

University of Limoges
ED 652 - Biologie, Chimie, Santé (BCS)
LABCiS UR22722

A thesis submitted to University of Limoges
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy (Ph.D.)
Discipline: Molecular and Cellular Aspects of Biology

Presented and defended by
Layla HAYMOUR

On December 13, 2022

**Study of the Expression and Role of PAMR1 in
Colorectal Cancer**

Thesis supervisors: Dr. Sébastien LEGARDINIER, Pr. Mona DIAB ASSAF,
and Pr. Abderrahman MAFTAH

JURY:

President of the jury

M. Daniel PETIT, Professor, University of Limoges

Reporters

Mrs. Ikram EL YAZIDI-BELKOURA, Professor, University of Lille

M. Laurent MOREL, Professor, University of Clermont Auvergne

Examiner

M. Mohamed OUZZINE, Research director of INSERM, University of Lorraine



To my family: Dad, Mom,
my sisters Dana and Noor,
and my brother Mohammad

“The important thing is to never stop questioning”
Albert Einstein

Acknowledgements

This dissertation is the result of the efforts and assistance of several persons to whom I am extremely grateful.

First of all, I take immense pleasure in thanking my thesis committee members. Thanks to the reporters Mrs. **Ikram EL-YAZIDI BELKOURA**, Professor at University of Lille, and Mr. **Laurent MOREL**, Professor at University of Clermont Auvergne for accepting and offering their time to review and evaluate this thesis. As examiner, thanks to Mr. **Mohamed OUZZINE**, research director of INSERM at University of Lorraine for his generous time invested to judge my thesis as well. Great thanks to Mr. **Daniel PETIT**, Professor at University of Limoges who accepted to chair my thesis committee.

My special thanks go to my thesis director, at University of Limoges, Pr. **Abderrahman MAFTAH**... thesis director like no other! Thank you for your supervision, guidance, help, and trust over the past three years. You have been always available and responsive to all my requests. You have done your best to ensure I am in good conditions to carry out my thesis. Thank you!

My sincere thanks go to Dr. **Sebastien LEGARDINIER**, lecturer at the University of Limoges, and my thesis supervisor. Thank you for your insightful advice and guidance during the completion of this project. Thank you for the essential scientific discussions that we had to develop and progress the research work. You have been always understanding, available, patient and so helpful. Above all, being in a good mood despite long working days and spending so many weekends in the Lab. I could never imagine having a better advisor. Thank you so much!

I would like as well to thank my thesis co-director at the Lebanese University, Pr. **Mona DIAB ASSAF**. Thank you for placing your trust and confidence in my abilities.

Thanks to Pr. **Vincent SOL**, previous director of PEIRENE laboratory for welcoming me into his Laboratory. Thanks to Pr. **Bertrand LIAGRE** as well, the current LABCiS director.

I would like to thank Mrs. **Nelly VALLAT**, the technical assistant within the team. Thank you for your technical help as well as the delicious crêpes!

I would like to thank Mrs. **Oumaima MAZOUR**, the last Ph.D. student that has joined the team. Thank you for the good times we spent together since your arrival and all our scientific and non-scientific discussions. I wish you all the best in your thesis. I hope you will achieve all your goals in life that you look forward to.

I express my gratitude to Dr. **Agnès GERMOT**, lecturer at the University of Limoges, and Mrs. **Emilie PINAULT**, an engineer at the University of Limoges, for their contribution to a part of my work.

Thanks to Mrs. **Rahima RACHADI** and Mrs. **Claire COLOMBEL LE FAOU** that have done their master's internship within our team. Thanks for your contribution in some experiments.

Thanks to Mrs. **Marlene DESCHUYTER**, the last graduated Ph.D. student within the team and an official Dr. now! Thank you for teaching me some experiments upon my arrival at the lab.

I would like to thank Mr. **Alain CHAUNAVEL**, Clinical Research Associate at the Biological Resources Center of University of Limoges Hospital. Thank you for giving your best and for your collaboration.

I would like to thank each and every member of **LABC*i*S Laboratory** for their direct and indirect contribution to my thesis.

This part wouldn't be complete without acknowledging some very special people to me, whether friends or family.

Jacquie MASSOUD my M2 internship and Ph.D. partner. Thank you for always thinking about me and checking up on me. Thank you for being always there during difficult times before good ones, always next to me. **Nader BAYDA**, I am thankful to have you as a friend in my life. Thank you for always helping me without hesitation. Thanks for both of you for all the good moments, laughs, and memories throughout time. I think the best thing that happened is defending our Ph.D. theses in December 2022. So now I can officially call you Doctors! What I wish you in life is more than what I could express in words. I hope we never fall out of touch.

Thanks to **Sylvana CHAMIEH**, **Rayan CHKAIR**, **Suzan GHADDAR**, and **Roy LAKKIS** for always being there. I wish you nothing but the very best in your life. Thanks to all friends that I have met in Limoges, the list is too long!

Hasan OMAIS, thank you for more than 8 years of friendship. I wish you all the best in your thesis as well.

Last but not least, my deepest and great thanks to my **Family**: Mom, Dad, Dana, Mohammad, and Noor, my support system. I find no words to thank you and to express my feelings. Thank you for giving me all courage, unfailing support, and being always by my side. This accomplishment wouldn't have been possible without you.

Rights

This creation is available under a Creative Commons contract:
« **Attribution-Non Commercial-No Derivatives 4.0 International** »
online at <https://creativecommons.org/licenses/by-nc-nd/4.0/>



Table of Contents

Acknowledgements	4
Rights	6
Table of Contents	7
List of Figures.....	9
List of Tables.....	10
Abbreviations	11
Abstract.....	13
Résumé.....	14
Introduction	16
Chapter 1: Cancers	17
I.1. General Overview of Cancer	17
I.1.1. Cancer from the very beginning – up to date	17
I.1.2. Cancer Nowadays	18
I.1.3. Hallmarks of Cancer	19
I.1.4. Classification of cancer.....	20
I.2. Colorectal Cancer	20
I.2.1. Epidemiology of colorectal cancer	20
I.2.2. Anatomy of colon and rectum	21
I.2.2.1. Colon	21
I.2.2.2. Rectum	21
I.2.3. Development of CRC.....	23
I.2.4. Classification of CRC.....	23
I.2.4.1. Dukes Classification.....	23
I.2.4.2. TNM classification	24
I.2.5. Risk Factors	27
I.2.6. Signs and Symptoms.....	28
I.2.7. Prevention and Screening	28
I.2.7.1. Prevention of CRC	28
I.2.7.2. Screening methods	28
I.2.8. Biomarkers of CRC.....	28
I.2.9. Oncogenes and Tumor suppressor genes.....	30
I.2.9.1. Oncogene mutational activation	30
I.2.9.2. Tumor suppressor gene inactivation	30
I.2.10. Tumorigenesis process in CRC/ Carcinogenesis process.....	30
I.2.10.1. Chromosomal instability (CIN).....	30
I.2.10.2. CpG island methylator phenotype (CIMP)	31
I.2.10.3. Microsatellite Instability (MSI).....	31
Chapter II. Glycosylation	34
II.1. Different glycosylation modifications	34
II.2. N-glycosylation.....	34
II.3. O-glycosylation	36
II.3.1. Different types of O-glycans	36
II.3.2. O-fucosylation mediated by POFUT1 and POFUT2.....	38

