Aus dem Pathologischen Institut der

Ludwig-Maximilians-Universität München

Direktor: Prof. Dr. med. Thomas Kirchner

in der DKTK Arbeitsgruppe „Oncogenic Signaling Pathways in Colorectal and Pancreatic Cancer“

Leiter: Dr.rer.nat. Peter Jung

Dissertation

zum Erwerb des Doctor of Philosophy (Ph.D.)

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität München

Wnt-driven O-glycosylation by LARGE2 in human colon and colorectal cancer

---

vorgelegt von

Vanessa Dietinger

aus Ansbach, Deutschland

am

18.06.2020

Gedruckt mit der Genehmigung der Medizinischen Fakultät der Ludwig-Maximilans-Universität München

First supervisor:	Prof. Dr. rer. nat. Heiko Hermeking
Second supervisor:	Dr. rer. nat. Peter Jung

Dean: **Prof. Dr. med. Dent. Reinhard Hickel**

Day of oral defense:

24.11.2020

“I was taught that the way of progress was neither swift nor easy”

Marie Curie

To everyone who supported me

Affidavit

Pathologisches Institut der LMU
Dietinger, Vanessa
Thalkirchner Str. 36
80337 München
Deutschland

I hereby declare, that the submitted thesis entitled

„Wnt-driven O-glycosylation by LARGE2 in human colon and colorectal cancer “

is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Munich, 30.11.2020

Vanessa Dietinger

Confirmation of congruency between printed and electronic version of the doctoral thesis

Pathologisches Institut der LMU
Dietinger, Vanessa
Thalkirchner Str. 36
80337 München
Deutschland

I hereby declare that the electronic version of the submitted thesis, entitled

“Wnt-driven O-glycosylation by LARGE2 in human colon and colorectal cancer”

Is congruent with the printed version both in content and format.

Munich, 30.11.2020

Vanessa Dietinger

Publications

Parts of this thesis have been published in the following article:

Dietinger, V.*, García de Durango, C.R.*, Wiechmann, S., Boos, S., Michl, M., Neumann, J., Hermeking, H., Küster, B., Jung, P.: Wnt-driven LARGE2 mediates laminin-adhesive O-glycosylation in human colonic epithelial cells and colorectal cancer. *Cell communication and signaling*, 2020.

* Equal contribution

Table of contents

Abbreviations

aa	*Amino acids*
ADF	Advanced DMEM/F12
ADO	Adenoma organoid lines
APC	Adenomatous polyposis coli
APS	Ammonium persulfate
Asp	Aspartic acid
BSA	Bovine serum albumin
ChIP	Chromatin immunoprecipitation
COAD	Colon adenocarcinoma
CRC	Colorectal cancer
DAPI	4‘6-diamidino-2-phenylindole
DC	Destruction complex
DEPC	Diethyl pyrocarbonate
DMEM	Dulbecco’s modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
E. *coli*	*Escherichia coli*
ECL	Enhanced chemiluminescence
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EMT	Epithelial-mesenchymal transition
EN	EGF & noggin containing medium
ER	Endoplasmic reticulum
ERT	ERT2
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
FITC	Fluorescin
gRNA	Guide(RNA)
GI	Gastrointestinal
GlcA	Glucuronic acid
gRNA	guideRNA
GSEA	Gene set enrichment analysis
gw	Gateway
HRP	Horseradish peroxidase
IHC	Immunohistochemistry
ISC	Intestinal stem cell
ISH	*In situ* hybridization
kDa	Kilo Dalton
KIRC	Kidney Renal Clear Cell Carcinoma

LBB	Laminin binding buffer
LG-Domain	Laminin-Globular Domain
LY	TGFβ-Inhibitor LY2157299
MCR	Mutation cluster region
mPDTO	Metastatic patient derived tumor organoid
mRNA	Messenger RNA
NAc	N-Acetlycysteine
NE	nTCF-ERT2
NGS	Next generation sequencing
NM	Normal mucosa
Nog	Noggin
O-GlcNAc	N-acetylglucosamine
O-GalNAc	N-acetylgalactosamine
ORF	Open reading frame
PAGE	Polyacryamide gelelectrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDO	Patient derived organoid
PDTO	Patient derived tumor organoid
PFA	Paraformaldehyde
PGE2	Prostaglandine E2
PRAD	Prostate adenocarcinoma
qPCR	Quantitative real-time PCR
READ	Rectal adenocarcinoma
RNA	Ribonucleic acid
RPMI	RPMI-1640 culture medium
R-spo	R-Spondin
SB	p38 Inhibitor SB202190
SD	Standard Deviation
SDS	Sodium dodecyl sulfate
SEM	Standard error of the mean
Ser	Serine
shRNA	Short hairpin RNA
TAE	Tris - Acetate – EDTA Buffer
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween20
TCGA	The cancer genome atlas
TEA	Triethanolamine
TEMED	Tetramethylethylenediamine
TF	Transcription factor
TGFβ	Transforming growth factor beta
Thr	Threonine

TNFα	Tumor necrosis factor alpha
WB	Western blotting
WREN	Wnt, R-spondin, EGF, Noggin containing medium
Xyl	Xylose
Y	Rho kinase inhibitor (Y27632)
α-DG	Alpha-Dystroglycan
β-DG	Beta-Dystroglycan
β-ME	Beta-Mercaptoethanol

1. Introduction

1.1. The human intestine

The mammalian intestinal lumen is covered with a single layer of epithelial cells, which perform the primary functions of the intestinal tract: digestion and absorption of nutrients. Potten and colleagues were the first to illustrate a concept for intestinal crypts and epithelium in mice, which is applicable to the human intestinal tract (Potten and Loeffler, 1990). Anatomically, the intestinal tract is divided in small intestine and colon. Furthermore, the small intestine can be subdivided in the duodenum, jejunum and ileum (**Figure 1A**) (HEATH, 1996). The small intestinal

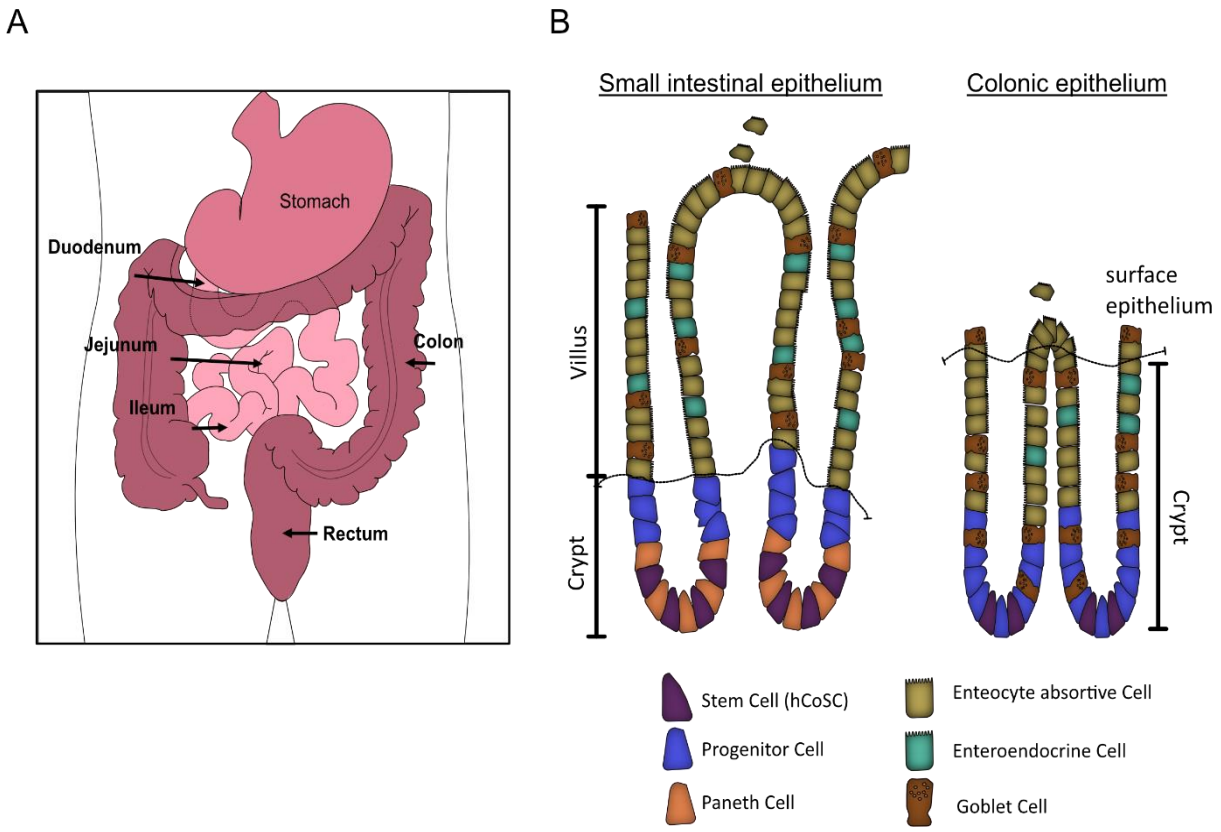


Figure 1: Structure of the human intestine. A) Division of the human gastrointestinal tract. The intestine is subdivided in Duodenum, Jejunum and Ileum (forming the small intestine, light red) and the colon (large intestine, dark red). **B**) Morphology of small intestinal epithelium and colonic epithelium. The small intestine is organized by crypts and villi. In the colon, the crypt bottom contains proliferative stem cells and progenitor cells. The colonic crypts are longer compared to the small intestinal crypts, and the colon is lined by a columnar surface epithelium instead of villi. In the small intestine, the crypt bottom contains proliferative stem cells and Paneth cells, flanked by progenitor cells, while the villi increase the luminal surface and consist of differentiated cells (Figure A according to the American cancer society, Figure B adapted from Pinto and Clevers (2005)).

epithelium is organized into large numbers of crypt-villus units, which renew themselves every 5 to 7 days and highly increase the absorptive surface. The longest villi are found in the duodenum and their length decreases along the intestinal tract (Simons and Clevers, 2011). The colon lacks villi completely but has a surface epithelium consisting of separated units of the so-called crypts of Lieberkühn (named after their discoverer Jonathan Lieberkühn) (Umar, 2010). These crypts harbor a population of proliferating epithelial stem cells at their bottom, which assure consistent self-renewal and replenishing of transit-amplifying progenitor (TA) cells. TA cells divide approximately 4 to 5 times before they differentiate into a terminal intestinal epithelial cell type (Potten and Loeffler, 1990). Several types of differentiated cells are present in the colonic crypt: mucous secreting goblet cells, hormone producing enteroendocrine cells and absorptive enterocytes, all moving upwards the crypt as the TA cells divide. As they reach the top, they undergo apoptosis and are shed into the lumen. Exclusively the small intestinal crypts contain differentiated Paneth cells. Responsible for antibacterial defense, they settle at the crypt bottom, where they reside for roughly 20 days instead of migrating in the upward direction (**Figure 1B**) (Roberts, 2000; van der Flier and Clevers, 2009; Darwich *et al.*, 2014).

The main driver for self-renewal of the intestinal epithelium is the Wnt-signaling pathway, which will be explained in detail in the following chapter. Studies found that intestinal stem cells, which reside at the crypt bottom, undergo cell cycle arrest and differentiation upon inhibition of Wnt activity, suggesting that target genes of this pathway are mainly expressed in non-differentiated stem/progenitor cells and are down-regulated in differentiated cells, thereby creating a Wnt-gradient along the vertical axis of crypt-villus units (Van de Wetering *et al.*, 2002). Altogether, the normal intestinal epithelium is tightly balanced between cell proliferation, differentiation, migration and apoptosis to ensure homeostasis.

1.2. Canonical Wnt signaling pathway

Originally discovered in Drosophila melanogaster, the wingless (Wnt) gene was found to affect developmental processes and it is highly conserved among species (Sharma, 1973; Rijsewijk *et al.*, 1987). Wnt signaling is one of the main drivers for

proper tissue development in embryos, stem cell proliferation and adult tissue maintenance. Soon after its discovery it was shown that aberrant activation of Wnt is implicated in mouse mammary tumorigenesis (Rijsewijk *et al.*, 1987) and augmented Wnt activity found in colon tumors is caused by mutations in key regulators of the pathway (Polakis, 1995).

The Wnt pathway is divided in β-Catenin dependent (canonical) and independent (non-canonical) signaling, defined by the composition of Wnt and Frizzled (Fzd) complexes, although the canonical pathway is preponderant for cell proliferation in

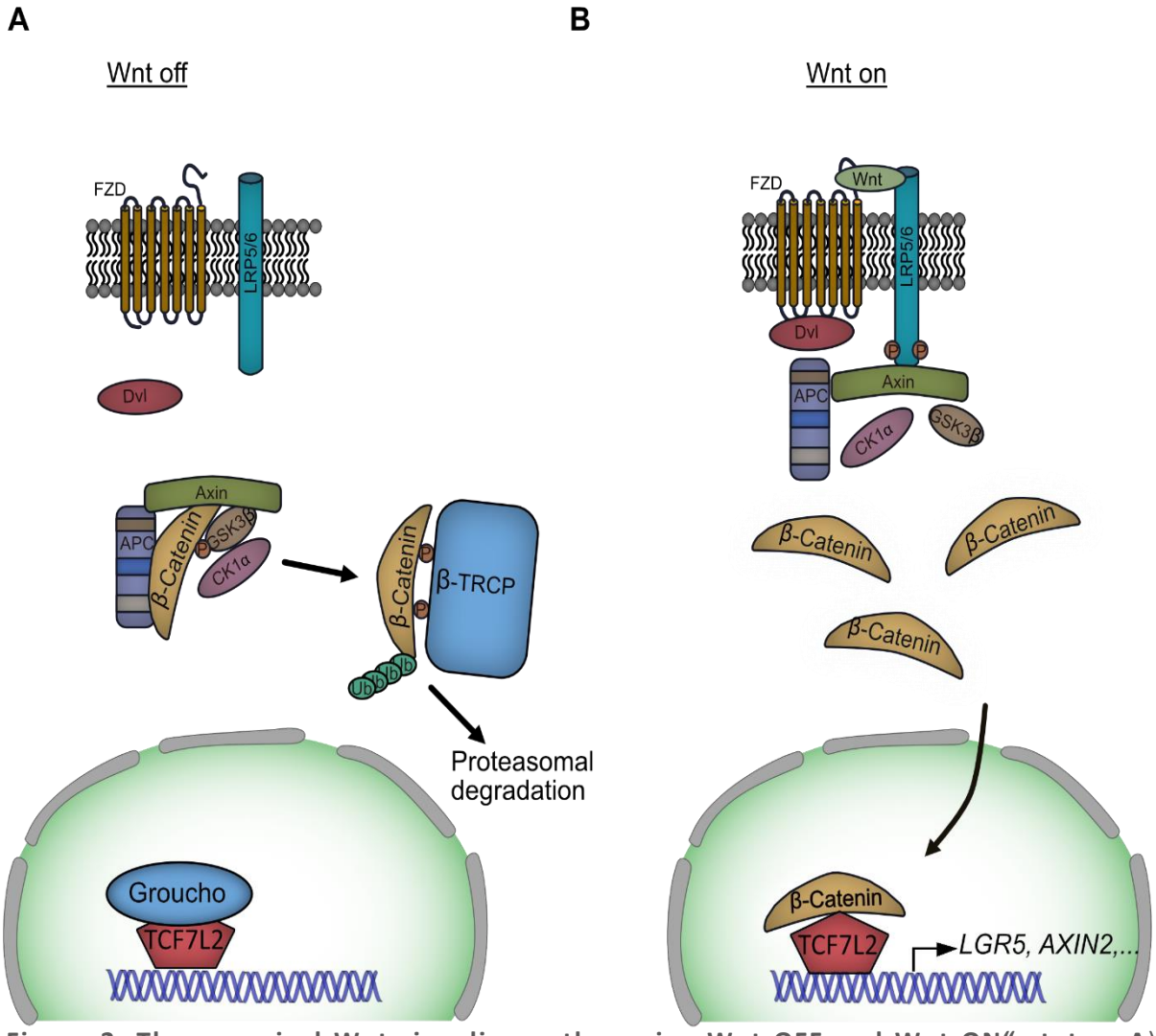


Figure 2: The canonical Wnt signaling pathway in „Wnt OFF and Wnt ON" states. A) **OFF**: If the Wnt receptors Frizzled and lipoprotein receptor-related protein (LRP) are not occupied by ligands, β-Catenin is bound by the destruction complex (DC) including AXIN, APC, GSK-3 and CKI, and β-TrCP. After β-Catenin is phosphorylated and ubiquitinated, it is degraded by the proteasome. In the nucleus, the transcription factor TCF7L2 remains repressed by Groucho and target gene expression is blocked. **B**) **ON**: Upon binding of Wnt ligands to the membrane receptors, AXIN is relocated to the cell membrane and no DC is formed. Intracellular β-Catenin levels increase, it translocates to the nucleus and activates expression of target genes by forming a TCF7L2/β-Catenin complex.

the intestine and cancer progression (Logan and Nusse, 2004).

In the absence of Wnt ligands, representing the inactivated state, the transcriptional activator β-Catenin is sequestered in the cytoplasm and degraded via the proteasomal pathway (**Figure 2A**). A destruction complex, including the tumor suppressors AXIN and adenomatous polyposis coli (APC), the constitutively active serine-threonine kinases GSK3 and CK1 and disheveled (DVL), binds β-Catenin for phosphorylation. Phosphorylated β-Catenin is recognized and bound by the E3 ubiquitin ligase β-TRCP. Consequently, ubiquitination and proteasomal degradation of β-Catenin can occur and prevents its translocation into the nucleus. In the absence of β-Catenin, the Wnt transcription factors of the TCF/LEF family repress gene expression by recruiting co-repressors, such as Groucho/TLE proteins (Daniels and Weis, 2005).

In a Wnt "ON" state, binding of secreted Wnt ligands to the FZD/LRP receptors at the cell membrane lead to the formation of a FZD/DVL complex and recruitment of AXIN from the destruction complex to the LRP receptor, where it becomes phosphorylated (**Figure 2B**). This prevents the formation of the functional complex and allows β-Catenin to accumulate in the cell and translocate into the nucleus, where it interacts with and displaces Groucho/TLE proteins from TCF/LEF factors to facilitate gene expression of Wnt targets like *LGR5*, *AXIN2* or *c-MYC* (Schneikert and Behrens, 2007; Nusse and Clevers, 2017). The best described member of the TCF family known to interact with β-Catenin in the colon is TCF7L2 (previously named TCF4). The second relevant member TCF1 lacks the N-terminal interaction domain but acts as an antagonist for TCF7L2 in a heterozygous *APC* mutation background (Van de Wetering *et al.*, 2002).

In 1998 Korinek *et al*. first described canonical Wnt signaling linked to the maintenance of adult epithelial stem cells in the intestine (Vladimir Korinek, Barker, Moerer, *et al.*, 1998). To further decipher the detailed regulation of homeostasis of intestinal stem cells by Wnt, multiple studies aimed for the identification of marker genes. By screening putative adult stem cells for β-Catenin target genes (van de Wetering *et al.*, 2002), the target gene *Lgr5* was detected in mouse intestinal crypts, located at the crypt bottom in colon or commingled with Paneth cells in the small intestine (Barker *et al.*, 2007). Long-term co-marker expression and lineage tracing of Lgr5+ cells expressing a LacZ reporter revealed that the labeled clones expand along

the vertical crypt axis into the villus compartment of the mouse small intestine. Consequently, Barker and colleagues could prove that all differentiated cell types present in the intestine are progeny of LGR5+ cells. Additional β-Catenin target genes like *AXIN2* (Yan *et al.*, 2001; Jho *et al.*, 2002) or Naked Cuticle homolog 1 (*NKD1*) (van Raay *et al.*, 2011) were later deciphered in detail and some serve today as bona-fide markers for intestinal stem cells e.g. achaete scute-like 2 (*ASCL2*) (van der Flier *et al.*, 2009), *EPHB2* (Merlos-Suárez *et al.*, 2011) or *PTK7* (Jung *et al.*, 2015).

The capacity of LGR5+ cells to renew cells in the intestine was also observed *ex vivo*. Embedding of Lgr5+ cells, purified from mouse crypts, in an extracellular matrix covered with a layer of culture medium providing stem cell niche factors EGF, Noggin and R-Spondin, allows growing of three-dimensional cell clusters which resemble the structure of the intestinal crypts, are lined by a villus-like epithelium, and were named as organoids (Sato *et al.*, 2009). The in-vitro growing organoids contained among the Lgr5+ cells all other differentiated cell types of the small intestinal epithelium, highlighting that Wnt active Lgr5+ cells from the intestine are bona-fide stem cells in *in vivo* and *ex vivo* settings.

Wnt signaling activity is essential for intestinal stem cell homeostasis, and components of the signaling pathway are frequently mutated in colorectal carcinomas. Evidence exists that upon mutation of APC, only stem cells expressing *LGR5* can promote the growth of intestinal adenoma, while differentiated cells compromised in APC functionality fail to do so (Barker *et al.*, 2009; Schepers *et al.*, 2012). Therefore, LGR5+ cells are widely considered as the cells-of-origin in intestinal cancers. However, if Wnt activation triggered by APC inactivation occurs simultaneously with the acquisition of additional mutations in *KRAS*, *TP53* and *SMAD4*, intestinal tumorigenesis can also develop from more differentiated cells (Tetteh *et al.*, 2016).

1.3. Incidence of colorectal cancer

The frequency and mortality of cancer is rapidly growing all over the world. With a population that is growing and aging, there is also a growing prominence of cancer as a leading cause of death. In 2018, 18.1 million new cases of cancer and 9.6 million

cancer related deaths were recorded worldwide (Bray *et al.*, 2018). Colorectal cancer (CRC) is the third most commonly diagnosed cancer with 1.8 million new cases and

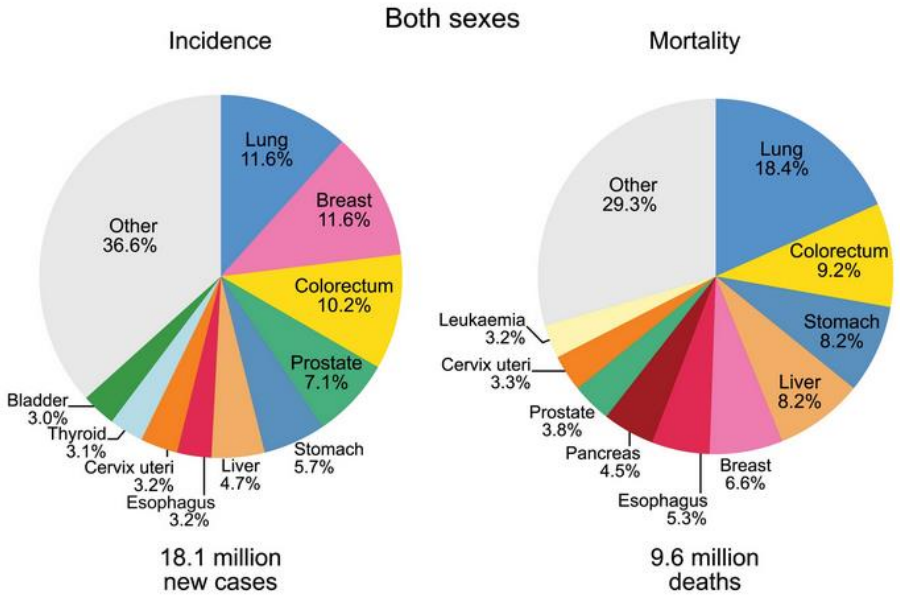


Figure 3: Incidence and Mortality of cancer in 2018. Pie chart representing the distribution of cases and mortality of the ten most common cancers in 2018, occurring at both sexes. Adapted from Bray et al., 2018.

the second leading cause of cancer death with 881.000 cases in 2018 (**Figure 3**), accounting for around 1 of 10 cancer cases and deaths among both, female and male patients (Bray *et al.*, 2018).

The incidence of CRC varies widely across world regions, is highest in Europe, Australia, Northern America and Eastern Asia, and correlates with the socioeconomic development of a population. Furthermore, colorectal cancer is contemplated as one of the clearest markers of epidemiological and nutritional transition, especially in western countries (Torre *et al.*, 2015). About 60% of CRC cases and deaths occur in well-developed countries with a higher average income. It is thus predicted, that over the next 15 years, the number of CRC will rise to more than 2.2 million new cases leading to over 1 million deaths per year. The increase in CRC incidence is associated with a number of dietary patterns and lifestyle factors, including processed meat, alcohol and elevated body fat due to reduced physical activity, all seen in the countries mentioned above (Perera, Thompson and Wiseman, 2012; Schlesinger *et al.*, 2017; Abar *et al.*, 2018). However, the same developed countries show reductions in the mortality rate linked to cancer, as they adopt improvements in cancer treatment and patient management (Fidler, Soerjomataram and Bray, 2016;

Arnold *et al.*, 2017).

Many years ago, research focused on treatment of advanced cancer, as the molecular pathogenesis remained unknown. Today, more studies focus on early detection of CRC or immune therapy of advanced stage CRC (Vogelstein *et al.*, 2013). However, until cancer deaths can be significantly reduced, greater efforts must be made toward early detection and prevention.

1.4. Genetic alterations leading to colorectal cancer

With the goal to understand and more differentially treat colorectal cancer, a rich history of research has revealed that CRC subtypes can be characterized by one of the following features: chromosomal instability (CIN), microsatellite instability (MSI) and a CpG island methylator phenotype (CIMP). In general, several critical mutations, consequently leading to aberrant activation or repression of WNT, RAS-MAPK, PI3K, TGFβ, p53 and other cell cycle regulating pathway components, are required for the phenotype of full-blown colorectal tumorigenesis. These so-called driver gene mutations typically activate oncogenes or cause loss-of-function of tumor suppressor genes (Vogelstein *et al.*, 1988). In the classical CIN pathway, which represents 65-70% of sporadic CRCs, the affected driver genes often suffer from mutations in one allele and loss of the non-mutated second allele (loss of heterozygosity, LOH). The occurrence and order of these mutations in the adenoma-carcinoma sequence was described by Fearon and Vogelstein (Fearon and Vogelstein, 1990) (**Figure 4**).

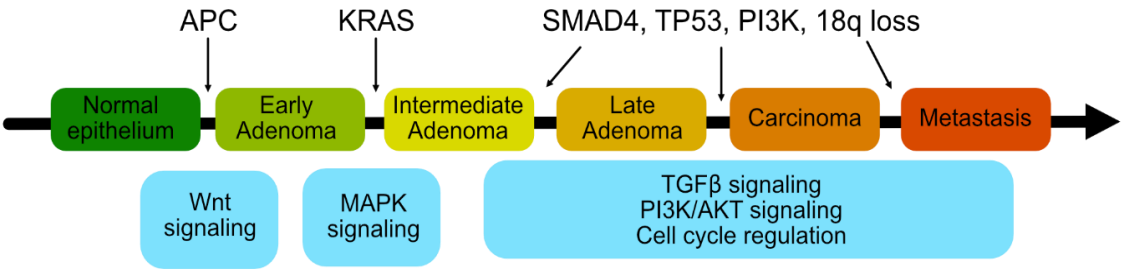


Figure 4: Adenoma-Carcinoma sequence model. Genetic model of colorectal tumor progression. Shown above are typical driver gene mutations and below the related pathway alterations. Illustration is adapted from Pinto and Clevers (2005).

In the early 1990s, several studies found that mutations in the Adenomatous

polyposis coli (*APC*) gene, located on chromosome 5 (5q21), are the underlying cause of familial adenomatous polyposis (FAP), in which numerous polyps form in the intestinal epithelium (Kinzler *et al.*, 1991; Powell *et al.*, 1992). Furthermore, “sporadic” proliferative but benign precursor lesions of dysplastic morphology (early adenomas/ small, neoplastic polyps), can arise from normal, not genetically pre-disposed epithelium. Their progression is spontaneously triggered by inactivating mutations in the *APC* gene. Besides somatic or germline mutations, hypermethylation of the *APC* promotor is also known to occur (Freitas *et al.*, 2018). *APC* mutations are also the initiating event in the classical pathway of colorectal tumorigenesis and can be found in roughly 70-80% of smallest adenomas, mostly located in the mutational cluster region (MCR). Adenomas have the potential risk to become cancerous if not surgically removed but only 10-15% of adenomas and polyps actually undergo progression to malignancy over years or decades (Powell *et al.*, 1992; Jass, 2007; Fearon, 2011).

Point mutations in *KRAS* often occur after *APC* mutations and constitutively activate the corresponding enzyme, which increases the MAPK-signaling pathway. Nearly 50% of colon cancers harbor activating mutations in KRAS, stimulating growth factor independency of adenomas and favor tumor progression (Haigis *et al.*, 2008). However, activated KRAS alone, without pre-existing *APC* mutations, is not sufficient to trigger the malignant transformation (Tariq and Ghias, 2016). When RAS is permanently activated, PI3K signaling activity is increased, further promoting cell growth and epithelial to mesenchymal transition (EMT) (Bos *et al.*, 1987; Fearon, 2011).

Another driver event in CRC tumorigenesis is the loss of heterozygosity (LOH) at the long arm in chromosome 18 (18q), which occurs in around 70% of cases. This frequency increases as the cancer progresses to a more advanced state. For instance, *DCC* (deleted in colorectal carcinoma) is located at 18q21.3 and encodes a transmembrane protein, which regulates cell growth in conjunction with Netrin-1. Netrin-1 is produced in the colonic crypts and its concentration decreases as the cells differentiate and move towards the surface. This way it regulates apoptosis of normal epithelial cells. In case of mutated or lost *DCC*, Netrin-1 cannot bind to *DCC*, leading to abnormal cell survival, which correlates with poorer survival in CRC (Vogelstein *et al.*, 1988; Jen *et al.*, 1994; Armaghany *et al.*, 2012).

LOH at chromosome 18 can also affect *SMAD4* (18q21.2), representing one of the later events in CRC carcinogenesis. Dysregulations in the TGFβ-signaling pathway are caused by loss of *SMAD4* or its partner transcription factors *SMAD2* and *SMAD3,* as well as mutations in e.g. *TGFBRII*. A third of all CRC cases show mutations causing inactivation of the TGFβ receptors or SMADs, which serve as modulators of cell growth in epithelial cells (Grady *et al.*, 1998). Furthermore, studies and mouse models harboring *Smad4* mutations revealed that defects or loss of *SMAD4* correlates with poor survival and distant metastasis, as well as increased transition to malignancy and drug resistance (Papageorgis *et al.*, 2011; Fleming *et al.*, 2013).

Another genetic key step in CRC development is the mutation of the tumor suppressor gene *TP53*, which is located on the short arm of chromosome 17 and normally activated under cellular stress. TP53 is a transcription factor and its regular targets include cell cycle inhibitors and pro-apoptotic factors. In nearly half of colorectal adenomas (45%) and carcinomas (43%), TP53 is found dysfunctional by a combination of missense mutations or loss of 17p (Vogelstein and Kinzler, 2004; Markowitz and Bertagnolli, 2009; Cooks *et al.*, 2013).

Overall, survival rates depend on the cancer progression at the point of diagnosis. Patients diagnosed with stage I CRC, can expect a 90% 5-year survival rate, compared to 70% at stage III and only 13% for patients already suffering from distant metastatic colorectal cancer (stage IV) (Siegel, DeSantis and Jemal, 2014).

Colorectal cancer is a heterogeneous disease, as pathologically similar tumors often differ in treatment response and patient survival. Until recently, the classification of CRC mainly relied on parameters such as microsatellite instability, as well as on the mutational status of KRAS and BRAF (Jass, 2007; Shen *et al.*, 2007). Other approaches to subtype CRC included clinical and histopathological features and gene expression data by using genome-wide expression profiling of large patient cohorts (Muzny *et al.*, 2012; Perez Villamil *et al.*, 2012; Xie *et al.*, 2012). However, a missing “gold-standard” for CRC subtyping impeded translational and clinical utility. In order to better characterize biological features of CRC core subtypes, linked to cellular phenotypes and clinical behavior, an international consortium was formed. In that study, several CRC cohorts were used to biologically define four consensus molecular subtypes (CMS) (Guinney *et al.*, 2015). The majority of BRAF-mutated MSI tumors were classified in the CMS1, showing hypermutations with a low prevalence

of somatic copy number alterations (SCNAs) and increased expression of genes associated with diffuse immune infiltration. CMS2 includes samples with copy number gains in oncogenes and losses of tumor suppressor genes, accompanied by low epithelial cell differentiation and high activation of Wnt and MYC. CMS3 categorized tumors display metabolic dysregulation combined with features of epithelial cells and mutations in *KRAS*. Tumors of the CMS4 are more mesenchymal, show stromal infiltration and display an upregulation of genes associated with EMT and activation of TGFβ signaling. By providing a common foundation of CRC subtype classification, which can be further connected to various “omics” and clinical data, this model can help to understand the details of tumor biology and facilitate targeted drug therapy.

1.5. Aberrant activation of Wnt signaling and the role of APC

The homeostasis of the intestinal epithelium is tightly regulated by Wnt signaling. Several mutational events can rattle up this balanced process, resulting in

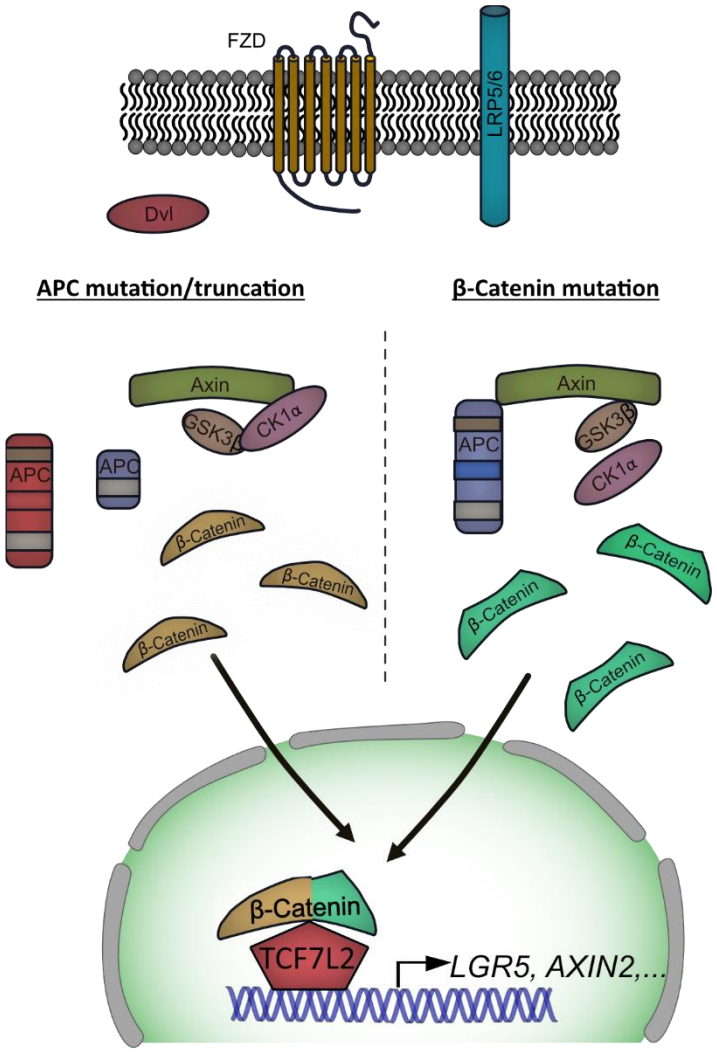


Figure 5: Aberrant activation of Wnt signaling. Perverted activation of Wnt can be triggered by mutations or truncations of the tumor suppressor *APC*, which restrains the formation of the destruction complex. Activating mutations of β-Catenin remove a phosphorylation site, which is needed for stabilization of β-Catenin. All scenarios lead to nuclear accumulation of β-Catenin and activation of the TCF7L2/β-Catenin complex, which drives Wnt target gene transcription.

tumorigenesis and colorectal cancer. Common initial events in CRC are inactivating mutations of the tumor suppressor APC, as well as activating mutations in *CTNNB1*, leading to nuclear accumulation of β-Catenin (**Figure 5**). Human CRC-derived cell lines often show enhanced levels of Wnt activity and consequently TCF/β-Catenin mediated transcription (Korinek *et al.*, 1997; Pinto and Clevers, 2005). Studies could prove that the same pathway which regulates the homeostasis of the intestinal epithelium, imposes a crypt/progenitor phenotype on colon carcinoma cells,

suggesting that Wnt target genes are the primary transforming event in CRC (Batlle, Henderson, Beghtel, van den Born, *et al.*, 2002; Van de Wetering *et al.*, 2002).

APC is a large protein (312 kDa), which interacts with many partners, most prominently AXIN and β-Catenin. As a tumor suppressor, it functions as part of the destruction complex but also sequesters nuclear β-Catenin into the cytoplasm for degradation. Necessary for proper binding and regulation of β-Catenin are three main structural motifs (**Figure 6**): Three 15-aminoacid (aa) repeats enable binding of β-Catenin combined with seven 20-aa repeats involved in binding and targeting for proteasomal degradation of β-Catenin. Additionally, *APC* harbors 3 SAMP repeats enabling AXIN binding (Rubinfeld *et al.*, 1993; Smith *et al.*, 1993; Peifer, Berg and Reynolds, 1994).

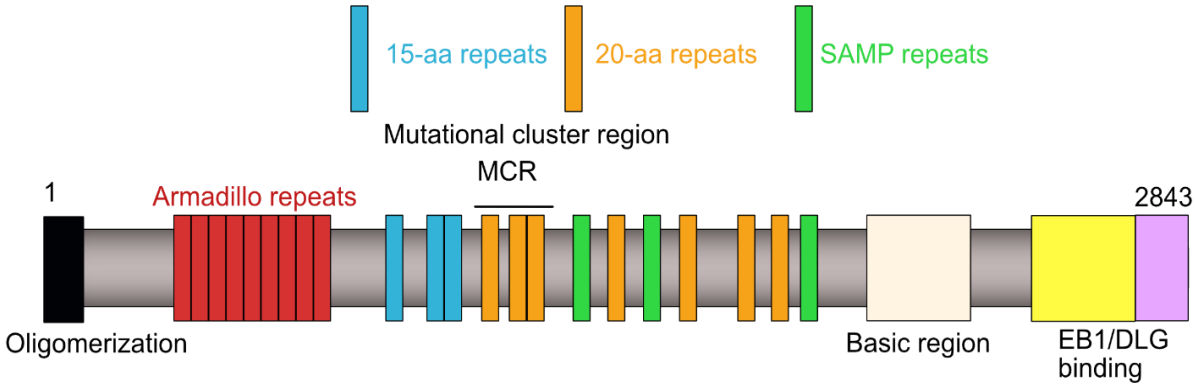


Figure 6: Functional domains of the APC gene. The N-terminus of the APC protein consists of an oligomerization domain (shown in black) and an armadillo region (red). In the center are 15- and 20-amino acid repeats (blue and orange) for binding and regulation of β-Catenin, as well as SAMP repeats (green) for binding of AXIN. At the C-terminus, APC has a basic domain (white) and sites for EB1 and DLG protein binding (yellow, purple).

More than 300 different *APC* mutations are known to be associated with colorectal cancer, mostly insertions and deletions leading to frameshift and premature STOP codons. Many of these mutations occur within the so-called mutational cluster region (MCR), located in exon 15, resulting in truncated transcripts of the gene which then encodes only one or several (usually up to 3) of the 20-aa repeats (**Figure 6**). Several studies showed that truncated APC proteins, retaining two or three out of the 7 20-aa repeats, are still able to regulate β-Catenin activity to some extent (Munemitsu *et al.*, 1995; Laurent-Puig, Béroud and Soussi, 1998). Many aggressive tumors with these types of mutant *APC* alleles can therefore still mediate cytoplasmic sequestration and/or ubiquitination of this transcriptional co-activator (Yang *et al.*,

2006). The biological outcome is an incomplete activation of the Wnt signaling pathway. Clinical evidence and experimental data from mouse models showed that longer *Apc* isoforms enhance polyposis, compared to shorter *Apc* versions, which supports the so-called “just right” hypothesis (Albuquerque *et al.*, 2002). This model suggests that the functionality of APC ought to be partially impaired to allow an intermediate strength of Wnt signaling and hence balanced nuclear accumulation of β-Catenin to an extent, which is more favorable for pre-malignant disease progression than a total functional ablation of APC that might impose exaggerated levels of oncogenic stress on cells.

1.6. Cellular mechanisms of glycosylation

Glycan structures are one of the basic cell components, regulating several physio-pathological processes and are created in an enzymatic progress by linking saccharides to proteins and lipids (Cummings, 2009; Varki and Gagneux, 2017). Glycoproteins usually carry one or more glycans on their polypeptide backbone, which are covalently attached by glycosyltransferases. Depending on their linkages, they are called N-glycans (nitrogen-linkage to an Asp residue) or O-glycans (oxygen linkage to a Ser or Thr residue). Finally modified glycoproteins function as surface proteins, transcription factors, receptors or structural proteins, adhesion proteins and ECM molecules (Clausen and Bennett, 1996; Bennett *et al.*, 2011; Oliveira-Ferrer, Legler and Milde-Langosch, 2017). In case of O-glycans, they are classified by the initiating monosaccharide linked to the amino acid residue. A common type of O-glycans called “Mucin-type”, is initiated via O-mannose and can be further extended by a variety of sugars to produce so-called “cores”, thereby allowing the formation of either linear or branched chain glycoproteins (**Figure 7**).

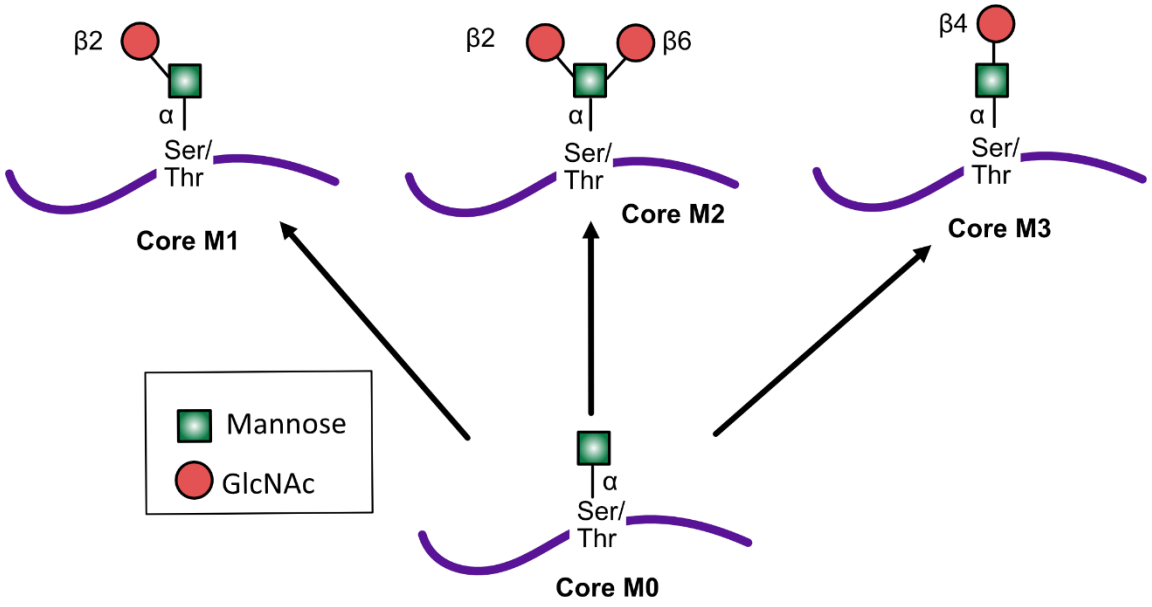


Figure 7: Set up of “Mucin type” O-mannosylated core structures. The biosynthesis is catalyzed in the ER where O-mannose is transferred to the hydroxyl-oxygen of Ser/Thr residues (Core M0). M0 can further be extended by a β1,2-linked (M1), β1,6-linked (M2) or β1,4-linked (M3) GlcNAc, depending on the active glycosyltransferase.

Mucin-type O-glycans are frequently found in secreted or membrane-associated glycoproteins. Another sort of O-glycans are the nucleocytoplasmic N-

acetylglucosamine (O-GlcNAc), N-acetylgalactosamine (O-GalNAc) and proteoglycans, like as heparan sulfate or hyaluronic acid (I. Brockhausen J. Yang, M. Lehotay, 2005; Ma and Hart, 2014; Brockhausen and Stanley, 2017). The stepwise assembly of O-glycans involves several glycosyltransferases throughout the ER-Golgi pathway (Colley KJ, Varki A, 2017) and produces various structural modifications required for protein folding, molecular trafficking, signal transduction, receptor activation or cell adhesion (Varki, 1993; Roth, 2002; Spiro, 2002; Moremen, Tiemeyer and Nairn, 2012). The absence or defect of a single glycosyltransferase or kinase can affect multiple protein modifications as each glycan-modification depends on numerous cell intrinsic factors and differs in a cell-, protein-, and site-specific way, which provides a tremendous molecular heterogeneity (Varki and Gagneux, 2017).

Alterations in glycosylation can be due to one or more factors. Deregulated expression or aberrant functionality of glycosyltransferases can lead to hyper- or hypo-glycosylation of proteins. Second, changes in the tertiary conformation of the peptide backbone can lead to non-functional glycan chains. Studies have shown that differences in the available donor substrates or abundance of other cofactors and sugars can influence the expression of glycan structures (Kellokumpu, Sormunen and Kellokumpu, 2002; Schietinger *et al.*, 2006; Aryal, Ju and Cummings, 2010). Defects in this posttranslational modifications are typical for diseases like congenital muscular dystrophies (Shaheen *et al.*, 2013), muscle-eye-brain-disease (Dwyer *et al.*, 2012), cell adhesion in e.g. early embryonic development (Lommel *et al.*, 2013) and cancer (Christiansen *et al.*, 2014; Pinho and Reis, 2015). Numerous studies have identified glycans involved in fundamental biological processes of cancer progression, where they are just as necessary as in normal tissues for inter-and intracellular signaling, regulation of metabolism, inflammation and cell-cell-adhesion. Exemplary, by directly interfering with the glycoprotein E-Cadherin, a major cell adhesion molecule, tumor cells can gain the ability to overcome adhesion and spread in surrounding tissue (Pinho *et al.*, 2011; Bassagañas *et al.*, 2014). Wnt/β-Catenin signaling and protein N-glycosylation have recently been reported to regulate the E-cadherin mediated intercellular adhesion (Sengupta, Bouchie and Kukuruzinska, 2010). Upon activation of β-Catenin, the N-glycosyltransferase DPAGT1 enhances the complexity of N-glycans on the ectodomains of E-cadherin. Moreover, dysregulations of DPAGT1 have been shown to affect the organization of adherens

junctions by E-cadherin (Vargas *et al.*, 2016). Additionally, DPAGT1 is involved in the N-glycosylation of the Wnt components LRP5/6 and Wnt3a, as overexpression of *DAPGT1* led to enhanced Wnt activity and increased secretion of hyperglycosylated Wnt3a and LRP5/6, thereby creating a positive feedback loop which may drive disease if activated (Sengupta *et al.*, 2013).

1.7. Dystrophin-associated glycoprotein 1 (DAG1)

The dystrophin-associated glycoprotein 1, short DAG1, was found to play a major role in muscle function and its malfunction leads to muscular dystrophy. First described in the late 1980s in the context of Duchenne muscular dystrophy, it was thought to be a complex of several independently encoded transmembrane, cytoskeletal and extracellular proteins (Koenig, Monaco and Kunkel, 1988). Only a few years later the group of Kevin Campbell deciphered that the glycoprotein complex originates from one and the same precursor protein, encoded by the same 5.8 kb mRNA and found in a variety of muscle and non-muscle tissue (Ibraghimov-Beskrovnaya *et al.*, 1992). The encoded 895 aa DAG1 precursor protein is inactive until posttranslational cleavage at the ectodomain into the extracellular subunit α-Dystroglycan (α-DG, residues 1-653, 156K) and the transmembrane subunit β-Dystroglycan (β-DG, residues 654-895, 43K) (**Figure 8**).

