmosphere in the lab open for discussions and advice. Thanks to Deborah for keeping the lab running.

Like Percival the butterfly: I couldn’t have done it without my friends. Veri, thank you for all your nice distractions and Tom for always giving me uplifting advice.

But now to those people that have made it possible. My family, to who this thesis is dedicated to. First Oma, I could not have done anything without your help and support through all my life. Studying as well as moving abroad to become a scientist would not have been possible without you. I hope you are looking down proudly now. The same holds true for my family. Mama, you are the best always being there for me whatever life threw at me, Stefan, thank you for keeping me on the ground reminding me who I am and Steffi, thank you for encouraging me to pursue my career. You are truly a woman I look up to admiring your multitasking skills.

Adam, I am grateful for understanding when I had to invest my time in my work and for taking over in dad mode, when I needed it. You have been and are continuing to be my rock, simple as that. Sarah and Robert, thank you for being full time childcare for Laila. I couldn't have finished my thesis without your limitless support and care. Thank you for welcoming me into your hearts.

Table of Contents

List of figures

List of Tables

List of abbreviations

°C	degrees celcius
ADC	antibody drug conjugate
ALP	alkaline phosphatase
APC	adenomatous polyposis coli
BMP	Bone morphogenic protein
BMP	Bone morphogenic protein
BMPR	Bone morphogenic protein receptor
bp	base pair
BSA	bovine serum albumin
C-clamp	cysteine clamp
C-terminus	carboxy-terminus
CBC	crypt base columnar
CFTR	cystic fibrosis transmembrane conductance regulator
ChIP	Chromatin Immunoprecipitation
CIMP	CpG island methylator phenotype
CIN	Chromosomal instablilty
CK1α	casein kinase 1α
CMS	Consensus molecular subtype
CMV	*cytomegalovirus*
CMV	cytomegalovirus
CRC	colorectal cancer
CtBM	C-terminal binding motif
CtBP	C-terminal binding protein
DAPI	4′,6-Diamidin-2-phenylindol
DEFA5	human Defensin 5
DMEM	dulbecco's modified eagle medium
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleoside triphosphates
DsbC	Disulphide bond isomerase C
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ECL	enzymatic chemiluminescence
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor

EGFR	EGF receptor
FACS	fluorescence-activated cell sorting
FAP	Familial adenomatous polyposis
FCS	Fetal calf serum
FDR	false discovery rate
FGF	fibroblast growth factor
FP	Fluorescence Polarisation
g	acceleration of gravity
GAPDH	Glyceraldehyde-3-phopsphate dehydrogenase
GBS	Groucho (TLE/GRG) binding site
GDP	guanosine diphophate
GndHCl	guanidine hydrochloride
GO	Gene Ontology
Grb2	growth factor receptor bound protein 2
GSEA	Gene set enrichment analysis
GSK3β	glycogen synthase kinase 3β
GTP	guanosine triphosphate
h	hour
HAT	Histone acetyl transferase
HBP1	Histone binding protein 1
HDAC	Histone deacetylase
HMG	High mobility group
HRP	horseradish peroxidase
IBD	inflammatory bowel disease
IC50	inhibitory concentration
IEX	Ion exchange chromatography
IPTG	isopropyl-ß-D-thiogalactopyranosid
IRES	internal ribosome entry site
ISC	intestinal stem cells
ITR	inverted terminal repeat
kDa	kilodalton
L	Liter
Lgr5	leucine-rich repeat containing G protein-coupled receptor 5
LOCE	lowest observed cellular effect
LRP5/6	low-density lipoprotein related receptor related protein 5/6
LRR	Leucine rich repeat
M	molar

mA	milliampere
MBP	Maltose binding protein
MCR	mutational cluster region
min	minute
ml	milliliter
mP	millipolarisation
MSI	microsatellite instability
MSS	microsatellite stability
mTOR	Target of Rapamycine
MW	molecular weight
N-terminus	amino-terminus
NICD	Notch intercellular domain
nm	nanometer
nM	nanomolar
nmol	nanomole
OD	optical density
ORF	Open reading frame
PBS	Phosphate buffered saline
PCP	Planar cell polarity
PCR	polymerase chain reaction
PEI	Polyethylenimine
PIP2	phosphatidylinositol (3,4)-bisphosphate
PIP3	phosphatidylinositol (3,4,5)-trisphosphate
pmol	picomole
PPS	3-[3-(1,1-bisalkyloxyethyl)pyridin-1-yl]propane-1-sulfonate
PVDF	polyvinylidene Fluoride
Rnf43	Ring finger protein 43
Rpm	rounds per minute
RT	room temperature
RT-PCR	Reverse transcription PCR
rtTA	reverse tetracycline-controlled transactivator
SCNA	somatic copy number alteration
SDS-PAGE	sodiumdodecylsulfate polyacrylamide gel electrophoresis
sec	second
SEC	size exclusion chromatography
shRNA	small hairpin ribonucleic acid
siRNA	small interfering ribonucleic acid

SOS	son of sevenless
T2D	Type 2 diabetes
TBS-T	Tris buffered saline – Tween
TC	Tetra cysteine
TCF	T cell factor
TCGA	The Cancer Genome Atlas
TF	transcription factor
TGF-ß	transforming growth factor ß
TLE/GRG	Transducin like Enhancer of Split/Groucho-related gene
TRE	Tetracycline response element
Tris	Tris(hydroxymethyl)-aminomethan
Tween	polyoxyethylene sorbitan monolaurate
U	unit
UEA-1	Ulex europaeus agglutinin 1
V	Volt
WRE	Wnt response element
X-Ray	Roentgen radiation
β-TrCP	β-transducin repeat containing E3 ubiquitin protein ligase
βDC	β-catenin destruction complex
µg	microgram
µl	microliter
µm	micrometer

Contributions and collaborations

- **Figure 14:** TCF7L2 alternative splicing isoforms in human CRC cell lines
  Experiment and primer design as well as exon abundance analysis initially carried out by Marc de la Roche laboratory, then repeated, reproduced and figure created by me.

- **Figure 15:** TCF7L2 is processed post-translationally independent of the proteasome and depending on TCF7L2 specific intrinsic sequences
  TCF7L2 C-terminal mutants (1, 2 and 3) and TCF7 variants cloned and Western blot performed by Marc de la Roche laboratory (Figure 12E).

- **Figure 17B**: TCF7L2 mutations mutual exclusivity and co-mutation spectrum refine a novel subset of MSI colon cancer
  Bioinformatic analysis initially carried out by Dr. Andrew Holding (CRUK, Cambridge), then specified for specific clinicopathological features and visualisation by me.

- **Figure 19A:** CRC cases with truncated TCF7L2 show better prognosis and validate in vitro
  R Code initially written by Dr. Andrew Holding (CRUK, Cambridge), then adapted by me for visualisation.

- **Figure 22A:** Gene expression profiling and validation of differentially expressed genes in vitro
  Gene expression analysis performed by Dr. Andrew Holding (CRUK, Cambridge), volcano plot visualisation criteria specified by me.

- **Figure 26:** Bacterial expression and production of Gremlin 1
  Expression and purification of human Gremlin 1 performed by Katharina Ravn and provided by the Marco Hyvönen laboratory

- **Figure 27F:** Optimal concentrations of recombinant R-spondin 1 and Gremlin 1 for organoid growth.
  Organoid cultures immune-fluorescently stained and visualised by Marc de la Roche and Helena Rannikmäe.

- **Figure 28:** Growth media formulated with R-spondin 1 or Gremlin 1 can sustain growth of various types of organoids
  Organoid cultures maintained and photographed by Marc de la Roche, Helena Rannikmäe and Thomas Foets.

- **Figure 29:** Structural predictions of R-spondin 1 mutagenesis
  Structural predictions performed by Dr. Sony Malhotra (University of Cambridge, Dept. of Biochemistry)

- **Figure 35:** FACS analysis of mesenteric lymph node derived T-cells show a LGR5 positive CD8+ and CD4+ cell population specifically sorted by SPNH R-spondin 1
  Isolation, staining, FACS analysis and RT-PCR was performed by Louise O'Brien (Maike de la Roche laboratory, CRUK, Cambridge)

Chapter 1 Introduction

1.1 Colorectal cancer (CRC)

Colorectal cancer (CRC) is the fourth leading cause of cancer deaths worldwide. Development of the disease from adenomas to full-blown adenocarcinoma is slow (up to decades) and accompanying somatic changes fuelling progression have stratified three major genetic pathways of disease development. However, classification according to these genetic pathways did not always resemble treatment success, where a novel approach to CRC classification according to consensus molecular subtypes appeared to be more promising [1]

1.1.1 Genetic pathways of CRC development

I - Conventional pathway:

(i) The hereditary form of CRC, familiar adenomatous polyposis (FAP), is caused by germline nonsense APC mutations of one allele and occurs with a frequency of 1-2% of all CRC cases [2]. A characteristic feature of FAP is loss of heterozygosity (LOH) through somatic mutation in the second allele [3], that leads to the initiation of tumour development.

(ii) Sporadic nonsense mutations in APC are thought to be the earliest, if not principal driver of the disease, that lead to malignant transformation of the colon epithelia. Such truncating mutations are found in 60-70% of all CRC cases [4]. CRC development along the adenoma to carcinoma sequence is fuelled by a relatively common series of genetic events that most often include chromosomal instability, oncogenic mutations in KRAS, SMAD4 and mutational inactivation of the tumour suppressor p53.

II – the serrated pathway:

A less common pathway of CRC development is characterised by the appearance of a serrated pattern of tumour growth.

(i) Traditional serrated tumours harbour either oncogenic BRAF or KRAS mutations and can manifest the CpG island methylator phenotype (CIMP) resulting in hyper methylation of genomic regions rich in cytosines and guanines. The traditional serrated pathway has not been well characterised, but is classed as a separate molecular subtype [5] and is up to 15% penetrant amongst CRCs.

(ii) Sessile serrated tumours occur when BRAF is mutated and aberrant methylation is present. On the one hand, this leads to epigenetic silencing of the p16 locus, resulting in CIMP, but stable microsatellites. MSS CIMP+ serrated tumours are rare with around 6-8% incidence. On the other hand, epigenetic silencing of the mismatch repair gene MLH1 can arise. Silencing of MLH1, by either epigenetic modification or mutation, leads to frameshift mutations of genes due to impaired post-replication repair especially in short tandem (microsatellite) repeats, hence they are classed as highly microsatellite instable (MSI-H). Sessile serrated tumours are characterised by the methylation of CpG islands and transcriptional repression of critical tumour suppressor genes and occur in 9-12% of colorectal cancers [6-8].

III – the microsatellite instability pathway:

A fifth genetic pathway of CRC development is known as Lynch syndrome (also referred to as hereditary non-polyposis colorectal cancer; HNPCC) that is characterised by germline mutations of DNA mismatch repair genes (for instance MSH2 and MLH1) [9]. Lynch Syndrome, is present in 2-5% of CRC and is characterised by a high number of intestinal adenomas, frequent frameshift mutations and the absence of CIMP that distinguishes it from MSI sessile serrated tumours. (Figure 1).

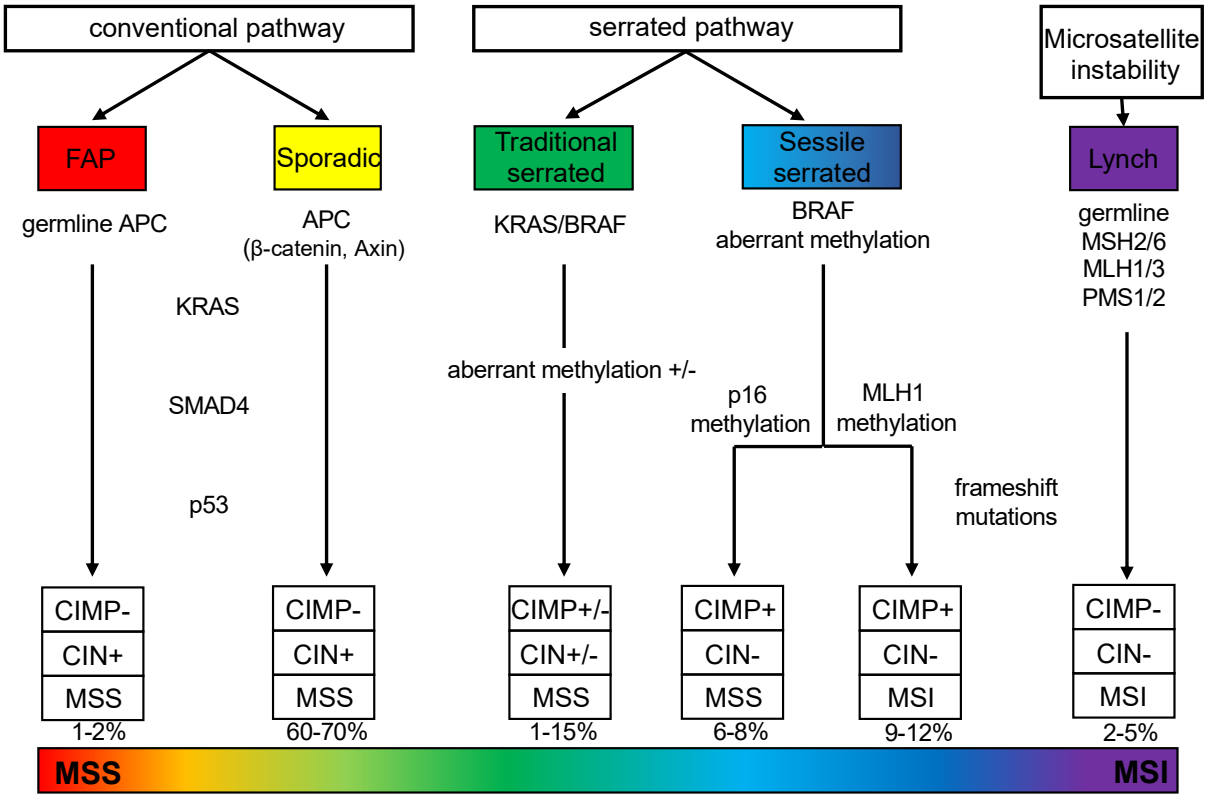


Figure 1. Spectrum of colorectal cancer development divided into different pathways.
I - Conventional pathway: the adenoma to carcinoma pathway follows the sequential accumulation of genetic mutations causing microsatellite stable (MSS) tumours with chromosomal instability (CIN) initiated mostly by APC mutation, hereditary in familial adenomatous polyposis (FAP) by germline mutation and sporadically in most CRC cases.
II - Serrated pathway can be divided in traditional serrated pathway and sessile serrated pathway resulting in microsatellite stable (MSS) or microsatellite instable (MSI) tumours, respectively, and can display CpG island methylator phenotype (CIMP), responsible for epigenetic silencing of tumour supressing genes.
III – Lynch syndrome is caused by germline mutation of MMR genes resulting in MSI without CIMP.

1.1.2 Consensus molecular subtype (CMS) classification of colorectal cancer

CMS classification is based on transcript expression data obtained from a large data sharing project. Combining genomic alterations investigated using different algorithms with clinical data, the CMS Consortium shows a novel und robust classification system dividing colorectal cancer into four molecular subtypes [1]. CMS1 (14%, microsatellite instability immune) is described as hypermutated, microsatellite unstable with immune activation and hallmarked by

BRAF mutations and CIMP positivity. CMS2 (37%, canonical) contains epithelial features and is marked by Wnt and myc signalling activation as well as a high amount of somatic copy number alterations (SCNA): this subtype is closest to the canonical pathway described in 1.2 with APC truncating mutations initialising tumorigenesis. CMS3 (13%, metabolic), is mostly microsatellite stable, low in SCNA, CIMP negative, KRAS activated and shows metabolic dysregulation. CMS4 (23%, mesenchymal) is high in SCNA and hallmarked through transforming growth factor–β (TGFβ) activation, increased angiogenesis, and stromal invasion. In 13 % of samples distinct features could not be applied (no classification) due to altering heterogeneity (Figure 2) [1].

CMS1	CMS2	CMS3	CMS4	no class
MSI immune	canonical	metabolic	mesenchymal	?
MSI hypermutation CIMP+	SCNA high	mostly MSS SCNA low CIMP-	SCNA high	?
BRAF mutations		KRAS mutations		?
immune infiltration and activation	Wnt and myc activation	metabolic misregulation	TGFß activation stromal infiltration angiogenesis	?

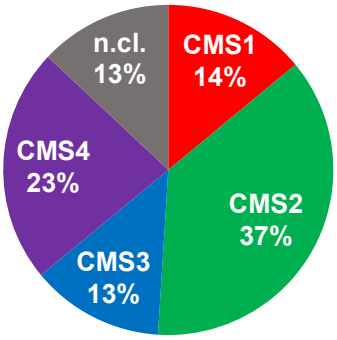


Figure 2. Consensus molecular subtype classification of CRC
Features and characteristics of each CMS group are listed. Proportion of cases grouped into respective CMS class depicted as pie chart with percentage. MSI=microsatellite instability, MSS= microsatellite stability, CIMP=CpG island methylator phenotype, SCNA=somatic copy number alteration. Figure adapted from [1].

1.2 signalling pathways in colorectal cancer

Colorectal cancer (CRC) development is due, primarily, to mutations in genes regulating two signalling pathways that control cell proliferation, 'stemness' and cell growth. In 80-90% of CRCs Wnt signalling pathway is affected and are considered the principle driver mutations of the disease [10], In 10-20% of CRC driving mutations occur within the EGFR/MEK/ERK signalling cascade. Deregulation of Notch signalling as well as PI3K signalling also contribute to carcinogenesis through alterations in cell differentiation, apoptosis and migration, however, these mutations are found later in tumour progression [11, 12].

1.2.1 Wnt signalling

The canonical Wnt signalling pathway is a well-established regulator of mammalian development [13] and controls the maintenance of a number of stem cell niches in adult tissues [14]. In the 'off' state, the key Wnt pathway regulator is a dedicated multi-protein destruction complex that constitutively targets the central effector of the pathway, β-catenin, to the proteosome. The key limiting components of this β-catenin destruction complex (βDC) are Axin and APC that interface β-catenin with the kinases GSK3β and CK1α. Phosphorylation of β-catenin at 4 residues in its N-terminus creates the so-called phospho-degron that functions as a receptor for the E3 ligase β-TrCP, leading to ubiquination and targeting to the proteosome. Binding of an afferent Wnt ligand to its cognate receptors, Frizzled and LRP, on the surface of a receptive cell activates the Wnt pathway by evoking an intracellular signalling cascade that culminates on the inhibition of the βDC. The mechanism for this inhibition remains to be fully resolved but relies on the direct interaction of Axin with Lrp5/6, the association of GSK3β, and the involvement of the Dishevelled proteins [15]. Attenuated phosphorylation stabilises β-catenin leading to its translocation into the nucleus and binding to the TCF/LEF family of transcription factors, predominantly TCF7L2, which is constitutively localized to promoters of Wnt pathway target genes [16]. The output of the Wnt signalling pathway is therefore a context-specific transcriptional program. Transcriptional targets not only include negative regulators such as Rnf43, a E3 Ubiqitin ligase, involved in Wnt receptor degradation [17, 18], but also a mechanism facilitating the increased abundance of Wnt receptors and therefore potentiating the Wnt signalling cascade. Binding of R-spondin to the Wnt target gene LGR5 and its LGR family members recruits Rnf43, impairing its negative function on Wnt receptors [19].

Colorectal cancer is the paradigm of deregulated Wnt pathway activity in disease. Greater than 80% CRCs harbour mutations in APC in a discrete 500 amino acid region of the protein known as the mutational cluster region (MCR). Invariably, the mutations lead to C-terminal truncations that remove its β-catenin and Axin interacting domains. The net result is inactivation of the βDC, accumulation of non-phosphorylated β-catenin and constitutive

activation of Wnt pathway target gene expression and malignant transformation of the intestinal epithelia. Less frequent (approximately 5-10% of CRC cases) are mutations of the phospho-degron phosphorylation sites that render β-catenin refractory to phosphorylation by the β-DC and proteasomal degradation [20]. Notably, APC and β-catenin mutations have been found to be mutually exclusive, distinctly categorising independent pathways of CRC development [21]. Less frequent are mutations in other Wnt pathway regulatory components are also found; for instance inactivating mutations in Axin [22] and RNF43 [23] (Figure 3).

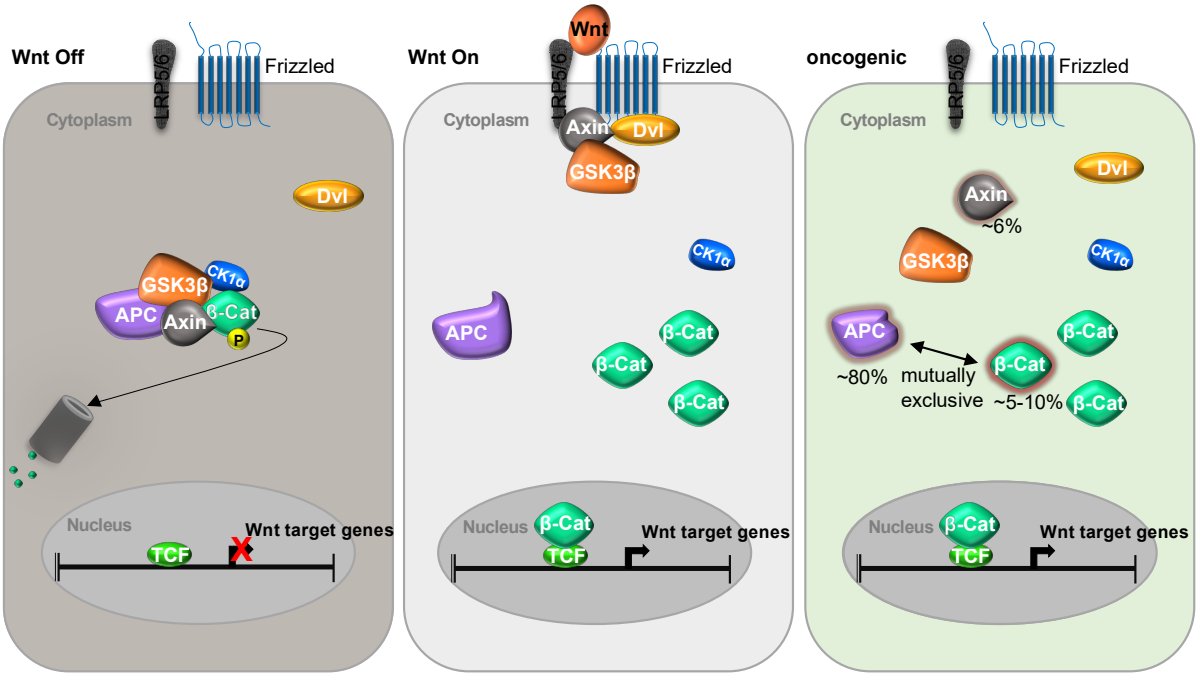


Figure 3: Canonical Wnt signaling – physiological normal and oncogenic
Left panel: In the absence of Wnt ligand, the key effector of the pathway, β-catenin, is phosphorylated by the β-catenin destruction complex (βDC) components CK1α and GSK3β resulting in its proteasomal degradation. *Middle panel:* Stimulation of the pathway when Wnt ligand binds to its receptor Frizzled results in βDC inhibition through the recruitment of Axin and GSK3β to the plasma membrane. Stabilisation of β-catenin leads to its translocation into the nucleus where it interacts with TCF/LEF family proteins, driving a transcriptional program. *Right panel:* truncating mutations of APC result in inactivation of the βDC which leads to constitutively elevated β-catenin levels and ligand independent expression of Wnt target genes, which is oncogenic. Less frequent and mutually exclusive to APC mutations of β-catenin itself leads to its accumulation and target gene transcription. In rare cases Axin harbours mutations compromising βDC function and oncogenic pathway output.

1.2.2 EGFR/MEK/ERK signalling pathway

De-regulated EGFR/MEK/ERK signalling is another frequent feature of CRC progression. The epidermal growth factor receptor EGFR is a receptor tyrosine kinase that dimerises when bound by ligands such as EGF. Autophosphorylation of the intracellular kinase domains creates a binding site for adaptor proteins growth factor receptor bound protein 2 (Grb2) and son of seven-less (SOS) [24]. Formation of the EGFR complex leads to activation of the small GTPase RAS (HRAS, NRAS, KRAS) via GDP to GTP exchange that binds to and activates RAF family (ARAF, BRAF, CRAF) kinases [25] that in turn phosphorylates and activates MEK which subsequently phosphorylates ERK1/2 protein. The phosphorylated ERK1/2 proteins drive transcription of genes regulating cell proliferation and survival (Figure 4) [26, 27].

The PI3K/Akt pathway is another arm of the EGFR signalling pathway. Class I PI3K is a heterodimer consisting of regulatory (p85) and catalytic (p110) subunits. Three genes, PIK3CA, PIK3CB and PIK3CD encode different isoforms of p110 subunits that convert phosphatidylinositol (3,4)-bisphosphate (PIP_2) lipids to phosphatidylinositol (3,4,5)-trisphosphate (PIP_3) [28]. The serine/threonine kinase Akt is activated by binding to PIP_3 at the plasma membrane. Akt phosphorylates and activates mTORC1 which in turn phosphorylates and activates the S6 Kinase. Subsequent phosphorylation of the S6 ribosomal protein promotes protein synthesis and cellular proliferation. This PI3K/Akt pathway is inhibited by the kinase activity of the PTEN tumour suppressor which converts PIP2 to PIP3 [29].

While EGFR mutations, amplifications, or copy number alterations have been reported as cancer drivers [30], oncogenic mutations in the KRAS proto-oncogene (G12D/C/V and G13D) are found in 40% of CRC cases [20]. In contrast, NRAS mutations are observed in 2.5% to 4.5% of CRCs [31]. Oncogenic BRAF mutations (V600E) are exclusive of KRAS mutations and are found in 8% of all CRCs. Commonly, oncogenic BRAF mutations feature in CRC tumours that display microsatellite instability (MSI) and high levels of CpG island methylation (the CpG island methylator phenotype; CIMP) [32] and are associated with poor prognosis [33].

Mutations leading to de-regulation of the PI3K/Akt pathway are often found in CRCs. Oncogenic activation of the pathway, either through mutations in PIK3CA and PIK3CB, or amplification of AKT, loss of the function of PTEN and mutational activation of mTORC1 have all been observed in CRC [34]. However, these alterations are thought to occur in later stages of CRC development [35-37](Figure 4).

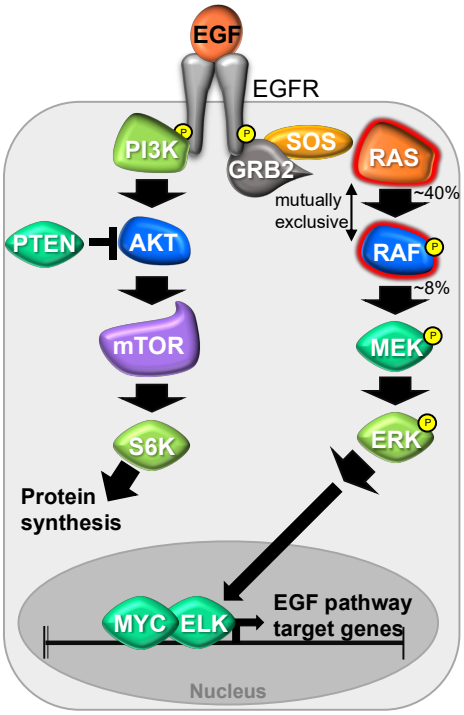


Figure 4: EGFR/MEK/ERK signalling and its alterations in CRC
See text above (1.2.2) for description of molecular signalling cascade. Sites of pathway mutations prevalent in CRC are underlaid in red.

1.3 Organoids as model systems for CRC

1.3.1 Organoids recapitulate i*n vivo* tissue organisation

The development of a robust system for culturing three-dimensional organ-like epithelia *in vitro* was first reported by the Clevers laboratory [38]. The basis of the technology was the correct combination and concentration of various growth factors in media formulations that sustain three-dimensional growth of murine intestinal epithelial tissue in a laminin/collagen/proteoglycan support matrix. Organoids recapitulate the cell types and epithelial architecture found *in vivo* - they resemble an epithelial tissue, with a crypt compartment harbouring the stem cell niche at the base [39] (Figure 5). More recently, organoid models have been developed for a variety of tissues that include gastric [40] cerebral [41] liver and pancreas [42, 43].

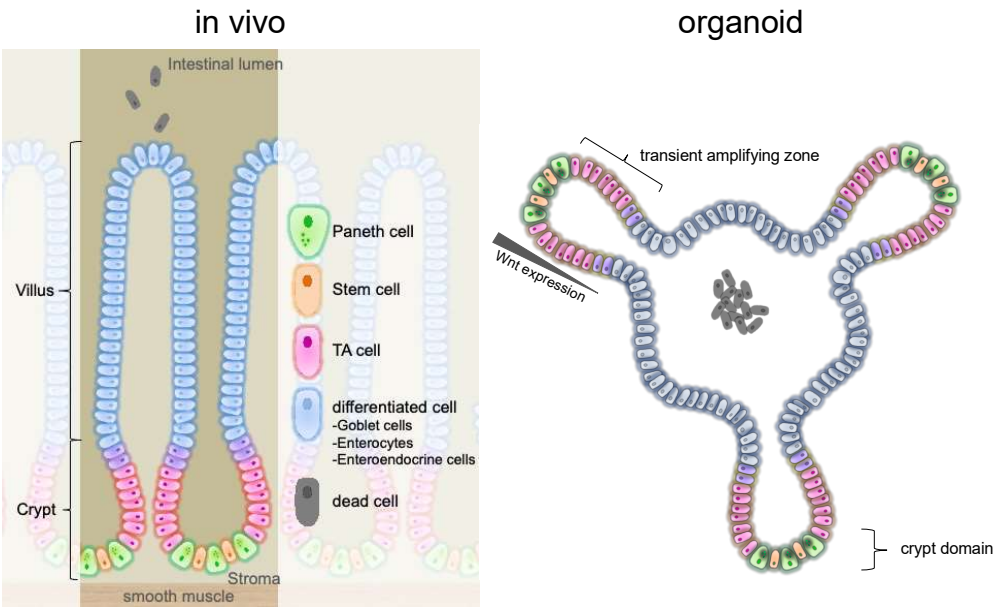


Figure 5. Organoids recapitulate the cell types and architecture of the intestinal epithelia
Schematic presentation of intestinal epithelial cell types and architecture *in vivo* (left), and organotypic cultures (right)

1.3.2 Applications of organoid culture

1.3.2.1 Organoids as models for cell and developmental biology

The development of organoid technology has tremendous potential to facilitate study within the fields of epithelial and cell biology by allowing interrogation of cell behaviour in a tissue context. Owing to their experimental and genetic tractability, organoids can be used to determine cell-autonomous molecular mechanisms regulating epithelial homeostasis and development; for example, the requirement of small intestinal stem cells for specific molecular signals from neighbouring Paneth cells and the underlying stroma for self-renewal [44, 45]. Whilst previous studies have established that a concentration gradient of Wnt ligand defines the stem cell niche, a more recent study using organoids determined that restriction of the stem cell compartment to the bottom of the crypt is maintained by the membrane bound Wnt ligands rather than the diffusion of extracellular protein. Moreover, the discovery of R-spondins and Gremlin-1/Noggin as essential stem cell niche growth factors was determined using organoids [46].

1.3.2.2 Organoids as disease models and for clinical use

One intriguing use of organoids is to model disease enabling precise molecular understanding of cellular origin and the identification of therapeutic targets. A wide range of disease models has been established so far, covering cancer and other genetic, as well as infectious, diseases [47]. Moreover, organoids derived from diseased human tissue have translational value. For instance, clinical biobanking of patient-derived cancer biopsies [48, 49] and therapeutic

screening platforms. Proof of principle studies show tumouroid response is similar to the clinical observations in drug screens and can retain tumour original genetics and histopathologic features [48, 50]. Furthermore, a biobank was developed covering different clinical stages and subtypes [49]. Hence, biobanks of organoids and tumouroids have a huge predictive potential to model and screen diseases especially cancer.

1.3.2.3 The potential of organoids in regenerative medicine

The successful culture of organoids holds promise for regenerative medicine. Studies using animal models have shown that *in vitro* cultured retinal and intestinal organoids can be orthotopically transplanted into mice [[51-53] [54],] and transplanted mouse colon organoids were able to regenerate injuries of the colon mucosa, pointing to a promising direction for the treatment of intestinal failure and IBD in humans [55, 56]. Even early steps in gene therapy approaches were successful using CRIPSR/Cas9 mediated repair of the CFTR mutation in intestinal organoids derived from cystic fibrosis patients [57]. Taken together, organoid and tumouroid culture have future potential to find new treatment options for various diseases, especially cancer.

1.3.3 Control of epithelial homeostasis

A plethora of past work has established that a limited set of signalling pathways have critical roles in controlling epithelial homeostasis. Here I discuss five major signalling pathways that choreograph intestinal epithelial tissue maintenance, regeneration, and repair: (Figure 6)

1.3.3.1 Wnt signalling

The canonical Wnt signalling pathway has been described in 1.2.1 and is the key regulator of intestinal epithelial homeostasis, by maintaining the stem cell niche (Figure 6).

1.3.3.1.1 Wnt signalling in stem cells and Paneth cells

In intestinal epithelia, Wnt signalling activity is restricted to the stem cell compartment in crypts. Through functional studies in the murine intestinal epithelia, somatic mutations in Wnt pathway components within stem cells are widely believed to initiate CRC tumorigenesis [58]. Paneth cells localise to the bottom of the intestinal crypt compartment and produce a range of antibacterial factors like lysozyme and defensins. Intriguingly, human defensin 5 (*DEFA5*) is a Paneth cell specific gene that is expressed by TCF7L2 mediated Wnt signalling activity [59]. Paneth cells express Wnt3 ligands and other growth factors such as Epidermal growth factor (EGF) and Notch ligands that drive cellular programs for proliferation and homeostasis in neighbouring stem cells [60]. However, it remains controversial whether Paneth cells are essential for stem cell maintenance and multipotency [44, 61, 62].

1.3.3.1.2 Non-canonical Wnt signalling

In addition to canonical Wnt signalling there are other non-canonical Wnt signaling pathways whose activities are independent of β-catenin levels [63]. Among these, the Wnt/Ca^{2+} signalling and Wnt/planar cell polarity (PCP) signaling are the best studied. The Wnt/Ca^{2+} signalling activates protein kinase C via phospholipase C and mediates increased intracellular Ca^{2+} [64]. The Wnt/PCP pathway drives remodeling of the cytoskeleton leading to a defined planar polarity of the cell [65] - the axis of polarity within the plane of an epithelial sheet [66].

1.3.3.2 Epidermal Growth Factor (EGF) Signalling

The EGF signalling cascade is described in detail in 1.2.2. The pathway drives a proliferative gene expression program mediated by transcription factors such as *myc* and *elk*. EGF signalling is essential for intestinal epithelial self-renewal (Figure 6) [24].

1.3.3.3 Bone Morphogenic Protein (BMP) Signalling

BMP signalling drives enterocyte differentiation, an activity in opposition to Wnt pathway specification of multipotency in the stem cell niche. Activation of the BMP receptors BMPR1 and BMPR2 leads to the phosphorylation of transcription factors SMAD1, SMAD5 and SMAD8, binding to the transcription factor SMAD4 and target gene expression [67]. The secreted proteins of the Dan family, Noggin, or its homologs such as Gremlin 1 function as BMP inhibitors thereby antagonising differentiation [68]. Gremlin 1 is expressed at the base of intestinal crypts and prevent intestinal epithelial stem cell differentiation [69, 70] (Figure 6).

1.3.3.4 Notch Signalling

The Notch signalling pathway is conserved across all clades of metazoans where it regulates proliferation, differentiation and cell death [71]. Engagement of the Notch ligands Jag or Dll expressed on the surface of one cell with Notch receptors expressed on the surface of a receiving cell leads to activation of γ-secretase protease. Proteolytic cleavage of the intracellular domain of the Notch receptor, the Notch intracellular domain (NICD), releases it into the cytoplasm. Translocation of the NICD to the nucleus and interaction with the transcription factor CSL drives the expression of Notch pathway target genes.

As part of the complex signalling network to balance epithelial homeostasis, the Wnt pathway directly represses Notch signalling by inhibiting CSL through Dvl binding [72]. Moreover, Notch signalling is particularly important to control murine intestinal epithelial stem cell differentiation by preventing secretory cell lineage differentiation [73]. This is likely exerted by Paneth cells surrounding intestinal stem cells (ISC) and expressing the Notch ligands Dll1 and Dll4 on the cell surface [44].

1.3.3.5 The Hedgehog/Gli pathway

Hedgehog (Hh) ligand engagement with patched family receptors (PTC) releases smoothend (SMO). Through a mechanism involving the primary cilium, the kinesins Cos2/Kif7 and Fu and the kinases GSK3 and CK1α lead to stabilisation of the Gli TFs that are normally targeted for proteasomal-mediated destruction [74]. Gli2 and Gli3 are the main signalling mediators in the intestinal epithelia [75] and both can undergo partial processing by the proteasome leading to the conversion from the active form of the transcription factor to a processed repressive form [76]. The contribution of proteasomal function to two signalling nodes regulating Hh/Gli pathway activity constitute interesting parallels with Wnt pathway regulation and are a topic of research in Chapter 3. In the intestine, Hedgehog signalling is active in enterocytes at the top of the crypt and throughout the villi where it drives a program of cell differentiation [77].

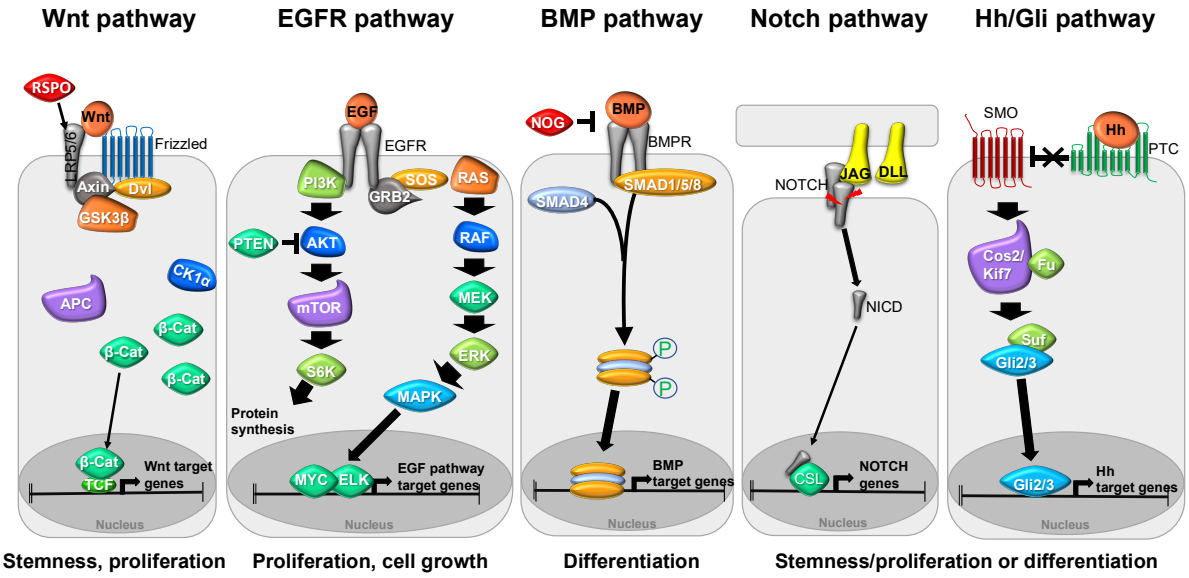


Figure 6. Signalling pathways regulating epithelial homeostasis
Multiple key signalling pathways govern intestinal homeostasis. Representation of the principal signalling cascades of Wnt, EGFR, BMP, Notch, Hedgehog (Hh)/Gli pathway that together control stem cell behaviour and intestinal homeostasis, see text for further details.
Lrp6, low-densitiy lipoprotein receptor-related protein 6; FZD, Frizzled; R-spondin , R-spondin; CK1α, casein kinase 1 α ; GSK3β, glycogen synthase kinase 3 β; Dvl, Dishevelled; APC, adenomatous polyposis coli; β-cat, β-catenin; TCF, T cell-specific transcription factor; EGF, epidermal growth factor; EGFR, EGF receptor; PI3K, phosphoinositide 3-kinase; MAPK, mitogen-activated protein kinase; BMP, bone morphogenetic protein; BMPR , BMP receptor; Dll, delta-like ligand; NICD, Notch intracellular domain; *red arrow*, γ-secretase; CSL, CBF1, suppressor of hairless; Hh, Hedgehog; PTC, Patched; SMO, Smoothend; Gli, glioblastoma.

1.4 Potentiation of Wnt signalling

1.4.1 The LGR family of Wnt pathway receptors

Leucine rich repeat containing G-protein coupled receptor (LGR) 5 is a 7-transmembrane receptor whose extracellular N-terminal domain consisting of 16 leucine rich repeats (LRRs) that fold into a horseshoe shape (Figure 7). It is part of the group B LGR protein family that are receptors for the R-Spondin (R-spondin) family of extracellular ligands. Mammalian cells express three LGR isoforms, LGR4, LGR5 and LGR6, but only LGR4 and LGR5 are expressed in the intestine.

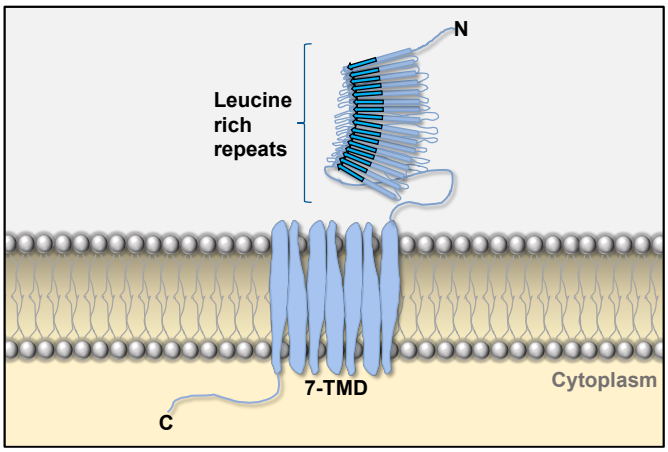


Figure 7. Schematic LGR protein structure.
LGR family proteins contain a 7 transmembrane domain and a large N-terminal extracellular domain composed of 16 leucine rich repeats (LRRs) that form the receptor for R-spondin (R-spondin) ligands.

1.4.2 LGR5 – stem cell marker and Wnt agonist

In the intestinal epithelia the Wnt target gene, *LGR5,* is uniquely expressed by stem cells located at the base of crypts. *In vivo* lineage tracing experiments that rely on activity of the *LGR5* gene promoter driving expression of the LacZ or eGFP transgenes demonstrated that LGR5 expressing cells give rise to all cell types found in the intestinal epithelia, establishing them as stem cells [78]. Similar experiments using the LGR5-eGFP reporter mouse strategy indicated LGR5 expression was a characteristic of stem cells from the mammary gland, liver, kidney, stomach, pancreas and other epithelial tissues [79-82]. Hence, LGR5 is a ubiquitous epithelial stem cell marker. In support of this notion, intestinal epithelia organoid cultures can be established from single LGR5 expressing stem cells [38].

1.4.3 LGR proteins potentiate Wnt pathway activity

LGR5 is a stem cell marker and functionally promotes stemness through its activity in the Wnt pathway [83]. Genetic deletion of LGR4, which is expressed throughout the crypt compartment and LGR5 in the mouse intestine specifically attenuated the Wnt target gene program [19]. Recent work has unearthed the mechanism for LGR protein function in the Wnt pathway:

Frizzled-LRP5/6 are co-receptors for Wnt ligands that activate the pathway leading to β-catenin stabilisation (Figure 6). The Wnt co-receptors associate with the Rnf43 ubiquitin ligase and constitutive ubiquitination drives their endocytosis and degradation [18] (Figure 8). R-spondins bind with high affinity to a heterodimer of LGR protein and Rnf43, leading to internalisation (Figure 8). The removal of Rnf43 from the cell surface stabilises the expression of the Wnt co-receptors and increases the cellular response to afferent Wnt ligand. It is important to note that cellular treatment with R-spondins (R-spondins) does not in itself activate the Wnt pathway, it merely potentiates pathway activity alongside Wnt ligand treatment.