II.3.3. POFUT1 mediated O-Fucosylation	40
II.3.3.1. Structure of POFUT1	40
II.3.3.2. EGF like domains.....	40
II.3.3.3. Malfunction of POFUT1	41
Chapter III. POFUT1- target proteins.....	42
III.1. Membrane and secreted glycoproteins O-fucosylated by POFUT1	42
III.2. NOTCH receptors	43
III.2.1. Notch signaling pathway.....	43
III.2.2. Structure of Notch 1.....	45
III.2.3. Notch and O-fucosylation	46
III.3. The Wnt signaling factor 1 (or WIF1).....	46
III.3.1. Wnt signaling pathway.....	46
III.3.2. Structure of WIF1	47
III.3.3. POFUT1-mediated O-fucosylation of mouse WIF1 by and its role in WIF1 secretion	48
III.4. PAMR1	49
III.4.1. Structure and isoforms of PAMR1	49
III.4.2. O-glycosylation of PAMR1 EGF-like domain.....	51
III.4.3. Biological role of PAMR1	51
III.5. Expression of recombinant glycoproteins such as POFUT1 and its target proteins for biochemical analyses and structure-function studies.....	52
III.5.1. Expression of glycoproteins in mammalian expression systems	52
III.5.2. Expression of glycoproteins in the baculovirus insect cell system	53
Chapter IV. PAMR1 in cancers.....	54
IV.1. Expression of PAMR1.....	54
IV.2. PAMR1 in breast and cervical cancer	54
IV.3. Mechanism of action of PAMR1 in cancer	55
IV.4. Signaling pathways linked to PAMR1	55
Chapter V. Objectives and Experimental Approaches	57
Chapter VI. Results	60
VI.1. PAMR1 and Colorectal Cancer.....	60
VI.2. POFUT1-mediated O-fucosylation in the baculovirus expression system	69
Chapter VII. General Discussion	72
Chapter VIII. Conclusion and Perspectives.....	78
Bibliography	81

List of Figures

Figure 1: Oldest evidence of human cancer, osteocarcinoma.....	17
Figure 2: Cancer today based on Globocan statistics.....	18
Figure 3: Cancer Hallmarks.....	19
Figure 4: Estimated number of incident and mortality cases of cancers worldwide.	20
Figure 5: Organization and histology of the distal part of alimentary canal.....	22
Figure 6: Colorectal neoplasia at different stages.	23
Figure 7: Stages of development of CRC	25
Figure 8: General representation of carcinogenesis process or adeno-carcinoma sequence.	33
Figure 9: The three main N-glycans families.....	35
Figure 10: Different O-glycosylations in mammals.....	37
Figure 11: Core structures of O-glycans	38
Figure 12: O-fucosylation of EGF repeat and TSR by POFUT1 and POFUT2 respectively...39	
Figure 13: Crystal structures of POFUT1 and POFUT2 binding their respective substrates .39	
Figure 14: Representation of a single EGF-like domain modified by various glycosyltransferases	41
Figure 15: Notch signaling pathway	44
Figure 16: Structure of the four NOTCH receptor paralogs.....	45
Figure 17: Canonical Wnt/ beta-catenin pathway.....	47
Figure 18: Structure of WIF1 and Shifted.....	48
Figure 19: Scheme representing the different isoforms of human PAMR1	50
Figure 20: Representation of different domains of human and mouse PAMR1	50

List of Tables

Table 1: Colorectal cancer staging system AJCC/TNM, Dukes, and Astler-Coller	26
Table 2: Modifiable and non-modifiable risk factors of CRC.....	27
Table 3: List of biomarkers used for diagnosis.....	29
Table 4: The genetic targets of MSI in CRC	32
Table 5: Table listing the potential human target protein of POFUT1	42

Abbreviations

ACC: Adrenocortical Carcinoma	GlcNAc: N-acetylglucosamine
ADAM: A Disintegrin and Metalloprotease	GSK-3β: Glycogen Synthase Kinase 3 β
AJCC: American Joint Committee on Cancer	GTP: Guanosine Triphosphate
APC: Adenomatous Polyposis Coli	HNPCC: Hereditary Non Polyposis Colorectal Cancer
Asn: Asparagine	ICD: Intracellular Domain
BAX: BCL2 Associated X, Apoptosis Regulator	JAG: Jagged
BRAF: B-Raf proto-oncogene, serine/threonine kinase	KIRC: Kidney renal clear cell carcinoma
CA19-9: Carbohydrate Antigen 19-9	KRAS: Kirsten Rat Sarcoma viral oncogene homolog 2
Ce: <i>Caenorhabditis elegans</i>	LAML: Acute Myeloid Leukemia
CEA: Carcino-Embryonic Antigen	LGG: Brain Lower Grade Glioma
CESC: Cervical and Endocervical carcinoma	LLO: Lipid-Linked Oligosaccharide
CHO: Chinese Hamster Ovary	LNR: Lin-12 Notch Repeats
CIMP: CpG Island Methylator Phenotype	LOH: Loss of Heterozygotie
CIN: Chromosomal Instability	Man: Mannose
CK1α/ϵ: Casein Kinase 1 α/ϵ	MAPK: Mitogen-activated protein kinases
CMP: Cytidine Mono-Phosphate	MESO: Mesothelioma
Cys: Cysteine	MLH1: MultL Homolog 1
COAD: Colon Adenocarcinoma	MLST8: Mammalian Lethal with SEC13 Protein 8
COADREAD: Colorectal Adenocarcinoma	MMP: Matrix Metallopreinase
CRC: Cancer Colorectal	MMR: MisMatch Repair
CUB: Cubulin domain	MSI: Microsatellite Instability
CXCL12: C-X-C motif chemokine ligand 12	MSS: Microsatellite Stable
DDD: Dewling-Degos Disease	mTORC: mammalian Target Of Rapamycin Complex 1
DG: Dystroglycan	NICD: Notch Intracellular Domain
DGC: Dystrophin Glycoprotein Complex	NLS: Nuclear Localization Signal
DII: Delta-like	NRR: Negative Regulatory Region
EGF-LD: Epidermal Growth Factor Like Domain	OGA: O-GlcNAcase
EOGT: EGF specific O-GlcNAc transferase	OGT: O-GlcNAc transferase
ESCA: Esophageal Carcinoma	PBS: Phosphate Buffered Saline
ER: Endoplasmic Reticulum	PCR: Polymerase Chain Reaction
FPA: Familial Adenomatous Polyposis	OV: Ovarian Serous Cystadenocarcinoma
Fuc: Fucose	p53: protein 53
FUT: Fucosyltransferase	PAMR1: Peptidase domain containing Associated with Muscle Regeneration 1
Gal: Galactose	PCPG: Pheochromocytoma, and Paraganglioma
GalNAc: N-acetylgalactosamine	PI3K: Phosphatidyl-Inositol-3-Kinase
GDP: Guanosine Diphosphate	PIK3CA: Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
Glc: Glucose	