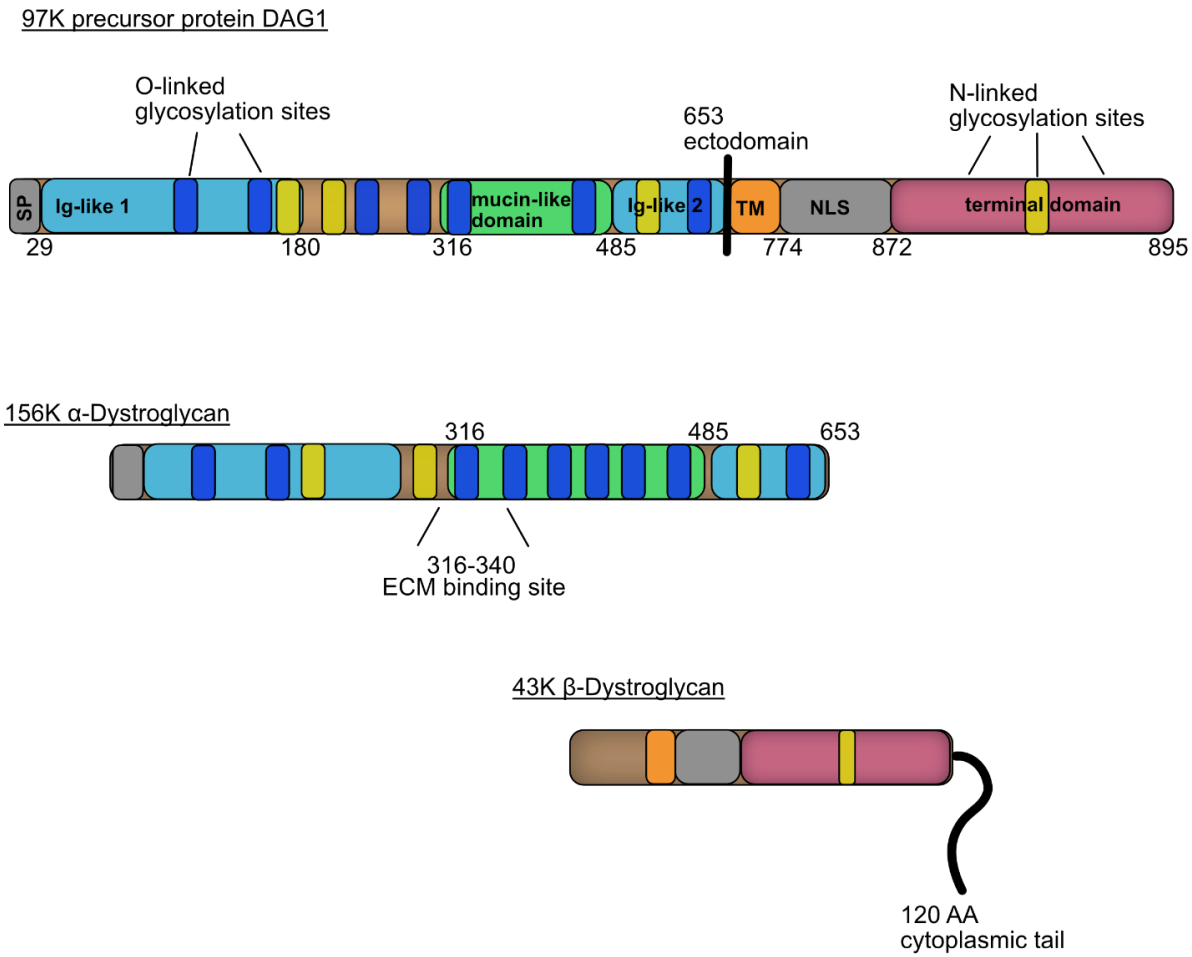


Figure 8: Representation of the DAG1 precursor protein and its processed subunits. The 97K sized pre-protein is posttranslational cleaved in two subunits. The ECM binding α-Dystroglycan contains a highly O-glycosylation (colored in dark blue) linked mucin-like domain (green) and two Ig-like domains (light blue). The transmembrane β-Dystroglycan is N-glycosylated (yellow) and harbors the transmembrane domain (TM, orange), as well as a nuclear localization site (NLS, grey), a terminal domain (red) and a cytoplasmic tail. Figure based on Ibraghimov-Beskrovnaya et al. (1992).

The first amino acids in the open reading frame of *DAG1* were mapped as the signal peptide, followed by two globular regions (Ig-like), containing three N-glycosylation sites. They are separated by a highly O-glycosylated mucin-like domain, including the ECM binding site of α-DG. By passing through the secretory pathway, α-DG is further post-translationally processed by the calcium dependent serine endoprotease furin, resulting as mature protein (residues 316-653) (Ervasti and Campbell, 1991; Bello and Darribère, 2018). β-DG is the C-terminus of the precursor protein and includes a transmembrane domain, the nuclear localization site and a single N-glycosylation site, terminated by a 120 aa cytoplasmic tail (Ibraghimov-Beskrovnaya *et al.*, 1992; Holt *et al.*, 2000) (**Figure 8**). Smaller size

transmembrane proteins were identified as variants of β-DG, resulting from aberrant procession. Intracellularly, β-Dystroglycan binds via its cytoplasmic tail to dystrophin, which in turn is associated with the actin cytoskeleton and other proteins, such as GRB2.

α-DG probably is one of the most well studied O-glycan modified proteins. The size of α-DG ranges from 120 to more than 250 kDa, depending on the amount of glycosylation. High MW forms of α-DG are found in skeletal and cardiac muscle membranes, whereas in brain and lung tissue, the size of α-DG was reduced to ~120 kDa. Therefore, the variability of glycosylation seems to depend on the tissue type (Ibraghimov-Beskrovnaya *et al.*, 1992). The degree of glycosylation is consequently directly associated with the affinity to bind ECM proteins containing LG-domains, like laminin, agrin or perlecan, forming the link between the cytoskeleton and the basal lamina (Michele and Campbell, 2003) (**Figure 9**).

In the past years, several studies have analyzed the physiological and pathological role of Dystroglycan. As the interaction of α-DG with integrins plays a significant role in early embryonic development and cancer, the expression pattern and function of glycoproteins like α-DG got more into focus. In particular, it was shown that the glycosylation pattern of α-DG changes during embryogenesis in chicken and mouse development and was compromised after day 6.5 upon gene disruption (Williamson *et al.*, 1997; Leschziner *et al.*, 2001). In multiple other studies aberrant glycosylation of α-DG was linked to dystroglycanopathies and tumors (Weir *et al.*, 2006; Akhavan *et al.*, 2012; Hohenester and Yurchenco, 2013; Bello *et al.*, 2015). The variability of α-DG glycosylation previously found in normal tissues is also present in tumors originating from different organs, but also within carcinoma cell lines from the same cancer entity (Losasso *et al.*, 2000). Some of the first studies investigated the expression pattern of α-DG in a panel of both normal and cancer cell lines from human mammary gland and human colon. The amounts of α-DG glycosylation were reduced in several cell lines, concluding that these alterations could have severe influence on the contact of epithelial cells to the basement membranes and therefore not only influence signal transduction and tumorigenic growth, but also invasiveness (Losasso *et al.*, 2000; Muschler *et al.*, 2002; Sgambato *et al.*, 2003).

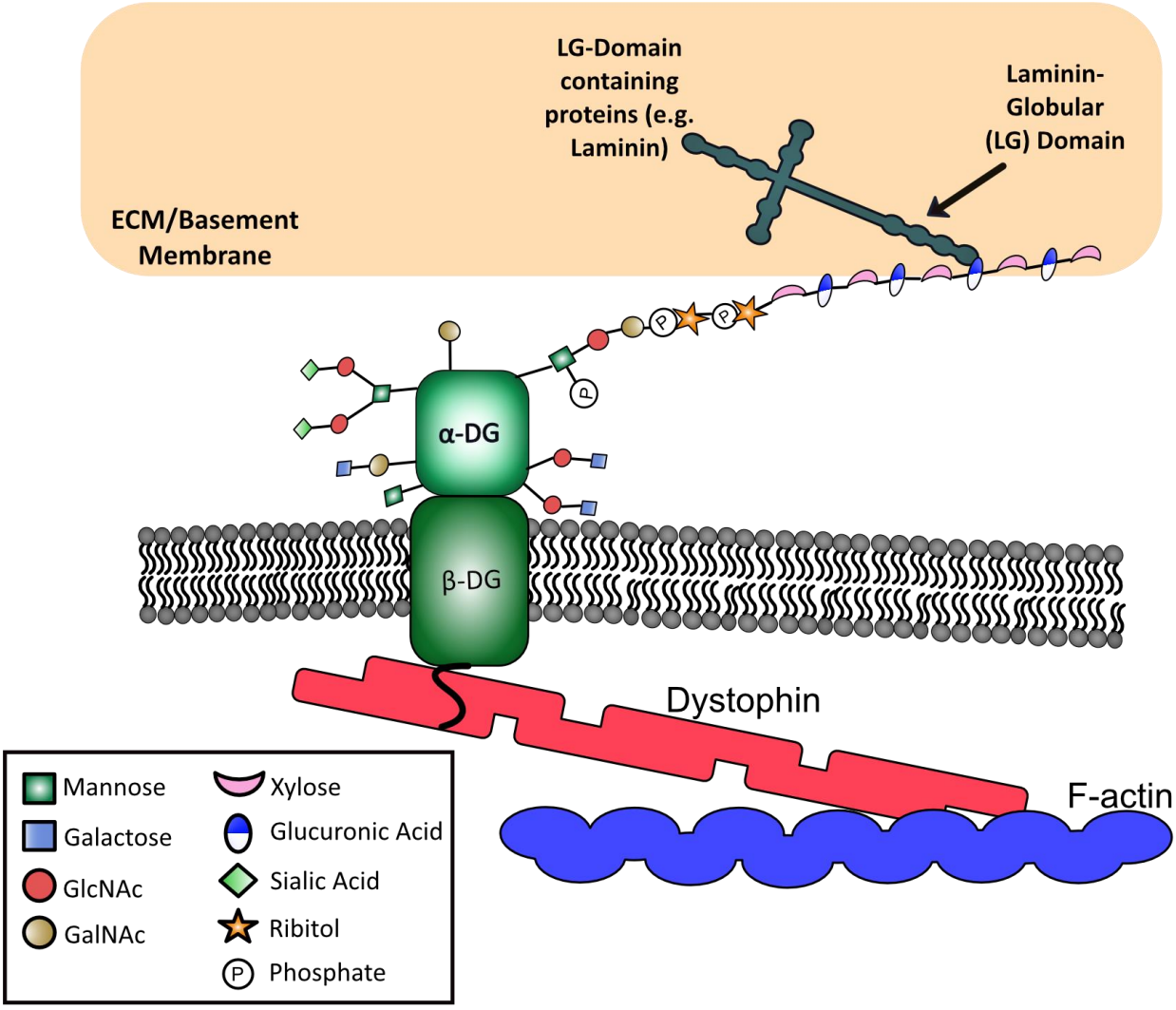


Figure 9: The Dystroglycan-Complex (DGC). The DGC consists of the heavily glycosylated, ECM binding α-DG, mediating interactions to the LG-Domain of proteins like laminin via repeating disaccharides. The transmembrane β-DG connects to the actin cytoskeleton and other intracellular proteins via dystrophin. Figure based on Sheikh et al. (2017).

In total there are 11 glycosyltransferases and kinases actively extending the O-mannose M3 core on α-DG to build up a glucosaminaminoglycan disaccharide (referred to as matriglycan), which is essential for α-DG to bind to laminin globular domains of ECM proteins (**Figure 9**). Another 12 enzymes are involved as indispensable donor substrates or regulatory subunits.

The initiating event in glycosylation of α-DG is the attachment of O-mannose at a Ser or Thr chain and building of the M3 core structure by the glycosyltransferases POMT1 or POMT2 and POMGNT2. The core is further extended by B3GALNT2 and

phosphorylated by the kinase POMK, all located in the ER. Phospho-ribitol transferases (FKTN and FKRP) elaborate the trisaccharide to enable addition of a single xylose (Xyl) and glucuronic acid (GlcA) by TMEM5 and B3GBT1, serving here as priming enzymes for the bifunctional glycosyltransferases LARGE1 and its paralog LARGE2 (GYLTL1B), which synthesize a high molecular weight matriglycan structure consisting of alternating Xyl and GlcA disaccharides (**Figure 10**). This matriglycan moiety is responsible for the laminin-binding capability of a-DG.

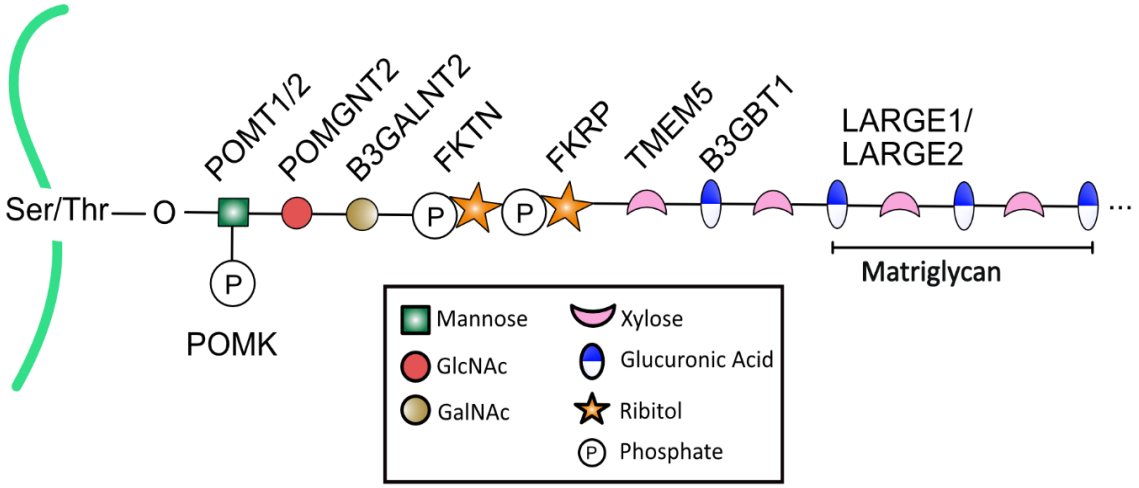


Figure 10: Representative composition of the M3 core functional O-glycan. The O-linked phosphorylated monosaccharide is prepared by POMT1/2, POMGNT2 and POMK. Extension and priming by FKTN, FKRP, TMEM5 and B3GBT1 enable the addition of, repeating Xyl-GlcA disaccharides (indicated as matriglycan) by LARGE1/2. As the composition is just a representative model, glycan symbol nomenclature differs from Varki et al (2015). Figure based on Sheikh et al. (2017).

2. Aim of the study

Wnt signaling plays a significant part in colorectal cancer (CRC) development, and in maintaining the self-renewal capacity of human colonic stem cells (hCoSC). Since activation of Wnt signaling in CRC is often caused by mutations in the *APC* gene, the first objective of this work was to explore new target genes of the Wnt signaling pathway after acute silencing of *APC*.

The second aim was to further characterize LARGE2, a bifunctional glycosyltransferase, which we found is driven by activation of Wnt-signaling. We describe the direct linkage between TCF7L2-dependent Wnt signaling and O-glycosylation of α-Dystroglycan in human colonic epithelia and colorectal cancer and the effect of Wnt-pathway deregulation on α-Dystroglycan functionality, which is the binding of laminin and laminin-domain containing proteins.

3. Materials

3.1. Chemicals and reagents

Application	*Chemical compound*	*Supplier*
2D Cell culture	4-Hydroxytamoxifen	Sigma, St. Louis, MO, USA
	DMEM	Invitrogen, Karlsruhe, Germany
	Doxycycline Hydrochloride	Sigma, St. Louis, MO, USA
	FBS	Gibco®, Life Technologies, Darmstadt, Germany
	Laminin-111	SantaCruz Biotechnology Inc., Dallas, TX, USA
	Lenti-X Concentrator	Takara, Kusatsu, Japan
	Lipofectamine LTX	Invitrogen, Karlsruhe, Germany
	Opti-MEM	Invitrogen, Karlsruhe, Germany
	PBS	Sigma, St. Louis, MO, USA
	Penicillin-Streptomycin (10.000 u/ml)	Gibco®, Life Technologies, Darmstadt, Germany
	Polybrene	Sigma, St. Louis, MO, USA
	Puromycin	Sigma, St. Louis, MO, USA
	PEI	Sigma, St. Louis, MO, USA
	RPMI	Invitrogen, Karlsruhe, Germany
3D cell culture	Matrigel	Corning, New York, NY, USA
	Advanced DMEM/F12	Invitrogen, Karlsruhe, Germany
	B27	Invitrogen, Karlsruhe, Germany

	NAc	Sigma, St. Louis, MO, USA
	EGF	Preprotech, Hamburg, Germany
	Noggin	Preprotech, Hamburg, Germany
	HEPES	Invitrogen, Karlsruhe, Germany
	GlutaMAX	Invitrogen, Karlsruhe, Germany
	PGE2	Sigma, St. Louis, MO, USA
	Normocin	Invitrogen, Karlsruhe, Germany
	SB	Sigma, St. Louis, MO, USA
	Y	Biozol Diagnostics, Eching, Germany
	LY	Selleckchem, Munich, Germany
	N2	Invitrogen, Karlsruhe, Germany
	Cell recovery solution	Corning, Corning, NY, USA
	Antibiotic-Antimycotic	Gibco®, Life Technologies, Darmstadt, Germany
Western blotting	SDS	Carl Roth, Karlsruhe, Germany
	Tris	Sigma, St. Louis, MO, USA
	APS	AppliChem, Darmstadt, Germany
	TEMED	Carl Roth, Karlsruhe, Germany
	Tris-Bis Acrylamide	Carl Roth, Karlsruhe, Germany
	Triton X-100	AppliChem, Darmstadt, Germany

	Bradford Reagent	Sigma, St. Louis, MO, USA
	WGA Agarose	Vector Labs, Burlingame, CA, USA
	TEA	Sigma, St. Louis, MO, USA
	DTT	Sigma, St. Louis, MO, USA
	NuPAGE™ 4-12% Bis-Tris Gels	Invitrogen, Karlsruhe, Germany
	LDS sample buffer (NuPAGE)	Invitrogen, Karlsruhe, Germany
	Protease Inhibitor	Sigma, St. Louis, MO, USA
	Immobilon ECL/HRP Substrate	Merck Millipore, Darmstadt, Germany
	SuperSignal™ west femto maximum sensitivity substrate	Thermo Scientific, Waltham, MA, USA
	Immobilon-P PVDF, 0.45µm Membrane	Merck Millipore, Darmstadt, Germany
	Methanol	Carl Roth, Karlsruhe, Germany
	PageRuler ™ Prestained Protein Ladder	Invitrogen, Karlsruhe, Germany
	Skim milk powder	Sigma, St. Louis, MO, USA
	BSA	Sigma, St. Louis, MO, USA
qRT-PCR	Fast SYBR® Green Master Mix	Invitrogen, Karlsruhe, Germany
	primaQUANT 2* qPCR CYBR Green Master Mix	Steinbrenner, Wiesenbach, Germany
In situ hybridization	DEPC	Sigma, St. Louis, MO, USA
	Paraformaldehyde	Sigma, St. Louis, MO, USA
	Proteinase K	Sigma, St. Louis, MO, USA

	Glycine	Sigma, St. Louis, MO, USA
	HCl	Sigma, St. Louis, MO, USA
	Acetic Anhydride	Sigma, St. Louis, MO, USA
	CHAPS Hydrate	Sigma, St. Louis, MO, USA
	Heparin Sodium Salt	Sigma, St. Louis, MO, USA
	Blocking Reagent	Roche, Penzberg, Germany
	EDTA	VWR, Radnor, PA, USA
	Formamide	Sigma, St. Louis, MO, USA
	Yeast RNA	Sigma, St. Louis, MO, USA
	BCIP/NBT Liquid Substrate	Roche, Basel, Switzerland
RNAscope	RNAscope 2.5 HD Assay - BROWN	Bio-Techne, Minneapolis, MS, USA
	Ammonium hydroxide 28-30%	Sigma, St. Louis, MO, USA
	Hematoxylin	Sigma, St. Louis, MO, USA
Cell transfection	RNAiMAX	Invitrogen, Karlsruhe, Germany
	Lipofectamine 2000	Invitrogen, Karlsruhe, Germany
	Lipofectamine LTX	Invitrogen, Karlsruhe, Germany
RNA Isolation	Trizol	Invitrogen, Karlsruhe, Germany
	Chloroform	Sigma, St. Louis, MO, USA
Cloning	LB- Medium	Carl Roth, Karlsruhe, Germany
	LB-Agar	Carl Roth, Karlsruhe, Germany
	dNTP solution Mix	NEB, Frankfurt a.M., Germany
	Ampicilin sodium salt	AppliChem, Darmstadt, Germany

	Kanamycin sulfate	Th. Geyer, Renningen, Germany
Immunohistochemistry	3,3‘-diaminobenzidine	Agilent, Santa Clara, CA, USA
	Crystal MausBlock	DCS, Hamburg, Germany
	Target retrieval solution	Agilent, Santa Clara, CA, USA
	Hematoxylin	Vector Labs, Burlingame, CA, USA

3.2. Enzymes

Application	*Chemical compound*	*Supplier*
Cell culture	Trypsin	Gibco®, Life Technologies, Darmstadt, Germany
	TripLE Select	Gibco®, Life Technologies, Darmstadt, Germany
	Dispase II	Stemcell Technologies, Vancouver, Canada
	Accutase	Stemcell Technologies, Vancouver, Canada
	Collagenase IV	Gibco®, Life Technologies, Darmstadt, Germany
Cloning	Gateway® BP Clonase	Invitrogen, Karlsruhe, Germany
	Gateway® LR Clonase	Invitrogen, Karlsruhe, Germany
	Phusion® High Fidelity Polymerase	NEB, Frankfurt a.M., Germany
	Q5® High Fidelity Polymerase	NEB, Frankfurt a.M., Germany
	Fast-AP Thermosensitive Alkaline Phosphatase	Invitrogen, Karlsruhe, Germany
	T4 Ligase	NEB, Frankfurt a.M., Germany
	Quick Ligase	NEB, Frankfurt a.M., Germany
	Taq DNA Polymerase	Biozym Scientific, Oldendorf, Germany
Restriction endonucleases	EcoRI-HF	NEB, Frankfurt a.M., Germany
	HindIII-HF	NEB, Frankfurt a.M., Germany
	AgeI-HF	NEB, Frankfurt a.M., Germany
	MfeI	NEB, Frankfurt a.M., Germany

	SgrAI	NEB, Frankfurt a.M., Germany
RNA Isolation	PureLink DNAse	Invitrogen, Karlsruhe, Germany

3.3. Kits

Application	***Chemical compound***	***Supplier***
DNA Isolation	GenElute Mammalian Genomic kit	Sigma, St. Louis, MO, USA
RNA Isolation	RNeasy ® Mini Kit	Qiagen, Hilden, Germany
	High Pure RNA Isolation Kit	Roche, Penzberg, Germany
	PureLink RNA purification Kit	Invitrogen, Karlsruhe, Germany
	Qubit RNA HS Assay	Invitrogen, Karlsruhe, Germany
cDNA Transcription	High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems, Foster City, CA, USA
In situ hybridization	DIG RNA Labeling Kit (SP6/T7)	Roche, Basel, Switzerland
Genotyping	Alt-R® Genome Editing Detection Kit	IDT Inc., Coralville, IA, USA
	CloneJET PCR Cloning Kit	Thermo Scientific, Waltham, MA, USA
Plasmid purification	NucleoSpin® Plasmid Kit	Macherey-Nagel, Düren, Germany
	PureLink ™ HiPure Plasmid Maxiprep Kit	Invitrogen, Karlsruhe, Germany
Chromatin-Immunoprecipitation	SimpleChIP® Enzymatic Chromatin IP Kit	Cell signaling technology, Danvers, MA, USA
Cell culture	LookOut Mycoplasma PCR detection Kit	Sigma Aldrich, St. Louis, MI, USA
RNAscope	RNAscope Probe Hs	Bio-Techne, Abingdon, UK

	LARGE2	
	RNAScope IntroPack 2.5 HD Brown Hs Kit	Bio-Techne, Abingdon, UK
	RNAScope Probe-Hs-LGR5	Bio-Techne, Abingdon, UK

3.4. Antibodies

Application	*Name*	*Number*	*Dilution*	*Distributor*	*origin*
Western Blotting	APC	Ali 12-28	1:500	SCBT	mouse
	Actin	A2066	1:2000	Sigma	rabbit
	Tubulin	T9026	1:2000	Sigma	mouse
	αDG	IIH6C4	1:500	Millipore	mouse
	DAG1	11017-1-AP	1:500	PTGLAB	rabbit
	Laminin	NB300-144	1:2000	Novus Bio	rabbit
	Anti-mouse HRP	715-035-150	1:10000	Dianova	donkey
	Anti-rabbit HRP	711-035-152	1:10000	Dianova	donkey
	Anti-mouse IRDye 760RD	925-32210	1:10000	LiCOR	goat
	Anti-rabbit IRDye 680RD	925-68071	1:10000	LiCOR	goat
In situ hybridization	Anti-digoxigenin	11093274910	1:2000	Roche	sheep
Immuno-histochemistry	αDG	IIH6	1:20	SCBT	mouse
	Biotinylated anti-mouse IgM	BA2020	1:100	Vector Labs	goat
ChIP	TCF7L2	C48H11	1:1000	CST	rabbit
FACS	PTK7_APC coupled	CCK-4	1:20	Miltenyi	mouse
	αDG	IIH6C4	1:100	Millipore	mouse
	EphB2_APC coupled	BDB564699	1:100	BD Biosciences	mouse
	Anti-mouse FITC	515-095-062	1:200	Jackson ImmunoResearch	sheep

3.5. DNA constructs and oligonucleotides

3.5.1. Plasmids

Name	***Description***	***Distributor***
LARGE2 ORF	In pENTR221 Gateway full ORF clone	GPCF, DKFZ Heidelberg
SMOC2 ORF	In pENTR221 Gateway full ORF clone	GPCF, DKFZ Heidelberg
pDONR™221	Gateway™ destination vector	ThermoFisher Scientific, Waltham, USA
pMD2.G	2nd generation envelope vector for virus production	A. Trumpp, DKFZ Heidelberg
psPAX2	2nd generation packaging vector for virus production	A. Trumpp, DKFZ Heidelberg
pSpCas9(BB)-2A-GFP	Cas9 from S. pyogenes with 2A-EGFP, and cloning backbone for sgRNA	Addgene plasmid #48138
pSpCas9(BB)-2A-GFP LARGE2 g1	Cas9 from S. pyogenes with 2A-EGFP, guide RNA1 for LARGE2 knockout	
pSpCas9(BB)-2A-GFP LARGE2 g2	Cas9 from S. pyogenes with 2A-EGFP, guide RNA2 for LARGE2 knockout	
pLentiCRISPR-E	Lentiviral vector LentiCRISPR V2, which contains eSpCas9 and puromycin cassette	Addgene plasmid #78852
pLentiCRISPR-E LARGE2 g1	LentiCRISPR V2 which contains eSpCas9 and puromycin cassette, guide RNA1 for LARGE2 knockout	
pLentiCRISPR-E LARGE2 g2	LentiCRISPR V2 which contains eSpCas9 and puromycin cassette, guide RNA2 for LARGE2 knockout	

pLentiCRISPR-E TCF4_BS g1	LentiCRISPR V2 which contains eSpCas9 and puromycin cassette, guide RNA1 for TCF4_BS knockout	
pLentiCRISPR-E TCF4_BS g2	LentiCRISPR V2 which contains eSpCas9 and puromycin cassette, guide RNA2 for TCF4_BS knockout	
pTRIPZ non silencing	Glycerol stock, negative TRIPZ Inducible Lentiviral shRNA Control	Dharmacon, RHS4743
pTRIPZ APC shRNA 1-3	Glycerol set, TRIPZ Inducible Lentiviral Human APC shRNA 1-3	Dharmacon, RHS5087
FUW-CMV-NTCF-ERT2	Tamoxifen inducible vector for NTCF-ERT2 expression	Eduard Batlle, IRB Barcelona
FUW-CMV-ERT2	Tamoxifen inducible vector for ERT2 expression	Eduard Batlle, IRB Barcelona
pV2Luc2	luciferase reporter plasmid with very low basal activity	H. Hermeking, LMU Munich
pRL-TK	Renilla Luciferase Reporter Vector	Promega, Madison, USA
pTRIPZ	Doxycycline Inducible lentiviral vector	ThermoFisher Scientific, Waltham, USA
pTRIPZ LARGE2	Doxycycline Inducible lentiviral vector for LARGE2 overexpression	

3.5.2. Taqman Probes

Gene	***Number***	***Distributor***
ANPEP	Hs00174265_m1	ABI, ThermoFisher
B2M	Hs99999907_m1	ABI, ThermoFisher
KRT20	Hs00300643_m1	ABI, ThermoFisher
LARGE2	Hs00293342_m1	ABI, ThermoFisher

LGR5	Hs00173664_m1	ABI, ThermoFisher
MKI67	Hs00606991_m1	ABI, ThermoFisher
PPIA	Hs99999904_m1	ABI, ThermoFisher
SMOC2	Hs00405777_m1	ABI, ThermoFisher

3.5.3. Primer

3.5.3.1. qPCR

	forward	***reverse***
ANPEP	TTGACCGCTCCGAGGTCTAT	CTCGCTGTACTGGTCCATCA
APC	TCCTAGAGGAGCCAAGCCAT	TTGAGTTTCGGCCAGGAGAC
ASCL2	GTGAAGCTGGTGAACTTGGGC	CAGCGTCTCCACCTTGCTCA
AXIN2	ATGGGATGATCTGTTGCAGAGGGA	TGTCATTTCCACGAAAGCACAGCG
B2M	TCCATCCGACATTGAAGTTG	ACACGGCAGGCATACTCAT
B3GALNT2	CCTGCAGATCAGTTGGCCT	CAGCCCCGATGAAGTGTCTT
B3GNT1_1	CCGCATGCCCATGAACAAAA	CGCAGGCCTGGCTGATTC
CA12	GACCTGCACAGTGACATCCT	GCTGGCGTCAGGATAAAGGT
CHGA	CAAGACCTCGCTCTCCAAGG	CTCTTCCACCGCCTCTTTCA
CTNNB	AGCTGACCAGCTCTCTCTTCA	CCAATATCAAGTCCAAGATCAGC
DAG1	GGAGAACCCAACCAGCGCCCAGAGC	CGGGTGATATTCTGCAGGGTGATGG
FKRP	CCATTGCTCCAAGATGGCG	AGTAGAAACAGGGGGCTTGC
FKTN	TGGCCCTTTTAACGCTGACA	TTAACTGCACGCCACTGTGT
KI67	ACGAGACGCCTGGTTACTATC	GCTCATCAATAACAGACCCATTTAC
KRT20	ACTCCAGACACACGGTGAACT	TGAAGTCCTCAGCAGCCAG
LARGE1	TTCTGGAACAGAGCTGCACG	TCCTCGGCCTCCCTCATAAT
LARGE2	GGCCCTCACGCCTTACGT	GTAGCGCAGGGTCTCGAATG

LGALS4	GGCTCAACGTGGGAATGTCT	ACCGCGGATTGAAGTGGAAG
LGR5	AATCCCCTGCCCAGTCTC	CCCTTGGGAATGTATGTCAGA
NKD1	ACTTTCGGCTGGAAGTGGC	GTGACCTTGCCGTTGTTGTC
NOTUM	CTGGTGGAACGCAAACATGG	CACACGGTCCACATTCAGGA
POMGNT2	CAAACCTGTGCTGGCTGC	GGTGAAAGCCTGAAGCCTGG
POMT1	CCCCGAAGTAAGAACTCCGT	GCTCAGTAACCCCATCCCAG
POMT2	TCAGACAAAGTGTGCCTCCG	CTCATCCCAACAGATGTGCG
PPIA	AGCATGTGGTGTTTGGCAAA	TCGAGTTGTCCACAGTCAGC
RXYLT1	CTTTGCCCCCTACAGCCTC	GACTGTTCTTGCGGCGATTT
SMOC2	AACTGGTCTGCGTGCATAA	CTCTCAGACTGTCCTCCAAATG
SP5	CAGGCCTTTCTCCAGGACC	GGTGTAAGGGGAAGCTCGTG
TFF3	GCAGGAAGCAGAATGCACCT	AGCTGAGATGAACAGTGCCT
WNT11	TCTTTGGGGTGGCACTTCTC	CTGCCGAGTTCACTTGACGA

3.5.3.2. ChIP PCR

	forward	***reverse***
LARGE2_intron1_1	GAAGGCTTCAGGTGCTCCG	GGGATCAAAGACACTGCGGG
LARGE2_intron1_2	CAGAGCAGGTGTCCCAGAC	GGATCAAAGACACTGCGGGA
LARGE2 upstream_1	ACAGCGATCACTACCCTGACT	CCCTGACCTTTGATCAGTTCC
LARGE2 upstream_2	TCATAGGTCAGGGCCCAGAA	CAGTGCGGGGCTTCAGTTAC
LGR5	AGGTCTGGTGTGTTGCTGAG	GAGTGACGTGGGGAAGTACT
MYC control	AAAAACGGGGTCAGAAGTCAGGAA	AGGTAAAGATTGGGGAAGCAGCAA
16q22	CTACTCACTTATCCATCCAG	ATTTCACACACTCAGACATC

GCTAC	ACAG

3.5.3.3. Genotyping

	forward	***reverse***
LARGE2 region		
Guide 1	GATTCCTCCTGTTCGGTGGG	GCGTGGAAAAGCTCTTTGCA
Guide 2	GGACATCTGCGGGAATCACA	GCGTGGAAAAGCTCTTTGCA
APC region		
Exon 7	CGACCGCCAATCGTACTGGAGG	GACAGCACATTGGTACTGAATGC
Exon 15, MCR	TGTAATCAGACGACACAGGAAGCAGA	GGGGACCACTTTGTACAAGAAAGCTGGGTG

3.5.3.4. In situ hybridization

	forward	***reverse***
Large2 probe 1	CTCTATGGACCGGCTGCAGATGCTG	GAGAATTAGACTAGCCCCAAACCAGC
Large2 probe 2	GTGTGATCCTGCTGTGGCTGGACAG	AGTCTGTGGATGAGTGACCCTGTGG
LARGE2 probe 1	GCCCTGTGCAGGCACTGGCCTG	AAGGGATAAACAATTCAAGATGG
LARGE2 probe 2	ACCCTGGCCGGCGGCTGCGAGC	CCTTGATCACAGCGTTGAAGATG

3.5.3.5. Cloning and PCR

	forward	***reverse***
CTNNB1 S33Y	AAAAAGCAGGCTTCACCATGGCTACTCAAGCTGATTTGATG	AGAAAGCTGGGTGTTACAGGTCAGTATCAAACCAGGC
attB universal primer	GGGGACAATTTGTACAAAAAAGCAGGCTTCAC	GGGGACCACTTTGTACAAGAAAGCTGGGTG
LARGE2 ORF M13	TGTAAAACGACGGCCAGT	CATGGTCATAGCTGTTTCCTG
LARGE2 gRNA1	CACCGGCCCCCGGGTCTCCGTCGA	AAACTCGACGGAGACCCGGGGGCC

LARGE2 gRNA2	CACCGAAGAGCATGGACTTCACCA	AAACTGGTGAAGTCCATGCTCTTC
LARGE2 Promotor	GCACAATCACGACGATCTG	CTCGTCCTGGTGGGAAACTC
TCF7L2	GCCGAATTCGCACAATCACGACGATCTG	GGGCTCAGGTTTCACCTCAAGATCTCCG
TCF7L2mut	GTGTCGAGCTCCCGCAGTGTCTTTGGCCCCGGGCCCGGGGACACTGG	GAGGGAGGAGGAGGGGGCCGGAATGCAGG

3.6. Buffers and solutions

Laminin Binding Buffer

- 140mM NaCl
- 1mM $MgCl_2$
- 1mM $CaCl_2$
- 19mM TEA
- pH 7.4

TBS-T (1L)

- 24.2g Tris, pH 7.6
- 80g NaCl
- 0.1% Tween20

Running Buffer (10x, 1L)

- 30g Tris
- 140g Glycine
- 100mL 10% SDS

Transfer Buffer (10x, 1L)

- 30g Tris
- 140g Glycine

WGA Buffer

- 50 mM Tris/HCl, pH 7.4
- 150 mM NaCl
- 1% Triton X-100

Vogelstein PCR Buffer (VB, 10x) (Vogelstein and Kinzler, 1999)

- 166mM NH4SO4
- 670 mM Tris, pH 8.8
- 67 mM MgCl2
- 100 mM β-ME

Vogelstein PCR Mix for colony PCR

- 23,25 µl ddH2O
- 2,5 µl 10x VB
- 1,5 µl dNTP Mix (10mM each)
- 1 µl fw primer
- 1 µl rv primer
- 1,5 µl DMSO
- 0,25 µl Taq Polymerase

2x Laemmli Buffer

- 125 mM Tris, pH 6.8
- 4% SDS
- 20% Glycerol
- 0.05% Bromphenol blue
- 2% DTT

10x Gitschier`s Buffer (10x GB):

- 1675 µl 2M Tris, pH 8.8
- 830 µl 1M $(NH4)_2SO4$
- 670 µl 0.5M $MgCl_2$
- 1825 µl ddH_2O

Cell lysis Buffer

- 5040 µl ddH_2O
- 600 µl 10x GB
- 300 µl 10% Triton X-100
- 60 µl β-ME

In-situ hybridization Buffers

Acetic anhydride solution

- 1.5% triethanolamine
- 0.15% HCl
- 0.6% acetic anhydride

Pre-hybridization solution

- 50% Formamide
- 5X SSC pH 4.5
- 2% Blocking Reagent
- 0.05% CHAPS
- 5mM EDTA
- 50 µg/ml Heparin
- 50 µg/ml yeast RNA

NTM Buffer

- 100mM Tris. pH 9.5
- 100mM NaCl
- 50mM $MgCl_2$

20x SSC

- 175.3 g NaCl
- 88.2 g Sodium citrate

WREN medium

- 50% Conditioned medium (Wnt3a, R-Spo3, Noggin)
- 50% ADF
- 10 mM Glutamax
- 10 mM HEPES
- 1x N2
- 1x B-27
- 1 mM N-Acetylcysteine
- 50 ng/mL EGF
- 7.5 µM p38 inhibitor (SB202190)
- 0.5 µM TGFβ inhibitor (LY2157299)
- 50 µg/ml Normocin
- 10 µM Rho kinase inhibitor (Y27632)

EN medium

- ADF
- 10 mM Glutamax
- 10 mM HEPES
- 1x N2
- 1x B-27
- 1 mM N-Acetylcysteine
- 50 ng/mL EGF
- 25 ng/ml Noggin
- 0.5 µM TGFβ inhibitor (LY2157299)
- 50 µg/ml Normocin
- 10 µM Rho kinase inhibitor (Y27632)

Tumor organoid medium

- ADF
- 10 mM Glutamax
- 10 mM HEPES
- 1x B-27
- 1 mM N-Acetylcysteine
- 50 ng/mL EGF
- 0.015 µM Prostaglandine E2
- 25 ng/ml Noggin
- 7.5 µM p38 inhibitor (SB202190)
- 0.5 µM TGFβ inhibitor (LY2157299)
- 50 µg/ml Normocin
- 10 µM Rho kinase inhibitor (Y27632)

3.7. Laboratory equipment

Device	*Supplier*
AxioPlan 2 Microscope System	Carl Zeiss, Oberkochen, Germany
BD FACS Aria™ III	Becton Dickinson, Franklin Lakes, NJ, USA
BD LSRFortessa™	Becton Dickinson, Franklin Lakes, NJ, USA
Berthold Orion II Microplate Luminometer	Titertek-Berthold, Pfortzheim, Germany
Binder Incubator	Binder, Tuttlingen, Germany
BioPhotometer Plus	Eppendorf, Hamburg, Germany
Cell culture flasks, Multiwell plates, Canonical tubes	Greiner Bio One, Frickenhausen, Germany
GeneAmp® PCR System 9700	Applied Biosystems, Foster City, CA, USA
Heraeus Megafuge 1.0R	ThermoFisher Scientific, Waltham, MA, USA
HTU Soni130 Sonicator	G. Heinemann, Schwäbisch Gmünd, Germany
Li-Cor Odyssey Fc	Li-Cor, Lincoln, NE, USA
LightCycler480	Roche, Penzberg, Germany
LSM 700 Laser Scanning Confocal Microscope	Carl Zeiss, Oberkochen, Germany
Mini Trans-Blot™ Cell	Bio-Rad Laboratories, Munich, Germany
Mini-PROTEAN Tetra Vertical Electroporesis Cell	Bio-Rad Laboratories, Munich, Germany
NanoLC UltiMate 3000 / quadrupole-Orbitrap QExactive HF mass spectormeter	Thermo Fisher Scientific, Waltham, MA, USA
ND 1000 NanoDrop Spectrophotometer	NanoDrop Products, Wilmington, DE, USA
Neubauer counting chamber	Carl Roth, Karlsruhe, Germany
Qubit 3.0 Fluorometer	ThermoFisher Scientific, Waltham, MA, USA
T100™ Thermal Cycler	Bio-Rad Laboratories, Munich, Germany
VectraPolaris	Perkin Elmer, Waltham, MA, USA
Shake'n' Stack Hybridisierungsofen	Hybaid limited, Ashford, UK

4. Methods

4.1. Generation of DNA constructs

For CRISPR/Cas9-mediated gene knockouts, guide RNAs were designed using the CRISPR design tool crispr.mit.edu (Zhang Lab, MIT 2013). Creation of an expression plasmid for LARGE2 gRNAs was done as described by Ran et al. (Ran *et al.*, 2013). Oligonucleotides encoding the 20-nt guide RNA (listed in 3.5.3.5) were phosphorylated and annealed. The lentiviral eCas9 expressing vector pLentiCRISPR-E and pSpCas9(BB)-2A-GFP vector were linearized by digestion with Esp3I. Oligo pairs and vector backbone were ligated using quick ligase (NEB) and transformed into E. coli Stbl3 bacteria. Insertions were verified by sequencing (Eurofins Genomics, Ebersberg, Germany).

For stable, doxycycline-inducible overexpression of genes, initially an empty lentiviral pTZgw vector was created. The pTRIPZ vector was digested with EcoRI-HF and AgeI-HF (NEB) to remove turboRFP followed by de-phosphorylation via Fast-AP (ThermoFisher). A Gateway cassette was cut out from pLenti CMV-Puro-Dest (Addgene #17452) by restriction with MfeI and SgrAI (NEB) and ligated into the pTRIPZ backbone, creating pTZ gw. The respective gene ORFs were either amplified via Phusion® High Fidelity Polymerase (NEB) from human cDNA (Primer listed in 3.5.3.5), creating amplicons with attB overhangs, which were first cloned into pDONR221 with Gateway® BP Clonase ® II (Invitrogen). Insertions were verified by sequencing (M13 fw/rv Primer, Eurofins Genomics, Ebersberg, Germany). Verified inserts were relocated to pTZgw via Gateway® LR Clonase ® II (Invitrogen) and standard procedures. To create pTZgw LARGE2, LARGE2-ORF plasmid was used as entry-vector for the final LR reaction according to the manufacturer's protocol. For induction, transduced cells were treated with 500 ng/mL Doxycycline.

To silence APC and activate Wnt-signaling, a set of three commercially available pTRIPZ lentiviral shRNAs targeting APC (RHS4740-EG324) and a non-silencing shRNA control (RHS4743) were purchased from Dharmacon.

To create the pV2Luc2 reporter construct, harboring the TCF7L2 binding site within the first intron of LARGE2, the appropriate *LARGE2* amplicon was amplified via Phusion® High Fidelity Polymerase (NEB) and the "TCF7L2" primer listed in

3.5.3.5 from SW480 genomic DNA. The pV2Luc2 backbone and amplicons were digested with EcoRI-HF and HindIII-HF (NEB), quick-ligated (NEB) to result in pV2Luc2-TCF7L2, which was transformed in Stbl3 bacteria. Positive clones were verified via colony-PCR in Vogelstein PCR Buffer (VB, see 3.6) and primer "LARGE2 Promotor" listed in 3.5.3.5.

To create pV2Luc2-TCF7L2mut, the NEB Q5® site directed mutagenesis Kit was used in combination with Phusion® High Fidelity Polymerase (NEB) to amplify from pV2Luc2-TCF7L2, using the TCF7L2mut primer from 3.5.3.5. The PCR product was afterwards processed as described by the manufacturer and transformed in NEB *E.coli* 5α. Positive clines were verified by PCR.

4.2. Bacterial cell culture

Bacterial transformation was done in the *E. coli* Stbl3 strain, except for site directed mutagenesis, for which 5α cells were used. Plasmid DNA was added to competent cells and incubated on ice for 30 minutes. After heat-shock for 45 seconds at 42°C, cells were placed on ice for 2 minutes and 500µl LB-medium without antibiotics added. After 1 hour incubation at 37°C, 200 µl of the cell suspension was plated on LB-plates, supplemented with the appropriate antibiotics.

To multiply transformed bacteria, LB-medium cultures were inoculated with single colonies and proper antibiotics. After incubation overnight at 37°C, cell suspensions were purified with either NucleoSpin Miniprep Kit or PureLink Maxiprep Kit, dependent on the culture volume.

4.3. Mammalian cell culture

4.3.1. 2D human cell lines

HT-29, RKO, SW480, SW620, LS174T, Colo205, and HEK293T were obtained from ATCC, HMEC-1 were a gift from Professor T. Grünewald (Institute of pathology, LMU, Germany), KM12 and KM12L4a were a gift from Professor I. Fidler (MD Anderson, TX, USA). Cells were cultured in DMEM, supplemented with 10% FBS, 100 U/mL Penicillin and 0.1 mg/mL Streptomycin (P/S). Exceptions are

SW480 cells, which were cultivated in RPMI-1640 medium (10% FBS, 1% P/S), HMEC-1 (cultivated in RPMI-1640 with 10% FBS, 1% P/S, 50 µg/µl EGF, 1 µg/mL Hydrocortisone) and RKO cells, cultivated in McCoys5A with 10% FBS, 1% P/S. Cells were kept in a humidified incubator (37°C, 5% CO_2) and subcultured as needed.

All cell lines were tested negative for Mycoplasma contamination by using the LookOut Mycoplasma PCR detection kit. For cryo-preservation, cells were resuspended in DMEM supplemented with 40% FBS and 10% DMSO and slowly cooled to -80°C. For long-term storage, cells were transferred to liquid nitrogen.

4.3.2. 3D culture

4.3.2.1. Organoid culture from human colonic mucosa

For isolation of crypts from patient material and 3D cultivation of colonic mucosal tissue, small tissue pieces were first incubated in 10 mM DTT for 5 min at room temperature, then transferred to 8 mM EDTA for 60 min at 4°C, and then vigorous shaken to enrich colonic crypts in the supernatant. 200-300 Crypts were directly cultured in 50 µl Matrigel and plated as drops in 24-well culture dishes. Embedded cells were overlaid with WREN medium, which was replaced with fresh medium every other day. For serial passaging, organoids were disaggregated in TripLE select for 5-10 min at 37°C, washed twice with ADF medium, and 10.000 cells per 50 µl Matrigel drop re-seeded. Ex vivo multi-lineage differentiation of human colonic organoids was achieved by replacing the WREN medium with EN media lacking Wnt3a, RSPO-1 and SB202190 for at least 72 hours.

4.3.2.2. Organoid culture from human tumor samples

Tissue pieces from primary CRC or liver metastasis were cut into small pieces and incubated with Normocin and Antibiotic-Antimycotic for 15 min at room temperature. Next, pieces were disaggregated in ADF, 5U/ml Dispase II, 75U/ml Collagenase IV and 10 µM Y-27632 for 30 min at 37°C and subsequently passed through a 1.2 mm needle using a syringe. Cells were treated with ammonium

chloride for 3 min and then washed twice with ADF. After resuspension of the tumor cells in Matrigel, varying cell densities were plated in 24-well tissue culture plates (50 µl Matrigel per well). Embedded tumor cells were overlaid with tumor organoid medium. Medium was replaced every 2-3 days. For serial passaging, embedded Organoids were disaggregated using 0.025% Trypsin for 10 min at 37°C and subsequently passed through a 0.8 mm needle by a syringe. After washing with ADF medium, cells and small cell aggregates were counted if needed, and re-seeded. For cryo-preservation, Organoids were disaggregated in 0,025% Trypsin using a P-1000 pipette, washed twice with ADF and resuspended in ADF supplemented with 40% FBS and 10% DMSO. For long-term storage, vials were transferred to liquid nitrogen.

4.3.3. Transduction of cell lines and organoids

For production of recombinant lentiviral particles, HEK293T cells were seeded in 6-wells 24 hours prior transfection. Cells were co-transfected with total 1 µg of 2nd generation packaging vectors pMD2.G and psPAX2 as well as 1 µg of the lentiviral vector encoding for the gene or guide RNA of interest, all diluted in 150 mM NaCl and PEI. Viral supernatants were collected after 24 and 48 hours, passed through 0.45 µm filters (Millipore) and concentrated using Lenti-X Concentrator. 3 Volumes of concentrator were added and the mixture incubated for 4 hours on ice, before centrifugation at 1,500 x g for 45 minute at 4°C. Virus particles were resuspended in 500 µl PBS to receive a 20x concentrated virus and used for infection in the presence of 8 mg/mL Polybrene. Alternatively, virus-containing supernatants were passed through 0.45 µm filters (Millipore) and directly used to infect cell lines in presence of 8 mg/mL Polybrene. After 16 hours the virus was removed and the next day cells were selected with 2 µg/mL Puromycin for 7 days.

To transduce tumor organoids, single cells in suspension were seeded on Matrigel-coated surfaces 4 hours prior infection.