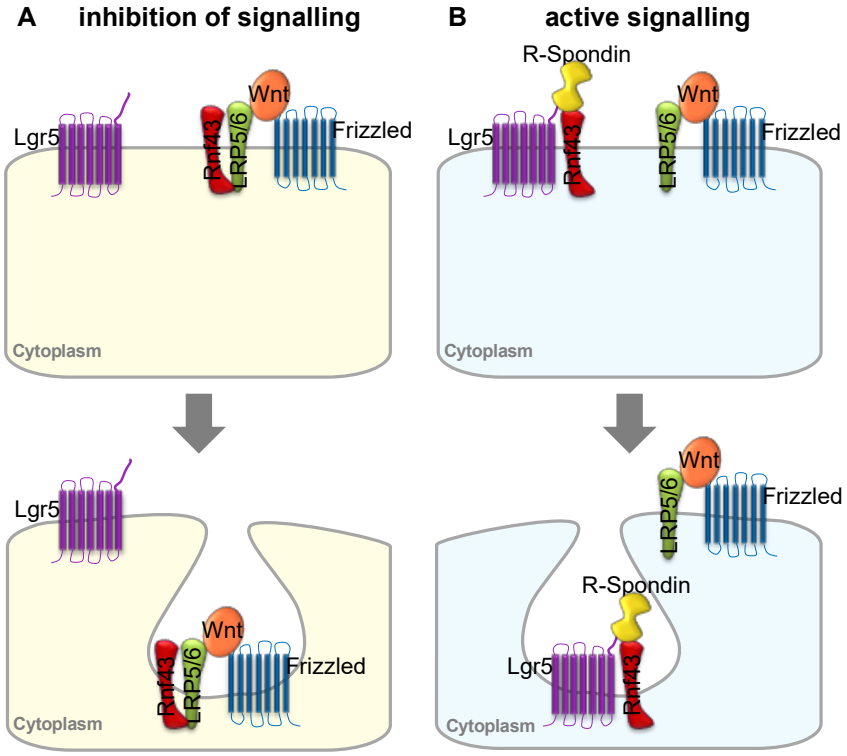


Figure 8: Mechanism of R-spondin mediated enhancement of Wnt signalling
(**A**) Binding of Wnt ligands to the Frizzled LRP5/6 complex activates the Wnt pathway which is attenuated by ubiquitination of Frizzled by the E3 ligase Rnf43 and internalisation of the Rnf43/LRP/Frizzled complex. (**B**) R-spondin nucleates a heterotrimer with LGR5 and Rnf43 that is, in turn, internalised, increasing the expression of cell surface Wnt receptor and potentiation Wnt pathway activity.

1.4.4 LGR5 in normal physiology and colon cancer

LGR5 plays its role in potentiating Wnt signalling activity in conjunction with its ligand R-spondin and Wnt pathway activity and is positively regulated by this feedback loop as a Wnt transcriptional target. In normal gut epithelium LGR5 is dispensable for epithelial homeostasis [84] but is necessary for creating different stem cells pools to create progeny of different secretory signatures [85]. Overexpression of LGR5 is a general feature of CRCs [86-94] and increased LGR5 levels are found at the invasive front of tumours [87, 89, 92]. Functionally,

this might be a consequence of increased Wnt pathway transcriptional activity; β-catenin levels are characteristically high at the invasive front of CRC tumours [95-97].

LGR5 expression throughout the course from adenoma to metastatic carcinoma might be a plastic phenomenon with expression being lost from adenoma to carcinoma and regained at the metastatic stage [98] LGR5 overexpression in CRC tumours correlates with a poor clinical outcome and treatment response [99, 100]. Cellular studies using cell lines and xenograft models have ascribed roles for LGR5 in promoting proliferative and migratory [101] and survival [102]. However, LGR5 function in CRC is controversial. Some studies have found that high LGR5 expression correlates with good prognosis [103] and loss of LGR5 expression during cancer development, likely through *LGR5* promotor methylation [104]. Moreover, the cellular function of LGR5 in promoting Wnt pathway activity has been recently disputed - Lgr5 depletion in mice and cell lines led to upregulation of the Wnt target genes Axin2, Ascl2 and CD44 [105-107].

Despite the controversy over its expression levels in CRC, cancer cells expressing LGR5 have emerged as a therapeutic target - antibody drug conjugates (ADCs) targeting LGR5 have found success in reducing proliferation and tumour size of engrafted CRC cells in mice [108]. Moreover, combining LGR5 ablation by CRISPR/Cas9 in combination with standard of care treatment in organoids has been effective for driving tumour regression [109].

1.4.5 LGR-R-spondin interaction

Indirect binding assays using fluorescently tagged versions of the R-spondins (isoforms 1-4) with LGR4 or LGR5 transfected cell lines indicate very high affinity interactions, approximate *Kd* of 2-5 nM [110]. Previous studies from the Pioszak laboratory have used bacterial expression systems to produce pure versions of R-spondin and indirect binding studies with these recombinant proteins confirmed binding affinities in the nM range [111]. Studies measuring direct binding affinities between LGR4 and LGR5 and the four R-spondin isoforms are lacking.

Binding of the R-spondin isoforms 1-4 to the ectodomains of LGR4 and LGR5 is mediated by the N-terminal Furin domains of R-spondin to the leucine rich repeats 1-10 (LRRs 1-10) of LGR4/5. The Furin domains of R-spondin s consist of a conserved cysteine-rich domain (CRD), divided into so-called head and rod modules. The head module is a loop emanating from a β-sheet structure that interacts with Rnf43 binding (see Chapter 1, Introduction). The rod module consists of two cysteine knots that create an interface for LGR4/5 binding (Figure 28). The R-spondin amino acids comprising the LGR4/5 binding interface are for the most part hydrophobic and include the invariant residues Phe106 and Phe110 that are thought to mediate the high affinity interaction with LGR (Figure 28) [112]. The other amino acid residues comprising the binding site are highly conserved and thus the R-spondins are not expected to

display overt binding specificity to LGR family members [113]. The hydrophilic residues that vary amongst R-spondin isoforms may explain the modest differences in binding affinities for LGR4 and LGR5 [112] and may provide a 'handle' for engineering R-spondin with different binding affinities for LGR family members.

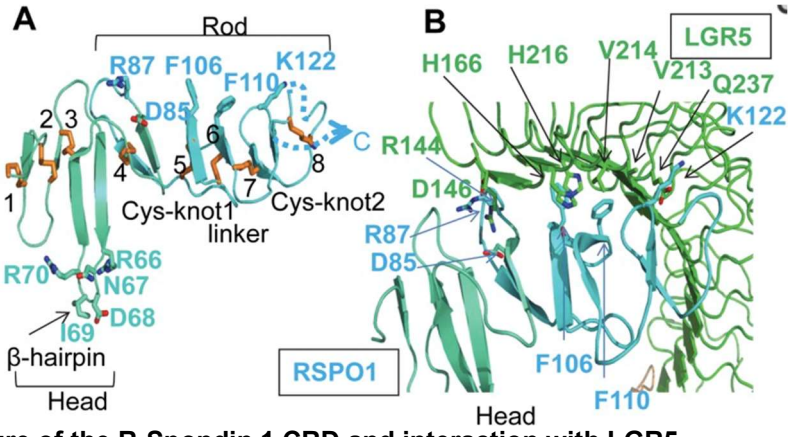


Figure 9: Structure of the R-Spondin 1 CRD and interaction with LGR5
(**A**) R-spondin 1 CRD consists of a head (blue-green) and a rod (blue) module. Disulphide bridges 5-8 within the rod module position residues required for LGR5 binding. (**B**) Binding interface between R-spondin 1 and LGR5 (green) with interacting amino acids as sticks.

1.5 TCF7L2 – key transcription regulating effector of the Wnt pathway

1.5.1 The TCF/LEF family

In mammals the TCF/LEF family of DNA binding proteins consists of TCF7 (formerly named TCF1), LEF1, TCF7L1 (TCF3) and TCF7L2 (TCF4) and function as the transcription factors (TFs) for the transmission of Wnt pathway activity to target gene expression. The TCF/LEF TFs harbour common structural features: an N-terminal interaction domain for β-catenin, a Transducin like Enhancer of Split/Groucho-related genes (TLE/GRG) corepressors binding site (GBS), and an homeobox (HMG-box) DNA-binding motif [114] as well as a nuclear localisation signal (NLS). TCF7 and TCF7L2 are unique amongst the family members in that they contain a C-terminal extension composed of a second DNA binding motif, termed the C-clamp and a C-terminal binding motif (CtBM) that has been shown, *in vitro*, to bind to the transcriptional corepressor CtBP [115] (Figure 10, 11).

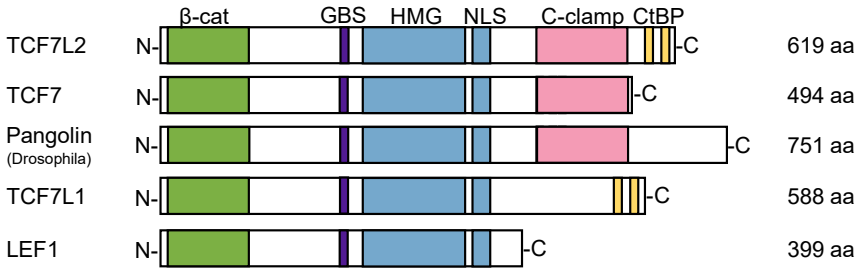


Figure 10: Schematic representation of TCF/LEF family members
All TCF/LEF proteins share a common domain structure with an N-terminal β-catenin binding domain, a GBS, a HMG domain, followed by a NLS. TCF7L2, TCF7 and the only TCF member in Drosophila Pangolin harbour an additional auxiliary DNA binding domain, the C-Clamp. TCF7L2 and TCF7L1 are the only family members with C-terminal binding protein interaction sites (CtBP). For all TCFs the the size of encoded protein is depicted and indicated in length of amino acids (aa).

In the absence of Wnt pathway activation, TCF/LEF TFs bind to target genes and repress Wnt target gene expression [116, 117] through their interaction with the transcriptional corepressors Groucho [118] and C-terminal binding protein (CtBP) [119]. Both Groucho and CtBP recruit nuclear corepressors, such as histone deacetylases (HDACs), to the promoters of Wnt targeted genes [120] facilitating transcriptional repression through 'closed' chromatin remodelling. Furthermore, the HMG-box binding repressor HBP1 has been reported to inhibit the Wnt signalling output by interacting with TCF7L2 [121] and inhibiting its binding to WREs and recruiting corepressive HDACs [122].

In the presence of Wnt pathway activity, increased β-catenin levels lead to its interaction with TCF/LEF TFs, displacing Groucho and recruiting a cohort of transcriptional co-activators such as histone acetyltransferases (HATs) CBP/p300 [123]. Consequent 'open' chromatin remodelling enabling access of the transcriptional machinery (Figure 11). More recently, one study has demonstrated TCF/LEF TF independent transcription of Wnt pathway target genes [124] and it will remain to be seen whether this data is generally applicable.

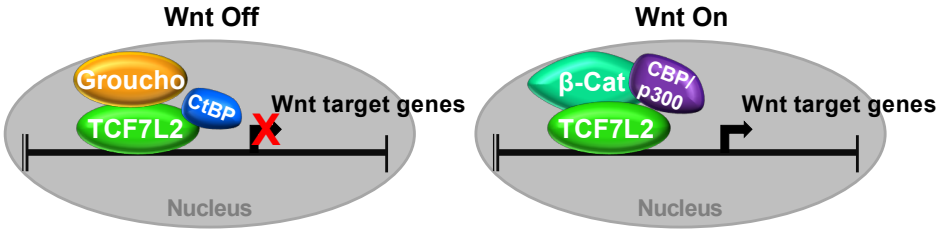


Figure 11: Regulation of Wnt pathway target gene expression by TCF7L2
Left panel - Wnt Off, no Wnt pathway activity; TCF7L2 is bound to the Wnt pathway co-repressors Groucho/TLE and CtBP actively repressing target gene expression. *Right panel* - Wnt On, Wnt pathway activity leads to binding of TCF7L2 by β-catenin, displacing Groucho/TLE and recruits coactivators such as CBP/p300 to drive Wnt target gene expression.

In vivo studies have established that TCF7L2 is the crucial Wnt pathway effector transcription factor in murine intestinal epithelial homeostasis, as$TCF7L2^{-/-}$ mice lack a defined intestinal epithelial stem cell compartment – gene ablation of other TCF/LEF family members in the murine intestine reveal much milder phenotypes. Gene ablation studies have shown that TCF7 is the principal regulator of Wnt signalling in T-cells whereas LEF1 appeared to be more functionally complex regulating tissue development in hair follicles, and proliferation in B-cells. [125-127].

1.5.2 TCF7L2 splice variants and their regulation

Expression of various forms of TCF7L2 protein is believed to arise from alternative splicing - up to 25 variants have been identified in human cells [USCS, Genome Browser]. The human *TCF7L2* gene consists of 18 exons with 6 alternatively spliced exons - exons 4, 6 and 14-17 (Figure 12). [128, 129]. Additionally, exon 10 contains two splice acceptor sites. Based on the domains within the encoded proteins that are altered as a result of alternative splicing, expressed TCF7L2 protein can be stratified into four categories:

(i) E-forms: TCF7L2-E proteins are approximately 70 kDa and as the largest of the isoforms, I term them the 'long forms'. TCF7L2-E forms harbour a C-terminal extension containing a second DNA binding domain, the C-clamp, as well as two interaction sites for the transcriptional corepressor C-terminal Binding Protein (CtBP) (Figure 12) [119].

(ii) S-forms: TCF7L2 S-forms are expressed as shorter proteins and while lacking most of the C-terminal extension, possess a partial C-clamp

(iii) M-forms (also known as B-forms): TCF7L2 M-forms and the short forms of the protein with around 55 kDa that lack the entire C-terminal extension (Figure 12)

(iv) N-forms: TCF7L2 N-forms are generated by a premature stop codon using a variable splice acceptor site at the beginning of exon 10. This isoform is truncated before the HMG DNA binding domain, and functions as dominant negative form of TCF7L2 by directing β-catenin away from target gene promotors [130].

Studies ascribing distinct functions for different TCF7L2 forms have determined they possess unique transcriptional activities for specific cohorts of Wnt pathway target genes [131]. While gene target cohorts of individual TCF7L2 proteins have, in part, been determined, it is not clear how the Wnt pathway transcriptional program is shaped by context-specific and combinatorial expression of TCF7L2 variants. One hypothesis is that expression of the TCF7L2 forms influences Wnt target gene expression by the presence of the C-clamp DNA binding domain.

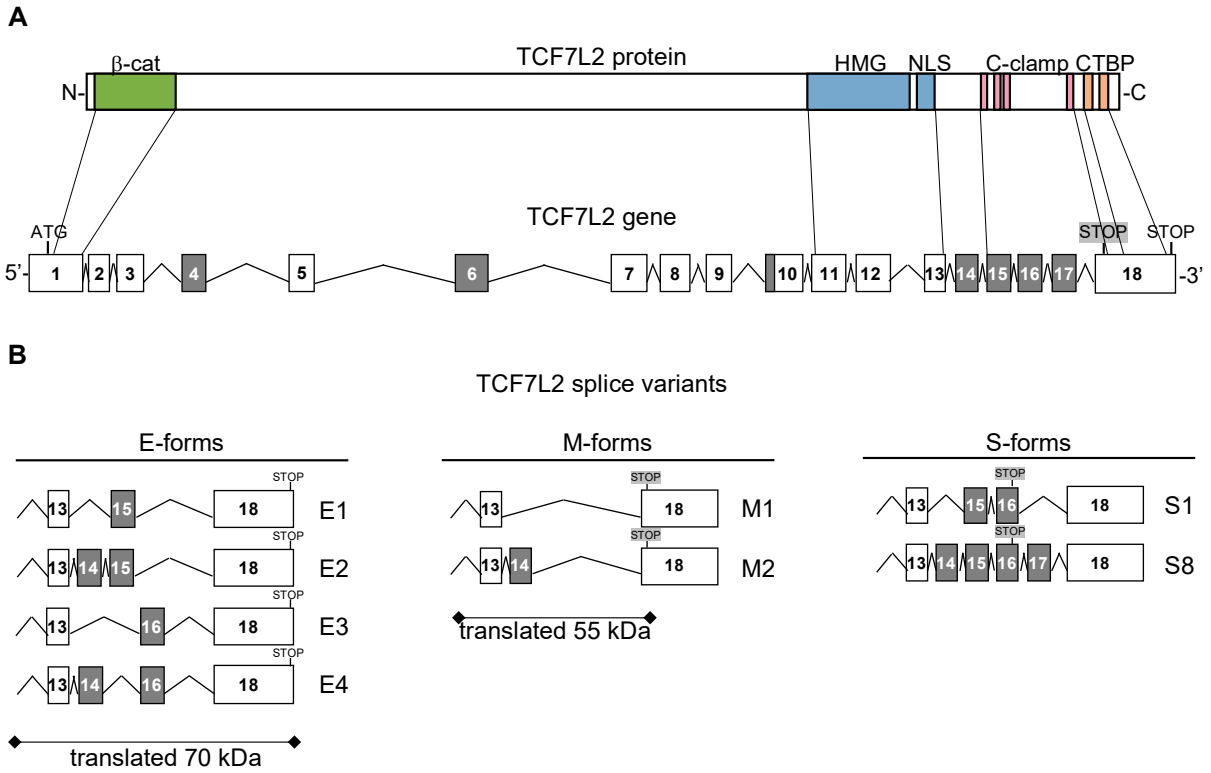


Figure 12: TCF7L2 alternative splicing isoforms in human CRC cell lines
(**A**) Simplified scheme correlating TCF7L2 protein to the encoding human *TCF7L2* gene exons (not to scale). The *TCF7L2* gene consists of 18 exons with 6 alternatively spliced exons indicated in grey. A premature stop codon generated through an alternative splice acceptor site in the M-forms is underlaid in grey. Alternative splice acceptor at the beginning of exon 10 leads to expression of the TCF7L2 N-form. (**B**) Close-up of the 3' end of the *TCF7L2* gene showing the exon structure leading to alternatively spliced TCF7L2 variants. Nomenclature according to [131], translated onto splice variants detected in human CRC cell lines [132].

1.5.3 The C-clamp motif

The C-clamp (cysteine-clamp) is a bipartite auxiliary 30-amino-acid DNA binding motif found in the TCF7L2 E-forms and in some forms of TCF7, the only other TCF/LEF family member containing the C-terminal extension. The C-clamp contains 4 cysteine residues that coordinate a Zn^{2+} ion which forms the DNA binding moiety and invariably the amino acid sequence motif C-R-A-R-F. Studies using expressed versions of TCF7 and TCF7L2 and their C-clamp mutants indicates the ability to binds promoters of Wnt pathway target genes containing so-called helper sites, GC-rich DNA sequences located within 9-11 bases from canonical Wnt response elements (WREs) [133]. WREs have the consensus sequence CCTTTGWW and are bound by the HMG box common to all TCF/LEF family members [134, 135]. Moreover, TCF1E (TCF7L) but not TCF1B (TCF7S) (without C-clamp) was able to activate LEF1 transcription as well as TCF7L2E but not TCF7L2M or TCF7L2S (M and S without C-clamp) activated CDX1 expression [131, 133].

The key point is that Wnt pathway target genes contain WREs that are bound by all versions of TCF7L2 and TCF/LEF family members. However, Wnt pathway target promoters that

contain both a WRE and helper site are exclusive targets for the C-clamp containing TCF7L2E and TCF7L proteins. Now it will be important to define the cell-specific composition of TCF7L2 proteins generated by alternative splicing and how this shapes Wnt pathway transcriptional responses and biological output.

1.5.4 Co-transcriptional activation and repression of TCF7L2

Another feature found of the TCF7L2-E C-terminus is a binding domain for the transcriptional coactivator p300, that was determined as an interactor in one *in vitro* study. Further investigation using a p300 overexpression system in cell lines found that the protein modulated expression of the Wnt pathway target gene CDX1 [136]. TCF7L2-E may also bind CtBP. The functional consequence of CtBP binding is controversial; one study has found that binding leads to both activation and repression of Wnt target gene transcription [137] while another study has found it does not bind to TCF7L2 directly and instead repressed Wnt pathway transcription through binding to APC and diverting β-catenin away from gene targets [138]. Taken together, there is a lack of consensus regarding the function of interactions between TCF7L2 and p300 or CtBP and the functional ramifications on expression of Wnt pathway target genes.

1.5.5 TCF7L2 in CRC

The role of TCF7L2 in intestinal oncogenesis is complex. TCF7L2 contributes to colorectal carcinogenesis by stimulating stem cell proliferation [139]. This oncogenic function of TCF7L2 is believed to be mediated through expression of known drivers of proliferation [83, 140-142]. However, a role for TCF7L2 as an oncogene is controversial. Some have found TCF7L2 to be dispensable for tumour initiation [143]. Moreover, the prevalence of loss-of-function TCF7L2 mutations in cancer suggest a tumour suppressive role for TCF7L2 [144-146]. Overexpression of TCF7L2 reduces growth of CRC cell lines, supporting this hypothesis [147] and TCF7L2 depletion leads to a slower G1/S transition although these cells shown enhanced migratory and invasive phenotypes [148].Taken together, TCF7L2's role in tumorigenesis appears to be ambiguous and its biology has yet to be fully elucidated.

TCF7L2 mutations and polymorphisms of TCF7L2 have been well studied in type 2 diabetes where the single nucleotide variant (SNV, rs7903146) provides the strongest correlation to T2D risk [149-151]. Moreover, alternatively spliced TCF7L2 variants have been assigned to T2D risk correlations with variants containing exon 4 providing the highest correlation [152, 153] in a tissue specific manner [128]. Although it is not understood how *TCF7L2* gene regulation influences T2D risk, there is evidence for a role of TCF7L2 expression in the regulation of glucose metabolism [154].

TCF7L2 has high mutation rates in CRC which may rationalise a link between T2D polymorphisms and colorectal cancer [155]. TCF7L2 contains a microsatellite within the coding region at the start of exon 18, adjacent to the splice acceptor site. The microsatellite consists of a string of 9 adenines (A9 microsatellite) and insertion/deletion mutations within lead to frameshifts and termination at new stop codons, most frequently resulting in a truncated protein of around 55 kDa (Figure 13). The exon 18 microsatellite mutation have been reported in MSI-high cancers along the gastrointestinal tract [156]. Analysis of CRC mutation signatures have identified A9 microsatellite mutations with high frequency in tumours with MSI-H and was even proposed to be positively selected for in contrast to non Wnt signalling dependent gastric and endometrial cancer [157, 158]. In contrast, a more recent study indicated A9 microsatellite mutations may be merely passenger mutations and do not contribute to tumorigenesis of MSI-H CRC [158, 159]. It is unknown what contribution the A9 microsatellite mutations to CRC aetiology.

Figure 13: TCF7L2 A9 microsatellite frameshift mutations
Frameshift mutations within the A9 microsatellite (red) create premature Stop codons and result in protein truncation to 55 kDa with loss of Lysine 485 (K485, green).

1.6 Aims of the thesis

TCF7L2 is the principle Wnt pathway transcription factor in the intestinal epithelia. However, its role in normal intestinal epithelial homeostasis and tumorigenesis is controversial. In this thesis, I will:

- Determine the cell specific expression of TCF7L2 variants in CRC cell lines and investigate the hypothesis of TCF7L2 variant generation through post-translational processing.
- Investigate TCF7L2 microsatellite mutations in CRC cohorts and determine the transcriptional and biological output of altering expression levels of TCF7L2 variants in cell lines.

I will also detail my work improving organoid technology that focusses on the activity of LGR5:

- Produce organoid media formulations of defined activity for the investigating TCF7L2 expression and activity.
- Develop an LGR5-specific R-spondin variant that specifically discriminates stem cells.

This will enable future work to study TCF7L2 variant biology in different Wnt regulated cell types, as so isolated stem cells and Paneth cells.

Chapter

2 Materials and Methods

2.1 Cell culture

All cell lines were maintained in Dulbecco´s Modified Eagle Medium (DMEM; Gibco) containing 10% Fetal Bovine Serum (FBS), 2 mM Glutamine, 2 mM Penicillin/Streptomycin at 37°C with 5% CO_2.

Cells were grown in cell culture dishes and flasks and split at 90 – 95% confluence using 0.05% Trypsin-EDTA reagent. Cell lines used in this thesis are outlined in Table 1. Genetic background regarding tumour driving mutations are listed in Table 2.

Table 1. cell lines used in this thesis.

Cell line	Origin	Microsatellite status
HEK293T	Human embryonic kidney (T-antigen)	n.a
L-cells Wnt3a	Murine fibroblast Wnt3a expression stable	n.a
HCT116	Human colon adenocarcinoma	MSI
RKO	Human colon adenocarcinoma	MSI
LOVO	Human colon adenocarcinoma	MSI
DLD-1	Human colon adenocarcinoma	MSI
HT29	Human colon adenocarcinoma	MSS
SW480	Human colon adenocarcinoma	MSS
C2C12	Murine myoblasts	n.a.

MSI: microsatellite instable, MSS: microsatellite stable, n.a.: not applicable

Table 2. genetic background of colorectal cancer cell lines

Cell line	CIMP	KRAS	BRAF	PIK3CA	PTEN	TP53	APC	CTNNB1
HCT116	+	G13D	wt	H1047R	wt	wt	wt	ΔS45/wt
RKO	+	wt	V600E	H1047R	wt	wt	wt	wt
LOVO	-	G13D/A14V	wt	wt	wt	wt	wt	wt
DLD-1	+	G13D	wt	E545K/D549N	wt	S241F	1427*	wt
HT29	+	wt	V600E	P449T	wt	R273H	1555*	wt
SW480	-	G12V	wt	wt	wt	R273H	1338*	wt

CIMP- CpG island methylator phenotype, wt- wildtype

2.2 Expression plasmids

Expression plasmids generated used throughout the course of this work are listed in Table 3. All plasmids generated during the course of the work were created using Gibson Assembly (ThermoFisher) and are described in Appendix 1.

The pR-spondin 1 construct for production of R-spondin 1 is based on a previously reported pETDuet-based vector [111] for dual expression of (i) a R-spondin 1-containing fusion protein consisting of the maltose binding protein transgene, a tetra-cysteine tag, a thrombin cleavage site, residues 21–145 of human R-spondin 1 and a hexa-histidine tag) and (ii) the *E. coli* disulphide bond isomerase, *DsbC*. I have inserted the coding sequence of the Avi-tag upstream of the R-spondin 1 transgene downstream of the MBP moiety by digestion of the vector with *BamH*I and *Not*I and Gibson assembly to create pR-spondin 1.

Plasmids used for quantifying Wnt pathway activity were the Super 8x TOPFlash (SuperTop; Addgene 12456) and the *Renilla* Luciferase control reporter vector pRL-SV40P (Promega). The expression construct, pHAT4-ΔN-Grem1, encoding the hexa-histidine tag, TEV protease cleavage site and amino acids 72–184 of human Gremlin 1 (Uniprot: O60565), has been previously described [160]. The human BMP2 bacterial expression plasmid pBAT-BMP2 and purification of the expressed protein has been detailed elsewhere [160].

Table 3: Expression plasmids used in this work.

Expressed transgene	Plasmid backbone	Source
myc-TCF7L2	pcDNA3	Addgene #16512
myc-TCF7L2-FL-FLAG	pcDNA3	This thesis
myc-TCF7L2-K485S fs-FLAG	pcDNA3	This thesis
myc-NFκB-FLAG	pcDNA3	This thesis
myc-NFκB	pcDNA3	This thesis
myc-TCF7L2 KKK->IPA	pcDNA3	de la Roche laboratory
myc-TCF7L2 ΔCT-internal	pcDNA3	de la Roche laboratory
myc-TCF7L2 ΔCT 5aa	pcDNA3	de la Roche laboratory
myc-TCF7-L	pcDNA3	de la Roche laboratory
myc-TCF7-S	pcDNA3	Addgene #40620
myc-TCF7L2-FL-FLAG	piggyBac	This thesis
myc-TCF7L2-K485S fs-FLAG	piggyBac	This thesis
MBP-Avi-R-spondin 1 21-145 His	pETDuet1	This thesis
MBP-TC-R-spondin 1 21-145 His	pETDuet1	This thesis
MBP-LGR4-LRR10-hagVLRB-His	pETDuet1	This thesis
MBP-LGR5-LRR10-hagVLRB-His	pETDuet1	This thesis
8x TOPFlash	SuperTop	Addgene #12456
Renilla	pRL-SV40P	Promega
ΔN-Grem1	pHAT4	Hyvönen laboratory
BMP2	pBAT	Hyvönen laboratory

2.3 RT PCR

Gene expression analysis using Reverse Transcription Polymerase Chain Reaction (RT-PCR) cell lines were cultured in 12 well plates and mRNA extraction was carried out using TRI Reagent® (Sigma Aldrich). Confluent cell cultures for testing were washed once with cold PBS (Gibco) then lysed with 500 µl TRI Reagent®. Cell monolayers were harvested by scraping and vigorous pipetting, transferred to 1.5 ml reaction tubes, vortexed and incubated for 5 minutes at ambient temperatures to facilitate nucleoprotein complex dissociation. Cell debris was removed by centrifugation at 2000 rpm for 5 minutes and supernatant transferred into clean 1.5 ml tubes. For phase separation, 100µl of chloroform was added to each sample followed by vigorous vortex and centrifugation at 12000 x g for 15 minutes at 4 °C. The upper aqueous phase was collected into a fresh 1.5 ml tube and RNA was precipitated by adding 250 µl of isopropanol. After a 10 min incubation at ambient temperatures, RNA was pelleted

by centrifugation at 12000 x g for 10 mins at 4 °C and the supernatant removed. The RNA pellet was washed twice with 500 µl 75 % ethanol and air dried for 20 minutes at room temperature. The resulting RNA pellet was reconstituted in 25-50 µl nuclease free water, concentration measured using a NanoDrop spectrophotometer and stored for future use at -80 °C. Purity was assessed by measuring A260/A280 nm and A260/A230 ratios. A ratio of 1.7-2.1 for A260/A280 and 1.8-2.2 for A260/A230 was set as acceptable limits for proceeding with cDNA synthesis.

cDNA was synthesised using the High-Capacity cDNA Reverse Transcription Kit (ThermoFisher) supplemented with RNase inhibitor according to manufacturer's instructions. Reverse transcription reactions were carried out using 1 µg of RNA per 20 µl reaction in a thermocycler (Biometra Thermocycler 3000) under the following conditions: 10 mins at 25°C, 120 mins at 37°C and 5 mins at 85°C followed by incubation at 4°C until further. cDNA was stored at -20°C until further use.

Quantitative reverse transcription PCR (qRT-PCR) was performed using 2 x FastSYBR™ Green Master Mix (Applied Biosystems). cDNA was diluted in water (1:2 dilution). qRT-PCR reactions were carried out using 1 µl of cDNA, 1 µl each of forward and reverse primers (stock concentrations of 400 nM), 5 µl FastSYBR™ Green Master Mix and nuclease free water to a final reaction volume of 10 µl. For amplification a QuantStudio 5 thermocycler (Applied Biosystems) was used with the following cycle: 95°C for 15 seconds and 60°C for 20 seconds for 40 cycles. Ct values were derived from fluorescent data using QuantStudio 5 software and $\Delta\Delta$Ct values derived using Excel relative to the housekeeping gene GAPDH.
Primers for RT PCR (Sigma, UK) are listed in Table 4. Primer sequences were derived from the Harvard University Primer Bank.

Gene expression analysis of primary mouse T-cells was carried out as follows: T-cell populations were isolated by fluorescence activated cell sorting (FACS) using fluorescent antibodies to the cell surface markers CD3, CD4 and CD8. Cells were spun down, and pellet stored in -80°C for further analysis by RT-PCR. RNA was extracted using the RNA extraction Kit for very low cell numbers (Qiagen) and expressed genes were quantified using Taqman RT-PCR probes (Applied Biosystems). In this instance, the control housekeeping gene was TBP Mm01277042_m1 and probes for assessed genes were LGR5 Mm01251802_m1, LGR4 Mm00554385_m1, and c-myc Mm00487804_m1. RT-PCR was kindly carried out by Louise O'Brien (Maike de la Roche laboratory, CRUK).

Table 4: PCR Primers for human genes used in this chapter

Gene	Sequence 5' → 3'	Application
TCF7L2 βcatBD for	ACGAACTGATTTCCTTCAA	splice variants/abundance
TCF7L2 βcatBD rev	CCAAACTTTCCCGGGATTT	abundance
TCF7L2 exon 3-4 for	GCCCGAACCCTCCATTTTCA	abundance
TCF7L2 exon 5 rev	AGCAGTGGCCATTTCATCTG	splice variants/abundance
TCF7L2 exon 5-6 for	TGTCAGCCCCCTCCCTT	abundance
TCF7L2 exon 6 rev	TGGCTCCTTTTCATGTCCCTG	abundance
TCF7L2 exon 9 for	GCTCTGACCGTCAATGCTTC	splice variants
TCF7L2 exon 14 for	CTTGCCTTTCACTTCCTCCG	abundance
TCF7L2 exon 15 rev	GCCGCACCAGTTATTCTGTT	abundance
TCF7L2 exon 15-16 for	CCCTTGCAGATGCAAATACTCC	abundance
TCF7L2 exon 16 rev	CTGCACGGTTTGCACCATAA	abundance
TCF7L2 exon 13-18 for	CGGGAGAGACCAATG	abundance
TCF7L2 exon 18 rev	TTGTATGTAGCGAACGCAC	splice variants/abundance
GAPDH for	TGCACCACCAACTGCTTAGC	abundance/GEP
GAPDH rev	GGCATGGACTGTGGTCATGAG	abundance/GEP
MAGEA12 for	GGAGACGAGCTTCCAAGTAGC	GEP
MAGEA12 rev	GCACGATCTGATTGTCGCC	GEP
DEFA5 for	AGACAACCAGGACCTTGCTAT	GEP
DEFA5 rev	GGAGAGGGACTCACGGGTAG	GEP
GATA4 for	CGACACCCCAATCTCGATATG	GEP
GATA4 rev	GTTGCACAGATAGTGACCCGT	GEP
SFRP5 for	TGCTGCACTGCCACAAGTT	GEP
SFRP5 rev	GTGCTCCATCTCACACTGGG	GEP
TBP	Mm01277042_m1 Taqman	T-cells
LGR5	Mm01251802_m1 Taqman	T-cells
LGR4	Mm00554385_m1 Taqman	T-cells
c-myc	Mm00487804_m1 Taqman	T-cells

(application splice variants: endpoint PCR for detection of splice variants, abundance: qRT-PCR for quantification of expressed splice variants, GEP: Gene Expression profiling TCF7L2-FL vs TCF7L2-S expressing cell lines, T-cells: RT-PCR for LGR5 postitive SPNH+ sorted population)

2.4 Protein extraction, Western blotting and quantification

Cells were cultured using 10 cm or 6 well dishes. They were washed twice with PBS and then lysed with 300 µl or 100 µl 1x RIPA buffer (Millipore) containing 1x EDTA free protease inhibitor cocktail (cOmplete®, Roche) and 1x phosphatase inhibitors (PhosSTOP™, Roche). After 5-minute incubation on ice cells, were harvested by scraping and transferred into a 1.5 ml microfuge reaction tube. The sample was sonicated for 15 minutes in a pre-cooled waterbath followed by centrifugation at 20 000 x g for 20 minutes at 4°C. The supernatant was collected, and the protein concentration determined using Biorad 5x Bradford reagent calibrated with a BSA standard. Samples of 40 ng of protein, 100 nM DTT and 1:4 NuPage LDS sample buffer (ThermoFisher) were denatured at 95°C for 5 minutes and loaded onto 12% denaturing SDS-gel. After PAGE, proteins were transferred onto a 0.45 µm pore PVDF membrane (Immobilon®-P, Merck) by 100 V for 1 hour at 4°C. Protein transfer was confirmed

with PonceauS solution (ThermoFisher). The membrane was blocked with 5% skimmed milk or 5% BSA according to primary antibody requirements. Primary and secondary antibodies are listed in table 5. ECL substrate (Pierce) or BCIP/NBT solution was used for development. Relative band intensities from Western blots were quantified using ImageJ™ software and its gel analysis tool. Data was imported into excel for quantification.

Table 5: Antibodies used in this chapter for Western blotting

Protein target	Conjugate	Species	Company	Cat. no	Dilution	Dilution buffer	Blocking buffer
myc-epitope tag		mouse	lab	9E10	1:1000	BSA	milk
TCF7L2 (C48H11)		rabbit	CST	2569	1:2500	milk	milk
TCF7 (C46C7)		rabbit	CST	2206	1:2000	milk	milk
GAPDH		rabbit	Sigma	G9545	1:5000	milk	milk
Vinculin		rabbit	CST	4650	1:5000	milk	milk
Mouse IgG	HRP	goat	Abcam	Ab6721	1:10000	milk	
Rabbit IgG	HRP	goat	Abcam	Ab6789	1:10000	milk	

All antibodies were incubated with membranes in PBS containing 0.2% Tween80 + 5% w/v skimmed milk except myc-epitope tag antibody where 5% BSA was used instead of skimmed milk

2.5 Expression and purification of R-spondin 1

Expression of the R-spondin 1 fusion protein from pR-spondin 1 was carried out in the host *E. coli* strain SHuffle® T7 *E. coli*/K12, following a protocol from the Pioszak laboratory [111]. An overnight starter culture was used to inoculate 2 × 1 L of Luria Broth (LB) media in 2 L baffled Erlenmeyer flasks and grown to an OD_{600} of 0.7 with shaking at 200 rpm. Flasks were transferred to a 16 °C shaker and expression was induced for 16 hours with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma Aldrich, cat. No I6758). Unless otherwise noted, all subsequent steps were carried out at 4 °C. Cells were harvested by centrifugation at 4000 × g for 10 minutes and the pellet was resuspended in 10 ml of resuspension buffer (50 mM Tris-HCl pH 7.5 containing 150 mM NaCl and 10% glycerol, protease inhibitor cocktail cOmplete® Roche) per 1 L culture. Cells were lysed in an Avestin EmulsiFlex C5 homogeniser and lysates were clarified by centrifugation at 48,000 g for 30 min. The supernatant was transferred to a 50 ml conical tube containing 2 ml of a 50% slurry of Ni^{2+}-agarose pre-equilibrated with resuspension buffer and absorption of MBP-R-spondin 1 to Ni^{2+}-agarose was carried out for 1 h with gentle agitation. The lysate - Ni^{2+}-agarose mixture was transferred to a

14 cm Econo-Pac column (cat. No 7321010, Bio-Rad), washed thrice with 10 ml of resuspension buffer containing 25 mM imidazole (cat. No I5513, Sigma Aldrich) and MBP-R-spondin 1 eluted in 1 ml fractions of resuspension buffer containing 265 mM imidazole.

Eluted fractions containing MBP-R-spondin 1 were pooled and concentrated using an Amicon Ultra-15 centrifuge filter (cat. No UFC9010, Millipore). Buffer exchange into disulphide reshuffling buffer (50 mM Tris base, pH 8.0 containing 150 mM NaCl, 5% glycerol, 1 mM EDTA, 5 mM reduced glutathione and 1 mM oxidized glutathione) was performed on a PD10 Sephadex G25 desalting column (cat. No GE17-0851-01, GE Healthcare). The eluate was diluted to 1 mg/ml with disulphide reshuffling buffer and incubated for 12 hours at room temperature. The reshuffled protein was injected onto a HiLoad Superdex 200 pg 16/600 column (cat. No GE28-9893-35, GE Healthcare) equilibrated in resuspension buffer and resolved fractions containing MBP-R-spondin 1 (identified by SDS-PAGE) were pooled and concentrated using an Amicon Ultra-14 centrifuge filter. Buffer exchange was carried out using a PD10 Sephadex G25 desalting column equilibrated in IEX buffer (50 mM Na^{2+}/K^{+} phosphate buffer pH 7.0 containing 5% glycerol). At this stage, MBP-R-spondin 1 fusion protein can be analysed to determine concentration (using the Bradford protein assay; Bio-Rad) and used in organoid growth media at a concentration of 25 nM (see Results).

MBP-R-spondin 1 was incubated with 5 U of bovine Thrombin (cat. No T6634, Sigma Aldrich) per mg protein for 8 hours at room temperature and the cleaved R-spondin 1 moiety was resolved using cation exchange chromatography (IEX) with a HiTrap SP HP 5 ml column on an ÄKTA Purifier equilibrated in IEX buffer (15 mM sodium phosphate buffer containing 5% glycerol) with a linear gradient to 1 M NaCl in IEX buffer. R-spondin 1-containing fractions (identified by SDS-PAGE) were pooled, concentrated and buffer exchanged using a PD10 Sephadex column equilibrated in phosphate buffered saline (PBS). The cellular activity of purified R-spondin 1 was determined by Wnt pathway reporter assay.

2.6 Luciferase reporter assay of R-spondin 1 activity

The cellular activities of MBP-R-spondin and R-spondin 1 fractions were determined by their ability to potentiate Wnt3A-induced pathway activity in HEK293T cells, designated as WPC50 (Wnt pathway activity potentiation constant), and defined as the concentration of added R-spondins sufficient for 50% of total Wnt pathway reporter activity. Cells were plated onto 24-well tissue culture plates (75,000 cells per well) in DMEM media supplemented with 10% fetal calf serum (Gibco) and transfected with 100 ng SuperTop and 10 ng pRL-SV40P (Renilla-luciferase) plasmids. Transfected cells were treated with Wnt3A-conditioned media (WCM; produced from Wnt3A-expressing L-cells; ATCC CRL-2646) diluted 1:4 with growth media and varying concentrations of R-spondin 1 (typically 1–150 nM). At 16 hours post-treatment, cells

were harvested, and reporter activities determined with the Promega Stop & Glo kit using a PHERAstar microplate reader. Normalised luciferase activity was determined by dividing firefly luciferase values by the *Renilla* luciferase values. Wnt pathway reporter assay values for each concentration of R-spondin were the average of four individual biological replicates and reported with standard deviation (SD). All WPC50 values were calculated using GraphPad Prism Software Version 8 by fitting the Hill equation to the dose-response data.

2.7 Statistical analysis

All statistic methods used in this thesis are described in the respective chapter method sections. P-values are indicated as asterisks indicating the level of significance. A p-value less than 0.05 is (*), less than 0.01 is (**) and less than 0.001 is (***). Significance threshold was 0.05 in all analyses performed.

Chapter

3 TCF7L2 is post-translationally processed

3.1 Introduction

TCF7L2 has been extensively studied as a Wnt signalling effector of target gene expression However, it is not known how expression levels of the various forms of the protein are regulated and what functions they may play in Wnt pathway target gene expression.

3.1.1 Aim of this chapter

The aims of Chapter 3 are to:

- Define cell-specific expression of the TCF7L2 splice variants in CRC cell lines.
- Determine the mechanism regulating the expression levels of the various TCF7L2 variants

3.2 Materials and methods

3.2.1 End point PCR

Amplification of TCF7L2 variants was performed by end point PCR using cDNA of the CRC cell lines DLD1, HCT116, RKO, LoVo, SW480 and HEK293T cells (as described in Chapter 2.3). PCR products were generated using primers outlined in Table 3 (Chapter 2) with application for splice variant determination and Phusion Hot Start polymerase (NEB) with the following specifications: Per 50 µl reaction 1 µl of cDNA was used along with 200µM dNTPs, 0.5 µM forward and 0.5µM reverse primer, 3 % DMSO in 1x GC-buffer. Amplification was carried out using a Biometra PCR cycler with the following program: 98°C for 30 seconds, 98°C for 5 seconds for denaturation, 65°C for 30 seconds for primer annealing, 72°C for 20 seconds for extension, 72°C for 10 minutes for final extension. This was repeated in 35 cycles followed by hold at 4°C.

PCR products were visualised after agarose gel electrophoresis on a 1% agarose gel using Ethidium Bromide (Sigma-Aldrich) and UV exposure.

3.2.2 Gel extraction of PCR products and sequencing

PCR fragments of TCF7L2 variant analysis were excised from an 1% agarose gel and purified using a gel extraction kit (Qiagen). The amplicons were quantified using a Nanodrop spectrometer and the fragments were sequenced at the sequencing facility at the Department of Biochemistry with primers used for PCR amplification. Obtained sequences were aligned with TCF7L2 cDNA variant sequences downloaded from the UCSC genome browser using Snapgene Software.