POFUT1: Protein O-fucosyltransferase type 1
POFUT2: Protein O-fucosyltransferase type 2
POGLUT: Protein O-glucosyltransferase
Pro : Proline
PTM: Post-Translational Modification
READ: Rectal Adenocarcinoma
RSEM: RNA-Seq Expectation-Maximization
Ser: Serine
SCUBE2: Signal Peptide, CUB Domain And EGF Like Domain Containing 2
Sf: *Spodoptera frugiperda*
SFRP2: Secreted Frizzled Related Protein 2
Sin1: Stress-activated protein kinase-interacting protein
SKCM: Skin Cutaneous Melanoma
SMAD: Suppressor of Mothers against Decapentaplegic
SMC1A: Structural Maintenance of Chromosome 1A
STAD: Stomach Adenocarcinoma
TBS: Tris Buffered Saline

TCF/LEF: T-cell factor/lymphoid enhancer factor
TCGA: The Cancer Genome Atlas
TF: Transcription Factor
TGCT: Testicular Germ cell Tumors
TGF β R: Transforming Growth Factor β Receptor
Thr: Threonine
TIMP: Tissue Inhibitors of Matrix Metalloproteinase
Tn: Thomsen-nouvelle
TNM: Tumor Node Metastasis
TPA: Tissue Polypeptide Antigen
TPS: Tissue Polypeptide-Specific antigen
TSR: Thrombospondin Repeat
UCS: Uterine Carcinosarcoma
UDP: Uridine Di-Phosphate
UVM: Uveal Melanoma
Val: Valine
WHO: World Health Organization
WIF1: Wnt Inhibitory Factor
Xyl: Xylose
 β -TrCP: Beta-transducin repeats-containing proteins

Abstract

Colorectal cancer (CRC) is becoming one of the most prevalent cancers worldwide. This necessitates better understanding the molecular mechanisms behind its occurrence and identifying biomarkers allowing CRC early detection. In this thesis work, we focused our attention on one POFUT1-target protein, namely PAMR1 (Peptidase domain-containing Associated with Muscle Regeneration 1), which is frequently inactivated in breast and cervical cancers and since considered as protein tumor suppressor. We thus wondered if PAMR1 could also exert a similar role in CRC.

Our *in silico* analysis showed a significantly reduced quantity of *PAMR1* in colorectal cancer tissues as early as stage I, indicating that PAMR1 might be an early biomarker of CRC. PAMR1 was also not detected at the protein level in the secretome of CRC cell lines, consistent with the very low transcripts levels expressed by these cells, as determined by qPCR. To study the effect of a supply or an increased expression of PAMR1 in CRC lines, two experimental approaches were carried out, namely exogenous treatments of CRC cell lines with addition of recombinant PAMR1 in growth medium and transient or stable PAMR1 overexpression in HT29 cell line. Using these two approaches, a reduction in HT29 cell proliferation and migration was shown, correlated to a potential tumor suppressor role of PAMR1 in CRC.

For the previous study, we had to use recombinant mouse PAMR1, stably produced by mammalian CHO cells, to treat the CRC lines. Indeed, we did not succeed in producing and purifying a sufficient quantity of recombinant human PAMR1, either in CHO cells or in the baculovirus-insect cell expression system. In the latter system, we showed the inability of POFUT1 from *Sf9* cells to modify the single EGF-like domain of PAMR1 with O-fucose, which could be one of the reasons for its instability and its strong propensity to degradation.

Keywords: PAMR1, colorectal cancer, Proliferation, Migration, POFUT1, *Spodoptera frugiperda*, EGF-like domain.

Résumé

Le cancer colorectal (CCR) est en train de devenir l'un des cancers les plus répandus dans le monde. Cela nécessite de mieux comprendre les mécanismes moléculaires à l'origine de son apparition et d'identifier des biomarqueurs pour une détection précoce du CCR. Dans ce travail de thèse, nous avons porté notre attention sur PAMR1 (protéine à domaine protéase associée à la régénération musculaire 1), une protéine cible de POFUT1 qui est fréquemment inactivée dans les cancers du sein et du col de l'utérus et considérée depuis comme une protéine suppresseur de tumeur. Nous nous sommes donc demandé si PAMR1 pouvait également exercer un rôle similaire dans le CCR.

Notre analyse *in silico* a montré une quantité significativement réduite de *PAMR1* dans les tissus du cancer colorectal dès le stade I, indiquant que PAMR1 pourrait être un biomarqueur précoce du CCR. PAMR1 n'a pas été détecté non plus au niveau protéique dans le sécrétome de lignées cellulaires de CCR, ce qui est cohérent avec les très faibles niveaux de transcrits exprimés par ces cellules, tels que déterminés par qPCR. Pour étudier l'effet d'un apport ou d'une expression augmentée de PAMR1 dans les lignées CCR, deux approches expérimentales ont été menées, à savoir des traitements exogènes des lignées cellulaires CCR avec ajout de PAMR1 recombinant dans le milieu de croissance et la surexpression transitoire ou stable de PAMR1 dans la lignée cellulaire HT29. Par ces deux approches, une réduction de la prolifération et de la migration des cellules HT29 a été montrée, en adéquation avec un potentiel rôle suppresseur de tumeur de PAMR1 dans le CCR.

Pour l'étude précédente, nous avons dû utiliser du PAMR1 recombinant de souris, produit de manière stable par des cellules de mammifères CHO, pour traiter les lignées CCR. En effet, nous n'avons pas réussi à produire et à purifier en quantité suffisante du PAMR1 humain recombinant, que ce soit dans les cellules CHO ou dans le système d'expression baculovirus-cellules d'insectes. Dans ce dernier système, nous avons montré l'incapacité de POFUT1 des cellules Sf9 à O-fucosyler le domaine EGF-like unique de PAMR1, ce qui pourrait être une des raisons de son instabilité et de sa forte propension à la dégradation.

Mots-clés: PAMR1, cancer colorectal, Prolifération, Migration, POFUT1, *Spodoptera frugiperda*, domaine de type EGF.

Chapter 1: Cancers

I.1. General Overview of Cancer

I.1.1. Cancer from the very beginning – up to date

Cancer is not a recent malignancy. It has been born ever since the creation of humans. The first detected cancer case was breast cancer, found in the Edwin Smith Papyrus, in 3000 BC, around 1.7 million years ago (Dunning et al., 1999). Other types of cancers were also described, in 1500 BC, including stomach, uterus, rectum, and skin cancers. At that time, to explain this phenomenon, Egyptians used the term “incurable disease”, or “the curse of God” (Hajdu, 2011). After that, Hippocrates was the first who explained cancer from a scientific point of view, considering it occurs as a result of remarkable presence or increase in the quantity of black bile in the body. “*Karkinos*” is the first word used to refer to “cancer” by Hippocrate (*Hippocrates, and Emile Littré 1839*) (Weiss, 2000), who was considered the father of medicine. He thought that the tumor resembles the crab in the way breast cancer spreads to the skin (Mitrus et al., 2012). The physician Celsus, later translated this word into *cancer*, the Latin word for crab.