4.3.4. Transfection of oligonucleotides and vector constructs

For transfection of cell lines with Oligonucleotides or plasmids, cells were seeded

24 hours prior transfection at 50% confluency. For 6-well format, the transfection reagent mix contained 4 µg DNA diluted in 500 µl Opti-MEM and 4 µl Plus reagent. After 10 minutes incubation at RT, 25 µl Lipofectamine LTX reagent were added and incubated for another 25 minutes before drop-wise application to the cells.

4.4. Luciferase Assay

For luciferase assays, 1 * 10^5 cells were seeded in 24-well plates 24 hours prior to transfection with 200 ng of the appropriate pV2Luc2 vector and 200 ng Renilla Luciferase as normalization control using Lipofectamine LTX. After 48 hours cells were lysed in 1x Lysis-Juice and 10 µl of lysate were transferred to a 96-well white flat bottom plate, mixed with 50 µl Beetle-Juice and measured on a Berthold Orion II Microplate Luminometer. To measure the transfection control, 50 µl of Renilla-Juice were added and measured the same way (Beetle-Juice Luciferase Assay Firefly; Renilla-Juice Luciferase Assay, both pjk Biotech). Cells were transfected three times and each sample was measured three times to exclude technical errors. Fluorescence intensities were analyzed using the SIMPLICITY software package (DLR).

4.5. Mutation detection

To detect inserted mutations after CRISPR/Cas9 editing, genomic DNA was isolated using GenElute Mammalian Genomic kit. Primers for PCR amplification using the Q5 polymerase Master Mix are shown in 3.5.3.3. Amplicons were denatured, reannealed and digested with T7E1 enzyme as specified in the Alt-R® Genome Editing Detection Kit protocol. Digested fragments were then visualized on a 2% agarose gel.

4.6. Chromatin-Immunoprecipitation (ChIP)

For analysis of chromatin and transcription factor TCF7L2/TCF4 binding sites, 4.5 * 10^6 SW480 cells were used per immunoprecipitation according to the

SimpleChIP® Enzymatic Chromatin IP Kit protocol. Cells were cross-linked with 37% Formaldehyde, harvested and cell nuclei first digested with Micrococcal Nuclease, followed by sonication on ice to 150-900 bp fragment length. Total 10 µg of fragmented DNA were incubated over night with either TCF7L2, control rabbit IgG or Histone H3 antibody. The precipitated chromatin fragments were treated with Proteinase K for 2 hours to remove the crosslinking, adjacent purified and analyzed via quantitative real time PCR, using primer listed in 3.5.3.2.

4.7. Single cell sorting and genotyping

48 hours post transfection with the plasmid of choice, cells were harvested in 10mM EDTA in PBS, washed twice with PBS and resuspended in 1 mL PBS supplemented with 5% FBS. 0.3 µg/ml DAPI were used to exclude dead cells during cell sorting. Single GFP-positive cells were sorted on a BD Aria III instrument in 96-well plates, filled with FBS supplemented DMEM and Normocin.

To screen for positive clones, confluent cells were harvested in 50 µl cell lysis buffer, supplemented with 5 µl Proteinase K, incubated 1 h at 55°C and subsequently boiled 20 min at 95°C. 5 µl of the cell lysate was directly used as template for genotyping via PCR (Phusion® High Fidelity Polymerase (NEB), (98°C, 10’’; [33 cycles: 98°C, 10’’; 66°C, 30’’; 72°C, 15’’] 72°C, 10’; 16°C, hold).

4.8. Immunolabeling of human cells for FACS analysis

For analysis of surface α-DG glycosylation, CRC cells were washed with PBS and incubated with 8 mM ETDA/PBS for 10 min on ice. Detached cells were washed in PBS and re-suspended in staining buffer (5% FBS in PBS). 500.000 cells were stained with anti- α-DG IIh6c4 antibody for 1 hour, washed with staining buffer, followed by labeling with a FITC-coupled secondary antibody for 45 minutes. Control cells were stained with the secondary antibody alone. All incubation steps were carried out on ice. Prior to FACS analysis, 0,3 µg/ml DAPI was added to exclude dead cells. Analysis was performed on a LSR Fortessa instrument.

For labeling and sorting of PDO cells, organoids were disaggregated using Dispase II, and resuspended in ADF supplemented with 10 mM HEPES, 10 mM

Glutamax, 5% FBS and 10 µM Y-27632. 250.000 cells were stained with an APC coupled anti-PTK7 antibody or EphB2 and sorted into FACS buffer. The brightest 14-15% PTK7+/EphB2+ organoid cells were defined as “high” fraction. Next 25-30% of cells were sorted as “low” population. The PTK7 or EphB2 negative population did not stain for PTK7. Sorting was performed on an Aria III instrument. Sorted cells were centrifuged for 10 min at 200*g at 4°C and re-suspended in 500 µl Trizol for RNA isolation.

4.9. RNA Isolation and transcription

RNA isolation from cells for qPCR was prepared using High Pure RNA Isolation Kit according to the manufacturer protocols. RNA was transcribed using High-Capacity cDNA Reverse Transcription Kit. qPCR was performed using the Fast SYBR Green Master Mix on a LightCycler480. Relative expression values were normalized to *PPIA* and *B2M* expression and calculated using the ΔΔCt method. Primer used for qPCR are listed in 3.5.3.1.

RNA from cell sorting (see section 4.8) was extracted via Trizol/Chloroform and purified using the RNeasy® MiniKit. RNA concentration was determined via Qubit RNA HS Assay on a Qubit 3.0 Fluorometer and qPCR performed on cDNA using TaqMan probes (3.5.2) and the primaQUANT qPCR Master Mix on the LightCycler480.

4.10. Protein analysis

4.10.1. Preparation of protein lysates

For protein analysis, Organoids were harvested in 1 mL cell recovery solution per 50 µl Matrigel and incubated 30 min on ice. After washing the cell pellet twice with cold PBS, cells were resuspended in WGA Buffer supplemented with Protease Inhibitors. 2D cells were directly harvested in WGA Buffer and Protease inhibitors. Lysates were incubated on ice for 15 min, then sonicated 5x 5 seconds with 50% amplitude and spun 15 min at 16.000*g. Protein concentration of supernatant was determined by Bradford Reagent and measurement of OD_{595}.

4.10.2. Precipitation of Glycoproteins

Prior incubation, WGA beads were equilibrated by washing three times with WGA Buffer without inhibitors. For precipitation of glycoproteins, 1-3 mg of whole cell lysate were incubated with 50 µl WGA agarose beads overnight at 4°C while rotating. For elution of proteins, Beads were first washed three times with WGA buffer to remove unbound proteins and glycoproteins were eluted by boiling 10 min at 70°C in 60 µl of 2x Laemmli buffer with 100 mM DTT.

4.10.3. Western Blot analysis

For western blotting, whole cell or glycoprotein-enriched lysates were boiled for 10 min at 95°C in 2x Laemmli buffer, quick spun and analyzed on 8-10% SDS-Acrylamide Gels via electroporation in 1x Running Buffer. Proteins were transferred on Immobilon PVDF membranes by using the Mini-Protean-electrophoresis system (Bio-Rad) in 1x Transfer Buffer overnight at 30 V and 4°C. After the transfer, membranes were incubated 1 h at room temperature in 5% skim milk-TBST to block unspecific antibody binding. Afterwards the membrane was incubated with a primary antibody (listed in section 3.4) overnight at 4°C. The next day, the membrane was washed three times for 15 min in TBST, followed by secondary antibody incubation for 1 hour at room temperature. Next, the membrane was washed three times for 15 min in TBST and imaged using Immobilon ECL Ultra solution for detection of HRP signals with a CCD camera (Li-Cor Odyssey Fc). For detection of α-DG signals, SuperSignal west femto substrate was used exclusively.

4.10.4. Laminin Overlay Assay

For detection of laminin binding proteins, PVDF membranes were blocked in 5% skim milk in laminin binding buffer (LBB) for 1 hour at room temperature. After washing in 3% BSA/LBB, membranes were incubated with 5 µg/mL Matrigel in 3%

BSA/LBB at 4°C overnight. Unbound residual Matrigel was washed off the membrane with 5% milk in LBB, followed by incubation in anti-laminin antibody. HRP signals from secondary antibodies were detected using Immobilon ECL Ultra Western HRP Substrate and a CCD camera (Li-Cor Odyssey Fc).

4.10.5. LC-MS/MS

Glycoprotein enriched lysates were prepared as described above (4.10.2). WGA bound glycoproteins were eluted with LDS sample buffer supplemented with 100 mM DTT for 10 min at 70°C, followed by alkylation with 55 mM CAA for 30 min at RT. Eluted glyco-proteins were purified and separated on 4-12% NuPAGE Bis-Tris Gels and subsequently split in six fractions for in-gel tryptic digestion according to standard procedures. After drying in a centrifugal evaporator, the samples were stored at -20 °C until LC-MS/MS analysis.

Samples were analyzed in triplicates (3 samples per condition) by S. Wiechmann and B. Küster (CIPSM, Technical University Munich).

Analysis of peptides from eluates was performed via nano-flow LC-MS/MS measurement using a nanoLC UltiMate 3000, coupled to a quadrupole-Orbitrap Q Exactive HF mass spectrometer (both ThermoFisher Scientific). Peptides were desalted on a trap column (100 µm × 2 cm, packed in-house with Reprosil-Pur C18-AQ 5 µm resin; Dr. Maisch) in 0.1 % FA at 5 µl/min and separated on an analytical column (75 µm × 40 cm, packed in-house with Reprosil-Pur C18-AQ, 3 µm resin; Dr. Maisch), using a 22 min linear gradient from 4-32 % solvent B (0.1 % formic acid, 5 % DMSO in acetonitrile) in solvent A (0. 1% formic acid, 5 % DMSO in water) at a flow rate of 300 nL/min. The mass spectrometer was operated in data dependent acquisition and positive ionization mode. MS1 spectra were acquired over a range of 360-1300 m/z at a resolution of 60,000 in the Orbitrap, by applying an AGC of 3e6 or maximum injection time of 10 ms. Up to 10 peptide precursors were selected for fragmentation by higher energy collision-induced dissociation (HCD; 1.7 m/z isolation window, AGC value of 1e5, maximum injection time of 25 ms) using 25 % normalized collision energy (NCE) and analyzed at a resolution of 15,000 in the Orbitrap.

4.10.6. Peptide quantification

Identification and qualification of peptides and proteins was performed using MaxQuant (Cox and Mann, 2008) (version 1.6.0.16), by searching the tandem MS dataset against all human protein sequences annotated in the Swissprot database (42356 entries, 06.06.2018), using the embedded search engine Andromeda (Cox *et al.*, 2011).

Carbamidomethylated cysteine was defined as fixed modification and oxidation of methionine and as variable modification N-terminal protein acetylation. As proteolytic enzyme Trypsin/P was specified and up to two missed cleavage sites were set as allowed.

Precursor and fragment ion tolerance were set to 4.5 ppm and 20 ppm. Minimal peptide length was defined as seven and all data was adjusted to 1% PSM and 1% protein FDR. Intensitiy-based absolute quantification (iBAQ) (Schwanhäusser *et al.*, 2011) assisted within MaxQuant. To filter out contaminants, reverse hits and protein groups which were only defined by one site, the Perseus software (Tyanova *et al.*, 2016) (version 1.6.2.3) was used. Additionally, only protein groups detected in two out of three replicates were considered for the analysis.

Datasets produced in this experiment are available in the PRIDE database using the accession number **PXD013800**

4.11. Migration Analysis

To analyze cell invasion, cells were deprived of serum for 24 hours before seeding. Cell inserts (0.8 µm pore size) were coated with Laminin-111 (1 µg/mL Laminin-111 from EHS lathrytic mouse tumor) and 10^5 cells seeded in serum free DMEM into the upper chamber.

In case of transmigration of cells through endothelium, 48 hours before cancer cell seeding, 10^5 HMEC-1 cells were seeded on non-coated transwells to build a monolayer. 24 hours later, 50 ng/mL TNFα was added, and endothelial cells were activated for another 24 hours. Then 10^5 cancer cells were seeded in serum free DMEM as described above.

As chemoattractant, DMEM, supplemented with 10% FBS, was placed in the

bottom chamber. After 24 to 48 hours, non-migrated cells from the upper side of the chamber were removed with a cotton swab. Migrated cells on the bottom side of the insert membrane were fixated using 4% PFA for 20 minutes, air dried and stained with 100 ng/mL DAPI. Pictures were taken with a Zeiss LSM 700 confocal microscope in five random fields per membrane. Cells were counted with ImageJ software. Each experiment was performed three times in duplicates, therewith the relative migration was normalized to the equivalent control cells.

4.12. Cell adhesion Analysis

For analysis of cell attachment on an endothelial layer, $2 * 10^5$ HMEC-1 cells were seeded in 200 µl supplemented RPMI medium per 48-well. 48 hours later, cancer cells were detached with Accutase, counted and a dilution of $0.75 * 10^6$ cells/mL in full DMEM prepared. 200 µl of the cancer cell suspension was added to the confluent endothelial layer to attach for the indicated time. Attachment was stopped by removing the supernatant and a double washing step in PBS to remove non-attached cells.

Cells were lysed in 50 µl 1x Passive Lysis Buffer (Promega, diluted in PBS) for 20 minutes at room temperature, with additional up and down pipetting to increase cell lysis. 10 µl of lysate were transferred as technical replicates to a 96-well white flat bottom plate and 50 µl of Beetle Juice (pjk Biotech) added. Read-out was performed immediately afterwards with Berthold Orion II Microplate Luminometer, using SIMPLICITY software package (DLR).

4.13. In situ hybridization

For detection of *large2* in mouse ileum normal mucosa and ApcMin-driven adenoma, RNA probes were amplified with DIG RNA Labeling Kit (Roche) from the pENTR221 Gateway full ORF clone according to the manufacturers' protocol. The RNA probe was purified using RNeasy Mini Kit (Qiagen). FFPE sections (5 µm) of mouse tissues were de-waxed and re-hydrated using standard procedures. Proteins were digested by treating the slides 15 min at room temperature with 0.2 N HCl, and cells permeabilized with 30 µg/mL Proteinase K in PBS at 37°C. The reaction was

stopped by rinsing the slides in 0.2% Glycine/PBS and washing twice in PBS. Sections were then post-fixed for 10 min in 4% PFA, washed three times in PBS and histone acetylation was performed with acetic anhydride solution. Next, slides were incubated for 1 hour at 65°C in pre-hybridization solution. Per labeling, 500 ng of RNA probe was diluted in pre-hybridization solution and incubated for 24 hours at 65°C. Post-hybridization washes were performed using 50% Formamide in 2x SSC Buffer pH 4.5, three times for 20 minutes at 65°C. In the following, sections were rinsed in Tris/NaCl buffer and blocked for 30 min at room temperature in blocking solution (0.5% blocking powder in Tris/NaCl). Sheep anti-DIG antibody (Roche) was diluted 1:2000 in blocking solution and incubated overnight at 4°C. Finally, samples were first washed in Tris/NaCl buffer and then in NTM buffer before developing the sections in BCIP/NBT Liquid Substrate for 24-72 hours.

4.14. RNAscope®

Detection of *LARGE2* and *LGR5* in human mucosa and adenoma tissue was performed via RNAscope® technology, as described in the manufacturer's protocol. In brief, tissue slides were dewaxed using standard procedures. Slides were treated with hydrogen peroxide for 10 minutes, washed in water, followed by boiling the slides for 10 seconds in water and 15 minutes in retrieval solution. Slides were rinsed with water, incubated for 3 minutes in 100% EtOH and then air-dried. Prior labelling, slides were incubated with protease plus for 30 minutes at 40°C, washed and then incubated with the RNA probe of interest in a humidified chamber for 2 hours at 40°C. Labelling was amplified with AMP chemicals 1-6, incubating 15-30 minutes per amplifier. Incubation with 1:1 mixed DAB-A and DAB-B for 10 minutes detected the signals of the RNA probe. After hematoxylin staining for 2 minutes, slides were washed and rinsed with ammonium water. Finally, samples were de-hydrated and mounted for imaging.

4.15. Immunohistochemistry staining

For Immunohistochemical (IHC) staining, 2 µm formalin-fixed, paraffin-embedded sections were used. Antigen retrieval was achieved with target retrieval solution

(S1699, Agilent Technologies) via microwave heating. Primary antibody IIh6 (Santa Cruz) at a concentration of 4 ug/ml (1:20 dilution) was incubated at room temperature for one hour. Biotinylated anti-mouse IgM (BA2020, Vector Labs) and streptavidin-HRP (RE 7104, Novocastra) were used for signal detection. Samples were developed via exposure to 3,3`-diaminobenzidine (DAB+, K3468, Agilent) and counterstained with hematoxylin (Vector Labs).

For IHC staining against on mouse FFPE tissue, Crystal MausBlock (Fa. DCS, ML125R015) was used to avoid secondary non-specific binding.

4.16. RNA Sequencing (Next generation sequencing)

RNA from HT-29 cells was isolated using the PureLink RNA Kit with DNase (both ThermoFisher) as described in the manufacturer's protocol. Quality of the isolated RNA as determined via Agilent BioAnalyzer 2100.

Preparation of Sequencing libraries and Sequencing was performed in Heidelberg, Germany at the DKFZ Genomics and Proteomics Core Facility.

Sequencing libraries were prepared using the TrueSeq Stranded mRNA Library Prep Kit for Illumina (New England Biolabs) according to the manufacturer's instructions and 50bp single-read sequencing was performed on the Illumina HiSeq 2000 v4.

For bioinformatic analysis, bcl2fastq 2.19.0.316 was used for base calling, low quality bases were removed with Fastq_quality_filter from the FASTX Toolkit 0.0.13 with 90 percent of the reads needing a quality phred score > 20. Homertools 4.7(Heinz *et al.*, 2010) was used for PolyA-tail trimming, and reads with a length < 17 were removed. Mapping of filtered reads to the human genome GRCh38 was done via TOPHAT2 (Kim *et al.*, 2013) and PicardTools 1.78 CollectRNASeqMetrics (https://broadinstitute.githu.io/picard/). Annotations and count data were generated by htseq-count (Anders, Pyl and Huber, 2015) using the gencode.v29.annotation.gtf (https://www.gencodegenes.org/) file. Differential gene expression analysis was performed with DESeq2 (Love, Huber and Anders, 2014) using default parameters. The input tables containing the replicates for the groups to compare, were created by a custom perl script, count numbers < 5 were removed.

Datasets produced in this experiment are available in the Gene Expression

Omnibus database using the accession number **GSE131575**.

4.17. Gene expression data and gene set enrichment analysis

Gene Set Enrichment Analysis (GSEA) was performed as previously described (Subramanian *et al.*, 2005), using the GSEApreRanked tool of the GSEA 3.0 software (Broad Institute, http://www.gsea-msigdb.org/gsea/index.jsp).

Pre-ranked gene lists from GDC-TCGA datasets of COAD, READ and PRAD, as well another set covering 566 CRC cases (Marisa *et al.*, 2013), were generated calculating the Pearson´s Coefficient (PCs) for each gene to determine its correlation with *LARGE2* levels. Equally, a *LARGE2* gene signature was derived from the PDX microarray data set described in Isella et al. (C Isella *et al.*, 2017) (source: GSE76402), covering 244 patients.

Samples were categorized within the CMS subtypes (Guinney *et al.*, 2015) and CRIS categories (Claudio Isella *et al.*, 2017), as described.

Gene sets of Wnt-up regulated genes were obtained from (Herbst *et al.*, 2014), PTK7+ and EPHB2+ human colonic stem cells signatures from (Jung *et al.*, 2011), intrinsic Wnt signature from (Michels *et al.*, 2019) and Wnt-β-Catenin Signaling from the Molecular Signatures database (MSigDB) (Liberzon *et al.*, 2015).

The Enrichment Score (ES) describes the degree to which a gene set S is overrepresented at the bottom or top of the ranked gene list. The normalized enrichment score (NES) can be used to compare analysis results across different gene sets. The false discovery rate (FDR) corresponding to each NES is the estimated chance that a gene set S with a given NES represents a false positive finding.

4.18. Statistics

Significant differences between two groups of biological replicates were calculated, by a two-tailed Student's t-test (unpaired, 5% FDR). Differences from qPCR data were shown as standard deviation ($n \geq 3$) or standard error ($n < 3$).

Correlation coefficients were calculated by Pearson's correlation. For the association of patient cohort subpopulations from CMS or CRIS classification, a one-way ANOVA test was applied (multiple comparison). p-values ≤ 0.5 were considered as significant; * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$, **** $p < 0.0001$, n.s. = not significant. Statistical analysis was performed via GraphPad Prism software (v7.01).

4.19. Clinical samples

Biological samples of fresh, normal and cancerous tissue specimen used for PDO and PDTO generation were received from individuals undergoing curative colectomy or partial hepatectomy at the Hospital Großhadern, LMU Munich. Samples were taken by a pathologist from residual resected tissue, which was not need for diagnostic purposes, and irreversibly anonymized. This procedure has been classified as uncritical by the ethical committee of the LMU Munich and was specifically approved for our projects (project-No. 591-16-UE and 17-771-UE).

Molecular characterization of PDTOs was done at the molecular diagnostics department at the Institute of Pathology of LMU (accredited after DIN 17020 (DAkkS)). In brief, the status of *KRAS*, *NRAS*, *BRAF*, and *PIK3CA* was examined as follows: Partial sequences of *BRAF* exon 15, *KRAS* and *NRAS* exons 2, 3, and 4, and *PIK3CA* exons 9 and 20 were amplified by PCR. Each PCR product was analyzed by established and accredited primer systems via pyrosequencing (Pyromark Q24 Advanced, Qiagen). For analysis of the microsatellite status in PDTOs, microsatellites of the NCI (National Cancer Institute) panel BAT25, BAT26, D5S346 (*APC* locus), D17S250 (*TP53* locus), and D2S123 were amplified by multiplex-PCR. PCR products were electrophoretically separated on the Genetic Analyzer (ABI 3130, ThermoFisher), and fragment length of PCR amplicons was compared between PDTO and matched normal tissue-derived DNA (fragment length polymorphism).

Anonymized colorectal cancer specimens (FFPE tissues) from patients that underwent surgical resection at the University of Munich between 1994 and 2017 (LMU, Munich, Germany) were obtained from the archives of the Institute of Pathology. Follow-up data were recorded prospectively by the Munich Cancer

Registry (data provided by J. Neumann, LMU, Munich, Germany). Specimens were anonymized, and the study was approved by the institutional ethics committee of the Medical Faculty of the LMU (18-105-UE).

For the tissue microarray (TMA) collection of locally invasive CRC, inclusion criteria were colorectal adenocarcinomas with bowel wall infiltration (T3). This collection consisted of 28 samples represented by tissue discs showing either the core tumor or the tumor invasion front.

FFPE tissue samples from the ApcMin mouse model were provided by the HH lab. Experimentation with these animals was approved by the Government of Upper Bavaria, Germany (AZ 55.2-1-54-2532-4-2014).

5. Results

5.1. Silencing of *APC* in HT-29 cells activates Wnt target gene expression

The partially functional 1555 amino acid variant of APC in HT-29 and Colo205 CRC cells encompasses three 20 AARs (3*20AAR) sufficient for partial β-Catenin inactivation, which translates to low cell-intrinsic Wnt activity (Yang *et al.*, 2006; Vijaya Chandra *et al.*, 2012). In order to stimulate an over-activation of Wnt signaling, we aimed for silencing of *APC* via a short hairpin RNA (shRNA), directed against Exon 4 (**Figure 11**).

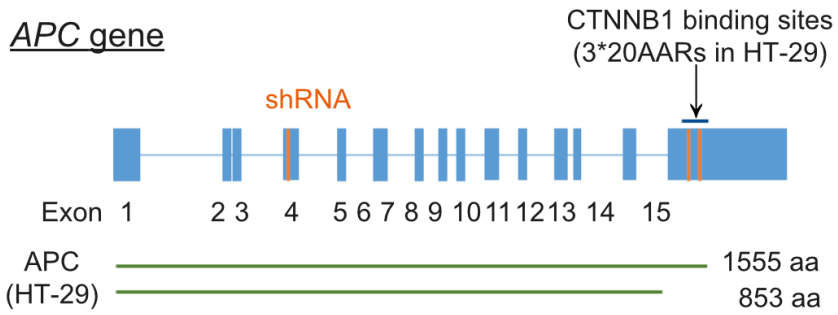


Figure 11: Scheme of the APC gene in HT-29 cells. Shown are the shRNA, targeting exon 4 and the β-Catenin binding sites within the MCR (both indicated in orange) as well as the truncated APC variants (indicated in green).

To silence *APC* in those cells, we equipped them with a doxycycline (DOX) inducible shRNA via lentiviral transduction (see 4.3.3). Upon treatment with 500 ng/mL DOX for 72 hours, we were able to reduce *APC* mRNA levels up to 16-fold (**Figure 12A**). The same treatment was sufficient to reduce the amount of APC protein in the two analyzed cell lines (**Figure 12B,C**). Consequential enhancement of Wnt-signaling led to increased expression of known Wnt target genes like *LGR5* and *AXIN2* (Jho *et al.*, 2002; Barker *et al.*, 2007).

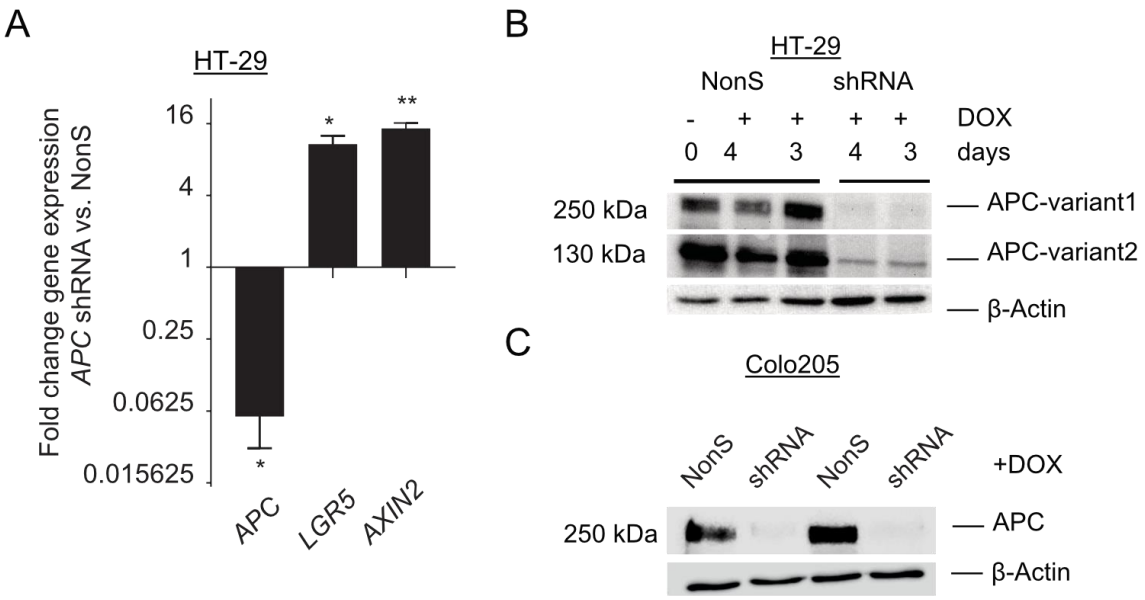


Figure 12: APC expression after silencing in HT-29 and Colo205 cells. A) qPCR analysis of indicated genes in HT-29 after 72-hour treatment with 500 ng/mL DOX. **B, C**) Immunoblot analysis of APC and β-Actin in HT-29 and Colo205 cells.

In order to reveal the genome-wide pattern of altered gene expression, de-regulated by silencing of *APC* in CRC cells, we performed RNA Sequencing of both wildtype HT-29 and HT-29 *APC* shRNA cell lines after 72 hours of treatment. Analysis of altered gene expression upon silencing of *APC* in CRC cells showed up to 205 upregulated genes, while the expression of 82 genes was downregulated (cutoff: >4-fold change, p-value <0.01) (**Figure 13**).

Besides the already known Wnt-targets *AXIN2*, *NOTUM*, *ASCL2*, *SP5* and *NKD1*, we found potentially new targets, such as *PTK7*, *TEAD2* and *LARGE2*, which were upregulated in that setting. *PTK7* was previously characterized as a surface marker of human colonic stem cells derived from primary tissue and human colonic organoids (Jung *et al.*, 2015), and our data point to *PTK7* as a putative direct or indirect target gene of the Wnt signaling pathway. *TEAD2*, a key player in the Hippo signaling pathway, promotes cell invasion by YAP1/TEAD2 regulation in hepatocellular carcinoma and is involved in EMT in breast cancer cells (Diepenbruck *et al.*, 2014; Zhou *et al.*, 2016).

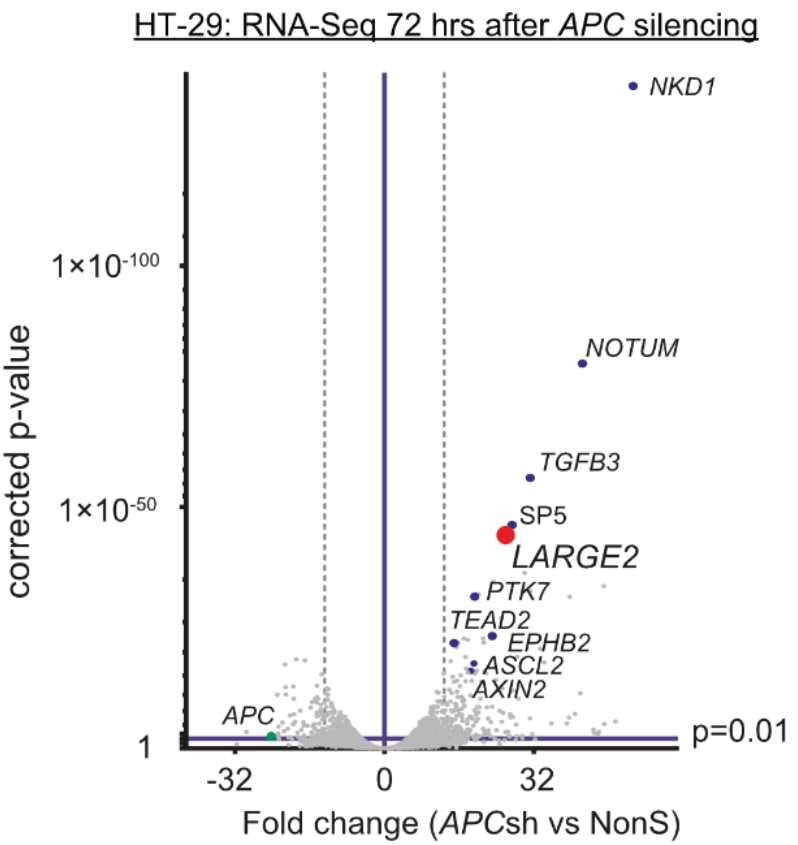


Figure 13: Volcano plot showing genes up- or downregulated upon APC silencing in HT-29 cells. Analysis via RNA-Sequencing from biological duplicates. Cut off: normalized p-value < 0.01 (blue line), Fold change > 4 (dotted grey line).

5.2. LARGE2 is a direct target gene of Wnt signaling in CRC

We focused our work on LARGE2, a bifunctional glycosyltransferase with xylosyl- and glucuronyltransferase activity (Fujimura *et al.*, 2005), since there had been no direct link between canonical Wnt signaling and O-glycosylation reported yet.

To further validate the regulation of *LARGE2* in CRC, we performed quantitative real-time PCR (qRT-PCR) upon silencing of *APC* in HT-29 and Colo205 cells. We could observe a strong increase in the levels of *LARGE2* and other well-described Wnt-target genes *LGR5*, *AXIN2*, *SP5*, *NKD1*, *NOTUM* and *WNT11*. Expression of *LARGE1*, the paralog of *LARGE2*, was very low in these cells and not affected by the shRNA-mediated knockdown of *APC* (**Figure 14A,B**).

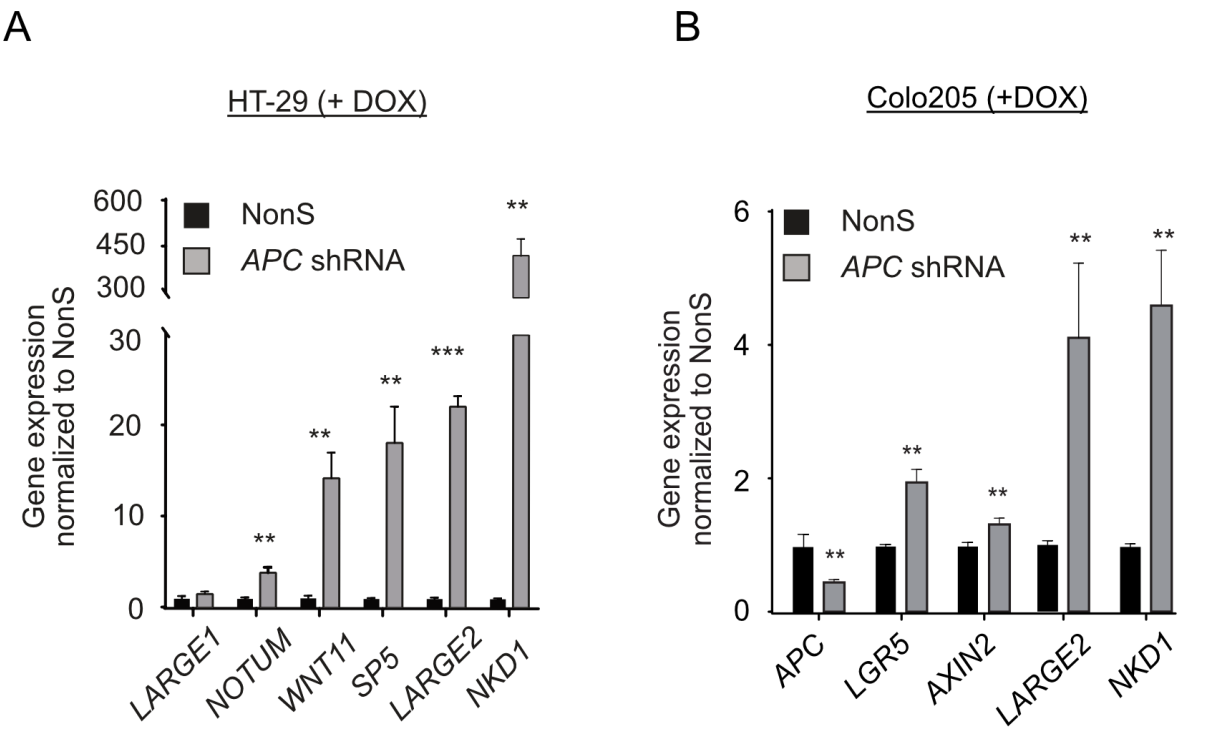


Figure 14: qRT-PCR analysis of the indicated genes upon APC silencing. A) HT-29 and **B**) Colo205 cells were treated with DOX (500ng/mL) for 72 hours. Results are shown as mean ± SD (n=3); *p<0.05, **p<0.01, *** p<0.001.

We next sought to examine whether activation of Wnt signaling apart from *APC* silencing in HT-29 cells triggers *LARGE2* expression. We stably transduced the cells with lentiviruses encoding for a DOX-inducible allele of oncogenic β-Catenin (CTNNB1-S33Y) (Kolligs *et al.*, 1999) or an empty control vector. Induction of ectopic CTNNB1-S33Y triggered the expression of *LARGE2* 8-fold higher when compared to the control cells. We also found other Wnt-targets such as *NKD1*, *AXIN2*, and *LGR5* significantly elevated after induction of oncogenic β-Catenin (**Figure 15**).

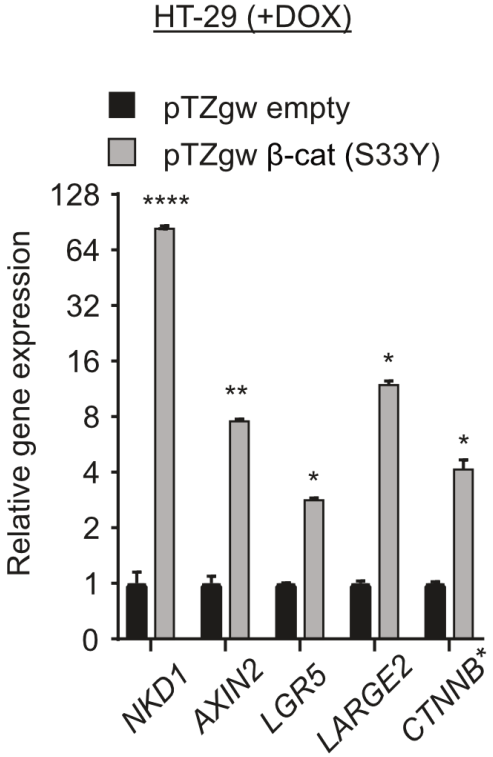


Figure 15: qRT-PCR analysis of HT-29 cells, expressing S33Y-mutated β-Catenin. Cells were treated 72 hours with 500ng/mL DOX. Displayed is mean ± SD (n=3); *p<0.05, **p<0.01, **** p<0.0001

Next, we aimed to block Wnt activity by interfering with β-Catenin/TCF functionality in CRC cells: The here used lentiviral vector system, kindly provided by Eduard Batlle (IRB Barcelona), encodes for a fusion protein of the CTNNB1-binding domain of the transcription factor TCF7L2 (nTCF: amino acids 1-90) fused to a tamoxifen-inducible version of the hormone-binding domain of the estrogen receptor ERT2, from here on referred to as "-NE". As a control, the estrogen receptor-hormone binding domain alone, referred to as "-ERT2", was used.

Cells expressing the ERT2 receptor alone experience no influence on Wnt signaling upon addition of 4-hydroxy-tamoxifen (4-OHT) to the culture media. In the cells stably expressing the nTCF-ERT2 construct, the estrogen receptor domain becomes activated upon addition of 4-OHT, and locates to the nucleus where the nTCF domain can sequester nuclear β-Catenin, thus preventing active binding of β-Catenin to endogenous TCF transcription factors and transcriptional activation of Wnt target genes (**Figure 16**).

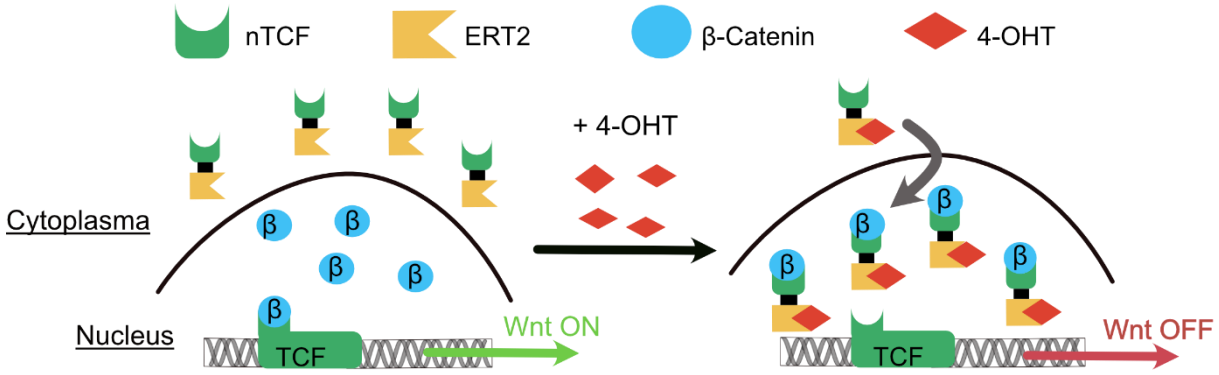


Figure 16: Schematic representation of a model to block active Wnt signaling in cells. Expression of nTCF-ERT2 fused receptors sequester nuclear β-Catenin upon activation via 4-OHT.

The CRC cell lines SW480 (microsatellite stable, mutant *APC*) and LS174T (microsatellite instable, mutant *CTNNB1*) both possess a high cell intrinsic Wnt activity. We transduced the cells with lentiviruses encoding the -NE or-ERT2 vector and treated them with 4-OHT for 12 to 48 hours before RNA isolation was performed. After 12 hours of 4-OHT treatment in LS174T-NE cells we already observed a downregulation of *LARGE2*, *LGR5, NKD1* and *AXIN2* expression when compared to their respective -ERT2 controls (**Figure 17A**), and we gained congruent results in SW480 cells after 24 hours (**Figure 17B**). These results suggest that *LARGE2* gene

expression depends on the activation state of canonical Wnt signaling in CRC cells.

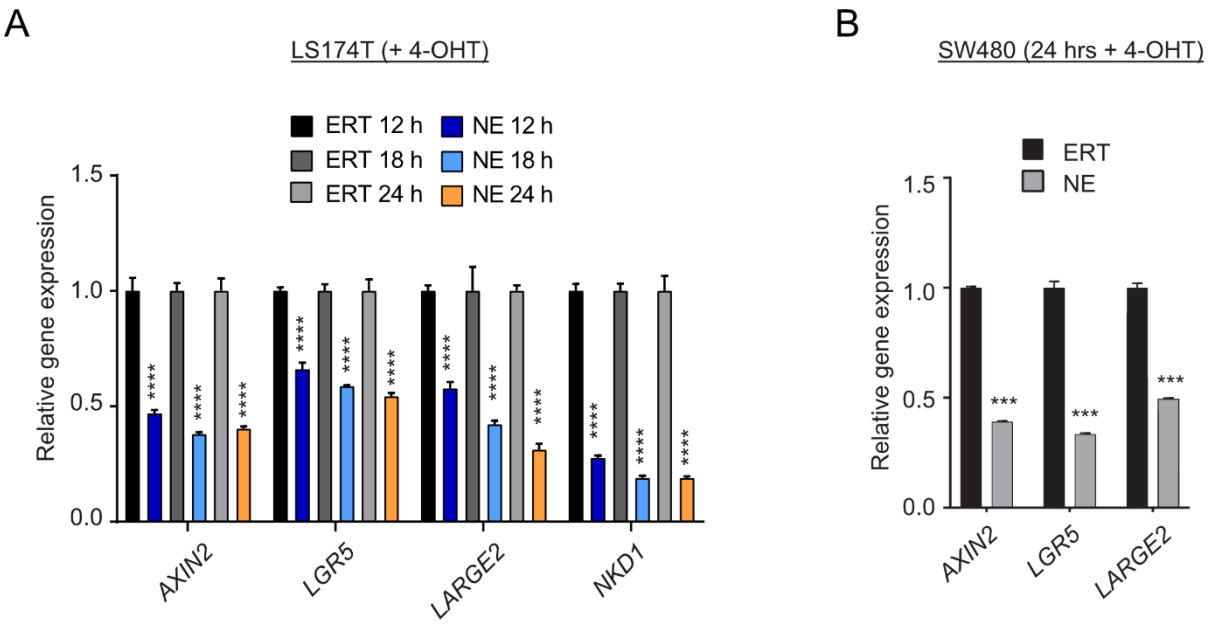


Figure 17: qRT-PCR gene expression analysis of CRC cells after blocking Wnt-signaling. Analysis of the indicated genes in **A**) LS174T cells after 12, 18 and 24 hours and **B)** in SW480 cells after 24 hrs of 4-OHT treatment. Results are compared to ERT control cells and depicted as mean ±SD (n=3); *** p<0.001, **** p<0.0001.

After that, we intended to study a possible regulation of *LARGE2* through the β-Catenin/TCF7L2 transcription factor complex. Therefore, we investigated whether the transcription factor TCF7L2 binds to the *LARGE2* genomic sequence. First, we performed an *in-silico* analysis of the *LARGE2* genomic region in order to identify putative TCF7L2 binding sites, by using publicly available ChIP-Seq data that had been generated and deposited by the Peggy Farnham laboratory in the UCSC genome browser tool (Rosenbloom *et al.*, 2013) (UCSC-ENCODE-hg19:wgEncodeEH000629). Indeed, we found the TCF7L2 binding motif "CTTTGATC" (van de Wetering *et al.*, 1997) within the first intron of *LARGE2* and another potential binding region in the proximal *PEX16* gene, ~5 kb upstream of *LARGE2* (**Figure 18A**).

To verify this data, we performed quantitative chromatin immunoprecipitation (qChIP) with an antibody directed against TCF7L2. In this assay, we succeeded to confirm binding of the Wnt-transcription factor to the *LARGE2* region of interest, whereas the second potential binding site in the adjacent *PEX16* gene did not show

any enrichment. To proof the accuracy of our ChIP experiment, we confirmed binding of TCF7L2 to its well-described recognition sites within *LGR5* and *SP5* (Hatzis *et al.*, 2008). A control region upstream of *c-MYC*, shown by others to lack any binding sites for TCF7L2 and β-Catenin (Yochum, 2011), was not enriched (**Figure 18B**).

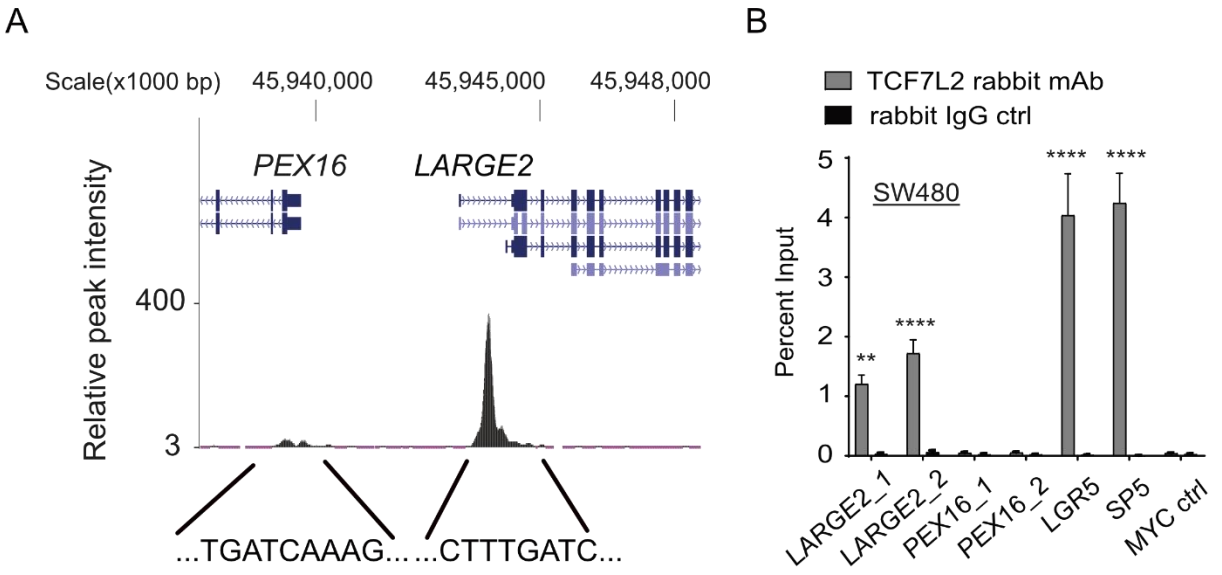


Figure 18: Analysis of TCF7L2 binding to the genomic locus of *LARGE2* and other genes. A) Representation of the LARGE2 genomic locus showing potential binding sites for TCF7L2 in HCT-116 cells. **Source**: HCT-116 TCF7L2 UC Davis ChIP-seq Signal from ENCODE/SYDH. **B**) qChIP analysis on the indicated genes in SW480 cells. The amount of DNA, in percent of the total chromatin input, immunoprecipitated with either TCF7L2 antibody or rabbit IgG control antibody. Results are shown as mean ±SD (n=3); ** p<0.01, **** p<0.0001

To provide proof for the functional importance of the confirmed TCF7L2 binding for Wnt-mediated *LARGE2* induction, we took advantage of CRISPR-Cas9 genome editing. Two target-specific designed guide RNAs (gRNAs 1 and 2), covering the TCF7L2 binding motif in *LARGE2* on either the both (+) or (–) strand, were delivered

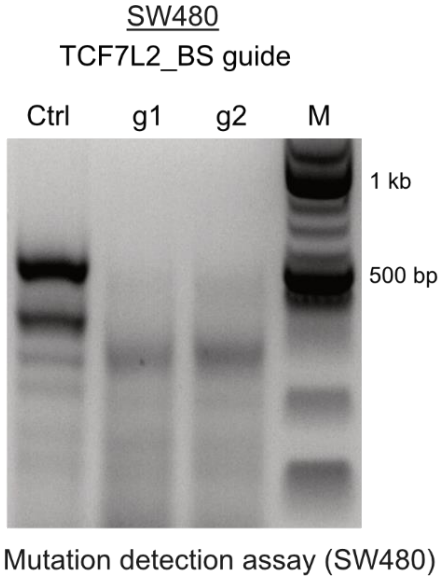


Figure 19: Visualization of the mutation detection assay on genomic DNA from stably transduced SW480 cell pools. Cells were edited via CRISPR/Cas9 and successful targeting of the TCF7L2_BS (guides: g1, g2) visualized against control (Ctrl) cells on an agarose gel.

to the cells combined with a Streptococcus pyogenes Cas9 derivative with improved specificity (eCas9) by lentiviral transduction. As control, a non-targeting tracrRNA was used and successful mutation of the targeted locus was confirmed by Alt-R Genome Editing Detection Kit (see 4.5, and **Figure 19**).