3.2.3 Proteasomal inhibition with MG-132

HEK293T cells were transfected with plasmids harbouring the Myc- and FLAG-epitope tagged TCF7L2 and NFκB transgenes as outlined in table 2 (Chapter 2) using PEI lipofection in 6 well plates at 80% confluency. After 24 hours of recovery, cells were treated with the proteasome activity inhibitor MG-132 (Sigma-Aldrich) at a concentration of 10 µM for 10 hours. After treatment cells were lysed and probed with Myc-epitope tag antibody.

3.3 Results

3.3.1 TCF7L2 is post-translationally processed into 55 kDa short form

3.3.1.1 Alternative splicing variants differ from protein variants

TCF7L2 protein is expressed as different variants produced by alternative splicing of the *TCF7L2* gene. The most prevalently expressed TCF7L2 transcript form is the TCF7L2-E variant that harbours the 4 functional protein domains for β-catenin binding at the N-terminus, a HMG box DNA binding domain including the nuclear localisation signal, the DNA binding C-clamp and CtBP binding sites at the very C-terminus (Figure 14A) resulting in a translated protein of 70 kDa. Alternative splicing also gives rise to the shorter TCF7L2-M form through alternative splicing of exons 14-17. TCF7L2-M forms do not contain the C-terminal extension consisting of the C-clamp and C-terminal CtBP interaction domains by virtue of the premature stop codon downstream of the 13-18 exon-exon junction. This shorter form gives rise to a protein of 55 kDa.

I analysed expression of TCF7L2 E- and M-forms by amplification of cDNA from various human colon cancer cell lines and the HEK293T cell line using splice-variant specific primers. An amplicon generated using primers spanning from the 5' β-catenin binding domain to exon 5 (Figure 14B, primer pair A) discriminates the presence of exon 4 that is specific to the TCF7L2E forms. A PCR product of 430 bp indicates the presence of exon 4, whereas the lack of exon 4 yields a product of 361 bp. Amplification of cDNA with primers from exons 9 to 18 (Figure 14B, primer pair B) not only discriminates long TCF7L2-E from short TCF7L2-M, but also TCF7L2E2 from TCF7L2E1/3 by the presence of exon 14-15 or 16, respectively. Here, a PCR product of 627 bp corresponds to TCF7L2E2 whereas a 576 bp amplicon indicates TCF7L2E1/3 and a 503 bp product, the TCF7L2M variant. The amplicon generated using primer pairs A and B from cDNA of various cancer cell lines as well as HEK293T cells in an endpoint PCR showed that there are 3 main splice variants expressed. Primer pair A to discriminate between E and M form reveals the presence of both (Figure 14C). Primer pair B results in three bands in each of the cell lines, indicating the presence of TCF7L2E2, TCF7L2E1/3 and TCF7L2M. Gel extraction of the PCR products and subsequent sequencing confirmed the presence of the variants depicted in Figure 14B.

In order to determine the expression amount of the previously detected splice variants, quantitative RT-PCR was performed with primers for each alternative exon, normalised to all expressed variants represented by the presence of the coding region for the β-catenin binding domain in exon 1 (Figure 14A). The 70 kDa 1 form (TCF7L2E2) together with 70 kDa 2 form (TCF7L2E1/3) were the predominant transcripts in the cell lines ranging from 70% in DLD1 to 89% in HCT116 of the total TCF7L2 transcripts. Expression of the shorter form of 55 kDa (TCF7L2M) is appeared to be expressed in a range from 11% to 30% of total TCF7L2 transcripts (Figure 14D).

Western blot analysis confirmed the fact that both forms of TCF7L2, the 70 kDa and 55 kDa, are also detectable when probing protein lysates of these cell lines with α-TCF7L2 antibody. Intriguingly, the ratio of the long 70 kDa form to the short 55 kDa forms is approximately equal, varying only slightly amongst the cell lines. The exception is the LoVo cell line which harbour a heterozygous truncating mutation in an exon 18 microsatellite leading to expression of the short form of 55 kDa – this is the subject of investigation in Chapter 4 of this work. Although HCT116 cells express the highest amounts of TCF7L2, the ratio between 70 kDa and 55 kDa forms remain consistent with the DLD1, RKO, SW480, and HEK293T cell lines (Figure 14E). Quantification of the protein variant abundance of 3 independent Western blots revealed that the ratio of 70 kDa to 50 kDa TCF7L2 around the 50% for each, with the exception in LoVo cells mainly expressing short TCF7L2 with 80% (Figure 14F). Taken together, expression analysis of the TCF7L2 splice variants correlates with the presence of the 70kDa and 55kDa forms of TCF7L2 protein. However, the data fails to reconcile the relatively equal levels of TCF7L2 protein with the expression ration of their encoding transcripts.

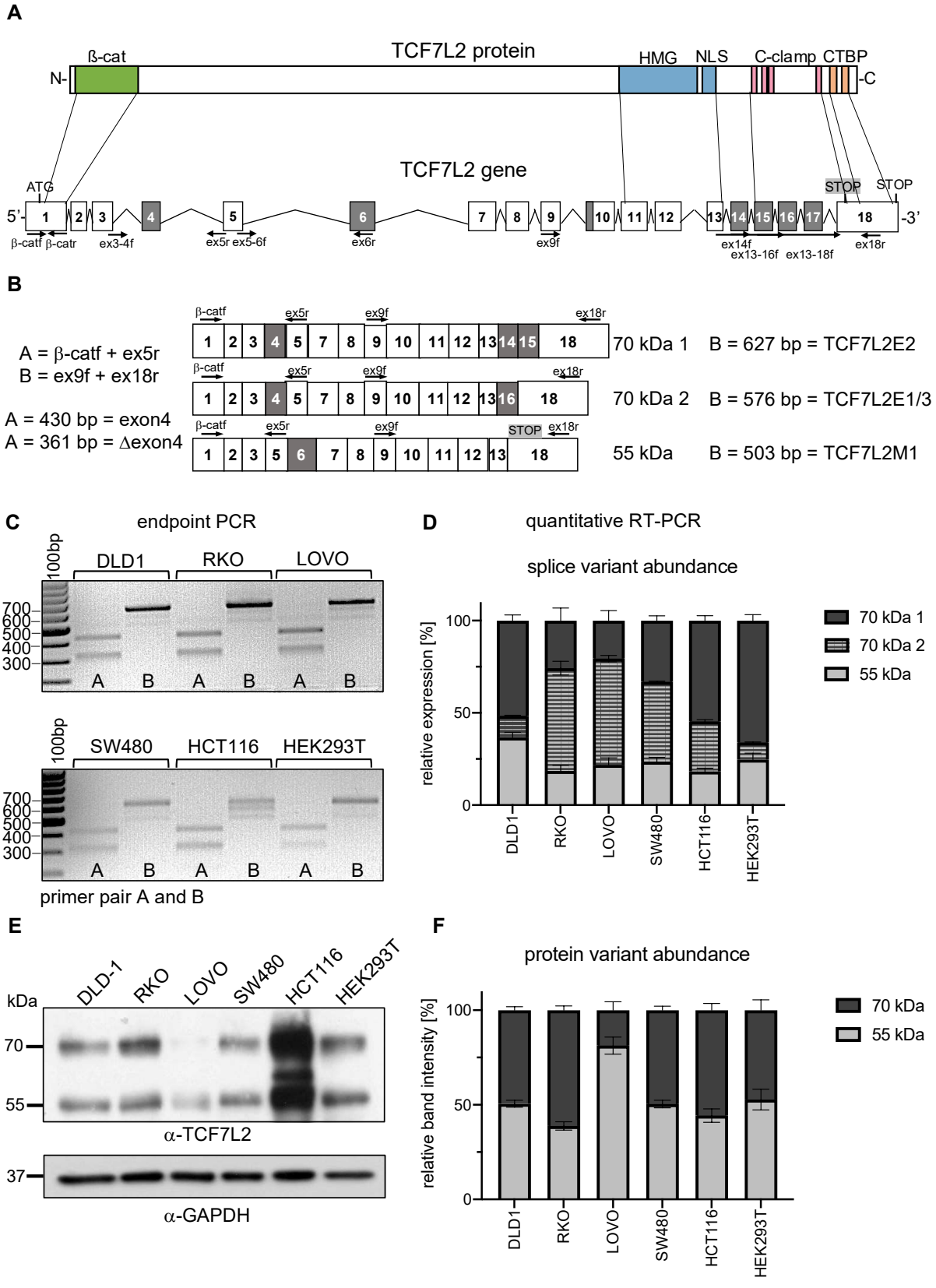

A
ß-cat
TCF7L2 protein
HMG
NLS
C-clamp
CTBP
N-
-C
TCF7L2 gene
ATG
STOP
STOP
5'
-3'
β-catf β-catr
ex3-4f
ex5r ex5-6f
ex6r
ex9f
ex14f
ex13-16f
ex13-18f
ex18r
B
A = β-catf + ex5r
B = ex9f + ex18r
A = 430 bp = exon4
A = 361 bp = Δexon4
70 kDa 1
B = 627 bp = TCF7L2E2
70 kDa 2
B = 576 bp = TCF7L2E1/3
55 kDa
B = 503 bp = TCF7L2M1
C
endpoint PCR
100bp
DLD1
RKO
LOVO
SW480
HCT116
HEK293T
700
600
500
400
300
A
B
primer pair A and B
D
quantitative RT-PCR
splice variant abundance
relative expression [%]
100
50
0
70 kDa 1
70 kDa 2
55 kDa
E
kDa
70
55
α-TCF7L2
37
α-GAPDH
F
protein variant abundance
relative band intensity [%]
70 kDa
55 kDa

Figure 14: Variation between in the relative levels of the TCF7L2 splice variants and the encoded proteins

(A) TCF7L2 protein structure aligned with TCF7L2 gene structure. TCF7L2 protein consists of β-catenin binging domain (β-cat, green) at N-terminus, HMG box DNA binding domain (HMG, blue), nuclear localisation signal (NLS, blue), C-clamp (pink) and CTBP binding domain (orange). Gene consists of 18 exons with alternatively spliced exons in grey (exon 4,6,14-17). Stop codons are indicated with if created by alternative splicing underlaid in grey. Primers for endpoint PCR as well as RT-PCR for splice variant abundance determination are depicted underneath the gene with arrows corresponding to their binding/spanning position. (**B**) Schematic composition of most abundant splice variants and primer composition for endpoint PCR for variant detection. Alternative exons are underlaid in grey and primer positions are marked by corresponding arrows. Presence of exon 4 indicates TCF7L2E forms, absence indicates TCF7L2M form. TCF7L2E forms are discriminated by either exon 14-15 or exon16 for TCF7L2E2 or TCF7L2E1/3, respectively. (**C**) Expression analysis of splice variants in different colon cancer cells and HEK293T cells. After RNA isolation cDNA was synthesised, and TCF7L2 was amplified with primers indicated in (B). agarose gel electrophoresis separated PCR products and fragments visualised. PCR product size was analysed and assigned to the expected fragment size as indicated. DNA fragments were excised, purified and sequenced to confirm their exon combination. (**D**) RT-PCR with indicated exon specific primers as depicted in (A). GAPDH primers were used as housekeeping gene to ensure cDNA quantity and integrity. Splice variant abundance was calculated using β-catenin binding site (β-catf+r) primer amplification cycle number (CT values) as relation to exon specific primer amplification cycle number. Sum of all variants was set 100% and variant expression was calculated for each cell line. TCF7L2E2 is 70 kDa 1, TCF7L2E1/3 is 70 kDa 2 and TCF7L2M is 55 kDa. (**E**) Expression analysis of TCF7L2 protein in colon cancer cells and HEK293T cells. After cell lysate was generated using RIPA buffer 40 µg of protein was subjected to SDS PAGE and transferred onto PVDF membrane. Probing with α-TCF7L2 antibody results in a 70 kDa and a 55 kDa protein form in each of the cell lines, with LoVo cells mostly expressing the short form of TCF7L2. GAPDH antibody was used as loading control. (**F**) Quantification of 3 independent Western blots from (E). Band intensity was measured using ImageJ and protein variant ratio from 70 kDa/55 kDa was calculated and depicted in % of total expressed protein. Each experiment was carried out in 3 replicates.

3.3.2 TCF7L2 is processed post-translationally

The discrepancy between relative ratios of alternatively spliced TCF7L2 transcripts and encoding protein levels lead me to investigate whether there was some alternative regulation of transcript or protein processing. I created transgenic versions of TCF7L2E and TCF7L2M that contained an N-terminally myc-tag. For size comparison the short TCF7L2 version lacking the C-terminal extension (amino acids 1-485) tagged with an N-terminal myc epitope tag and a C-terminal FLAG tag (Figure 15A). Transfection of these cDNA clones into HEK293T cells and subsequent Western blotting experiments, using α-myc antibodies showed both protein sizes of 70 kDa and 55 kDa corresponding to full length and short TCF7L2-M in myc-TCF7L2 expression samples. Importantly, the transgenic version of TCF7L2-E and TCF7L2-M do not contain introns, excluding alternative splicing and indicating that the long form is processed to the short form post-translationally.

To determine whether a site-specific protease was responsible for processing of the expressed myc-TCF7L2 protein to the short form, I created a modified myc-TCF7L2 transgenic expression vector to include a C-terminal FLAG epitope tag. In this way, I could identify the approximately 15 kDa C-terminus of myc-TCF7L2-FL-FLAG by Western blot. Surprisingly, in cells transfected with myc-TCF7L2-FL-Flag, only TCF7L2 full length was detected; the inclusion of the C-terminal FLAG-tag abrogated processing. The shorter

transgenic version of myc- and FLAG-epitope tag version of TCF7L2 was detected at its expected size of 55 kDa in myc-TCF7L2-short-FLAG expressing cells (Figure 15B). The size of this short version of TCF7L2 corresponds to the TCF7L2M form, narrowing the potential processing site to a range between amino acid 480-490, perhaps involving the 'KKK' motif encoded by the exon 18 microsatellite. The processed short version of TCF7L2 lacks the C-terminal extension containing the CtBP interaction domain and part of the C-clamp.

3.3.3 TCF7L2 processing is independent from the proteasome

The domain structure of TCF7L2 is similar to that of the transcription factors Gli3 and NFκB, containing an N-terminal regulatory domain, an HMG DNA, an adjacent structured domain followed by a proline and serine rich sequence, followed by a stretch of lysine residues and phosphorylation sites within a potential degron sequence. Interestingly, both Gli3 and NFκB share a unique processing mechanism that converts a functional long form of the transcription factor into a shorter form with unique transcriptional properties – both have been shown to be post-translationally processed by the proteasome from the C-terminus, resulting in an active short form of NFκB and a repressive short version of Gli3.

To test the hypothesis that TCF7L2 is processed by the proteasome, I generated plasmids for transgenic expression of NFκB, in the same format as for TCF7L2, containing either an N-terminal myc tag or both the N-terminal myc tag and a C-terminal FLAG tag. Nuclear localisation of all transcription factors was confirmed using fluorescent staining probing with myc-tag antibody (data not shown).

Myc-TCF7L2, myc-TCF7L2-FL-Flag, myc-NFκB as well as myc-NFκB-Flag expression plasmids were transfected into HEK293T cells. After 24 hours of recovery, cells were treated with the proteasome inhibitor MG-132. At 10 hours post-treatment, cells were lysed, and protein lysates were probed with α-myc antibody. In cells transfected with myc-NFκB two bands were detected corresponding to full length 105 kDa as well as the processed 50 kDa form. In samples treated with the proteasomal inhibitor MG-132 processing was significantly impaired indicated by the loss of intensity of the 50 kDa band, of approximately 90%. Notably, myc-NFκB-Flag was processed to the same degree as the myc- NFκB protein and proteasomal processing of myc-NFκB-Flag was likewise inhibited by MG-132 treatment (Figure 15C left). In contrast, the two bands at 70 kDa and 55 kDa, corresponding to full length and processed myc-TCF7L2, are maintained in samples from cells treated with MG-132 at approximately the same ratio as untreated cells. Additionally, full-length TCF7L2-Flag appeared to be stabilised under MG-132 proteasomal inhibition through an unkown process (Figure 15C right). My data indicates that TCF7L2 undergoes post-translational processing through a unique proteasome-independent mechanism (Figure 15C right), that needs to be confirmed by further experiments with other proteasomal inhibitors.

3.3.4 TCF7L2 processing relies on specific sequences within the protein.

In order to determine the mechanism of TCF7L2 post-translational processing, I interrogated sufficiency of conserved amino acid features that are unique to the protein. The other long form of the TCF/LEF family, TCF7, contains C-terminal extension similar to that of TCF7L2 and is also expressed as long and short protein variants. *Drosophila* express a single TCF/LEF family member, Pangolin (Pan) that contains the C-terminal extension and corresponds to the long form of TCF/LEF family proteins. Alignment of TCF7L2, TCF7 and Pan at the region targeted by the post-translational processing activity indicates a sequence unique to TCF7L2, 'CT-internal', and a stretch of three lysine residues 'KKK' common to the 'long' forms of TCF7L2 and TCF7 (Figure 15D).

Mutant versions of TCF7L2 were created as follows:

Mutant 1 - Changing the KKK amino acids to isoleucine-proline-alanine (KKK→IPA)

Mutant 2 - Deleting the 5 amino acids at the C-terminus (ΔCT 5aa)

Mutant 3 - Deleting the TCF7L2-specific CT-internal sequence (ΔCT-internal).

Western blot analysis of expressed mutants 1-3 indicated all three versions were impaired for post-translational processing (Figure 15E), while still being located nuclear (data not shown).

I also determined whether TCF7 was processed, similar to TCF7L2 by expressing the long and short forms of the protein. Western blot analysis indicates that expressed 'long; version of TCF7 is not processed post-translationally demonstrated by the lack of detectable protein at the expected size of the 'short' version (42 kDa; Figure 15F). Therefore, the expressed protein variants of TCF7 are likely due to alternative splicing, as previously reported. *Drosophila* Pan failed to express in mammalian cells and could therefore not been investigated for processing.

I conclude that TCF7L2 is potentially processed by a novel post-translational mechanism.

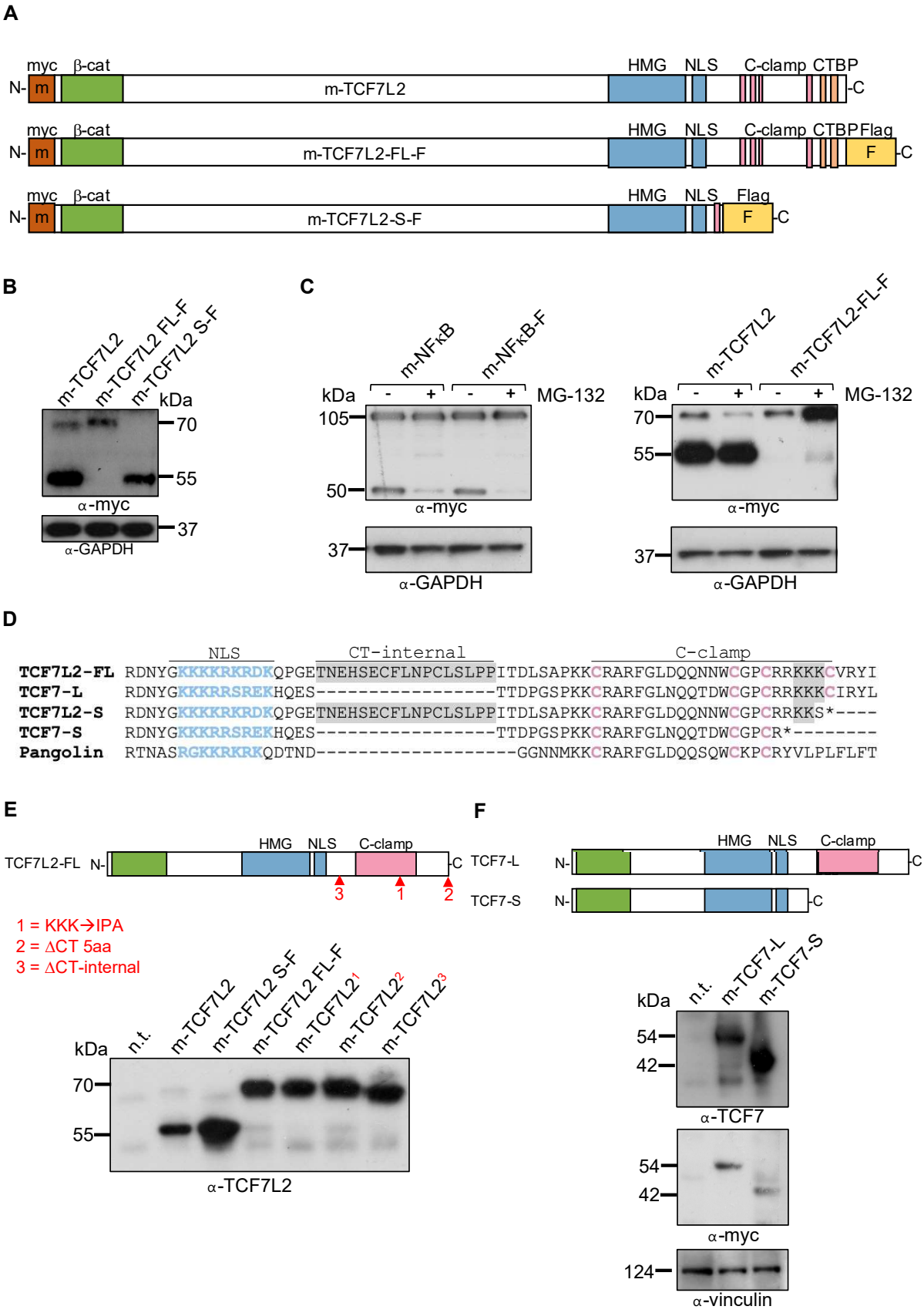

A
myc β-cat
HMG NLS C-clamp CTBP
N- m m-TCF7L2 -C
myc β-cat
HMG NLS C-clamp CTBP Flag
N- m m-TCF7L2-FL-F F -C
myc β-cat
HMG NLS Flag
N- m m-TCF7L2-S-F F -C
B
m-TCF7L2
m-TCF7L2 FL-F
m-TCF7L2 S-F
kDa
70
55
α-myc
37
α-GAPDH
C
m-NFκB
m-NFκB-F
kDa - + - + MG-132
105
50
α-myc
37
α-GAPDH
m-TCF7L2
m-TCF7L2-FL-F
kDa - + - + MG-132
70
55
α-myc
37
α-GAPDH
D
NLS
CT-internal
C-clamp
TCF7L2-FL RDNYGKKKKRKRDKQPGETNEHSECFLNPCLSLPPITDLSAPKKCRARFGLDQQNNWCGPCRRKKKCVRYI
TCF7-L RDNYGKKKRRSREKHQES----------TTDPGSPKKCRARFGLNQQTDWCGPCRRKKKCIRYL
TCF7L2-S RDNYGKKKKRKRDKQPGETNEHSECFLNPCLSLPPITDLSAPKKCRARFGLDQQNNWCGPCRRKKS*----
TCF7-S RDNYGKKKRRSREKHQES----------TTDPGSPKKCRARFGLNQQTDWCGPCR*--------
Pangolin RTNASRGKKRKRKQDTND----------GGNNMKKCRARFGLDQQSQWCKPCRYVLPLFLFT
E
TCF7L2-FL N- HMG NLS C-clamp -C
3
1
2
1 = KKK→IPA
2 = ΔCT 5aa
3 = ΔCT-internal
kDa
n.t.
m-TCF7L2
m-TCF7L2 S-F
m-TCF7L2 FL-F
m-TCF7L2 1
m-TCF7L2 2
m-TCF7L2 3
70
55
α-TCF7L2
F
TCF7-L N- HMG NLS C-clamp -C
TCF7-S N- -C
kDa
n.t.
m-TCF7-L
m-TCF7-S
54
42
α-TCF7
54
42
α-myc
124
α-vinculin

Figure 15: TCF7L2 is processed post-translationally independent of the proteasome and depending on TCF7L2 specific intrinsic sequences
(**A**) Schematic presentation of TCF7L2 expression clones (not to scale). Myc-TCF7L2 (m-TCF7L2), myc-TCF7L2-full length-Flag (m-TCF7L2-FL-F) and myc-TCF7L2-short-Flag (m-TCF7L2-S-F) were generated to investigate TCF7L2 processing. Myc tag illustrated in brown, β-catenin binding domain (β-cat) in green, Homeobox mobility group (HMG) DNA binding domain and nuclear localisation signal in blue, C-clamp DNA binding domain in pink, C-terminal binding protein (CTBP) interaction sites in orange and Flag-tag in yellow. (**B**) HEK293T cells expressing m-TCF7L2, m-TCF7L2-FL-F or m-TCF7L2-S-F were lysed, and protein detected by probing with α-myc antibody. m-TCF7L2 expressing cells show 70 kDa and 55 kDa band consistent with processing products. m-TCF7L2-FL-F only shows one band of 70 kDa indicating no processing and m-TCF7L2-S-F shows one band of 55 kDa confirming the size of short TCF7L2. (**C**) myc-NFκB (m-NFκB) and myc-NFκB-Flag (m-NFκB-F) expression plasmid were transfected into HEK293Ts and proteasomal activity inhibited by MG-132 treatment. Protein was detected my probing with α-myc antibody after lysis and revealed processing of NFκB without MG132 treatment resulting in a band of 105 kDa and 50 kDa. Processing was inhibited in cells treated with MG-132 indicated by mostly one band of 105 kDa present. myc-TCF7L2 (m-TCF7L2) and myc-TCF7L2-FL-Flag expressing HEK293T cells treated with MG-132 as well and protein detected by α-myc antibody. m-TCF7L2 showed two bands, at 70 kDa and 55 kDa, confirming post-translational processing in cells not treated with MG-132. Cells treated with proteasomal inhibitor also show 2 bands at 70 kDa and 55 kDa, suggesting processing was not inhibited. Expression of m-TCF7L2-FL-F revealed no processing by the presence of only 70 kDa protein without MG-132, and inhibition of proteasomal activity mostly shows full length protein with a faint band at 55 kDa. (**D**) Protein sequence alignment of TCF7L2-FL, TCF7L2-S, TCF7-L, TCF7-S and TCF7L2 fly homolog Pangolin. NLS indicated in blue, TCF7L2 specific additional sequence underlaid in gray, C-clamp indicated by Cysteines in pink. Microsatellite resulting in tripe lysine stretch KKK underlaid in gray. (**E**) Protein expression analysis of TCF7L2 mutants as indicated in scheme. Mutant 1: substitution mutants of KKK→IPA, mutant 2: deletion of last 5 amino acids from C-terminus ΔCT 5aa, mutant 3: deletion of CT-internal sequence ΔCT-internal. Positive control for processing activity m-TCF7L2, negative control m-TCF7L2-FL-F, size control for processing product m-TCF7L2-S-F. All mutants resulted in bands of only 70 kDa when probing of cell lysate with α -TCF7L2 antibody. (**F**) Analysis of processing activity in cells expressing TCF7-long or TCF7-short. Cells transiently expressing either TCF7-L or TCF7-S were lysed, and protein detected with α-TCF7 or α-myc antibody. TCF7-L expressing cells showed one band of 54 kDa corresponding to long TCF7, TCF7-S expression cells showed one band of expected 42 kDa, suggesting that TCF7 is not post-translationally processed. Each experiment was carried out in 3 replicates.

3.4 Discussion

3.4.1 TCF7L2 variants generated by alternative splicing differ from protein variants

My investigation of cell-specific TCF7L2 isoform expression led me to conclude that two of the TCF7L2 E-forms are the prominent isoforms in colorectal cancer cell lines and HEK293T cells. I could also detect one of the TCF7L2M-forms which was expressed at 4-fold to 10-fold less in cell lines. The relative expression levels of the TCF7L2 transcripts expression were in contrast to protein levels of the TCF7L2 variants. I could detect two TCF7L2 protein species of 70kDa and 55kDa corresponding to TCF7L2E and TCF7L2M forms. These observations led me to question whether transcriptional splicing was the predominant mechanisms leading to a relatively even expression levels 70 kDa and 55 kDa versions of TCF7L2.

3.4.2 TCF7L2 is processed post-translationally

Expression of the 'long' 70 kDa encoding transgenic myc-TCF7L2 led to two TCF7L2 protein species, the 70 kDa and 55 kDa forms. As the transgene did not contain introns, I excluded transcriptional splicing as a mechanism and hypothesised that TCF7L2 is processed post-translationally. The accepted paradigm has been that alternative splicing explains the relative expression of different TCF7L2 variants. Therefore, the identification of a TCF7L2 post-translation processing machinery is a key step forward in our understanding of the regulation of Wnt pathway target gene expression and its biological outputs.

3.4.3 Processing of TCF7L2 requires intrinsic sequence signals and is independent of the proteasome

I determined whether TCF7L2 was processed by a similar mechanism to Gli3 and NFκB which relies on the proteasome [161]. This wasn't the case; while proteasome inhibition attenuated NFκB processing, it had no effect on TCF7L2 expression. One phenomenon I observed was the enrichment of the processed 55 kDa TCF7L2 in MG-132 treated samples. This may be due to accumulation of components of the post-translational processing machinery. Another unusual result was the enhanced stabilisation of unprocessed Flag-tagged full length TCF7L2 under MG132 treatment, which could potentially indicate a proteasomal contribution to TCF7L2 turnover. With both of the phenomenons, their identification will provide an interesting avenue of research once the underlying mechanisms have been identified. As a first step, my initial studies were geared towards identifying sequence determinants within TCF7L2 that were responsible for its processing. I showed that the addition of a FLAG-tag at the C-terminus abrogates processing and created a series of mutations that implicates the C-terminus, the 'KKK' motif and CT-internal sequences as critical. In particular, the CT-internal sequence is unique to TCF7L2 amongst all TCF/LEF family members, is juxtaposed to the processing site and may therefore be key for the processing activity. Future studies will include further

mutagenesis studies, for instance inserting the CT-internal sequence into TCF7 and testing whether this leads to a 'processable' version of the protein. It will also be important to refine the initial studies and identify the exact site of processing, accompanied by protein mass spectrometry studies.to determine the exact scissile site.

Another potential TCF7L2 processing mechanism that will be examined is self-cleavage. Self-cleavage of a transcription factor is a well-recognised phenomenon, e.g. the ER bound protein MYRF harbours a chaperone domain capable of catalysing auto-proteolysis leading to translocation of the N-terminal moiety into the nucleus and activating gene expression [162]. Experimentally, this would involve the determination of TCF7L2 protein self-cleavage *in vitro* using purified proteins.

Altogether, identifying the post-translational processing machinery and underlying mechanism will be a fruitful future avenue of research.

3.4.4 Conclusion

The principal mechanism regulating the expression ratio of the short and long forms of TCF7L2 could be post-translational processing, owing to intrinsic sequence features unique to this protein. In addition, a short variant of TCF7L2 is produced by a truncating mutation, which genetic and biological function is investigated in the next chapter.

Chapter

4 TCF7L2 defines a novel CRC subset

4.1 Introduction

4.1.1 Aim of this chapter

- Define the frequency of A9 microsatellite mutations in CRC cohorts via bioinformatic analysis of large, publicly accessible datasets.
- Characterise the transcriptional and biological output of altering expression levels of TCF7L2 variants. I will establish an inducible transgenic system that enables control over the relative expression levels of the TCF7L2E and TCF7L2M variants and characterise the consequence on cell proliferation, clonogenicity, cell cycle progression, survival and Wnt pathway target expression.

4.2 Materials and Methods

4.2.1 Bioinformatic analysis

4.2.1.1 Analysis of mutation frequencies in CRC

Mutation and clinical data from a targeted sequencing of a large cohort (1134 CRC samples from 1099 patients) of colorectal cancer tumours (the MSKCC cohort) [163] was analysed using the program R. All programming codes used for R are listed in Appendix 2.

TCF7L2 mutations were analysed using cBioPortal tools (cBioportal.org) and data was presented with Graph Pad Prism 8. The frequency of *TCF7L2*, *APC* and *BRAF* mutations were analysed using cBioPortal by plotting gene name (y-axis) versus MSI Score (x-axis), and relevant mutations were highlighted as protein change. Analyses were visualised using Graph Pad Prism.

The cBioportal oncoprinter tool was used to determine genetic interactions amongst known CRC driver mutations in *TCF7L2*, *APC* and *BRAF*.

Mutational data and clinical data files were downloaded from the cBioportal server (https://www.cbioportal.org/study/clinicalData?id=crc_msk_2017) imported into R and merged according to tumour sample barcodes. Quality control at this step included the elimination of overlapping data. Frequency of mutations co-occurring with MSI TCF7L2 frameshift or microsatellite indels (K485Sfs*23) were calculated and depicted alongside other driver mutations.

4.2.1.2 Kaplan Meier survival probability

Patient survival data from the MSKCC cohort was merged with mutational data according to tumour sample barcode and filtered for Kaplan Meier survival probability assessment by assigning the following parameters to each sample: living status was scored as 10, Deceased status was set to 1, absence of mutation was set to 0 and presence of mutation was set to 1. Filtering was applied to TCF7L2 A9 microsatellite mutations and BRAFV600E missense mutation. Survival probability of the MSKCC cohort was plotted in a Kaplan Meier curve using the ggsurvplot package.

4.2.1.3 Transcriptome profiling of CRCs with TCF7L2 A9 microsatellite mutations

First, mutational characteristics of MSKCC data was verified in TCGA COAD dataset.

Stratification of gene expression in CRC cases harbouring specific mutational signatures was carried out using the Bioconductor packages analysis tool TCGABiolinks. The MSKCC dataset was not linked to gene expression data. I therefore analysed transcriptome profiles of CRC cases containing TCF7L2 A9 microsatellite mutations from the TCGA COAD dataset. Data was queried for the data category “Transcriptome Profiling” and data type “gene expression quantification” from experimental strategy “RNA seq” and workflow type “HTSeq-Counts”.

Files were downloaded from the TCGA database (https://portal.gdc.cancer.gov/projects/TCGA-COAD) and a data matrix generated (COADmatrix). This dataset was then merged with MSKCC patient data and gene expression specific to cases with truncated TCF7L2 was quantified using the BioConductor DESeq2 package. Differentially expressed genes were visualised using the BioConductor enhanced volcano package. Lists of differentially expressed genes along with fold change, *p*-value (t-test derived) and adjusted *p*-value (Benjamini-Hochberg FDR correction) was then exported and saved as excel file for use in pathway enrichment analysis (below).

4.2.1.4 Pathway enrichment analysis

To identify categories in cellular processes of differentially expressed Gene Ontology enrichment analysis was performed using the Gene ontology enrichment analysis and visualisation tool (GOrilla). Following the instructions on the website an adjusted gene list was uploaded with log2 fold change >0.585 and a tree map was generated for cellular process analysis. Functional differences attributed to specific gene expression signatures was performed using Bioconductor ReactomePA and GSEA functions.

4.2.2 Generation of TCF7L2 expressing cell lines

4.2.2.1 Cloning of inducible TCF7L2 expression plasmids

Inducible expression of either full length or truncated TCF7L2 was achieved by transferring myc-TCF7L2-FL-FLAG or myc-TCF7L2-K485S-fs-FLAG from a pcDNA3 backbone (de la Roche laboratory) into the tet-on system of the piggyBac expression vector (de la Roche laboratory) by Gibson Assembly® (ThermoFisher). An inducible expression system was generated with an N-terminal myc epitope tag and C-terminal FLAG-tag, either encoding full length TCF7L2, or truncated TCF7L2 ending with a Lysine to Serine substitution at position 485.

4.2.2.2 Generation of stable cell lines with tuneable expression of TCF7L2 transgenes

The microsatellite instable CRC cell lines HCT116 and RKO as well as HEK293T cells were plated in 6 well plates at $6x10^5$ cells per well and cultured to approximately 80% confluency. Cells were transfected with 1 µg piggyBac expression plasmid encoding the transgene and the fluorescence marker mCherry, 1 µg rtTA plasmid also encoding a Hygromycin resistance gene and 0.25 µg *piggybac* transposase per well using either Lipofectamine™ 2000 (Invitrogen) for HCT116 and RKO according to manufacturer's instructions or polyethylenimine (PEI, Sigma) for HEK293T. For PEI transfection, 5 µl of PEI was used per 1 µg of DNA. Lipid DNA complexes were incubated at ambient temperatures for 20 minutes, then dispensed onto cells in complete culture media. The next day media was changed to

fresh culture media. Three days after transfection Hygromycin B was added to the cultures to select for rtTA containing clones - 100 µg/ml for HCT116 and HEK293T and 150 µg/ml for RKO.

After 10-14 days of Hygromycin B selection, expression of full length or truncated myc-TCF7L2 and mCherry was induced by addition of 1 µg/ml Doxycycline for 48h with one dish as a non-induced control. mCherry expression was confirmed by fluorescence microscopy and cells were trypsinised and stored in 10 % FBS/PBS on ice after filtering through a 35 µm cell strainer mesh cap. Cells expressing mCherry were sorted using a Sony SH800S cell sorter. For each transgenic cell lines, 50000 cells were collected into wells of a 12-well plate containing complete culture media. Confirmation of established cell lines was further confirmed via Western blotting using antibodies raised against the the myc epitope tag.

4.2.3 Proliferation assay

HEK293T, HCT116 and RKO cell lines containing the full length or truncated TCF7L2 transgenes were seeded into 6 well plates at densities of $3x10^4$ cells per well. Transgene expression was induced after 24 hours culture by addition of 1 µg/ml Doxycycline to each well. Media containing Doxycycline (Dox) were changed every 48 h and cell growth monitored over a 7-day time period until confluent. Each cell line, along with no Dox controls, was analysed in three technical and biological triplicates and relative growth rate was calculated relative to controls. Significance was assessed using multiple grouped t-tests.

4.2.4 Cell clonogenic assays

Clonogenic assays were carried out in 6-well plates containing cells seeded at a density of 300 cells per well. After 24 hours of culture, transgene expression was induced by addition of 1 µg/ml Doxycycline/well. Media were changed every 48h and cell growth monitored over a 14 and 21-day time period. Cells were fixed in 100 % methanol for 5-10 minutes and then stained with a 0.5 % crystal violet solution containing 20% methanol for 10 minutes at RT. Excess stain was removed by washing plates with tap water prior to imaging. Colonies were counted in ImageJ using via the particle analysis tool and data analysed in Excel as the number of colonies for expressed TCF7L2 variants relative to non-expressing controls. Each cell line was analysed in three technical and biological triplicates. Significance was assessed using multiple grouped t-tests.

4.2.5 Cell cycle analysis

For determination of steady-state cell cycle phase distribution $1x10^6$ HEK293T, HCT116 or RKO cells either expressing full length TCF7L2 or truncated TCF7L2 or no Dox control were treated for 24 hours harvested by trypsin incubation (1x Trypsin-EDTA), washed once in PBS

and pelleted by centrifugation at 350xg for 4 min. Cells were dropwise resuspended in 70% ice cold ethanol and stored at -20°C for 3 days minimum to quench mCherry fluorescence. Cells were centrifuged at 850 x g for 3 min, resuspended in 1 ml cold PBS, filtered through a 100 µm mesh cap and centrifuged again at 850 x g for 3 min. Pellets were resuspended in 500 µl PBS containing 100µg/ml RNase A (Sigma Aldrich) and 25 µl of 1 mg/ml Propidium Iodine solution (PI) was added to yield a final concentration of 50 µg/ml. After a 30-minute incubation at ambient temperature, cells were analysed on a BD Accuri C6 Plus Flow Cytometer and the cell cycle phase distribution was annotated using FlowJo software via its model fitting tool. Each cell line was analysed in three technical and biological triplicates. Significance was assessed using multiple grouped t-tests.

4.2.6 Annexin V staining

To analyse the proportion of steady state numbers of apoptotic cells within cell populations Annexin V staining using the BD Bioscience Annexin V-FITC Kit was performed according to the manufacturers protocol. Briefly, cells grown in 6 well plates were trypsinised and washed twice with PBS, then resuspended in 1x binding buffer at $1x10^6$ cells/ml. A volume of 100 µl of the cell solution ($1x10^5$ cells) was transferred into 1.5 ml reaction tubes and 5 µg of recombinant Annexin V was added. After 15-minute incubation at ambient temperature 5 µl Annexin V-FITC was added and cells were incubated for 15 minutes at ambient temperature in the dark. A volume of 400 µl of 1x binding buffer was added and cells were analysed within 1 hour using the BD Accuri C6 flow cytometer. For half of the samples live/dead staining was performed using eFlour780 fixable viability dye (1:1000, ThermoFisher). Before starting the Annexin protocol, cells were resuspended in PBS and stained with eFlour780 in 1:1000 dilution for 30 minutes at 4°C in the dark and washed once with 1x Annexin binding buffer. Proportions of apoptotic cells were evaluated using FlowJo software gated on viable mCherry expressing cells comparing FITC signal intensities. Significance was assessed using multiple grouped t-tests.

4.3 Results

4.3.1 The *TCF7L2* exon 18 microsatellite is a prevalent mutational target in MSI CRC

4.3.1.1 Truncating mutations of TCF7L2 define a unique colon cancer subtype

The stretch of three lysines I investigated in Chapter 3 is encoded by a stretch of 9 adenine residues (A9) adjacent to the splice acceptor site in *exon18* of the *TCF7L2* gene. The A9 microsatellite has been reported as a frequent target for mutation in MSI colorectal cancers [158]. However, the molecular and functional impact of the A9 mutations on TCF7L2 gene function, expression and protein function have not been elucidated.

I analysed CRC mutational data from a recent targeted sequencing study (MSKCC, Cancer Cell, 2018, cBioPortal), combined with the TCGA COAD data cohort for TCF7L2 mutations and genetic interactions (Figure 16). Amongst the 1134 patient samples analysed, I identified TCF7L2 mutations in 108 cases; 74 were truncating mutations, 27 missense mutations, 4 multiple hit mutations, 2 inframe mutations and one TCF7L2 fusion. There were 812 cases that had no TCF7L2 mutations, and 214 cases were not profiled (Figure 16A). TCF7L2 truncations were the most prevalent mutation, comprised approximately 70% of all TCF7L2 mutations including the A9 mutations, followed by missense mutations (25%). In contrast, multiple hit -, inframe- and fusion mutations are underrepresented (Figure 16A).

I determined whether tumours with *TCF7L2* truncating mutations displayed microsatellite instability. Therefore, a MSI Score above 10 was considered MSI, below 10 MSS [164]. Using cBioportal's plot function filtering for MSI score I found that the majority of cases displayed MSS (67% of all truncating mutations). However, specifically, the TCF7L2 truncating mutation K485Sfs* which appeared to be the most frequent truncating mutation within the A9 microsatellite, was invariably in CRC cases with MSI high scores (87% of MSI cases are K485Sfs*).

The vast majority of CRCs contain oncogenic APC mutations (>95%). As expected, my analysis indicated APC truncating mutations mainly appear in MSS CIN distal cancers and are found on the lower MSI Score scale (96%), whereas a fourth of all BRAFV600E missense mutations were MSI (25%). Taken together, TCF7L2 truncating microsatellite mutations, especially K485Sfs* mutations, appear in cases that are MSI positive (Figure 16B).