Ancients back then considered that once cancer has spread, there is no curable treatment, and the intervention could be more harmful. Galen, a 2nd-century Greek doctor, discussed the resection of cancer in its early stages. However, surgeries were primitive and complicated. Major advances in cancer surgeries started in the 19th and early 20th centuries. William Arbuthnot-Lane is one of the colon cancer surgeons who practiced in London at the turn of the 20th century. At that time, cancer started to be more scientifically understood. Several researchers conducted experiments to confirm that specific chemicals in the workplace and environment were linked to the carcinogenic process, demonstrating the multistage and multifactor nature of the tumorigenesis process.

This investigation and explanation of cancer went through a long path up to a century ago, when Boveri's referred to cancer as a genetic disease (Manchester, 1995). With the development of research, new visions are highlighted, especially nowadays with genetics analysis. These analyses keep open-ended questions that could be resolved only by research and more developed science. Hoping to come up with a clear overview of this devastating disease.

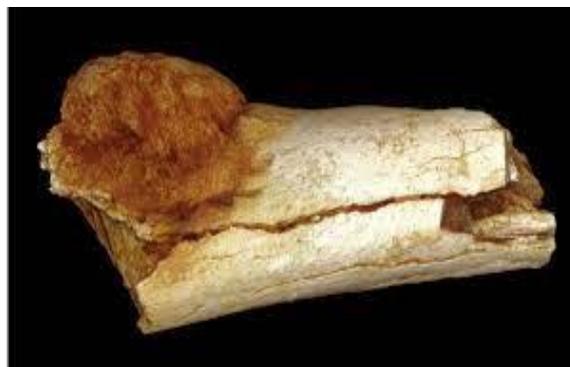


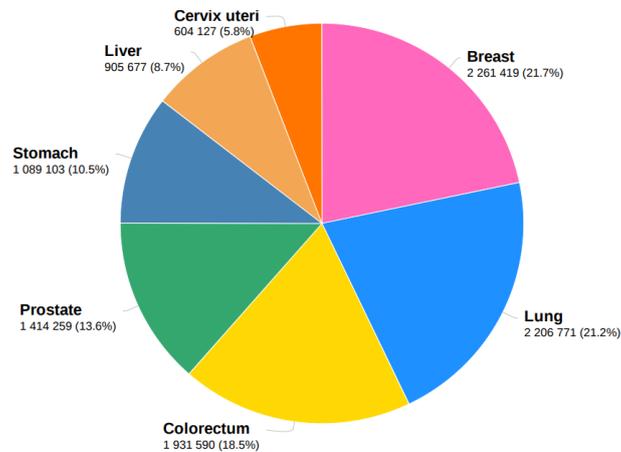
Figure 1: Oldest evidence of human cancer osteocarcinoma. According to National geographic reports

I.1.2. Cancer Nowadays

According to the World Health Organization (WHO), cancer is counted a leading cause of death worldwide, after cardiovascular disease (Bray et al., 2021), accounting for nearly 10 million deaths in 2020.

By referring to the latest Global Cancer Statistics 2020 (WHO), the estimated occurring cancer cases in the world was 19 292 789 cases, where breast cancer ranks first place (Figure 2). Mortality occurred in 50% of the incident cases (9 958 133 cases), in both genders and all ages. Lung cancer ranks the first place in mortality cases (Figure 2). The distribution of cancer differs among population and countries due to various factors affecting it.

Estimated number of new cases in 2020, worldwide, both sexes, all ages



Estimated number of deaths in 2020, worldwide, both sexes, all ages

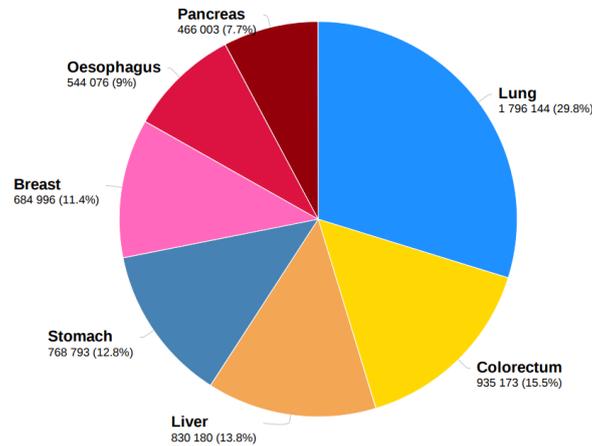


Figure 2: Cancer today based on Globocan statistics

Latest statistics estimating new cancer and new death cases worldwide for both sexes in 2020.

I.1.3. Hallmarks of Cancer

Cancer arises by the excessive proliferation of abnormal cells beyond its normal limit. These cells form a mass called tumor. Tumoral cells tend to break up from the tumor and evade to nearby tissues and can reach all body organs, through blood and lymphatic vessels, this is called metastasis.

Neoplastic diseases occur as a result of genetic instability (Giacomini et al., 2005) that is considered one of the new hallmarks of cancer that enable it to progress toward tumorigenicity and malignancy (Hanahan and Weinberg, 2011) (Figure 3). In addition to the other six hallmarks of cancers (Sustaining proliferative signaling, Evading growth suppressors, Resisting cell death, Enabling replicative immortality, Inducing angiogenesis, Activating invasion and metastasis) (Hanahan and Weinberg, 2000), **reprogramming cellular metabolism** and **avoiding immune destruction** are the two **emerging hallmarks** that were introduced among the acquired capabilities (Hanahan and Weinberg, 2011). “Tumor-promoting inflammation” besides “genome instability and mutation” are the enabling capabilities fundamental for the activation of the eight hallmarks of cancer.

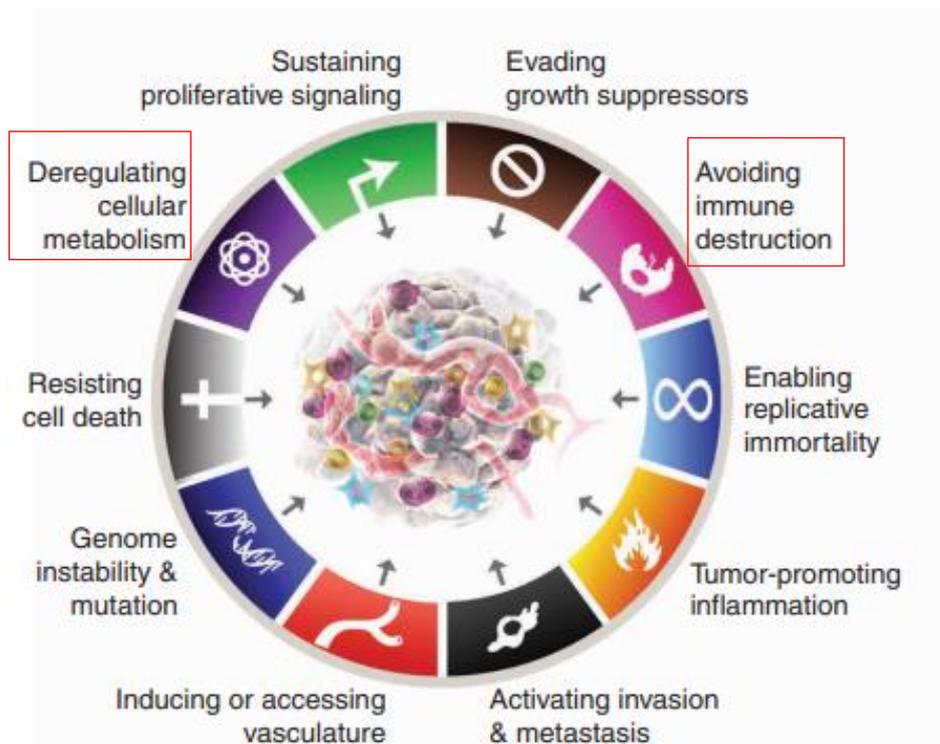


Figure 3: Cancer Hallmarks. According to (Hanahan and Weinberg, 2011)