The effect of *LARGE2*-specific TCF7L2 binding site targeting was determined via qRT-PCR in SW480 cells and PDTO1. As expected, targeting of the binding site led to 4- fold (g1) to 5-fold (g2) decreased expression of *LARGE2* in SW480 cells (**Figure 20A**), whereas expression of the TCF7L2 targeted genes *NKD1* and *AXIN2* was not impaired. In PDTO1, the mRNA levels of *LARGE2* were reduced by 3-fold (g2) and 4-fold (g1), while *AXIN2* was not affected (**Figure 20B**).

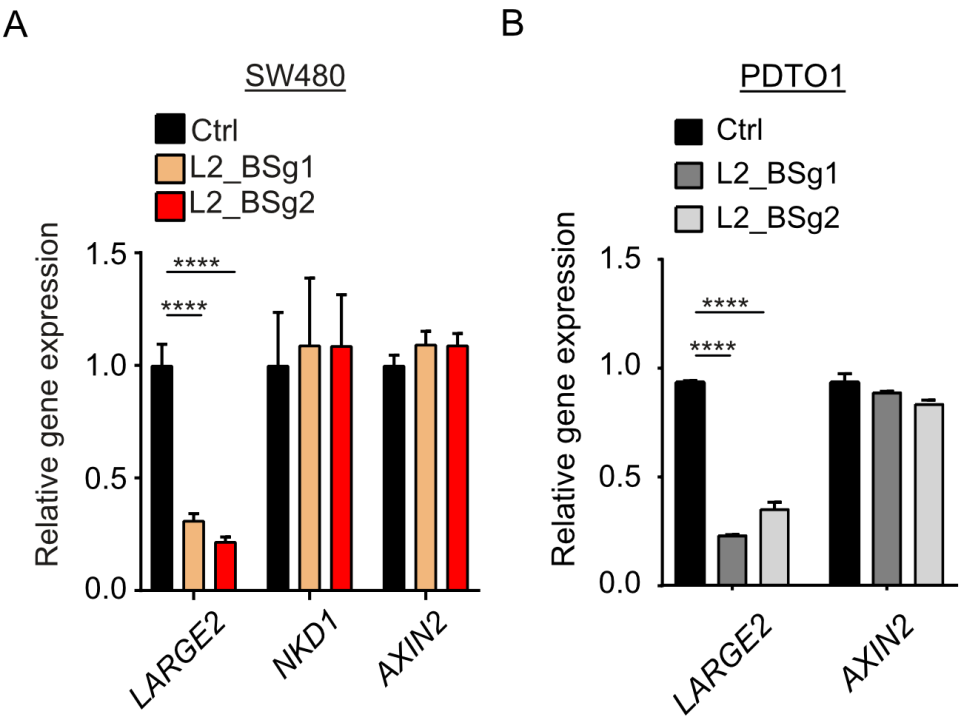


Figure 20: qRT-PCR analysis of the indicated genes upon editing the TCF7L2 binding sire (BS) in *LARGE2* via CRISPR/Cas9. **A**) SW480 n=5 and **B**) PDTO1 n=2. Results are depicted as mean ± SD. **** p<0.0001.

To proof that editing of the TCF7L2 binding site within the *LARGE2* intron also abolishes occupation by the transcription factor, we used qChIP analysis on DNA fragments precipitated with an antibody directed against TCF7L2. Compared to the results from SW480 wt cells, *LARGE2* intron 1-fragments were not enriched in SW480 cells with a mutated TCF7L2 binding site (compare **Figure 21** and **Figure 18B**).

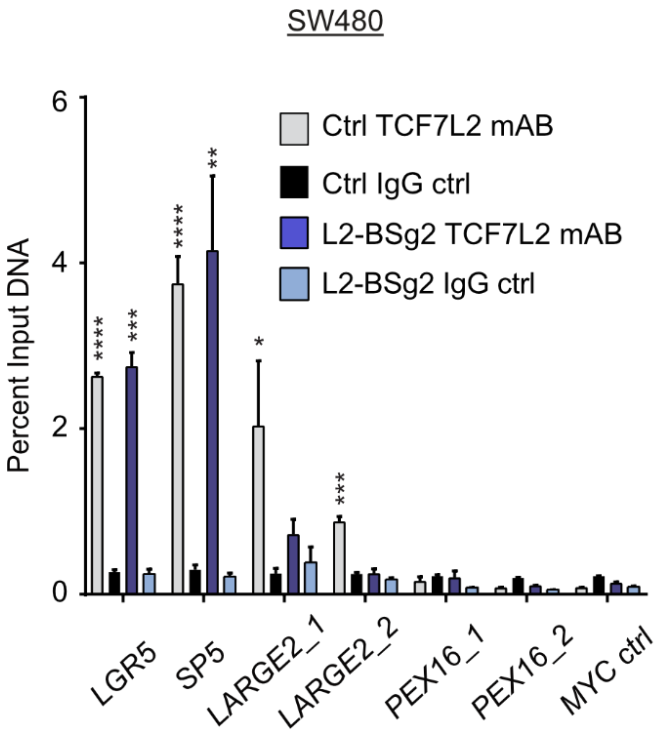
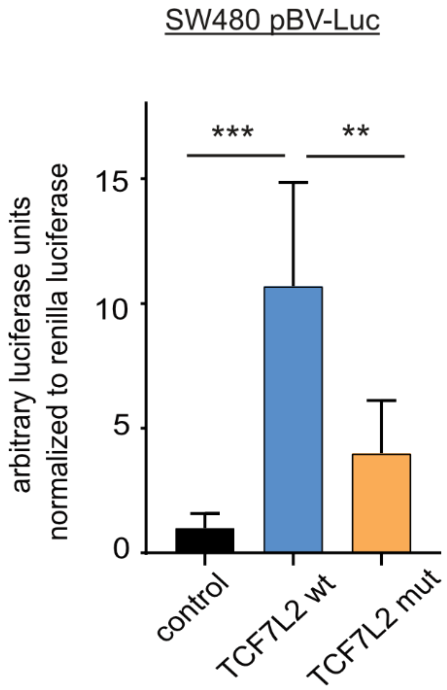


Figure 21: qChIP analysis on SW480 cells. Cells are either wild-type Ctrl) or mutant for TCF7L2 binding site in *LARGE2* (L2-BSg2). The amount of DNA, in percent of the total chromatin input, immunoprecipitated with either TCF7L2 antibody or rabbit IgG control antibody. Results are shown as mean ±SD (n=3). * p<0.1, ** p<0.01, *** p< 0.001, **** p< 0.0001.

To further validate our results, we cloned a ~600 bp DNA fragment, covering the TCF7L2 binding site of *LARGE2,* into the luciferase reporter construct pBV-Luc (=TCF7L2 wt) (Plasmid kindly provided by Heiko Hermeking). High intrinsic Wnt activity in SW480 cells induced the luciferase activity up to 10-fold compared to the empty reporter vector. In accordance with previous results, a site directed mutagenesis of the "CTTTG**AT**C" TCF7L2 binding motif to "CTTTG**GC**C" (Korinek *et al.*, 1997), significantly restricted the activity of the luciferase reporter (compared to the TCF7L2 wt fragment) (**Figure 22**). Taken together, our data show that *LARGE2* represents a direct target gene of the WNT transcription factor TCF7L2, and binding of TCF7L2 to its recognition motif in the first intron of *LARGE2* is essential to drive *LARGE2* gene expression upon WNT activation in CRC cells.



5.3. LARGE2 correlates with active Wnt

Figure 22: Luciferase reporter assay in SW480 cells. Cells were transfected with equimolar amounts of control pBV-Luc vector, the vector containing wt or mutated TCF7L2_BS. A Renilla luciferase vector was used as a transfection control., luciferase activity was measured 36 hrs after transfection. Results are shown as mean (n=3) ± SD. ** p < 0.01, *** p < 0.001

signaling and hCoSC gene expression in CRC

CRC is a very heterogeneous disease, and different CRC subclasses are described according to their distinct molecular and pathophysiological characteristics. One of these classification systems distinguishes between four consensus molecular subtypes (CMS1 to 4) (Guinney *et al.*, 2015). We wondered whether the expression of *LARGE2* is associated with any of those categories. By analyzing two independent CMS classified CRC patient cohorts, TCGA-COAD (Cancer Genome Atlas and The Cancer Genome Atlas Network, 2012) and the cohort from Marisa et al., released in 2013 (Marisa *et al.*, 2013), we found *LARGE2* mRNA levels highest in the CMS2 subtype, characterized by active WNT/c-MYC signaling (**Figure 23A,B**).

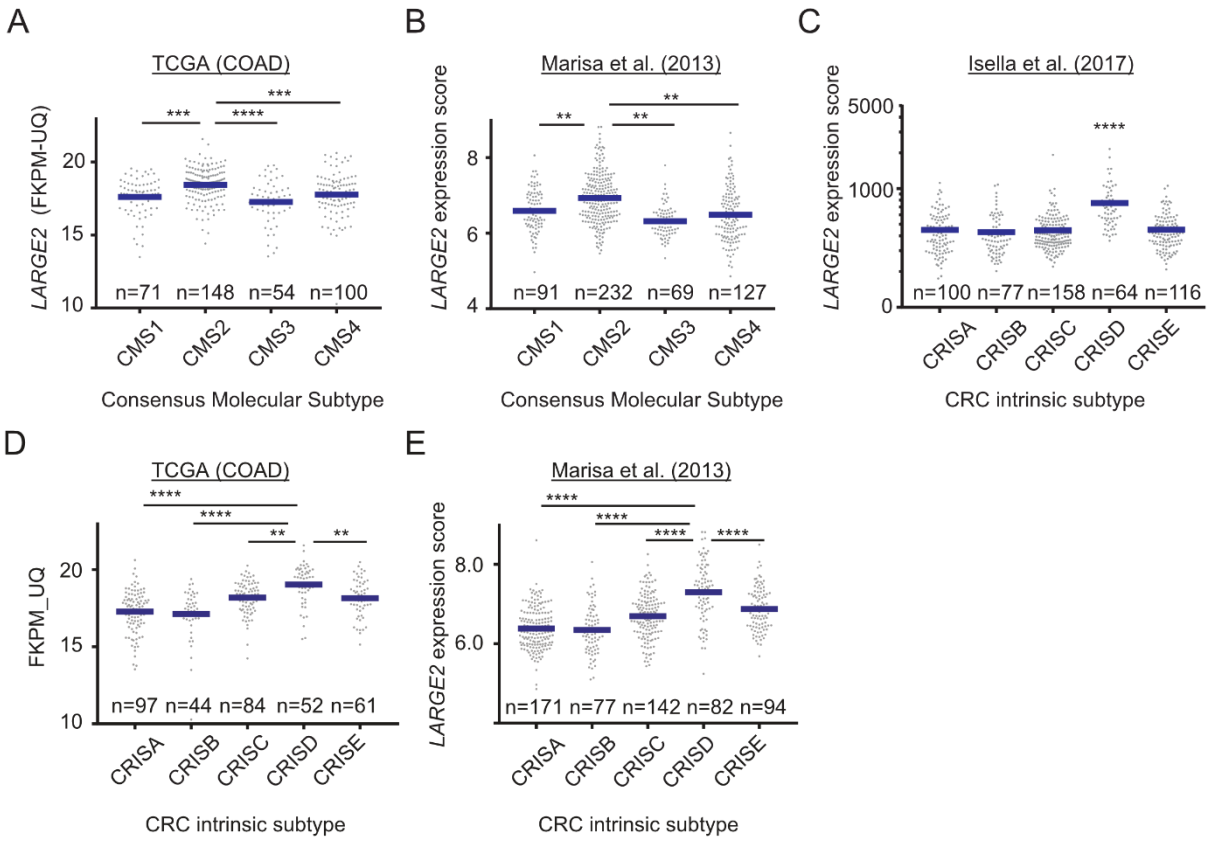


Figure 23: LARGE2 gene expression in CRC subtypes. Classification in CMS of **A**) TCGA COAD dataset and **B**) data from Marisa et al. (2013). **C**) Division of PDX samples from Isella et al. (2017) **D**) TCGA COAD and **E**) Marisa et al. (2013) according to CRIS. Shown are the mean levels of LARGE2 expression (horizontal blue line). Asterisks define the significance of CMS2, or CRIS-D compared to other subtypes. Significance was calculated by one-way ANOVA: ** p< 0.01, *** p<0.001, **** p<0.0001.

Since stromal content within primary tumor samples seriously biases the transcriptional classification of CRC tissue (Calon *et al.*, 2015), a recent study defined human colorectal cancer epithelial-cell specific intrinsic subtypes (CRIS A-E) of 515 xeno-transplanted human tumor samples (PDXs), derived from 244 CRC patients, by performing RNA deconvolution (Claudio Isella *et al.*, 2017). When analyzing this dataset, we found that the highest expression of *LARGE2* occurs in the CRIS subset D, well defined by high Wnt activity and an intestinal stem cell (ISC) phenotype (**Figure 23C**). Employing the TCGA-COAD and Marisa patient cohorts, re-classified according to CRIS classes, we found *LARGE2* highly enriched in the CRIS-D subtype also in these two datasets (**Figure 23D,E**).

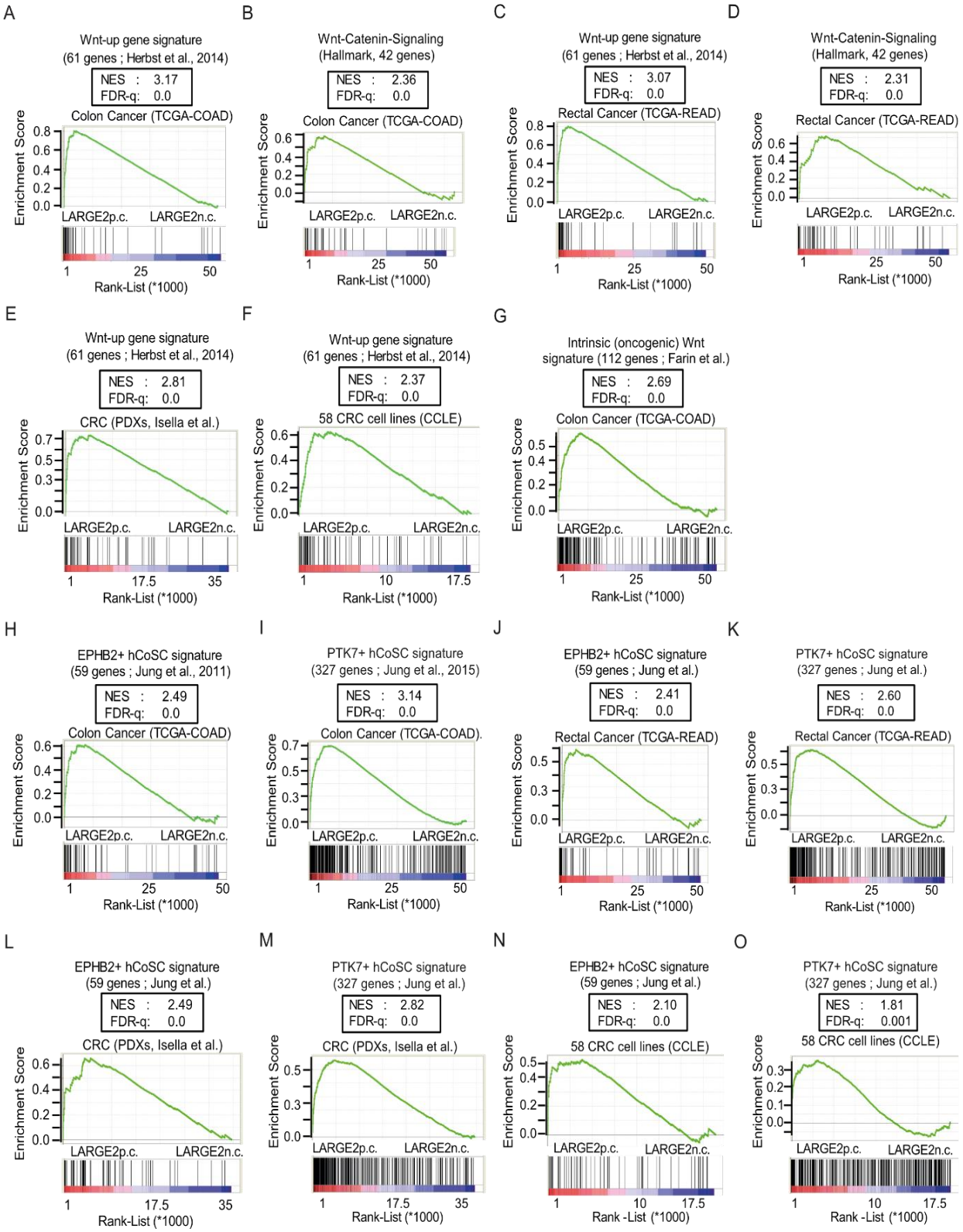
Next, we used publicly available NGS RNA sequencing data derived from 458 colon cancer (COAD) and 167 rectal cancer (READ) patients, provided by the NCI-GDC (https://gdc.cancer.gov/), and generated pre-ranked gene lists according to their Pearson-correlation with the expression of *LARGE2*. The same approach was used for the CRIS-PDXs gene expression data set and for CRC cell line data (Cancer Cell Line Encyclopedia, CCLE) (Barretina *et al.*, 2012; Claudio Isella *et al.*, 2017).

Using gene set enrichment analysis (GSEA) we compared gene-sets specifying genes upregulated upon active Wnt-signaling, upon siRNA-mediated β-Catenin silencing in CRC cells (Herbst *et al.*, 2014) and alternatively a set of genes de-regulated by different modes of β-Catenin activity modulation in 4 different cell lines (Liberzon *et al.*, 2015), with the expression of *LARGE2* in different data sets.

We could confirm that genes from these Wnt-signatures are positively correlated with *LARGE2* expression in the used TCGA-COAD and -READ datasets. Moreover, we saw the same positive correlation in the PDX cohort and CCLE dataset. (**Figure 24A-F**).

Additionally, a recently published description of an oncogenic/intrinsic gene set by Michels et al. (Michels *et al.*, 2019) was also enriched among the genes positively correlating with *LARGE2* expression in the TCGA-COAD dataset (**Figure 24G**). Furthermore, we studied these *LARGE2* signatures with respect to the distribution of two published gene sets defining $EPHB2^{high}$ and $PTK7^{high}$ human colonic stem cells (hCoSCs), since Wnt-signaling is mainly active in non-differentiated intestinal cells

and maintains stemness in the colonic epithelium and in CRC. In fact, both hCoSC-specific gene sets were enriched among the genes positively correlating with



LARGE2 expression in the TCGA-COAD and -READ data, as well as in the Isella et al. PDX cohort and CCLE data sets (**Figure 24G-N**).

Figure 24 (see previous page): GSEA from CRC cohorts shows correlation between Wnt and hCoSC signatures and *LARGE2* expression. **A-G**) Two independent Wnt-gene sets show a positive correlation to the *LARGE2* expression in datasets from TCGA-COAD, TCGA-READ, PDXs cohort and CRC cell lines from CCLE database. **H-O**) Gene sets which specify EPHB2^{high} or PTK7^{high} hCoSCs show positive correlation to LARGE2 expression. **NES**: normalized enrichment score, **FDR-q**: False discovery rate q-value.

Moreover, we analyzed and compared the gene expression of *LARGE2* and bona-fide Wnt genes in different CRC cell lines. HT-29 and Colo205, characteristic for low intrinsic Wnt activation (Rosin-Arbesfeld *et al.*, 2003), and RKO cells, which are wild-type for *APC* and *CTNNB1* (Scholer-Dahirel *et al.*, 2011) show relatively low levels of *LARGE2*, *LGR5*, *AXIN2* and *NKD1.* In comparison, SW480 and SW620 cells, highly Wnt-active due to shorter *APC* alleles or LS174T cells, which express oncogenic mutant β-Catenin, display higher expression levels of the afore mentioned genes (**Figure 25**).

Overall, these analyses show, that colonic and rectal tumors with highest expression of *LARGE2* are characterized by active Wnt signaling, an ISC phenotype, and a hCoSC gene expression program.

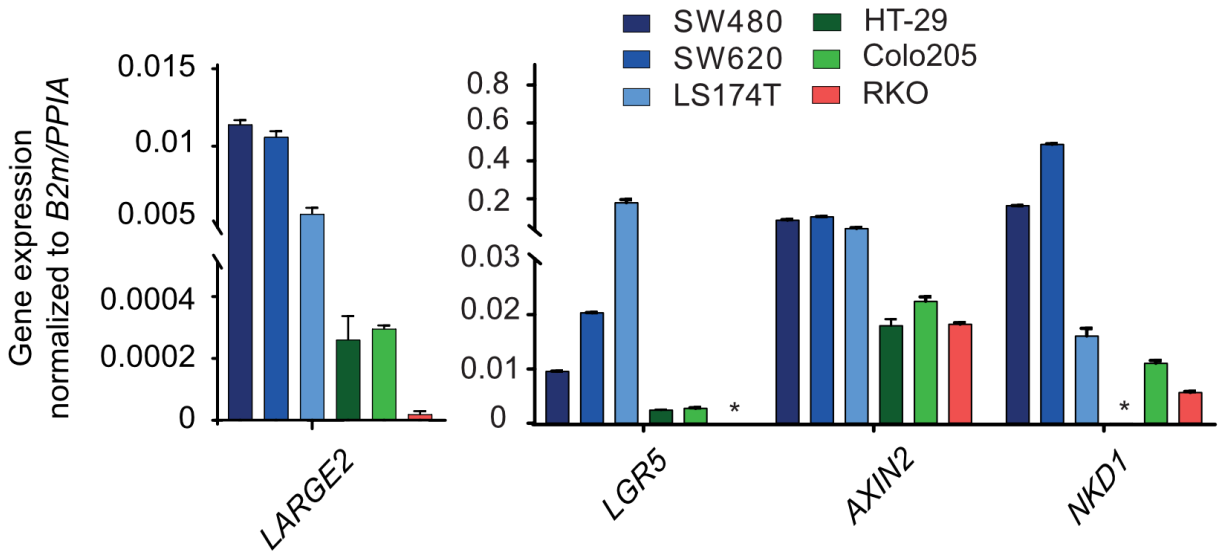


Figure 25: qRT-PCR analysis of Wnt target genes and LARGE2 expression in CRC cell lines. SW480, SW620 and LS174T cells (all high in intrinsic Wnt-signaling) show higher expression of LAREG2 and bona-fide Wnt-target genes when compared to cells with low intrinsic Wnt activity (HT-29, Colo205 and RKO). Results are shown as mean ± SD from biological duplicates. *: gene expression not detectable by qRT-PCR.

5.4. LARGE2 is essential and sufficient for matriglycan formation on α-Dystroglycan in CRC

In order to assess the importance of LARGE2 activity for α-Dystroglycan (α-DG) glycosylation in colorectal cancer cells, we generated targeted knockouts (KO) of LARGE2 in SW480 cells, its lymph node metastatic derivative SW620, and in a patient-derived CRC organoid line from primary colorectal cancer (CIN, APC and TP53 mutant) via CRISPR/Cas9-mediated genome editing. In our experimental approach, delivery of eCas9 and two different guide RNAs was achieved by stable transduction of cells using lentiviral particles. Both gRNAs are targeting different recognition sequences within the *LARGE2* open reading frame (**Figure 26A**) and, as a control, a non-targeting guide RNA was used. Successful mutation of the targeted *LARGE2* loci was confirmed by the Alt-R Genome Editing Detection Kit (**Figure 26B**).

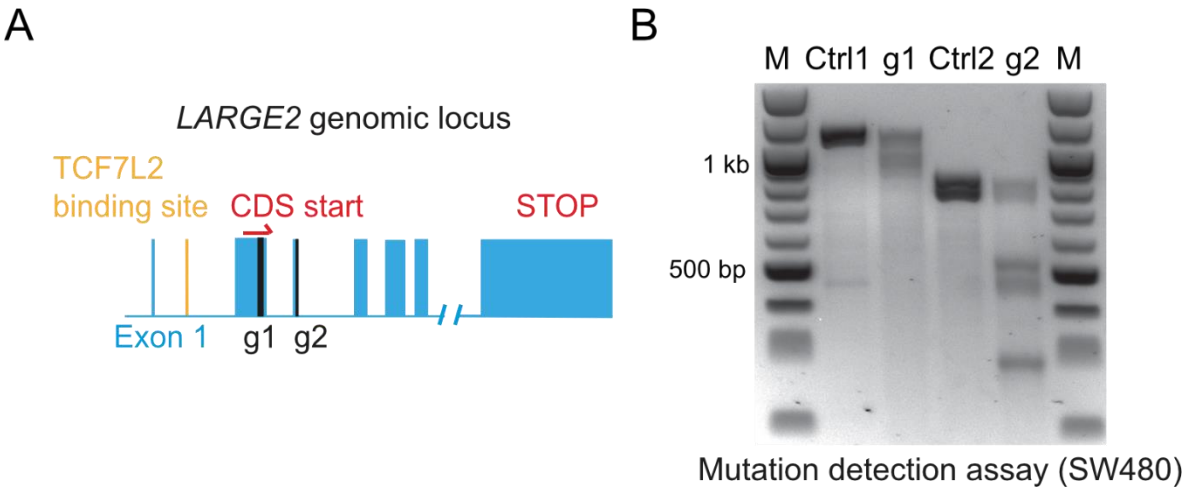


Figure 26: CRISPR/Cas9 mediated targeting of LARGE2. **A**) LARGE2 exons pictured as blue boxes, gRNAs 1 and 2 target sites in black, TCF7L2 binding site in yellow and CDS start and stop in red. **B**) Mutation detection assay on genomic DNA from SW480 cell pools, targeted via gRNA1 and gRNA2 in LARGE2, compared to control.

Due to the lack of commercially available antibodies which would detect LARGE2 protein in immuno-blotting analysis, we affinity enriched glycoprotein fractions from LARGE2 wild type and LARGE2 mutant cell-derived whole protein lysates by using wheat germ agglutinin (WGA) agarose (4.10.2). An antibody, directed against O-glycosylated α-DG glycan epitopes (IIh6c4) allowed us to examine potential effects of LARGE2 inactivation on the glycosylation status of α-DG. Targeting of LARGE2 almost completely diminished α-DG glycosylation in SW480/SW620 CRC cells and in the primary tumor organoid. With respect to molecular weight of the α-DG signal, we

could observe a downwards shifted signal for α-DG in PDTO1 guide 1. However, this might be due to incomplete targeting of the *LARGE2* locus in this organoid pool (**Figure 27**), since organoids transduced to express guide 2 showed a loss of α-DG O-glycosylation, similar to what we had observed in SW480 and SW620 cells (guide

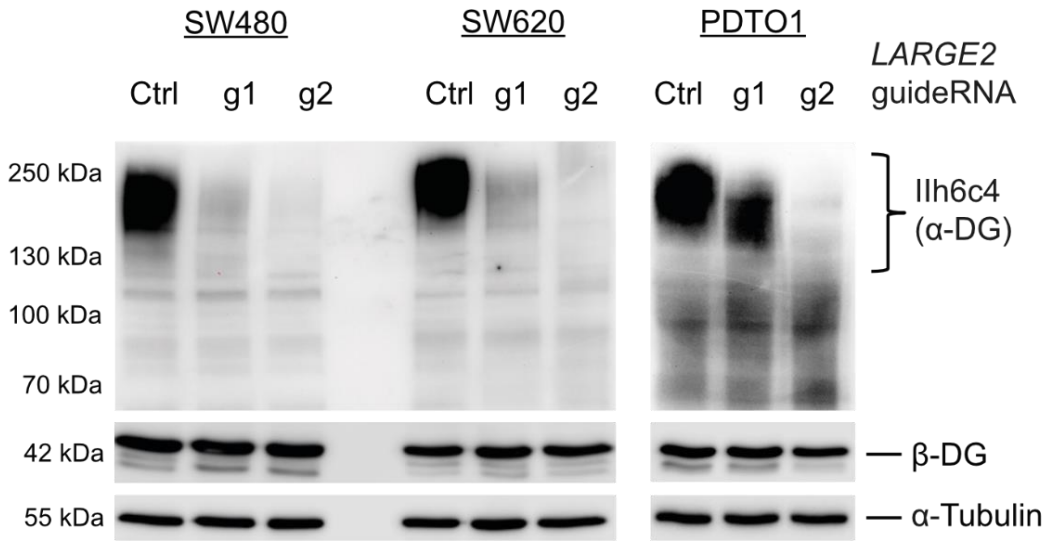


Figure 27: Knockout of LARGE2 in SW480/SW620 cells and primary tumor organoid. KO of LARGE2 using two different gRNAs resulted in vanished signals for O-glycosylation of α-DG. Lysates of WGA-enriched glycoproteins were used to detect α-DG and whole cell lysates (WCL) used to detect β-DG and α-Tubulin.

1 and guide 2).

Notably, the loss of LARGE2 functionality did not affect the levels of β-Dystroglycan (β-DG), the second splice product of the pre-protein DAG1. These data show that LARGE2 is essential for glycosylation of α-DG in chromosomally instable CRC cell lines and patient-derived CRC organoids.

Next, we wondered if LARGE2 activity alone is sufficient to mediate O-glycosylation of α-DG in CRC cells characterized by low Wnt signaling activity and hence low endogenous expression of *LARGE2*. Consequently, we stably transduced HT-29 cells with lentiviruses carrying a doxycycline-inducible cDNA encoding for LARGE2 and induced its expression for 72 hrs. As we could observe, induction of *LARGE2* was ~10-25-fold higher, when compared to Wnt-active SW480 and PDTO1 cells or Wnt-activated HT-29 cells (**Figure 28**). This higher than physiological level of *LARGE2* expression (overexpression) represents a limitation of this approach. However, also previous studies which focused on the identification of LARGE2 substrates other than α-DG performed their analyses after *LARGE2* overexpression

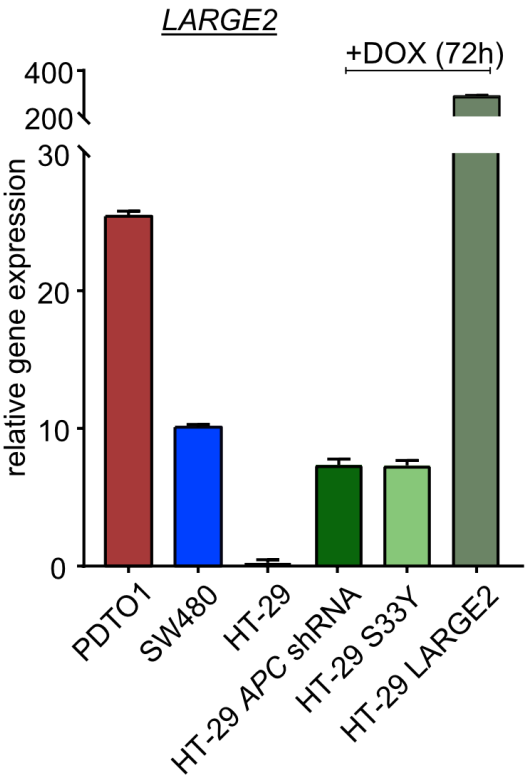


Figure 28: Comparison of endogenous and ectopic *LARGE2* gene expression. Expression of *LARGE2* in the PDTO1 and the CRC cell line SW480, in comparison to wild type HT-29 cell and the DOX-inducible derivates HT-29 APC shRNA, HT-29 β-cat S33Y, or HT-29 LARGE2. Results are shown as mean ± SD from technical duplicates.

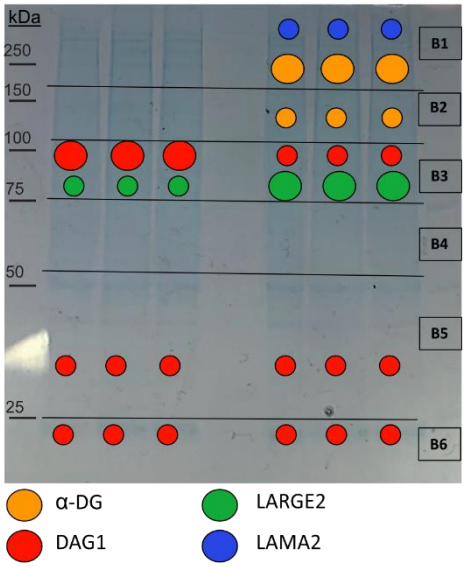
(K.-I. I. Inamori *et al.*, 2016)

For example an ectopically expressed LARGE2 enzyme has been reported to also O-glycosylate laminin-binding proteins like GPC4 in mouse ES cells (K. I. Inamori *et al.*, 2016). To shed light on the identity of proteins whose glycosylation is affected by *LARGE2* overexpression in HT-29 cells or proteins which display affinity to glycosyl structures, we used quantitative mass spectrometry (qLC-MS/MS; in collaboration with Prof. Dr. Küster and Svenja Wiechmann, TUM Freising). To distinguish between these two classes of proteins, we electrophoretically pre-fractionated WGA IP-enriched proteins into 6 molecular weight (MW) windows (0-25, 25-55, 50-75, 75-110, 110-160, and 160-300 kDa) prior to qLC-MS/MS analysis (**Figure 29A**) in order to account for the shift in molecular weight after augmented matriglycan attachment by ectopic LARGE2.

Endogenous LARGE2 was detected in HT-29 control cells only in the 75-110 kDa MW window, in line with to its calculated MW of 81.8 kDa (source: www.uniprot.org). DAG1 peptides were detected in both samples between 75-110 kDa; which represent the peptides of the core protein DAG1. Notably, the intensity of non-glycosylated DAG1 peptides at this MW was overall decreased in the *LARGE2* overexpression

sample when compared to the control, which pointed to enhanced post-translational

A



B

HT-29 WGA-IP fraction for LC-MS/MS	pTZgw-empty Protein ID (iBAQ)	pTZgw-*LARGE2* Protein ID (iBAQ)
B1 (160 < kDa)		DAG1* (54.0E+06) LAMA2 (0.45E+06)
B2 (110 - 160 kDa)		DAG1* (11.3E+06)
B3 (75 - 110 kDa)	DAG1 (83.4E+06) LARGE2 (3.5E+07)	DAG1 (11.9E+06) LARGE2**(1.5E+11)
B4 (55 - 75 kDa)		
B5 (25 - 55 kDa)	DAG1 (15.9E+06)	DAG1 (25.0E+06)
B6 (< 25 kDa)	DAG1 (37.2E+06)	DAG1 (29.5E+06)

*: all detected DAG1 peptides matched alpha-DG
**: intensity maximum of ectopic LARGE2

Figure 29: LC-MS/MS analysis on HT-29 cells, overexpressing LARGE2 compared to control. **A**) Lysates enriched for glycoproteins via WGA were separated on SDS-Page and split in 6 fractions according to their molecular weight. Shown in color are detected proteins of interest in respect to their intensity. **B**) iBAQ (Intensity Based Absolute Quantification) of the proteins of interest.

processing of DAG1 by LARGE2. Peptides detected between 0-55 kDa matched the β-DG subunit. These results confirm our immunoblot analysis of WCLs (**Figure 27**). Importantly, peptides matching the α-DG subunit of cleaved DAG1, but not β-DG, were detected at 110-160 kDa and with very high intensity at 150-300 kDa exclusively upon ectopic expression of *LARGE2* (**Figure 29B**). This points to α-DG as the main catalytical substrate of LARGE2 in HT-29 CRC cells. GPC4, described by Inamori *et al*. as a LARGE2 substrate in mouse embryonic stem cells, was not detected in lysates derived from the HT-29 cell line.

Interestingly, in the 160-300 kDa glycoprotein fraction of *LARGE2*-overexpressing cells we found LAMA2 (Laminin subunit alpha 2) specific peptides that were not detectable in control cell fractions. This suggests that LAMA2 co-precipitates with the WGA-α-DG complex. Other detected proteins, which were found exclusively in the *LARGE2* overexpression samples, are provided in 10.1.

In the performed immune-blot analysis of HT-29 cells expressing ectopic *LARGE2*, we observed a strong immuno-reactivity against glycosylated α-DG at a molecular

weight of ~200-250 kDa which was almost absent in control HT-29 cells. In the Laminin-Overlay (OL) assay (see 4.10.4), we observed laminin binding proteins in the WGA purified glycoprotein fractions from *LARGE2* overexpressing cells, which showed a high MW similar to what we had found for glycosylated α-DG itself. We

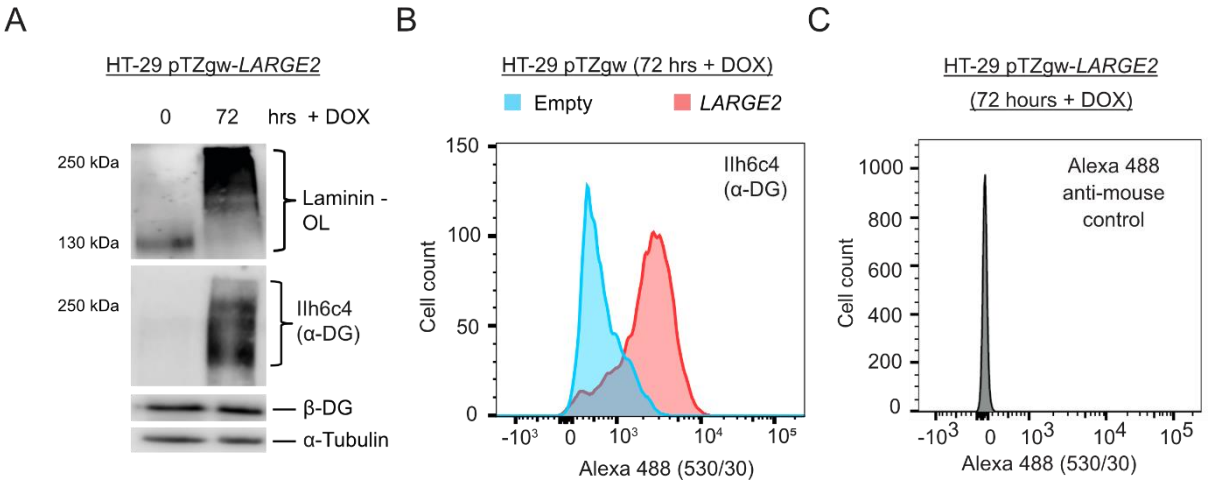


Figure 30: LARGE2 mediates functional O-glycosylation of α-DG in HT-29 cells. **A)** Immunoblot analysis of α-DG and laminin overlay (Laminin-OL) as well as **B**) analysis via flow cytometry in HT-29 cells upon ectopic expression of LARGE2. Cells were treated with 500ng/mL DOX for 72 hours. **C**) anti-mouse antibody control for flow cytometry analysis.

suggest that those detected laminin-binding proteins correspond to matriglycan chain-attached versions of α-DG (**Figure 30A**).

Furthermore, we verified a higher surface abundance of glycosylated α-DG on HT-29 cells upon overexpression of *LARGE2*, quantified by immuno-labeling on viable cells and single cell analysis by flow cytometry (**Figure 30B,C**).

From these data, we conclude that LARGE2 activity is essential and sufficient for the attachment of a high molecular weight matriglycan structure on α-DG, and this confers a binding capacity to laminin in CRC cells.

5.5. Wnt signaling modulates functional O-glycosylation of α-Dystroglycan via induction of *LARGE2* in CRC

Our work showed that *LARGE2* is a direct target of the Wnt signaling pathway. Next, we sought to clarify whether the modulation of Wnt activity in CRC cells is

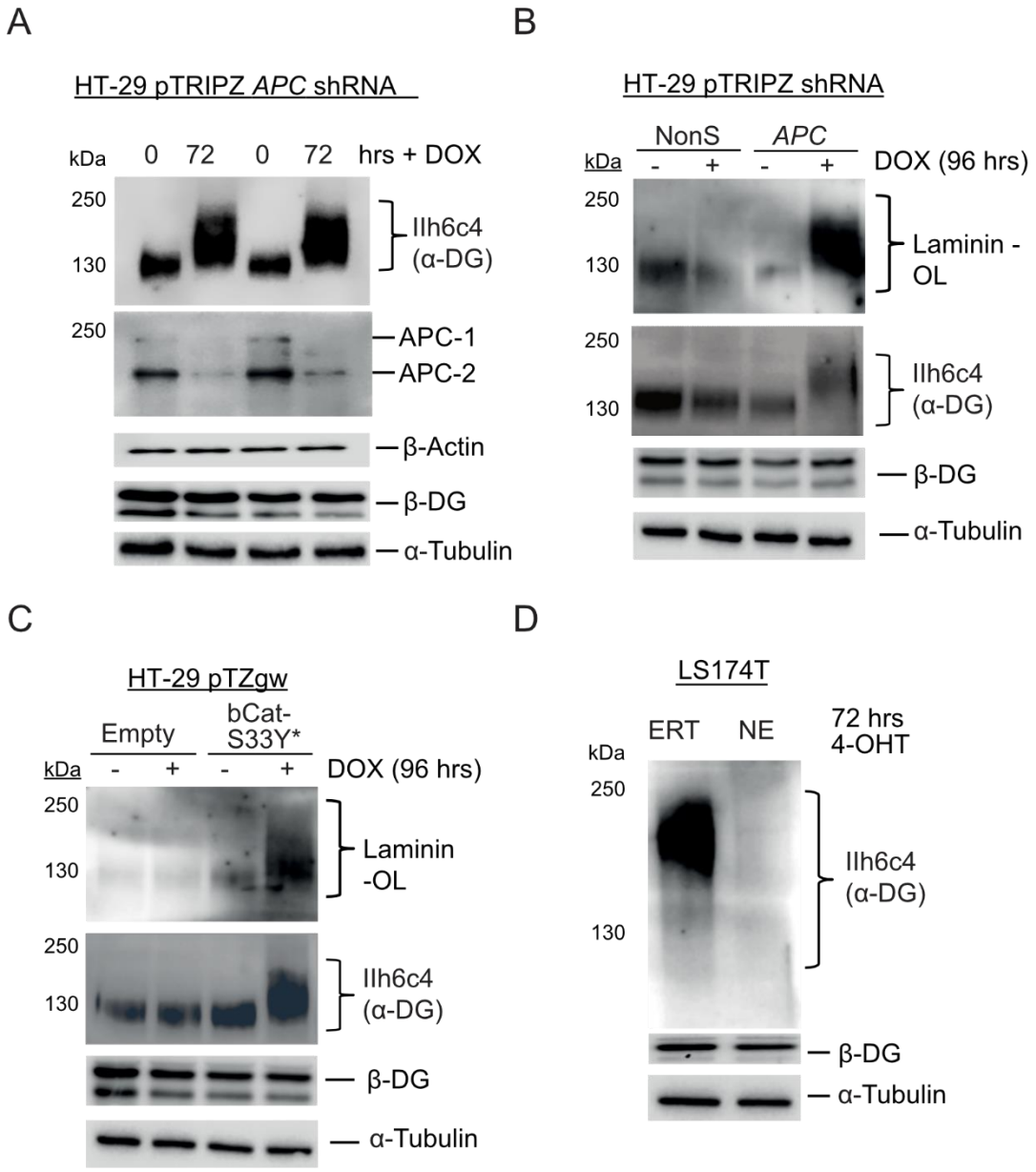


Figure 31: Modulation of Wnt signaling influences functional O-glycosylation of α-DG in CRC. A-C) Immunoblot analysis of O-glycosylated α-DG and laminin binding capacity by Laminin OL in HT-29 cells after silencing of *APC* or expression (A,B) of mutant *CTNNB1* (C). **D**) Analysis of O-glycosylated α-DG in LS174T-NE cells upon 4-OHT treatment for 72 hours, compared to their ERT control. α-DG and laminin were analyzed on WGA-purified lysates, APC; Actin, β-DG and Tubulin were analyzed on WCL.

sufficient to substantially influence the functional O-glycosylation of α-Dystroglycan. Indeed, we observed that conditional, shRNA-mediated knockdown of *APC* or, alternatively, ectopic expression of oncogenic *CTNNB1-S33Y* in HT-29 cells led to an augmented α-DG glycosylation and increased laminin-binding capacity (**Figure 31A-C**).

Moreover, we analyzed the O-glycosylation status of α-DG after acute blockade of Wnt signaling via addition of 4-OHT for 72 hours to LS174T-NE and -ERT cell lines.

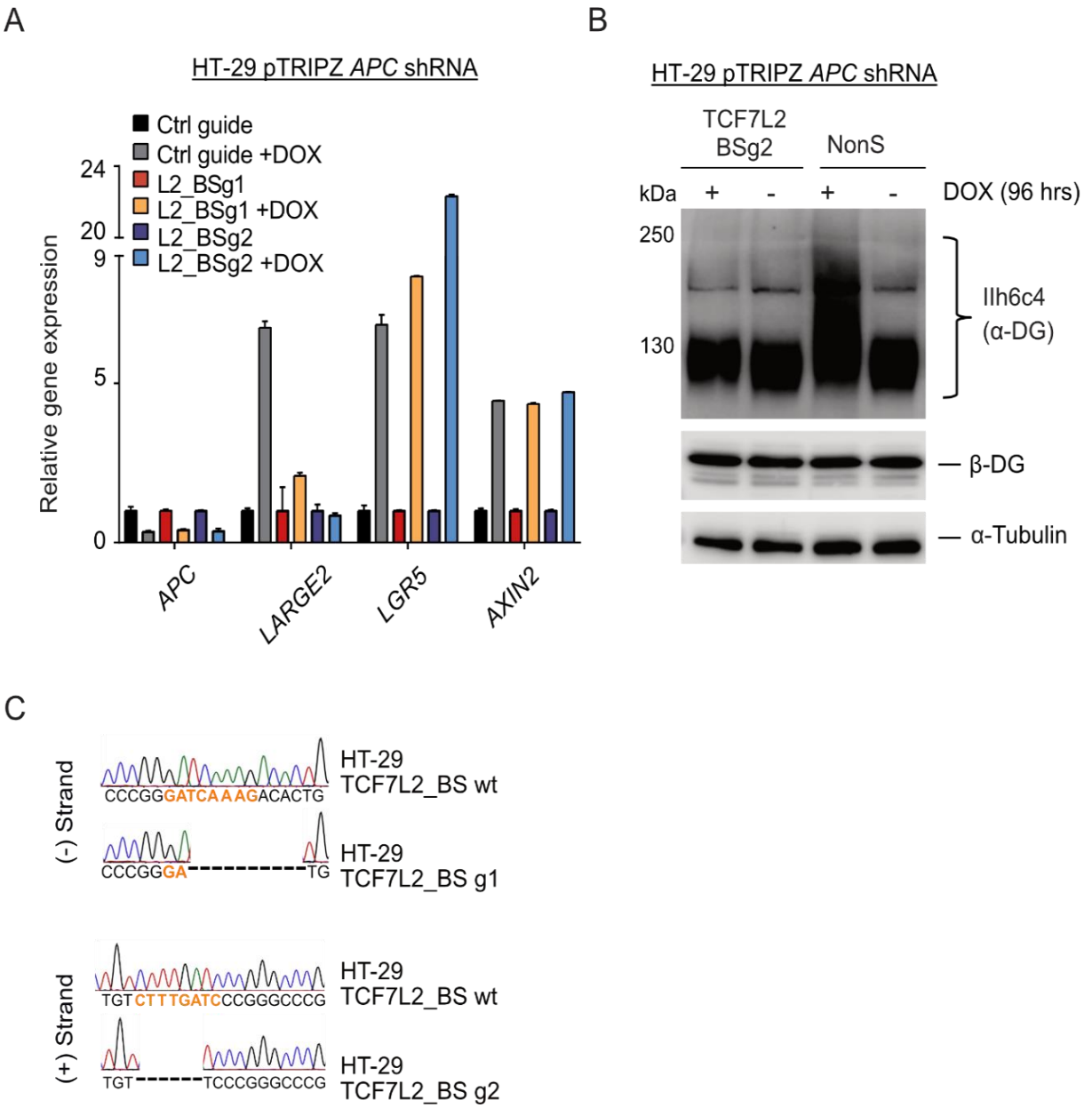


Figure 32: Modification of TCF7L2_BS in HT-29 inhibits induction of *LARGE2* upon silencing of APC. A) qRT-PCR analysis of indicated genes in HT-29 APC shRNA after mutation of TCF7L2_BS. Shown are means ± SD from technical replicates. **B**) Immunoblot analysis of glycosylated α-DG in HT-29 cells wild type or mutant for TCF7L2_BS after 96 hours of DOX treatment. WCL was used for analysis of β-DG and α-Tubulin. **C**) Sanger Sequencing of HT-29 wild type or mutants for TCF7L2_BS gRNA1 and

We noticed a strongly reduced O-glycosylation of α-DG upon Wnt blockade when compared to the respective control CRC cells, which only expressed the ERT2 domain (**Figure 31D**). Importantly, we did not observe changes in mRNA expression of DAG1 or β-DG protein levels upon modulation of Wnt signaling in the analyzed CRC cell lines (**Figure 31**).