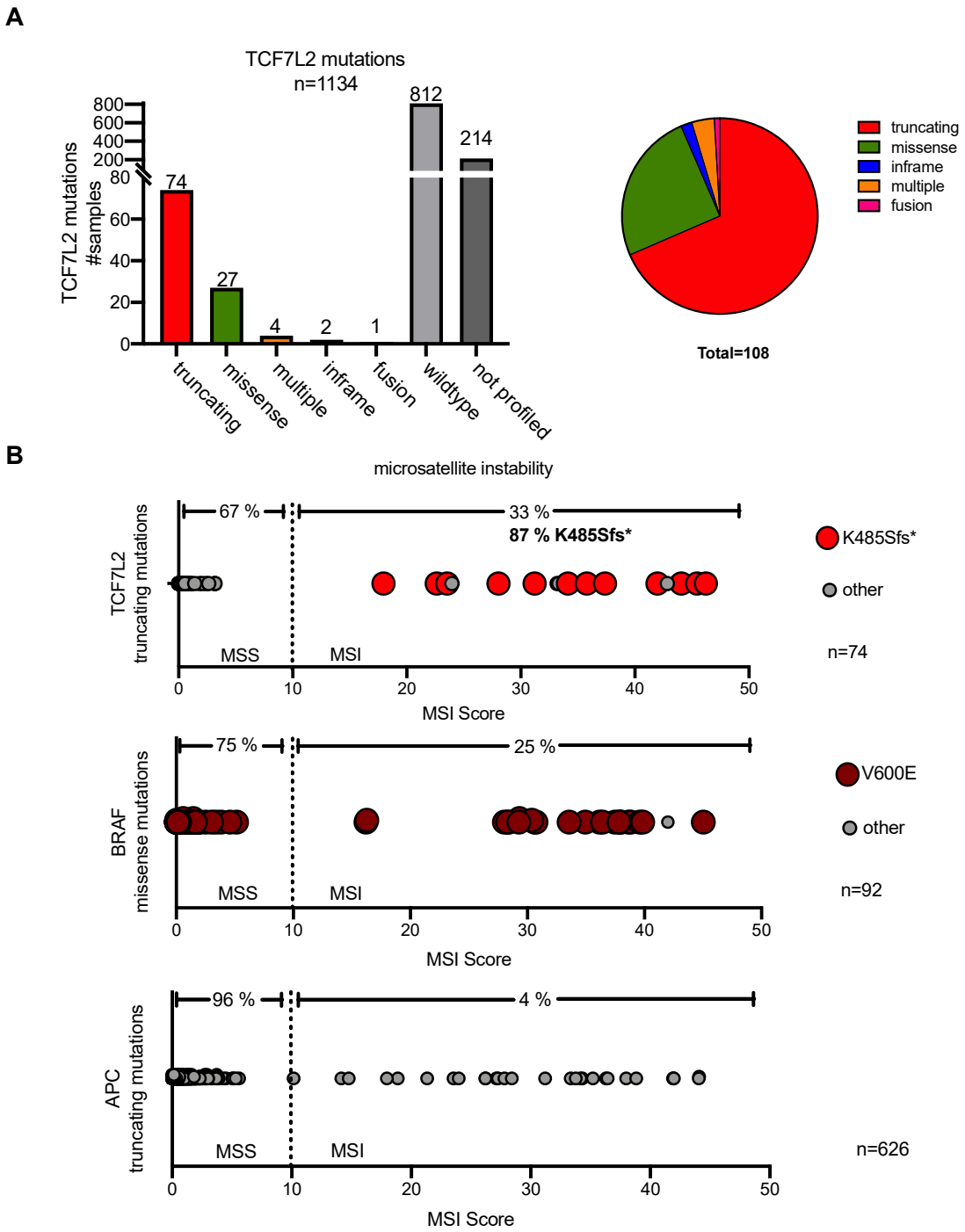


Figure 16: TCF7L2 mutations in colon cancer and their microsatellite instability.
(**A**) quantitative analysis of TCF7L2 mutations in 1134 samples (targeted sequencing of metastatic colon adenocarcinoma MSKCC, cBioPortal) reveal a frequency of 9.5% with mainly truncating mutations. Other mutations are indicated as colour coded bars, light grey and dark showing wildtype TCF7L2 or not profiled samples (left). Pie diagram visualisation of relative abundance of occurring TCF7L2 mutation types. (**B**) MSI Score plotted against TCF7L2 truncating mutations, BRAF missense mutation or APC truncating mutations. TCF7L2 K485Sfs* highlighted in red, BRAFV600E in dark red large points.

I determined the coincidence and exclusivity of TCF7L2, BRAF and APC mutations within individual CRC tumours using cBioPortal's online algorithm by customising the filter for MSI BRAFV600E mutations as well as truncating TCF7L2 mutations and truncating APC mutations. These data were then plotted using the Oncoprint function and *p-values* calculated (derived from one-sided Fisher Exact Test). As expected, BRAF missense mutations are mutually exclusive to APC truncating mutations (*p<0.001*). Interestingly, *TCF7L2 A9* mutations showed a significant co-occurrence with APC truncating mutations (*p=0.002*) and were mutually exclusive with BRAFV600E missense mutations (*p=0.026*) (Figure 17A). Whereas CRCs with *TCF7L2 A9* mutations co-occur with oncogenic APC mutations in approximately 80% of cases, the remaining 20% of cases invariably harbour mutations in Wnt signalling components such as β-catenin or other Wnt pathway regulatory components. The characteristic mutational signature of *TCF7L2 A9* mutations suggests a unique, stratified, subset of CRCs, as most of the APC/Wnt mutated cancers don't display MSI, which is the definition of the described TCF7L2 A9 mutation.

I also compared mutational signatures between MSI CRCs with *TCF7L2 A9* mutations and the BRAFV600 - the commonest 10 co-occurring mutations in each CRC subset and their frequencies are also unique (Figure 17B), supporting the notion that *TCF7L2 A9* mutant CRCs stratify a subset of MSI CRCs. Taken together, my bioinformatics data indicates that TCF7L2 truncating mutations define a novel subset of MSI CRCs that are wild type for BRAF and contain oncogenic Wnt pathway activity (Figure 17C).

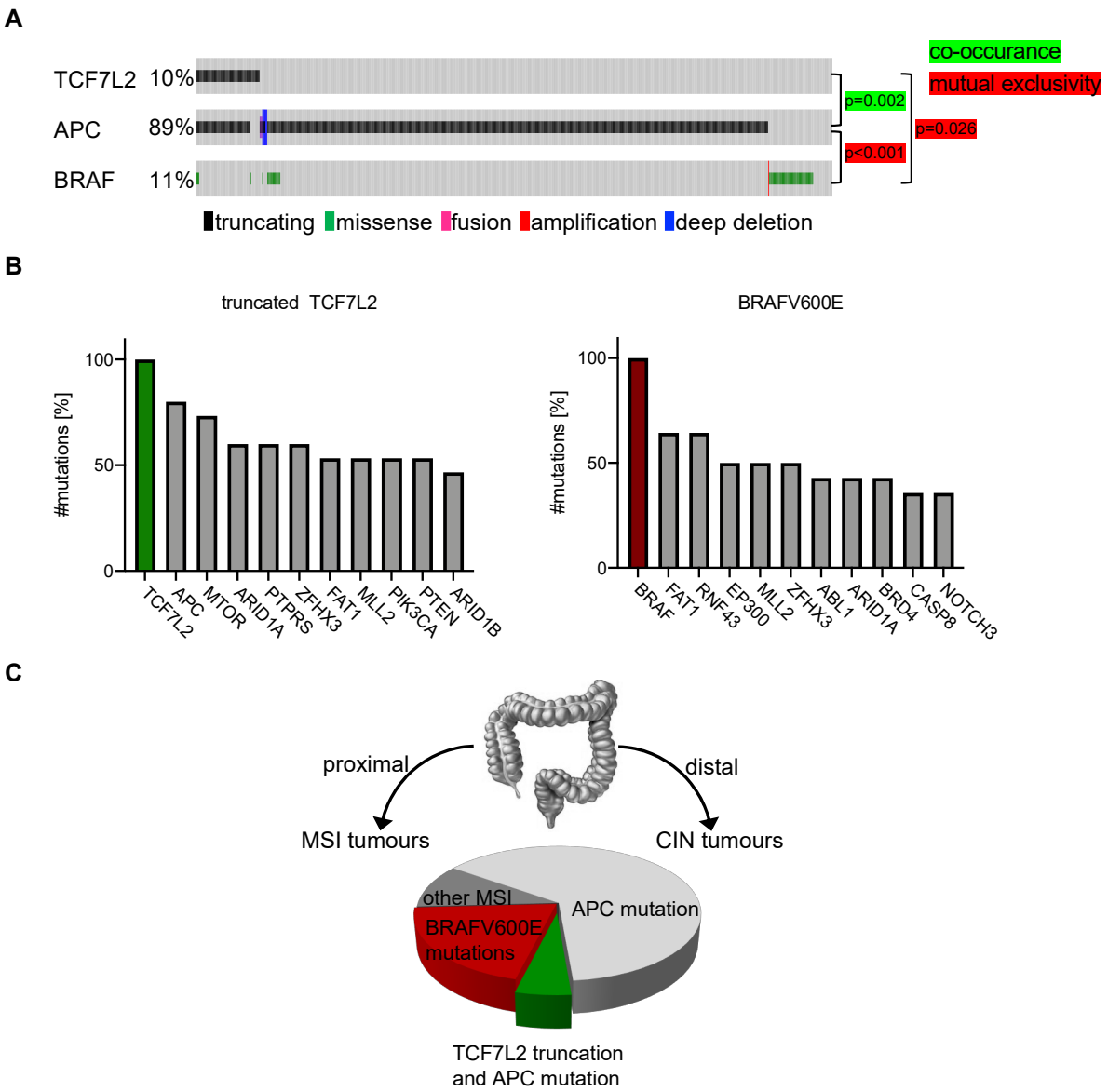


Figure 17: TCF7L2 mutations mutual exclusivity and co-mutation spectrum refine a novel subset of MSI colon cancer.
(**A**) Oncoprint plot of TCF7L2 truncating mutations, BRAFV600E missense mutations and APC truncating mutations reveal mutual exclusivity of TCF7L2 truncations and BRAFV600E along with co-occurrence with APC truncations. BRAFV600E and APC truncating mutations are mutually exclusive. p-values indicated (one-sided Fisher Exact Test), red as mutual exclusivity, green as co-occurrence. (**B**) most frequently occurring co-mutations of either TCF7L2 truncating mutations, or BRAFV600E mutations with each set to 100% show distinct mutational patterns with some overlapping mutations in different frequencies. (**C**) TCF7L2 truncations define novel subset of MSI colon cancers mutated in APC, but exclusive of BRAFV600E.

4.3.2 Generation of *in vitro* system to investigate TCF7L2 truncating mutations

To investigate the molecular function of truncated TCF7L2 in MSI colon cancer *in vitro,* stable MSI colon cancer cell lines for expressing of either full length TCF7L2 (TCF7L2-FL) or the truncated TCF7L2 found in *TCF7L2 A9* CRCs (TCF7L2-S) were generated. The transgenic system used the *tet*-on system, enabling tuneable transgene expression with treatment of cells with doxycycline (Dox) and is genomically integrated using the *piggybac* transposon system

(Figure 18A). The two expression vectors contained the two transgenic versions of TCF7L2, short and long, each with a C-terminal FLAG tag in order to abrogate post-translational processing (Chapter 3). The cell lines used were the MSI cell lines HCT116 and RKO as well as HEK293T. To validate the stable cell lines for inducible TCF7L2-FL and TCF7L2-S, protein lysates of each cell line were probed for induced protein expression via Western blot. All cell lines show modest induction (+Dox) of transgenic TCF7L2-FL, or TCF7L2-S, approximately 3-5-fold over endogenous levels, indicated by corresponding bands at 70 kDa and 55 kDa, respectively (Figure 18B).

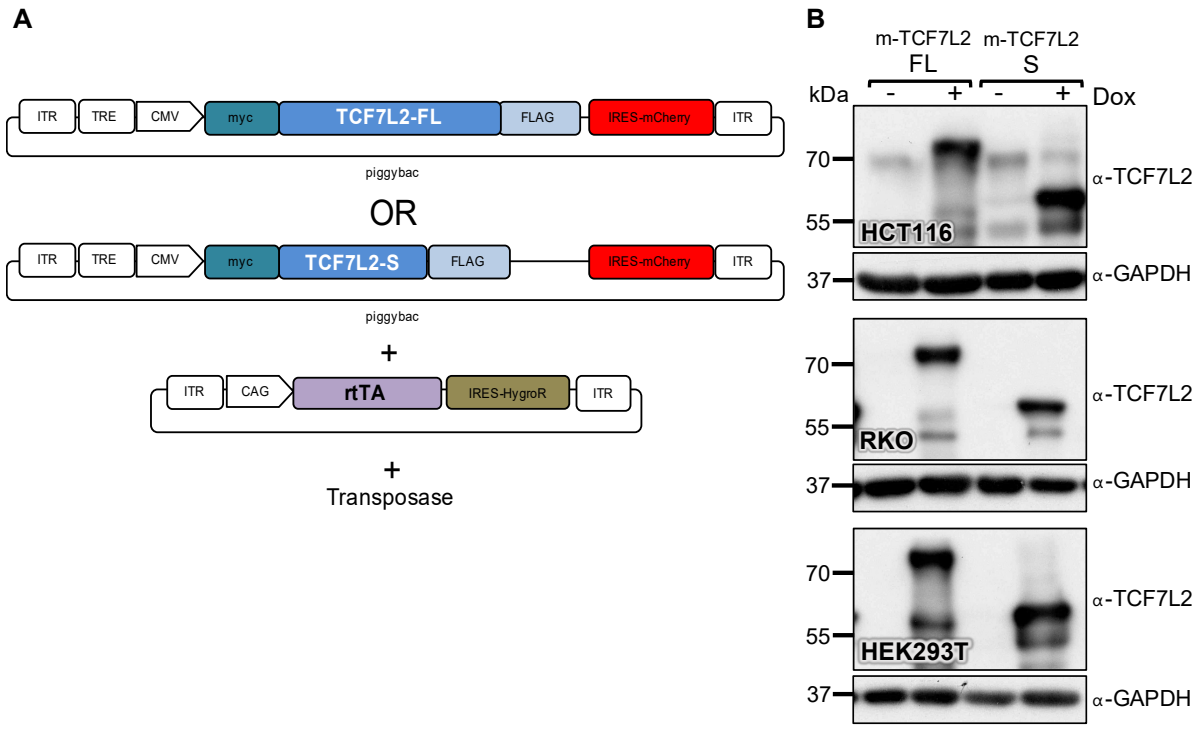


Figure 18: Cell line generation with stable Doxycycline inducible piggyBac system
(**A**) PiggyBac transposon plasmid system for inducible expression of TCF7L2 full length (TCF7L2-FL) and truncated K485S* (TCF7L2-S). The system is composed of three plasmids allowing genomic integration, doxycycline inducibility and visual detection by mCherry. (**B**) Validation of generated cell lines by Western blot show Dox induction of TCF7L2-FL (70 kDa) and TCF7L2-S (55 kDa) by probing with α-TCF7L2 antibody. α-GAPDH used as loading control, 40 µg protein/lane.

4.3.3 Truncating mutations of TCF7L2 show better prognosis and validate *in vitro*

To investigate clinical features of TCF7L2 truncated cases patient survival data downloaded from cBioPortals MSKCC was analysed. Kaplan Meyer curves were generated discriminating between TCF7L2 truncating frameshift mutations within the microsatellite, BRAFV600E mutated cases as well as all other cases not mutated in both, TCF7L2 nor BRAF. Overall survival of patients with truncated TCF7L2 was significantly better compared to all other cancer cases not mutated in TCF7L2 or BRAF. Excluding the possibility of a microsatellite effect, BRAFV600E mutated cases were also plotted resulting in slightly worse prognosis than

not mutated cancers. Therefore, TCF7L2 truncations classify MSI tumours towards a more positive survival prognosis in this analysed dataset of 1049 cases (Figure 19A).

To determine the consequence of TCF7L2-FL and TCF7L-S expression on proliferation and clonogenic growth, TCF7L2-FL or TCF7L2-S were expressed in the transgenic HCT116, RKO and HEK293T cell lines. Cell were cultured for 7-days, at sub-confluency to avoid contact inhibition, counted and analysed (Figure 19B). While HCT116 and RKO cells expressing TCF7L2-S showed a significant reduction in cell number, expression of TCF7L2-FL had no effect. There were no changes in HEK293T cell proliferation with the expression of either TCF7L2 transgene (Figure 19B).

I tested survival of TCF7L2-FL and TCF7L2-S expressing cells using colony formation assays. Transgenic cell lines were grown from single cells for up to 14 days, with and without expression of TCF7L2-FL or TCF7L2-S, fixed, stained and quantified. TCF7L2-S expression led to a significant reduction in clonogenic growth in both HCT116 and RKO cells whereas full length TCF7L2-FL expression had no effect (Figure 19C). Clonogenic growth was unaltered in TCF7L2-FL or TCF7L2-S expressing HEk293T cell lines (Figure 19C).

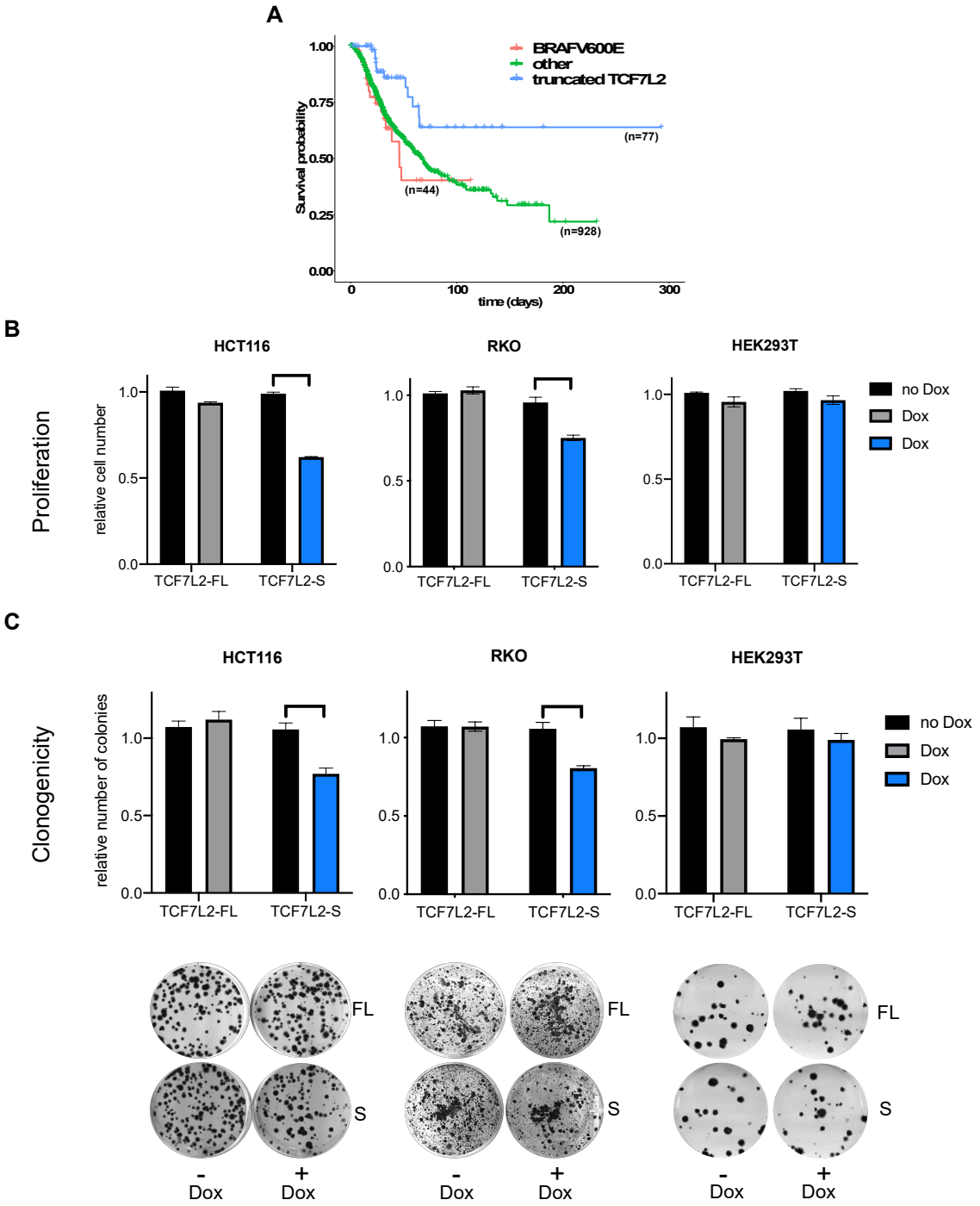


Figure 19: CRC cases with truncated TCF7L2 show better prognosis and validate in vitro. (**A**) Kaplan Meyer plot of TCF7L2 truncated cases (frameshift in microsatellite), BRAFV600E (red) and nonTCF7L2/BRAF mutated cases (green) indicate higher survival probability for TCF7L2 patients (blue). (**B**) Growth curves of HCT116, RKO and HEK293T induced by Doxycycline to express either TCF7L2 full length (TCF7L2-FL) or TCF7L2 truncated (TCF7L2-S for short). TCF7L2-S induced HCT116 and RKO cells show less proliferation (blue bar), compared to TCF7L2-FL (grey bar). HEK293T cells show no change in proliferation. (**C**) Colony formation assay of HCT116, RKO and HEK293T induced by Doxycycline to express either TCF7L2-FL or TCF7L2-S. TCF7L2-S induced HCT116 and RKO cells show less clonogenic growth compared to TCF7L2-FL. Each experiment was carried out in triplicates and *p*-values calculated using multiple t-test, indicated as asterisks for significance.

4.3.4 Expression of TCF7L2-S induces G1 arrest and increases apoptosis

Cell cycle analysis of HCT116 cells indicated that TCF7L2-S expression significantly increased the proportion of cells in the G1, from approximately 75% to 80%, concurrent with a corresponding reduction of cells in S-phase (from 20% to 15%). In contrast, expression of full length TCF7L2 had no effect (Figure 20A). Similar results were seen for RKO cells, where TCF7L2-S expression increased cells in G1 from 45% in controls to 55%, with no change upon TCF7L2-FL expression. I observed no changes in cell cycle with expression of either version of TCF7L2 in HEK293T cells (Figure 20A). Therefore, expression of TCF7L2-S, but not TCF7L2-FL attenuates proliferation through G1 arrest in MSI CRC cell lines but not in HEK293T cells.

I examined cell survival in the transgenic HCT116, RKO and HEK293T cell lines using an Annexin V antibody. TCF7L2-S expression led to increased levels of apoptotic cells in HCT116 cells (increase from 10% to 15% in expressing cells) and RKO cells (increase from 18% to 25%) whereas TCF7L2-FL expression has no effect. Neither TCF7L2-FL or TCF7L2-S expression had any effects on cell death in transgenic HEK293T cells (Figure 20B).

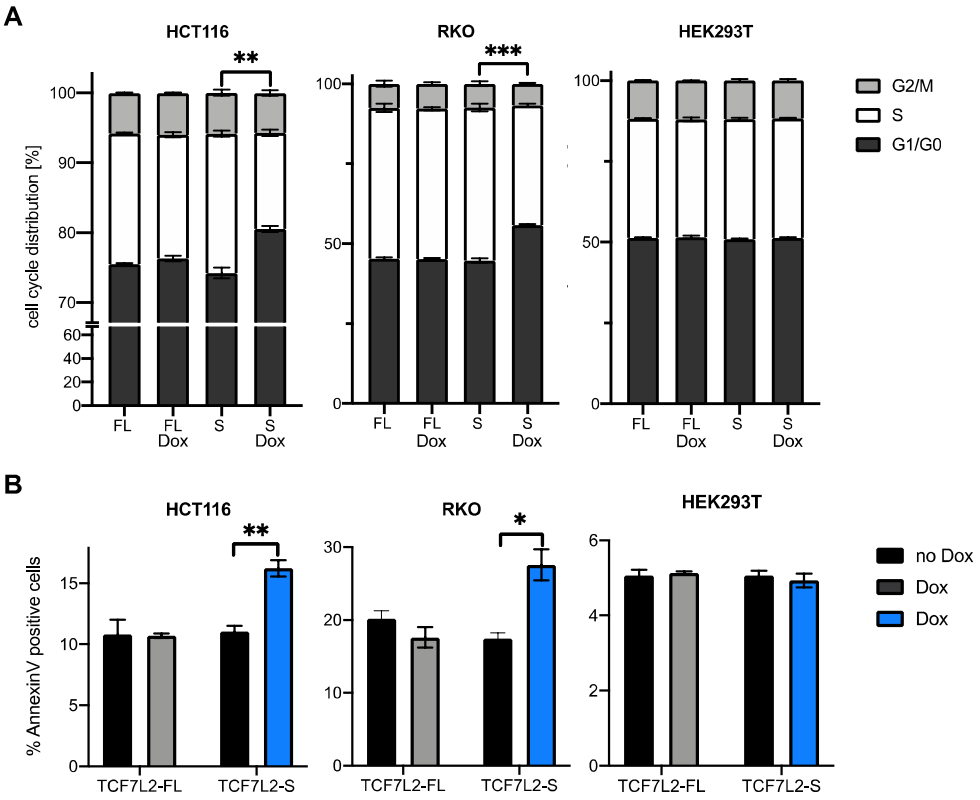


Figure 20: short TCF7L2 induces G1 arrest and increases apoptosis
(**A**) cell cycle distribution analysis of HCT116, RKO and HEK293T cells expressing TCF7L2-full length (FL) or TCF7L2-short (S) with and without Dox. HCT116 and RKO undergo G1 arrest and reduction of S-phase upon TCF7L2-S induction. HEK293T cells don't change their cell cycle phase distribution in neither TCF7L2-F or -S expression. (**B**) Detection of apoptotic cells through Annexin V staining in HCT116, RKO and HEK293T cells upon expression of TCF7L2-FL or TCF7L2-S. TCF7L2-S expressing HCT116 and RKO cells show more apoptosis compared to TCF7L2-F. HEK293T showed no change in apoptosis in both, TCF7L2-FL or TCF7L2-S. Each experiment was carried out in triplicates and *p*-values calculated using multiple t-test, indicated as asterisks for significance.

4.3.5 Gene set enrichment analysis of *TCF7L2 A9* mutant tumours

To gain insight into gene expression specific to *TCF7L2 A9* mutant CRCs, the transcriptional signature was determined from the TCGA dataset (see 2.8) through analysis using the gene set enrichment tool in R (GSEA) and mapped according to *p*-value adjusted with Benjamini-Hochberg correction (FDR). Genes categorised by the cellular processes 'Signalling by rho GTPases', 'mRNA-Splicing' and 'Programmed Cell Death' were enriched with the most significance (Figure 21).

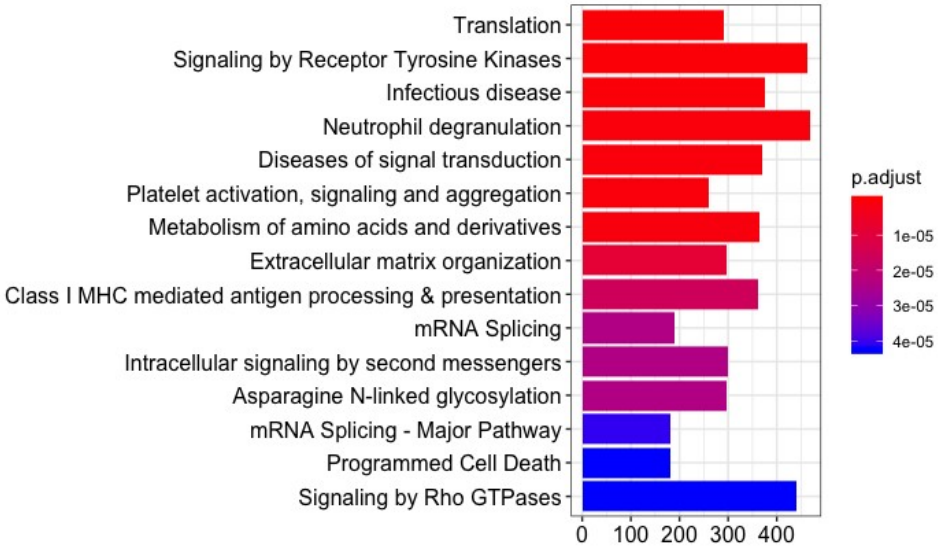


Figure 21: Gene set enrichment analysis of truncated TCF7L2 specific expressed genes
Differentially expressed genes in TCF7L2 truncated tumour samples (TCGA COAD) were analysed using GSEA algorithm in R and plotted in ranking of p-value. Signalling by Rho GTPases, Programmed Cell Death and mRNA Splicing showed highly significant p-values.

4.3.6 Differential gene expression of MAGEA12 and DEFA5 discriminates *TCF7L2 A9* subset CRCs

In order to identify potential biomarkers for the *TCF7L2 A9* subset CRC, I carried out transcriptome profiling and gene expression quantification analysis on the TCGA-COAD dataset (Figure 22A). Notably, MAGEA12 expression was significantly reduced, approximately 20-fold, whereas DEFA5 expression was increased approximately 5-fold (Figure 22A). Including MAGEA12 and DEFA5, 30 genes were selected for further *in vitro* investigation. 10 of the most down- and upregulated, respectively, were chosen for determination of relative gene expression analysis by qRT-PCR, given PCR primers were available and the log2 fold change had been assigned a significant p-value.

Out of the 20 genes selected, I validated differential expression of MAGEA12 and DEFA5 in the transgenic HCT116, RKO and HEK293T cell lines by qRT-PCR. MAGEA12 showed a reduction of approximately 90% in HCT116 cells expressing TCF7L2-S whereas no change was observed with expression of TCF7L2-FL. Similarly, expression of TCF7L2-S in RKO cells led to a 60% decrease in MAGEA12 mRNA levels and no change with expression of TCF7L2-

FL. In contrast, transgenic HEK293T cells showed no modulation of MAGEA12 expression with expression of either TCF7L2-FL or TCF7L2-S (Figure 22B left).

DEFA5 mRNA levels were increased with TCF7L2-S expression in both HCT116 and RKO cell lines, approximately 3.5-fold and 2.5-fold, respectively. There were no changes in DEFA5 levels with TCF7L2-FL expression, nor in HEK293T cells expressing either version of TCF7L2 (Figure 22B right). Taken together, differential expression of MAGEA12 and DEFA5 are hallmarks of *TCF7L2 A9* CRCs and this can be modelled by expressing TC7L2-S in transgenic HCT116 and RKO cell lines. As MAGEA12 has been reported to contribute to tumorigenesis in melanoma its downregulation in this *in vitro* model of CRCs is especially of interest. DEFA5 expression has been validated to be restricted to Paneth cells within the intestinal epithelium and appears to be targeted by truncated TCF7L2, suggesting a role in tissue integrity.

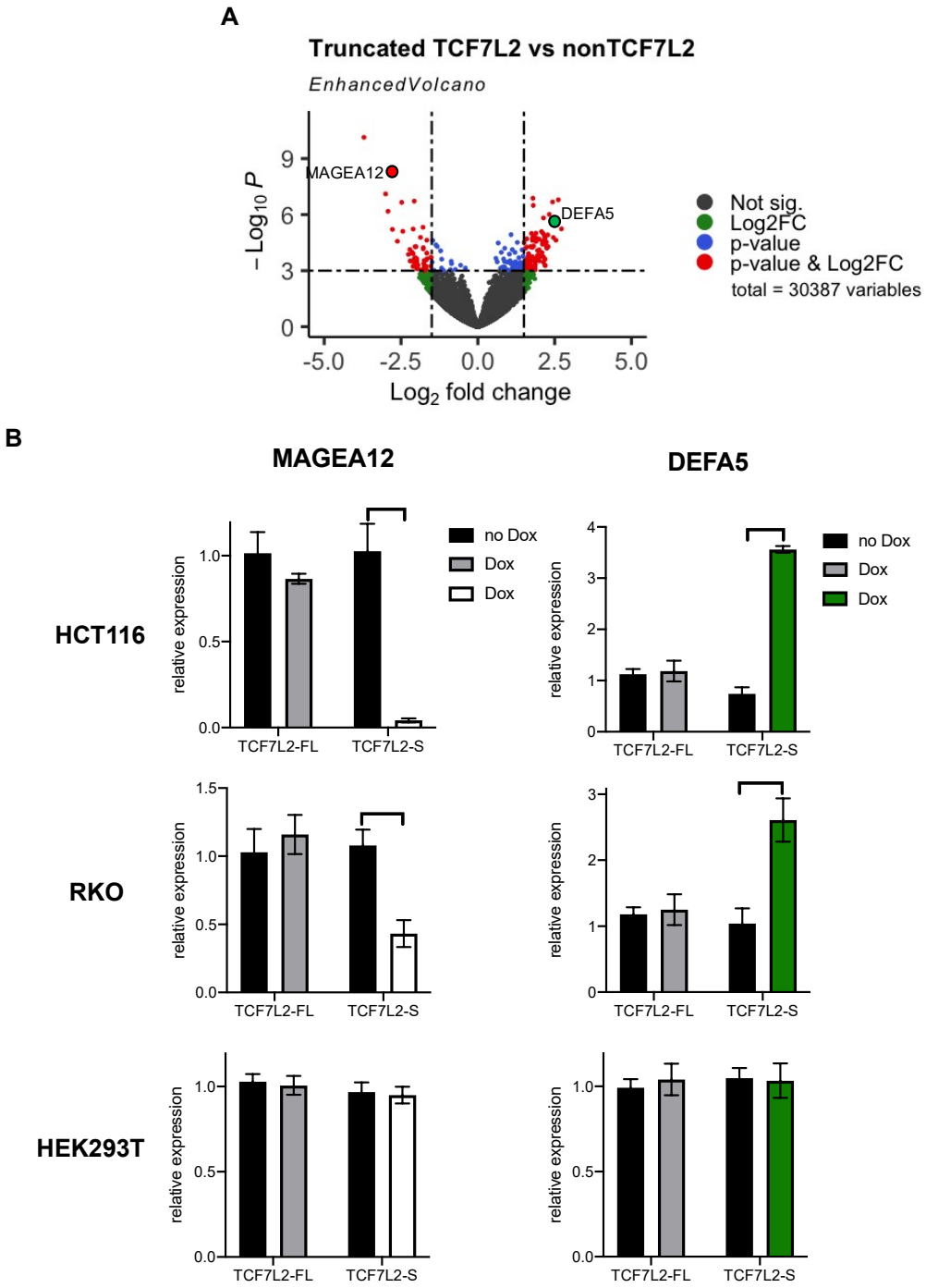


Figure 22: Gene expression profiling and validation of differentially expressed genes in vitro
(**A**) Volcano plot of differentially expressed genes in TCF7L2 truncated tumour cases in TCGA-COAD data mapped with MSKCC mutational data compared to tumour cases with wildtype (non)TCF7L2, n=30387, p-value cut-off=10^{-4}, log2fc cut-off =1.5, threshold for significance p=10^{-3}. Validated genes MAGEA12 and DEFA5 indicated as red and green dot, respectively. (**B**) qRT-PCR of MAGEA12 and DEFA5 expression in HCT116, RKO and HEK293T cells Dox-inducibly expressing TCF7L2 full length (TCF7L2-FL) or TCF7L2 truncated (TCF7L2-S). MAGEA12 reduced 90% in HCT116 and 60% in RKO cells upon TCF7L2-S expression, DEFA5 increased 3.5 times and 2.5 times in HCT116 and RKO, respectively. HEK293T did not change in MAGEA12 or DEFA5 neither in TCF7L2-FL nor TCF7L2-S. Each experiment was carried out in triplicates and *p*-values calculated using multiple t-test, indicated as asterisks for significance.

4.3.7 Gene Ontology (GO) enrichment analysis of differentially expressed genes

Gene expression profiling identified differentially expressed genes in *TCF7L2 A9* CRCs. I extended this analysis to determine CRC-specific biological processes/molecular functions to find potential clustering in certain pathways. The existing list of differentially expressed genes was sorted by fold change (cut-off 1.5) and *p*-value ($p<0.05$) and submitted to GOrilla Gene Ontology enrichment tool (Figure 23A). TCF7L2 has been reported to play a role in glucose metabolism [154], and this is reflected by truncated TCF7L2 affecting glycogen synthesis, potentially by differential regulation of IGF2 and ADIPOQ. Highly significant enrichment was found for genes associated with RNA metabolism, in particular RNA processing and nucleolar RNAs (Figure 23A). This is consistent with studies establishing cross-talk between the Wnt signalling and RNA processing pathways [165]. GATA4 and SFRP5, negative regulators of Wnt pathway activity were significantly enriched genes regulating BMP signalling (Figure 23A). I performed qRT-PCR to quantify GATA4 and SFRP5 transcript levels in transgenic HCT116, RKO and HEK293T cell lines expressing either TCF7L2-FL or TCF7L2-S but found no change (Figure 23B).

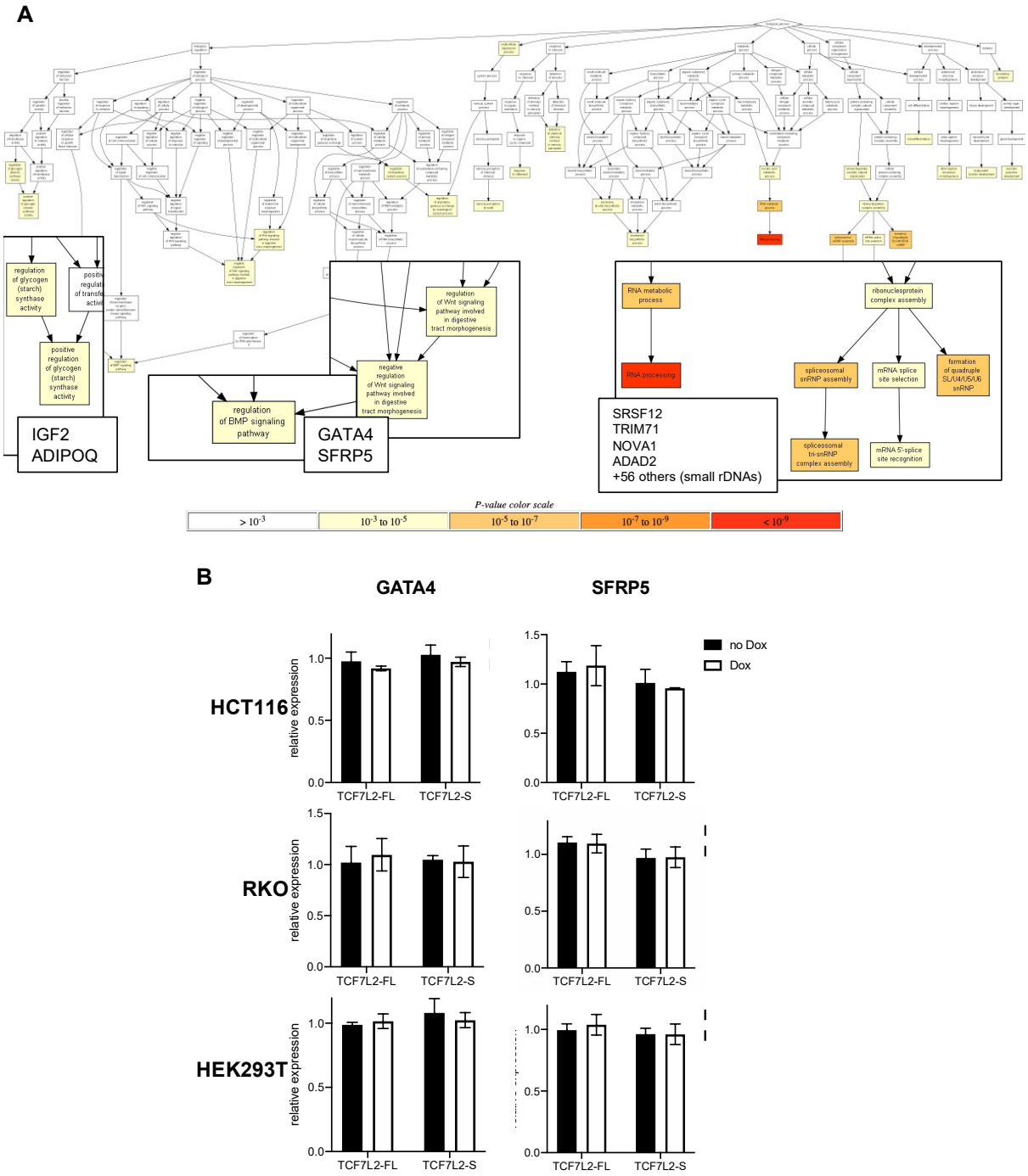


Figure 23: Gene Ontology enrichment of TCF7L2 truncated differentially expressed genes
(**A**) GOrilla Gene Ontology enrichment of genes differentially expressed genes in TCF7L2 truncated tumour cases. Visualised as biological process tree and filtered for significant p-value. Enriched genes indicated. (**B**) qRT-PCR of GATA4 and SFRP5 in HCT116, RKO and HEK293T cells expressing full length TCF7L2 or truncated TCf7L2 after Dox induction relative to GAPDH expression and no Dox control. Each experiment was carried out in triplicates.

4.4 Discussion

4.4.1 TCF7L2 defines a novel subset of MSI CRC

Here I describe a unique subset of microsatellite instable CRCs that is characterised by specific frameshift mutations in the exon 18 microsatellite of *TCF7L2*. The mutations lead to expression of a truncated version of the TCF7L2 protein. These truncating mutations are usually caused by a frameshift mutation within the *TCF7L2 A9* microsatellite, leading to a premature stop codon. The majority of all TCF7L2 mutations are truncating, the result of insertion/deletions, frameshift and translational termination at a new stop codon. My analysis of CRC tumours harbouring *TCF7L2* A9 mutations revealed that these stratify a unique subset that is consistently (i) MSI, (ii) show high occurrence of oncogenic APC truncating mutations and (iii) are mutually exclusive of BRAFV600 mutations. Analysis of the most frequent co-occurring mutations of the mutually exclusive groups *TCF7L2* A9 mutations and BRAFV600E MSI also highlight the distinct nature of these MSI-H subsets. While previous studies have identified *TCF7L2* A9 mutations in CRC [157-159], none have appreciated the fact that these are a distinct stratified subclass of CRCs with unique clinical and phenotypic features, like better prognosis and differential gene expression.

The different mutations co-occurring in the TCF7L2 truncated cases not only support the finding of a distinct CRC subset, but also give potential insights into relevant biology. For example, mTOR mutations are the second most frequent mutation in the TCF7L2 A9 mutant tumours. Oncogenic hotspots of mTOR mutations are mostly within the FAT domain, FRB domain or Kinase domain resulting in a hyperactivation and drug resistance [166]. mTOR mutations co-occurring with TCF7L2 truncating mutations however are mainly in its N-terminal region (Heat domain) or between functional domains and are of unknown significance (data not shown due to lack of validation at time of submission). mTOR has been shown to be involved in apoptosis regulation besides protein synthesis, migration and invasion [167]. My working hypothesis regarding mTOR mutations accompanied by TCF7L2 truncations is that it negatively regulates the mTOR pathway leading to increased apoptosis.

The importance of stratifying different types of CRCs has been demonstrated by the consensus molecular subtype (CMS) classification system [1]. A more detailed understanding of the transcriptional and genetic properties of different CRC subtypes has shaped treatment options and clinical decision making. Here I show that TCF7L2 truncating mutations define a distinct new subtype of CRCs; based on the data from MSKCC and TCGA datasets, the subset does not phenotypically match any of the established CMS classes [1]. A future goal will be to carry out deeper analysis of the *TCF7L2* A9 mutant CRCs in order to define them as a new classification,

4.4.2 Development of *in vitro* system for determining functions of TCF7L2 and the *TCF7L2 A9* mutant form.

In this chapter I describe an *in vitro* system I generated to model the cellular effects of the *TCF7L2 A9* truncating mutations. The tuneable system allowed me to quantify cellular phenotypes and gene expression in transgenic cell lines. In order to investigate if this phenotype is restricted to MSI CRC, APC mutated MSS cell lines should be similarly investigated.

4.4.3 Truncating mutations of TCF7L2 in CRCs are associated with better prognosis

I used the MSKCC dataset to determine that the presence of *TCF7L2* A9 mutant CRCs predicted a better overall survival that all other cancer subtypes. The presence of the BRAFV600E mutation has been reported as an indicator of poor prognosis and this is recapitulated in my analysis by significantly lower survival probability in CRCs containing the BRAFV600E mutation [168].

I then determined whether I could model the phenotypic consequence of expression of the truncated, short form of TCF7L2 at the cellular level. Indeed, it did - I found that CRC cell lines expressing the truncated form of TCF7L2 exhibited reduced proliferation and clonogenic growth. The cell line results align with patient survival data and may provide some explanation as to why *TCF7L2* A9 mutant CRC cases are more benign.

I also examined other cellular consequences of mutant TCF7L2 expression. A recent study suggested that loss of TCF7L2 plays a role in the rate of the G1/S cell cycle transition [148]. I found that mutant TCF7L2 expression in CRC cell lines lead to a significant increase in the G1 phase of the cell cycle concurrent with a decrease in S phase. Moreover, I found an increase in apoptotic cells in the TCF7L2-S expression.