The eight hallmarks of cancer with the two enabling fundamental characteristics

I.1.4. Classification of cancer

Cancer can be either liquid or solid cancer. Liquid tumors or blood cancer are either leukemias or lymphoma. Leukemia affects the blood and bone marrow whereas lymphomas (Hodgkin's and non-Hodgkin's lymphomas) affect the lymphatic system. Solid cancers are classified into carcinomas and sarcomas. The majority of human cancers (80%) are of epithelial origin, thus named carcinomas. Carcinomas are tumors arising from epithelial cell layer of the gastrointestinal tract (mouth, esophagus, stomach small and large intestine...) as well as skin, mammary gland, liver, pancreas, lung, ovaries, gallbladder, and urinary bladder. The other types of cancer are of mesenchymal origin, named sarcomas. In addition to the non-epithelial tumors. Colorectal cancer falls into adenocarcinoma category that reflects the epithelial functions associated with epithelial and not squamous cell carcinoma (*The Biology of cancer by Robert A. Weinberg, n.d.*).

I.2. Colorectal Cancer

I.2.1. Epidemiology of colorectal cancer

Colorectal cancer is becoming the most prevalent cancer worldwide. According to the Global cancer statistics 2020, colorectal cancer counts for 10% of the occurring cancers (1931590 cases), harboring/ranking the third place in terms of incidence, after breast (11.7%) and lung cancer (11.4%). Colorectal cancer (CRC) is the third most commonly occurring cancer in men and the second most commonly occurring cancer in women (Pitchumoni and Broder, 2020). Colorectal cancer is the second deadly cancer (935173 cases) (9.4%) after lung cancer (18%). (Figure 4)

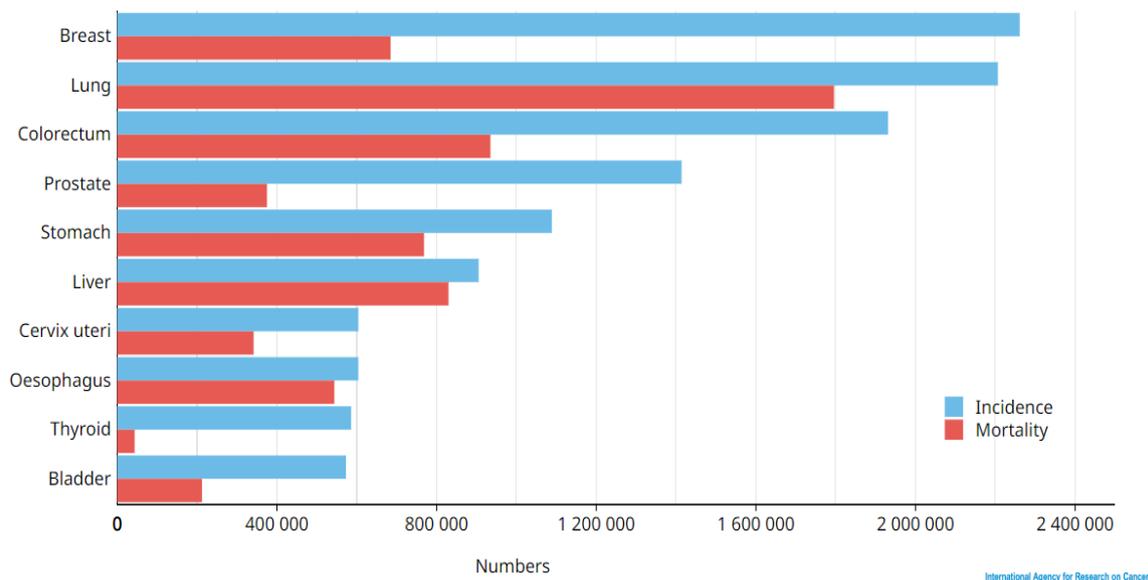


Figure 4: Estimated number of incident and mortality cases of cancers worldwide.

Estimated number of incident and mortality cases of cancers worldwide, both sexes and all ages. According to GLOBOCAN 2020. World Health Organization (<https://gco.iarc.fr>)

I.2.2. Anatomy of colon and rectum

I.2.2.1. Colon

The colon or the large intestine, is the distant part of the digestive system. It follows the small intestine, having the cecum at its proximal part and the anal canal at its distal part. After food digestion, the non-digested/absorbed food arrive to the colon. The colon dehydrates food residues, absorbs electrolytes and vitamins and moves it towards the rectum before being eliminated outside the body through the anal canal in the form of solid waste. The wall of the colon is formed of several distinctive layers from the lumen outward as follows (Figure 5):

- **Mucosa:** comprised of a lining epithelium, a layer of connective tissue known as the lamina propria, and a smooth muscle-filled muscularis mucosa. These structures vary in the various sections of the alimentary canal to accommodate their various roles. The mucosa comprises crypts of Lieberkühn which traverse all the colon. Protection, absorption, and secretion are the mucosa's three primary functions.
- **Submucosa:** consists of irregular connective tissue. It is rich in blood and lymphatic vessels.
- **Muscularis propria:** formed of two concentric layers of smooth muscles. The superficial layer is made up of longitudinal fibers in contrary to the underlying layer formed of circular ones. Contractions of the muscularis externa mix and propel the contents of the digestive tract.
- **Subserosa:** It is an adipose and vascularized tissue that is encompassed by the serosa and the visceral peritoneum in the outer colon layer. It is covered in tiny fatty deposits known as omental appendages.

This horse shoe shaped arc is of 150cm in length and it is subdivided into four main segments at the basis of its anatomic location:

- **Ascending Colon:** it starts superiorly to the cecum and bends into 90 degrees forming right colic flexure (or hepatic flexure). It measures 8 to 15 cm in length.
- **Transverse segment:** It is 50 cm segment that extends from the ascending colon and bends into another 90 degrees forming the Left colic flexure (or splenic flexure).
- **Descending colon:** it is 12 cm segments that descends inferiorly to intersect with the sigmoid colon.
- **Sigmoid colon:** it is the terminal part of the colon, of 40 cm in length. It descends to intersect with the rectum.