Interestingly, we could find that the induction of *LARGE2* expression upon silencing of *APC* in HT-29 cells depends on the integrity of the TCF7L2 binding site (BS) in the first intron of *LARGE2* (**Figure 32A,C**), as a mutated TCF7L2 binding site in HT-29 cells prevented the induction of *LARGE2* upon *APC* silencing. Accordingly, increased functional matriglycan formation on α-Dystroglycan was only detectable in *APC* silenced cells with an intact TCF7L2 binding site, while mutation of TCF7L2_BS left *LARGE2* expression and the MW of α-DG unchanged (**Figure 32B**). We obtained congruent results in pools of Wnt-active SW480 cells and PDTO1, where mutation of TCF7L2_BS with guide1 and guide2 resulted in the loss of α-DG glycosylation (**Figure 33A,B**).

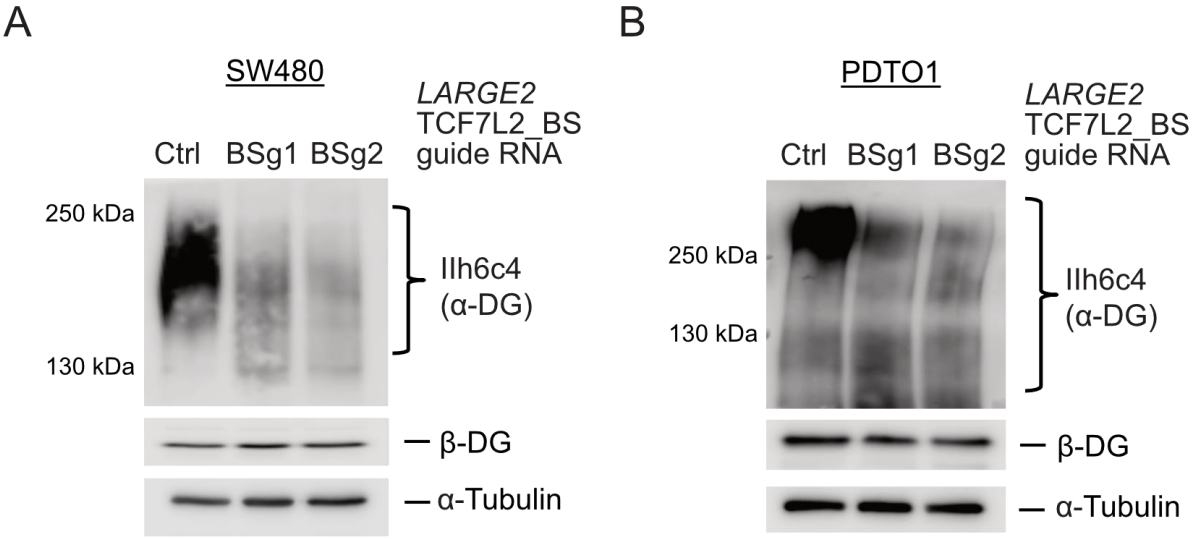


Figure 33: Modification of TCF7L2_BS in CRC cell and tumoroid line. Immunoblot analysis of glycosylated α-DG in SW480 (**A**) and PDTO1 (**B**) after knockout of TCF7L2_BS with 2 gRNAs. α-Tubulin and β-DG were analyzed from WCL.

In summary, our data suggest that Wnt pathway activity, by modulating *LARGE2* gene expression via the TCF7L2 binding site in the first intron, affects the O-glycosylation status of α DG, regardless of the CRC cell line employed in this study.

5.6. *LARGE2* expression and O-glycosylation of α-DG is enriched in hCoSCs

Considering our results pointing to *LARGE2* expression as part of a hCoSC-gene expression program and self-renewing hCoSCs depending on active Wnt signaling (Pinto *et al.*, 2003), we aimed for a deeper analysis in hCoSC-enriched organoids. Therefore, my colleague Dr. Cira García de Durango isolated human colonic mucosal crypts (see 10.2) from fresh tissue specimen and cultured non-differentiated hCoSC-enriched organoids in a 3-dimensional extracellular matrix (Matrigel®) overlaid with self-renewal promoting WENR culture media (4.3.2.1) similar to what has been described before (Jung *et al.*, 2011).

Her following analysis revealed that stem cell enriched human colonic organoids (**Figure 34A**) show overall high expression levels of *LARGE2* and other stem cell genes like *LGR5* and *SMOC2*. Leaving out of the culture medium Wnt ligands (Wnt-3A and RSPO1) to trigger multi-lineage cell differentiation (so-called EN medium), resulted in strong decrease of *LARGE2* expression (~100-fold) and other bona-fide Wnt target genes, while *LARGE1* and *DAG1* mRNA levels remained unchanged. As expected, expression levels of intestinal differentiation markers *KRT20* (pan-differentiation) *ANPEP* (enterocytes), *TFF3* (goblet cells) and *CHGA* (enteroendocrine cells) were increased in differentiated PDOs (**Figure 34B**).

O-glycosylated α-DG protein was detected in hCoSC-enriched organoids at a MW of ~ 130 kDa. However, we observed diminished O-glycosylation of α-DG and a reduction of β-DG protein levels upon PDO differentiation (**Figure 34C**).

As previously reported, self-renewing human PDOs display a heterogeneity of stem and progenitor cells, which can be specified by high or low surface abundance of the inactive protein tyrosine kinase PTK7 (Jung *et al.*, 2015). The effect of differentiation of PDOs was indicated by lower protein levels of PTK7 (**Figure 34C**).

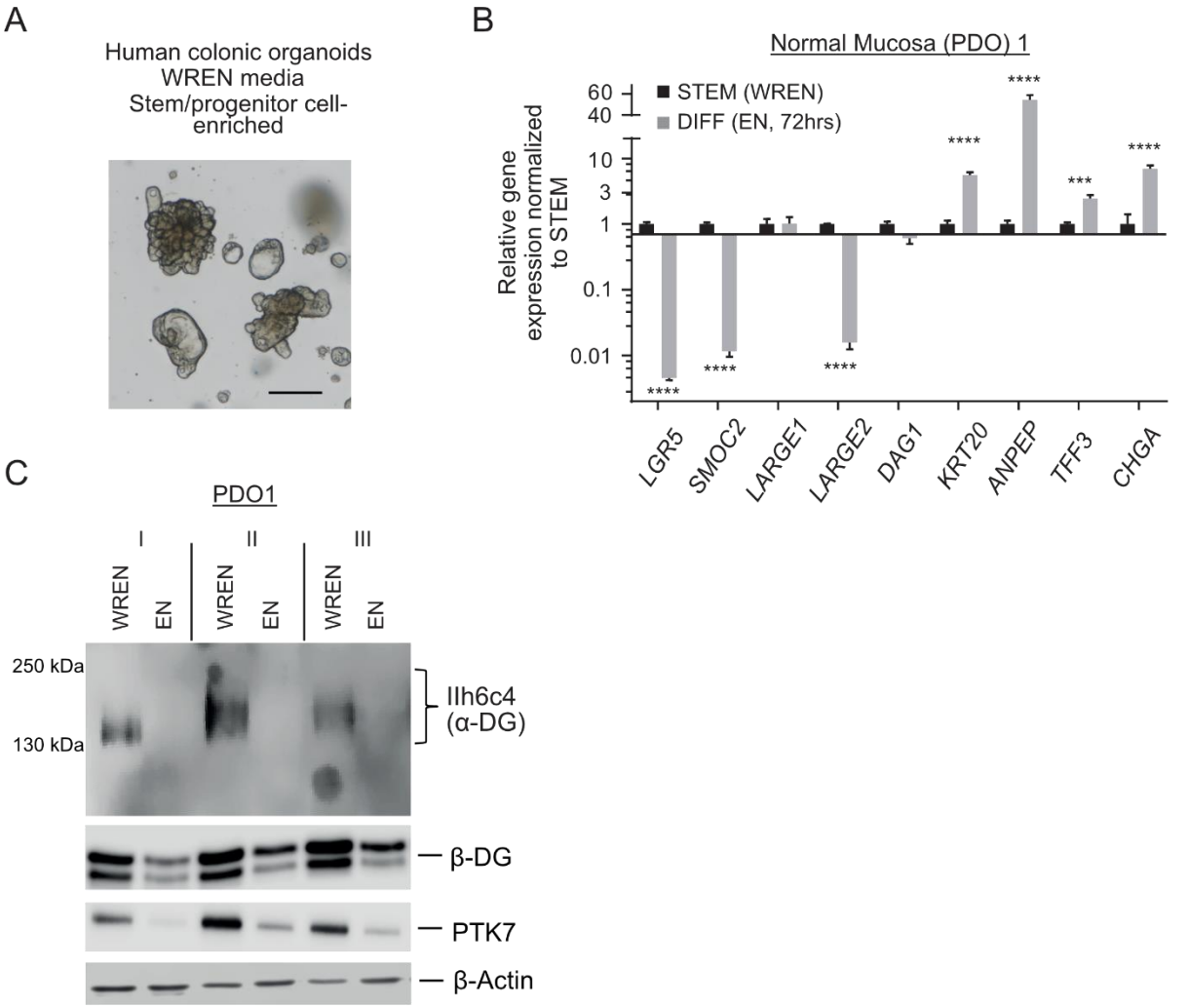


Figure 34: Gene expression in stem-like and differentiated human colonic organoids. **A**) Human colonic organoids (PDO) in Matrigel, cultivated in self-renewing medium (WREN). Scale bar represents 20 µm. **B**) qRT-PCR analysis of indicated genes in PDO1, maintained in WREN or EN medium. Results are depicted as mean ± SD, all experiments analyzed three times. ** p < 0.01, *** p< 0.001, **** p < 0.0001. **C**) Immunoblot analysis of α-DG from WGA enriched lysates, derived from stem or differentiated PDO1 cells after 72 hours. PTK7, β-DG and Actin were analyzed on WCL.

Additionally, Dr. Jung performed flow cytometry-assisted cell sorting (FACS) in order to isolate PDO populations according to their PTK7 surface level. The expression of *LARGE2* in these different subpopulations (**Figure 35A**) was then analyzed via quantitative PCR (4.7). Due to the low yield of RNA from FACS sorted cells, gene specific Taqman assays (ABI) were used for gene expression analysis.

The cell population with highest levels of PTK7 surface abundance (PTK7-hi) was

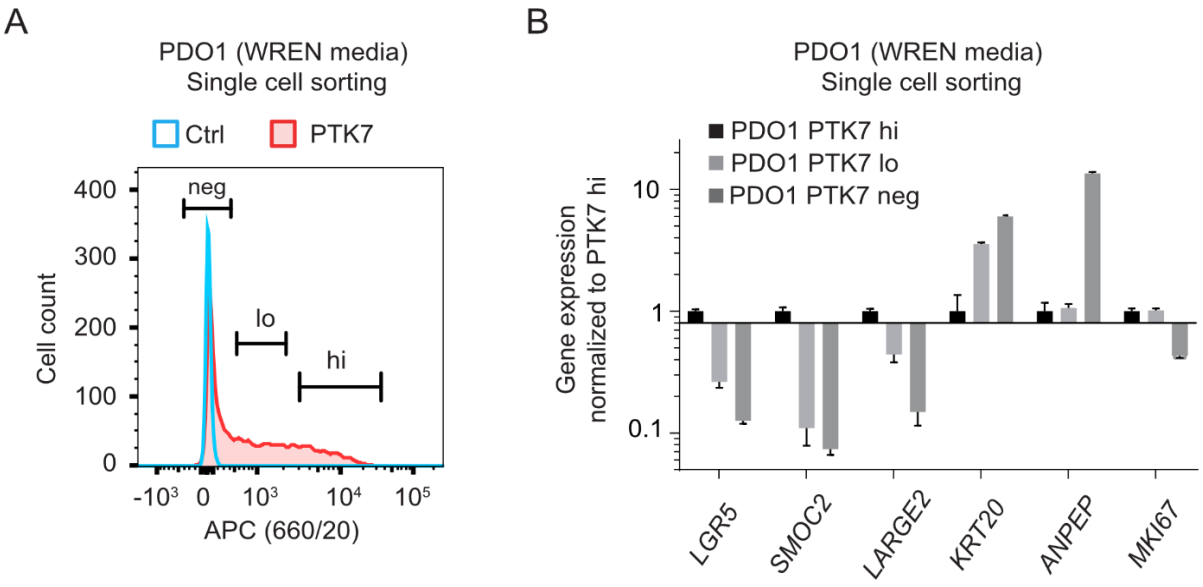


Figure 35:Expression of LARGE2 according to PTK7 surface abundance in PDOs. **A**) FACS profile from in vitro cultured human PDO cells, maintained in WREN medium, stained with PTK7 antibody (APC-coupled). DAPI staining was performed to define PTK7 negative control fraction (Ctrl). **Neg**: PTK7 negative, **lo**: PTK7 low, **hi**: PTK7 high. **B**) TaqMan™ qRT-PCR of indicated genes from PDO1 cell fractions, isolated via FACS (from A). Shown is mean ± SD. Each sample was measured three times.

characterized by strongest expression of *LARGE2*, *LGR5* and *SMOC2,* when compared to PTK7-medium or low fraction. In contrast, the lowest *LARGE2* mRNA levels were detected in surface-PTK7 negative fraction. This fraction also harbored the highest expression of the differentiation markers *KRT20* and *ANPEP* (**Figure 35B**).

Similar results were obtained when comparing sub-fractions of human colonic crypts sorted according to their surface abundance of EPHB2. *EPHB2* is highly expressed at the bottom of crypts, where ISC are located (Jung *et al.*, 2011) (**Figure 36A**). The highest expression of *LARGE2*, *LGR5* and *SMOC2* was observed in this stem cell enriched EPHB2-hi human colonic epithelial subpopulation.

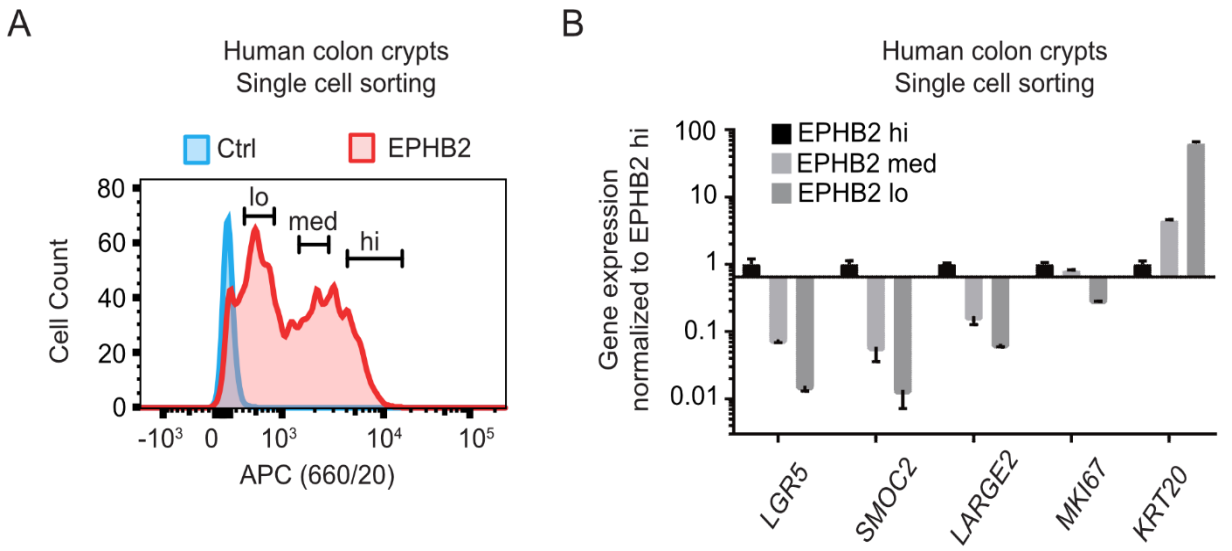


Figure 36: Expression of LARGE2 according to EPHB2 surface abundance in human colonic crypts. **A**) FACS profile from in vitro cultured human colonic crypts, stained with EPHB2 antibody (APC-coupled). DAPI staining was performed to define EPHB2 negative fraction (Ctrl). Lo: EPHB2 low, med: EPHB2 medium, hi: EPHB2 high cell fraction. **B**) TaqMan™ qRT-PCR of indicated genes from human colon crypt cell fractions, isolated via FACS (from A). Shown is the mean ± SD. Each sample was measured three times.

As expected, the same population showed the lowest mRNA levels of *KRT20*. Gene expression of *LARGE2* and the intestinal stem cell markers *LGR5* and *SMOC2* is directly correlated with the abundance of EPHB2, which is representative of the cell differentiation state, while *KRT20* gene expression is inversely correlated with EPHB2 levels (**Figure 36B**).

5.7. A differentiation status dependent gradient of the LARGE2/α-DG axis intestinal tissue is conserved among species

Next, we analyzed sections of formalin-fixed paraffin-embedded (FFPE) human colonic and rectal mucosa for O-glycosylated α-Dystroglycan by immunohistochemistry (IHC), using the glycosylation-sensitive antibody IIh6c4. The strongest levels of immunoreactivity were found in the membrane of stem/progenitor cells at the crypt bottom, while differentiated cells in the upper crypt did not show this staining pattern (**Figure 37A,B**). To verify whether these findings are conserved among species, we additionally performed IHC staining on FFPE sections of mouse intestinal epithelia. We could confirm a similar gradient of immunoreactivity in the

membrane of mouse crypts, which mainly occurred at the bottom of the crypts, while the differentiated cells in mouse villi did not show any signal (**Figure 37C**).

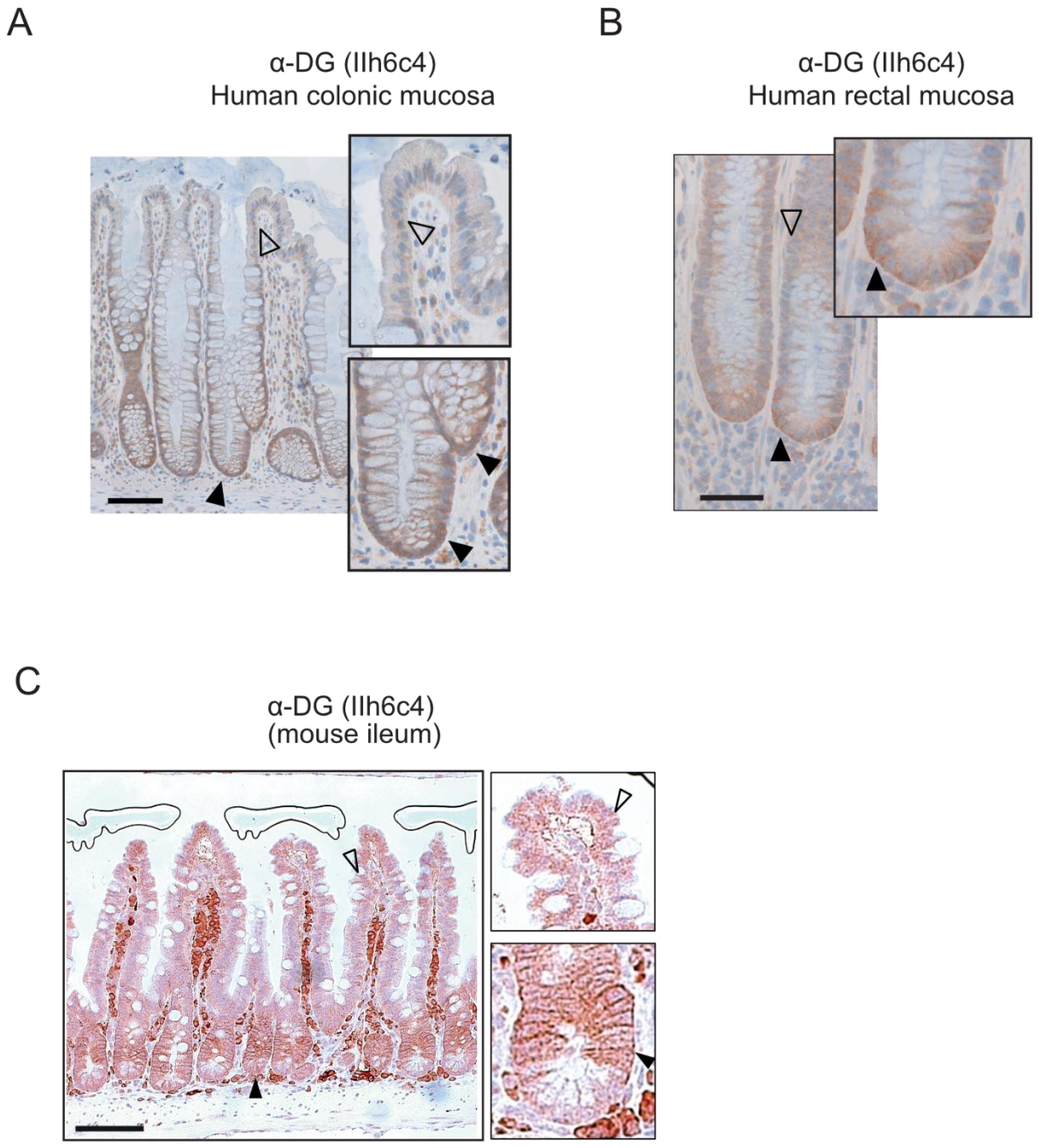


Figure 37:α-DG staining on human and mouse FFPE tissue via IHC. **A**) IHC analysis of glycosylated α-DG (IIH6c4 antibody) on human colonic and rectal mucosa and **B**) mouse ileum. Black arrowheads: crypt base specific staining, clear arrowheads: fading or loss of staining. Scale bars indicate 100 µm (A, C) and 50 µm (B)

As already mentioned earlier, due to the lack of a commercially available antibody against LARGE2, and failure of the classical *in-situ* hybridization protocol on human FFPE tissue, we performed RNAscope ® *in-situ* hybridization (4.14) to detect and locate mRNA with a up to single molecule sensitivity in the human colonic mucosa. We employed specially designed *LARGE2* and *LGR5* specific probe sets on human mucosa sections. This method allowed us to precisely detect mRNA molecules for each target, visualized as brown spots within the tissue after signal development with 3'-Diaminobenzidine (DAB).

Expression of *LGR5* and *LARGE2* was exclusively detected at the crypt bottom in human colon (**Figure 38A,B**), whereas the negative control, targeting the bacterial *DapB* gene, showed no signal overall (**Figure 38C**).

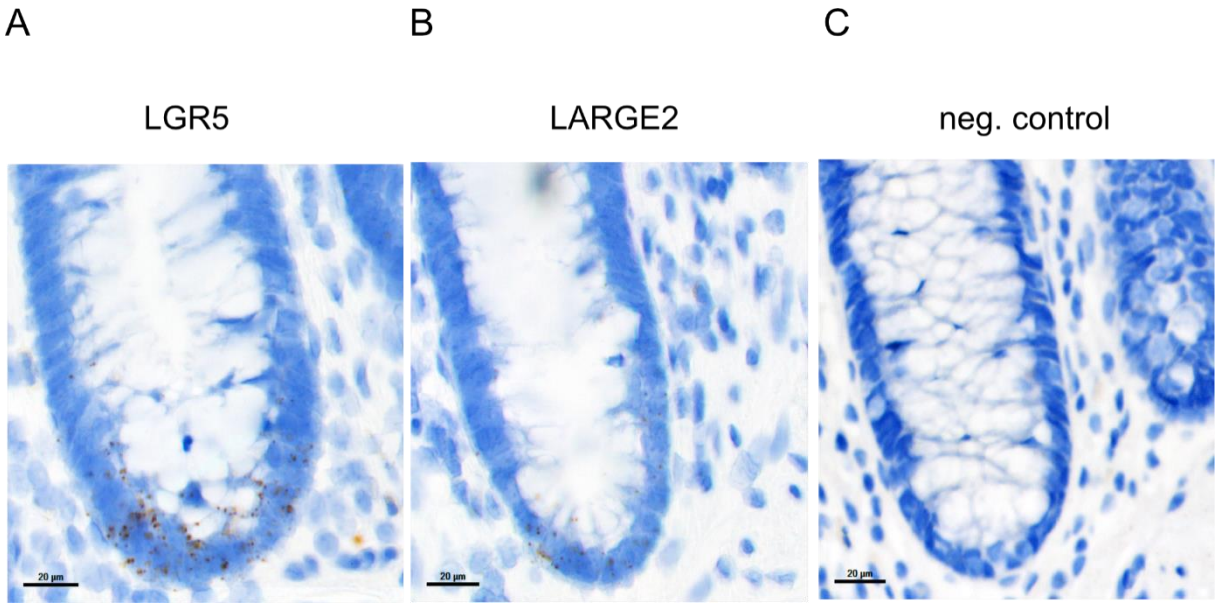


Figure 38: RNAscope in situ hybridization on human colonic FFPE sections. **A-C**) Detection of *LGR5*, *LARGE2* mRNA molecules and a negative control in human colonic mucosa tissue and in human colonic adenoma (**D-F**). Note that brown dots indicate the expression of the indicated gene. As a negative control, a commercial probe set directed against a bacterial gene was used for hybridization.

Next, we performed classic *in-situ* hybridization with a *Large2* directed RNA probe on FFPE sections of mouse intestinal tissue (**Figure 39A**). Here we could detect expression of *Large2* mRNA primarily at the bottom half of the intestinal crypts, and the signal faded towards the top. Intestinal villi, containing terminally differentiated cells, did not show any signal for *Large2* expression.

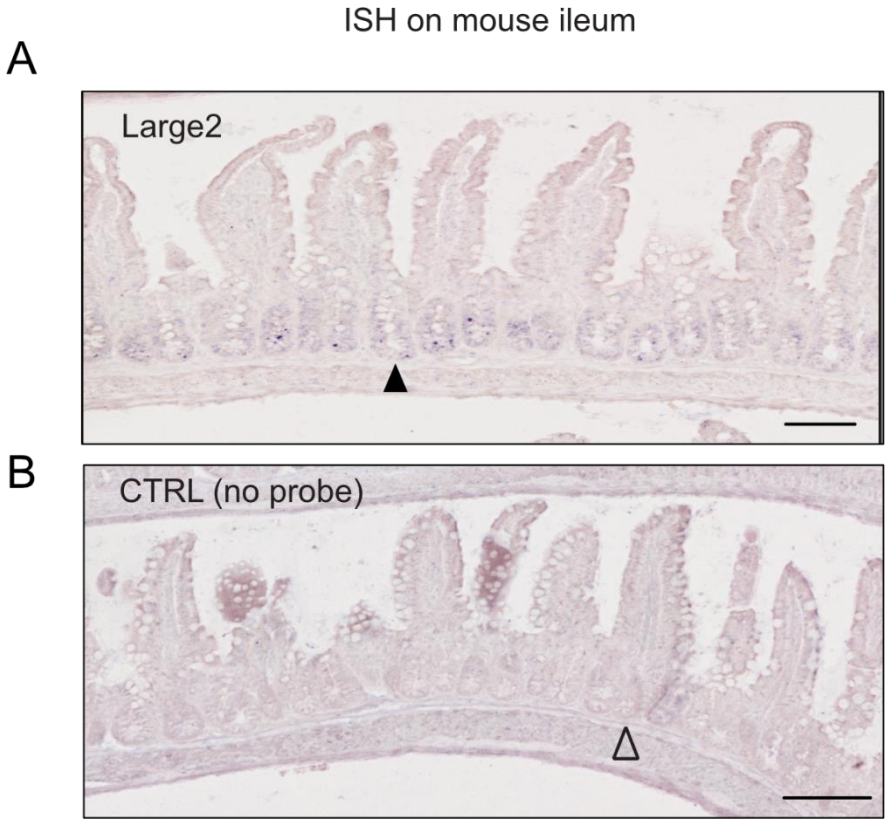


Figure 39: In-situ hybridization with RNA probes on mouse ileum. **A**) A specific probe was used to detect Large2 mRNA on mouse intestinal Ileum, completed by a control section, where no probe was used (**B**).

5.8. *LARGE2* expression and α-DG glycosylation in Wnt active mouse and human adenoma cells

Since classical adenoma are characterized by active Wnt signaling, we sought to address the status of LARGE2 and O-glycosylated a-DG in this early disease context. Using public microarray data, including patient-matched human colonic mucosa-adenoma pairs (GSE8671, (Sabates-Bellver *et al.*, 2007)), we were able to determine a higher *LARGE2* expression in the adenoma specimen (**Figure 40A**). Next, we tested a *LARGE2* probe on human adenoma FFPE sections using the RNAscope assay, as described above. We could detect expression of *LARGE2* (visualized as brown spots) within adenoma cells (**Figure 40B**), while the negative probe showed no signals (**Figure 40C**).

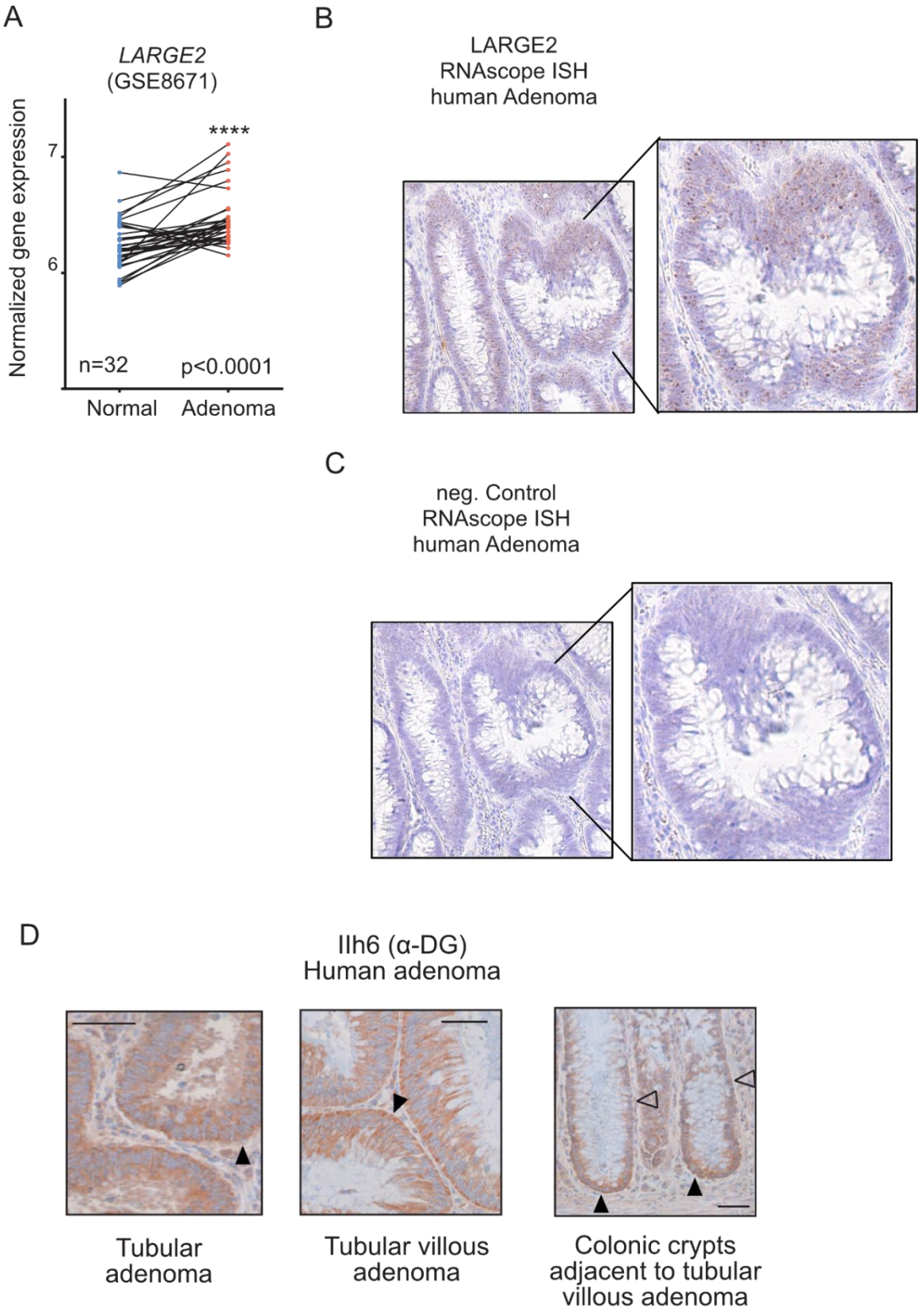

A
LARGE2
(GSE8671)
Normalized gene expression
7
6
****
n=32
p<0.0001
Normal
Adenoma
B
LARGE2
RNAscope ISH
human Adenoma
C
neg. Control
RNAscope ISH
human Adenoma
D
IIh6 (α-DG)
Human adenoma
Tubular adenoma
Tubular villous adenoma
Colonic crypts adjacent to tubular villous adenoma

Figure 40 (see previous page): LARGE2 expression and O-glycosylation of α-DG in human adenoma. A) LARGE2 expression from microarray data of 32 patient-matched human colonic mucosa-adenoma pairs. **B,C**) RNAscope ISH on human FFPE adenoma sections with specific probes for LARGE2 and a negative control. **D**) IHC staining for O-glycosylated α-DG on human tubular (villous) adenoma and crypts from FFPE sections. Black arrowheads indicate LARGE2 and α-DG specific staining, clear arrowheads indicate missing or vanished signals. Significance was calculated with Student's t-test.

Moreover, we performed immunohistochemistry staining for α-DG with a glycosylation sensitive antibody on different human FPPE sections of adenomas. We could observe basal and membranous reactivity in tubular and tubular-villous adenomas, normally characterized by an intrinsically activated Wnt-program (**Figure 40D**, left and middle). Benign crypts in the vicinity of the tubular villous adenoma showed O-glycosylation restricted to the stem cell compartment at the crypt bottom (**Figure 40D**, right).

To confirm the conserved expression of *LARGE2* and the connected O-glycosylation of α-Dystrogylcan in adenomas, we performed classic *Large2* in-situ hybridization and IHC staining on FFPE sections derived from APCmin mice (Moser, Pitot and Dove, 1990). The results showed a strong signal for *Large2* mRNA in the

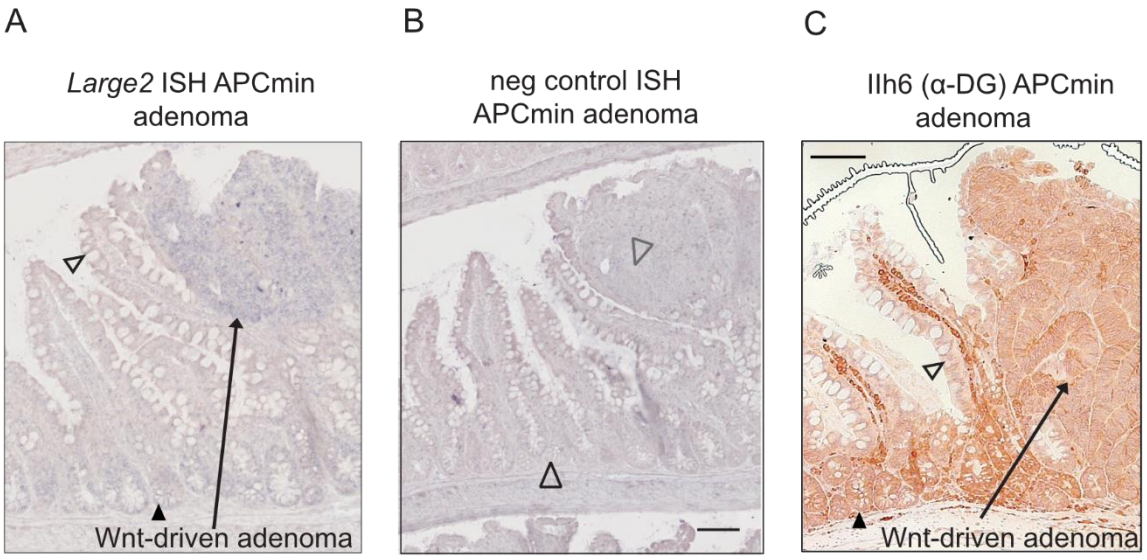


Figure 41: Detection of Large2 expression and O-glycosylation of α-DG on FFPE murine Ileum. **A, B**) In-situ hybridization with Large2 specific or negative probe on Ileum from APCmin mice. **C**) IHC staining for O-glycosylated α-DG on APCmin FFPE sections. Black arrowheads indicate Large2 and α-DG specific staining, clear arrowheads indicate missing or vanished signals.

Wnt-driven adenoma, in addition to the expected gradient-like staining pattern in adjacent crypt-villus units, whereas the negative control showed no signal (**Figure 41A,B**). Likewise, the IHC staining gave a strong reactivity for basal and membranous glycosylated α-DG (**Figure 41C**).

Taken together, these data show that the extent of LARGE2-mediated O-glycosylation of α-DG correlates with the differentiation status of human colonic and mouse intestinal epithelial cells, and *LARGE2* gene expression and α-DG O-glycosylation mainly occurs in the Wnt-driven stem/progenitor cell compartment of the human colonic and mouse intestinal epithelium, as well as in Wnt-driven adenoma.

5.9. LARGE2 gene expression is elevated in primary and liver metastatic CRC

When analyzing expression of *LARGE2* in TCGA cohorts from COAD and READ patients (Cancer Genome Atlas and The Cancer Genome Atlas Network, 2012), we discovered elevated levels of *LARGE2* in advanced colorectal adenocarcinoma, compared to normal colonic tissue (**Figure 42A,B**). Accordingly, patient matched data showed a higher expression of *LARGE2* in the cancerous specimen than in normal tissue (**Figure 42C, D**).

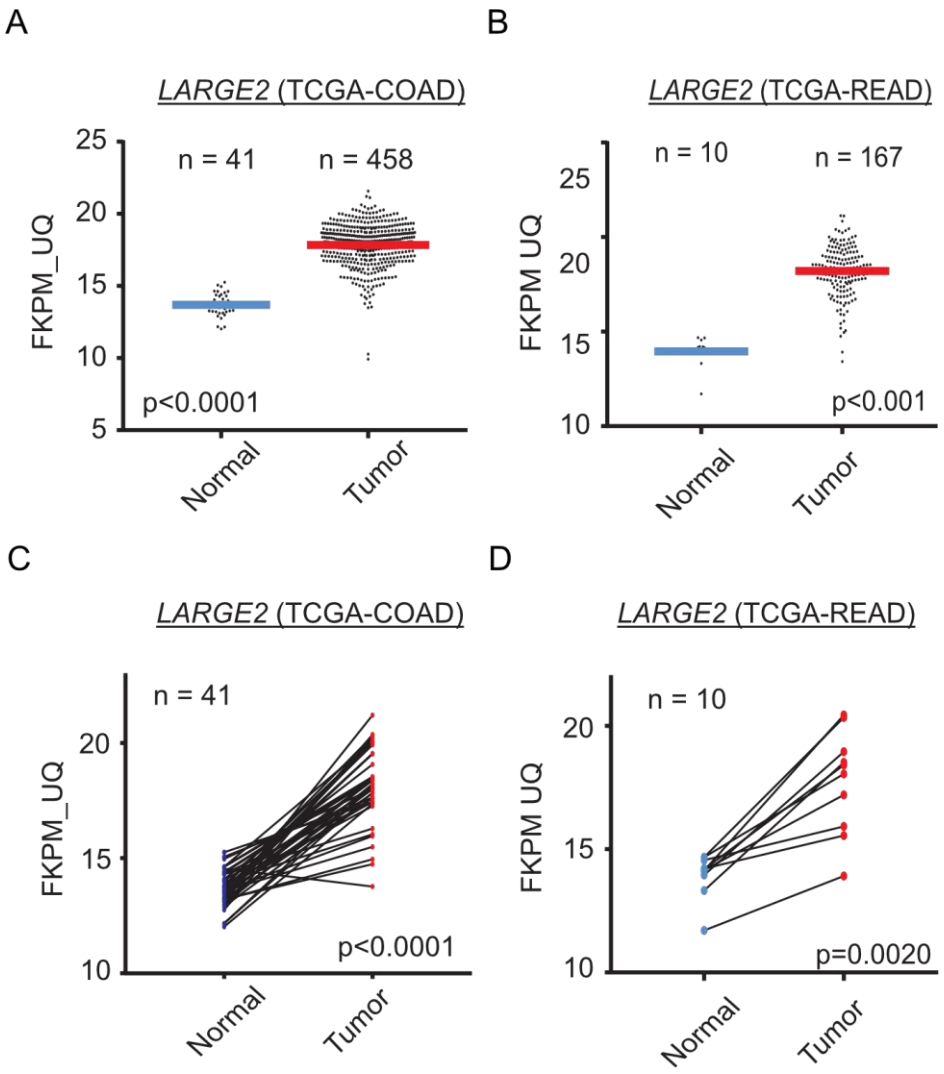


Figure 42: LARGE2 gene expression analysis on human TCGA COAD and READ datasets. **A, B**) Comparison of LARGE2 expression between normal and tumor tissue in COAD and READ. **C, D**) Comparison of LARGE2 expression in patient matched CRC and normal tissue. Significance was calculated via Student's t-test.

To experimentally validate these data on purely epithelial cancer cells, we analyzed PDTOs derived from different individuals suffering from primary or liver metastatic colorectal cancer disease. To assure that all tested organoids are independent on ectopic Wnt-activation, only microsatellite stable PDTO lines, cultivated at least 2 months without Wnt3a and R-Spondin, were chosen for analysis (see 10.3).

6 of the 10 PDTOs were derived from primary tumors after curative colectomy, the other 4 were obtained from individuals suffering from liver metastatic CRC after

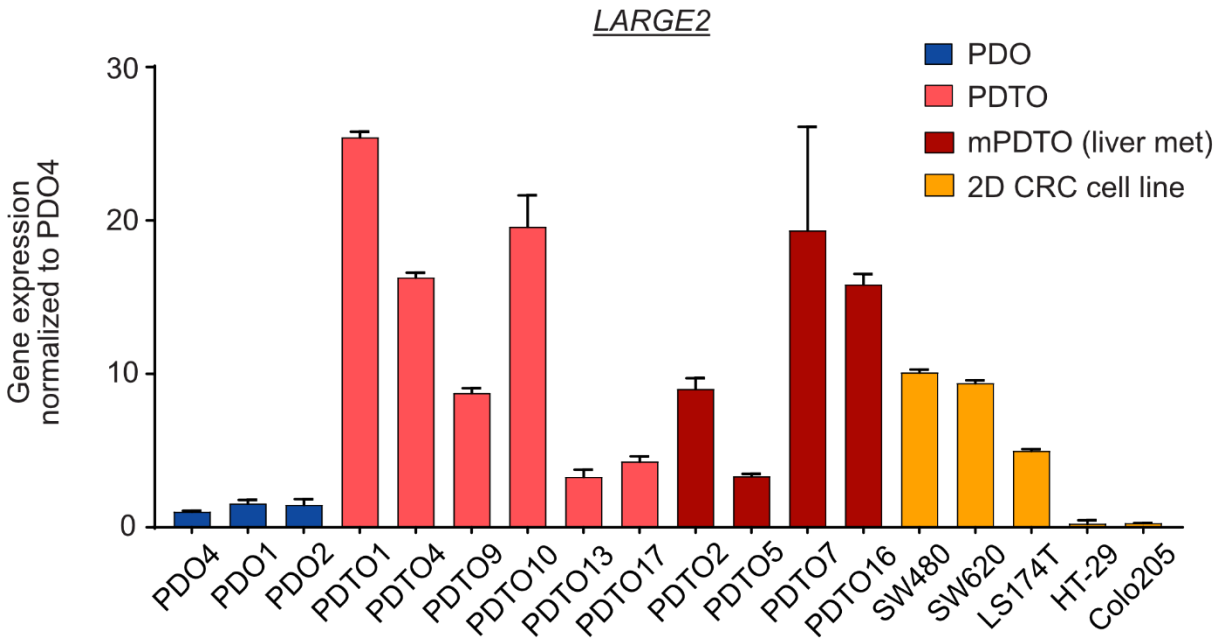


Figure 43: LARGE2 expression in primary and liver metastatic CRC cells. qRT-PCR analysis of LARGE2 in PDOs, PDTOs and CRC cell lines. PDOs: colonic organoids (blue), PDTOs: primary tumor organoids (light red), mPDTOs: liver metastatic tumor organoids (dark red), CRC CL: CRC cell line (yellow).

partial hepatectomy. To approximately rank the level of *LARGE2* expressed in tumor organoid cultures, we compared them to three PDO lines, maintained in Wnt-ligands rich (=WREN) culture medium. We also compared the CRC cell lines previously used in this project.

Concordant with data derived from TCGA, we observed overall elevated *LARGE2* mRNA levels in cancer tissue-derived organoids when compared to normal mucosa-derived PDOs, even if the gene expression of *LARGE2* in PDTOs appeared to be more heterogeneous. This might be explained by different levels of intrinsic Wnt activation in PDTOs. We also noted that *LARGE2* expression in cell lines with very low Wnt-activity (HT-29 and Colo205) was much weaker when compared to PDTOs and PDO gene expression levels (**Figure 43** and **Figure 25**). However, we cannot exclude that the very different culture conditions used for 2D and 3D CRC cells differentially influence expression of LARGE2. Therefore, a direct comparison of gene expression between these two model-systems must be taken with caution.

Since we did not observe any clear patterns in mRNA expression between our primary and metastatic PDTOs, we took advantage of a small cohort (n=12) of non-metastatic (M0) and liver metastatic (M1) primary CRC tissues, provided by Dr. Marlies Michl (LMU Munich). We analyzed RNA samples isolated from FFPE tumor areas via TaqMan™ qRT-PCR. In M1 cases, *LARGE2* gene expression tended to be higher, but we could not find a significant difference when compared to M0 (p=0.083) (**Figure 44A**). Similarly, analysis of RNA-Seq data from TGCA-COAD (M0=365 cases, M1=71 cases) cohort showed no differential expression of *LARGE2* between M0 and M1 (**Figure 44B**), suggesting that mRNA levels of *LARGE2* are not generally elevated or reduced in metastatic CRC tissue compared to non-metastatic tissue.