4.4.4 Gene expression profiling and biomarkers for TCF7L2 A9 mutant CRCs

I identified a number of differentially expressed genes in the *TCF7L2 A9* mutant CRCs, two of which I validated using RT-PCR - MAGEA12, which was reduced in the subset and human defensin 5 (DEFA5) which displayed significantly increased expression levels. It will be interesting to determine the consequence of altered expression of the two genes. Previous work has shown that suppression of MAGEA12 expression has a pro-apoptotic effect in HCT116 colon cancer cells dependent on wildtype TP53 [169] and may go some way to explaining why TCF7L2-S expression led to increased apoptosis. A potential mechanism by which TCF7L2 regulates MAGEA12 expression has yet to be determined as MAGEA12 does not harbour a TCF/LEF recognition site in its promotor and so, may not be a direct Wnt pathway target gene.

In the intestinal epithelia, DEFA5 expression is restricted to Paneth cells and its gene promoter contains canonical TCF7L2 binding sites [59, 170]. Interestingly, my results indicate it is predominantly targeted by TCF7L2-S and not TCF7L2-FL in both the transgenic cell lines and the *TCF7L2* A9 mutant CRCs. On a note and as expected, this analysis did not reveal Wnt pathway genes amongst the differentially expressed genes. Due to the C-terminal truncation of TCF7L2 in this subset, the N-terminus of the protein, containing the β-Catenin and canonical DNA binding HMG domain is unaltered, and is therefore capable of binding to Wnt pathway gene promotors in the same way as the wildtype, full length form does.

My investigations of MAGEA12 and DEFA5 expression support the notion that the mutant TCF7L2 A9 version of the protein discriminates a specific cohort of Wnt pathway downstream genes. Therefore, C-terminal interaction partners might have gene targeting determining properties. Another option of discrimination is a conformational influence of truncated TCF7L2 bound to β-catenin sterically allowing interaction of a co-activator or the inhibition of co-repressing factors. The presence of a helper site, adjacent to the WRE bound by the C-clamp, does not mean it is a transcriptional activator.

Future work could involve validation of MAGEA12 and DEFA5 expression as biomarkers for the *TCF7L2* A9 mutant CRCs. Alongside full genetic and clinical characterisation of this CRC subset by sequencing more tumour samples, the output of such work will inform clinical decision making. ChIP-sequencing experiments using cells expressing either the full length or truncated version of TCF7L2 tagged with myc-epitope will further give insights into the nature of potential differential downstream gene populations specifically driven by full length or truncated TCF7L2, respectively. The results can therefore discover more distinct features of this subset, elucidating its molecular classification.

4.4.5 Gene ontology enrichment of TCF7L2-S specific genes

My analysis of *TCF7L2 A9* mutant gene expression indicated significant enrichment of genes associated with pathways of glycogen synthesis, negative regulation of Wnt signalling pathway, regulation of BMP signalling and RNA-processing.

Of special interest is the significant crosstalk between negative regulation of Wnt signalling, (responsible for stem cell homeostasis and proliferation) and the regulation of BMP pathway (executing programs for cellular differentiation). Associated genes with significant downregulation include GATA4, (reported to negatively regulate cell differentiation in pancreatic carcinoma [171]) and SFRP5 secreted Frizzled-related protein, (a Wnt signalling antagonist that has been shown to be epigenetically silenced in breast and kidney cancer [172, 173]). These results suggest that cells with truncated TCF7L2 show a well differentiated phenotype which has been reported in a study investigating the impact of TCF7L2 truncating frameshift mutation [159]. However, using the *in vitro* cell line system created in this study,

these genes did not change upon truncated TCF7L2 expression. Moreover, for the limited dataset available this concludes to be an interesting finding and accumulation of data from other sources will enable to resolve the *TCF7L2 A9* mutant phenotype.

4.4.6 Conclusion

This chapter presents data supporting the stratification of a novel subset of MSI CRCs that is defined by specific truncating *TCF7L2 A9* microsatellite mutations. Contrary to the expectation based on literature, these tumours display to be more benign as validated *in silico* and *in vitro*. The biology of TCF7L2's C-terminus therefore appears to be way more complex influencing stem cell homeostasis programs as well as crosstalk with differentiation pathways. I have carried out initial characterisation of the subset as well as the cellular consequences of mutant TCF7L2 (TCF7L2-S) expression. It will be important to further establish how the mutant TCF7L2 contributes to CRC pathology which will require appropriate disease modelling. The next chapter focusses on the establishment of an organoid system cultured with growth factors of defined cellular activity. Organoids will provide an experimentally tractable system to understand TCF7L2 processing and CRC mutations in a complex cellular system.

Chapter

5 Growth factors of defined cellular activity for use in organoid culture systems

5.1 Introduction

5.1.1 Media formulation is the basis for organoid technologies

The critical component of organoid technology is the formulation of culture media with essential growth factors, in particular, R-spondin, that potentiates Wnt pathway activity in epithelial stem cells and Gremlin 1 (or Noggin) that inhibits differentiation cues from the BMP pathway. Currently, the production of growth factors relies on eukaryotic expression systems to ensure correct disulphide bond formation, glycosylation and macromolecular folding. Typically, in eukaryotic expression systems, growth factors are secreted into the cell culture media from the expression hosts and conditioned media containing the growth factor of choice is then diluted directly into organoid culture media [38, 39, 45, 174]. R-spondin 1 and Gremlin 1/Noggin expressing cell lines are not widely available. Moreover, the use of conditioned media raises problems with batch-to-batch variation of growth factor activity as well as the presence of unknown serum factors present in cell culture media that will affect organoid growth. Purified growth factors are commercially available, however, these are expensive for medium to large scale applications such as genetic and chemical library screening or clinical biobanking. Moreover, commercial growth factors preparations retain impurities such as media serum components preventing accurate and reproducible determination of their cellular activities.

5.1.2 Aims of this Chapter

Goals of this chapter are to

- establish the protocol for production of highly pure recombinant Gremlin 1 and R-spondin 1 from bacterial expression
- Develop quality control measure to ensure reproducible cellular activities of the growth factors.
- These will enable a reproducible organoid system for further development of an approach to isolate LGR5 expressing (stem) cell populations and subsequently facilitate the investigation of TCF7L2 biology within different epithelial cell types.

5.2 Materials and Methods

5.2.1 Organoid media components

Components of the respective media formulations for each organoid type used in the study, listed in Table 7, are as follows: Base media - Advanced DMEM F12 (cat No 12634028, ThermoFisher) containing 1X penicillin/streptomycin and 10 mM HEPES buffer pH 7.0 (cat. No. 15630080, ThermoFisher). Media supplements included are listed in Table 6. All organoids were grown in growth factor reduced, serum free Matrigel (cat. No. 356237, Corning).

Table 6: Base media supplements used for organoid growth in this Chapter

Supplement	Company	Catalog number
N21-MAX (B27)	R&D Systems	AR008
N2	ThermoFisher	17502048
N-acetyl-L-cysteine (NAC)	Sigma-Aldrich	A9165
human gastrin 1	Sigma-Aldrich	G9020
A83-01	Sigma-Aldrich	SML0708
SB202190	Sigma-Aldrich	S7067
prostaglandin E2 (PGE2)	Sigma-Aldrich	P0409
nicotinamide	Sigma-Aldrich	N3376
Primocin	InvivoGen	ant-pm-1
R-spondin 1 (aa 31–263)	R&D Systems	4645-RS-025
Gremlin 1	R&D Systems	5190-GR-050
FGF-2	R&D Systems	233-FB-010
Neuregulin-1 (NRG-1)	R&D Systems	396-HB/CF
EGF	R&D Systems	236-EG-200

5.2.2 Expression and purification of R-spondin 1

Expression of the R-spondin 1 fusion protein and the purification protocol is outlined in detail in Chapter 2.5

5.2.3 Expression and purification of Gremlin 1

Human Gremlin 1 was expressed from a T7-based His-tag vector, His-ΔN-Gremlin 1 (encoding amino acids 72–184 of Gremlin 1), in the *E. coli* strain BL21 (DE3) as follows: cells from glycerol stock were plated on L-agar plates with 100 µg/ml of ampicillin and grown overnight. Cells from the plates were grown in baffled 2 L flasks in 1 L of 2xYT media until an OD_{600} of 0.8 at which point Gremlin 1 expression was induced with 400 µg/ml IPTG for three hours at 37 °C. Cells were harvested by centrifugation for 20 minutes at 4,000 × g and stored frozen at −20 C. The frozen pellet from 1 L of Gremlin 1 expressing culture was resuspended in lysis buffer (50 mM Tris base pH 8.0, 2 mM EDTA pH 8.0, 10 mM DTT, 0.5% Triton-X100) and lysed using an Emulsiflex C5 homogeniser. The lysate was incubated for 20 minutes at 25 °C with DNase I (Sigma) and 4 mM magnesium chloride and clarified by centrifugation for

30 minutes at 15,000 × g. The supernatant was discarded and the pellet fraction containing insoluble inclusion bodies was washed by resuspension in lysis buffer followed by centrifugation for 30 minutes at 15,000 × g. The pellet was re-washed with lysis buffer containing 1 M NaCl and finally with lysis buffer without Triton-X100. The pellet was resuspended in 5 ml of 100 mM Tris(2-carboxyethyl)phosphine (TCEP), to which 15 ml of 50 mM Tris-HCl pH 8.0, 0.5 mM EDTA containing 8 M guanidine hydrochloride (GndHCl) was added (resulting in a final concentration of 6 M GndHCl and 25 mM TCEP). The resuspended pellet was incubated for 15 minutes at 25 °C, clarified by centrifugation (20 min, 15000 × g) and the supernatant diluted to 100 ml with 6 M urea, 20 mM HCl. The solubilised protein solution was then mixed rapidly with 900 ml refolding buffer (50 mM Tris pH 8.0, 1 M 3-[3-(1,1-bisalkyloxyethyl)pyridin-1-yl]propane-1-sulfonate (PPS), 50 mM ethylenediamine, 2 mM EDTA, 2 mM cysteine and 0.2 mM cystine) and left for 7 days at 4 °C to allow the correct formation of the disulphide bonds.

After centrifugation (15,000 × g for 30 mins), the pH of the refolded protein solution was adjusted to 7.0 and loaded onto a 5 ml HiTrap SP HP column equilibrated with binding buffer (20 mM Hepes, pH 7.0). Gremlin 1 was eluted using a linear, 75 ml gradient, of 0% to 100% elution buffer (20 mM Hepes, pH 7.0, 1 M NaCl) and peak fractions, analysed by SDS-PAGE, were pooled. At this stage of the purification, Gremlin 1 can be used in small intestinal epithelial organoid media at a concentration of 25 nM.

Further purification of Gremlin 1 proceeds with the addition of acetonitrile and trifluoroacetic acid to a concentration of 10% and 0.1%, respectively. The sample was loaded onto an ACE 5 C8-300 (4.6 × 250 mm) reversed phase column (HiChrom) equilibrated with binding buffer (10% ACN, 0.1% TFA). Proteins were eluted with a linear 20 ml gradient of 20% to 40% ACN containing 0.1% TFA. Peak fractions with pure protein (analysed by SDS-PAGE) were pooled and their concentration determined using absorbance at 280 nm with a calculated molar absorption coefficient of 11,460 $mol^{-1} \cdot cm^{-1}$. Aliquots of 500 µg pure Gremlin 1 were dried under vacuum in a centrifugal concentrator and stored at −80 °C. The protein is stable for at least 24 months when stored in this way. Prior to use, purified Gremlin 1 was reconstituted in PBS at a concentration of 1 mg/ml and cellular activity determined by alkaline phosphatase assay (see below). Both, production of BMP2 as well as expression and purification of ΔN-Gremlin 1 used in this thesis, have been performed by Katharina Ravn and kindly provided by the Hyvonen lab.

5.2.4 Cellular reporter assay of R-spondin 1 activity

The cellular activities of MBP-R-spondin and R-spondin 1 fractions were determined as described in Chapter 2.6 and WPC50 calculated from technical triplicates of 3 independent experiments. Significance was assessed using multiple grouped t-tests.

5.2.5 Cellular Gremlin 1 activity

The cellular activity of purified Gremlin 1 is defined as the IC_{50} for inhibition of BMP2-induced cellular activity using the alkaline phosphatase assay (ALP assay) [175]. Optimal concentration of BMP2 for use in Gremlin 1 activity assays was determined as follows: C2C12 cells (low passage number is critical) were plated into 96-well plates and cultured in DMEM containing 10% FBS. A serial dilution of BMP2 (0 nM–15 nM) was added to the cells and incubated for 48 hrs. Cells were washed with PBS and lysed in 0.56 mM 2-amino-2-methyl-propan-1-ol and 1.0% SDS, pH 10.0. ALP activity was measured using the substrate *p*-nitrophenyl phosphate (Sigmafast pNPP pellets; Sigma Aldrich) and absorbance read at 405 nm using a PHERAstar FS plate reader. The EC_{50} value for BMP2-induced ALP activity was determined as 5.8 ± 0.28 nM with saturating activity at approximately 12 nM. Activity values for each concentration of BMP2 were the average of 3 individual experimental replicates reported with SD. The sub-maximal value of 8 nM BMP2 was chosen for use in Gremlin 1 activity assays.

Determination of the Gremlin 1 IC_{50} for inhibition of BMP2-induced ALP activity was carried out as above, using serial dilutions of Gremlin 1 (1–25 nM) in the presence of 8 nM BMP2. Activity values for each concentration of Gremlin 1 were the average of 4 individual experimental replicates reported with SD. All kinetic parameters for ALP assays were calculated using GraphPad Prism Software Version 8 by fitting the Hill equation to the dose-response data.

5.2.6 Endotoxin assay

Concentrations of endotoxin in the R-spondin 1 and Gremlin 1 preparations were determined using the ToxinSensor system (GenScript) according to manufacturer's protocol. Values were determined from at least two individual preparations of R-spondin 1 and Gremlin 1 fractions, measured with two technical replicates and reported as average values with SD. A maximum endotoxin concentration of 0.5 EU/ml in organoid media was set as the Lowest Observed Cellular Effect (LOCE) threshold for detectable cellular cytokine production and NF-kB activation, previously determined in cell culture assays [176].

5.2.7 Derivation of organoid cultures

Murine small intestinal epithelial organoids and colon epithelial organoids (murine and human) were derived according to Sato *et al*. [174] and Sato *et al*. [39], respectively. Murine mammary gland organoids were derived as in Nguyen-Ngoc *et al*. [177]. A summary of all media formulations used in the study is shown in Table 7. All derived organoid cultures were grown in 48-well CellStar plates (Sigma-Aldrich), embedded in a 25 µL droplet of Matrigel (Corning), cultured in 200 µL of the corresponding organoid media in an incubator at 37 °C and 5% CO_2.

All murine tissue used in the study was collected from euthanized mice by trained technicians within the Wellcome Trust Gurdon Institute facility, following UK Home Office approved guidelines. Murine intestinal organoids were grown, maintained, stained and data kindly provided by Helena Rannikmae. Mammary gland organoids were established in cultures by Marc de la Roche. Human colon epithelial- and colon cancer-derived organoids were from the Addenbrooke's Hospital Tissue Biobank donated under informed, explicit, patient consent. Human tissue obtained for the study was under generic Biobank ethical approval and all applicable institutional guidelines and regulations regarding the use of human tissue were followed. Human tissue was processed, and organoid cultures established by Thomas Foets.

5.2.8 Optimisation of R-spondin 1 and Gremlin 1 concentration in media formulations

Optimisation of R-spondin 1 and Gremlin 1 concentrations for use in media for culture of murine small intestinal epithelial organoids were determined by growing organoids in individual wells with variable concentrations of R-spondin 1 (0–100 nM) or Gremlin 1 (0–25 nM) in otherwise fully-supplemented organoid growth media. Growth characteristics (multiplicity and crypt number) of organoids were tested 8 days post-plating. Organoid multiplicity was determined as the average organoid count for each treatment from 8 different wells, reported with SD. Organoid crypt number was determined as the average number of crypts from 10 organoids in individual wells, averaged across 8 wells, reported with SD. Statistical differences in organoid multiplicity and crypt number amongst the individual concentrations and growth media conditions were determined by two-tailed T-test, calculated using Excel software.

5.2.9 Fluorescent detection of molecular markers in organoids

Organoids were seeded and cultured in eight-chamber µ-slides (Thistle Scientific) for three days and fixed in 4% formaldehyde solution and molecular markers were visualised following the protocol from Fatehullah [178]. Molecular probes used for fluorescent detection and indirect immunofluorescence in organoids were as follows: Rhodamine-conjugated Ulex europeaus agglutinin 1 (UEA; Vector Laboratories), Alexa Fluor 647 Phalloidin (Thermo Fisher Scientific), and antibodies raised against β-catenin (BD Transduction Laboratories, 1:500) or Ki67 (Thermo Fisher Scientific, 1:250). The secondary antibodies were Goat Anti-Mouse IgG Alexa Fluor 555 and Goat Anti-Rabbit IgG Alexa Fluor 488 (both from BD Transduction Laboratories, 1:500), respectively. Slides were mounted using DAPI Fluoromount G (Southern Biotech).

Fluorescent images were taken by a Zeiss LSM 700 laser scanning confocal microscope using an oil-immersion 40x objective. Images were processed using Zen and ImageJ software.

Culture and indirect immunofluorescence of these organoids was performed by Marc de la Roche.

5.3 Results

The protocol for the production and validation of R-spondin 1 and Gremlin 1 production is described in Figure 24.

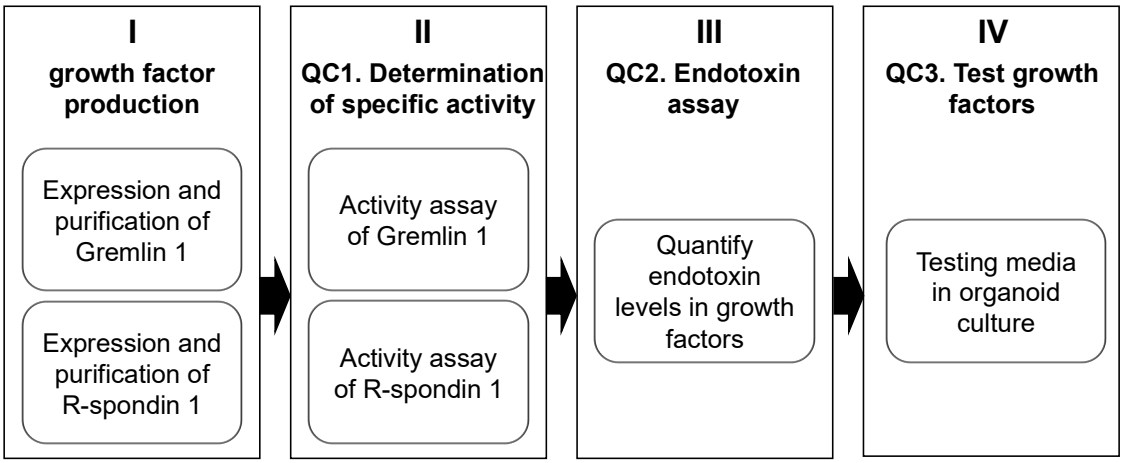


Figure 24: Overview of the procedure
The generation of recombinant R-spondin 1 and Gremlin and subsequent characterisation steps for use in organoid media are as follows: Step I - Recombinant growth factors R-spondin 1 and Gremlin 1 are expressed in bacteria and purified. Step II – Cellular activities of Gremlin 1 and R-spondin 1 were determined from corresponding reporter assays. Step III – A quality control step is used to ensure minimum endotoxin levels (from growth factor expressing bacteria) are present in the preparations using the limulus amoeboid lysate (LAL) assay. Step IV - media formulated with defined activities of R-spondin 1 and Gremlin 1 is tested for the ability to support growth of murine intestinal epithelial organoids. I have further outlined abridged purification protocols for Gremlin 1 and R-spondin 1 in Materials and Methods, ideally suited to laboratories without access to production-scale chromatography equipment, for production of MBP-R-spondin 1 and Gremlin 1 of sufficient purity and cellular activity for organoid culture.

5.3.1 Production of R-spondin 1

The expression vector for R-spondin 1 (Figure 25A) has been modified from a previous report [111] to include the coding sequence for the 9 amino acid Avi-tag fused to the N-terminus of the R-spondin 1 that confers enhanced solubility and specific activity of the expressed protein.

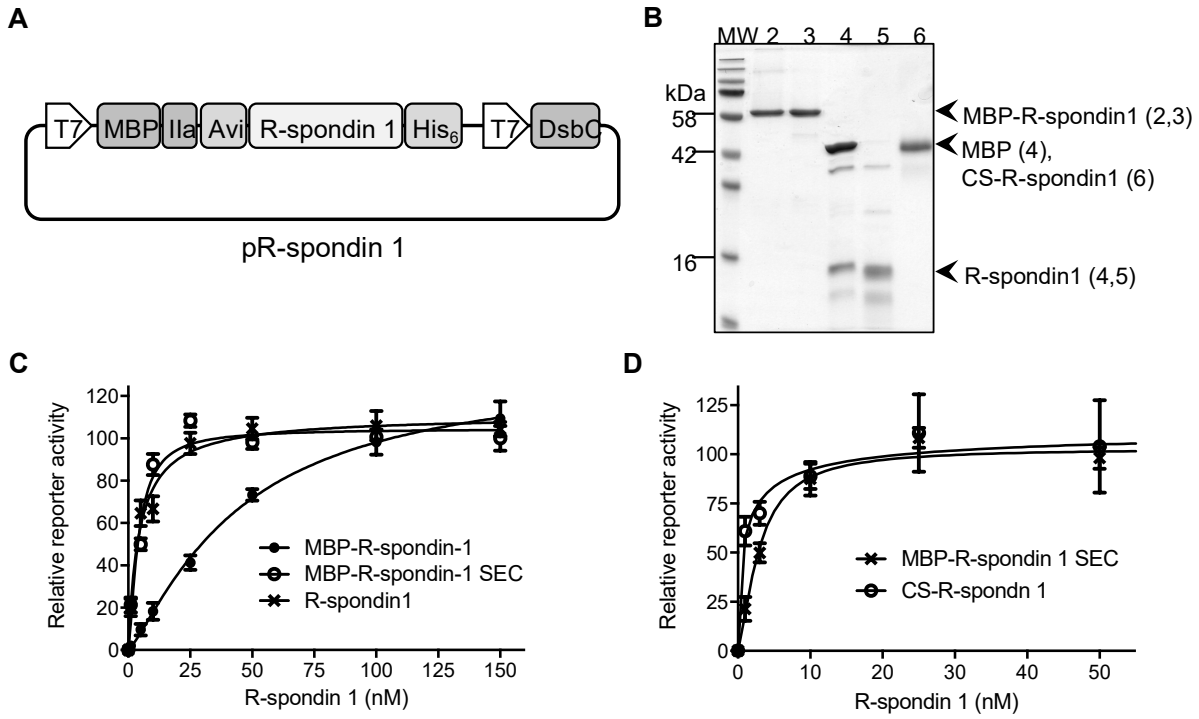


Figure 25: Bacterial expression and production of R-spondin 1
(**A**) Format of the pR-spondin 1 expression vector based on the pETDuet backbone. pR-spondin 1 is a modification of a previously R-spondin 1 expression vector10. (**B**) SDS-PAGE analysis of fractions from generation of R-spondin 1. For each fraction, 2 µg of protein was loaded. Lane numbers (top and in brackets) are: 2 – MBP-R-spondin 1 before SEC, 3 - MBP-R-spondin 1 after SEC, 4 – products of Thrombin cleavage, 5- final R-spondin 1 preparation and 6 – CS-R-spondin 1. MW, molecular weight standards. (**C**) Cellular activity of R-spondin 1 fractions measured as WPC50s: circles - MBP-R-spondin 1 before SEC, WPC50 = 45 ± 9.9 nM, diamonds - MBP-R-spondin 1 after SEC, WPC50 = 4.0 ± 0.53 nM and squares - final R-spondin 1 preparation, WPC50 = 4.5 ± 0.80 nM. (**D)** Cellular activities of MBP-R-spondin 1 and CS-R-spondin 1 measured as WPC50s: circles - MBP-R-spondin 1 after SEC, squares - CS-R-spondin 1, WPC50 = 1.2 ± 0.69 nM. To delineate between the MBP-R-spondin 1 and CS-R-spondin 1 dose-response data sets, I have limited presentation of the results to protein concentrations up to 50 nM.

R-spondin is an extracellular protein and its activity is particularly sensitive to the correct configuration of disulphide linkages. In addition to the R-spondin 1 expression cassette, pR-spondin 1 drives expression of *DsbC*, a disulphide isomerase that promotes the correct configuration of disulphide linkages. R-spondin 1 was expressed in the NEB Shuffle® T7 *E. coli* strain that also expresses additional *DsbC* in the cytoplasm. Moreover, following lysis of expressing bacteria and batch purification by Nickel-NTA agarose, expressed R-spondin 1 was subject to an *in vitro* disulphide shuffling step in the presence of reduced and oxidised glutathione. Nonetheless, a subsequent SEC step removes approximately 60% of the inactive R-spondin 1 protein aggregates found in the void volume. Non-aggregated MBP-R-spondin 1 resolves in fractions corresponding to the correct molecular mass of 58 kDa and is an essentially pure preparation demonstrated by migration as a single band of protein by SDS-

PAGE (Figure 25B). Cellular activities of the MBP-R-spondin 1 fractions, measured as WPC50s, increase greater than 10-fold with SEC, from 45 ± 9.9 nM to 4.0 ± 0.53 nM (Figure 25C; Table 7). The purified MBP-R-spondin 1 can be used to supplement organoid media at a concentration of 25 nM.

Table 7: Cellular activities and endotoxin contamination in R-spondin 1 and Gremlin 1 fractions.

Protein	Cellular activity* (nM)	Endotoxin levels (EU/mg protein)	Endotoxin levels in organoid media EU/ml	Estimated cost per litre of organoid media (£)**
MBP-R-spondin 1 before SEC	45 ± 9.9	1.01	0.001	n/a
MBP-R-spondin 1 after SEC	4.0 ± 0.53	4.22	0.005	<10
R-spondin 1	4.5 ± 0.80	17.41	0.026	<10
CS-R-spondin 1	1.2 ± 0.69	9.18	0.005	>5,000
Gremlin 1	6.4 ± 0.65	6.66	0.002	<10
CS-Gremlin 1	6.0 ± 0.33	not detected	not detected	>3,500

*Cellular activities for R-spondin 1 are calculated as WPC50 and for Gremlin 1 are calculated as IC50 for BMP2-induced ALP. **Cost of materials.

Removal of the MBP moiety by Thrombin cleavage followed by cation exchange chromatography produces pure R-spondin 1 with a WPC50 of 4.5 ± 0.80 nM, similar to MBP-R-spondin 1 after SEC (Figure 25C; Table 6). R-spondin 1 migrates on SDS-PAGE gels as a single band at the expected molecular mass of 16 kDa (Figure 25B). Commercially sourced R-spondin 1 (CS-R-spondin 1) retains 118 amino acids at the C-terminus not present in bacterially derived R-spondin 1 and migrates as a 42 kDa protein instead of the predicted 26 kDa, presumably due to glycosylation (Figure 25B). The WPC50 of CS-R-spondin 1 was determined to be 1.2 ± 0.69 nM, based on the assumption of a single protein species with a molecular mass of 42 kDa (Figure 25D; Table 7). A typical yield of purified MBP-R-spondin 1 and R-spondin 1 is approximately 2.5 mg and 1 mg of protein per L of Luria broth. For organoid culture, 2.5 mg or 1 mg of the respective proteins is sufficient to supplement approximately 3 L of organoid media at a concentration of 25 nM (Table 8). Typical levels of contaminating endotoxin in the final MBP-R-spondin 1 and R-spondin 1 preparations are similar to CS-R-spondin 1 and >20-fold less than the LOCE threshold value of 0.5 EU/ml endotoxin when diluted into organoid culture media (Table 7).

Table 8: Media components for the culture of various organoid types.
Organoids derived from murine small intestinal epithelial, colon epithelial, mammary gland and human colon epithelia and patient-matched colon carcinoma were formulated as previously described[38, 39, 177] and include optimised concentrations of bacterially derived R-spondin 1 and Gremlin 1. *Base media is DMEM/F12 media containing 2 mM glutamine, 10 mM HEPES and 0.5 U/ml Penicillin/streptomycin.

Component	Source of derived organoids				
	murine small intestinal epithelial	murine colon epithelial	murine mammary gland	Human colon epithelia	Human colon carcinoma
Base media*	1×	1×	1×	1×	1×
B27 suppl.	1×	1×	1×	1×	1×
N2 suppl.	1×	1×	1×	1×	1×
NAC	1.25 mM		—	1 mM	1 mM
EGF	50 ng/ml	50 ng/ml	—	50 ng/ml	50 ng/ml
R-spondin 1	25 nM	25 nM	25 nM	25 nM	25 nM
Gremlin 1	25 nM	25 nM	25 nM	25 nM	25 nM
WCM	—	0.5×	—	0.5×	—
NRG-1	—	—	100 ng/ml	—	—
A83-01	—	—	—	500 nM	500 nM
SB202190	—	—	—	10 µM	10 µM
PGE2	—	—	—	10 mM	10 mM
FGF				100 ng/ml	100 ng/ml
Gastrin I	—	—	—	10 nM	10 nM
Nicotinamide	—	—	—	10 mM	10 mM
Primocin	—	—	—	1×	1×

For the production of MBP-R-spondin 1 from 2 L of Luria broth a time commitment of two days is estimated and material cost of <£10 per litre of organoid media (Table 7). In contrast, the cost of CS-R-spondin 1 for one litre of organoid media is >£5,000 per litre.

5.3.2 Production of Gremlin 1

The vector for bacterial expression of recombinant human Gremlin 1 (His-ΔN-Gremlin 1; Figure 26A) and the purification procedure has previously been described [160]. Bacterially expressed recombinant Gremlin 1 associates with inclusion bodies requiring suspension of the protein under denaturing conditions and a refolding/disulphide shuffling step for generation of natively folded protein. Subsequent anion exchange chromatography leads to an essentially pure preparation of Gremlin 1 that can be used in organoid media formulations at a concentration of 25 nM.

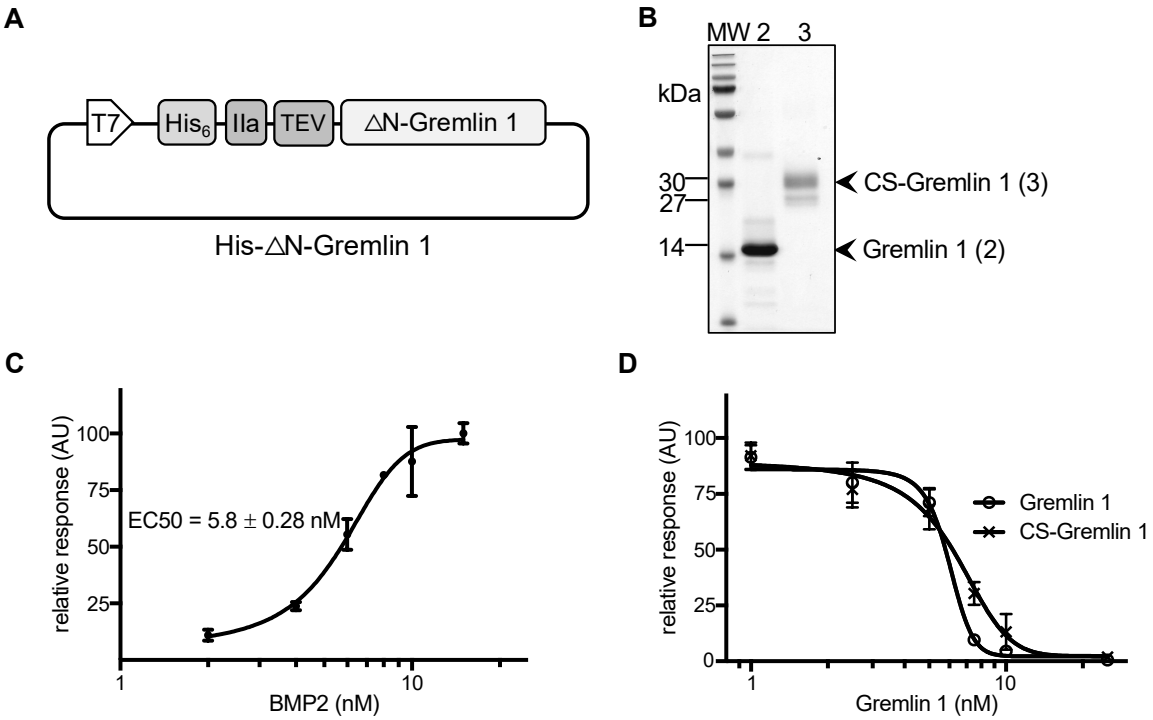


Figure 26: Bacterial expression and production of Gremlin 1
(A) Format of the His-ΔN-Gremlin 1 plasmid for production of Gremlin 14. (**B)** SDS-PAGE of purified Gremlin 1 (lane 2) and commercially sourced (CS)-Gremlin 1 (lane 3). For each protein, 2 µg was loaded. MW, molecular weight standards. (**C)** and (**D**) Activity assay for purified Gremlin 1 based on its inhibition of BMP2-induced ALP activity in C2C12 cells. (**C**) Determination of BMP2 activity to establish a suitable concentration for use in Gremlin 1 assays. (**D**) Inhibition of ALP activity, induced with 8 nM BMP2 treatment, by Gremlin 1 and CS-Gremlin 1. IC50 values for Gremlin 1 and CS-Gremlin were determined as 6.4 ± 0.65 nM and 6.0 ± 0.33 nM respectively.

A final reverse phase chromatography step resolves a highly pure fraction of Gremlin 1 shown by a single band on SDS-PAGE migrating at the expected molecular mass of 14 kDa (Figure 26B). By contrast, the 30 kDa commercially-sourced Gremlin 1 (CS-Gremlin 1) protein, consisting of amino acids 25–284, migrates as two bands on SDS-PAGE gels, at 27 kDa and 30 kDa (Figure 26B). The reduced mobility and diffuseness of CS-Gremlin 1 migration is typical of glycosylated proteins derived from eukaryotic expression systems. CS-Gremlin 1 cellular activity was calculated assuming a single protein species with a molecular mass of 30 kDa.

The cellular activity of bacterially derived Gremlin 1 is measured as the IC_{50} for the inhibition of BMP2-induced ALP activity in C2C12 cells. I first assayed BMP2 ALP activity and determined an EC_{50} value of 5.8 ± 0.28 with a saturating concentration of approximately 12 nM BMP2 (Figure 26C). A sub-saturating value of 8 nM BMP2 was chosen for Gremlin 1 assays (Figure 26C). The IC_{50} value for bacterially expressed Gremlin 1 is 6.4 ± 0.65 nM, comparable to the IC_{50} value for CS-Gremlin of 6.0 ± 0.33 nM (Figure 26C; Table 7). Endotoxin levels for Gremlin 1 preparations are approximately 250-fold lower that the LOCE threshold (Table 7). A typical yield of pure Gremlin 1 is approximately 10 mg per L of 2-YT broth culture. For the

culture of small intestinal epithelial organoids, 10 mg of Gremlin 1 is sufficient for approximately 30 L of media at a concentration of 25 nM. For the production of Gremlin 1 a time commitment of two days is estimated and material cost of <£10 per litre of organoid media (Table 7). Typically, the cost for one litre of organoid media containing CS-Gremlin 1 at a concentration of 25 nM is >£3,500.

5.3.3 Use of R-spondin 1 and Gremlin 1 in organoid media formulation

The ability of bacterially derived R-spondin 1 and Gremlin 1 and the corresponding commercially sourced growth factors, CS-R-spondin 1 and CS-Gremlin 1 to sustain growth of murine small intestinal epithelial organoids was compared. No significant differences equal to or below a threshold value of $p < 0.01$ were found for activity measurements between R-spondin 1 and CS-R-spondin; both were sufficient, at a minimum concentration of 5 nM, to support the growth of organoid cultures (Figure 27A). Crypt multiplicity in organoids was analysed as a more sensitive probe of R-spondin 1 activity. The addition of bacterially derived R-spondin 1 or CS-R-spondin 1 to organoid media resulted in a similar number of crypts per organoid after 8 days of growth (Figure 27B). There were no statistically significant differences in crypt multiplicity amongst the R-spondin 1 treatment groups. Similar results were obtained for CS-R-spondin 1, except that treatment of organoids with 50 nM of the protein led to a significantly higher crypt multiplicity than with 5 nM, but not with 25 nM (Figure 27B). Routinely 25 nM R-spondin 1 were used in organoid culture media.

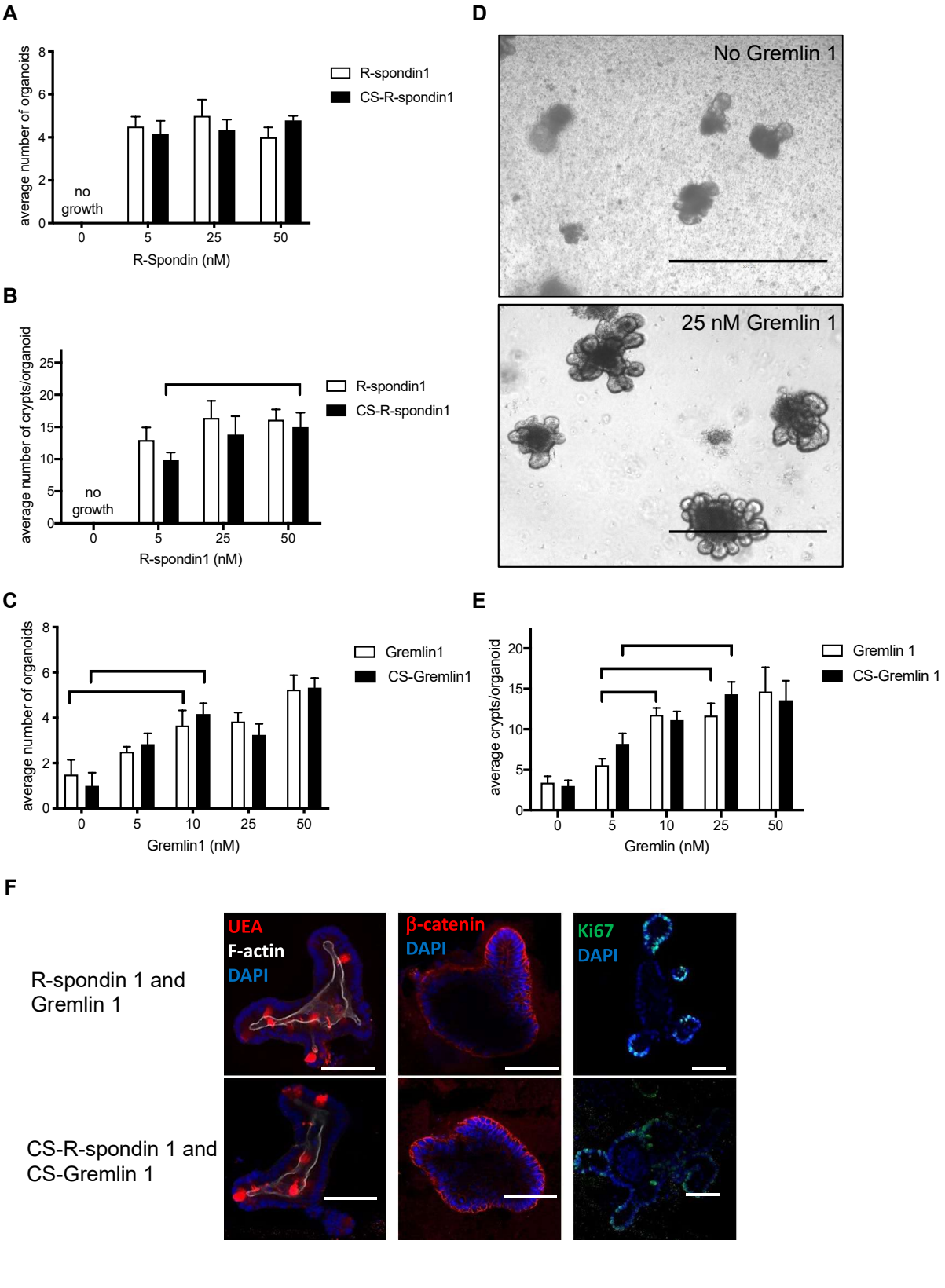

A
average number of organoids
R-spondin1
CS-R-spondin1
no growth
R-Spondin (nM)
B
average number of crypts/organoid
R-spondin1
CS-R-spondin1
no growth
R-spondin1 (nM)
C
average number of organoids
Gremlin1
CS-Gremlin1
Gremlin1 (nM)
D
No Gremlin 1
25 nM Gremlin 1
E
average crypts/organoid
Gremlin 1
CS-Gremlin 1
Gremlin (nM)
F
R-spondin 1 and Gremlin 1
CS-R-spondin 1 and CS-Gremlin 1
UEA
F-actin
DAPI
β-catenin
DAPI
Ki67
DAPI

Figure 27: Optimal concentrations of recombinant R-spondin 1 and Gremlin 1 for organoid growth.
(**A**) Media formulated with R-spondin 1 or CS-R-spondin 1 equally sustains growth of murine small intestinal epithelial organoids at a minimal concentration of 5 nM. Organoid multiplicity was determined as the average count of organoids that grew under each treatment condition averaged across 8 biological replicates. There were no significant differences in organoid multiplicity between the treatment concentrations. No organoid growth was observed in the absence of R-spondin 1 or CS-R-spondin 1. (**B**) Crypt multiplicity per organoid when grown for 8 days in media supplemented with either R-spondin 1 or CS-R-spondin 1 determined as the average crypt count for 10 organoids with each treatment condition, averaged across 8 biological replicates. There was no significant difference in crypt multiplicity amongst R-spondin 1 treatments. For CS-R-spondin ndin 1, *indicates significant difference in crypt multiplicity between concentrations of 50 nM and 5 nM at $p < 0.01$, No organoid growth was observed in the absence of R-spondin 1 or CS-R-Spondin 1. (**C**) Gremlin 1- or CS-Gremlin 1-supplemented media equally sustains maximal organoid growth at a minimum concentration of 10 nM. Organoid multiplicity was determined as the average count of organoids that grew under each treatment condition averaged across 8 biological replicates. *Significant difference in organoid multiplicity between 0 nM and 10 nM treatment groups, $p < 0.01$. Note that Gremlin 1 is not strictly required in growth media as some organoid growth is maintained in its absence. (**D**) Representative examples of organoids cultured in media lacking Gremlin 1 (upper panel) or containing 25 nM Gremlin 1. Scale bar, 200 µm. (**E**) Addition of 10 nM of Gremlin 1 or 25 nM CS-Gremlin 1 is sufficient for maximal crypt multiplicity per organoid after 8 days growth. Average crypt number is the average crypt count for 10 organoids with each treatment condition, averaged across 8 biological replicates. * and ** denote significant differences between 10 nM and 25 nM Gremlin 1 treatment groups and the 5 nM Gremlin 1 treatment group at $p < 0.01$ and $p < 0.001$, respectively. There were no significant differences in crypt multiplicity amongst 10, 25 and 50 nM Gremlin 1 or CS-Gremlin treatment groups. (**F**) Comparison of molecular markers in organoids grown in bacterially derived (top panels) and commercially sourced growth factors (lower panels). Left panels – detection of secretory cells and F-actin using Rhodamine-conjugated UEA and Alexa Fluor 647-conjugated Phalloidin. Middle panels – indirect immune-fluorescent detection of β-catenin. Right panels - indirect immune-fluorescent detection of Ki67.

The addition of Gremlin 1 to organoid media is not strictly required for growth; media formulations that lack Gremlin 1 and other BMP signalling inhibitors (such as Noggin) can sustain growth of murine intestinal epithelial organoids for up to 2–3 passages as previously described [38]. Gremlin 1 or CS-Gremlin 1 have similar activities in organoid media and at 10 nM either protein sustains organoid culture (Figure 27C). Analysis of crypt multiplicity after 8 days growth indicates that minimum concentrations of 10 nM Gremlin 1 or 25 nM CS-Gremlin 1 yield the maximum number of crypts per organoid (Figure 27D, E). Routinely 25 nM of Gremlin 1 were used in organoid media formulations, sufficient for sustaining growth for at least 20 serial passages.

The cellular localisation of selected molecular markers within the epithelial monolayer of organoids culture in media containing either bacterially derived or commercially sourced R-spondin 1 and Gremlin 1 were compared (Figure 27F). No differences in the cellular localisation or distribution of the following molecular probes were detected: UEA, phalloidin, DAPI and antibodies raised against β-catenin or the Ki67 antigen.