I.2.2.2. Rectum

The rectum, distal part of the colon. It starts by the rectosigmoid junction and ends by the anus. It is of about 13-15 cm. The pelvic diaphragm runs perpendicular to the rectal junction and the anal junction, constricting the solid wastes before being excreted outside on the body. This small cavity is divided into three or four parts segmented by valve of Houston. The external muscular sheath that surrounds the rectum allows it to be modified and shorten in length. It comprises the same four layers as the colon. (Figure 5)

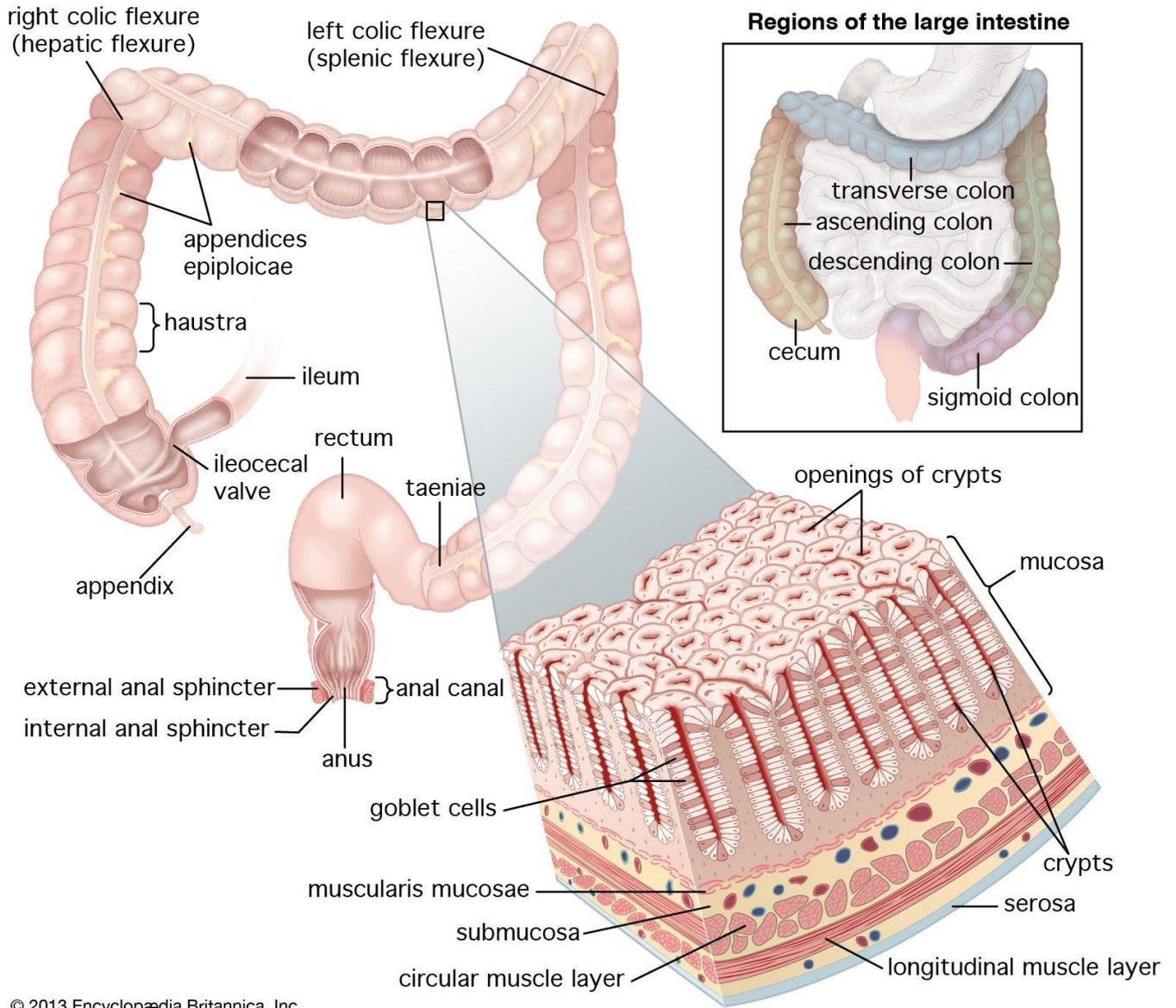


Figure 5: Organization and histology of the distal part of alimentary canal

I.2.3. Development of CRC

Colorectal cancer, or colon cancer, occurs in the colon or rectum. It is a major health problem that develops over a long period of time (5-10) years, starting as benign precancerous polyp (adenoma) (Figure 6a). These polyps are localized growths in the intestinal mucosa (Figure 6b). Excessive cell division can be combined to accumulation of genetic mutations (KRAS, p53...) and/or epigenetic alterations resulting in cytological and histological dysplasia (Figure 6c). If not diagnosed early and removed by surgery, they can acquire the ability to invade beyond the colon/colorectal wall and reach up different body organs such as lungs and liver, this is called metastasis.



Nature Reviews | Disease Primers

Figure 6: Colorectal neoplasia at different stages. According to (Kuipers et al., 2015)

I.2.4. Classification of CRC

I.2.4.1. Dukes Classification

In 1932, an new classification of colorectal cancer (Bowel's cancer) was devised by the British pathologist Cuthbert Dukes (Dukes, 1932). This classification has three main subgroups:

- Dukes A: refers to the presence of the tumor on the inner bowels wall.
- Dukes B: refers to the invasion of the tumor through the muscle layer.
- Dukes C: refers to the migration and invasion into nearby/distant lymph nodes.
- Dukes D: or advanced bowel cancer, refers to the invasion into other body organs or what is called metastasis.

This classification has been modified by the Americans Astler and Coller in 1954 (Astler and Coller, 1954) into a more adapted classification, and more precise description especially for stages B and C. The tumor is:

Stage A: Limited to mucosa

Stage B1: Extending into muscularis layer but not penetrating through it.

Stage B2: Penetrating through muscularis propria.

Stage C1: Extending into muscularis propria but not penetrating through it. Nodes involved

Stage C2: Penetrating through muscularis propria. Nodes involved

Stage D: Distant metastatic spread

This system is now no longer of clinical interest for patients' cancer characterization and it has been replaced by TNM classification.

I.2.4.2. TNM classification

The WHO classification made it possible distinguish group of patients according to their oncological or pathological situation, this provides clear view to establish better targeted treatment and appropriate patients care according to the stage of cancer. The TNM classification, established by **AJCC**, American Joint Committee on Cancer (cancerstaging.org), has been updates in 2018 to the newest Eighth edition (Amin et al., 2017), to a more "personalized" cancer patients classification. TNM is a topographic classification based on the three criteria represented by the acronym **TNM**:

T: refers to the extent and the **size of the primary tumor** with values ranging from 0 to 4 (T1, T2, T3, and T4). T0 indicates the absence of primary tumor, whereas T4 indicates the invasion of the tumor of visceral peritoneum.

N: refers to the invasion of tumoral cells to nearby and/or distant **lymph nodes** with values ranging from 0 to 2 (N0, N1, N2). N0 indicates the absence of tumoral cells in lymph nodes whereas N2 indicates the involvement of lymph nodes located nearby or distant from the tumor.

M: refers to the presence of **Metastasis**. M0 represents the absence of metastases and M1 represents one or more metastases.