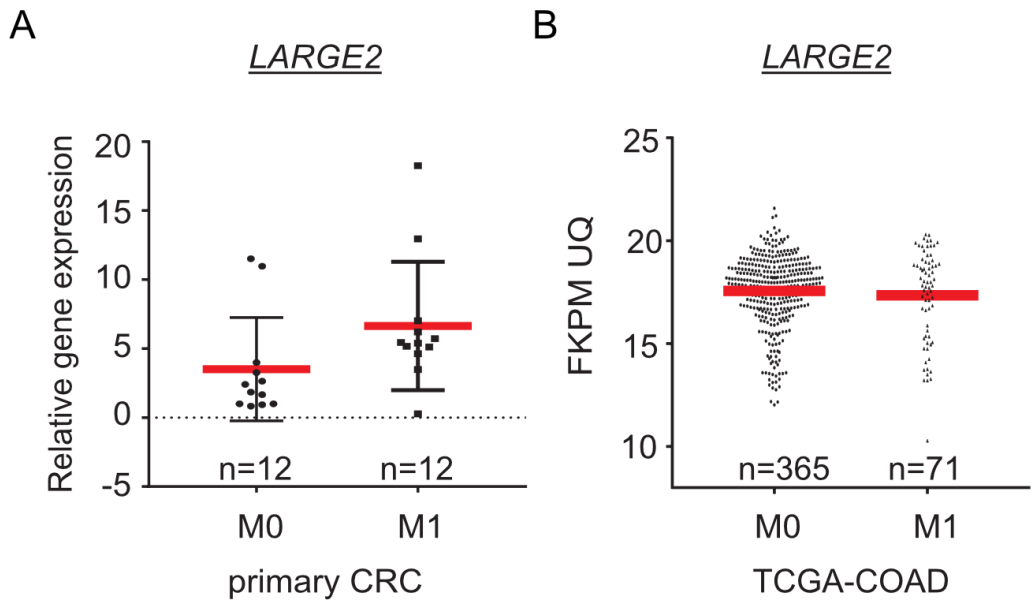


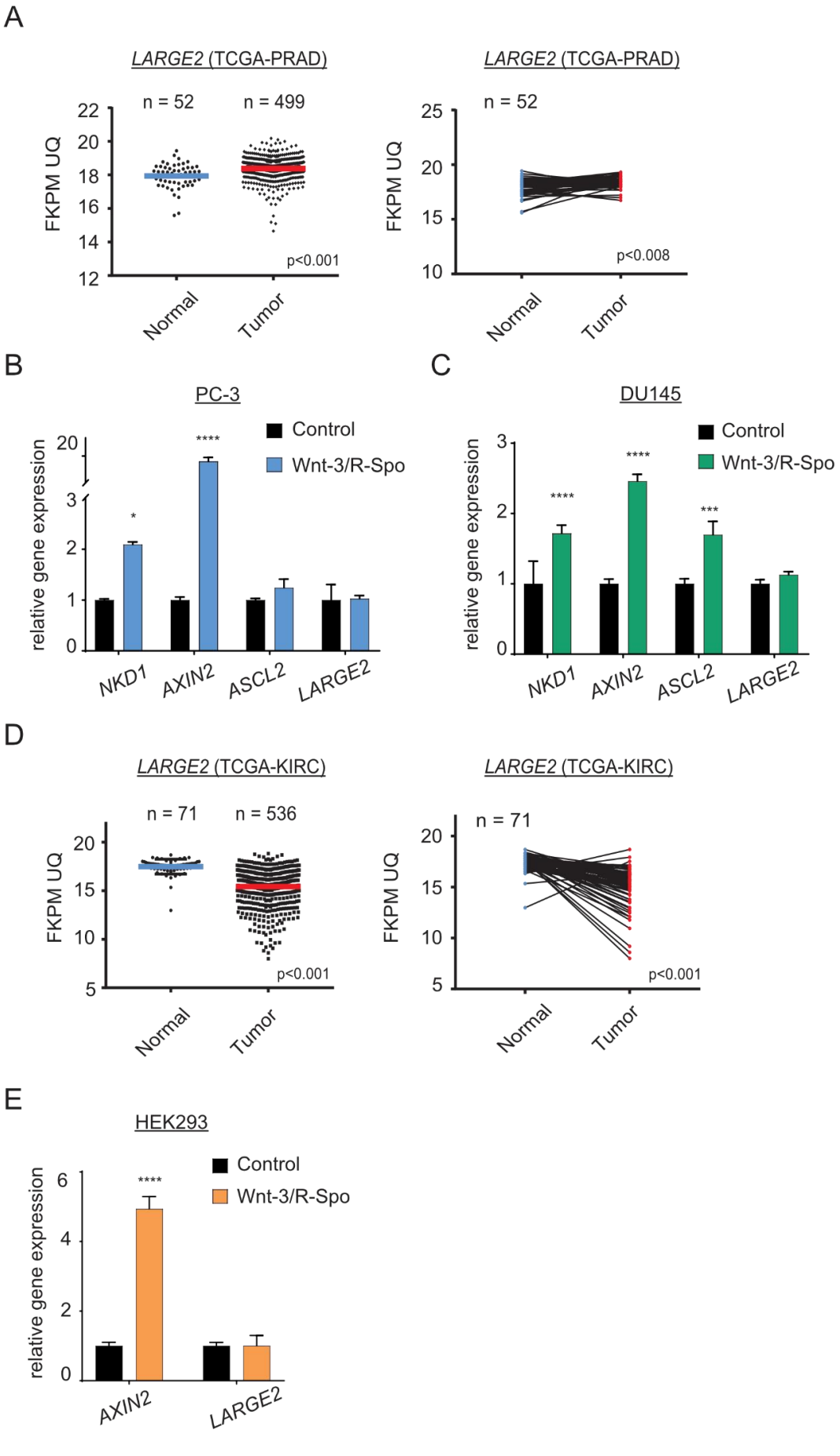
Figure 44: Expression of LARGE2 in M0 and M1 CRC cases. **A**) TaqMan™ qRT-PCR analysis in primary FFPE tissue comparing M0 and M1 CRC (n=12). **B**) TCGA-COAD RNASeq expression analysis comparing M0 and M1 CRC for LAREG2 analysis.

5.10. *LARGE2* expression in kidney and prostate cancer

As has been reported in previous studies, *LARGE2* is the main regulator of α-DG glycosylation in prostate (PRAD) and clear cell renal cell carcinoma (KIRC) (Esser *et al.*, 2013; Miller *et al.*, 2015). In contrast to CRC, our expression analysis in TCGA cohorts revealed that *LARGE2* is only mildly upregulated in PRAD and even downregulated in KIRC compared to normal tissue (**Figure 45A,D**). Contrary to CRC, genetic alterations which activate Wnt signaling, do only occur in a smaller fraction of these tumor entities (35% in PRAD and 7% in KIRC, compared to >90% in CRC) (Sanchez-Vega *et al.*, 2018).

To further address whether Wnt activation is able to trigger *LARGE2* expression in a kidney or prostate cancer cell background, we treated PC-3 and DU145 prostate cancer cells and HEK293 embryonic kidney cells with conditioned media, containing Wnt-3a and R-Spondin, in order to increase their Wnt activity as has been described previously (Kim *et al.*, 2008; Lu *et al.*, 2009). We could observe that unlike other Wnt targets, such as *AXIN2*, *NKD1* and *ASCL2, LARGE2* was not upregulated upon Wnt activation in these cell lines (**Figure 45B,C,E**). We cannot exclude that *LARGE2* expression is responsive to Wnt-stimulation in other prostate and kidney cancer cell lines, but our results indicate that the Wnt-/LARGE2/α-DG pathway might be tumor type-specific.

Figure 45 (see next page): LARGE2 gene expression is not elevated in KIRC and PRAD cohorts. A) LARGE2 gene expression analysis on TCGA-PRAD dataset, comparing normal and tumor samples (left panel) and matched normal und tumor tissues (right panel). **B,C**) qRT-PCR analysis of the indicated genes in prostate cancer cell lines PC-3 and DU145 after treatment with Wnt-3a and R-Spondin conditioned medium for 32 hours. **D**) *LARGE2* expression analysis on TCGA-KIRC dataset, comparing normal and tumor samples (left panel) and matched normal und tumor tissues (right panel). **E**) qRT-PCR analysis in HEK293 embryonic kidney cells after treatment with Wnt-3a & R-Spondin medium for 32 hours. Shown are mean values ±SD (n=3). * p<0.05, *** p<0.001, **** p<0.0001.


A
LARGE2 (TCGA-PRAD)
n = 52
n = 499
FKPM UQ
p<0.001
Normal
Tumor
LARGE2 (TCGA-PRAD)
n = 52
FKPM UQ
p<0.008
Normal
Tumor
B
PC-3
Control
Wnt-3/R-Spo
relative gene expression
NKD1
AXIN2
ASCL2
LARGE2
C
DU145
Control
Wnt-3/R-Spo
relative gene expression
NKD1
AXIN2
ASCL2
LARGE2
D
LARGE2 (TCGA-KIRC)
n = 71
n = 536
FKPM UQ
p<0.001
Normal
Tumor
LARGE2 (TCGA-KIRC)
n = 71
FKPM UQ
p<0.001
Normal
Tumor
E
HEK293
Control
Wnt-3/R-Spo
relative gene expression
AXIN2
LARGE2

5.11. Colorectal tumor organoids show a heterogenous pattern of α-DG O-glycosylation

In order to better address the status of α-DG glycosylation in our primary and metastatic PDTO panel, we analyzed 9 PDTO lines via immunoblot on electrophoretically separated WGA-purified protein lysates. In 7 out of 9 PDTOs we saw a clear signal for IIh6c4 reactivity, the molecular weight (MW) of O-glycosylated α-DG varying between 130 kDa and 250 kDa (**Figure 46A-C**). PDTO1, PDTO16 and PDTO17 showed signals at highest MW compared to other, even though the signal of PDTO16 is rather weak compared to PDTO17 (**Figure 46**C) PDTO13 and PDTO2 only showed a weak signal for α-DG glycosylation but at a physiological MW around 130 kDa (**Figure 46B**), while PDTO4 is negative for any IIh6c4 staining (**Figure 46**A). PDTOs 7 and 10 show a strong signal for O-glycosylated α-DG, distributed between <100 kDa and ~130 kDa.

For the PDTOs 1 and 10, as well as the metastatic PDTO16, we observed high levels of *LARGE2* mRNA and distinct high MW O-glycosylation of α-DG. In PDTO13, *LARGE2* expression was lower when compared to other PDTO samples, and here the level of α-DG glycosylation was low as well, which is in accordance with a LARGE2 dependent amount of α-DG glycosylation in these samples. Interestingly, this correlation was not observed in all analyzed cases. The PDTOs 2, 4 and 7 express high levels of *LARGE2*, but showed no or near physiological levels glycosylation, whereas PDTOs 5 and 17 displayed high MW variants of α-DG O-glycosylation even though they expressed only mildly elevated *LARGE2* mRNA (**Figure 43** and **Figure 46**). Hence, other factors besides LARGE2 might affect the status of glycosylated α-DG in a fraction of colorectal tumors.

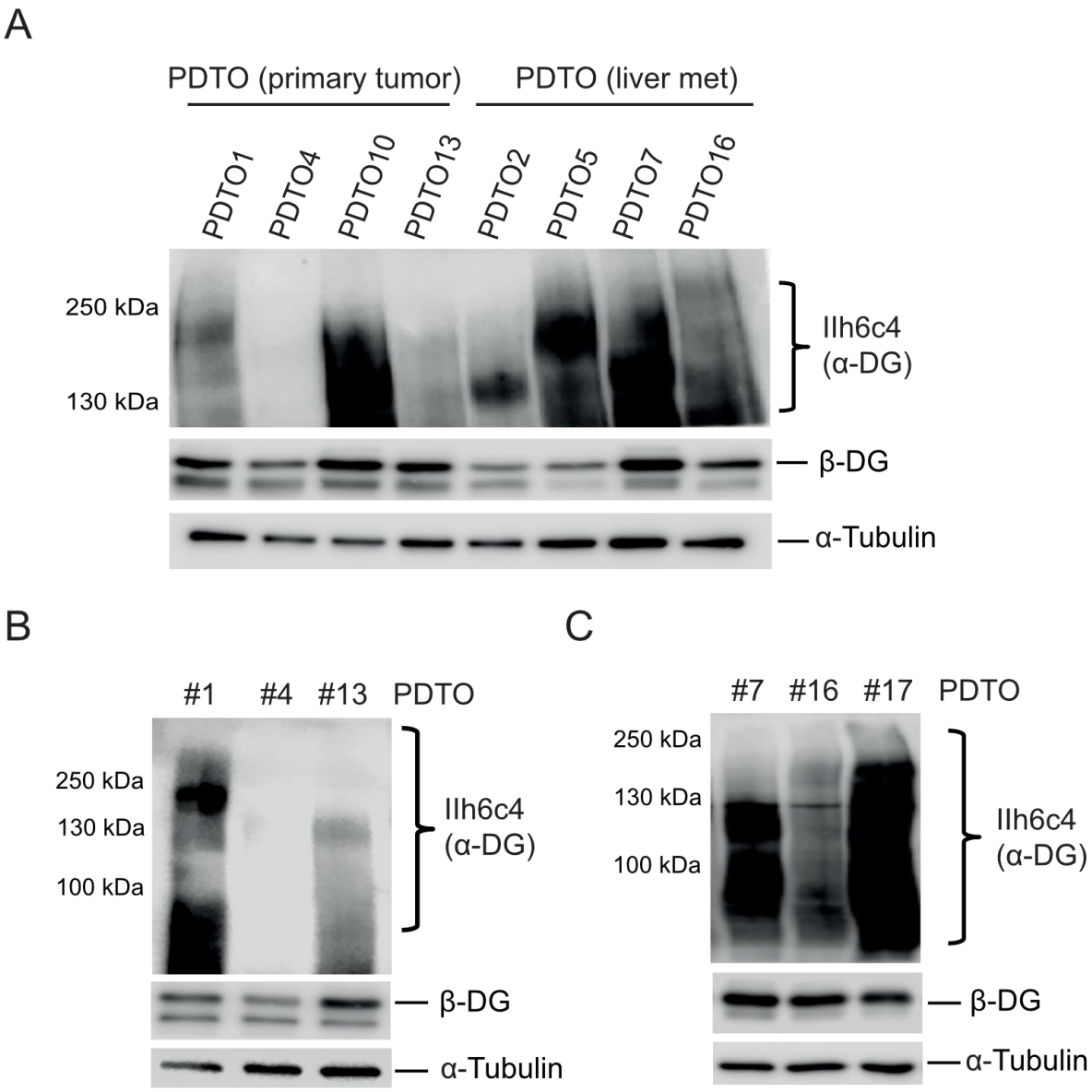


Figure 46: Immunoblot analysis of CRC PDTOs. **A**) Immunoblot analysis of primary and metastatic CRC tumor organoids for O-glycosylated α-DG. **B**, **C**) Additional assessment of primary and liver metastatic PDTOs regarding their status of α-DG glycosylation. All lysates for glycosylation analysis were enriched via WGA affinity purification, β-DG and α-Tubulin were detected on WCL.

To compare the α-DG O-glycosylation status of our PDTO cultures to tumor tissue, we performed IHC analysis for O-glycosylated α-DG on sections of FFPE tumors, from which the PDTOs had been derived. Primary tumors PDTO 1, 10 and 17 show strong membranous reactivity (**Figure 47A**), concordant with the immunoblot analysis and staining was absent in PDTO4 as expected. Similar results were obtained when staining the FFPE sections of the metastatic tumor samples (**Figure 47B**), where a weak staining was observed in PDTO2 and 7, and stronger reactivity for PDTO5 and

16.

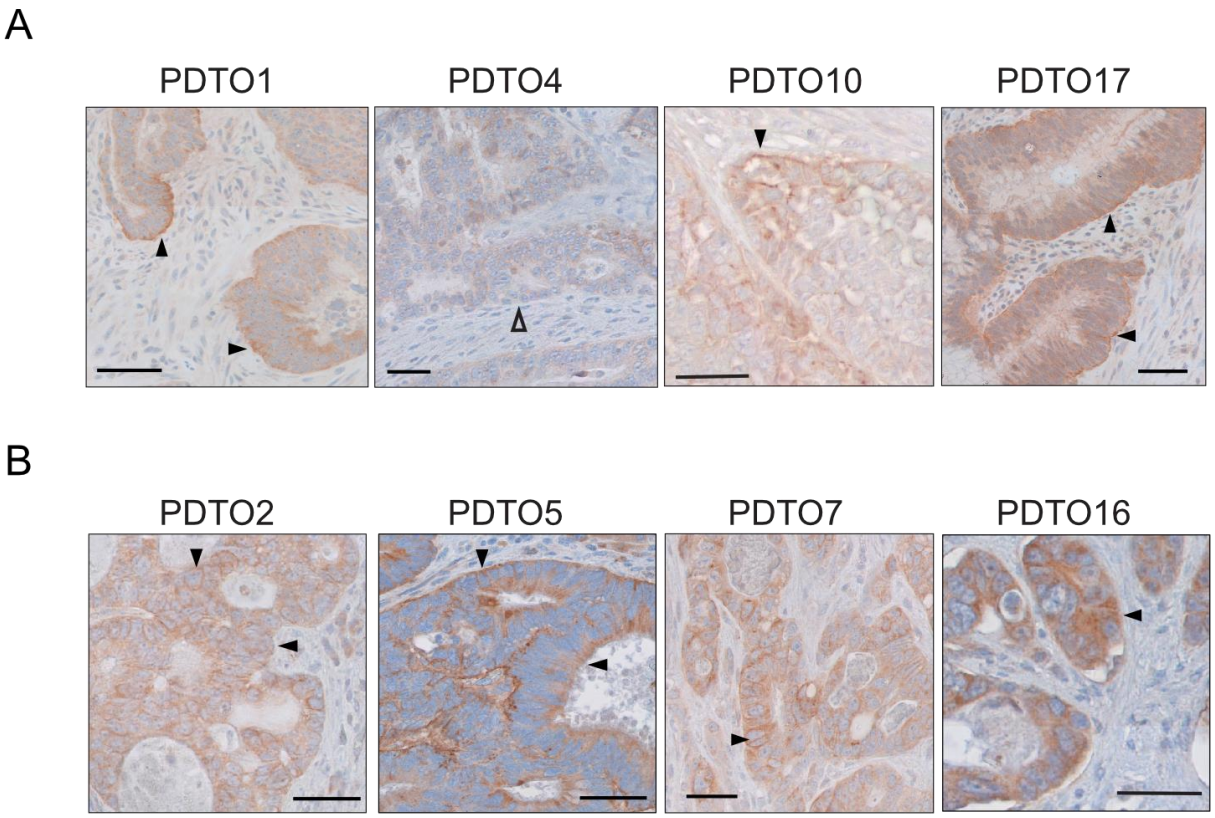


Figure 47: Immunohistochemistry staining for glycosylated α-DG in FFPE tumor sections. A) Selection of analyzed primary tumors and **B**) cohort of metastatic tumor samples embedded in FFPE. Tissues were stained with IIh6c4 antibody.

Our data suggest that LARGE2-dependent matriglycan formation on α-DG is a common event in primary and metastatic CRC. Nevertheless, alternative cell intrinsic mechanisms, presumably acting upstream of LARGE2 in the α-DG O-glycosylation cascade, which affect the ultimate complexity of the α-DG attached matriglycan structure cannot be excluded.

5.12. Functional α-DG glycosylation in cell-adhesion of liver metastatic CRC

To address the status of O-glycosylated α-DG in a well-characterized model of CRC metastasis to the liver, we used the KM12c and KM12-L4a CRC cell line-model from I. J. Fidler (Morikawa *et al.*, 1988). KM12c, which have poor capability to form metastasis in the liver, showed no detectable glycosylation of α-DG nor binding of

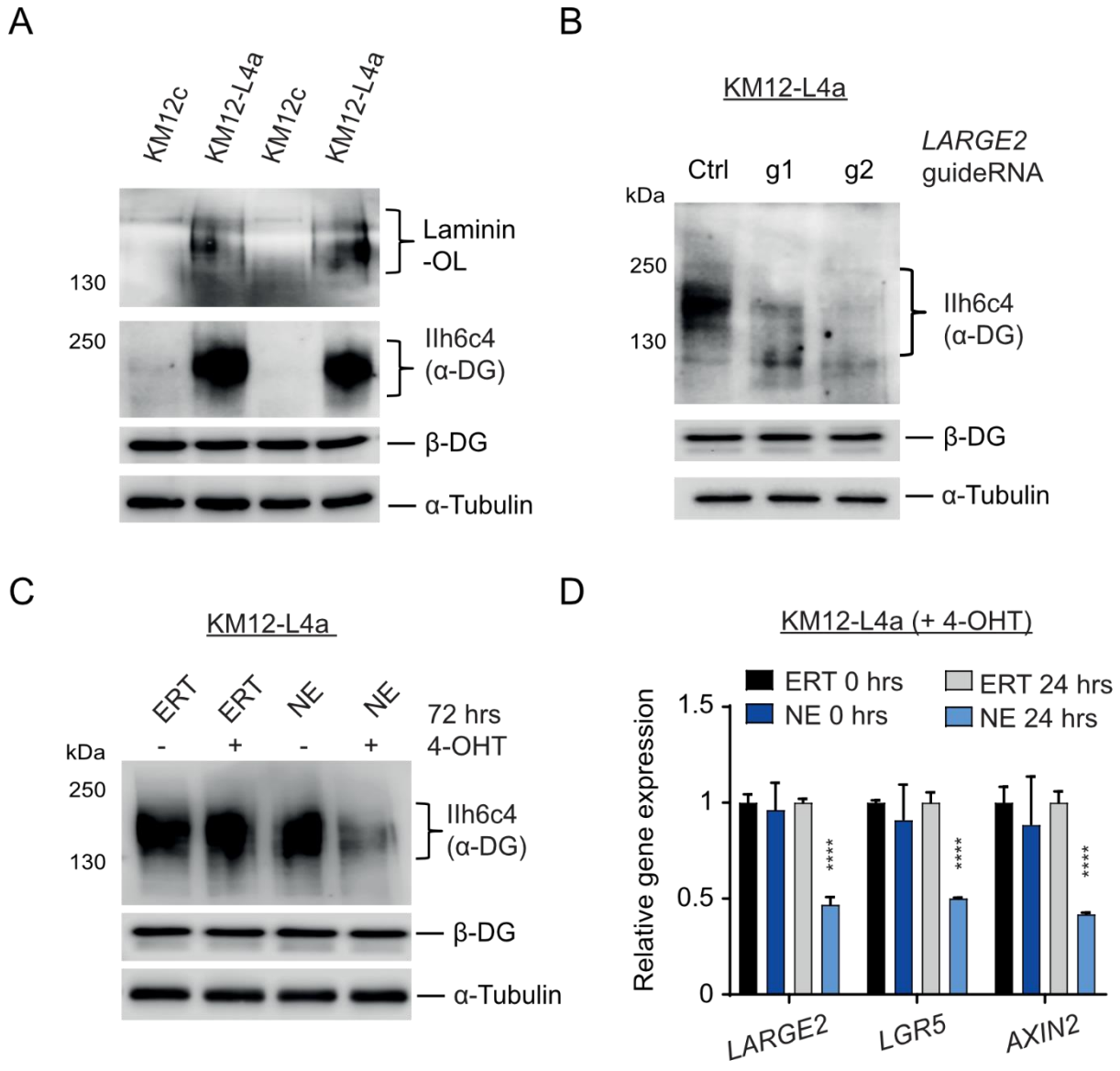


Figure 48: LARGE2 and α-DG glycosylation in a model of liver metastatic CRC cells. **A**) Immunoblot analysis of glycosylated α-DG in KM12c cells and its liver metastatic derivative KM12-L4a. **B**) Analysis of O-glycosylated α-DG after knockout of LARGE2 in CRC cell line KM12L4a **C**) after treatment of KM12-L4a with 4-OHT for 72 hours. α-DG and laminin were analyzed on WGA-purified lysates, β-DG and α-Tubulin were analyzed on WCL. **D**) qRT-PCR of KM12-L4a-NE cells after 0- and 24-hours treatment with 4-OHT compared to ERT controls. **** $p < 0.0001$.

laminin, whereas the liver metastatic derivative KM12-L4a cells displayed high MW of α-DG glycosylation and the capacity to bind laminin (**Figure 48A**). CRISPR/Cas9 knockout of LARGE2 consequently led to loss of α-DG glycosylation in KM12-L4a cells (**Figure 48**B). Additionally, we could observe that the expression of *LARGE2* and the O-glycosylation of α-DG in KM12-L4a cells can be downregulated via acute blockade of Wnt-signaling in KM12-L4a-NE cells after treatment with 4-OHT (**Figure 48C,D**).

Moreover, we investigated the adhesive capacity of KM12-L4a cells, stably transduced to express a luciferase (luc), in dependence on *LARGE2* integrity and hence the presence of high MW matriglycan on a-DG. The cells wild-type or KO for LARGE2 were seeded on a confluent layer of HMEC-1 endothelial cells and the adhesion capacity was determined after 0, 45 and 90 minutes. We observed that the loss of O-glycosylated α-DG interferes with the cell ability to rapidly attach to the endothelial layer (**Figure 49**).

From the results of these analyses, we can conclude that the liver metastasis derivative CRC cell line KM12-L4a has higher O-glycosylation of α-DG compared to the parental cell line KM12c, and that abolishing the LARGE2/α-DG axis reduces the adhesive ability of metastatic KM12-L4a cells. In a published study from Jones et al., it was reported that KM12-L4a cells show enhanced matrix adhesiveness compared to KM12c (Jones *et al.*, 2002). Our data hint to a potential role of O-glycosylated α-DG in the adhesion of circulating CRC cells to the laminin-rich ECM of blood vessels. Future studies should therefore determine the biological consequence of enhanced

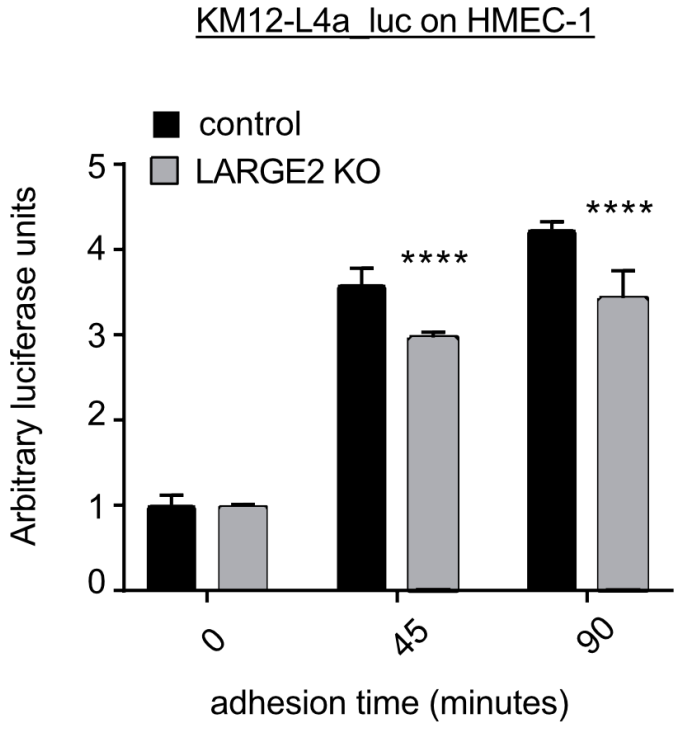


Figure 49: LARGE2 dependent adhesion of KM12-L4a to endothelial cells. Adhesion of KM12-L4a cells, stably transduced with a luciferase (luc)-encoding lentivirus and either 1122 wild-type or LARGE2 KO, was quantified at the indicated time points by measuring luc activity. T=0 represents the baseline control and was also used for sample normalization. Results are shown as mean ± SD (n=3). **** p<0.0001.

laminin-anchorage in the context of high Wnt/LARGE2/α-DG signaling for the liver colonization capacity of CRC cells.

5.13. *LARGE2* and α-DG in CRC cell migration and invasiveness

O-glycosylated α-DG has been found to interfere with cellular migration and hence cancer cell invasiveness in renal, prostate and breast carcinoma cells (Bao *et al.*, 2009). Therefore, we tested the migratory capacity of CRC cell lines LS174T and SW620 via migration assay using laminin-111 coated transwell cell culture chambers. We could observe that upon knockout of LARGE2 and the consequential loss of α-DG glycosylation, the *in vitro* cell migration capacity was significantly improved (**Figure 50A,B**). Concordantly, the formation of higher MW matriglycan on α-DG upon conditional ectopic expression of *LARGE2* in HT-29 cells that was confirmed above (as observed in **Figure 30A**), interferes with cell migration in the same experimental setting (**Figure 50C**). Taken together, this data point to a cell migration/invasion-inhibitory role of the Wnt-driven LARGE2/α-DG signaling pathway

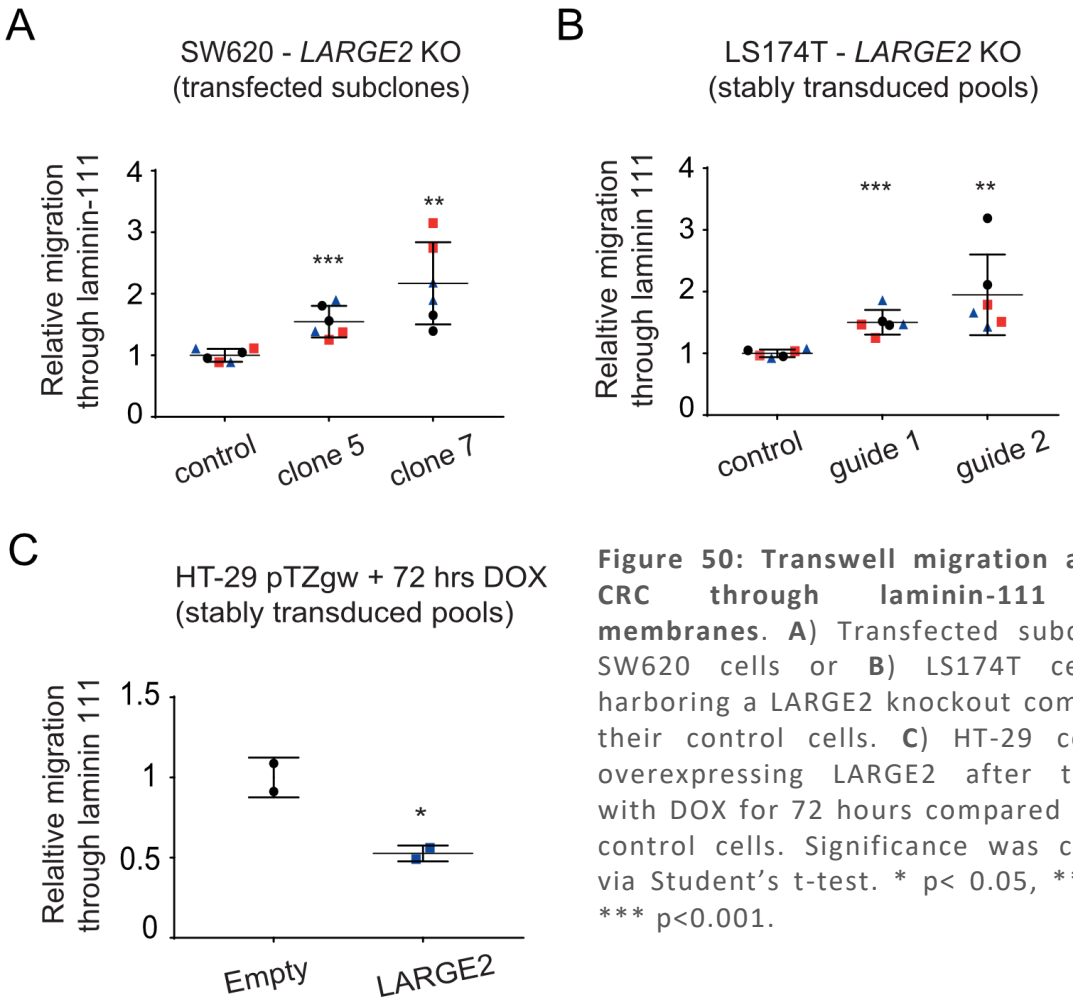


Figure 50: Transwell migration assays of CRC through laminin-111 coated membranes. **A**) Transfected subclones of SW620 cells or **B**) LS174T cell pools harboring a LARGE2 knockout compared to their control cells. **C**) HT-29 cell pools overexpressing LARGE2 after treatment with DOX for 72 hours compared to empty control cells. Significance was calculated via Student's t-test. * p< 0.05, ** p<0.01, *** p<0.001.

in CRC cells. This suggests that LARGE2-dependent α-DG O-glycosylation might act in a rather tumor suppressive way, e.g. at tumor invasion fronts where tumor dissemination occurs.

Since the *in vitro* assay used here does not reproduce all facets of an *in vivo* interaction between CRC cells and the tumor microenvironment (TME) during tumor progression, we set out to compare the levels of O-glycosylated a-DG between stroma-facing tumor areas and their central mass. To achieve this, we performed IHC staining for glycosylated α-DG on a FFPE tissue microarray (TMA) cohort of 28 locally invasive colorectal tumors (stage T3), showing either the main tumor or the patient-matched tumor invasion front. For quantification, we developed a scoring system (0=no staining, 1=weak cytoplasmic plus membranous staining, 2=moderate cytoplasmic plus membranous staining, 3=strong basal cytoplasmic plus

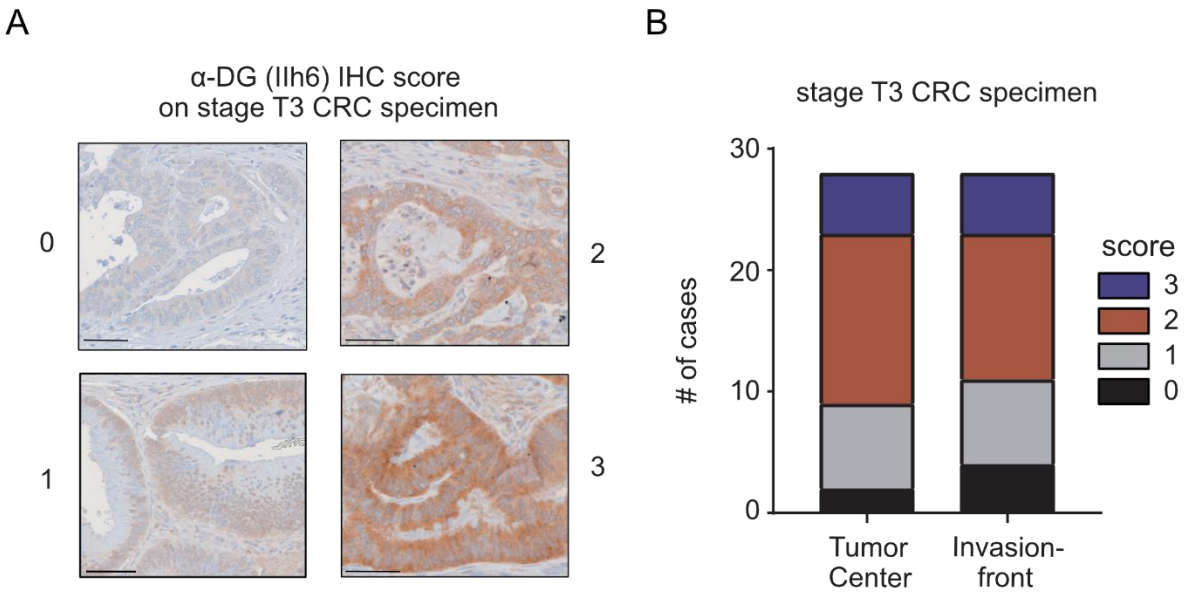


Figure 51: O-glycosylation of a-DG is a frequent feature of T3 CRC specimen. A) Exemplary scoring of α-DG membrane staining intensity in main tumor areas and invasion fronts of stage T3 CRC cases (n=28). **B**) Distribution of the membrane staining scores in main tumor areas and invasion fronts of the 28 stage T3 cases. Scalebar represents 50 µm for 0-2, 40µm for 3.

membranous staining (**Figure 51A**).

Our analysis revealed that 9 out of 28 main tumor area showed no or weak staining (score 0 or 1) for glycosylated α-DG, while moderate to strong staining was observed in 19 cases, which represents the majority. Moreover, the tumor invasion fronts show a similar intensity distribution compared to the central tumor area (**Figure 51B**), suggesting that expression of *LARGE2* and consequently O-glycosylation of α-DG is

not exclusively restricted to the tumor front and also occurs in the tumor center. Importantly, the majority of invasive tumors analyzed here was positive for O-glycosylated α-DG at the invasion front, suggesting that the Wnt/LARGE2/α-DG axis does not *per se* prevent local CRC spread and presumably affects tumor dissemination in a context-dependent manner.

A previous study showed that in the context of prostate cancer, *LARGE2* expression is repressed by the EMT-mediating transcription factors SNAIL and ZEB, subsequently decreasing α-DG glycosylation (Q Huang *et al.*, 2015). However, in colorectal cancer *SNAIL* is positively regulated by elevated Wnt signaling and supports partial EMT (Brabletz *et al.*, 2005; Horvay *et al.*, 2011) of cells high in nuclear β-Catenin at tumor invasion front. Indeed, we found two independent sets of EMT-related genes enriched among the genes positively correlated with LARGE2 in CRC by performing GSEA analysis (**Figure 52**).

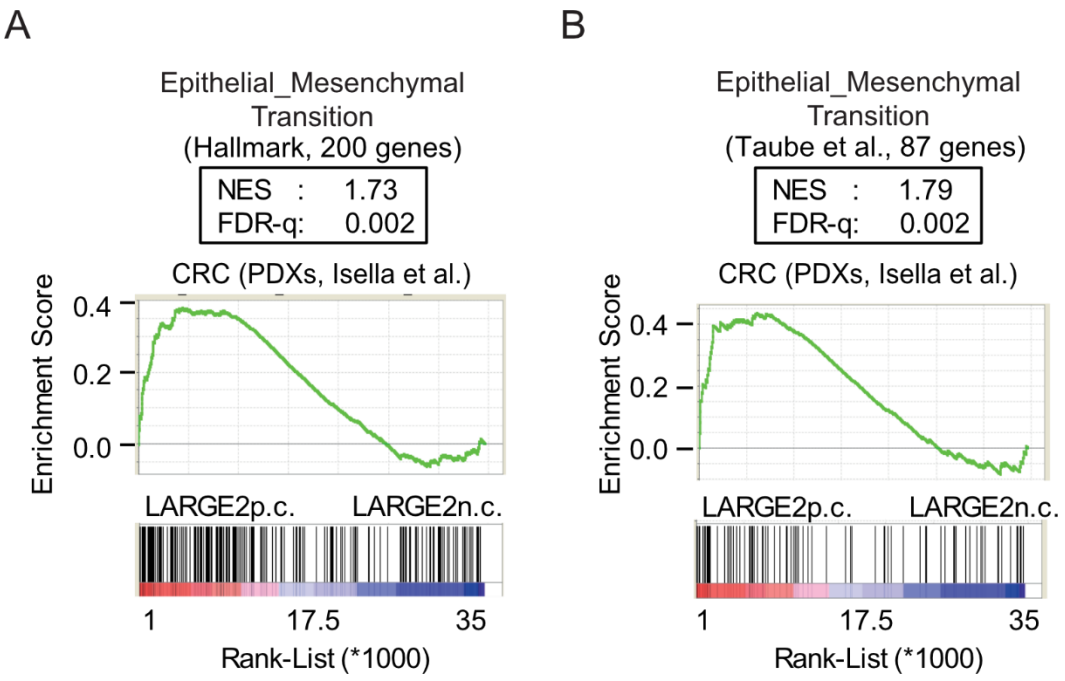


Figure 52: GSEA on LARGE2 using Epithelial-Mesenchymal Transition (EMT) Hallmark gene sets. Analysis was performed on PDX dataset from Isella et al. using the Hallmark gene sets from **A**) the Broad institute and **B**) from Taube et al. NES: normalized enrichment score, FDR-q: False discovery rate q-value.

From our data, we hypothesize that strong Wnt activity in migrating CRC cells might interfere with SNAIL and/or ZEB1-mediated down-modulation of LARGE2 in this disease context. Therefore, low abundance or loss of O-glycosylated α-DG in

CRC grounds either on low/mild Wnt/LARGE2/α-DG signaling or, alternatively, on LARGE2-independent molecular alterations.

6. Discussion

6.1. The O-glycosyltransferase LARGE2 is regulated by Wnt signaling

Development and homeostasis of tissues depend on a sophisticated interplay between factors necessary for cell proliferation, differentiation, migration, cell-matrix adhesion and apoptosis (Logan and Nusse, 2004). Even before the critical role of Wnt signaling in gut development and tissue homeostasis was uncovered, the relevance of Wnt in colorectal cancer had been recognized (Munemitsu *et al.*, 1995). Independent observations about somatic mutations in the tumor suppressor gene *APC* in hereditary colorectal cancer (CRC) led to the conclusion, that inactive APC triggers constitutively active Wnt signaling in CRC cells (Kinzler *et al.*, 1991; Rubinfeld *et al.*, 1993; Korinek *et al.*, 1997). As aberrant activation of Wnt was found to be the initiative event responsible for the initiation of most sporadic CRC and familial adenomatous polyposis (FAP) (Powell *et al.*, 1992), multiple studies aimed to identify disease relevant genes either directly or indirectly driven by Wnt. As part of this thesis, we set out for identification of so far unknown genes triggered by cell-intrinsic activation of oncogenic Wnt signaling in CRC cells.

Many colorectal cancer cells retain one truncated APC allele with a mutational cluster region (MCR) encoding two or three 20-aa repeats, necessary for functional binding and regulation of β-Catenin, whereas the second allele is often much shorter or lost completely (Yang *et al.*, 2006; Vijaya Chandra *et al.*, 2012). The “just right” Wnt hypothesis (Albuquerque *et al.*, 2002) states that APC truncating mutations, tolerating an intermediate activity of β-Catenin and Wnt, are more favorable for tumor progression, compared to maximal Wnt activation due to the loss of all β-Catenin binding sites in shorter truncated APC versions, as for example implemented in the widely used APCmin mouse model (Moser, Pitot and Dove, 1990; Su *et al.*, 1992; Pollard *et al.*, 2009). Complete loss of Apc functionality in mice abolishes cell migration and enhanced levels of cell death (Sansom *et al.*, 2004), which is likely because of exaggerated oncogenic stress in this scenario. Nevertheless, experimentally creating an acute complete loss of APC, which then leads to rapid stabilization and nuclear accumulation of β-Catenin, is an elegant setting for the identification of oncogenic Wnt signaling-driven genes. With the aim to overstimulate

the ligand-independent, intrinsic Wnt signaling activity and investigate the aftermath on CRC gene expression, we chose a short hairpin RNA (shRNA) to downmodulate *APC* mRNA and protein in a CRC cell line with low oncogenic Wnt background activity. In studies preceding our own work, different approaches were used to identify genes commonly regulated by Wnt/β-Catenin signaling in the background of CRC. The use of β-Catenin silencing short interfering (si)RNAs, as reported by Herbst et al., resulted in deregulation of more than 4000 genes (Herbst *et al.*, 2014). Even though, β-Catenin plays a significant role in the Wnt signaling pathway, it is also involved in the regulation of cell-cell adhesion via cadherin-catenin complexes (Nelson, 2008). Notably, expression loss of β-Catenin does not occur in CRC. Therefore, we decided on a different, more disease-representative setting for this thesis in order to decipher the set of genes de-regulated downstream of oncogenic Wnt signaling.

As expected, upon down-modulation of *APC* in HT-29 cells we observed an induction of bona-fide Wnt target genes like *NKD1* or *LGR5* and a gene encoding for the bi-functional glycosyltransferase LARGE2 (GYLTL1B). Long-term culturing of the cells resulted in heavily decelerated growth (data not shown). At a molecular level, more than the “just right” Wnt activity leads to extensive nuclear accumulation of β-Catenin, consequentially increasing oncogenic stress, leading to enhanced proliferation, increased apoptosis or aberrant mitosis due to over activation of downstream signaling pathways (Kim *et al.*, 2000; Sansom *et al.*, 2004). The outcome is arrested cell proliferation and programmed cell death, concordant to what we observed in long-term cultures of HT-29 cells with permanently silenced APC.

Accumulation of β-Catenin in the nucleus due to non-functional APC or stabilizing mutations in the β-catenin encoding gene *CTNNB1* itself prevent its destruction in the cytoplasm and lead to augmented formation of nuclear β-Catenin/TCF complexes (Morin *et al.*, 1997; Kolligs *et al.*, 1999). In mammals there are three known members of the TCF family present: TCF-1 (van de Wetering *et al.*, 1991), TCF-3 and TCF-4 (Castrop, Van Norren and Clevers, 1992). Whereas *TCF-1* and *TCF-3* mRNA is expressed during early embryonic development, *TCF-4* (today official designated and from here on referred to as *TCF7L2* for the human tissue background) expression is restricted to the intestinal epithelium (Vladimir Korinek, Barker, Willert, *et al.*, 1998) and interacts with β-Catenin to drive gene expression (Korinek *et al.*, 1997). Analysis

of the intestine from mouse models harboring a full body Tcf-4 knockout revealed reduced proliferative activity in the small intestine leading to an extensively stretched epithelial layer, which ultimately causes neonatal lethality (V Korinek *et al.*, 1998). TCF7L2 is essential to maintain the undifferentiated phenotype in intestinal epithelial cells but also sustains a crypt-progenitor-like phenotype in colorectal cancer. This finding from van de Wetering et al. proved that the inhibition of β-Catenin/TCF complex formation, via the expression of a dominant-negative form of TCF (dnTCF) factors, blocks cell intrinsic Wnt-activity in CRC cells, resulting in cell cycle arrest and differentiation (Van de Wetering *et al.*, 2002). Moreover, they observed that the Wnt driven genetic program in CRC includes several genes expressed in the proliferative compartment of normal colonic crypts, such as *EPHB2*, *CD44*, *c-MYC* (van de Wetering *et al.*, 2002). We could show that a TCF7L2 canonical binding motif "CTTTGATC" (Hatzis *et al.*, 2008) in the first intron of *LARGE2* is essential for binding of TCF7L2 to this region and critical for Wnt-driven induction of *LARGE2*. We therefore conclude that LARGE2 is a direct target of the canonical Wnt signaling pathway (Kikuchi *et al.*, 2011).

Several studies demonstrated that the balance between cell proliferation and intercellular adhesion in epithelial tissue is regulated by a network of Wnt/β-Catenin signaling, E-cadherin adhesion and protein N-glycosylation (Tang *et al.*, 2012; Vargas *et al.*, 2016). Wnt ligands themselves are a family of secreted glycoproteins and comprise 19 members in mice and men. They represent hydrophobic proteins with a number of well-described N-glycosylation sites (Gross and Boutros, 2013). Other studies have identified Wnt-regulated N-glycosyltransferases, which play a potential role in cell adhesion and metastasis. Expression of *DPAGT*-1 is upregulated upon binding of the β-Catenin/TCF complex to the promotor and enhances the N-glycosylation status of E-cadherin, which in turn controls cell-cell adhesion (Sengupta *et al.*, 2013). Highly N-glycosylated E-cadherin acts as a suppressor of metastasis in melanoma cells (Yoshimura *et al.*, 1996), even though the tumor-suppressive function of E-cadherin seems to depend on the specific modified N-glycosylation site (Carvalho *et al.*, 2016). The Wnt-regulated glycosyltransferase ST6GAL1 is overexpressed in gastric, colonic, and pancreatic adenocarcinomas. By synthesis of sialic acids on N-glycans of TNFR1, ST6GAL1 inhibits homeostatic apoptosis (Alexander *et al.*, 2018). The association of *ST6GAL1* gene expression with the Wnt

target genes *LGR5* and *AXIN2* goes in line with the positive correlation of the Wnt gene program or the ISC gene signature and LARGE2, as we revealed during my Ph.D. thesis. To our knowledge, this is the first time a direct connection between canonical Wnt signaling and O-linked glycosylation in human colonic epithelium and CRC has been demonstrated.

6.2. Regulation of LARGE2 and its role in functional O-glycosylation of α-Dystroglycan in different tissues

LARGE2 was previously identified to have a high functional similarity to its homolog LARGE (now referred to as LARGE1), a glycosyltransferase responsible for the post-translational synthesis of so-called matriglycan on α-Dystroglycan (α-DG) (Kanagawa *et al.*, 2004; Fujimura *et al.*, 2005). Albeit the proteins differ substantially in their transcript size, both are localized within the Golgi apparatus and, consistent with their function in O-glycosylation, they form part of a highly conserved gene family. Both genes are in principle expressed in most tissues. However; *LARGE1* mRNA abundance is higher in heart, brain, skeletal muscle, stomach and pancreas, while relatively stronger *LARGE2* expression has been demonstrated for placenta, pancreas, mammary glands, kidney and prostate tissues (Brockington *et al.*, 2005; Fujimura *et al.*, 2005; Grewal *et al.*, 2005). In contrast to previous results on small intestine, where *LARGE1* was prominently expressed, (Fujimura *et al.*, 2005), we observed substantially higher *LARGE2* levels in all analyzed hCoSCs organoids and PDTO samples in vitro, as well as in CRC cell lines. Different approaches to determine *LARGE1* expression levels via Taqman probe-based qPCR or by using multiple probes for CYBR green-based qRT-PCR resulted in Ct values of ~31-34 (data not shown), which is indicative of a very low *LARGE1* mRNA abundance in human colonic epithelial cells and CRC. Since LARGE1 could not compensate the observed loss of O-glycosylation on α-DG upon LARGE2 depletion, and LARGE2 ectopic expression alone was sufficient for functional α-DG in this tissue context, we assume that LARGE2 provides the driving enzymatic activity for matriglycan synthesis in the human colon. According to previous publications, LARGE2 cannot compensate for the lack of LARGE1 in MDC1D patients (Fujimura *et al.*, 2005). Hence, we suggest that the major catalyst of matriglycan synthesis an α-DG is

dependent on the tissue background. Furthermore, we did not observe any influence on *LARGE1* expression upon Wnt-activation or Wnt-blockade, indicating that albeit the functional similarity of both LARGE glycosyltransferases, they are regulated by different signaling pathways.

The capability of the extracellular Dystroglycan subunit α-Dystroglycan to bind LG-domain containing proteins embedded in the ECM and to communicate this anchorage to the cytoskeleton via the dystrophin complex, is highly dependent on the status of post-translational α-DG O-glycosylation (Kanagawa *et al.*, 2004). Both LARGE1 and LARGE2 glycosyltransferases can mediate this modification *in vivo* (Inamori *et al.*, 2014). The loss of LARGE1 in muscle cells results in the loss of α-DG glycosylation and laminin binding capacity of cells (Kanagawa *et al.*, 2004), but the functionality of α-Dystroglycan can be rescued by overexpression of *LARGE1* as it was shown in neural stem cells (Zhang, Zhang and Hu, 2011). Similarly, former studies in a mouse myoblasts cell line observed increased laminin-binding activity after overexpression of *LARGE2* (Grewal *et al.*, 2005). However, in human embryonic kidney cells, overexpression of *LARGE2* was found to support the synthesis of matriglycan more effectively than ectopic *LARGE1* (Fujimura *et al.*, 2005).