5.3.4 Application of the protocol to organoids from other tissues

To test the bacterially derived growth factors in their ability to support organoid growth, organoids derived from a number of tissue types in corresponding media formulations (Table 7) were cultured. These included murine colon epithelia, $APC^{min/-}$ small intestinal tumours and mammary glands (Figure 28A–C) as well as organoids derived from clinical biopsies of human colon epithelia and patient-matched colon cancer tumours (Figure 28D).

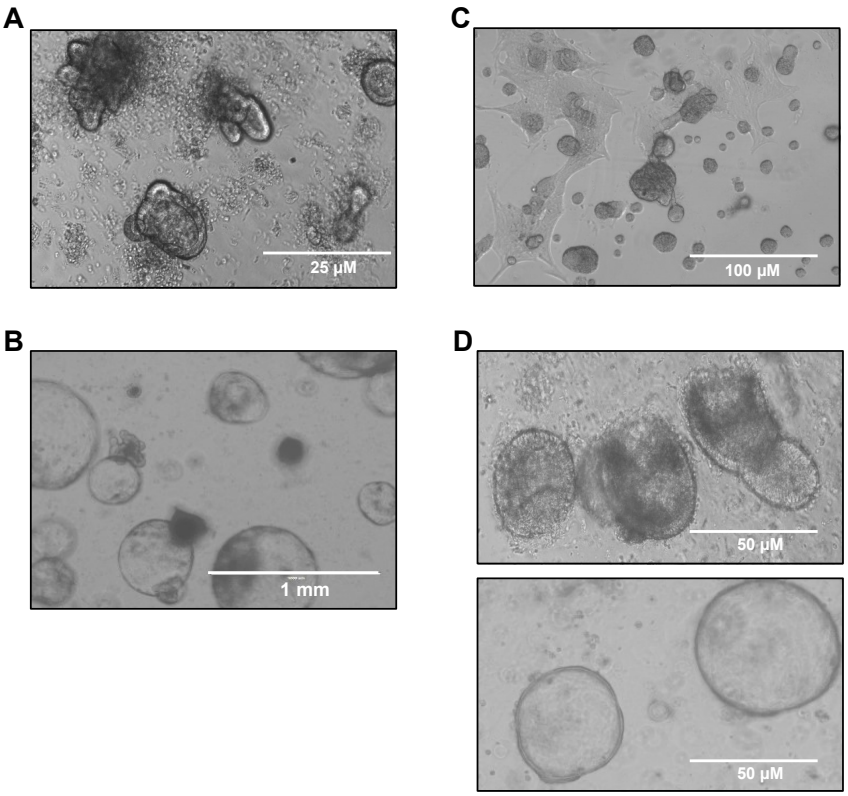


Figure 28: Growth media formulated with R-spondin 1 or Gremlin 1 can sustain growth of various types of organoids
The corresponding media formulations (Table 2) were used for the culture of organoids derived from: (**A**) murine colon epithelia, (**B**) murine APCmin/− small intestinal tumours, (**C**) murine mammary gland and (**D**) human biopsy-derived colon epithelia (top panel) and human colon adenocarcinoma (bottom panel).

In all cases, except for the human biopsy-derived colon epithelial organoids, the derived cultures were expanded a minimum of four passages and frozen for long-term storage in liquid nitrogen. Organoids were thawed, re-cultured and passaged prior to imaging (Figure 28A–D). Human colon epithelial organoids were derived from clinical biopsies and passaged 4 times prior to imaging, as growth and recovery rates after freeze-thaw cycles for these organoids are low, preventing efficient expansion of the cultures. Media containing bacterially expressed R-spondin 1 and Gremlin 1 supports the growth of all other types of organoids for greater than 20 serial passages

5.4 Discussion

Organoid cultures are now a widely used research tool. Current implementation of organoid technology has graduated from proof-in-principle studies to hypothesis testing in the field of epithelial physiology, stem cell homeostasis, and the investigations of genetic diseases such as cancer [48] and cystic fibrosis [179], as this culture method can recapitulate epithelial and tumour microenvironments [180] and host-virus interactions [181]. This chapter reports the production of Gremlin 1 and R-spondin 1 derived from bacterial expression for use in organoid media. Additionally, it demonstrates that bacterially derived R-spondin 1 and Gremlin 1 are comparable to the commercially sourced growth factors in terms cell proliferation and differentiation, as differences in organoid growth or molecular marker localisation within the organoid epithelia with culture in media were not detected. Furthermore, the ability of media formulations containing bacterially derived R-spondin 1 and Gremlin 1 to sustain the growth of organoids derived from a number of tissue types such as murine colon, mammary gland and $APC^{min/-}$ small intestinal tumours and human colon epithelial and colon adenocarcinoma biopsies was demonstrated.

The production of bacterially-derived R-spondin 1 and Gremlin 1 using these protocols has major advantages over conventional sources of the growth factors derived from conditioned media or commercial sources: (i) the R-spondin 1 and Gremlin 1 preparations are highly pure and devoid of impurities such as serum growth factors that may impact organoid growth; (ii) the use of precise cellular activities of the growth factors in organoid media minimises batch-to-batch variation; and (iii) a substantial reduction in cost of organoid culturing. The protocol for generation of media containing recombinant R-spondin 1 and Gremlin 1 is therefore a genuine progression in the ability to generate reproducible growth conditions for both routine organoid culture and large-scale applications such as genetic and chemical screens or clinical tissue biobanking previously reported for limited sample sets of colon cancers [48, 49]. In addition, application of bacterially derived growth factors to the culture of organoids derived from a range of tissues including murine small intestine and colon epithelia, mammary gland and human colon epithelia and colon cancer biopsies was validated.

The use of organoid media formulations with defined cellular activities of growth factors will spur the emergence of new applications for organoids that, in particular, require large-scale, cost-effective and defined culture. This refined system of organoid and tumouroid culture will enable me to study the effect of truncating mutations of TCF7L2 in a defined system that is free from variables created by serum contaminants and batch to batch differences in cellular activity. Using growth factors of defined activity will allow the modulation of Wnt signalling within the intestinal crypt compartment in order to gain insight into functional differences

between full length and truncated TCF7L2 in stem cells as well as Paneth cells. Additionally, it will allow the study of cell type specific regulation of TCF7L2 processing.

One caveat of my protocols is the requirement for an animal cell-derived surrogate extracellular matrix. Currently, reagents such as Matrigel™, are produced from Engelbreth-Holm-Swarm mouse sarcoma cells and may contain serum impurities that impact organoid growth. However, synthetic matrices have emerged in the literature and show substantial promise for supporting organoid growth [182]. Further development of synthetic matrices and their use in conjunction with our protocols will ultimately lead to a fully defined organoid culture system.

Chapter

6 Ligand based LGR5 specific potentiation of Wnt signalling

6.1 Introduction

In Chapter 5 I described the generation of R-spondin 1 from bacterial expression. The growth factor is key for potentiating Wnt pathway activity in the stem cell compartment [183]. Cells found throughout the intestinal epithelia and in organoids stem cell niche express LGR4, however, only stem cells express LGR5. The ligands for LGR family members are R-spondins (Chapter 1). The four members of the R-spondin family are relatively unspecific for binding to the LGR family members [113]. Having established the bacterial R-spondin 1 expression system, I next determined whether I could engineer the protein for LGR5 specificity. In this way, I could use it as a reagent for discriminating stem cells from a range of tissues that exclusively express LGR5 [78, 79, 81, 184].

6.1.1 Aims of this chapter

- Engineer a version of R-spondin that displays enhanced binding to LGR5 and reduced binding to LGR4
- Develop an expression system for the generation of fluorescently labelled forms of the engineered R-spondin
- Determine whether fluorescently labelled, engineered R-spondin can be used to discriminate stem cells expressing LGR5.
- This will enable the investigation of TCF7L2 variant biology with cell type specificity, for instance by probing TCF7L2 bound promotors in stem and Paneth cells in ChIP experiments.

6.2 Materials and Methods

6.2.1 Protein concentration determination

The concentration of purified recombinant proteins was determined using Bradford reagent (Biorad; Philadelphia, USA). The solution was calibrated using a standard absorbance-concentration curve with bovine serum albumin (BSA). Concentration determination was also determined by amino acid hydrolysis (amino acid analysis service, Department of Biochemistry).

6.2.2 SDS-PAGE and Coomassie staining

Proteins were resolved by SDS-polyacrylamide gel electrophoresis as outlined in Chapter 2.4. For staining of protein gels the Instant Blue Coomassie-based staining reagent (Expedeon) was used. The gel was stained for 15 to 20 min in staining solution and subsequently destained using de-ionised water.

6.2.3 Detection of purified biotinylated proteins

Biotinylated proteins (6.2.5) resolved on a SDS-polyacrylamide gel were transferred onto PVDF membranes and proteins were visualised using Ponceau red. The membranes were then incubated in blocking buffer (5 % BSA in TBS-T) for 1 h followed by incubation in blocking buffer containing diluted (1:500) streptavidin-alkaline phosphatase conjugate. Membranes were washed three times in TBS-T for 10 min and detection biotinylated proteins was carried out by BCIP/NBT solution. All procedures were carried out at room temperature.

6.2.4 Expression and purification of human recombinant R-spondin 1, LGR4 and LGR5 ectodomains

Overexpression and purification of R-spondin 1 is outlined in Chapter 2.5.

The plasmids containing the ectodomains of LGR4 or LGR5 with a N-terminal MBP and a C-terminal His_6 tag were expressed under the same conditions as R-spondin, however, induction was carried out by the addition of 0.2 mM IPTG. LGR4 and LGR5 ectodomains consisting of an MBP solubility tag, LRR1-LRR10 of LGR4/5 followed by a LRR domain from the hagfish VLRB protein and a C-terminal His_6-tag were overexpressed in bacteria. Hybrid VLRB LRR has previously been shown to increase solubility of LRR-containing proteins and prevent their aggregation [185]. Due to the occurrence of Leucin rich repeats and Cysteine bonds within the folded ectodomain of LGR4 and LGR5, protein was expressed in Shuffle® cells co-expressing the disulphide bond isomerase DsbC (expressed from a second ORF driven by the T7 bacterial promotor) and purified using Ni^{2+} NTA Agarose.

LGR5 was re-purified using an ÄKTA purifier with an HisTrap 5 ml column, fractionated with an imidazole gradient in order for better resolutions of the protein. Protein fractions were

desalted with PD10 desalting columns (GE Healthcare) pre-equilibrated in gel filtration buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol) according to the manufacturer´s instructions. For future use of both LGR4 and LGR5 ectodomains clean fractions were pooled, concentrated and buffer changed to PBS. Due to contaminants in LGR5 fractions, equal amounts of 80 kDa LGR ectodomain by gel quantification were assayed. The equation calculated from the gel quantification was then used to calculate needed input amounts for R-spondin 1-LGR4/5 interaction assays. Purified proteins were used for pulldown and fluorescence polarisation experiments in PBS along with 1x protease inhibitor cocktail (cOmplete®, Roche).

6.2.5 *In vivo* biotinylation of Avi-tagged R-spondin 1

The plasmid containing Avi-tagged R-spondin 1 (pR-spondin1, Chapter 5) was transformed into BL21(DE3) co-expressing the bacterial biotin ligase BirA. Cells were grown in LB media containing 100 µg/ml ampicillin (to select for the presence of the plasmid) and 10 µg/ml chloramphenicol (BL21 resistance) and 50 µM D-Biotin. Expression was induced by addition of 0.4 mM IPTG at an OD of 0.6-0.7 for approximately 14 hours 16°C followed by purification (Chapter 2.5).

6.2.6 Fluorescent labelling of R-spondin

R-spondin 1 containing a N-terminal tetracysteine tag instead of the Avi-tag was fluorescently labelled by the di-arsenic dye Flash-EDT2. Purified protein was incubated with 10 µM FlAsh-EDT2 reagent, 25-fold molar excess of TCEP and 1 mM EDTA in PBS for 90 minutes at room temperature with gentle agitation. Free dye was removed by desalting using a PD-10 Sephadex G25 column (GE Healthcare) and protein concentration was determined by amino acid analysis. Degree of labelling was determined by spectrophotometry using a nanodrop at 280 nm and 508 nm and calculated using the corresponding protein and dye molecular extinction coefficients. Labelled proteins were visualised by in gel fluorescence detection at 488 nm on a Typhoon™ Scanner (Amersham).

6.2.7 Pulldown assay

R-spondin 1 - LGR binding was investigated using pulldown experiments. MBP-LGR4, MBP-LGR5 or MBP (75 µg each) was immobilised onto 30 µl amylose beads by incubation for 60 minutes with gentle agitation at 4°C in 500 µl Buffer P (50 mM Tris/Cl pH8.0, 150 mM NaCl, 5 % glycerol, 1 mM EDTA). After three washes with Buffer P, 20 µg of FlAsh labelled R-Spondin 1 was added to the beads and incubated for 60 minutes with gentle agitation at 4°C in 500 µl Buffer P. Beads were then washed three times to wash out unbound protein and transferred into 1.5 ml spin columns. MBP-LGR4/5-R-spondin 1 complexes were eluted with 50 µl MBP-

Elution buffer (50 mM Tris/Cl pH 8.0, 150 mM NaCl, 10 mM maltose) for 5 minutes at room temperature followed by centrifugation for 2 minutes at 1000 rpm. A 15 % SDS-PAGE was used to resolve samples alongside 10 % of total protein input (2µg) as control. Bands were visualised using a Silver staining Kit (Invitrogen). Quantification was carried out of three independent assays and significance calculated using multiple t-test.

6.2.8 Fluorescence Polarisation assay

Fluorescence polarisation was used as a quantitative assay of R-spondin-LGR binding. Purified MBP-LGR4 or MBP-LGR5 was titrated into black 384-well microplates with each well containing 20 nM FlAsh labelled R-spondin 1. The range of added MBP-LGR proteins was from 5 nM to 700 nM in 50 mM Tris/Cl buffer (pH 8.0) to a final volume of 20 µl. Fluorescein (5 nM) was used as a calibration control with a value of 35 mP, well within the detection rage for fluorescent, FlAsh-labelled R-spondins. Fluorescence polarisation experiments were carried out using a PHERAStar FS microplate reader (BMG Labtech). Significance was calculated using multiple t-test.

6.2.9 Isolation and dissociation of crypts from murine small intestine

Mice (wildtype B6) were sacrificed by cervical dislocation and small intestine was removed and washed with cold PBS. Around 10 cm of intestine measured from stomach was opened longitudinally, cut into around 5 mm pieces and further washed with cold PBS until the supernatant was clear. The tissue fragments were then incubated in 20 ml 2 mM EDTA with PBS for 30 min with gentle agitation at 4°C. After removal of PBS-EDTA, the tissue fragments were vigorously suspended by using a 20 ml syringe with cold PBS. The supernatant contained the villous fraction and was collected. The sedimented tissue fragments were resuspended with 10 ml 2 mM EDTA/PBS. After further vigorous suspension and sedimentation, the resulting supernatant was collected, and this was repeated 6 times. Crypt enriched fractions were evaluated by inverted light microscopy. These fractions were pooled, resuspended with cold PBS/5% FBS after centrifugation at 1200 rpm for 5 min and passed through a 70-µm cell strainer (BD Bioscience) to remove residual villous material. Isolated crypts were centrifuged at 600 rpm for 2 min to remove single cells. The final fraction consisted of essentially pure crypts which were used for enzymatic dissociation.

Isolated crypts were pelleted by centrifugation at 1200 rpm for 5 min, resuspended in 1x Dispase solution along with 1 µg/ml DNase and incubated at 37 for 20 min. The degree of dissociation was checked by light microscopy. Cells were strained through a 40 µm mesh followed by centrifugation at 800 rpm for 5 min and resuspended in cold FACS buffer (5% FBS in PBS).

Freshly isolated primary single cells were incubated with (i) 50 nM FlAsH-labelled or biotinylated Avi-tagged R-spondin 1 and fluorescently conjugated streptavidin and (ii) EpCAM-PE antibody (1:200) for 1 h on ice in the dark. After washing with FACS buffer cells were visualised using an Accuri C6 Flow Cytometer or sorted by (FACS Core Facility CRUK Cambridge Institute).

6.3 Results

6.3.1 Mutagenesis of R-spondin 1

Sequence comparisons between published data of all 4 R-spondin isoforms and LGR4 and bio-informatic analysis of the R-spondin-LGR4/5 binding interfaces by Dr. Sony Malhotra (University of Cambridge, UK) identified a series R-spondin 1 amino acid residue substitutions that were predicted to increase binding affinity to LGR5 and decrease affinity to LGR4. The four mutations selected were S78D P77H N88G H108Q (Figure 29A-D). The resultant mutagenized R-spondin 1 protein is termed SPNH due to modifications of selected amino acids. Specifically, the Serine 78 to Aspartate mutation was predicted to confer an additional salt bridge to residue Arginine 96 of LGR5 (Figure 29E, F). The mutation at Histidine 108 to the Glutamine was predicted to may influence the positioning of the interacting residues to LGR5 due to a slight sterical change, whereas the Asparagine 88 change to Glycine removed a salt bridge to LGR4 but not LGR5. Finally, the Proline 77 to Histidine substitution was chosen as it was predicted to create an additional interaction to both LGR4 and LGR5 (Figure 29).

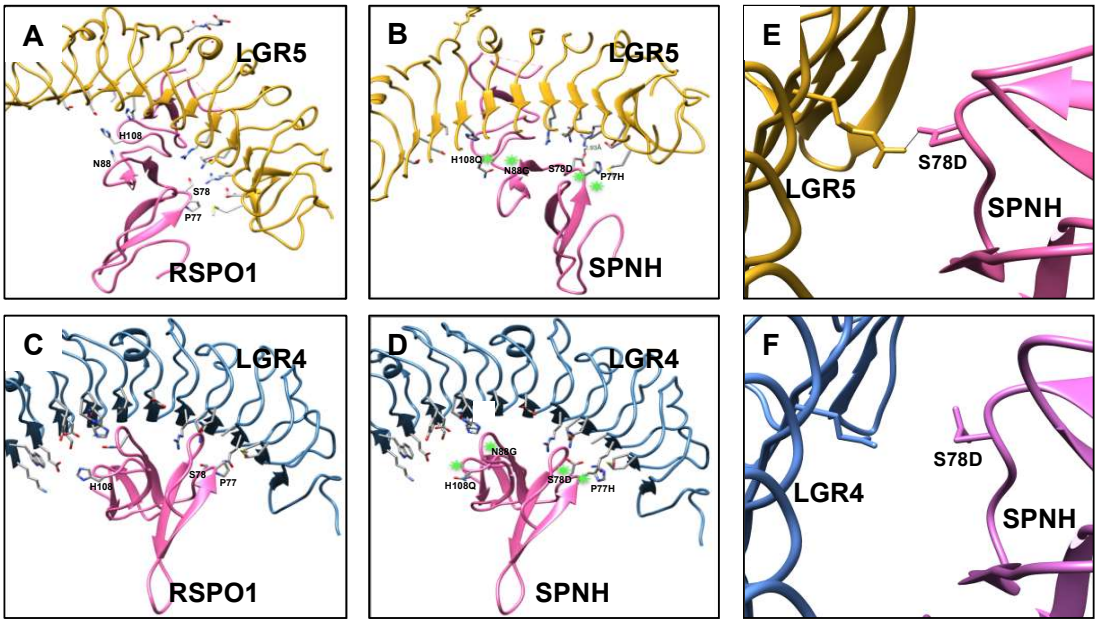


Figure 29: Structural predictions of R-spondin 1 mutagenesis
Bio-informational analysis of published structures of R-spondin 1 bound to LGR4 and LGR5 ectodomains revealed four mutations within R-spondin 1 to change its affinity to both, LGR4 and LGR5. (**A**) Structure of R-spondin 1 bound to LGR5 (**B**) Predicted structure of the R-spondin 1 SPNH mutant bound to LGR5. (**C**) R-spondin 1 bound to LGR4 (**D**) Predicted structure of the R-spondin 1 SPNH mutant bound to LGR4 (**E**) The consequence of the R-spondin 1 S78D mutation showing the newly created salt bridge to amino acid 96 of LGR5 that is not present (**F**) in the predicted binding of R-spondin 1 SPNH to LGR4.

6.3.2 Purification of TC-R-spondin 1 21-145 His_6

The bacterial plasmids created for expression consisted of R-spondin-1, R-spondin 1 F106A (F106A; containing a mutations that abrogates LGR4/5 binding [186]) or SPNH encoding a C-terminal 6x-polyhistidine tag and an N-terminal region consisting of Maltose binding protein, Thrombin cleavage site and tetracysteine (TC) tags (MBP-Th-TC-R-spondin 1; modified from Pioszak AA, Oklahoma University, USA; Figure 30A). The established R-spondin 1 protein purification procedure (Chapter 2.5) yielded the purified proteins at the expected size of 56 kDa (Figure 30B). Thrombin cleavage (Figure 30B) followed by ion exchange chromatography yielded pure TC-R-spondin 1 proteins (Figure 30C).

The purified R-spondin 1 proteins contained the TC tag that was labelled with FlAsH-EDT2 for *in vitro* binding and functional assays (Figure 30D, E).

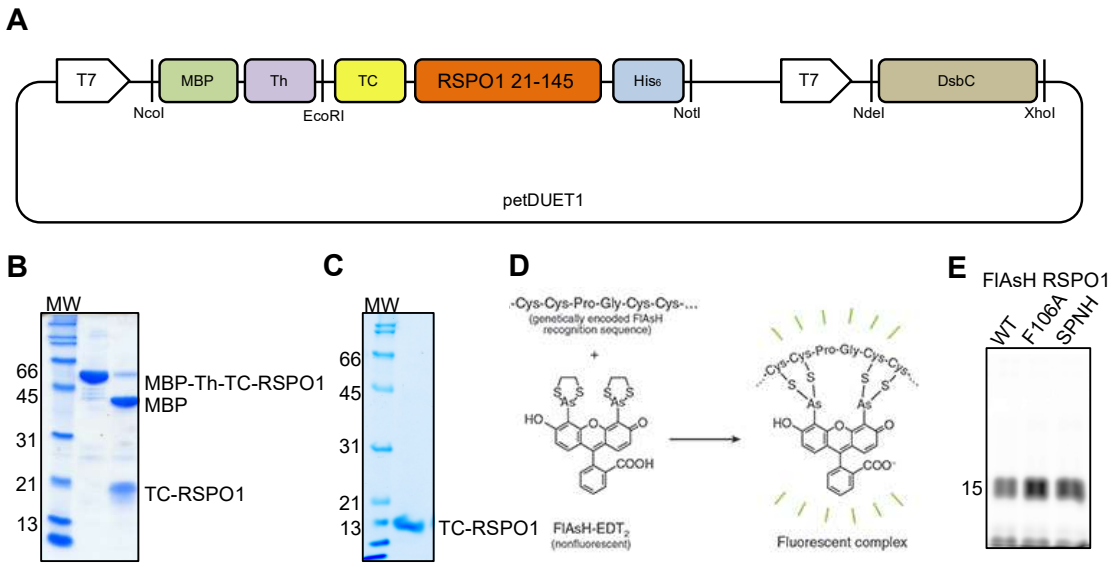


Figure 30: Purification and fluorescent labelling of TC-R-spondin 1
(**A**) Expression construct for R-spondin 1 CRD domain in the vector petDUET1. N-terminally tagged with Maltose Binding Protein (MPB), followed by a Thrombin cleavage site and tetracysteine tag (TC) for fluorescent labelling, C-terminus with His_6-tag (**B**) Over-expressed MBP-Th-TC-R-spondin 1 was purified using Ni^{2+}-NTA agarose and eluted with imidazole. MBP tag was cleaved by Thrombin, (**C**) Thrombin protease and MBP protein was removed by ion exchange chromatography and clean TC-R-spondin 1 fractions were collected and pooled. Protein purity was determined by Coomassie staining after SDS-PAGE. (**D, E**) TC-R-spondin 1 protein was fluorescently labelled with FlAsH-EDT2 reagent and protein visualised by fluorescence detection after SDS-PAGE.

6.3.3 Purification of LGR4 and LGR5 ectodomains

Pure but low yields of the bacterially expressed LGR4 and LGR5 ectodomains were generated for binding assays (Figure 31). Ni^{2+}-NTA batch purification of LGR4 yielded adequate amounts of protein of 80 kDa (Figure 31B). Yields of the LGR5 ectodomain were considerably lower and the preparation contained more contaminants (Figure 31C). LGR4 elution resulted in a single peak of 5 fractions (Figure 31D) whereas the LGR5 ectodomain eluted in a much broader peak of 10 fractions. Both wash steps show strong bands of 80 kDa along with other

proteins, indicating mis-folded or aggregated LGR5 protein that was not sufficiently bound to the Ni^{2+} resin resulting in premature elution with low imidazole concentrations and compromising yields (Figure 31E).

Due to contaminants in LGR5 fractions, equal amounts of 80 kDa LGR ectodomain by gel quantification were assayed (Figure 31F). The equation calculated from the gel quantification was then used to calculate needed input amounts for R-spondin 1-LGR4/5 interaction assays.

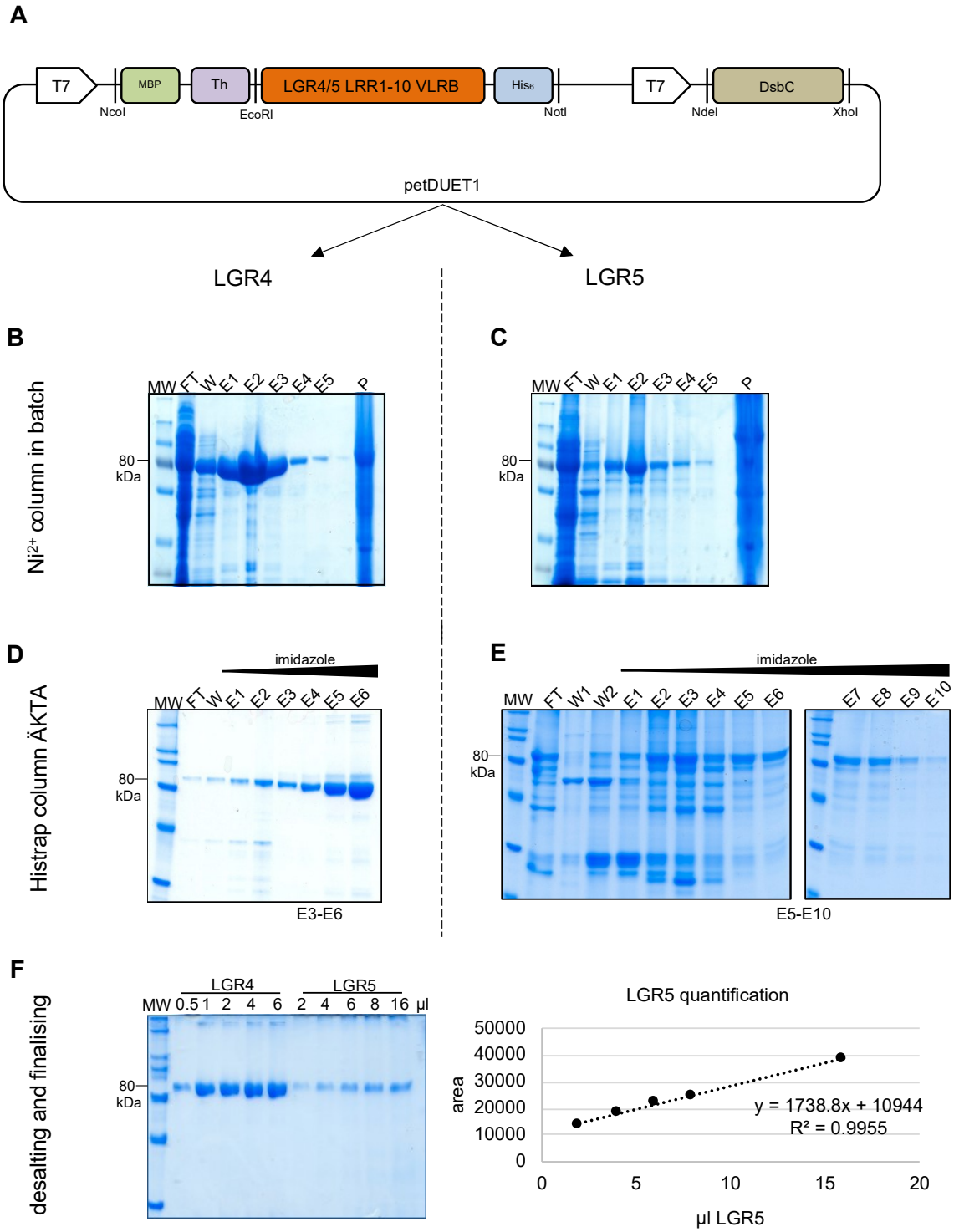


Figure 31: Purification of LGR4 and LGR5 ectodomains
(**A**) scheme of LGR4 or LGR5 ectodomain expression plasmid. Both ectodomains are expressed with a N-terminal MBP and a C-terminal His_6-tag (**B**) 10% SDS PAGE of Ni^{2+}affinity chromatography in batch of MBP-LGR4 visualised with InstantBlue (**C**) 10% SDS PAGE of Ni^{2+}affinity chromatography in batch of MBP-LGR5 visualised with InstantBlue (**D**) Further purification from (B) using a Histrap column FPLC eluted with a linear imidazole gradient., collected fractions indicated (**E**) Further purification from (C) using a Histrap column FPLC eluted with a linear imidazole gradient, collected fractions indicated (**F**) SDS-PAGE of final purification product of MBP-LGR4 and MBP-LGR5 in increasing amounts (left) for relative quantification of LGR5 to LGR4 concentration (right).
MW: molecular weight standard, FT: flow through, E1: elution fraction 1, W: wash, P: pellet

6.3.4 Validation of biological activity

Functional validation of the recombinant FlAsH-labelled R-spondin 1 proteins were tested for their ability to potentiate Wnt pathway activity using 100 nM of the recombinant FlAsH-R-spondin 1 variants. FlAsH-R-spondin 1 and the SPNH mutant were equally potent in potentiating Wnt pathway activity, whereas the FlAsH-R-spondin 1 F106A protein was inactive (Figure 32) reflecting its inability to bind either LGR4 or LGR5.

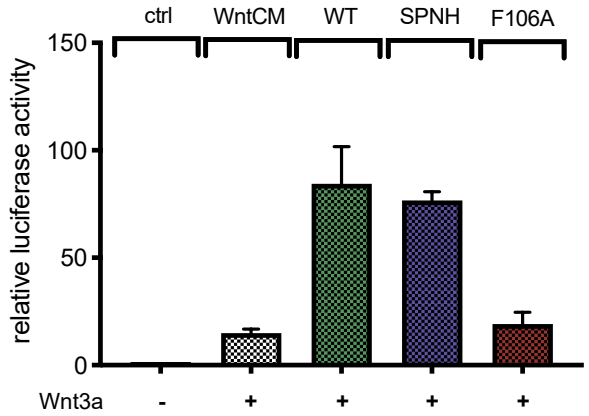


Figure 32: recombinant FlAsH labelled R-spondin 1 increases Wnt signalling upon Wnt induction

TCF/β-catenin mediated reporter assay of HEK293T cells transfected with Super8xTOPFlash and pRL-CMV incubated with Wnt3a conditioned media (Wnt3a) along with 100 nM FlAsH R-spondin 1 WT or SPNH indicated that both recombinant proteins were capable to exert their function leading to elevated levels of Wnt receptors at the membrane resulting in increased signalling output. FlAsH R-spondin 1 F106A mutant didn't potentiate Wnt signalling activity upon Wnt induction confirming no binding to LGR receptors.

6.3.5 Binding assays

Semi-quantitative binding studies were carried by pulldown with His-purified MBP, MBP-LGR5 or MBP-LGR4 immobilised onto amylose beads. The capture beads were incubated with R-spondin 1, R-spondin F106 and R-spondin 1 SPNH (15 kDa) and bound proteins visualised by SDS-PAGE chromatography followed by Silver stain. 10 % of each protein used as was loaded for amount reference (Figure 33A). R-spondin 1 was bound to both MBP-LGR4 or MBP-LGR5, with a greater concentration of protein eluted from LGR4 than LGR5, giving a greater LGR4/LGR5 ratio. R-spondin 1 SPNH was detected in both MBP-LGR4 and MBP-LGR5 as well, however, compared to R-spondin 1, there was less in LGR4 and more in LGR5 bound samples, giving a smaller LGR4/LGR5 ratio. This indicates that R-spondin 1 SPNH binds preferentially to LGR5, compared to LGR4, ectodomains (Figure 33B). R-spondin 1 F106 failed to be isolated by either, MBP-LGR4 or MBP-LGR5 proving its inability to bind to the ectodomains and thus served as a negative control. To control for non-specific binding, MBP only was incubated with all three R-spondin 1 forms and didn't result in any detectable R-spondin 1 protein. All pulldown assays were quantified from 3 independent experiments.

The binding affinities of the R-spondin 1 proteins to LGR5 and LGR4, investigated via fluorescence polarisation (FP) in a pilot experiment (n=1; Figure 33C), indicated that R-spondin 1 had a greater binding affinity to LGR5 relative to LGR4 (Figure 33C). Importantly, SPNH had enhanced binding affinity to LGR5, and weakened binding to LGR4 (Figure 33C). Note, owing to poor binding of SPNH to LGR4, the assay did not reach binding saturation and the apparent K_d value was modelled based on the data that could be derived. Taken together, the SPNH binding to LGR5 was favoured over LGR4 binding by at least two orders of magnitude (Figure 33C).

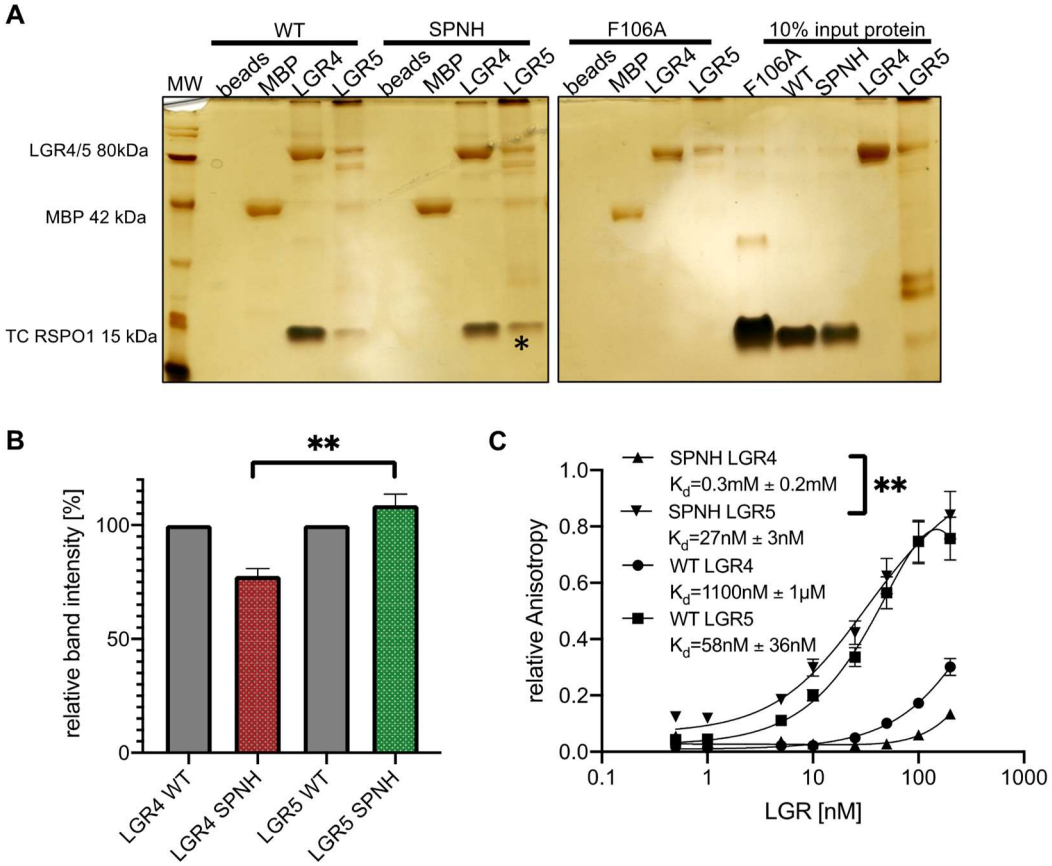


Figure 33: SPNH mutant R-spondin 1 binds stronger to LGR5 and weaker to LGR4
(**A**) The F106A mutant failed to bind both LGR4 and LGR5. Whereas both, WT and SPNH R-spondin 1 bound both LGR4 and 5; more SPNH was recovered from LGR5 binding than for LGR4 binding (asterisk). (**B**) Quantification of 3 independent pulldown experiments by analysing band intensities of R-spondin relative to LGR4/5 ectodomain. (**C**) Fluorescence polarisation of FlAsH labelled R-spondin 1 protein, wildtype or SPNH mutant in the presence of different concentrations of LGR protein titrant. Denote *p-values* (**) are at the 0.006 of significance.

6.3.6 Discrimination of intestinal epithelial stem cells with R-spondin SPNH

A key aim of this chapter was to generate an R-spondin-based probe selective for LGR5-expressing stem cells and tumour stem cells. Dissociated murine intestinal epithelial cells were mixed with either 50 nM FlAsH-labelled R-spondin 1 protein, the non-binding F106A mutant or SPNH. Flow cytometry was performed. The SPNH positive population as well as

cells from the SPNH depleted population were isolated by FACS but resulted in insufficient cell numbers for future gene expression analysis by RT-PCR (n=1) (Figure 34).

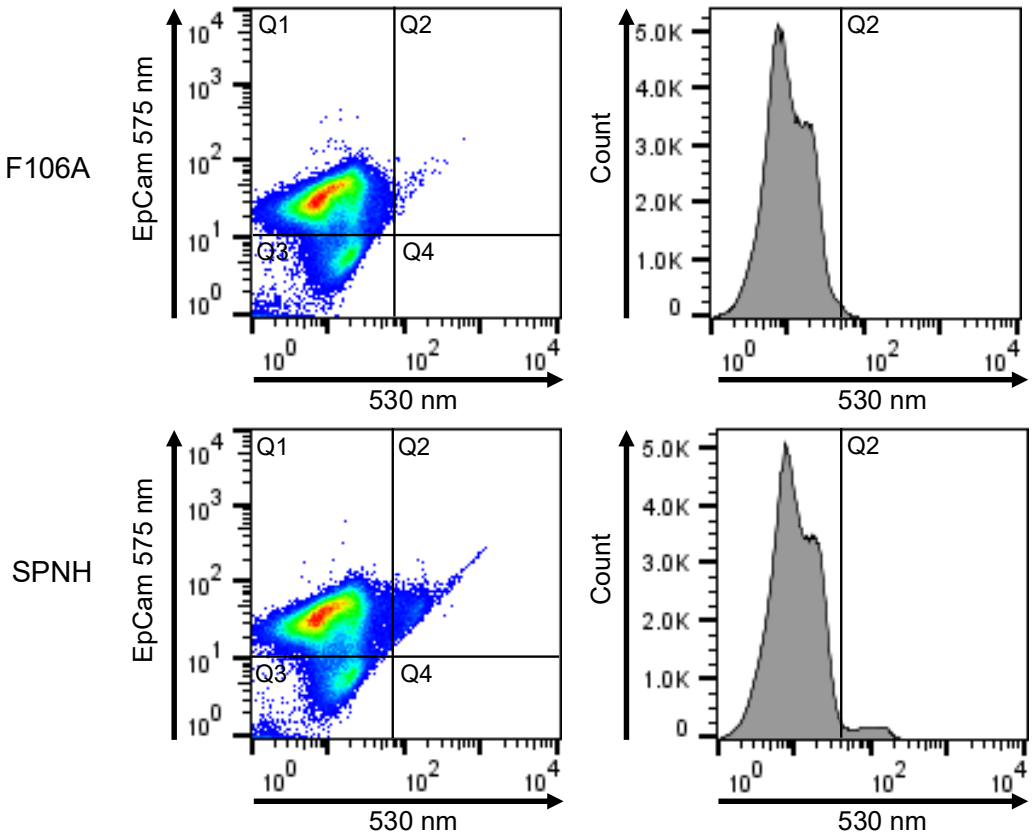


Figure 34: FACS analysis of intestinal single cells reveals SPNH R-spondin 1 mutant to be suitable to specifically label LGR5 positive intestinal cells
Staining of crypt cells with 100 nM FlAsH labelled SPNH R-spondin 1 and EpCam-PE resulted in a population of green fluorescently labelled cells compared to samples stained with non-binding FlAsH-F106A R-spondin 1.

6.3.7 Sorting of immune cells

Although LGR5 has emerged as a ubiquitous marker of epithelial stem cells, it is not clear whether multipotent LGR5 expressing cell populations exist in other tissues. LGR5-expressing immune cells in mice harbouring the LGR5-GFP allele have not been identified [187]. In collaboration with the laboratory of Maike de la Roche (CRUK Cancer Institute), specific T-cell populations expressing LGR5 were identified using FlAsH R-spondin variants. FACS sorting of T-cells derived from inguinal/brachial and mesenteric lymph nodes was carried out on cells stained with FlAsH labelled R-spondin 1, F106A or SPNH and the T-cell surface markers CD3, CD4 and CD8. The FACS protocol was established to discriminate live, singlet cells that expressed the T-cell marker CD8 and were bound by FlAsH-SPNH. T-cells isolated from blood and bone marrow did not contain a population interacting with SPNH (data not shown). However, lymph node-derived T-cells contained a small, discrete population of CD4+ and CD8+ SPNH+ cells (Figure 35). SPNH+ T-cells were also present in the thymus (data not

shown) and a much larger SPNH+ T-cell population was present in the mesenteric lymph nodes (Figure 35). Within the lymph node cells, the F106A control detected 0.8 % of cells (Figure 35A) whereas SPNH detected 3.7 % of the cells (Figure 35A). SPNH positive and SPNH depleted populations were isolated, mRNA extracted and LGR5 analysed by qRT-PCR; unsurprisingly, the SPNH-positive cells contained 9-12-fold more LGR5 transcripts than SPNH-negative cell (Figure 35B). Further qRT-PCR experiments were carried out to phenotypically characterise the SPNH-positive population. Not changed in the expression of LGR4 or c-myc (Figure 35B), transcripts were enriched encoding the Wnt pathway target genes *Tcf1*, and *Lef1* (not shown), which, among other expressed genes, characterise effector memory T-cells [188].