Based on TNM subclassification, colorectal cancer can be classified into 4 stages (I-IV) (Table Figure 7, Table 1)

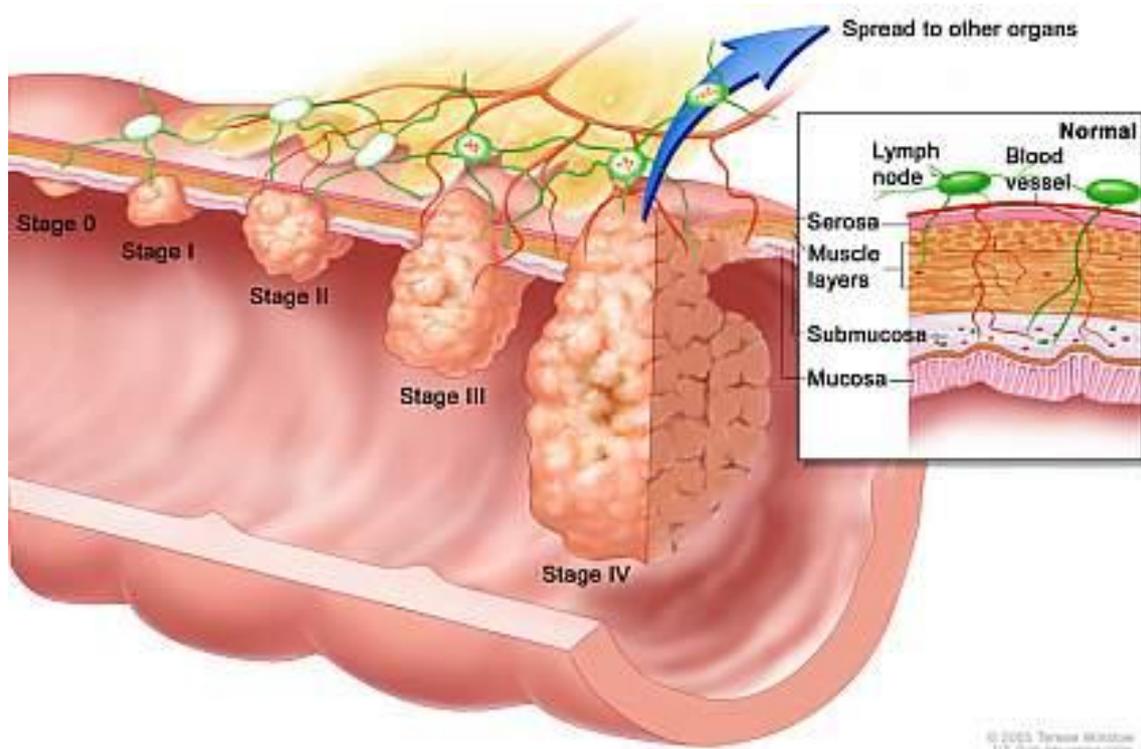


Figure 7: Stages of development of CRC

Colorectal cancer is developed in a localized organ, stage I, up to stage IV, the metastatic stage, where it spreads through the entire body. (<https://www.nih.gov/research-training/advances-colorectal-cancer-research>)

Table 1: Colorectal cancer staging system AJCC/TNM, Dukes, and Astler-Coller

	AJCC/TNM	Dukes	Astler-Coller
Stage 0	Tis, N0, M0	—	—
Stage I	T1-T2, N0, M0	A	A, B1
Stage IIA	T3, N0, M0	B	B2
Stage IIB	T4a, N0, M0	B	B2
Stage IIC	T4b, N0, M0	B	B3
Stage IIIA	T1-T2, N1, M0 T1, N2a, M0	C	C1
Stage IIIB	T3-T4a, N1, M0 T2-T3, N2a, M0 T1-T2, N2b, M0	C	C1, C2
Stage IIIC	T4a, N2a, M0 T3-T4, N2b, M0 T4b, N1-N2, M0	C	C2, C3
Stage IV	Any T, Any N, M1a Any T, Any N, M1b	—	D

It is the most modern and precise system for colorectal cancer. Numbers from 0 to 4 indicate cancer severity. According to (Centelles, 2012).

I.2.5. Risk Factors

Several risk factors promote the occurrence and development of colorectal cancer. They can be grouped into modifiable and non-modifiable ones (Balchen and Simon, 2016) (Table 2):

Table 2: Modifiable and non-modifiable risk factors of CRC

Modifiable Risk factors	Non-modifiable Risk Factors
Dietary pattern: low consumption of vegetables and fibers high meat intake (Angelo et al., 2015).	Familial history of colorectal polyps: Familial Adenomatous Polyps (Jasperson et al., 2010) and hereditary nonpolyposis colon cancer (HNPCC)(Jasperson et al., 2010; Sehgal et al., 2014).
Obesity and overweight: mostly abdominal obesity among both males and females (Corley et al., 2008).	Inflammatory Bowel Disease (IBD) (Kim, 2014).
Cigarette smoking and excessive alcohol consumption (Cho et al., 2015).	Diabetes type 2
Lack of physical exercise (Shaw et al., 2018).	Age: Risk increases after age 50 years (Byrne, 2017).

Controlling the modifiable risk factors lead to a decrease in the chance of developing colorectal cancer contrary to the non-modifiable ones.

The arise of COVID-19 epidemic three years ago (2019) was exhausting for the healthcare system. Covid-19 was linked with adverse effect on all especially old people and/or those with underlying/sensitive health conditions (hypertension, obesity, diabetes...) (Aboueshia et al., 2021). Cancer patients were highly vulnerable for this disease due to their weak immune system as a result of therapies (chemotherapy, radiotherapy, ...). However, a study showed there is no significant difference between cancer and non-cancer patients with Covid-19 in terms of complications and mortality, but cancer patients were hospitalized for a longer time (Aboueshia et al., 2021). Nevertheless, the management of colorectal cancer was affected by this pandemic. The reduction of primary care consultation led to a subsequent reduction of diagnosed colorectal cancer cases. As well as the challenging ability to deliver treatment to already diagnosed patients. Surgeries were limited too to reduce the risk of covid-19 transmission. The high number diagnosed patients with high prognosis as well as the emergency-admitted operations of colorectal cancer reflect the worse outcome of Covid-19 (late diagnosis and more progression of colorectal cancer) (Morris et al., 2021). Thus, several factors affect the development of this malignancy, despite the novel developed targeted therapies strategies. The necessity of designing early diagnosis methods is a major challenge nowadays.

I.2.6. Signs and Symptoms

The occurrence of this disease is accompanied by several signs and symptoms (Astin et al., 2011):

- Change in bowel's habit: prolonged diarrhea, constipation, blood in the stool
- Loss of weight
- Abdominal pain
- Anemia
- The feeling of fatigue and weakness

Note that these signs and symptoms could also be present in patients with benign condition. However, they are critically more pronounced to be selected in high-risk patients. In addition, they can be diagnosed once the tumor has developed in neighboring tissue, thus making the screening and treatment much more difficult. Then, prevention is a must especially for high-risk people without hesitating with periodical inspections.