LARGE2 was first implicated in cancer a few years ago, when Esser et al. investigated its role in functional glycosylation in prostate and pancreatic cancer. Their findings suggest that the decreased functional glycosylation of α-Dystroglycan, frequently observed in pancreatic and prostate cancer, is due to reduced expression levels of *LARGE2* (Esser *et al.*, 2013). Similar observations have been made in epithelial cancers from breast, lung and cervix where the loss of LARGE resulted in diminished laminin binding to the ECM (de Bernabé *et al.*, 2009). Rescue experiments, which restored higher levels of *LARGE2* expression, resulted in functional hyperglycosylation of α-DG (Esser *et al.*, 2013). We could confirm that the complexity and laminin-binding capacity of O-glycosylated α-DG in CRC can be modulated through the ectopic expression of *LARGE2*. The BRAF-mutated (V600E) and chromosomally instable (CIN) HT-29 CRC cell line harbors very long (three 20-aa repeats) truncated versions of APC, a consequentially low Wnt activity and *LARGE2* expression, accompanied by α-DG hypoglycosylation. Congruent with results from other investigators, who overexpressed either *LARGE1* or *LARGE2* in brain and prostate (Zhang, Zhang and Hu, 2011; Esser *et al.*, 2013), ectopic

expression of *LARGE2* in HT-29 CRC cells gave rise to high molecular glycan structures on α-DG and dramatically enhanced its laminin-binding capacity. More important, we could demonstrate increased endogenous *LARGE2* expression, matriglycan formation on α-DG, and enhanced laminin binding after triggering intrinsic Wnt signaling either via shRNA mediated APC silencing or, alternatively, via ectopic expression of oncogenic CTNNB1-S33Y in this cellular background. This demonstrates that oncogenic Wnt signaling pathway-triggered expression of endogenous LARGE2 is able to enhance α-DG functionality, and this highlights that our observations do not represent an artifact of higher-than physiologically relevant ectopic *LARGE2* levels.

Just recently, Inamori and colleagues identified the heparan sulfate proteoglycan glypican-4 (GPC4) as a novel substrate for functional modification of ectopically expressed Large2 besides α-DG in mouse embryonic stem cells (ES cells) (K.-I. I. Inamori *et al.*, 2016). The IIH6-reactive cell surface protein was detected by mass spectrometry. Overexpression of *Large2* in mouse ES cells, lacking either Dag1, Pomt1 or Fktn, resulted in functional protein-modification of Gpc4 and laminin-binding ability, suggesting that the Large2-mediated modification of the alternate substrate is independent of O-mannosylation and Fktn-modification in this cellular context. We tried to validate these results by analyzing WGA purified, glycoprotein enriched lysates of HT-29 cells overexpressing *LARGE2* via LC-MS/MS. However, unlike the situation in mouse ES cells, we could not detect GPC4 among the mapped peptides of any sample from the mass spectrometry analysis. Aside from DAG1 we detected LAMA2, encoding for the Laminin subunit alpha 2, which was exclusively present in *LARGE2* overexpression cells. LAMA2 itself only contains N-linked glycosylation sites (source: www.uniprot.org) but through the globular G-domain of the alpha-2 chain it binds to the matriglycan moiety of α-Dystroglycan. We therefore concluded that the high MW O-glycosylation of α-DG synthesized by LARGE2 led to a co-precipitation of LAMA2. We tried to confirm the LARGE2-mediated O-glycosylation of other proteins, some of which were detected exclusively in the *LARGE2* overexpressing samples in LC-MS/MS. By electrophoresis of cell lysates through a gel containing co-polymerized WGA, selectively retarded O-glycosylated proteins can be visualized even if only one or few sugar residues are attached (Kubota, Fujioka and Takekawa, 2017). However, we could verify any additional candidates by using

this technique (data not shown). Even though this could also be due to technical issues, we hypothesize that most other identified candidates are not O-glycosylated themselves, but show high affinity towards highly glycosylated α-DG in suspension. This would also explain why those proteins, in contrast to α-DG, were exclusively detected in the WGA-precipitated samples after overexpression of *LARGE2* and did not show the enormous shift in MW as we had observed for α-DG. Nonetheless, we could assign several of the identified peptides corresponding to DAG1 to the immunoglobulin (Ig)-like domains 1 and 2 of α-DG (see 10.5). These Ig-like domains show a similar structure and amino acid composition as the domains of immunoglobulins, which are indispensable for the recognition of surface molecules and cell adhesion (Williams and Barclay, 1988). A large fraction of the α-DG Ig-Domain specific peptide sequences were only present in low MW-windows of the control cells, but were almost exclusively found in the high MW-fractions derived from the *LARGE2* overexpressing cells. We hypothesize that this size shift was due to increased O-glycosylation of α-DG in this scenario. So far, several studies have identified and mapped the complex pattern of O-glycosylation on mammalian α-DG from muscle biopsies (Gomez Toledo *et al.*, 2012), as well as in-depth characterization and site mapping of the LARGE-mediated modification of phosphorylated O-mannose on α-DG (Yoshida-Moriguchi *et al.*, 2010). Importantly, Ashikov et al. showed that LARGE2 biochemically generates the same type of glycan polymer structures on α-DG as LARGE (Ashikov *et al.*, 2012). However, future experiments are needed to map and characterize the O-glycan structures on α-DG synthesized by LARGE2 in the context of Wnt-driven CRC and other tumor entities.

6.3. hCoSCs show elevated *LARGE2* expression and O-glycosylated α-DG

Laminin, fibronectin and collagen represent the main components of the basement membrane in the intestinal epithelium and are believed to play an important role in cell homeostasis (Simon-Assmann et al., 2010). At the intestinal crypt base, distinct isoforms of the ECM components laminin, as well as collagen, fibronectin and glycosaminoglycans (e.g. perlecan) are enriched and support the functionality of intestinal stem cells (ISC) (Laurie, Leblond and Martin, 1982). Complementary findings have shown that a laminin-based designer matrix meets the demands for

organoid formation and ISC functionality best (Gjorevski et al., 2016). We showed that human colonic stem/progenitor cells feature a Wnt/LARGE2-dependent, laminin-binding matriglycan structure on α-DG, which reveals a potentially important molecular trait of hCoSCs in this context. The migration of differentiating intestinal cells towards the lumen is achieved by dynamic epithelial-substratum interactions, and involves integrins, heparin sulfate proteoglycans and the ECM components laminin, fibronectin and collagen (Patterson and Watson, 2017). Recently published work illustrates that Dystroglycan and integrins collaborate in myelin stabilization on peripheral nerves, and they play a redundant role during the laminin-dependent epithelial polarization of epiblasts (Bello *et al.*, 2015; Li *et al.*, 2017). Interestingly, mouse models of β-integrin in the intestine do not show any alterations of epithelial anchorage to the BM (Jones *et al.*, 2006), suggesting that several redundant cell-matrix interactions are involved in intestinal epithelial homeostasis.

Cell adhesion can be differentially affected by altered expression of laminins and integrins, loss of cell adhesion molecules or rearrangement of the ECM, all strongly associated with angiogenesis, cell migration and differentiation as well as tumorigenesis (Litvinov *et al.*, 1997; Balda and Matter, 2003). Within this thesis we analyzed *LARGE2* gene expression of FACS sorted human colonic crypts according to their *EPHB2* surface levels. EPHB2 represents a Wnt target, which is positively regulated by β-Catenin/TCF signaling in the intestinal epithelium and highest expressed in the stem cell compartment (Batlle, Henderson, Beghtel, Van den Born, *et al.*, 2002). Differentiated cells show very low expression levels of this surface marker, thereby creating an *EPHB2* expression gradient along the vertical crypt-villus axis. We found that human colonic stem/progenitor cells feature a higher expression of *LARGE2* than terminally differentiated cells of the upper crypt, creating a gradient similar to EPHB2. Matrix metalloproteinases (MMPs) can cleave and degrade ECM proteins, ensuring proper remodeling during intestinal homeostasis, whereas deregulations of MMP activity is associated with inflammation and cancer (Bamba *et al.*, 2003; Bonnans, Chou and Werb, 2014). In epithelial cells and colorectal cancer, EphB/ephrin signaling regulates the formation of E-cadherin based adhesions (Cortina *et al.*, 2007; Genander *et al.*, 2009). Even though cell adhesion can be promoted by EphB/ephrin-B activity, site specific activation of the MMP ADAM10 induces shedding of E-Cadherin and consequently disrupts cell-cell adhesion at cell

interfaces (Solanas *et al.*, 2011). The combination of both mechanisms may be relevant in tissues as the intestine, to ensure the proper localization of EphB positive cells at the crypt bottom and to prevent the intermingling with the EphB negative and ephrin positive, differentiated cells at the surface (Batlle, Henderson, Beghtel, Van den Born, *et al.*, 2002). In most epithelial tissues, E-Cadherin is co-expressed with EpCAM, which has also been shown to impair E-Cadherin mediated cell adhesion, subsequently promoting cell proliferation and motility during the progression of malignancies (Litvinov *et al.*, 1997; Winter *et al.*, 2003). EpCAM is highly expressed in the proliferative cell compartment of intestinal crypts, when compared to intestinal differentiated cells and was found to be upregulated in epithelial cancers compared to the normal epithelium (Trzpis *et al.*, 2007). There are multiple cell-matrix interactions occurring simultaneously, which contributes to the stable homeostasis of the intestinal epithelium in development and disease. With respect to the findings in this project, we propose that the Wnt/LARGE2 dependent formation of laminin-binding matriglycan on α-DG plays an additionally supportive role in the maintenance and regulation of intestinal stem/progenitor cell attachment to their laminin-rich basement membrane. The downregulation of *LARGE2,* which occurs during colonic epithelial differentiation, presumably accounts for the diminished membranous O-glycosylation of α-DG we observed in the upper regions of human colonic crypts and in differentiated PDOs. This progressive loss of matriglycan during cell differentiation could therefore facilitate or be supportive to the epithelial cell migration along the vertical crypt axis. A loss of basement membrane adhesion due to diminished α-DG-laminin-binding might also contribute to the detachment of terminally differentiated cells during the shedding into the intestinal lumen at the end of their life cycle. However, in order to fully shed light on the functional relevance of the here identified Wnt/LARGE2/α-DG axis for intestinal epithelial homeostasis, future studies on mouse models which lack Large2 expression in adult intestinal epithelial cells are necessary.

6.4. Colorectal tumors show an elevated, but heterogeneous expression of *LARGE2* and α-DG O-glycosylation

In our cohort of primary and metastatic CRC PDTOs, we observed an overall elevated level of *LARGE2* when compared to normal colonic stem cells. In normal

stem cells, extrinsic Wnt signaling, caused by R-spondin and Wnt3a ligands, drives the expression of *LARGE2*. However, our bioinformatics analysis revealed that a recently published gene set, indicative of oncogenic, intrinsic Wnt activation due to *APC* mutations rather than a gene set specific for receptor-driven extrinsic Wnt activation positively correlates with *LARGE2* expression in CRC. Therefore, loss or truncation of APC might augment the expression of some Wnt target genes, including *LARGE2*, beyond the physiologic levels found in benign colonic stem cells. In CRC, Wnt activating mutations e.g. in *APC* or *CTNNB1* represent a very common event and occur in more than 90% of all cases (Sanchez-Vega *et al.*, 2018). In contrast, aberrant activation of Wnt signaling by these genetic alterations only play a minor role in kidney and prostate cancer (less than 5% of cases). When compared to normal tissue, our analysis revealed that the expression of *LARGE2* in PRAD is nearly unchanged and in KIRC even down modulated. Frequent downregulation of *LARGE2* has been identified as a main reason for α-DG hypoglycosylation during progression of prostate and clear cell renal cell carcinoma (Esser *et al.*, 2013; Miller *et al.*, 2015). Considering the influence of Wnt signaling on the expression of *LARGE2* we observed in hCoSCs and CRC but not in PRAD or KIRC TCGA cohorts, we propose that the identified Wnt/LARGE2/α-DG axis plays a tissue-specific and rather unique role in the intestinal tissue context.

Several PDTOs of our living biobank show elevated levels of O-glycosylation of α-DG, despite having moderate *LARGE2* expression levels when compared to other PDTOs with a similar complexity of O-glycosylation of α-DG. This indicates that other mechanisms and enzymes apart from the Wnt/LARGE2/α-DG axis, can decide over the amount and complexity of functional O-glycosylation of Dystroglycan. As introduced, the physiological generation of functionally O-glycosylated α-DG is a complex process and multiple enzymes and substrates are needed to extend the Ser or Thr residue and create the laminin binding matriglycan. NGS analysis from TCGA (Cancer Genome Atlas and The Cancer Genome Atlas Network, 2012) data showed missense or truncating mutations of one or more involved genes in approximately 18% of CRC cases (see 10.6 and 10.7), which can prevent or partially inhibit the formation of matriglycan in the context of high Wnt activity. The analyzed PDTOs and CRC cell lines from our cohort did not show any major mRNA de-regulation of these factors when compared amongst each other (data not shown), despite the observed

heterogeneous levels of α-DG glycosylation. This implicates that mutations rather than downmodulation or epigenetic silencing of these factors occur in CRC. By performing immunohistochemistry staining for O-glycosylated α-DG on a stage T3 (locally invasive tumor) CRC cohort in this thesis, we could show that the levels of O-glycosylated α-DG in CRC are low or absent in some cases, whereas two thirds show intermediate to strong glycosylation. This verifies the heterogeneous distribution of α-DG O-glycosylation that we have observed in our PDTO collection. In breast cancer cells, loss of membranous α-DG glycosylation was observed upon proteolytic cleavage of α-DG and β-DG by matrix metalloproteinases (MMPs) (Singh *et al.*, 2004). Wnt signaling was shown to promote the expression of MMPs in colon cancer and epithelial cells (Crawford *et al.*, 1999; Wu, Crampton and Hughes, 2007), which suggests that similar mechanisms could lead to proteolytic cleavage of α-DG in some CRC cells and might at least partially contribute to the here detected heterogeneity of O-glycosylation. Furthermore, increased levels of the endoprotease furin lead to extensive shedding of α-DG glycosylation. In this context, furin was previously shown to contribute in cancer progression (Bassi *et al.*, 2003). However, future studies should aim to fully uncover the LARGE2-independent regulatory processes in CRC, which influence the extent, complexity and functionality of matriglycan structures attached to α-DG.

6.5. Regulation of *LARGE2* by Hypoxia and SNAIL in different tissues and cancer types

The hypoxia transcription factor HIF-1α plays an important role in helping cells adapt to hypoxic conditions (Dengler, Galbraith and Espinosa, 2014). Studies named multiple target genes of HIF-1α, for instance the transcription factor SNAIL (Imai *et al.*, 2003). Stabilization of *SNAIL* helps cells to undergo epithelial-to-mesenchymal transition (EMT), where epithelial cells lose their cell-cell adhesion by e.g. suppression of E-cadherin (Imai *et al.*, 2003) and gain a more invasive, mesenchymal phenotype (Lundgren, Nordenskjöld and Landberg, 2009; Jackstadt *et al.*, 2013). This process can be observed in multiple sorts of tissues and cancers (Xu *et al.*, 2015; Choi *et al.*, 2017). Moreover, Mani et al. confirmed that induction of EMT via SNAIL contributes to the generation of cells, which show properties of stem cells in

the human mammary epithelium (Mani *et al.*, 2008). Even though many studies deciphered the role of hypoxia in EMT, only a few tried to understand how hypoxia influences the regulation of glycosyltransferases, and the effects on the glycome and tumor progression (reviewed in (Vajaria and Patel, 2017)). Huang and colleagues, found *LARGE2* negatively regulated by SNAIL and ZEB1 in prostate cancer (Qin Huang *et al.*, 2015), leading to α-DG hypoglycosylation. Our *in silico* analysis of the *LARGE2* sequence revealed a transcription factor binding site for SNAIL and overexpression of *SNAIL* in Caco-2 CRC cells lead to downregulation of *LARGE2* (data not shown, samples contributed by Dr. Markus Kaller, H. Hermeking lab, LMU Munich). Additional analysis in prostate, clear cell renal cell, breast and lung cancer, where *LARGE2* expression and α-DG glycosylation are known to be downregulated (Esser *et al.*, 2013; Miller *et al.*, 2015), found the expression of the EMT related gene *ZEB1* negatively correlated with *LARGE2*. In contrast to these tumor entities, our analysis on CRC cohorts showed a rather positive correlation between EMT related genes and *LARGE2,* as well as a strong correlation between *LARGE2* expression and Wnt signaling. Accordingly, the mesenchymal-type CRC cell lines SW480 and SW620, characterized by high Wnt activity, strong expression of EMT-related genes and high invasive capacity (Hewitt *et al.*, 2000), expressed substantially higher levels of *LARGE2* when compared to normal colonic stem cells. Moreover, Wnt signaling itself is involved in the induction of EMT in tumor cells (Naishiro *et al.*, 2001; Kim, 2002) and *SNAIL* levels are positively regulated by Wnt signaling in embryogenesis and CRC, presumably via an AXIN2-dependent mechanism (ten Berge *et al.*, 2008; Wu *et al.*, 2012). Our observation that high Wnt activity drives *LARGE2* gene expression in CRC suggests that endogenous levels of *SNAIL*, in contrast to what has been observed in prostate cancer, are not able to substantially repress *LARGE2* in this Wnt-active tissue background. The outstanding status of the Wnt/*LARGE2* connection we observed in CRC might give a glimpse in the delicate regulation of *LARGE2* by Wnt and EMT factors in different tissue entities.

6.6. Modulation of LARGE2 synthesized matriglycan on a-DG affects CRC cell adhesion and migration

From previously published work we know that defects of several

glycosyltransferases can impede the synthesis of functional matriglycan, thereby leading to hypoglycosylation of α-DG. Lower than physiologic levels of α-DG glycosylation, as observed in prostate, breast and brain cancer, provided evidence that a partial loss of laminin anchorage commonly occurs in aggressive tumors derived from these tissues. This leads to augmented cellular migration and invasiveness which are well-accepted characteristics of disease progression (Bao *et al.*, 2009; Akhavan *et al.*, 2012; Esser *et al.*, 2013). With respect to the biological outcome of LARGE2/α-DG, we obtained congruent results in *ex vivo* laminin invasion assays, as knockout of LARGE2 in two different CRC cell lines favored cell migration. Complementary, the ectopic expression of *LARGE2* and the consequential generation of higher MW O-glycosylation of α-DG in HT-29 cells interfered with cellular migration through laminin-coated membranes. However, one should be aware that the here used model of transwell migration has several limitations. The ECM, which CRC cells face at the tumor-stroma interface, is of much higher complexity than laminin-coated membranes can mimic. Furthermore, is it not possible to completely discriminate between active cell migration and passive, gravity-based transition of the cells through the laminin-coated membrane pores, which might occur due to reduced laminin adhesiveness after the loss of α-DG O-glycosylation. Despite these limitations, our *in vitro* experimental data point to Wnt/LARGE2/a-DG signaling as a rather migration/invasion-limiting signaling pathway frequently active in the background of CRC.

In CRC, a group of Wnt target genes, which include *LGR5* (Barker *et al.*, 2007) or *ASCL2* (van der Flier *et al.*, 2009) are upregulated at a very early step in carcinogenesis and remain expressed throughout tumor progression (Brabletz *et al.*, 2005). Some evidential studies have proven that the permanent expression of several Wnt target genes is crucial to maintain a stem-like character of cancer cells and drive tumor progression (He *et al.*, 1998; Zhang *et al.*, 2001), creating a “cancer stem cell” (CSC) population. Lineage tracing experiments revealed that these CSCs are important for adenoma and full-blown tumor self-renewal in CRC (Schepers *et al.*, 2012; Shimokawa *et al.*, 2017). A more recent work from Frederic de Sauvage and colleagues demonstrated that mouse colonic tumors enriched in Lgr5+ cells show an increased capacity to metastasize to the liver (de Sousa e Melo *et al.*, 2017). While a treatment for depletion of Lgr5+ cells in primary tumors impeded tumor growth,

tumors were able to restore a Lgr5+ cell population after treatment had been discontinued, which ultimately led to re-initiation of tumor growth. Moreover, the Lgr5+ cells were indispensable for the outgrowth and maintenance of CRC-derived metastasis. This raises the possibility that the colonic tumor niche could be more favorable for cancer cellular plasticity to occur, whereas the liver might provide a much more restrictive microenvironment, which largely prevents or minimizes the re-programming of non-CSCs into LGR5+ CSCs. Our data demonstrate that the expression of well-described hCoSC specific genes, originally derived from human colonic epithelial cell sub-populations high in EPHB2 or PTK7 surface abundance (Merlos-Suárez *et al.*, 2011; Jung *et al.*, 2015), positively correlates with the expression of *LARGE2* in several CRC cohorts. In addition, high levels of *LARGE2* are frequently found in tumors, which are highly positive for LGR5 and belong to the Wnt-driven and intestinal stem cell-like molecular subtype in CRC. These bioinformatical data, together with our observations on liver metastatic PDTOs, let us speculate that the Wnt/LARGE2/a-DG axis might not provide a mere tumor-suppressive functionality to CRC cells.

Several studies provided strong evidence for a critical role of the tumor-laminin adhesion in CRC metastasis. Analysis of liver and lung metastatic CRC revealed, that expression of integrin α2 contributes to successful formation of metastasis (Yoshimura *et al.*, 2009). Additionally, BCAM and its ligand LAMA5 were found to be highly expressed in the context of liver metastatic colorectal cancer, and therapeutic targeting of LAMA5-binding by BCAM helped to prevent hepatic colonization (Bartolini *et al.*, 2016). Recently, Ganesh et al. described that the expression of the laminin-adhesive surface molecule *L1CAM* partially overlaps with high *LGR5* mRNA levels in primary and liver metastasis-derived CRC cells, and L1CAM is critical for the propagation of orthotopically transplanted carcinomas in mouse models, as well as for metastatic spread of CRC cells to the liver (Ganesh *et al.*, 2020). Importantly, this group demonstrated that CRC cells with high surface abundance of the human colonic stem cells markers EPHB2 and CD44 are enriched in *L1CAM* gene expression, which is in line to what we observed for *LARGE2* gene expression in sub-populations of human colonic epithelial cells sorted according to their EPHB2 surface abundance. Therefore, it is likely that Wnt/LARGE2/αDG signaling is also very active in *L1CAM*+ CRC cells, which possess a high capacity for liver

colonization and growth. We therefore speculate that L1CAM and LARGE2/αDG, which both represent laminin-binding factors, play a supportive role in CRC liver metastasis outgrowth.

In stark contrast to colorectal cancer, decreased expression of *L1CAM* in pancreatic ductal carcinoma (PDAC) is associated with enhanced stemness and tumorigenicity, leading to a more aggressive phenotype and poor patient outcome (Cave *et al.*, 2020). This discrepancy between different tumor entities could also hold true for the LARGE2/αDG axis, which is driven by Wnt signaling and correlates with intestinal stemness feature in CRC, but whose activity is frequently down modulated in other types of cancer, such as pancreas, prostate, and renal cancer (Esser *et al.*, 2013; Miller *et al.*, 2015). We suggest that the actual contribution of L1CAM and αDG to cancer progression might be dependent on the tissue origin of tumor cells and presumably on tumor stage. However, additional experiments are needed to provide solid evidence for our hypothesis.

A just recently published study reported in very detail, that a non-CSC (Lgr5-) state represents an intermediate cellular phenotype in the multi-step process of CRC-derived liver metastasis (Fumagalli et al., 2020). These data show that, in fact, non-CSCs are the major metastasis seeding cells, and these seeds have to regenerate a LGR5+ cancer stem cell population which then drives further metastatic outgrowth. In support of this hypothesis LGR5+ cells were present in the primary tumor and in the liver metastatic population, but not in actually active migrating, blood circulating, or extravasating tumor cells (Fumagalli *et al.*, 2020). The laminin adhesion of tumor cells is a very dynamic process, which is tightly regulated at different stages of metastatic cancer progression. Once tumor cells leave their primary niche, colonic CSCs lose their “LGR5” stemness status according to Fumagalli and colleagues, and presumably this is accompanied by reduction of laminin adhesion due to downregulation of factors such as LARGE2 or L1CAM, whose expression is strongly associated with the LGR5 status. At the metastatic site, LGR5-low cancer cells acquire cellular plasticity, spawn LGR5+ CRC cells, and thereby they might re-establish a gene expression program for augmented laminin adhesion, including *L1CAM* and Wnt-driven *LARGE2*, in the outgrowing tumor seed.

By referring to this model, we hypothesize that metastasized LGR5- CRC cells re-gain stemness and increase their laminin adhesion capacity via the Wnt/LARGE2/α-

DG axis once they have successfully invaded the foreign organ. In support of this hypothesis, we observed that PDTOs derived from metastatic CRC show elevated *LARGE2* gene expression, normal hCoSC-like or even further elevated complexity of O-glycosylated α-DG and a basal plus membranous localization of this laminin-binding feature. Together with the appearance of functional α-DG in the liver-metastatic KM12c cell derivative KM12-L4a (Morikawa *et al.*, 1988), our data suggest that *LARGE2* expression and O-glycosylation of α-DG might drive one of several ECM-adhesive features of CRC cells beneficial for liver metastatic tumor spread.

In conclusion and according to our findings, the here described Wnt/LARGE2/α-DG axis might play a context dependent dual role in CRC disease by partially interfering with primary tumor dissemination and, at a later stage of tumor spread, by supporting the metastatic growth of CRC cells which have reached or passed the laminin-rich capillary system of the liver. Future studies should experimentally address this concept further on an *in vivo* model system, which closely recapitulates the complete liver metastatic process. The orthotopic transplantation of human or mouse tumor cells into the colonic colon wall of mice via veterinary endoscope-guided injection has been shown recently to meet the demands for this type of study (Roper *et al.*, 2017). Transplanting LARGE2 wild type versus LARGE2-KO isogenic organoid pairs by this method could bring further insight into the biological relevance of Wnt/LARGE2/αDG signaling for CRC progression and metastasis. Overall, a better understanding of the complex interplay between CRC cells and their adjacent ECM at the primary and liver metastatic tumor site might open new avenues for preventive and curative therapeutic strategies.

7. Summary

Wnt signaling plays a major role in the development and life-long self-renewal of the intestinal epithelium. Aberrant Wnt activation drives uncontrolled cell growth and represents the initiation step for a majority of colorectal cancer (CRC) cases. In-depth understanding of Wnt signaling effector pathways helps to define candidate therapeutic targets for CRC treatment.

In this Ph.D thesis, we revealed that expression of the O-glycosyltransferase LARGE2 encoding gene is upregulated after activation of Wnt signaling. Intrinsic Wnt activation by silencing of APC, by ectopic expression of an oncogenic β-catenin mutant, or the inhibition of Wnt signaling via β-catenin sequestration affects the expression of *LARGE2* in CRC. We found the binding motif “CTTTGATC” within the first intron of the *LARGE2* gene, and binding of TCF7L2 to this genomic locus is necessary to trigger Wnt-driven *LARGE2* expression. Moreover, we could demonstrate that O-glycosylation of the outer membrane protein α-dystroglycan (α-DG) by LARGE2 is essential and sufficient for its laminin-binding capacity. Importantly, this cell biological process, which represents an important mode of epithelial cell-to-matrix interaction, is triggered by Wnt signaling in CRC cells.

By performing experiments on normal human colon organoids and FFPE human colonic tissue samples, we found that *LARGE2* is highly expressed in the stem/progenitor cell sub-population, and multi-lineage differentiation of colonic stem cells reduces *LARGE2* levels and abolishes O-glycosylation of α-DG. Data from CRC cohorts, a living biobank of microsatellite stable tumor organoids, and classic CRC cell lines showed elevated expression of *LARGE2* when compared to normal colonic epithelial cells. Furthermore, *LARGE2* expression correlates with the Wnt signaling driven molecular subtype and an intestinal stem cell-like phenotype in publicly available data from CRC cohorts. We detected high molecular weight forms of O-glycosylated α-DG in primary and liver-metastasized patient-derived tumor organoids, and laminin-binding α-DG was enriched in a cell line model for CRC liver-metastasis. Functionally, modulation of the LARGE2/α-Dystroglycan axis affected CRC cell migration through laminin-coated membranes and adhesion of CRC cells on an endothelial cell monolayer.

Taken together, our results illustrate that the Wnt-driven expression of LARGE2

mediates the functional laminin-binding O-glycosylated α-Dystroglycan. Moreover, we suggest that activation of Wnt signaling increases cell adhesion to laminin-rich extracellular matrices in human colonic epithelial cells and in CRC at different disease stages.

8. Zusammenfassung

Der Wnt Signalweg spielt eine Schlüsselrolle in der Entwicklung und Erneuerung des intestinalen Epithelgewebes. Abnorme Aktivierung fördert unkontrolliertes Zellwachstum und ist in der Mehrheit der Erkrankungen am kolorektalen Karzinom (CRC) der initiierende Schritt. Ein umfassendes Verständnis des Wnt Signalweges erleichtert somit die Validierung therapeutischer Ziele für Krebsbehandlungen.

In dieser Arbeit fanden wir heraus, dass die bi-funktionale Glykosyltransferase LARGE2 durch unterschiedliche Aktivierungen des Wnt Signalwegs hochreguliert wird. Wnt-Aktivierung mittels ektopischer Expression einer onkogenen β-Catenin-Mutante oder Stummschaltung des Tumor-Suppressors APC sowie die Unterbrechung des Signalweges durch Sequestrierung von β-Catenin beeinflussten die Expression von *LARGE2* in CRC Zellen. Wir fanden eine spezifische Bindungs-Sequenz „CTTTGATC“ des Transkriptionsfaktors TCF7L2 innerhalb des ersten Introns von *LARGE2* und bewiesen die Notwendigkeit einer funktionellen Bindung von TCF7L2, um die Wnt-gesteuerte *LARGE2*-Expression auszulösen. Darüber hinaus haben wir festgestellt, dass LARGE2 durch Schaffung und Aufrechterhaltung der funktionellen O-Glykosylierung von α-Dystroglycan (α-DG) für die Affinität und Adhäsion der Darmkrebszellen an Laminin in der Extrazellulären Matrix unabdinglich ist.

Durch Experimente an Organoiden aus gesundem humanen Kolon-Gewebe und Kolongewebeproben in FFPE konnten wir nachweisen, dass *LARGE2* in der Stammzell-Population der Krypten stark exprimiert wird. Die Differenzierung der Stammzellen führte zu einer verringerten Expression von *LARGE2* und verminderter O-Glykosylierung von α-DG. Unsere Kohorte an CRC Organoiden und CRC Zelllinien zeigte im Vergleich zu normalem Darmgewebe eine erhöhte Expression von *LARGE2*. Eine Analyse der Expressionsdaten von öffentlichen CRC-Kohorten ergab, dass eine erhöhte Expression von *LARGE2* hauptsächlich in Fällen des Wnt-getriebenen Subtyps des kolorektalen Karzinoms auftritt und mit der Expression von Markern für intestinale Stammzellen in positiver Korrelation steht.

Außerdem konnten wir hochmolekulare Formen von α-Dystroglycan Glykosylierung in primären und lebermetastasierten Tumor Organoiden sowie in einem CRC-Zelllinienmodell für Lebermetastasen nachweisen. Die Migration der CRC-Zellen

durch eine laminin-beschichtete Membran sowie die Adhäsion auf Endothelzellen konnte durch Modulation der Wnt / LARGE2 / α-Dystroglycan-Achse beeinflusst werden.

Zusammengenommen zeigen unsere Ergebnisse, dass die Wnt-gesteuerte Expression von *LARGE2* die O-Glykosylierung von α-Dystroglycan vermittelt, welche für die funktionelle Laminin-Anbindung notwendig ist. Deshalb vermuten wir, dass die Aktivierung des Wnt-Signalwegs die Zelladhäsion an der Basalmembran in CRC durch verstärkte Laminin-Verankerung erhöht.

9. References

Abar, L. *et al.* (2018) 'Height and body fatness and colorectal cancer risk: an update of the WCRF–AICR systematic review of published prospective studies', *European Journal of Nutrition*, 57(5), pp. 1701–1720. doi: 10.1007/s00394-017-1557-1.

Akhavan, A. *et al.* (2012) 'Loss of cell-surface laminin anchoring promotes tumor growth and is associated with poor clinical outcomes.', *Cancer research*. NIH Public Access, 72(10), pp. 2578–88. doi: 10.1158/0008-5472.CAN-11-3732.

Albuquerque, C. *et al.* (2002) 'The "just-right" signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade', *Hum Mol Genet*, 11(13), pp. 1549–1560. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12045208.

Alexander, K. L. *et al.* (2018) '716 - Glycosyltransferase ST6GAL-I Associates with WNT Signaling Molecules and is a Characteristic Marker of Mucosal Epithelial Stem Cells', *Gastroenterology*. Elsevier BV, 154(6), p. S-148-S-149. doi: 10.1016/S0016-5085(18)30912-0.

Anders, S., Pyl, P. T. and Huber, W. (2015) 'HTSeq--a Python framework to work with high-throughput sequencing data', *Bioinformatics*, 31(2), pp. 166–169. doi: 10.1093/bioinformatics/btu638.

Armaghany, T. *et al.* (2012) 'Genetic alterations in colorectal cancer.', *Gastrointestinal cancer research : GCR*. International Society of Gastrointestinal Oncology, 5(1), pp. 19–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22574233 (Accessed: 15 October 2019).

Arnold, M. *et al.* (2017) 'Global patterns and trends in colorectal cancer incidence and mortality.', *Gut*. BMJ Publishing Group, 66(4), pp. 683–691. doi: 10.1136/gutjnl-2015-310912.

Aryal, R. P., Ju, T. and Cummings, R. D. (2010) 'The endoplasmic reticulum chaperone cosmc directly promotes in vitro folding of T-synthase', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 285(4), pp. 2456–2462. doi: 10.1074/jbc.M109.065169.

Ashikov, A. *et al.* (2012) 'LARGE2 generates the same xylose-and glucuronic acid-containing glycan structures as LARGE'. doi: 10.1093/glycob/cws153.

Balda, M. S. and Matter, K. (2003) 'Epithelial cell adhesion and the regulation of gene expression', *Trends in Cell Biology*. Elsevier Ltd, pp. 310–318. doi: 10.1016/S0962-8924(03)00105-3.

Bamba, S. *et al.* (2003) 'Matrix metalloproteinase-3 secretion from human colonic subepithelial myofibroblasts: Role of interleukin-17', *Journal of Gastroenterology*. Springer, 38(6), pp. 548–554. doi: 10.1007/s00535-002-1101-8.

Bao, X. *et al.* (2009) 'Tumor suppressor function of laminin-binding alpha-dystroglycan requires a distinct beta3-N-acetylglucosaminyltransferase', *Proc Natl*

Acad Sci U S A, 106(29), pp. 12109–12114. doi: 10.1073/pnas.0904515106.

Barker, N. *et al.* (2007) 'Identification of stem cells in small intestine and colon by marker gene Lgr5', *Nature*, 449(7165), pp. 1003–1007. doi: 10.1038/nature06196.

Barker, N. *et al.* (2009) 'Crypt stem cells as the cells-of-origin of intestinal cancer', *Nature*. Nature Publishing Group, 457(7229), pp. 608–611. doi: 10.1038/nature07602.

Barretina, J. *et al.* (2012) 'The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity', *Nature*, 483(7391), pp. 603–607. doi: 10.1038/nature11003.

Bartolini, A. *et al.* (2016) 'BCAM and LAMA5 Mediate the Recognition between Tumor Cells and the Endothelium in the Metastatic Spreading of KRAS-Mutant Colorectal Cancer', *Clin Cancer Res*, 22(19), pp. 4923–4933. doi: 10.1158/1078-0432.CCR-15-2664.

Bassagañas, S. *et al.* (2014) 'Pancreatic Cancer Cell Glycosylation Regulates Cell Adhesion and Invasion through the Modulation of α2β1 Integrin and E-Cadherin Function', *PLoS ONE*. Edited by Y. St-Pierre. Public Library of Science, 9(5), p. e98595. doi: 10.1371/journal.pone.0098595.

Bassi, D. E. *et al.* (2003) 'Increased furin activity enhances the malignant phenotype of human head and neck cancer cells', *American Journal of Pathology*. American Society for Investigative Pathology Inc., 162(2), pp. 439–447. doi: 10.1016/S0002-9440(10)63838-2.

Batlle, E., Henderson, J. T., Beghtel, H., van den Born, M. M. W., *et al.* (2002) 'Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB', *Cell*. Cell Press, 111(2), pp. 251–263. doi: 10.1016/S0092-8674(02)01015-2.

Batlle, E., Henderson, J. T., Beghtel, H., Van den Born, M. M. W., *et al.* (2002) 'β-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/EphrinB', *Cell*. Cell Press, 111(2), pp. 251–263. doi: 10.1016/S0092-8674(02)01015-2.

Bello, V. *et al.* (2015) 'The dystroglycan: Nestled in an adhesome during embryonic development', *Developmental Biology*. Elsevier, 401(1), pp. 132–142. doi: 10.1016/j.ydbio.2014.07.006.

Bello, V. and Darribère, T. (2018) 'Dystroglycan', in *Encyclopedia of Signaling Molecules*. Springer International Publishing, pp. 1457–1469. doi: 10.1007/978-3-319-67199-4_101578.

Bennett, E. P. *et al.* (2011) 'Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family'. doi: 10.1093/glycob/cwr182.

ten Berge, D. *et al.* (2008) 'Wnt signaling mediates self-organization and axis formation in embryoid bodies', *Cell Stem Cell*, 3(5), pp. 508–518. doi: 10.1016/j.stem.2008.09.013.

de Bernabé, D. B. V *et al.* (2009) ‘Loss of α-dystroglycan laminin binding in epithelium-derived cancers is caused by silencing of LARGE’, *Journal of Biological Chemistry*, 284(17), pp. 11279–11284. doi: 10.1074/jbc.C900007200.

Bonnans, C., Chou, J. and Werb, Z. (2014) ‘Remodelling the extracellular matrix in development and disease’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 786–801. doi: 10.1038/nrm3904.

Bos, J. L. *et al.* (1987) ‘Prevalence of ras gene mutations in human colorectal cancers’, *Nature*. Nature Publishing Group, 327(6120), pp. 293–297. doi: 10.1038/327293a0.

Brabletz, T. *et al.* (2005) ‘Opinion: migrating cancer stem cells - an integrated concept of malignant tumour progression’, *Nat Rev Cancer*, 5(9), pp. 744–749. doi: 10.1038/nrc1694.

Bray, F. *et al.* (2018) ‘Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries’, *CA: A Cancer Journal for Clinicians*. American Cancer Society, 68(6), pp. 394–424. doi: 10.3322/caac.21492.

Brockhausen, I. and Stanley, P. (2017) ‘Chapter 10 O-GalNAc Glycans’, *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, 1, pp. 1–9. doi: 10.1101/glycobiology.3e.010.

Brockington, M. *et al.* (2005) ‘Localization and functional analysis of the LARGE family of glycosyltransferases: significance for muscular dystrophy’, *Human Molecular Genetics*. Oxford University Press, 14(5), pp. 657–665. doi: 10.1093/hmg/ddi062.

Calon, A. *et al.* (2015) ‘Stromal gene expression defines poor-prognosis subtypes in colorectal cancer’, *Nature Genetics*. Nature Publishing Group, 47(4), pp. 320–329. doi: 10.1038/ng.3225.

Cancer Genome Atlas, N. and The Cancer Genome Atlas Network (2012) ‘Comprehensive molecular characterization of human colon and rectal cancer’, *Nature*, 487(7407), pp. 330–337. doi: 10.1038/nature11252.

Carvalho, S. *et al.* (2016) ‘Preventing E-cadherin aberrant N-glycosylation at Asn-554 improves its critical function in gastric cancer’, *Oncogene*. Nature Publishing Group, 35(13), pp. 1619–1631. doi: 10.1038/onc.2015.225.

Castrop, J., Van Norren, K. and Clevers, H. (1992) *A gene family of HMG-box transcription factors with homology to TCF-1*, *Nucleic Acids Research*.

Cave, D. D. *et al.* (2020) ‘TGF-β1 secreted by pancreatic stellate cells promotes stemness and tumourigenicity in pancreatic cancer cells through L1CAM downregulation.’, *Oncogene*. doi: 10.1038/s41388-020-1289-1.

Choi, B. J. *et al.* (2017) ‘Hypoxia induces epithelial-mesenchymal transition in colorectal cancer cells through ubiquitin-specific protease 47-mediated stabilization of Snail: A potential role of Sox9’, *Sci Rep*, 7(1), p. 15918. doi: 10.1038/s41598-017-15139-5.

Christiansen, M. N. *et al.* (2014) 'Cell surface protein glycosylation in cancer', *PROTEOMICS*. John Wiley & Sons, Ltd, 14(4–5), pp. 525–546. doi: 10.1002/pmic.201300387.

Clausen, H. and Bennett, E. P. (1996) *A family of UDP-GalNAc: polypeptide iV-acetylgalactosaminyl-transferases control the initiation of mucin-type O-linked glycosylation*, *Glycobiology*. Available at: https://academic.oup.com/glycob/article-abstract/6/6/635/635426 (Accessed: 1 April 2020).

Colley KJ, Varki A, K. T. (2017) 'Cellular Organization of Glycosylation In: Essentials of Glycobiology', *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, pp. 4–11. doi: 10.1101/glycobiology.3e.004.

Cooks, T. *et al.* (2013) 'Mutant p53 Prolongs NF-κB Activation and Promotes Chronic Inflammation and Inflammation-Associated Colorectal Cancer', *Cancer Cell*, 23(5), pp. 634–646. doi: 10.1016/j.ccr.2013.03.022.

Cortina, C. *et al.* (2007) 'EphB-ephrin-B interactions suppress colorectal cancer progression by compartmentalizing tumor cells', *Nature Genetics*. Nature Publishing Group, 39(11), pp. 1376–1383. doi: 10.1038/ng.2007.11.

Cox, J. *et al.* (2011) 'Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment', *Journal of Proteome Research*, 10(4), pp. 1794–1805. doi: 10.1021/pr101065j.

Cox, J. and Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature Biotechnology*. Nature Publishing Group, 26(12), pp. 1367–1372. doi: 10.1038/nbt.1511.

Crawford, H. C. *et al.* (1999) 'The metalloproteinase matrilysin is a target of β-catenin transactivation in intestinal tumors', *Oncogene*. Nature Publishing Group, 18(18), pp. 2883–2891. doi: 10.1038/sj.onc.1202627.

Cummings, R. D. (2009) 'The repertoire of glycan determinants in the human glycome', *Molecular BioSystems*. The Royal Society of Chemistry, pp. 1087–1104. doi: 10.1039/b907931a.

Daniels, D. L. and Weis, W. I. (2005) 'β-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation', *Nature Structural and Molecular Biology*, 12(4), pp. 364–371. doi: 10.1038/nsmb912.

Darwich, A. S. *et al.* (2014) 'Meta-analysis of the turnover of intestinal epithelia in preclinical animal species and humans', *Drug Metabolism and Disposition*. American Society for Pharmacology and Experimental Therapy, pp. 2016–2022. doi: 10.1124/dmd.114.058404.

Dengler, V. L., Galbraith, M. D. and Espinosa, J. M. (2014) 'Transcriptional regulation by hypoxia inducible factors', *Critical Reviews in Biochemistry and Molecular Biology*. NIH Public Access, pp. 1–15. doi: 10.3109/10409238.2013.838205.

Diepenbruck, M. *et al.* (2014) 'Tead2 expression levels control the subcellular distribution of yap and taz, zyxin expression and epithelial-mesenchymal transition',

Journal of Cell Science. Company of Biologists Ltd, 127(7), pp. 1523–1536. doi: 10.1242/jcs.139865.

Dwyer, C. A. *et al.* (2012) 'RPTPζ/phosphacan is abnormally glycosylated in a model of muscle-eye-brain disease lacking functional POMGnT1', *Neuroscience*. NIH Public Access, 220, pp. 47–61. doi: 10.1016/j.neuroscience.2012.06.026.

Ervasti, J. M. and Campbell, K. P. (1991) 'Membrane organization of the dystrophin-glycoprotein complex', *Cell*. Cell Press, 66(6), pp. 1121–1131. doi: 10.1016/0092-8674(91)90035-W.

Esser, A. K. *et al.* (2013) 'Loss of LARGE2 disrupts functional glycosylation of α-dystroglycan in prostate cancer.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 288(4), pp. 2132–42. doi: 10.1074/jbc.M112.432807.

Fearon, E. R. (2011) 'Molecular genetics of colorectal cancer', *Annu Rev Pathol*, 6, pp. 479–507. doi: 10.1146/annurev-pathol-011110-130235.

Fearon, E. R. and Vogelstein, B. (1990) 'A genetic model for colorectal tumorigenesis.', *Cell*, 61(5), pp. 759–67. doi: 10.1016/0092-8674(90)90186-i.

Fidler, M. M., Soerjomataram, I. and Bray, F. (2016) 'A global view on cancer incidence and national levels of the human development index', *International Journal of Cancer*, 139(11), pp. 2436–2446. doi: 10.1002/ijc.30382.

Fleming, N. I. *et al.* (2013) 'SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer', *Cancer Research*, 73(2), pp. 725–735. doi: 10.1158/0008-5472.CAN-12-2706.

van der Flier, L. G. *et al.* (2009) 'Transcription Factor Achaete Scute-Like 2 Controls Intestinal Stem Cell Fate', *Cell*, 136(5), pp. 903–912. doi: 10.1016/j.cell.2009.01.031.

van der Flier, L. G. and Clevers, H. (2009) 'Stem Cells, Self-Renewal, and Differentiation in the Intestinal Epithelium', *Annual Review of Physiology*, 71(1), pp. 241–260. doi: 10.1146/annurev.physiol.010908.163145.

Freitas, M. *et al.* (2018) 'A novel DNA methylation panel accurately detects colorectal cancer independently of molecular pathway', *Journal of Translational Medicine*. BioMed Central Ltd., 16(1), p. 45. doi: 10.1186/s12967-018-1415-9.

Fujimura, K. *et al.* (2005) 'LARGE2 facilitates the maturation of α-dystroglycan more effectively than LARGE', *Biochemical and Biophysical Research Communications*, 329(3), pp. 1162–1171. doi: 10.1016/j.bbrc.2005.02.082.

Fumagalli, A. *et al.* (2020) 'Plasticity of Lgr5-Negative Cancer Cells Drives Metastasis in Colorectal Cancer', *Cell Stem Cell*. Elsevier BV, 26(4), pp. 569-578.e7. doi: 10.1016/j.stem.2020.02.008.

Ganesh, K. *et al.* (2020) 'L1CAM defines the regenerative origin of metastasis-initiating cells in colorectal cancer', *Nature Cancer*. Springer Science and Business Media LLC, 1(1), pp. 28–45. doi: 10.1038/s43018-019-0006-x.

Genander, M. *et al.* (2009) 'Dissociation of EphB2 Signaling Pathways Mediating

Progenitor Cell Proliferation and Tumor Suppression', *Cell*. NIH Public Access, 139(4), pp. 679–692. doi: 10.1016/j.cell.2009.08.048.

Gjorevski, N. *et al.* (2016) 'Designer matrices for intestinal stem cell and organoid culture', *Nature*, 539(7630), pp. 560–564. doi: 10.1038/nature20168.

Gomez Toledo, A. *et al.* (2012) 'O-Mannose and O-N-acetyl galactosamine glycosylation of mammalian α-dystroglycan is conserved in a region-specific manner'. doi: 10.1093/glycob/cws109.

Grady, W. M. *et al.* (1998) 'Mutation of the type II transforming growth factor-beta receptor is coincident with the transformation of human colon adenomas to malignant carcinomas.', *Cancer research*, 58(14), pp. 3101–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9679977 (Accessed: 27 August 2019).

Grewal, P. K. *et al.* (2005) 'Characterization of the LARGE family of putative glycosyltransferases associated with dystroglycanopathies', *Glycobiology*, 15(10), pp. 912–923. doi: 10.1093/glycob/cwi094.

Gross, J. C. and Boutros, M. (2013) 'Secretion and extracellular space travel of Wnt proteins', *Current Opinion in Genetics and Development*, pp. 385–390. doi: 10.1016/j.gde.2013.02.017.

Guinney, J. *et al.* (2015) 'The consensus molecular subtypes of colorectal cancer', *Nat Med*, 21(11), pp. 1350–1356. doi: 10.1038/nm.3967.

Haigis, K. M. *et al.* (2008) 'Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon', *Nature Genetics*, 40(5), pp. 600–608. doi: 10.1038/ng.115.

Hatzis, P. *et al.* (2008) 'Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells.', *Molecular and cellular biology*, 28(8), pp. 2732–44. doi: 10.1128/MCB.02175-07.

He, T. C. *et al.* (1998) 'Identification of c-MYC as a target of the APC pathway', *Science*, 281(5382), pp. 1509–1512. doi: 10.1126/science.281.5382.1509.

HEATH, J. (1996) 'EPITHELIAL CELL MIGRATION IN THE INTESTINE', *Cell Biology International*. Wiley-Blackwell Publishing Ltd, 20(2), pp. 139–146. doi: 10.1006/cbir.1996.0018.

Heinz, S. *et al.* (2010) 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', *Mol Cell*, 38(4), pp. 576–589. doi: 10.1016/j.molcel.2010.05.004.

Herbst, A. *et al.* (2014) 'Comprehensive analysis of beta-catenin target genes in colorectal carcinoma cell lines with deregulated Wnt/beta-catenin signaling', *BMC Genomics*, 15, p. 74. doi: 10.1186/1471-2164-15-74.