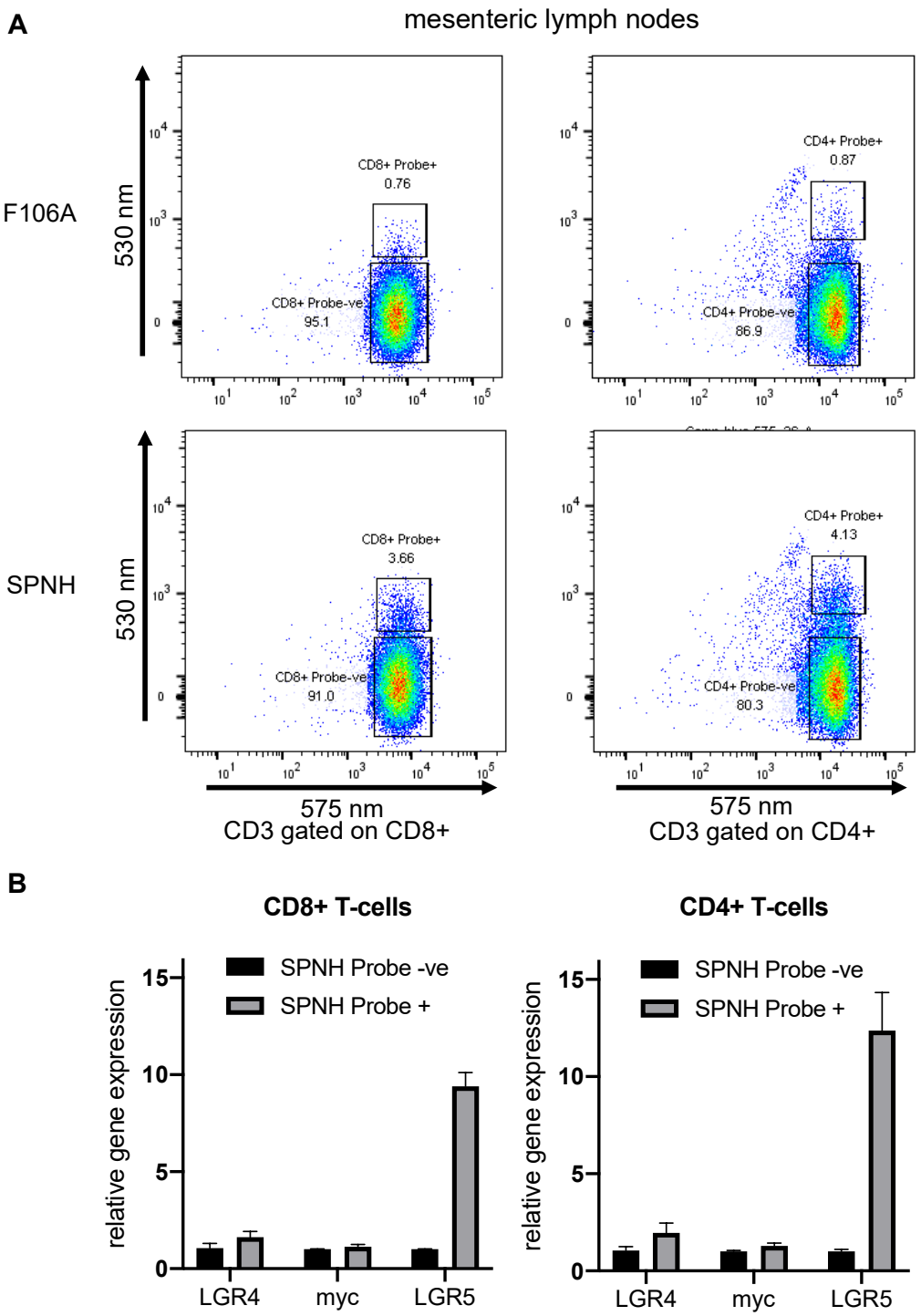


Figure 35: FACS analysis of mesenteric lymph node derived T-cells show a LGR5 positive CD8+ and CD4+ cell population specifically sorted by SPNH R-spondin 1

T-cells derived from mesenteric lymph nodes were sorted with FlAsH labelled non-binding F106A or SPNH R-spondin 1 and gated on the T-cell marker CD8 or CD4. (**A**) F106A staining resulted in a positive population of 0.8 % of live cells for CD8+ cells and 0.9 % of live cells for CD4+ cells. T-cells sorted with FlAsH labelled SPNH R-spondin 1 and gated on CD8+ or CD4+ living cells showed an enrichment of a population of 3.7 % or 4.2 %, respectively. (**B**) qRT-PCR detected LGR5 enrichment of 9 fold in CD8+ SPNH+ populations compared to CD8+ SPNH- population and 12 fold in CD4+ SPNH+ populations compared to CD4+ SPNH- population. Gene expression analysis by qRT-PCR for LGR4 and c-myc didn't show changes in expression in neither CD8+ nor CD4+ sorted T-cell populations.

6.3.8 Adapting detection method - Avi tag

I carried out further method development for fluorescent labelling of R-spondin 1 proteins as FlAsH labelling of tetracysteine tagged R-spondin 1 proteins which was problematic owing to the high disulphide bond content of the protein. The alternate strategy relied the incorporation of an N-terminal Avi tag (MBP-Th-Avi-R-spondin 1 wildtype = pR-spondin1, SPNH and F106A) that become biotinylated when the protein was expressed in the *E.coli* strain BL21(DE3) expressing the biotin ligase BirA and grown in the presence of biotin. Biotinylation of the MBP-Th-Avi-R-spondin 1 proteins was verified by Western blot and probing PVDF membranes with AP conjugated Streptavidin; purified, biotinylated MBP tagged protein yielded an expected band of 56 kDa before cleavage of the MBP moiety and 16 kDa for the biotinylated R-spondin 1 proteins after cleavage (Figure 36A). The biotinylated protein was then treated with streptavidin-PE. purified by anion exchange chromatography and R-spondins activity was tested for Wnt pathway activity potentiation using the TOPFlash reporter assay (Figure 36B). I observed no differences in WPA in the presence and absence of Streptavidin-PE, confirming that the 290 kDa streptavidin moiety did not affect function. All expressed R-spondin 1 proteins whether biotinylated or not, or bound to Streptavidin-PE, had similar activities (Figure 36B).

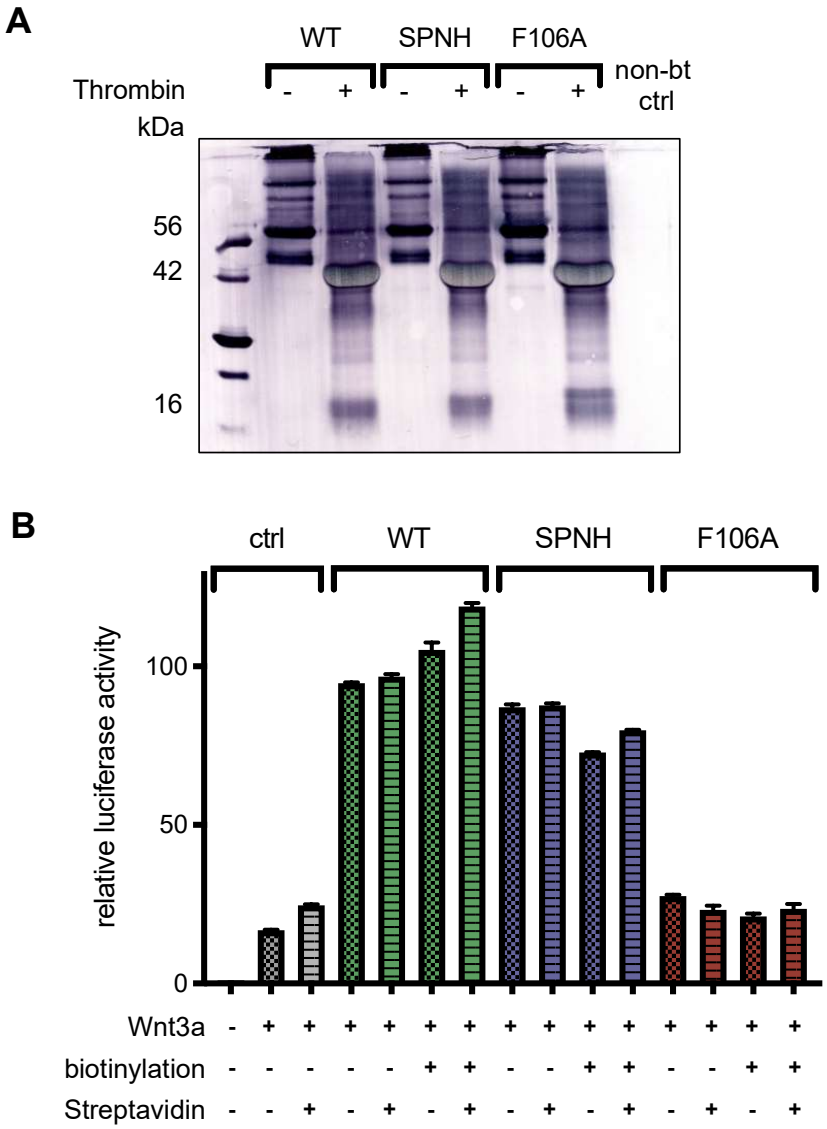


Figure 36: purified biotinylated R-spondin 1 increases Wnt signalling upon Wnt induction regardless of streptavidin binding
(**A**) Purified biotinylated MBP-Th-Avi-R-spondin 1 wildtype, SPNH and F106A mutants and Thrombin cleavage. Biotinylation of the purified protein was validated by Western blotting and probing with Streptavidin-AP. (**B**) TCF/β-catenin mediated reporter (TopFlash) assays of HEK293T cells treated with Wnt3A and 100 nM biotinylated (bt) or non-biotinylated R-spondin 1 WT or SPNH or R-spondin 1 F106A mutant.

6.4 Discussion

6.4.1 SPNH – a probe for cells expressing the LGR5 stem cell marker

Bioinformatic and structural analysis of RSPO1-LGR5 binding have been used to engineer the SPNH mutant form of RSPO1, designed to increases binding affinity towards LGR5 and reduce affinity to LGR4. Data derived from a comprehensive set of *in vitro* and biological assays show that SPNH; (1) demonstrates enhanced binding affinity and specificity for LGR5 relative to R-spondin 1 (2) recombinant SPNH is active and potentiates the activity of the Wnt signalling pathway and (3) SPNH selectively discriminates for an LGR5 enriched cell population of T-cells and a population of dissociated intestinal crypt cells.

While my data conclusively demonstrates the increased window for LGR5 specificity from 58 nM to 27 nM for SPNH, the very rough Kd values I obtained using fluorescence polarisation should be further refined using a different analytical method such as Isothermal Titration Calorimetry and/or Microscale Thermophoresis. Both methods will require further optimisation of the expression and purification protocols for LGR receptor ectodomains and RSPOs variants. The high content of disulphide bonds is an issue as it is crucial that RSPO1 is produced in a properly folded state. Hence, an alternate protein expression system may be required, for instance the use of eukaryotic cell expression systems to optimise protein yield. Again, this may pose problems in terms of inappropriate and heterogeneous patterns of glycosylation on the expressed proteins.

To prove that SPNH is specifically binding to LGR5 a simpler quality control system should be in place. For instance, a cell line stably expressing functional surface exposed LGR5 ectodomain that can then be visualised by labelled SPNH in FACS or in immunohistochemical staining of the cell line. Creation of this system was intended by transfection of a LGR5 ectodomain fusion to the transmembrane domain of the V2R receptor (LGR5-V2R-GFP, kind gift from the MG Caron, Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710, USA) into HEK293T cells that endogenously express a neglectable amount of LGR5. Visualisation of the fusion protein revealed aggregated internalised protein rather than constitutive cell surface expression [189], making it unsuitable as a quality control system. Protein LGR5-LGR4 fusions were realised, however also appeared to be aggregated internalised clusters.

One innovation I made for RSPO1 variant production was the use of the Avi-tag system to generate biotinylated protein. The tractability afforded with the streptavidin linkage will expand in a number of modalities – for instance fluorescent or enzymatic conjugates – thereby increasing potential applications.

One goal of generating the SPNH variant was to develop a reagent for detecting LGR5 expressing stem cells from epithelial tissues. Function validation of intestinal epithelial stem cell population using the SPNH probe should be tested by the ability to seed organoids [38].
One system in which I was able to validate the utility of the SPNH probe was in the isolation of LGR5-expressing T cells, in collaboration with the laboratory of Maike de la Roche. New insights can now be gained into the SPNH+ T-cell population(s), their identity and their role in the immune system.
LGR5 is not only a stem cell marker but potentially a biomarker for colorectal cancer cells, as many studies confirm an overexpression in colorectal tumour samples. One avenue that could be explored is whether SPNH can function as a targeting molecule for directed delivery of streptavidin-conjugated cellular toxophores to LGR5 over-expressing cells. This approach can be explored alongside an ADC treatment option.

6.4.2 The outcome

Taken together, these data demonstrate that LGR5-expressing cells can be discriminated using an engineered version of R-Spondin. Further development of the production of labelled SPNH protein will no doubt provide future applications for the technology in the field of cancer and stem cell research.
Due to technical obstacles and the lack of reproducibility in the production of fluorescently labelled R-Spondin protein with detectable binding to LGR ectodomains and *ex-vivo* single cells, this promising project was transferred to Cambridge Enterprise to receive biochemical improvement and commercial expertise.

Chapter
7 Conclusion

7.1 Conclusion

TCF7L2 is a key transcription factor regulating the Wnt dependent genetic program in intestinal epithelia. TCF7L2 is expressed as long and short variants thought to be the product of alternative splicing. The present study demonstrates that instead, the long form of the TCF7L2 protein is produced that is then processed, post-translationally to the short form. The mechanism underpinning the expression ratio of long and short forms of TCF7L2 is important as it defines the Wnt pathway transcriptional program and its biological outputs.

TCF7L2 variants differ from each other by the absence or presence of the C-terminal domain comprising the auxiliary DNA binding domain of the C-clamp and the interaction sites with CtBP. Therefore, the C-terminus of TCF7L2 appears to have important function in regulating stem cell homeostasis and interplay with other signalling pathways influencing differentiation and other cellular processes.

This study demonstrates that in colorectal cancer, exon 18 microsatellite mutations in *TCF7L2* lead to the expression of a truncated protein that is functionally identical to the short, processed form. I demonstrate that the expressed mutant short TCF7L2 endows CRC cells with specific biological properties.

To study TCF7L2 processing as well as the functional impact of TCF7L2 variants in complex tissue resembling systems I have developed a protocol for the precise and reproducible culture of organoids through the generation of the growth factors R-spondin and gremlin of defined activity. Ultimately, this will be useful for determining TCF7L2 biology within the context of intestinal epithelia.

I have extended my studies of R-spondin by engineering a form of the protein that displays specificity for LGR5 binding and not LGR4 binding. The engineered R-spondin 1 will be useful for future studies geared towards identifying stem cells and targeting cancer stem cells.

7.2 Future perspectives

The results of this study have expanded our view of TCF7L2 activity in stem cell homeostasis, differentiation and cancer biology and have given rise to future arenas of research. These include determining:

7.2.1 The molecular mechanism of TCF7L2 processing and its regulation

My work demonstrates that TCF7L2 post-translational processing appears to be uniquely specific to TCF7L2. It will of course be essential to determine the mechanism. Whether this involves a site-specific protease is responsible or even through a self-cleaving mechanism could be determined by isolating subcellular fractions with processing activity and carrying mass spectrometry for identification. Once the mechanism is revealed further studies can begin to uncover mechanism and regulation.

7.2.2 How TCF7L2 A9 mutations impact CRC pathology

Identifying TCF7L2 truncating mutations in clinical cohorts and their correlation with expressed markers will ultimately lead to a new CRC classification – this will pave the way to subset specific treatment options and clinical decision-making.

7.2.3 The functional impact of TCF7L2 long to short expression ratios

This study has revealed that altering the cellular ratio of long and short forms of TCF7L2 impacts gene expression. The question remains whether different intestinal epithelial cell types display different ratios and whether this drives cell specific functions. This will need to be examined in an experimentally tractable complex epithelial model such as organoids. My work refining organoid culturing technology will pave the way for reproducible experiments elucidating the requirements and mechanisms underlying changes in the long to short TCF7L2 ratio and the corresponding effect on cell-specific epithelial biology.

7.2.4 How the TCF7L2 C-terminus is co-regulated by interacting factors?

As this thesis shows that the lack of the C-terminus is resulting in a different cellular response, the biology of the C-terminus should be studied in the future. Not only the biology and target gene preferences of TCF7L2 variants harbouring a C-clamp is affected by short TCF7L2 variants. Interaction of co-repressing or co-activating factors at the C-terminus of TCF7L2 will be interesting to investigate by co-immunoprecipitations and mass-spectrometric analysis. Further, these interactions may also be regulated, impacting full length TCF7L2 biology itself.

7.3 Concluding remarks

This work shows that truncated TCF7L2 has tumour suppressive function via a variety of cellular processes. Further studies will pave the way to find therapeutic entry points and to apply this diagnostically for tumour subset specific treatment options.

8 References

1. Guinney, J., et al., *The consensus molecular subtypes of colorectal cancer.* Nat Med, 2015. **21**(11): p. 1350-6.
2. Groden, J., et al., *Identification and characterization of the familial adenomatous polyposis coli gene.* Cell, 1991. **66**(3): p. 589-600.
3. Plawski, A., et al., *Familial adenomatous polyposis of the colon.* Hered Cancer Clin Pract, 2013. **11**(1): p. 15.
4. Vogelstein, B., et al., *Genetic alterations during colorectal-tumor development.* N Engl J Med, 1988. **319**(9): p. 525-32.
5. Bettington, M., et al., *The serrated pathway to colorectal carcinoma: current concepts and challenges.* Histopathology, 2013. **62**(3): p. 367-86.
6. Issa, J.P., *CpG island methylator phenotype in cancer.* Nat Rev Cancer, 2004. **4**(12): p. 988-93.
7. Toyota, M., et al., *CpG island methylator phenotype in colorectal cancer.* Proc Natl Acad Sci U S A, 1999. **96**(15): p. 8681-6.
8. Weisenberger, D.J., et al., *CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer.* Nat Genet, 2006. **38**(7): p. 787-93.
9. Thibodeau, S.N., et al., *Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1.* Cancer Res, 1998. **58**(8): p. 1713-8.
10. Polakis, P., *The many ways of Wnt in cancer.* Curr Opin Genet Dev, 2007. **17**(1): p. 45-51.
11. Al-Sohaily, S., et al., *Molecular pathways in colorectal cancer.* J Gastroenterol Hepatol, 2012. **27**(9): p. 1423-31.
12. Sancho, E., E. Batlle, and H. Clevers, *Signaling pathways in intestinal development and cancer.* Annu Rev Cell Dev Biol, 2004. **20**: p. 695-723.
13. Clevers, H., *Wnt/beta-catenin signaling in development and disease.* Cell, 2006. **127**(3): p. 469-80.
14. Ring, A., Y.M. Kim, and M. Kahn, *Wnt/Catenin Signaling in Adult Stem Cell Physiology and Disease.* Stem Cell Rev, 2014.
15. MacDonald, B.T. and X. He, *Frizzled and LRP5/6 receptors for Wnt/beta-catenin signaling.* Cold Spring Harb Perspect Biol, 2012. **4**(12).
16. Gao, C., G. Xiao, and J. Hu, *Regulation of Wnt/beta-catenin signaling by posttranslational modifications.* Cell Biosci, 2014. **4**(1): p. 13.

17. Hao, H.X., et al., *ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner.* Nature, 2012. **485**(7397): p. 195-200.
18. Koo, B.K., et al., *Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors.* Nature, 2012. **488**(7413): p. 665-9.
19. de Lau, W., et al., *Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling.* Nature, 2011. **476**(7360): p. 293-7.
20. Inamura, K., *Colorectal Cancers: An Update on Their Molecular Pathology.* Cancers (Basel), 2018. **10**(1).
21. Sparks, A.B., et al., *Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer.* Cancer Res, 1998. **58**(6): p. 1130-4.
22. Schatoff, E.M., B.I. Leach, and L.E. Dow, *Wnt Signaling and Colorectal Cancer.* Curr Colorectal Cancer Rep, 2017. **13**(2): p. 101-110.
23. Giannakis, M., et al., *RNF43 is frequently mutated in colorectal and endometrial cancers.* Nat Genet, 2014. **46**(12): p. 1264-6.
24. Wee, P. and Z. Wang, *Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways.* Cancers (Basel), 2017. **9**(5).
25. Hallberg, B., S.I. Rayter, and J. Downward, *Interaction of Ras and Raf in intact mammalian cells upon extracellular stimulation.* J Biol Chem, 1994. **269**(6): p. 3913-6.
26. Hill, C.S. and R. Treisman, *Transcriptional regulation by extracellular signals: mechanisms and specificity.* Cell, 1995. **80**(2): p. 199-211.
27. Gaestel, M., *MAPKAP kinases - MKs - two's company, three's a crowd.* Nat Rev Mol Cell Biol, 2006. **7**(2): p. 120-30.
28. Alessi, D.R., et al., *Characterization of a 3-phosphoinositide-dependent protein kinase which phosphorylates and activates protein kinase Balpha.* Curr Biol, 1997. **7**(4): p. 261-9.
29. Stambolic, V., et al., *Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN.* Cell, 1998. **95**(1): p. 29-39.
30. Krasinskas, A.M., *EGFR Signaling in Colorectal Carcinoma.* Patholog Res Int, 2011. **2011**: p. 932932.
31. Takane, K., et al., *DNA methylation epigenotype and clinical features of NRAS-mutation(+) colorectal cancer.* Cancer Med, 2017. **6**(5): p. 1023-1035.
32. Lochhead, P., et al., *Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication.* J Natl Cancer Inst, 2013. **105**(15): p. 1151-6.
33. Summers, M.G., et al., *BRAF and NRAS Locus-Specific Variants Have Different Outcomes on Survival to Colorectal Cancer.* Clin Cancer Res, 2017. **23**(11): p. 2742-2749.

34. Papadatos-Pastos, D., et al., *The role of the PI3K pathway in colorectal cancer.* Crit Rev Oncol Hematol, 2015. **94**(1): p. 18-30.
35. Samuels, Y., et al., *High frequency of mutations of the PIK3CA gene in human cancers.* Science, 2004. **304**(5670): p. 554.
36. Wang, Q., et al., *PIK3CA mutations confer resistance to first-line chemotherapy in colorectal cancer.* Cell Death Dis, 2018. **9**(7): p. 739.
37. Nassif, N.T., et al., *PTEN mutations are common in sporadic microsatellite stable colorectal cancer.* Oncogene, 2004. **23**(2): p. 617-28.
38. Sato, T., et al., *Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche.* Nature, 2009. **459**(7244): p. 262-5.
39. Sato, T., et al., *Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium.* Gastroenterology, 2011. **141**(5): p. 1762-72.
40. McCracken, K.W., et al., *Modelling human development and disease in pluripotent stem-cell-derived gastric organoids.* Nature, 2014. **516**(7531): p. 400-4.
41. Lancaster, M.A., et al., *Cerebral organoids model human brain development and microcephaly.* Nature, 2013. **501**(7467): p. 373-9.
42. Broutier, L., et al., *Culture and establishment of self-renewing human and mouse adult liver and pancreas 3D organoids and their genetic manipulation.* Nat Protoc, 2016. **11**(9): p. 1724-43.
43. Hindley, C.J., L. Cordero-Espinoza, and M. Huch, *Organoids from adult liver and pancreas: Stem cell biology and biomedical utility.* Dev Biol, 2016. **420**(2): p. 251-261.
44. Sato, T., et al., *Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts.* Nature, 2011. **469**(7330): p. 415-8.
45. Sato, T. and H. Clevers, *Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications.* Science, 2013. **340**(6137): p. 1190-4.
46. Farin, H.F., et al., *Visualization of a short-range Wnt gradient in the intestinal stem-cell niche.* Nature, 2016. **530**(7590): p. 340-3.
47. Bartfeld, S., et al., *In vitro expansion of human gastric epithelial stem cells and their responses to bacterial infection.* Gastroenterology, 2015. **148**(1): p. 126-136 e6.
48. van de Wetering, M., et al., *Prospective derivation of a living organoid biobank of colorectal cancer patients.* Cell, 2015. **161**(4): p. 933-45.
49. Fujii, M., et al., *A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis.* Cell Stem Cell, 2016. **18**(6): p. 827-38.

50. Vlachogiannis, G., et al., *Patient-derived organoids model treatment response of metastatic gastrointestinal cancers.* Science, 2018. **359**(6378): p. 920-926.
51. Assawachananont, J., et al., *Transplantation of embryonic and induced pluripotent stem cell-derived 3D retinal sheets into retinal degenerative mice.* Stem Cell Reports, 2014. **2**(5): p. 662-74.
52. Mandai, M., et al., *iPSC-Derived Retina Transplants Improve Vision in rd1 End-Stage Retinal-Degeneration Mice.* Stem Cell Reports, 2017. **8**(4): p. 1112-1113.
53. Shirai, H., et al., *Transplantation of human embryonic stem cell-derived retinal tissue in two primate models of retinal degeneration.* Proc Natl Acad Sci U S A, 2016. **113**(1): p. E81-90.
54. Yui, S., et al., *Functional engraftment of colon epithelium expanded in vitro from a single adult Lgr5(+) stem cell.* Nat Med, 2012. **18**(4): p. 618-23.
55. Hong, S.N., et al., *Concise Review: The Potential Use of Intestinal Stem Cells to Treat Patients with Intestinal Failure.* Stem Cells Transl Med, 2017. **6**(2): p. 666-676.
56. Fujii, M., S. Sugimoto, and T. Sato, *Linking human intestinal scaffolds and organoids to combat intestinal failure.* Nat Med, 2020. **26**(10): p. 1517-1518.
57. Schwank, G., et al., *Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients.* Cell Stem Cell, 2013. **13**(6): p. 653-8.
58. Barker, N., et al., *Crypt stem cells as the cells-of-origin of intestinal cancer.* Nature, 2009. **457**(7229): p. 608-11.
59. Salzman, N.H., et al., *Protection against enteric salmonellosis in transgenic mice expressing a human intestinal defensin.* Nature, 2003. **422**(6931): p. 522-6.
60. Bevins, C.L. and N.H. Salzman, *Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis.* Nat Rev Microbiol, 2011. **9**(5): p. 356-68.
61. Kim, T.H., S. Escudero, and R.A. Shivdasani, *Intact function of Lgr5 receptor-expressing intestinal stem cells in the absence of Paneth cells.* Proc Natl Acad Sci U S A, 2012. **109**(10): p. 3932-7.
62. Durand, A., et al., *Functional intestinal stem cells after Paneth cell ablation induced by the loss of transcription factor Math1 (Atoh1).* Proc Natl Acad Sci U S A, 2012. **109**(23): p. 8965-70.
63. Semenov, M.V., et al., *SnapShot: Noncanonical Wnt signaling pathways.* Cell, 2007. **131**(7).
64. Kohn, A.D. and R.T. Moon, *Wnt and calcium signaling: beta-catenin-independent pathways.* Cell Calcium, 2005. **38**(3-4): p. 439-46.
65. Simons, M. and M. Mlodzik, *Planar cell polarity signaling: from fly development to human disease.* Annu Rev Genet, 2008. **42**: p. 517-40.

66. Jones, C. and P. Chen, *Planar cell polarity signaling in vertebrates.* Bioessays, 2007. **29**(2): p. 120-32.
67. He, X.C., et al., *BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling.* Nat Genet, 2004. **36**(10): p. 1117-21.
68. Dionne, M.S., W.C. Skarnes, and R.M. Harland, *Mutation and analysis of Dan, the founding member of the Dan family of transforming growth factor beta antagonists.* Mol Cell Biol, 2001. **21**(2): p. 636-43.
69. Gomes Fernandes, M., et al., *BMP-SMAD Signaling Regulates Lineage Priming, but Is Dispensable for Self-Renewal in Mouse Embryonic Stem Cells.* Stem Cell Reports, 2016. **6**(1): p. 85-94.
70. Kosinski, C., et al., *Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors.* Proc Natl Acad Sci U S A, 2007. **104**(39): p. 15418-23.
71. Schweisguth, F., *Regulation of notch signaling activity.* Curr Biol, 2004. **14**(3): p. R129-38.
72. Collu, G.M., et al., *Dishevelled limits Notch signalling through inhibition of CSL.* Development, 2012. **139**(23): p. 4405-15.
73. Pellegrinet, L., et al., *Dll1- and dll4-mediated notch signaling are required for homeostasis of intestinal stem cells.* Gastroenterology, 2011. **140**(4): p. 1230-1240 e1-7.
74. Huntzicker, E.G., et al., *Dual degradation signals control Gli protein stability and tumor formation.* Genes Dev, 2006. **20**(3): p. 276-81.
75. de Santa Barbara, P., G.R. van den Brink, and D.J. Roberts, *Molecular etiology of gut malformations and diseases.* Am J Med Genet, 2002. **115**(4): p. 221-30.
76. Wang, B., J.F. Fallon, and P.A. Beachy, *Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb.* Cell, 2000. **100**(4): p. 423-34.
77. van den Brink, G.R. and J.C. Hardwick, *Hedgehog Wnteraction in colorectal cancer.* Gut, 2006. **55**(7): p. 912-4.
78. Barker, N., et al., *Identification of stem cells in small intestine and colon by marker gene Lgr5.* Nature, 2007. **449**(7165): p. 1003-7.
79. Barker, N., et al., *Lgr5(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro.* Cell Stem Cell, 2010. **6**(1): p. 25-36.
80. Huch, M., et al., *Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis.* EMBO J, 2013. **32**(20): p. 2708-21.
81. Barker, N., et al., *Lgr5(+ve) stem/progenitor cells contribute to nephron formation during kidney development.* Cell Rep, 2012. **2**(3): p. 540-52.

82. Plaks, V., et al., *Lgr5-expressing cells are sufficient and necessary for postnatal mammary gland organogenesis.* Cell Rep, 2013. **3**(1): p. 70-8.
83. van de Wetering, M., et al., *The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells.* Cell, 2002. **111**(2): p. 241-50.
84. Tian, H., et al., *A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable.* Nature, 2011. **478**(7368): p. 255-9.
85. Buczacki, S.J., et al., *Intestinal label-retaining cells are secretory precursors expressing Lgr5.* Nature, 2013. **495**(7439): p. 65-9.
86. Wang, Y., C.Q. Jiang, and L.F. Fan, *Correlation of Musashi-1, Lgr5, and pEGFR expressions in human small intestinal adenocarcinomas.* Tumour Biol, 2015. **36**(8): p. 6075-82.
87. Wu, X.S., H.Q. Xi, and L. Chen, *Lgr5 is a potential marker of colorectal carcinoma stem cells that correlates with patient survival.* World J Surg Oncol, 2012. **10**: p. 244.
88. Fan, X.S., et al., *Expression of Lgr5 in human colorectal carcinogenesis and its potential correlation with beta-catenin.* Int J Colorectal Dis, 2010. **25**(5): p. 583-90.
89. He, S., et al., *Expression of Lgr5, a marker of intestinal stem cells, in colorectal cancer and its clinicopathological significance.* Biomed Pharmacother, 2014. **68**(5): p. 507-13.
90. Hsu, H.C., et al., *Overexpression of Lgr5 correlates with resistance to 5-FU-based chemotherapy in colorectal cancer.* Int J Colorectal Dis, 2013. **28**(11): p. 1535-46.
91. McClanahan, T., et al., *Identification of overexpression of orphan G protein-coupled receptor GPR49 in human colon and ovarian primary tumors.* Cancer Biol Ther, 2006. **5**(4): p. 419-26.
92. Takahashi, H., et al., *Significance of Lgr5(+ve) cancer stem cells in the colon and rectum.* Ann Surg Oncol, 2011. **18**(4): p. 1166-74.
93. Uchida, H., et al., *Overexpression of leucine-rich repeat-containing G protein-coupled receptor 5 in colorectal cancer.* Cancer Sci, 2010. **101**(7): p. 1731-7.
94. Baker, A.M., et al., *Characterization of LGR5 stem cells in colorectal adenomas and carcinomas.* Sci Rep, 2015. **5**: p. 8654.
95. Brabletz, T., et al., *Nuclear overexpression of the oncoprotein beta-catenin in colorectal cancer is localized predominantly at the invasion front.* Pathol Res Pract, 1998. **194**(10): p. 701-4.
96. Brabletz, T., et al., *Variable beta-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10356-61.

97. Suzuki, H., et al., *Nuclear beta-catenin expression at the invasive front and in the vessels predicts liver metastasis in colorectal carcinoma.* Anticancer Res, 2008. **28**(3B): p. 1821-30.
98. Morgan, R.G., E. Mortensson, and A.C. Williams, *Targeting LGR5 in Colorectal Cancer: therapeutic gold or too plastic?* Br J Cancer, 2018. **118**(11): p. 1410-1418.
99. Chen, Q., et al., *Prognostic value of LGR5 in colorectal cancer: a meta-analysis.* PLoS One, 2014. **9**(9): p. e107013.
100. Jiang, Y., et al., *Lgr5 expression is a valuable prognostic factor for colorectal cancer: evidence from a meta-analysis.* BMC Cancer, 2015. **15**: p. 948.
101. Hirsch, D., et al., *LGR5 positivity defines stem-like cells in colorectal cancer.* Carcinogenesis, 2014. **35**(4): p. 849-58.
102. Hsu, H.C., et al., *LGR5 regulates survival through mitochondria-mediated apoptosis and by targeting the Wnt/beta-catenin signaling pathway in colorectal cancer cells.* Cell Signal, 2014. **26**(11): p. 2333-42.
103. de Sousa, E.M.F., et al., *Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients.* Cell Stem Cell, 2011. **9**(5): p. 476-85.
104. Su, S., et al., *Lgr5 Methylation in Cancer Stem Cell Differentiation and Prognosis-Prediction in Colorectal Cancer.* PLoS One, 2015. **10**(11): p. e0143513.
105. Garcia, M.I., et al., *LGR5 deficiency deregulates Wnt signaling and leads to precocious Paneth cell differentiation in the fetal intestine.* Dev Biol, 2009. **331**(1): p. 58-67.
106. Kinzel, B., et al., *Functional roles of Lgr4 and Lgr5 in embryonic gut, kidney and skin development in mice.* Dev Biol, 2014. **390**(2): p. 181-90.
107. Walker, F., et al., *LGR5 is a negative regulator of tumourigenicity, antagonizes Wnt signalling and regulates cell adhesion in colorectal cancer cell lines.* PLoS One, 2011. **6**(7): p. e22733.
108. Junttila, M.R., et al., *Targeting LGR5+ cells with an antibody-drug conjugate for the treatment of colon cancer.* Sci Transl Med, 2015. **7**(314): p. 314ra186.
109. Shimokawa, M., et al., *Visualization and targeting of LGR5(+) human colon cancer stem cells.* Nature, 2017. **545**(7653): p. 187-192.
110. Glinka, A., et al., *LGR4 and LGR5 are R-spondin receptors mediating Wnt/beta-catenin and Wnt/PCP signalling.* EMBO Rep, 2011. **12**(10): p. 1055-61.
111. Moad, H.E. and A.A. Pioszak, *Reconstitution of R-spondin:LGR4:ZNRF3 adult stem cell growth factor signaling complexes with recombinant proteins produced in Escherichia coli.* Biochemistry, 2013. **52**(41): p. 7295-304.

112. Chen, P.H., et al., *The structural basis of R-spondin recognition by LGR5 and RNF43.* Genes Dev, 2013. **27**(12): p. 1345-50.
113. Carmon, K.S., et al., *R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/beta-catenin signaling.* Proc Natl Acad Sci U S A, 2011. **108**(28): p. 11452-7.
114. Cadigan, K.M. and M.L. Waterman, *TCF/LEFs and Wnt signaling in the nucleus.* Cold Spring Harb Perspect Biol, 2012. **4**(11).
115. Arce, L., N.N. Yokoyama, and M.L. Waterman, *Diversity of LEF/TCF action in development and disease.* Oncogene, 2006. **25**(57): p. 7492-504.
116. Brannon, M., et al., *A beta-catenin/XTcf-3 complex binds to the siamois promoter to regulate dorsal axis specification in Xenopus.* Genes Dev, 1997. **11**(18): p. 2359-70.
117. Bienz, M., *TCF: transcriptional activator or repressor?* Curr Opin Cell Biol, 1998. **10**(3): p. 366-72.
118. Cavallo, R.A., et al., *Drosophila Tcf and Groucho interact to repress Wingless signalling activity.* Nature, 1998. **395**(6702): p. 604-8.
119. Valenta, T., J. Lukas, and V. Korinek, *HMG box transcription factor TCF-4's interaction with CtBP1 controls the expression of the Wnt target Axin2/Conductin in human embryonic kidney cells.* Nucleic Acids Res, 2003. **31**(9): p. 2369-80.
120. Chen, G., et al., *A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in Drosophila development.* Genes Dev, 1999. **13**(17): p. 2218-30.
121. Sampson, E.M., et al., *Negative regulation of the Wnt-beta-catenin pathway by the transcriptional repressor HBP1.* EMBO J, 2001. **20**(16): p. 4500-11.
122. Bollaert, E., A. de Rocca Serra, and J.B. Demoulin, *The HMG box transcription factor HBP1: a cell cycle inhibitor at the crossroads of cancer signaling pathways.* Cell Mol Life Sci, 2019. **76**(8): p. 1529-1539.
123. Hecht, A., et al., *The p300/CBP acetyltransferases function as transcriptional coactivators of beta-catenin in vertebrates.* EMBO J, 2000. **19**(8): p. 1839-50.
124. Doumpas, N., et al., *TCF/LEF dependent and independent transcriptional regulation of Wnt/beta-catenin target genes.* EMBO J, 2019. **38**(2).
125. van Es, J.H., et al., *A critical role for the Wnt effector Tcf4 in adult intestinal homeostatic self-renewal.* Mol Cell Biol, 2012. **32**(10): p. 1918-27.
126. Korinek, V., et al., *Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4.* Nat Genet, 1998. **19**(4): p. 379-83.
127. van Noort, M. and H. Clevers, *TCF transcription factors, mediators of Wnt-signaling in development and cancer.* Dev Biol, 2002. **244**(1): p. 1-8.

128. Prokunina-Olsson, L., et al., *Alternative splicing of TCF7L2 gene in omental and subcutaneous adipose tissue and risk of type 2 diabetes.* PLoS One, 2009. **4**(9): p. e7231.
129. Van de Wetering, M., et al., *Extensive alternative splicing and dual promoter usage generate Tcf-1 protein isoforms with differential transcription control properties.* Mol Cell Biol, 1996. **16**(3): p. 745-52.
130. Kennell, J.A., et al., *T-cell factor 4N (TCF-4N), a novel isoform of mouse TCF-4, synergizes with beta-catenin to coactivate C/EBPalpha and steroidogenic factor 1 transcription factors.* Mol Cell Biol, 2003. **23**(15): p. 5366-75.
131. Weise, A., et al., *Alternative splicing of Tcf7l2 transcripts generates protein variants with differential promoter-binding and transcriptional activation properties at Wnt/beta-catenin targets.* Nucleic Acids Res, 2010. **38**(6): p. 1964-81.
132. Duval, A., et al., *The human T-cell transcription factor-4 gene: structure, extensive characterization of alternative splicings, and mutational analysis in colorectal cancer cell lines.* Cancer Res, 2000. **60**(14): p. 3872-9.
133. Atcha, F.A., et al., *A unique DNA binding domain converts T-cell factors into strong Wnt effectors.* Mol Cell Biol, 2007. **27**(23): p. 8352-63.
134. Giese, K., A. Amsterdam, and R. Grosschedl, *DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1.* Genes Dev, 1991. **5**(12B): p. 2567-78.
135. van Beest, M., et al., *Sequence-specific high mobility group box factors recognize 10-12-base pair minor groove motifs.* J Biol Chem, 2000. **275**(35): p. 27266-73.
136. Hecht, A. and M.P. Stemmler, *Identification of a promoter-specific transcriptional activation domain at the C terminus of the Wnt effector protein T-cell factor 4.* J Biol Chem, 2003. **278**(6): p. 3776-85.
137. Fang, M., et al., *C-terminal-binding protein directly activates and represses Wnt transcriptional targets in Drosophila.* EMBO J, 2006. **25**(12): p. 2735-45.
138. Hamada, F. and M. Bienz, *The APC tumor suppressor binds to C-terminal binding protein to divert nuclear beta-catenin from TCF.* Dev Cell, 2004. **7**(5): p. 677-85.
139. Clevers, H., *The intestinal crypt, a prototype stem cell compartment.* Cell, 2013. **154**(2): p. 274-84.
140. Yochum, G.S., R. Cleland, and R.H. Goodman, *A genome-wide screen for beta-catenin binding sites identifies a downstream enhancer element that controls c-Myc gene expression.* Mol Cell Biol, 2008. **28**(24): p. 7368-79.
141. Tuupanen, S., et al., *The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling.* Nat Genet, 2009. **41**(8): p. 885-90.

142. Giakountis, A., et al., *A Positive Regulatory Loop between a Wnt-Regulated Non-coding RNA and ASCL2 Controls Intestinal Stem Cell Fate.* Cell Rep, 2016. **15**(12): p. 2588-96.
143. Hrckulak, D., et al., *Wnt Effector TCF4 Is Dispensable for Wnt Signaling in Human Cancer Cells.* Genes (Basel), 2018. **9**(9).
144. Cancer Genome Atlas, N., *Comprehensive molecular characterization of human colon and rectal cancer.* Nature, 2012. **487**(7407): p. 330-7.
145. Sjoblom, T., et al., *The consensus coding sequences of human breast and colorectal cancers.* Science, 2006. **314**(5797): p. 268-74.
146. Wood, L.D., et al., *The genomic landscapes of human breast and colorectal cancers.* Science, 2007. **318**(5853): p. 1108-13.
147. Tang, W., et al., *A genome-wide RNAi screen for Wnt/beta-catenin pathway components identifies unexpected roles for TCF transcription factors in cancer.* Proc Natl Acad Sci U S A, 2008. **105**(28): p. 9697-702.
148. Wenzel, J., et al., *Loss of the nuclear Wnt pathway effector TCF7L2 promotes migration and invasion of human colorectal cancer cells.* Oncogene, 2020. **39**(19): p. 3893-3909.
149. Grant, S.F., et al., *Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes.* Nat Genet, 2006. **38**(3): p. 320-3.
150. Lyssenko, V., et al., *Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes.* J Clin Invest, 2007. **117**(8): p. 2155-63.
151. Saxena, R., et al., *Common single nucleotide polymorphisms in TCF7L2 are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals.* Diabetes, 2006. **55**(10): p. 2890-5.
152. Pradas-Juni, M., et al., *Differential transcriptional and posttranslational transcription factor 7-like regulation among nondiabetic individuals and type 2 diabetic patients.* Mol Endocrinol, 2014. **28**(9): p. 1558-70.
153. Prokunina-Olsson, L., et al., *Tissue-specific alternative splicing of TCF7L2.* Hum Mol Genet, 2009. **18**(20): p. 3795-804.
154. Oh, K.J., et al., *TCF7L2 modulates glucose homeostasis by regulating CREB- and FoxO1-dependent transcriptional pathway in the liver.* PLoS Genet, 2012. **8**(9): p. e1002986.
155. Hazra, A., et al., *Association of the TCF7L2 polymorphism with colorectal cancer and adenoma risk.* Cancer Causes Control, 2008. **19**(9): p. 975-80.
156. Kim, M.S., et al., *Frameshift mutations of Wnt pathway genes AXIN2 and TCF7L2 in gastric carcinomas with high microsatellite instability.* Hum Pathol, 2009. **40**(1): p. 58-64.

157. Duval, A., et al., *Variable mutation frequencies in coding repeats of TCF-4 and other target genes in colon, gastric and endometrial carcinoma showing microsatellite instability.* Oncogene, 1999. **18**(48): p. 6806-9.
158. Duval, A., et al., *Frequent frameshift mutations of the TCF-4 gene in colorectal cancers with microsatellite instability.* Cancer Res, 1999. **59**(17): p. 4213-5.
159. Ruckert, S., et al., *T-cell factor-4 frameshift mutations occur frequently in human microsatellite instability-high colorectal carcinomas but do not contribute to carcinogenesis.* Cancer Res, 2002. **62**(11): p. 3009-13.
160. Kisonaite, M., X. Wang, and M. Hyvonen, *Structure of Gremlin-1 and analysis of its interaction with BMP-2.* Biochem J, 2016. **473**(11): p. 1593-604.
161. Palombella, V.J., et al., *The ubiquitin-proteasome pathway is required for processing the NF-kappa B1 precursor protein and the activation of NF-kappa B.* Cell, 1994. **78**(5): p. 773-85.
162. Li, Z., Y. Park, and E.M. Marcotte, *A Bacteriophage tailspike domain promotes self-cleavage of a human membrane-bound transcription factor, the myelin regulatory factor MYRF.* PLoS Biol, 2013. **11**(8): p. e1001624.
163. Yaeger, R., et al., *Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer.* Cancer Cell, 2018. **33**(1): p. 125-136 e3.
164. Middha, S., et al., *Reliable Pan-Cancer Microsatellite Instability Assessment by Using Targeted Next-Generation Sequencing Data.* JCO Precis Oncol, 2017. **2017**.
165. Bordonaro, M., *Crosstalk between Wnt Signaling and RNA Processing in Colorectal Cancer.* J Cancer, 2013. **4**(2): p. 96-103.
166. Murugan, A.K., *mTOR: Role in cancer, metastasis and drug resistance.* Semin Cancer Biol, 2019. **59**: p. 92-111.
167. Castedo, M., K.F. Ferri, and G. Kroemer, *Mammalian target of rapamycin (mTOR): pro- and anti-apoptotic.* Cell Death Differ, 2002. **9**(2): p. 99-100.
168. Chen, D., et al., *BRAFV600E mutation and its association with clinicopathological features of colorectal cancer: a systematic review and meta-analysis.* PLoS One, 2014. **9**(3): p. e90607.
169. Yang, B., et al., *MAGE-A, mMage-b, and MAGE-C proteins form complexes with KAP1 and suppress p53-dependent apoptosis in MAGE-positive cell lines.* Cancer Res, 2007. **67**(20): p. 9954-62.
170. van Es, J.H., et al., *Wnt signalling induces maturation of Paneth cells in intestinal crypts.* Nat Cell Biol, 2005. **7**(4): p. 381-6.
171. Gong, Y., et al., *GATA4 inhibits cell differentiation and proliferation in pancreatic cancer.* PLoS One, 2018. **13**(8): p. e0202449.

172. Kawakami, K., et al., *Secreted frizzled-related protein-5 is epigenetically downregulated and functions as a tumor suppressor in kidney cancer.* Int J Cancer, 2011. **128**(3): p. 541-50.
173. Veeck, J., et al., *Epigenetic inactivation of the secreted frizzled-related protein-5 (SFRP5) gene in human breast cancer is associated with unfavorable prognosis.* Carcinogenesis, 2008. **29**(5): p. 991-8.
174. Sato, T. and H. Clevers, *Primary mouse small intestinal epithelial cell cultures.* Methods Mol Biol, 2013. **945**: p. 319-28.
175. Katagiri, T., et al., *Bone morphogenetic protein-2 converts the differentiation pathway of C2C12 myoblasts into the osteoblast lineage.* J Cell Biol, 1994. **127**(6 Pt 1): p. 1755-66.
176. Schwarz, H., et al., *Residual endotoxin contaminations in recombinant proteins are sufficient to activate human CD1c+ dendritic cells.* PLoS One, 2014. **9**(12): p. e113840.
177. Nguyen-Ngoc, K.V., et al., *3D culture assays of murine mammary branching morphogenesis and epithelial invasion.* Methods Mol Biol, 2015. **1189**: p. 135-62.
178. Fatehullah, A., P.L. Appleton, and I.S. Nathke, *Cell and tissue polarity in the intestinal tract during tumourigenesis: cells still know the right way up, but tissue organization is lost.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1629): p. 20130014.
179. Dekkers, J.F., et al., *A functional CFTR assay using primary cystic fibrosis intestinal organoids.* Nat Med, 2013. **19**(7): p. 939-45.
180. Baker, K., *Organoids Provide an Important Window on Inflammation in Cancer.* Cancers (Basel), 2018. **10**(5).
181. Hui, K.P.Y., et al., *Tropism, replication competence, and innate immune responses of influenza virus: an analysis of human airway organoids and ex-vivo bronchus cultures.* Lancet Respir Med, 2018. **6**(11): p. 846-854.
182. Gjorevski, N., et al., *Designer matrices for intestinal stem cell and organoid culture.* Nature, 2016. **539**(7630): p. 560-564.
183. Farin, H.F., J.H. Van Es, and H. Clevers, *Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells.* Gastroenterology, 2012. **143**(6): p. 1518-1529 e7.
184. Barker, N., S. Tan, and H. Clevers, *Lgr proteins in epithelial stem cell biology.* Development, 2013. **140**(12): p. 2484-94.
185. Kim, H.M., et al., *Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran.* Cell, 2007. **130**(5): p. 906-17.

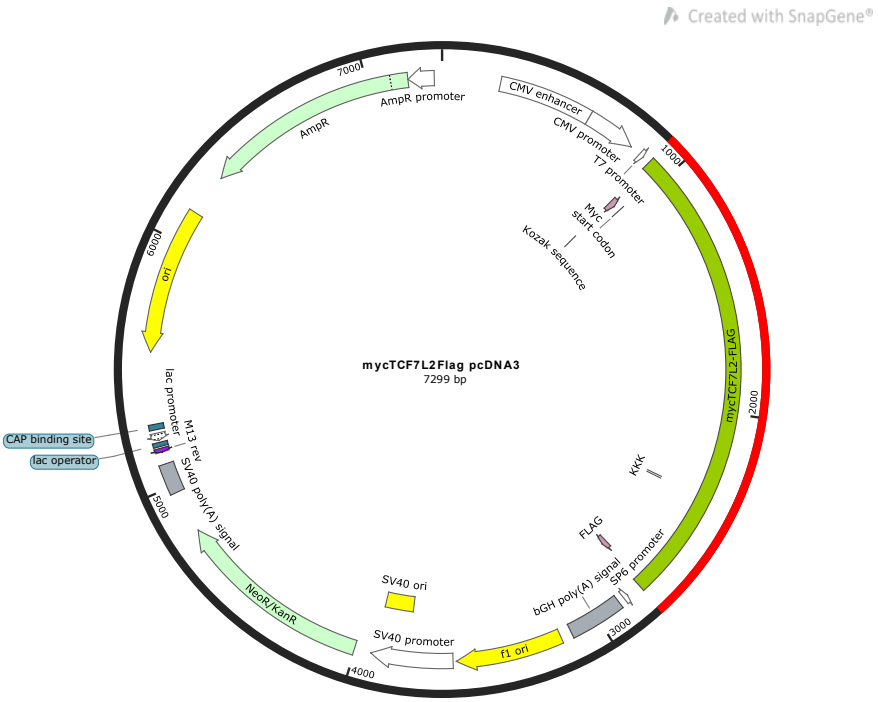
186. Xie, Y., et al., *Interaction with both ZNRF3 and LGR4 is required for the signalling activity of R-spondin.* EMBO Rep, 2013. **14**(12): p. 1120-6.
187. Vroegindeweij, E., et al., *Characterization of Lgr5-positive epithelial cells in the murine thymus.* Eur J Immunol, 2013. **43**(5): p. 1243-51.
188. Gattinoni, L., C.A. Klebanoff, and N.P. Restifo, *Paths to stemness: building the ultimate antitumour T cell.* Nat Rev Cancer, 2012. **12**(10): p. 671-84.
189. Snyder, J.C., et al., *Constitutive internalization of the leucine-rich G protein-coupled receptor-5 (LGR5) to the trans-Golgi network.* J Biol Chem, 2013. **288**(15): p. 10286-97.

9 Appendix

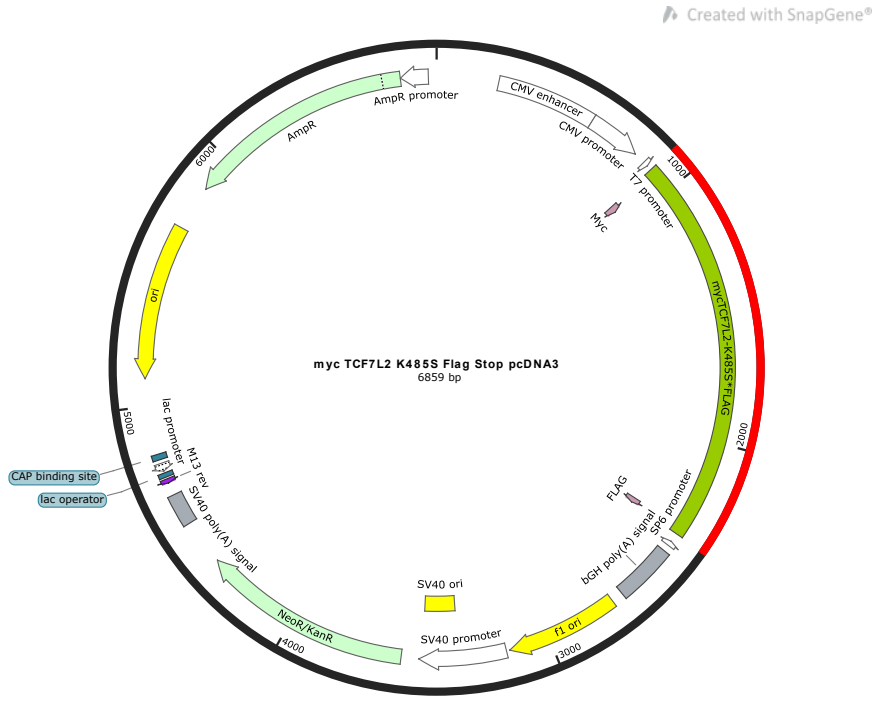
Appendix 1

Plasmids generated in this thesis

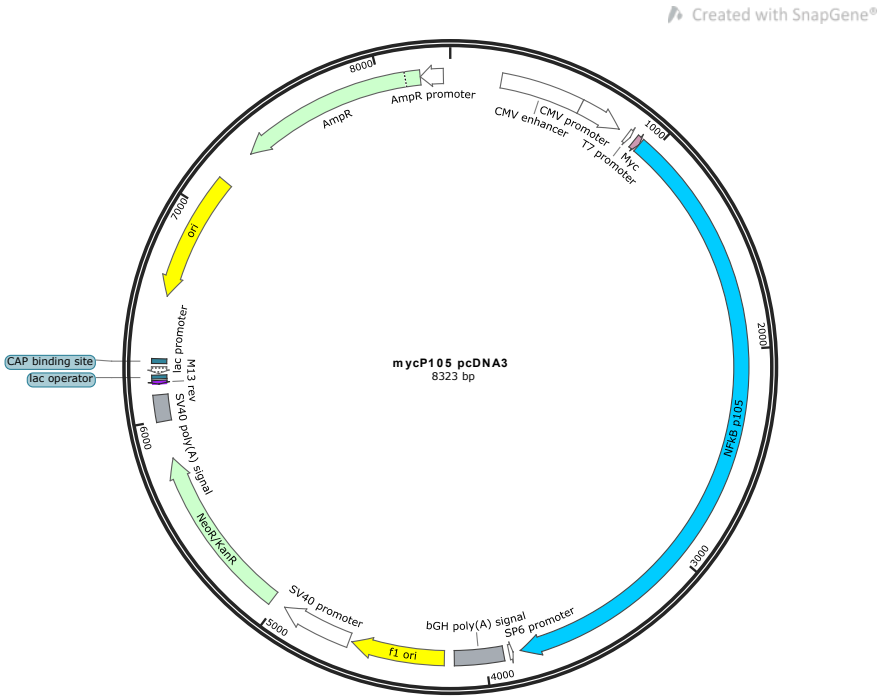
1. myc-TCF7L2-full length FLAG pcDNA3



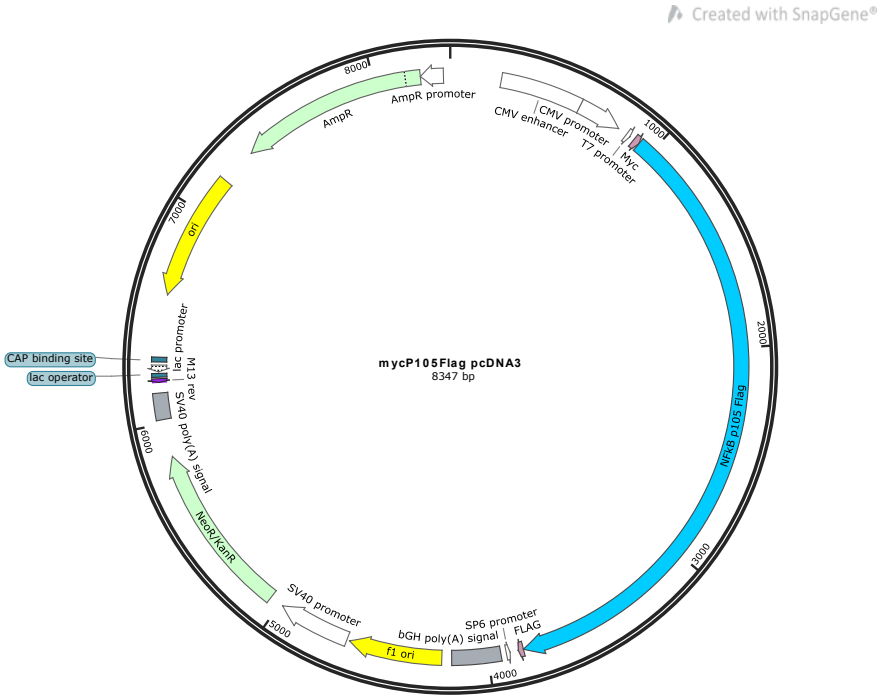
2. myc-TCF7L2-K485S Stop FLAG pcDNA3



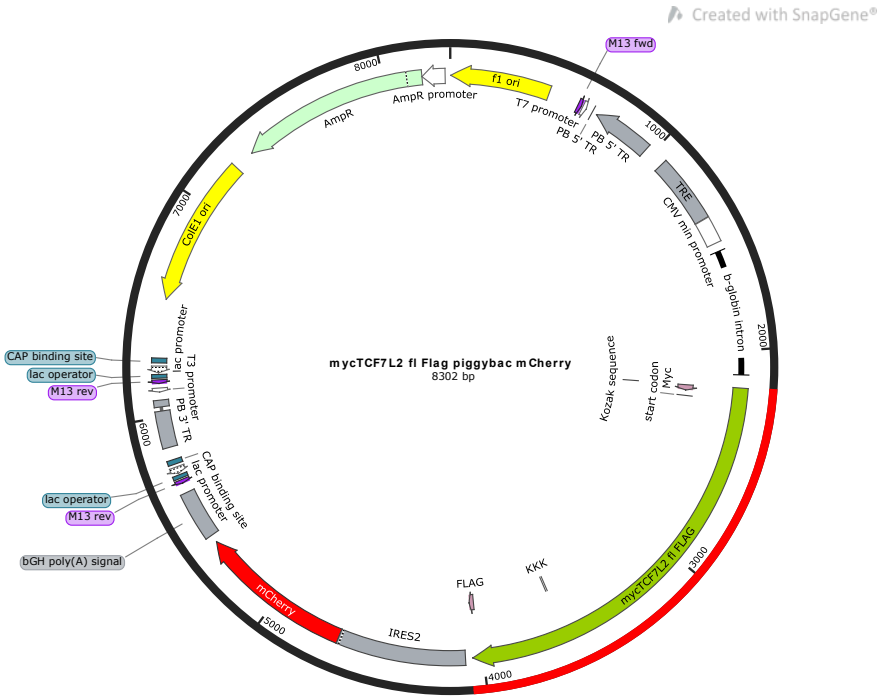
3. myc-NFκB pcDNA3



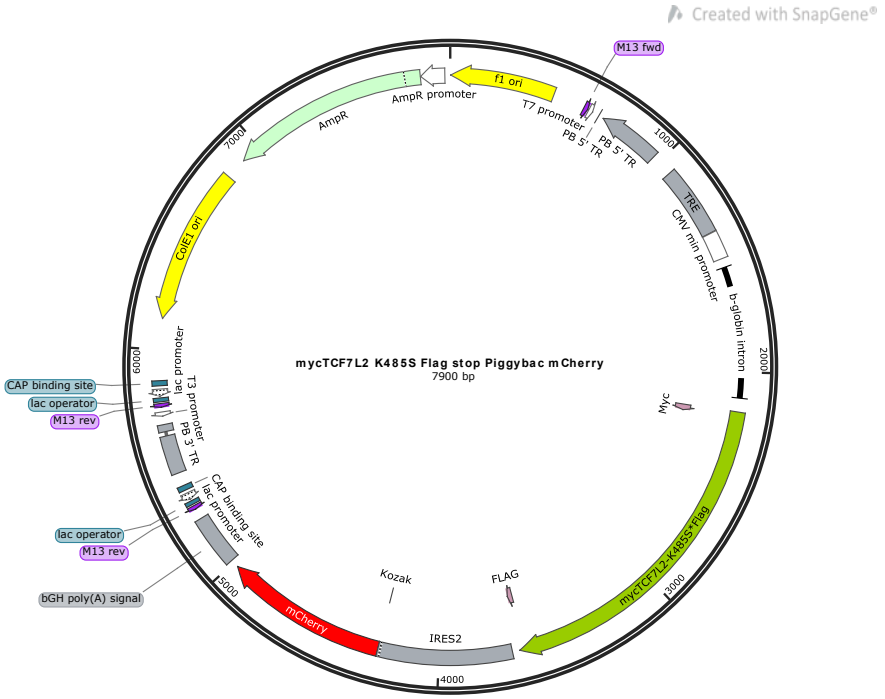
4. myc-NFκB FLAG pcDNA3



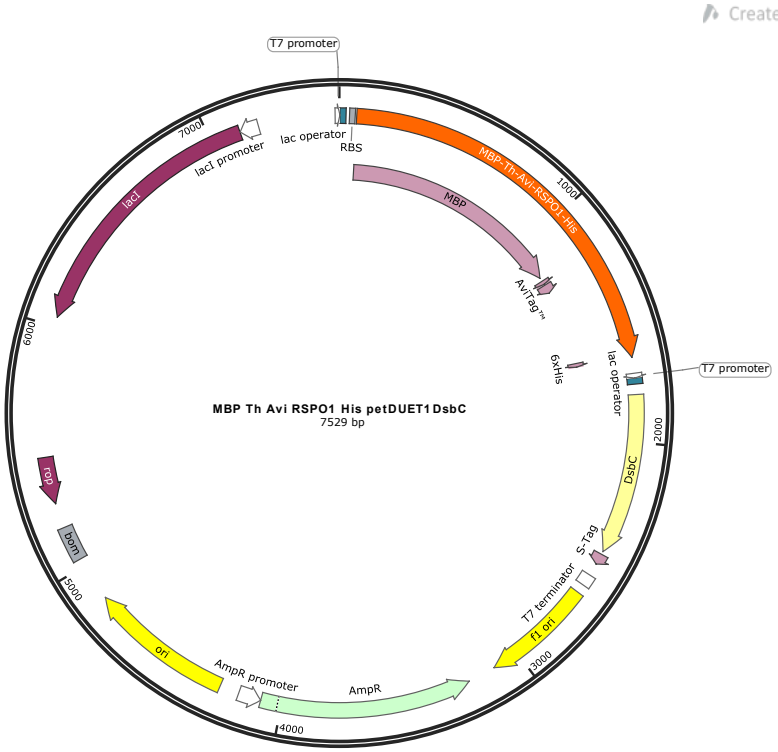
5. myc-TCF7L2-full length FLAG PiggyBac



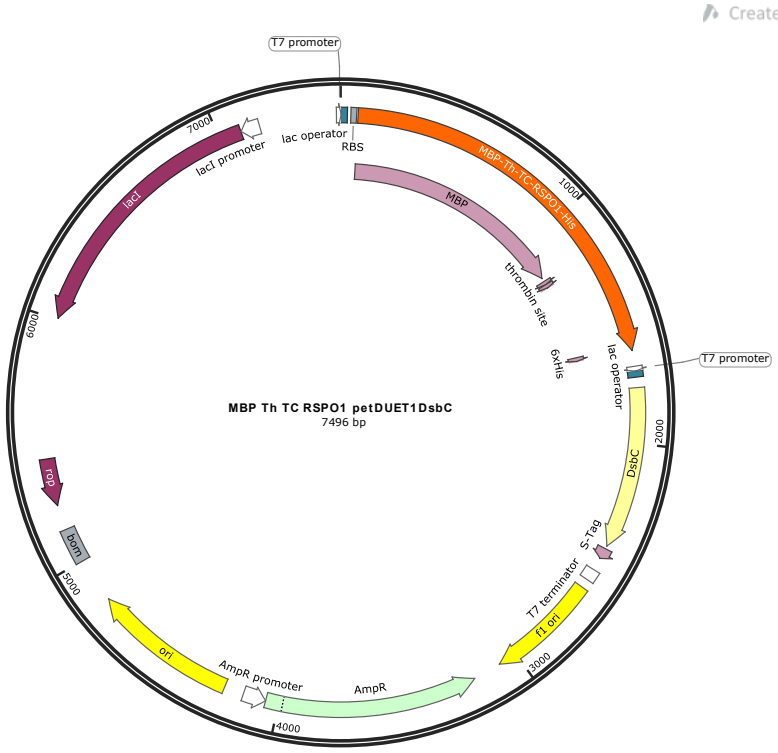
6. myc-TCF7L2-K485S Stop FLAG PiggyBac



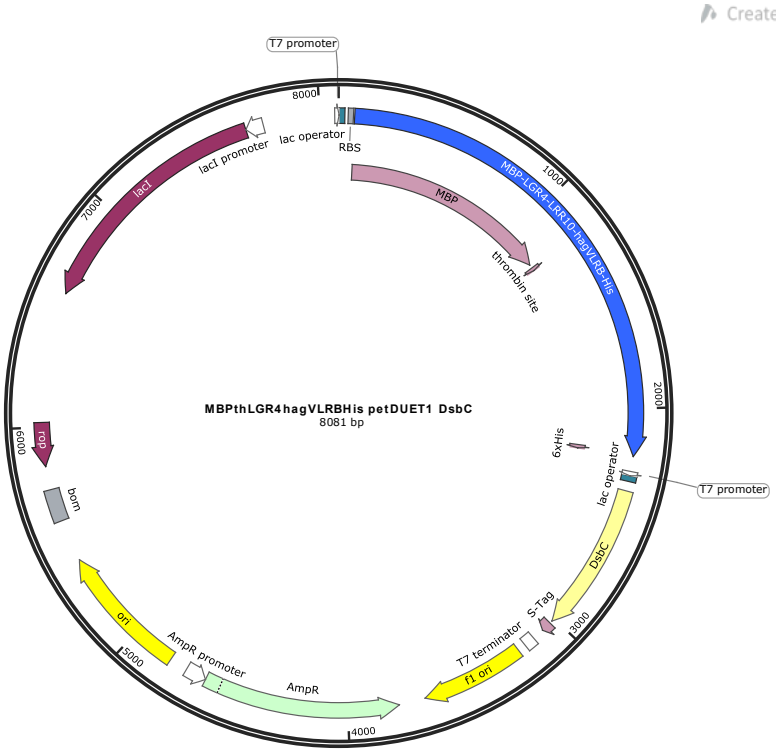
7. MBP-Avi-RSPO1 21-145-His petDUET1 DsbC



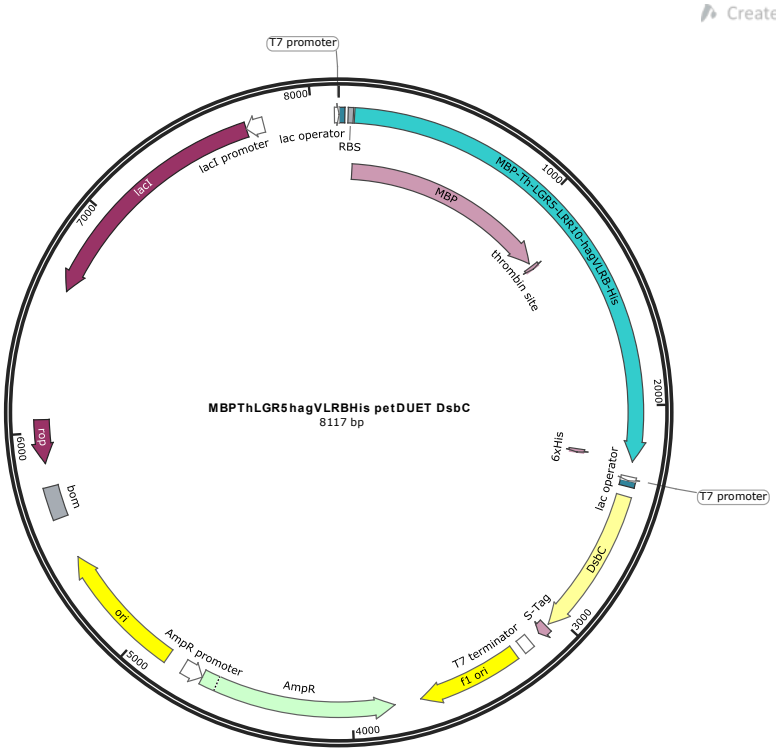
8. MBP-TC-RSPO1 21-145-His petDUET1 DsbC



9. MBP-LGR4-LRR10-hagVLRB-His petDUET1 DsbC



10. MBP-LGR5-LRR10-hagVLRB-His petDUET1 DsbC



Appendix 2

Bioinformatics

1. Mutations and clinical features

```
setwd("/Users/mu251/Dokumente/phd/TCF7L2/Bioinformatics_crc_msk_2018-2")
muts1<-read.table("data_mutations_mskcc.txt", header=TRUE,sep="\t")
muts2<-read.table("data_mutations_extended.txt", header=TRUE,sep="\t")
muts<-rbind(muts1,muts2)
crc<-read.table("crc_msk_2018_clinical_data.tsv", header=TRUE,sep="\t")
mergedMuts<-merge(muts, crc, by.x = "Tumor_Sample_Barcode", by.y = "Sample.ID")
View(mergedMuts)
TCF7L2<-muts[muts$Hugo_Symbol=="TCF7L2",]
TCF7L2bc<-as.character(TCF7L2$Tumor_Sample_Barcode)
TCF7L2bcmuts<-muts[muts$Tumor_Sample_Barcode %in% TCF7L2bc,]

TCF7L2unique<-
TCF7L2bcmuts[!duplicated(cbind(TCF7L2bcmuts$Hugo_Symbol,TCF7L2bcmuts$Tumor_S
ample_Barcode)),]

mutTable<-table(TCF7L2unique$Hugo_Symbol)
View(mutTable<-table(TCF7L2unique$Hugo_Symbol))
barplot(mutTable[order(-mutTable)][1:30], font.axis=1,cex.names =0.7, las=2)

#in mergedMuts looking to also sort it on MSI.Status
TCF7L2<-mergedMuts[mergedMuts$Hugo_Symbol=="TCF7L2",]
TCF7L2bc<-as.character(TCF7L2$Tumor_Sample_Barcode)
TCF7L2bcmuts<-mergedMuts[mergedMuts$Tumor_Sample_Barcode %in% TCF7L2bc,]

TCF7L2unique<-
TCF7L2bcmuts[!duplicated(cbind(TCF7L2bcmuts$Hugo_Symbol,TCF7L2bcmuts$Tumor_S
ample_Barcode)),]

mutTable<-table(TCF7L2unique$Hugo_Symbol)
View(mutTable<-table(TCF7L2unique$Hugo_Symbol))
barplot(mutTable[order(-mutTable)][1:30], font.axis=1,cex.names =0.7, las=2)

#applied in frameshift mutations
TCF7L2fs<-mergedMuts[mergedMuts$Hugo_Symbol=="TCF7L2" &
mergedMuts$Consequence=="frameshift_variant" & mergedMuts$MSI.Status=="MSI",]
TCF7L2fsbc<-as.character(TCF7L2fs$Tumor_Sample_Barcode)
TCF7L2fsbcmuts<-mergedMuts[mergedMuts$Tumor_Sample_Barcode %in% TCF7L2fsbc,]
TCF7L2fsunique<-
TCF7L2fsbcmuts[!duplicated(cbind(TCF7L2fsbcmuts$Hugo_Symbol,TCF7L2fsbcmuts$Tum
or_Sample_Barcode)),]

mutfsTable<-table(TCF7L2fsunique$Hugo_Symbol)
View(mutfsTable<-table(TCF7L2fsunique$Hugo_Symbol))
barplot(mutfsTable[order(-mutfsTable)][1:15], font.axis=1,cex.names =0.7, las=2)

#super filter for K485 only
TCF7L2fs<-mergedMuts[mergedMuts$Hugo_Symbol=="TCF7L2" &
mergedMuts$Consequence=="frameshift_variant" & mergedMuts$MSI.Status=="MSI" &
mergedMuts$HGVSp_Short=="p.K485Sfs*23",]
```

```
TCF7L2fsbc<-as.character(TCF7L2fs$Tumor_Sample_Barcode)
TCF7L2fsbcmuts<-mergedMuts[mergedMuts$Tumor_Sample_Barcode %in% TCF7L2fsbc,]
TCF7L2fsunique<-
TCF7L2fsbcmuts[!duplicated(cbind(TCF7L2fsbcmuts$Hugo_Symbol,TCF7L2fsbcmuts$Tumor_Sample_Barcode)),]

mutfsTable<-table(TCF7L2fsunique$Hugo_Symbol)
View(mutfsTable<-table(TCF7L2fsunique$Hugo_Symbol))
barplot(mutfsTable[order(-mutfsTable)][1:15], font.axis=1,cex.names =0.7, las=2)

#next step is to see what mutations there are in nonTCF7L2 mutated cancers (sorting for markers for different sub types)
#HOW are BRAF mutations characterised
BRAF<-mergedMuts[mergedMuts$Hugo_Symbol=="BRAF",]
BRAFbc<-as.character(BRAF$Tumor_Sample_Barcode)
BRAFbcmuts<-mergedMuts[mergedMuts$Tumor_Sample_Barcode %in% BRAFbc,]

BRAF<-mergedMuts[mergedMuts$Hugo_Symbol=="BRAF" &
mergedMuts$MSI.Status=="MSI" & mergedMuts$HGVSp_Short=="p.V600E",]
BRAFbc<-as.character(BRAF$Tumor_Sample_Barcode)
BRAFbcmuts<-mergedMuts[mergedMuts$Tumor_Sample_Barcode %in% BRAFbc,]

BRAFunique<-
BRAFbcmuts[!duplicated(cbind(BRAFbcmuts$Hugo_Symbol,BRAFbcmuts$Tumor_Sample_Barcode)),]

mutTable<-table(BRAFunique$Hugo_Symbol)
View(mutTable<-table(BRAFunique$Hugo_Symbol))
barplot(mutTable[order(-mutTable)][1:30], font.axis=1,cex.names =0.7, las=2)


#only BRAF V600E MSI mutations
BRAFV600E<-mergedMuts[mergedMuts$Hugo_Symbol=="BRAF" &
mergedMuts$HGVSp_Short=="p.V600E" & mergedMuts$MSI.Status=="MSI",]
BRAFV600Ebc<-as.character(BRAFV600E$Tumor_Sample_Barcode)
BRAFV600Ebcmuts<-mergedMuts[muts$Tumor_Sample_Barcode %in% BRAFV600Ebc,]
BRAFV600Eunique<-
BRAFV600Ebcmuts[!duplicated(cbind(BRAFV600Ebcmuts$Hugo_Symbol,BRAFV600Ebcmuts$Tumor_Sample_Barcode)),]

mutfsTable<-table(BRAFV600Eunique$Hugo_Symbol)
View(mutfsTable<-table(BRAFV600Eunique$Hugo_Symbol))
barplot(mutfsTable[order(-mutfsTable)][1:10], font.axis=1,cex.names =0.7, las=2)
```

2. Gene expression analysis and pathway enrichment

````
title: TC7FL2 mutational truncation vs TCGA-COAD with correction for BRAF mutations
  as confounder,
author: "Andrew Holding"
date: "03/07/2019"
output: pdf_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```


#Download Data

```{r, eval=FALSE}
if (!requireNamespace("BiocManager", quietly = TRUE))
    install.packages("BiocManager")


if(!requireNamespace("TCGAbiolinks",quietly=TRUE)) {
        BiocManager::install("TCGAbiolinks")
        library("TCGAbiolinks")
    } else {
        library("TCGAbiolinks")
    }



```


##Start Download
```{r, eval=FALSE}
query <- GDCquery(project = "TCGA-COAD",
                  data.category = "Transcriptome Profiling",
                  data.type = "Gene Expression Quantification",
                  experimental.strategy = "RNA-Seq",
                  workflow.type = "HTSeq - Counts"
                  )
GDCdownload(query, method = "api", files.per.chunk = 10)
data <- GDCprepare(query)
```


##Extract expression matrix
```{r eval=FALSE}
install.packages("mime")


if (!requireNamespace("DT", quietly = TRUE))
    install.packages("DT")

library(DT)
````
````

```
library(SummarizedExperiment)
datatable(as.data.frame(colData(data)),
        options = list(scrollX = TRUE, keys = TRUE, pageLength = 5),
        rownames = FALSE)


COADMatrix <- assay(data,"HTSeq - Counts") #was raw_count
```

##Save/Load Data to disk
```{r}

#saveRDS(COADMatrix, file="COADCountMatrx.RDS")
COADMatrix<-readRDS( file="COADCountMatrx.RDS")
```

#Differential Expression
##Prepare sample sheets

This time we use a more complex sample sheet using mutational status from cBioPortal

```{r}


#https://www.cbioportal.org/results/oncoprint?Action=Submit&RPPA_SCORE_THRESHOLD=2.0&Z_SCORE_THRESHOLD=2.0&cancer_study_list=coadread_tcga_pan_can_atlas_2018&case_set_id=coadread_tcga_pan_can_atlas_2018_sequenced&data_priority=0&gene_list=BRAF%2520TCF7L2&geneset_list=%20&genetic_profile_ids_PROFILE_MUTATION_EXTENDED=coadread_tcga_pan_can_atlas_2018_mutations&tab_index=tab_visualize
cBioPortal<-read.csv("PATIENT_DATA_oncoprint.tsv",na.strings = "",stringsAsFactors=FALSE,sep="\t")
colnames(cBioPortal)<-gsub("\\.","-",colnames(cBioPortal) )

matchedData<-substr(colnames(COADMatrix),1,12)[substr(colnames(COADMatrix),1,12) %in% colnames(cBioPortal)]

cBioPortal[is.na(cBioPortal)]<-"None"

coldata<-t(cBioPortal[c(3,4),matchedData])
rownames(coldata)<-substr(rownames(coldata),1,12)
colnames(coldata)<-c("BRAF","TCL7F2")
matchedcolData<-coldata

samplesNameLookup<-colnames(COADMatrix)
names(samplesNameLookup)<-substr(colnames(COADMatrix),1,12)

matchedMatrix<-COADMatrix[,samplesNameLookup[matchedData]]

rownames(matchedcolData)<-as.character(samplesNameLookup[rownames(coldata)])

```


##Run DeSeq2
```{r, eval=FALSE}
```

```
if (!requireNamespace("BiocManager", quietly = TRUE))
   install.packages("BiocManager")

BiocManager::install("DESeq2")
library("BiocParallel")
register(MulticoreParam(3)) #use 3 cores


all.equal(colnames(matchedMatrix),rownames(matchedcolData))
colnames(matchedMatrix)<-NULL
rownames(matchedcolData)<-NULL

library("DESeq2")
dds <- DESeqDataSetFromMatrix(countData = matchedMatrix,
                              colData = coldata,
                              design = ~ BRAF + TCL7F2)
dds

keep <- rowSums(counts(dds) >= 10) >= 10 #(remove lowly expressed genes)
dds <- dds[keep,]


dds <- DESeq(dds, betaPrior=TRUE, parallel = TRUE)
#https://support.bioconductor.org/p/105087/
```

```
#Save to Disk/Load from Disk
```{r}
#saveRDS(dds,file="002_dds.RDS")
dds<-readRDS(file="002_dds.RDS")
```

```
#Process results
```{r, eval=FALSE}
rN<-resultsNames(dds)
res <- results(dds, contrast = list( rN[10],c(rN[6:9]) ), listValues=c(1,-1/4) )
#contrast of TCF7L Truncating mutation vs all other TCF7L status + BRAF status as
confounder
```
#Save to Disk/Load from Disk
```{r}
#saveRDS(res,file="002_res.RDS")
saveRDS(res,file="002_res.RDS")
res<-readRDS(file="002_res.RDS")
```


##MA-Plot (to check what up/down mean on this plot!!)
```{r}
library(DESeq2)
plotMA(res,ylim=c(-5,5))
```

```
if (!requireNamespace("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install("XVector")
BiocManager::install("EnhancedVolcano")
library("EnhancedVolcano")
EnhancedVolcano(res,
    lab = rownames(res),
    x = 'log2FoldChange',
    y = 'pvalue',
    xlim = c(-5, 5),
    ylim = c(0,11),
    FCcutoff = 1.5,
    pCutoff = 10e-4,
    title = 'Truncated TCF7L2 vs nonTCF7L2',
    cutoffLineType = 'twodash',
    cutoffLineWidth = 0.8,
    pointSize = 1.0,
    labSize = 0,
    colAlpha = 1,
    legendLabels=c('Not sig.','Log2FC','p-value',
      'p-value & Log2FC'),
    legendPosition = 'right',
    legendLabSize = 16,
    legendIconSize = 5.0,
    gridlines.major = FALSE,
    gridlines.minor = FALSE)


```

```
##DE gene list
```{r}
GOI<-res$padj<0.05
GOI[is.na(GOI)]<-FALSE
if (!requireNamespace("BiocManager", quietly = TRUE))
    install.packages("BiocManager")

BiocManager::install("org.Hs.eg.db")

eg2sym<-function (x)
{
    library(org.Hs.eg.db)
    if (!exists("eg2symmap")) {
        eg2symmap <<- as.list(org.Hs.egSYMBOL[mappedkeys(org.Hs.egSYMBOL)])
    }
    x <- as.character(x)
    out <- eg2symmap[x]
    names(out) <- x
    out2 <- sapply(out, function(x) {
        if (is.null(x)) {
            return(NA)
        }
        else {
            return(x[1])
        }
```
```

```
    })
    out3 <- unlist(out2)
    out <- setNames(out3, names(out))
    return(out)
}
resultsList<-cbind(eg2sym(eg2sym(rownames(res[GOI,]))),as.data.frame(res[GOI,4:5]))
colnames(resultsList)<-c("","Stat","adj p-vlaue" )

resultsList
```
```{r}
#entrez conversion

ens2eg<-function (x)
{
    library(org.Hs.eg.db)
    if (!exists("ens2egmap")) {
        ens2egmap <<- as.list(org.Hs.egENSEMBL2EG)
    }
    out <- ens2egmap[x]
    names(out) <- x
    out2 <- sapply(out, function(x) {
        if (is.null(x)) {
            return(NA)
        }
        else {
            return(x[1])
        }
    })
    out3 <- unlist(out2)
    out <- setNames(out3, names(out))
    return(out)
}
#for example:
ens2eg("ENSG00000179348")
#now for all genes that are changed
a<-read.table("all genes ensemble.csv",header = FALSE)
head(a)
#put them into vector
b<-as.vector(unlist(a))

head(b)

#translate into entrezID
ens2eg(b)
```
```{r}
##pathway enrichment analysis


if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install(version = "3.12")
BiocManager::install("zlibbioc")
BiocManager::install("XVector")
```

```
BiocManager::install("reactome.db")
BiocManager::install("ReactomePA")
BiocManager::install("enrichPathway")
library(ReactomePA)
gene <- c("11171", "8243", "112464", "2194",
          "9318", "79026", "1654", "65003",
          "6240", "3476", "6238", "3836",
          "4176", "1017", "249","4312","8318","10874","55143","55388","991")
yy = enrichPathway(gene, pvalueCutoff=0.05)
head(as.data.frame(yy))
barplot(yy, showCategory = 8)
emapplot(yy)
cnetplot(yy, categorySize="pvalue", foldChange=gene)
dotplot(yy, showCategory=15)

##pathway enrichment analysis


if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install("ReactomePA")
BiocManager::install("enrichPathway")
library(ReactomePA)
gene <- ens2eg(b)
yy = enrichPathway(gene, pvalueCutoff=0.05)
head(as.data.frame(yy))
barplot(yy, showCategory = 15)
emapplot(yy)
cnetplot(yy, categorySize="pvalue", foldChange=gene)
dotplot(yy, showCategory=15)
```
```{r}

```

3. survival probability analysis

```
setwd("/Users/mu251/Dokumente/phd/TCF7L2/Bioinformatics_crc_msk_2018-2/survival")

install.packages("graphlayouts")
#Downloaded from Metastatic Colorectal Cancer (MSKCC, Cancer Cell 2018) on cBioPortal
survial<-read.csv("PATIENT_survival.tsv",sep="\t")
mutationList<-read.csv("mutation_list_table.tsv",sep="\t", stringsAsFactors = FALSE)
BRAFmutationList<-read.csv("BRAF_mutation_list_tabletable.tsv",sep="\t", stringsAsFactors
= FALSE)

rownames(survial)<-paste(survial[,1],"_",survial[,2])
survivalKM<-t(survial[,-1:-2])
library('survival')

colnames(survivalKM)[3]<-'time'
colnames(survivalKM)[4]<-'status'
colnames(survivalKM)[6]<-'mutation'

survivalKMFiltered<-survivalKM[!survivalKM[,3]=="",]
survivalKMFiltered[,3]<-as.numeric(survivalKMFiltered[,3])
survivalKMFiltered[survivalKMFiltered[,4]=='LIVING',4]<-'0'
survivalKMFiltered[survivalKMFiltered[,4]=='DECEASED',4]<-'1'
survivalKMFiltered[survivalKMFiltered[,6]=='',6]<-'0' #Set all without any mutation as 0
survivalKMFiltered[!survivalKMFiltered[,6]=='0',6]<-'1' #Any mutation set to 1

survivalKMFilteredDF<-data.frame(survivalKMFiltered,stringsAsFactors = FALSE)
survivalKMFilteredDF[,c(3,4,6)]<-sapply(survivalKMFilteredDF[,c(3,4,6)],as.numeric)

km1 <- survfit(Surv( time, status) ~ mutation  , data = survivalKMFilteredDF)
plot(km1,main="Mutated TCF7L2 vs Normal TCF7L2")


#Lets be more specific

mutationList[,c(2,4)]

processedMutationList<-mutationList[,c(2,4)]
processedMutationList[processedMutationList[,2]=="K485Sfs*23",2]<-1
processedMutationList[processedMutationList[,2]=="C486Vfs*8",2]<-1
processedMutationList[processedMutationList[,2]=="K485Vfs*8",2]<-1
processedMutationList[!processedMutationList[,2]=="1",2]<-0

TCL7F2short<-processedMutationList[processedMutationList[,2]==0,1]

shortVsLong<-survivalKMFilteredDF
shortVsLong[,6]<-0

#shortVsLong[rownames(survivalKMFilteredDF) %in% gsub("-", ".",
substr(TCL7F2short,1,9)),6]<-1
#shortVsLong[!rownames(survivalKMFilteredDF) %in% gsub("-", ".",
substr(TCL7F2short,1,9)),6]

TCL7F2shortRows<- gsub("-", ".", substr(TCL7F2short,1,9))[gsub("-", ".",
substr(TCL7F2short,1,9)) %in% rownames(survivalKMFilteredDF)]
```

```
shortVsLong[TCL7F2shortRows,6]<-1

km2 <- survfit(Surv( time, status) ~ mutation  , data = shortVsLong)
plot(km2, main="shortTCF7L2 vs not-TCF7L2")

#install.packages("ggfortify")
library(ggfortify)
library(grid)

autoplot(km2) +theme_classic()

library(survminer)
surv_pvalue(km2)

#Filter on +/-BRAF

BRAFmutationList[,c(2,4)]
processedBRAFMutationList<-BRAFmutationList[,c(2,4)]
processedBRAFMutationList[processedBRAFMutationList[,2]=="V600E",2]<-1
processedBRAFMutationList[!processedBRAFMutationList[,2]=="1",2]<-0
V600E<-processedBRAFMutationList[processedBRAFMutationList[,2]==0,1]

V600EvsNot<-survivalKMFilteredDF
V600EvsNot[,6]<-0

V600ERows<- gsub("-", ".", substr(V600E,1,9))[gsub("-", ".", substr(V600E,1,9)) %in%
rownames(survivalKMFilteredDF)]
V600EvsNot[V600ERows,6]<-1

km3 <- survfit(Surv( time, status) ~ mutation  , data = V600EvsNot)
plot(km3,main="BRAF V300E vs not-V300E")
autoplot(km3) +theme_classic()
surv_pvalue(km3 )

#Filter on +/-BRAF and TCF7L2
shortVsLong[,6]
V600EvsNot[,6]

BRAFvsTCF<-survivalKMFilteredDF
BRAFvsTCF[,6]<-"Neither"
BRAFvsTCF[TCL7F2shortRows,6]<-"ShortTCL7F2"
BRAFvsTCF[V600ERows,6]<- "V300E"
BRAFvsTCF[intersect(TCL7F2shortRows,V600ERows),6]<-"Both"
km4 <- survfit(Surv( time, status) ~ mutation  , data = BRAFvsTCF)
plot(km4,main="BRAF vs TCF")
autoplot(km4) +theme_classic()
#surv_pvalue(km4 )

#Filter on +/-BRAF and TCF7L2
shortVsLong[,6]
V600EvsNot[,6]

BRAFvsTCF<-survivalKMFilteredDF
BRAFvsTCF[,6]<-"Neither"
BRAFvsTCF[TCL7F2shortRows,6]<-"ShortTCL7F2"
```

```
BRAFvsTCF[V600ERows,6]<- "BRAFV600E"
km4 <- survfit(Surv( time, status) ~ mutation  , data = BRAFvsTCF)
plot(km4,main="BRAF vs TCF")
autoplot(km4)
ggsurvplot(km4, size=2, censor.size=8,
          censor.col="black",
          legend="none",
          xlab="time (days)",
          legend.labs=c("BRAFV600E","other","truncated TCF7L2"),
          legend.title="",
          font.main=c(20,"bold"),
          font.x=c(26,"bold"),
          font.y=c(26,"bold"),
          font.tickslab = c(26,"bold"),
          font.legend=c(26,"bold"))
```