Hewitt, R. E. *et al.* (2000) 'Validation of a model of colon cancer progression', *J Pathol*, 192(4), pp. 446–454. doi: 10.1002/1096-9896(2000)9999:9999<::AID-PATH775>3.0.CO;2-K.

Hohenester, E. and Yurchenco, P. D. (2013) 'Laminins in basement membrane

assembly', *Cell Adh Migr*, 7(1), pp. 56–63. doi: 10.4161/cam.21831.

Holt, K. H. *et al.* (2000) 'Biosynthesis of dystroglycan: processing of a precursor propeptide', *FEBS Letters*. John Wiley & Sons, Ltd, 468(1), pp. 79–83. doi: 10.1016/S0014-5793(00)01195-9.

Horvay, K. *et al.* (2011) 'Wnt signaling regulates Snai1 expression and cellular localization in the mouse intestinal epithelial stem cell niche', *Stem Cells Dev*, 20(4), pp. 737–745. doi: 10.1089/scd.2010.0188.

Huang, Qin *et al.* (2015) 'The glycosyltransferase LARGE2 is repressed by Snail and ZEB1 in prostate cancer.', *Cancer biology & therapy*. Taylor & Francis, 16(1), pp. 125–136. doi: 10.4161/15384047.2014.987078.

Huang, Q *et al.* (2015) 'The glycosyltransferase LARGE2 is repressed by Snail and ZEB1 in prostate cancer', *Cancer Biol Ther*, 16(1), pp. 125–136. doi: 10.4161/15384047.2014.987078.

I. Brockhausen J. Yang, M. Lehotay, S. O. und S. I. (2005) 'Pathways of Mucin O-Glycosylation in Normal and Malignant Rat Colonic Epithelial Cells Reveal a Mechanism for Cancer-Associated Sialyl-Tn Antigen Expression', *Biological Chemistry*, 382(2).

Ibraghimov-Beskrovnaya, O. *et al.* (1992) 'Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix', *Nature*, 355(6362), pp. 696–702. doi: 10.1038/355696a0.

Imai, T. *et al.* (2003) 'Hypoxia attenuates the expression of E-cadherin via up-regulation of SNAIL in ovarian carcinoma cells', *American Journal of Pathology*. American Society for Investigative Pathology Inc., 163(4), pp. 1437–1447. doi: 10.1016/S0002-9440(10)63501-8.

Inamori, K.-I. I. *et al.* (2016) 'LARGE2-dependent glycosylation confers laminin-binding ability on proteoglycans', *Glycobiology*, 26(12), pp. 1–13. doi: 10.1093/glycob/cww075.

Inamori, K. *et al.* (2014) 'Endogenous glucuronyltransferase activity of LARGE or LARGE2 required for functional modification of alpha-dystroglycan in cells and tissues', *J Biol Chem*, 289(41), pp. 28138–28148. doi: 10.1074/jbc.M114.597831.

Inamori, K. I. *et al.* (2016) 'LARGE2-dependent glycosylation confers laminin-binding ability on proteoglycans', *Glycobiology*, 26(12), pp. 1284–1296. doi: 10.1093/glycob/cww075.

Isella, C *et al.* (2017) 'Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer', *Nat Commun*, 8, p. 15107. doi: 10.1038/ncomms15107.

Isella, Claudio *et al.* (2017) 'Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer', *Nature Communications*. Nature Publishing Group, 8(1), p. 15107. doi: 10.1038/ncomms15107.

Jackstadt, R. *et al.* (2013) 'AP4 is a mediator of epithelial-mesenchymal transition and metastasis in colorectal cancer', *J Exp Med*, 210(7), pp. 1331–1350. doi: 10.1084/jem.20120812.

Jass, J. R. (2007) 'Classification of colorectal cancer based on correlation of clinical, morphological and molecular features', *Histopathology*. John Wiley & Sons, Ltd (10.1111), 50(1), pp. 113–130. doi: 10.1111/j.1365-2559.2006.02549.x.

Jen, J. *et al.* (1994) 'Allelic Loss of Chromosome 18q and Prognosis in Colorectal Cancer', *New England Journal of Medicine*, 331(4), pp. 213–221. doi: 10.1056/NEJM199407283310401.

Jho, E. H. *et al.* (2002) 'Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway', *Mol Cell Biol*, 22(4), pp. 1172–1183. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11809808.

Jones, R. G. *et al.* (2006) 'Conditional deletion of beta1 integrins in the intestinal epithelium causes a loss of Hedgehog expression, intestinal hyperplasia, and early postnatal lethality', *J Cell Biol*, 175(3), pp. 505–514. doi: 10.1083/jcb.200602160.

Jones, R. J. *et al.* (2002) 'Elevated c-Src is linked to altered cell-matrix adhesion rather than proliferation in KM12C human colorectal cancer cells', *Br J Cancer*, 87(10), pp. 1128–1135. doi: 10.1038/sj.bjc.6600594.

Jung, P. *et al.* (2011) 'Isolation and in vitro expansion of human colonic stem cells', *Nat Med*, 17(10), pp. 1225–1227. doi: 10.1038/nm.2470.

Jung, P. *et al.* (2015) 'Isolation of Human Colon Stem Cells Using Surface Expression of PTK7', *Stem Cell Reports*. The Authors, 5(6), pp. 979–987. doi: 10.1016/j.stemcr.2015.10.003.

Kanagawa, M. *et al.* (2004) 'Molecular recognition by LARGE is essential for expression of functional dystroglycan', *Cell*, 117(7), pp. 953–964. doi: 10.1016/j.cell.2004.06.003.

Kellokumpu, S., Sormunen, R. and Kellokumpu, I. (2002) 'Abnormal glycosylation and altered Golgi structure in colorectal cancer: dependence on intra-Golgi pH', *FEBS Letters*. John Wiley & Sons, Ltd, 516(1–3), pp. 217–224. doi: 10.1016/S0014-5793(02)02535-8.

Kikuchi, A. *et al.* (2011) 'New Insights into the Mechanism of Wnt Signaling Pathway Activation', in *International Review of Cell and Molecular Biology*. Elsevier Inc., pp. 21–71. doi: 10.1016/B978-0-12-386035-4.00002-1.

Kim, D. *et al.* (2013) 'TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions', *Genome Biol*, 14(4), p. R36. doi: 10.1186/gb-2013-14-4-r36.

Kim, K. *et al.* (2000) 'Overexpression of β-catenin induces apoptosis independent of its transactivation function with LEF-1 or the involvement of major G1 cell cycle regulators', *Molecular Biology of the Cell*. American Society for Cell Biology, 11(10), pp. 3509–3523. doi: 10.1091/mbc.11.10.3509.

Kim, K. (2002) ‘DIRECT EVIDENCE FOR A ROLE OF β-CATENIN/LEF-1 SIGNALING PATHWAY IN INDUCTION OF EMT’, *Cell Biology International*. Academic Press, 26(5), pp. 463–476. doi: 10.1006/cbir.2002.0901.

Kim, K. A. *et al.* (2008) ‘R-Spondin family members regulate the Wnt pathway by a common mechanism’, *Mol Biol Cell*, 19(6), pp. 2588–2596. doi: 10.1091/mbc.E08-02-0187.

Kinzler, K. W. *et al.* (1991) ‘Identification of FAP locus genes from chromosome 5q21’, *Science*. American Association for the Advancement of Science, 253(5020), pp. 661–665. doi: 10.1126/science.1651562.

Koenig, M., Monaco, A. P. and Kunkel, L. M. (1988) ‘The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein’, *Cell*, 53(2), pp. 219–228. doi: 10.1016/0092-8674(88)90383-2.

Kolligs, F. T. *et al.* (1999) ‘Neoplastic transformation of RK3E by mutant beta-catenin requires deregulation of Tcf/Lef transcription but not activation of c-myc expression’, *Mol Cell Biol*, 19(8), pp. 5696–5706. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10409758.

Korinek, V. *et al.* (1997) ‘Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma’, *Science*, 275(5307), pp. 1784–1787. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9065401.

Korinek, Vladimir, Barker, N., Moerer, P., *et al.* (1998) ‘Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4’, *Nature Genetics*, 19(4), pp. 379–383. doi: 10.1038/1270.

Korinek, V *et al.* (1998) ‘Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4’, *Nat Genet*, 19(4), pp. 379–383. doi: 10.1038/1270.

Korinek, Vladimir, Barker, N., Willert, K., *et al.* (1998) ‘Two Members of the Tcf Family Implicated in Wnt/β-Catenin Signaling during Embryogenesis in the Mouse’, *Molecular and Cellular Biology*. American Society for Microbiology, 18(3), pp. 1248–1256. doi: 10.1128/mcb.18.3.1248.

Laurent-Puig, P., Béroud, C. and Soussi, T. (1998) *APC gene: database of germline and somatic mutations in human tumors and cell lines*, *Nucleic Acids Research*. Available at: http://perso.curie.fr/tsoussi (Accessed: 5 April 2020).

Laurie, G. W., Leblond, C. P. and Martin, G. R. (1982) ‘Localization of type IV collagen, laminin, heparan sulfate proteoglycan, and fibronectin to the basal lamina of basement membranes’, *Journal of Cell Biology*. The Rockefeller University Press, 95(1), pp. 340–344. doi: 10.1083/jcb.95.1.340.

Leschziner, A. *et al.* (2001) ‘Neural Regulation of α-Dystroglycan Biosynthesis and Glycosylation in Skeletal Muscle’, *Journal of Neurochemistry*. John Wiley & Sons, Ltd, 74(1), pp. 70–80. doi: 10.1046/j.1471-4159.2000.0740070.x.

Li, S. *et al.* (2017) ‘Integrin and dystroglycan compensate each other to mediate laminin-dependent basement membrane assembly and epiblast polarization’, *Matrix Biol*, 57–58, pp. 272–284. doi: 10.1016/j.matbio.2016.07.005.

Liberzon, A. *et al.* (2015) 'The Molecular Signatures Database (MSigDB) hallmark gene set collection', *Cell Syst*. NIH Public Access, 1(6), pp. 417–425. doi: 10.1016/j.cels.2015.12.004.

Litvinov, S. V. *et al.* (1997) 'Epithelial cell adhesion molecule (Ep-CAM) modulates cell-cell interactions mediated by classic cadherins', *Journal of Cell Biology*. The Rockefeller University Press, 139(5), pp. 1337–1348. doi: 10.1083/jcb.139.5.1337.

Logan, C. Y. and Nusse, R. (2004) 'THE WNT SIGNALING PATHWAY IN DEVELOPMENT AND DISEASE', *Annual Review of Cell and Developmental Biology*. Annual Reviews, 20(1), pp. 781–810. doi: 10.1146/annurev.cellbio.20.010403.113126.

Lommel, M. *et al.* (2013) 'Protein O-mannosylation is crucial for E-cadherin-mediated cell adhesion', *Proceedings of the National Academy of Sciences of the United States of America*, 110(52), pp. 21024–21029. doi: 10.1073/pnas.1316753110.

Losasso, C. *et al.* (2000) 'Anomalous dystroglycan in carcinoma cell lines', *FEBS Letters*, 484(3), pp. 194–198. doi: 10.1016/S0014-5793(00)02157-8.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biol*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Lu, W. *et al.* (2009) 'Suppression of Wnt/beta-catenin signaling inhibits prostate cancer cell proliferation', *Eur J Pharmacol*, 602(1), pp. 8–14. doi: 10.1016/j.ejphar.2008.10.053.

Lundgren, K., Nordenskjöld, B. and Landberg, G. (2009) 'Hypoxia, Snail and incomplete epithelial-mesenchymal transition in breast cancer', *British Journal of Cancer*. Nature Publishing Group, 101(10), pp. 1769–1781. doi: 10.1038/sj.bjc.6605369.

Ma, J. and Hart, G. W. (2014) 'O-GlcNAc profiling: From proteins to proteomes', *Clinical Proteomics*. BioMed Central Ltd., pp. 1–16. doi: 10.1186/1559-0275-11-8.

Mani, S. A. *et al.* (2008) 'The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells', *Cell*, 133(4), pp. 704–715. doi: 10.1016/j.cell.2008.03.027.

Marisa, L. *et al.* (2013) 'Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value', *PLoS Med*, 10(5), p. e1001453. doi: 10.1371/journal.pmed.1001453.

Markowitz, S. D. and Bertagnolli, M. M. (2009) 'Molecular origins of cancer: Molecular basis of colorectal cancer', *N Engl J Med*, 361(25), pp. 2449–2460. doi: 10.1056/NEJMra0804588.

Merlos-Suárez, A. *et al.* (2011) 'The intestinal stem cell signature identifies colorectal cancer stem cells and predicts disease relapse', *Cell Stem Cell*, 8(5), pp. 511–524. doi: 10.1016/j.stem.2011.02.020.

Michele, D. E. and Campbell, K. P. (2003) 'Dystrophin-glycoprotein complex: post-translational processing and dystroglycan function', *J Biol Chem*, 278(18), pp.

15457–15460. doi: 10.1074/jbc.R200031200.

Michels, B. E. *et al.* (2019) 'Human colon organoids reveal distinct physiologic and oncogenic Wnt responses', *J Exp Med*, 216(3), pp. 704–720. doi: 10.1084/jem.20180823.

Miller, M. R. *et al.* (2015) 'Downregulation of dystroglycan glycosyltransferases LARGE2 and ISPD associate with increased mortality in clear cell renal cell carcinoma', *Mol Cancer*, 14, p. 141. doi: 10.1186/s12943-015-0416-z.

Moremen, K. W., Tiemeyer, M. and Nairn, A. V. (2012) 'Vertebrate protein glycosylation: Diversity, synthesis and function', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 448–462. doi: 10.1038/nrm3383.

Morikawa, K. *et al.* (1988) 'Influence of organ environment on the growth, selection, and metastasis of human colon carcinoma cells in nude mice', *Cancer Res*, 48(23), pp. 6863–6871. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2846163.

Morin, P. J. *et al.* (1997) 'Activation of β-catenin-Tcf signaling in colon cancer by mutations in β-catenin or APC', *Science*, 275(5307), pp. 1787–1790. doi: 10.1126/science.275.5307.1787.

Moser, A. R., Pitot, H. C. and Dove, W. F. (1990) 'A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse', *Science*, 247(4940), pp. 322–324. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2296722.

Munemitsu, S. *et al.* (1995) 'Regulation of intracellular β-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 92(7), pp. 3046–3050. doi: 10.1073/pnas.92.7.3046.

Muschler, J. *et al.* (2002) 'A Role for Dystroglycan in Epithelial Polarization', *Cancer Res.* American Association for Cancer Research, 59(Suppl.)(7_Supplement), pp. 1757s-1764s.

Muzny, D. M. *et al.* (2012) 'Comprehensive molecular characterization of human colon and rectal cancer', *Nature*. NIH Public Access, 487(7407), pp. 330–337. doi: 10.1038/nature11252.

Naishiro, Y. *et al.* (2001) 'Restoration of epithelial cell polarity in a colorectal cancer cell line by suppression of beta-catenin/T-cell factor 4-mediated gene transactivation.', *Cancer research*, 61(6), pp. 2751–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11289158 (Accessed: 10 April 2020).

Nelson, W. J. (2008) 'Regulation of cell-cell adhesion by the cadherin-catenin complex', in *Biochemical Society Transactions*. NIH Public Access, pp. 149–155. doi: 10.1042/BST0360149.

Nusse, R. and Clevers, H. (2017) 'Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities.', *Cell*. Elsevier, 169(6), pp. 985–999. doi: 10.1016/j.cell.2017.05.016.

Oliveira-Ferrer, L., Legler, K. and Milde-Langosch, K. (2017) 'Role of protein

glycosylation in cancer metastasis', *Seminars in Cancer Biology*. Academic Press, 44, pp. 141–152. doi: 10.1016/J.SEMCANCER.2017.03.002.

Papageorgis, P. *et al.* (2011) 'Smad4 inactivation promotes malignancy and drug resistance of colon cancer', *Cancer Res*, 71(3), pp. 998–1008. doi: 10.1158/0008-5472.CAN-09-3269.

Patterson, A. M. and Watson, A. J. M. (2017) 'Deciphering the Complex Signaling Systems That Regulate Intestinal Epithelial Cell Death Processes and Shedding', *Front Immunol*, 8, p. 841. doi: 10.3389/fimmu.2017.00841.

Peifer, M., Berg, S. and Reynolds, A. B. (1994) 'A repeating amino acid motif shared by proteins with diverse cellular roles', *Cell*. Elsevier, pp. 789–791. doi: 10.1016/0092-8674(94)90353-0.

Perera, P. S., Thompson, R. L. and Wiseman, M. J. (2012) 'Recent Evidence for Colorectal Cancer Prevention Through Healthy Food, Nutrition, and Physical Activity: Implications for Recommendations', *Current Nutrition Reports*, 1(1), pp. 44–54. doi: 10.1007/s13668-011-0006-7.

Perez Villamil, B. *et al.* (2012) 'Colon cancer molecular subtypes identified by expression profiling and associated to stroma, mucinous type and different clinical behavior', *BMC Cancer*. BioMed Central, 12, p. 260. doi: 10.1186/1471-2407-12-260.

Pinho, S. S. *et al.* (2011) 'Modulation of E-cadherin function and dysfunction by N-glycosylation', *Cellular and Molecular Life Sciences*, pp. 1011–1020. doi: 10.1007/s00018-010-0595-0.

Pinho, S. S. and Reis, C. A. (2015) 'Glycosylation in cancer: mechanisms and clinical implications', *Nature Reviews Cancer*, 15(9), pp. 540–555. doi: 10.1038/nrc3982.

Pinto, D. *et al.* (2003) 'Canonical Wnt signals are essential for homeostasis of the intestinal epithelium', *Genes Dev*, 17(14), pp. 1709–1713. doi: 10.1101/gad.267103.

Pinto, D. and Clevers, H. (2005) 'Wnt, stem cells and cancer in the intestine', *Biology of the Cell*, 97(3), pp. 185–196. doi: 10.1042/BC20040094.

Polakis, P. (1995) 'Mutations in the APC gene and their implications for protein structure and function', *Current Opinion in Genetics & Development*. Elsevier Current Trends, 5(1), pp. 66–71. doi: 10.1016/S0959-437X(95)90055-1.

Pollard, P. *et al.* (2009) 'The Apc 1322T mouse develops severe polyposis associated with submaximal nuclear beta-catenin expression', *Gastroenterology*, 136(7), pp. 2204–2213. doi: 10.1053/j.gastro.2009.02.058.

Potten, C. S. and Loeffler, M. (1990) 'Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt', *Development*, 110(4).

Powell, S. M. *et al.* (1992) 'APC mutations occur early during colorectal tumorigenesis', *Nature*. Nature Publishing Group, 359(6392), pp. 235–237. doi: 10.1038/359235a0.

van Raay, T. J. *et al.* (2011) 'Naked1 antagonizes wnt signaling by preventing nuclear accumulation of β-catenin', *PLoS ONE*. Public Library of Science, 6(4). doi:

10.1371/journal.pone.0018650.

Ran, F. A. *et al.* (2013) 'Genome engineering using the CRISPR-Cas9 system', *Nature Protocols*, 8(11), pp. 2281–2308. doi: 10.1038/nprot.2013.143.

Rijsewijk, F. *et al.* (1987) 'The Drosophila homology of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless', *Cell*. Cell, 50(4), pp. 649–657. doi: 10.1016/0092-8674(87)90038-9.

Roberts, D. J. (2000) 'Molecular mechanisms of development of the gastrointestinal tract', *Developmental Dynamics*. John Wiley & Sons, Ltd, 219(2), pp. 109–120. doi: 10.1002/1097-0177(2000)9999:9999<::AID-DVDY1047>3.0.CO;2-6.

Roper, J. *et al.* (2017) 'In vivo genome editing and organoid transplantation models of colorectal cancer and metastasis', *Nat Biotechnol*, 35(6), pp. 569–576. doi: 10.1038/nbt.3836.

Rosenbloom, K. R. *et al.* (2013) 'ENCODE data in the UCSC Genome Browser: year 5 update', *Nucleic Acids Res*. Nucleic Acids Res, 41(Database issue), pp. D56-63. doi: 10.1093/nar/gks1172.

Rosin-Arbesfeld, R. *et al.* (2003) 'Nuclear export of the APC tumour suppressor controls beta-catenin function in transcription', *EMBO J*, 22(5), pp. 1101–1113. doi: 10.1093/emboj/cdg105.

Roth, J. (2002) 'Protein N-glycosylation along the Secretory Pathway: Relationship to organelle topography and function, protein quality control, and cell interactions', *Chemical Reviews*. American Chemical Society , 102(2), pp. 285–303. doi: 10.1021/cr000423j.

Rubinfeld, B. *et al.* (1993) 'Association of the APC gene product with β-catenin', *Science*. American Association for the Advancement of Science, 262(5140), pp. 1731–1734. doi: 10.1126/science.8259518.

Sabates-Bellver, J. *et al.* (2007) 'Transcriptome profile of human colorectal adenomas', *Mol Cancer Res*, 5(12), pp. 1263–1275. doi: 10.1158/1541-7786.MCR-07-0267.

Sanchez-Vega, F. *et al.* (2018) 'Oncogenic Signaling Pathways in The Cancer Genome Atlas Article Oncogenic Signaling Pathways in The Cancer Genome Atlas', *Cell*, 173, pp. 321–337. doi: 10.1016/j.cell.2018.03.035.

Sansom, O. J. *et al.* (2004) 'Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration', *Genes and Development*. Cold Spring Harbor Laboratory Press, 18(12), pp. 1385–1390. doi: 10.1101/gad.287404.

Sato, T. *et al.* (2009) 'Single Lgr5 stem cells build crypt–villus structures in vitro without a mesenchymal niche', *Nature*. Nature Publishing Group, 459(7244), pp. 262–265. doi: 10.1038/nature07935.

Schepers, A. G. *et al.* (2012) 'Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas', *Science*. American Association for the Advancement of Science, 337(6095), pp. 730–735. doi: 10.1126/science.1224676.

Schietinger, A. *et al.* (2006) 'A mutant cheperone converts a wild-type protein into a tumor-specific antigen', *Science*. American Association for the Advancement of Science, 314(5797), pp. 304–308. doi: 10.1126/science.1129200.

Schlesinger, S. *et al.* (2017) 'Adult weight gain and colorectal adenomas—a systematic review and meta-analysis', *Annals of Oncology*, 28(6), pp. 1217–1229. doi: 10.1093/annonc/mdx080.

Schneikert, J. and Behrens, J. (2007) 'The canonical Wnt signalling pathway and its APC partner in colon cancer development.', *Gut*. BMJ Publishing Group, 56(3), pp. 417–25. doi: 10.1136/gut.2006.093310.

Scholer-Dahirel, A. *et al.* (2011) 'Maintenance of adenomatous polyposis coli (APC)-mutant colorectal cancer is dependent on Wnt/beta-catenin signaling', *Proc Natl Acad Sci U S A*, 108(41), pp. 17135–17140. doi: 10.1073/pnas.1104182108.

Schwanhäusser, B. *et al.* (2011) 'Global quantification of mammalian gene expression control', *Nature*. Nature Publishing Group, 473(7347), pp. 337–342. doi: 10.1038/nature10098.

Sengupta, P. K. *et al.* (2013) 'Coordinate regulation of N-glycosylation gene DPAGT1, canonical Wnt signaling and E-cadherin adhesion', *Journal of Cell Science*, 126(2), pp. 484–496. doi: 10.1242/jcs.113035.

Sengupta, P. K., Bouchie, M. P. and Kukuruzinska, M. A. (2010) 'N-glycosylation gene DPAGT1 is a target of the Wnt/β-catenin signaling pathway', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc., 285(41), pp. 31164–31173. doi: 10.1074/jbc.M110.149195.

Sgambato, A. *et al.* (2003) 'Dystroglycan expression is frequently reduced in human breast and colon cancers and is associated with tumor progression.', *The American journal of pathology*. American Society for Investigative Pathology, 162(3), pp. 849–60. doi: 10.1016/S0002-9440(10)63881-3.

Shaheen, R. *et al.* (2013) 'A truncating mutation in B3GNT1 causes severe Walker-Warburg syndrome', *Neurogenetics*. Springer, 14(3–4), pp. 243–245. doi: 10.1007/s10048-013-0367-8.

Sharma, R. P. (1973) 'Wingless, a new mutant in D. melanogaster', *Dros. Inf. Service*, 50, p. 134.

Shen, L. *et al.* (2007) 'Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer', *Proceedings of the National Academy of Sciences of the United States of America*, 104(47), pp. 18654–18659. doi: 10.1073/pnas.0704652104.

Shimokawa, M. *et al.* (2017) 'Visualization and targeting of LGR5+ human colon cancer stem cells', *Nature*. Nature Publishing Group, pp. 1–21. doi: 10.1038/nature22081.

Siegel, R., DeSantis, C. and Jemal, A. (2014) 'Colorectal cancer statistics, 2014', *CA: A Cancer Journal for Clinicians*. American Cancer Society, 64(2), pp. 104–117. doi: 10.3322/caac.21220.

Simon-Assmann, P. *et al.* (2010) 'The role of the basement membrane as a modulator of intestinal epithelial-mesenchymal interactions', in *Progress in Molecular Biology and Translational Science*. Elsevier B.V., pp. 175–206. doi: 10.1016/B978-0-12-381280-3.00008-7.

Simons, B. D. and Clevers, H. (2011) 'Stem cell self-renewal in intestinal crypt', *Experimental Cell Research*. Academic Press Inc., pp. 2719–2724. doi: 10.1016/j.yexcr.2011.07.010.

Singh, J. *et al.* (2004) 'Proteolytic enzymes and altered glycosylation modulate dystroglycan function in carcinoma cells', *Cancer Res*, 64(17), pp. 6152–6159. doi: 10.1158/0008-5472.CAN-04-1638.

Smith, K. J. *et al.* (1993) 'Association between Wild Type and Mutant APC Gene Products', *Cancer Research*. American Association for Cancer Research, 53(12), pp. 2728–2731.

Solanas, G. *et al.* (2011) 'Cleavage of E-cadherin by ADAM10 mediates epithelial cell sorting downstream of EphB signalling', *Nat Cell Biol*, 13(9), pp. 1100–1107. doi: 10.1038/ncb2298.

de Sousa e Melo, F. *et al.* (2017) 'A distinct role for Lgr5+ stem cells in primary and metastatic colon cancer', *Nature*, 543(7647), pp. 676–680. doi: 10.1038/nature21713.

Spiro, R. G. (2002) 'Protein glycosylation: nature, distribution, enzymatic formation, and disease', *Glycobiology*, 12(4), pp. 43R-56R. Available at: https://academic-oup-com.emedien.ub.uni-muenchen.de/glycob/article/12/4/43R/590954 (Accessed: 1 April 2020).

Su, L. K. *et al.* (1992) 'Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene', *Science*, 256(5057), pp. 668–670. doi: 10.1126/science.1350108.

Subramanian, A. *et al.* (2005) 'Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles', *Proc Natl Acad Sci U S A*, 102(43), pp. 15545–15550. doi: 10.1073/pnas.0506580102.

Tang, X. *et al.* (2012) 'Roles of N-glycosylation and lipidation in Wg secretion and signaling', *Developmental Biology*. Academic Press Inc., 364(1), pp. 32–41. doi: 10.1016/j.ydbio.2012.01.009.

Tariq, K. and Ghias, K. (2016) 'Colorectal cancer carcinogenesis: a review of mechanisms.', *Cancer biology & medicine*. Chinese Anti-Cancer Association, 13(1), pp. 120–35. doi: 10.28092/j.issn.2095-3941.2015.0103.

Tetteh, P. W. *et al.* (2016) 'Generation of an inducible colon-specific Cre enzyme mouse line for colon cancer research', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(42), pp. 11859–11864. doi: 10.1073/pnas.1614057113.

Torre, L. A. *et al.* (2015) 'Global cancer statistics, 2012', *CA: A Cancer Journal for Clinicians*, 65(2), pp. 87–108. doi: 10.3322/caac.21262.

Tyanova, S. *et al.* (2016) 'The Perseus computational platform for comprehensive analysis of (prote)omics data', *Nature Methods*. Nature Publishing Group, 13(9), pp. 731–740. doi: 10.1038/nmeth.3901.

Umar, S. (2010) 'Intestinal stem cells', *Current Gastroenterology Reports*. Current Medicine Group LLC 1, pp. 340–348. doi: 10.1007/s11894-010-0130-3.

Vajaria, B. N. and Patel, P. S. (2017) 'Glycosylation: a hallmark of cancer?', *Glycoconjugate Journal*. Springer New York LLC, pp. 147–156. doi: 10.1007/s10719-016-9755-2.

Vargas, D. A. *et al.* (2016) 'The Integrated Role of Wnt/β-Catenin, N-Glycosylation, and E-Cadherin-Mediated Adhesion in Network Dynamics', *PLoS Computational Biology*. Public Library of Science, 12(7). doi: 10.1371/journal.pcbi.1005007.

Varki, A. (1993) 'Biological roles of oligosaccharides: all of the theories are correct', *Glycobiology*, 3(2), pp. 97–130. Available at: https://academic.oup.com/glycob/article/3/2/97/822943 (Accessed: 1 April 2020).

Varki, A. and Gagneux, P. (2017) 'Chapter 7 Biological Functions of Glycans', *Essentials of Glycobiology, 3rd edition*. Cold Spring Harbor Laboratory Press, (Chapter 20), p. Chapter 7. doi: 10.1101/GLYCOBIOLOGY.3E.007.

Vijaya Chandra, S. H. *et al.* (2012) 'A common role for various human truncated adenomatous polyposis coli isoforms in the control of beta-catenin activity and cell proliferation', *PLoS One*. Public Library of Science, 7(4), p. e34479. doi: 10.1371/journal.pone.0034479.

Vogelstein, B. *et al.* (1988) 'Genetic Alterations during Colorectal-Tumor Development', *New England Journal of Medicine*. Massachusetts Medical Society , 319(9), pp. 525–532. doi: 10.1056/NEJM198809013190901.

Vogelstein, B. *et al.* (2013) 'Cancer genome landscapes.', *Science (New York, N.Y.)*, 339(6127), pp. 1546–58. doi: 10.1126/science.1235122.

Vogelstein, B. and Kinzler, K. W. (1999) 'Digital PCR.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 96(16), pp. 9236–41. doi: 10.1073/pnas.96.16.9236.

Vogelstein, B. and Kinzler, K. W. (2004) 'Cancer genes and the pathways they control', *Nature Medicine*, 10(8), pp. 789–799. doi: 10.1038/nm1087.

Weir, M. L. *et al.* (2006) 'Dystroglycan loss disrupts polarity and beta-casein induction in mammary epithelial cells by perturbing laminin anchoring.', *Journal of cell science*. NIH Public Access, 119(Pt 19), pp. 4047–58. doi: 10.1242/jcs.03103.

van de Wetering, M. *et al.* (1991) 'Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box.', *The EMBO Journal*. Wiley, 10(1), pp. 123–132. doi: 10.1002/j.1460-2075.1991.tb07928.x.

van de Wetering, M. *et al.* (1997) 'Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF', *Cell*, 88(6), pp. 789–799.

Available at: http://www.ncbi.nlm.nih.gov/pubmed/9118222.

van de Wetering, M. *et al.* (2002) 'The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells', *Cell*, 111(2), pp. 241–250. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12408868.

Van de Wetering, M. *et al.* (2002) 'The ??-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells', *Cell*. Cell Press, 111(2), pp. 241–250. doi: 10.1016/S0092-8674(02)01014-0.

Williams, A. F. and Barclay, A. N. (1988) *THE IMMUNOGLOBULIN SUPERFAMILY-DOMAINS FOR CELL SURFACE RECOGNITION1,2*, *Ann. Rev. Immunol*. Available at: www.annualreviews.org (Accessed: 8 April 2020).

Williamson, R. A. *et al.* (1997) *Dystroglycan is essential for early embryonic development: disruption of Reichert's membrane in Dag1-null mice*, *Human Molecular Genetics*.

Wu, B., Crampton, S. P. and Hughes, C. C. W. (2007) 'Wnt Signaling Induces Matrix Metalloproteinase Expression and Regulates T Cell Transmigration', *Immunity*. NIH Public Access, 26(2), pp. 227–239. doi: 10.1016/j.immuni.2006.12.007.

Wu, Z. Q. *et al.* (2012) 'Canonical Wnt suppressor, Axin2, promotes colon carcinoma oncogenic activity', *Proceedings of the National Academy of Sciences of the United States of America*, 109(28), pp. 11312–11317. doi: 10.1073/pnas.1203015109.

Xie, T. *et al.* (2012) 'A comprehensive characterization of genome-wide copy number aberrations in colorectal cancer reveals novel oncogenes and patterns of alterations', *PLoS ONE*. Public Library of Science, 7(7). doi: 10.1371/journal.pone.0042001.

Xu, X. *et al.* (2015) 'Snail Is a Direct Target of Hypoxia-inducible Factor 1alpha (HIF1alpha) in Hypoxia-induced Endothelial to Mesenchymal Transition of Human Coronary Endothelial Cells', *J Biol Chem*, 290(27), pp. 16653–16664. doi: 10.1074/jbc.M115.636944.

Yan, D. *et al.* (2001) 'Elevated expression of axin2 and hnkd mRNA provides evidence that Wnt/β-catenin signaling is activated in human colon tumors', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 98(26), pp. 14973–14978. doi: 10.1073/pnas.261574498.

Yang, J. *et al.* (2006) 'Adenomatous polyposis coli (APC) differentially regulates beta-catenin phosphorylation and ubiquitination in colon cancer cells', *J Biol Chem*. 2006/06/27, 281(26), pp. 17751–17757. doi: 10.1074/jbc.M600831200.

Yochum, G. S. (2011) 'Multiple Wnt/ss-catenin responsive enhancers align with the MYC promoter through long-range chromatin loops', *PLoS One*, 6(4), p. e18966. doi: 10.1371/journal.pone.0018966.

Yoshida-Moriguchi, T. *et al.* (2010) 'O-Mannosyl phosphorylation of alpha-dystroglycan is required for laminin binding', *Science*. American Association for the Advancement of Science, 327(5961), pp. 88–92. doi: 10.1126/science.1180512.

Yoshimura, K. *et al.* (2009) ‘Integrin alpha2 mediates selective metastasis to the liver’, *Cancer Res*, 69(18), pp. 7320–7328. doi: 10.1158/0008-5472.CAN-09-0315.

Yoshimura, M. *et al.* (1996) ‘Aberrant glycosylation of E-cadherin enhances cell-cell binding to suppress metastasis’, *Journal of Biological Chemistry*, 271(23), pp. 13811–13815. doi: 10.1074/jbc.271.23.13811.

Zhang, T. *et al.* (2001) ‘Evidence that APC regulates survivin expression: A possible mechanism contributing to the stem cell origin of colon cancer’, *Cancer Research*, 61(24), pp. 8664–8667.

Zhang, Z., Zhang, P. and Hu, H. (2011) ‘LARGE Expression Augments the Glycosylation of Glycoproteins in Addition to α-Dystroglycan Conferring Laminin Binding’, *PLoS ONE*. Edited by R. Chammas. Public Library of Science, 6(4), p. e19080. doi: 10.1371/journal.pone.0019080.

Zhou, Y. *et al.* (2016) ‘The TEAD family and its oncogenic role in promoting tumorigenesis’, *International Journal of Molecular Sciences*. MDPI AG. doi: 10.3390/ijms17010138.

10. Supplemental Material

10.1. Proteins detected by qLC-MS/MS exclusively enriched in HT-29 pTZ LARGE2 cells

Detected in 3/3 samples	MW window [1=high; 6=low]
LAMA2	1
NAA35	2
FANCA	2
FAM83H	3
AHSG	4
EDEM2	4
FUT10	4
SIGIRR	4
EHD2	4
RCN1	5
PSMA8;PSMA7	5
IDH3G	5
S100A4	6
RAB5A	6

Suppl. Table 1: qLC-MS/MS analysis after ectopic expression of LARGE2 in HT-29 cells. Analysis was performed as described in the methods section.

10.2. Information on the patient-derived organoid models used in this study

ID	TISSUE ORIGIN	PATIENT BACKGROUND
NM1 /PDO1	colonic mucosa (colon ascendens)	CRC
NM2 /PDO2	rectal mucosa	CRC
NM3 /PDO3	caecal mucosa	CRC
NM4 /PDO4	colonic mucosa (colon ascendens)	CRC

Suppl. Table 2: List of patient derived organoids from human benign intestinal tissue. **PDO**: Patient derived organoid, **NM**: normal mucosa, **CRC**: colorectal cancer

10.3. Information on the patient-derived tumor organoid models used in this study

ID	TISSUE ORIGIN	LOCATION OF PRIMARY TUMOR	MSI/MSS STATUS	KRAS	NRAS	BRAF	PIK3CA
PDTO1	primary tumor	rectum	MSS	wt	wt	wt	wt
PDTO2	liver metastasis	colon	MSS	wt	wt	wt	wt
PDTO4	primary tumor	caecum	MSS	G12D	wt	wt	wt
PDTO5	liver metastasis	colon	MSS	wt	wt	wt	wt
PDTO7	liver metastasis	rectum	MSS	wt	wt	wt	wt
PDTO9	primary tumor	colon	MSS	G13D	wt	wt	wt
PDTO10	primary tumor	caecum	MSS	wt	wt	wt	wt
PDTO13	primary tumor	rectum	MSS	wt	G12R	wt	E545A
PDTO16	liver metastasis	colon	MSS	wt	wt	wt	wt
PDTO17	primary tumor	colon	MSS	G12R	wt	wt	wt

Suppl. Table 3: List of patient derived organoids from human CRC tissue. **PDTO**: Patient derived tumor organoid. Missing numbers refer to PDTOs not analyzed in this study and these PDTOs were either microsatellite instable (MSI) or showed a limited ex vivo growth capacity insufficient for the experiments in this study. **MSS**: microsatellite stable, **wt**: wild type

10.4. Information on the T3 CRC cases from a TMA CRC used in this study

							score αDG membrane staining	
ArrayID	**T**	**Localization**	**Progression type**	**Follow up**	**Age at Tumor**	**Gender**	**Main tumor area**	**Invasion-front**
1	3	20,9		10,68	70	M	0	0
2	3	21,8	MR	5,75	73	M	2	2
3	3	18,2		9,89	57	W	1	1
4	3	18,2		8,75	57	M	1	1
5	3	18,7		8,77	75	W	2	2
6	3	18,7		7,92	70	M	1	2
7	3	18,7		8,39	53	W	2	2
8	3	18,2		8,07	66	M	1	2
9	3	18,7		10,68	64	M	3	3
10	3	18		10,53	71	M	1	0
11	3	18,7		10,60	73	M	3	3
12	3	18,7		10,12	76	M	0	0
13	3	18,4		9,94	81	W	2	2
14	3	18		5,46	77	M	2	2
15	3	18,7		9,14	60	M	3	3
16	3	18		5,85	89	W	2	1
17	3	18,7		5,77	81	W	2	1
18	3	18,2		5,27	68	M	3	3
19	3	18,6		5,19	83	W	1	1
20	3	18,6		9,56	74	M	2	2
21	3	18,7		10,20	71	W	2	3
22	3	18,7		5,91	83	W	2	1
23	3	18,7		8,94	65	W	3	2
24	3	18,7		5,05	85	W	2	2
25	3	18,7		9,72	63	W	2	2
26	3	18,4		8,59	57	M	1	0
27	3	18,4		8,27	70	W	2	2
28	3	18,2		8,15	48	M	2	1

Suppl. Table 4: CRC Stage T3 cohort. A tissue microarray (TMA) assembled at our institution (Institute of Pathology, LMU Munich) was used to immunohistologically examine the intensity and pattern of α-DG by using the glycosylation specific antibody IIh6 (Santa Cruz). Staining was performed as described in the Materials and Methods section and the quartile scoring mask (0-3) is shown in the results section. **Localization** of the primary tumor is given according to the International Classification of Disease (ICD) coding system. 18: Colonic tumor, 20: Rectal tumor, 21: Anal tumor. **Follow up**: survival follo up in years after initial diagnosis. Scoring of α-DG staining (0-3) was

performed on TMA discs representing the main tumor area or case-matching TMA discs representing the tumor invasion front.

10.5. Information on the identified peptides encoding DAG1 in qLC-MS/MS

	Peptide Sequence	Start/End Position in protein		MW in control cells	MW in LARGE2 overexpression cells
alpha-DG					
Ig like domain 1	SFRVTIPTDLIASSGDIIK	77	95		
	VTIPTDLIASSGDIIK	**80**	**95**	**5**	**1**
	EALPSWLHWDSQSHTLEGLPLDTDKGVHYISVSATR	**102**	**137**	**n/a**	**5**
	GVHYISVSATR	**127**	**137**	**n/a**	**5**
S6 like domain	SFSEVELHNMK	216	226		
	LVPVVNNR	227	234		
	LFDMSAFMAGPGNAK	235	249		
	KVVENGALLSWK	250	261		
	VVENGALLSWK	251	261		
	LGCSLNQNSVPDIHGVEAPAR	262	282		
	EGAMSAQLGYPVVGWHIANK	283	302		
Ig-like domain 2	**GGEPNQRPELK**	**490**	**500**	**3**	**1-3**
	NHIDRVDAWVGTYFEVK	**501**	**517**	**3**	**1**
	IPSDTFYDHEDTTTDK	518	533		
	IPSDTFYDHEDTTTDKLK	**518**	**535**	**3**	**1-3**
	LREQQLVGEK	**540**	**549**	**3**	**1-3**

	EQQLVGEK	**542**	**549**	**3**	**1**
	SWVQFNSNSQLMYGLPDSSHVGK	**550**	**572**	**3**	**1**
	HEYFMHATDK	**573**	**582**	**3**	**1**
	GGLSAVDAFEIHVHR	**583**	**597**	**3**	**1,2**
	AKFVGDPALVLNDIHKK	610	626		
	FVGDPALVLNDIHK	**612**	**625**	**3**	**1,2**
	FVGDPALVLNDIHKK	**612**	**626**	**3**	**1-3**
	KLAFAFGDR	**632**	**640**	**5,6**	**1**
	Peptide Sequence	**Start/End Position in protein**		**control**	**LARGE2 over-expression**
beta-DG					
Trans-membrane domain	EQIAGLSR	672	679		
	EQIAGLSRR	672	680		
	RIAEDDGKPR	680	689		
	IAEDDGKPR	681	689		
	IAEDDGKPRPAFSNALEPDFK	681	701		
	PAFSNALEPDFK	690	701		
	PAFSNALEPDFKATSITVTGSGSCR	690	714		
	ATSITVTGSGSCR	702	714		
	HLQFIPVVPPR	715	725		
Nuclear locatisation site	LTLEDQATFIK	783	793		
	LTLEDQATFIKK	783	794		
	GVPIIFADELDDSKPPPSSSMPLILQEEK	795	823		
Terminal domain	SPPPYVPP	888	895		

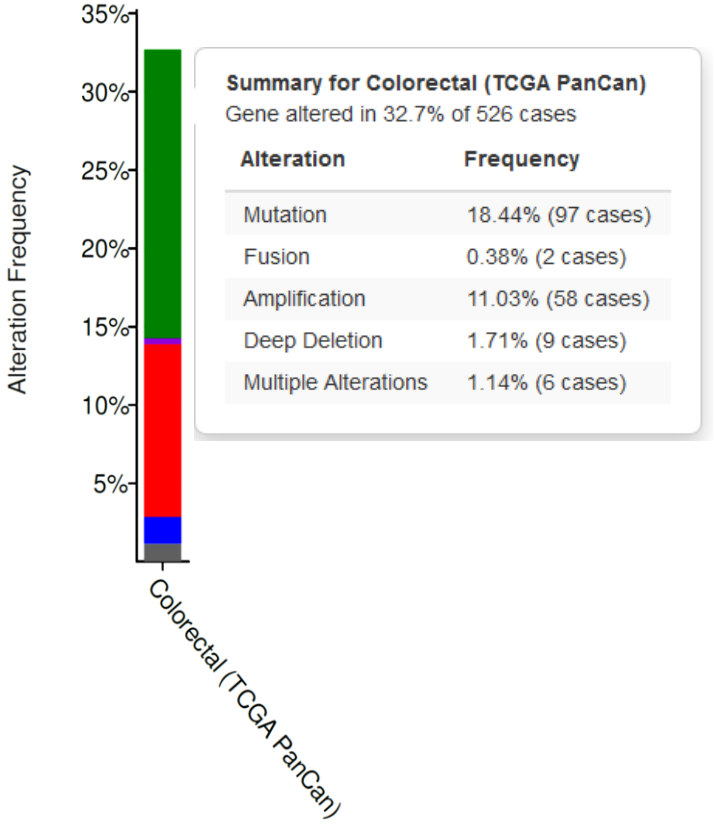
Suppl. Table 5 (see previous page): DAG1 peptide MW shifts after overexpression of LARGE2 in HT-29 cells. Shown are DAG1 peptides as detected by LC-MS/MS. **MW window**: 1=high MW, 6=low MW.

10.6. TCGA analysis of proteins involved in functional glycosylation of α-DG in CRC

% altered in gene	Missense Mutation	Truncating Mutation	Fusion	Amplification	Deep Deletion
DAG1	1,52%	0,38%			0,38%
POMT1	2,47%	0,19%		0,19%	
POMT2	2,47%	0,19%			
POMGNT1	1,71%	0,38%			
POMGNT2	2,47%	0,19%		0,19%	
MGAT5B	2,66%	0,95%		0,76%	
B3GALNT2	0,38%			0,57%	
POMK	0,76%	0,38%	0,19%	3,61%	0,19%
FKTN	1,14%	0,76%			0,19%
FKRP	0,57%				
RXYLT1	0,38%	0,57%			
B4GAT1	1,33%				0,19%
LARGE1	3,04%	0,38%		0,19%	0,76%
LARGE2	1,71%	0,76%		0,19%	
CHST10	3,04%	0,57%			
DPM1	0,57%	0,19%		5,70%	
DPM2	0,19%				
DPM3				0,57%	
DOLK	1,33%	0,19%			
CRPPA	1,52%	0,57%		0,38%	
HK1	2,47%	0,19%	0,19%	0,38%	0,19%
MPI	0,76%	0,38%			
PMM2	0,38%	0,19%			0,19%
GMPPB	0,38%			0,38%	
FURIN	2,47%	0,76%		0,38%	0,19%

Suppl. Table 6: TCGA PanCancer Atlas analysis. We analyzed the genetic status of the indicated genes within the TCGA cohort of 526 colorectal cancer cases accessible via the cBioPortal (*www.cBioPortal.org*). Percentages give the frequency of a given genetic alteration found in 527 CRC cases. Note that LARGE1 and LARGE2 plus 23 factors known to be involved in α-DG glycosylation were each affected in only a small (3 percent and less) fraction. Overall, approximately 18.5% of CRC cases harbored mutations in one or several of these genes, and 1.7% of patients were affected by deep deletions on one of the respective gene loci.

10.7. Graphical summary of altered genes encoding proteins involved in O-glycosylation of α-DG



Suppl. Figure 1: Graphical summary of TCGA PanCancer Atlas analysis. Analysis of Colorectal cancer TCGA cohort. 18.44% of all cases show one or more mutations in the genes involved in functional glycosylation of α-DG (see gene list in Suppl. Table 6).

11. Acknowledgements

First, I want to thank Dr. Peter Jung for supervision and giving me the opportunity to perform my Ph.D. thesis within his group. He gave me the chance to learn plenty of new techniques and improve my technical skills during the past years. Thanks to his never-ending motivation and knowledge, he always supported me with new ideas for the project and my way of scientific thinking was greatly influenced and inspired by discussions with him.

I am also grateful to Prof. Dr. Heiko Hermeking, who agreed to supervise my thesis and always supported my studies with scientific input and help.

Special thanks to Prof. Dr. Dr. Jens Neumann for his collarboration and supply with patient material to set-up our organoid biobank, as well as Dr. Marlies Michl who contributed human cohort samples. I also want to thank Svenja Wiechmann and Prof. Dr. Bernhard Küster for their collaboration and help with sample preparation and analysis of LC-MS/MS.

A very big thank you goes to Ursula Götz, who supported me with protocols, guidance, advice and especially chemicals, whenever I needed something last minute. I also want to thank the rest of the Hermeking lab, for the good times, support in the lab and helpful ideas, as well as other colleagues in the institute who always lend an ear to me.

Besides, I want to thank my colleagues, firstly Sophie Boos, my companion since day 1. Without you, I probably would have never made it this far and I am so thankful that we always had each other's back. A big thank you also goes to Dr. Cira „this mexican" García de Durango for her support, stamina and the joy she brought to our office. I also want to thank Leon Loevenich, the male colleague we were waiting for, who complemented our office with fun and knowledge since the day he started. I think we were a great team and I am so happy that, even on bad days, we always found a way to cheer each other up. Luckily, we never really started the "Mimimimi-Jar" ☺

Additionally I want to thank my friends who always had an open ear for my concerns and supported me along the way, especially my bestie Hanni, Ela and Ani.

Above all, I want to thank my parents for their emotional support throughout the years, the never-ending love and faith in me, as well as Jonas for his shoulder to cry on, patience, understanding, motivation and love.