I.2.7. Prevention and Screening

I.2.7.1. Prevention of CRC

The incidence and mortality of colorectal cancer can be diminished by primary or secondary prevention (Roncucci and Mariani, 2015). Primary care is illustrated by reducing/modifying/inhibiting factors that make people susceptible to high-risk colorectal cancer. This can be done by changing the dietary pattern into high fiber - low meat (red meat) diet, and consumption of dietary micronutrients especially vitamins A, C, and E known to have an anti-oxidant and anti-cancer effect. In addition to increasing physical activity and limiting/reducing smoking and alcohol consumption (Chan and Giovannucci, 2010). Even though colorectal cancer is among the highest in terms of morbidity, but if screened early it has a high risk to be cured.

I.2.7.2. Screening methods

There are three different methods for early detection of CRC: stool-based, imaging, and endoscopic tests. Stool-based tests have the capability of detecting asymptomatic cancerous lesions despite of indicating many false-positive results. However, imaging and endoscopy tests detect precancerous lesions also that can be removed, thus being much more efficient (Hadjipetrou et al., 2017). These tests must be periodically made especially for people with a high risk of developing CRC, thus repressing its development when recognized in early stages. In the case of a positive result, the appropriate treatment will be given for the patient according to several factors. For colorectal cancer, there are blood-based screening methods or biomarkers detection for diagnosis of this cancer.

I.2.8. Biomarkers of CRC

Early detection can be also through the identification of biomarkers of CRC. These biomarkers can be used for personalized therapy and prognosis of CRC. However, there is no specific biomarkers marking CRC exclusively. Among these biomarkers used often to diagnose CRC, besides others, the glycoproteins CEA and CA-19-9. Carcinoembryonic Antigen (CEA) is a blood-based biomarker commonly known to have high sensitivity and specificity for colorectal cancer (Tatsuta et al., 1989) (Tiernan et al., 2013). This cell-surface anchored glycoprotein is usually produced by mucosal cells during fetal development and involved in cell adhesion.

However, high serum levels of CEA could be seen in other cases such as pancreatic, lung, breast, mucinous ovarian cancers, etc (Perkins et al., 2003) (Kankanala and Mukkamalla, 2022). The level of CEA is elevated as well in non-cancerous malignancies (Ruibal Morell, 1992) or even non-malignant cases such as in heavy smokers (Sajid et al., 2007). However, 20-30% of cancer patients show no pronounced increase in CEA despite of their advanced stages (Duffy et al., 2003). Another commonly known cell surface cancer biomarker, CA 19-9 (Carbohydrate Antigen or cancer antigen 19-9 or sialylated Lewis antigen) is used for CRC screening where in 55.4% of the cases could be associated with elevated CEA levels (Stikma et al., 2014). Other serum biomarkers: CA 242, CA 72-4, CA 50, TPA (tissue polypeptide antigen), TPS (tissue polypeptide-specific antigen) and tissue inhibitor of metalloproteinase 1 (TIMP-1) are not limited for this malignancy (Duffy et al., 2003) (Table 3). Recently, Chabanais et al. demonstrated the overexpression of POFUT1 could be considered a novel CRC biomarker (Chabanais et al., 2018). All the above ensure the importance of determining an early specific CRC biomarker for early detection and targeted treatment.

Table 3: List of biomarkers used for diagnosis. The table is adapted from (Lech, 2016)

Biomarker	Applications
CEA	Screening Prognostic factor Follow up
CA 19-9	All
CA 72-4	All
CA 242	All
CA 195	All
CYFRA 21-1	All
MSI	Prognostic factor
18qLOH	Prognostic factor
p53 gene	Prognostic factor
KRAS	Prognostic factor Predictive factor
BRAF	Prognostic factor Predictive factor
PIK3CA	Predictive factor
PTEN	Predictive factor
UGT1A1	Predictive factor
VEGF	All
TPA, TPS	All
Ezrin	All
DNA ploidy	All
TS	Prognostic factor Prognostic factor
TP	All
DPD	Prognostic factor

I.2.9. Oncogenes and Tumor suppressor genes

The initiation and progression of human tumor (tumorigenesis), from adenoma to carcinoma, is a multi-step process. Colorectal cancer arises as a result of oncogene mutational activation in combination with tumor suppressor gene mutational inactivation. The latter predominates since oncogenes mutation exerts an effect even in “recessive” state or heterozygous phenotype. Accumulation of mutation in at least four genes leads to the formation of malignant tumor. The order of occurrence of these mutations with respect to each other define the biological characteristics or feature of the tumor.

I.2.9.1. Oncogene mutational activation

Oncogenes are mutant genes that in their normal nonmutant/non-altered state direct synthesis of proteins that positively regulate proliferation. One important genetic alteration to be found in colorectal cancer is *ras* gene mutation. This mutation is found in about 50% of colorectal carcinomas and predicted to be the initiation step of colorectal tumor development (Bos et al., 1987). In general, oncogenes are activated by point mutation resulting in amplification or oncogenes rearrangement. Among the first oncogenes identified to be amplified: *neu*, *c-myc*, *c-myb* in primary colorectal tumors (Finley et al., 1989).

I.2.9.2. Tumor suppressor gene inactivation

Tumor-suppressor genes encode proteins that in their normal state negatively regulate proliferation. Inactivation of tumor suppressor gene was first predicted to be linked to an allelic loss. The loss of 5q chromosome is associated with tumorigenesis of Familial Adenomatous Polyposis (FAP) as well as pronounced loss in APC loci in hereditary and sporadic colon cancer (Okamoto et al., 1990). The loss of large portion of chromosome 17p is commonly seen in colorectal cancer as well as brain, breast, lung, and bladder. This region is known to bear the tumor suppressor gene *p53* (Baker et al., 1989). Another example is the loss of chromosome 18q containing *DCC*, a tumor suppressor gene codes for protein involved in cell-cell/extracellular matrix interactions (Fearon et al., 1990).

Other ways of dysregulation of tumorigenesis-related genes are also present. The process of activation of some genes and inactivation of others can be referred to different mechanisms.

I.2.10. Tumorigenesis process in CRC/ Carcinogenesis process

Three known molecular mechanisms/ genetic instability mechanisms are behind the occurrence of, most if not all cases, colorectal cancer, whether independently or in a combination, namely: chromosomal instability (CIN), CpG island methylator phenotype (CIMP), and microsatellite instability (MSI). (Tariq and Ghias, 2016)

I.2.10.1. Chromosomal instability (CIN)

Chromosomal instability is a sequence of accumulative gene mutations that are predicted to occur in a specific order leading to changing adenoma into carcinoma. It occurs in 65-70% of sporadic CRC. This process comprises aneuploidy, an imbalance in gene copy number and Loss of Heterozygosity (LOH) (Pino and Chung, 2010). In this sequence, proto-oncogene KRAS and B-Raf proto-oncogene serine/threonine kinase (*BRAF*) are activated whereas tumor suppressor genes are inactivated including *APC* (Adenomatous Polyposis Coli), an important molecular actor in the regulation of Wnt signaling pathway, (loss of chromosome region 5q21), *p53* (chromosome region 17p13), and *DCC* netrin 1 receptor (*DCC*), SMAD family member (*SMAD2* and *SMAD4*) (chromosome region 18q). As predicted, this sequence/ process is

initiated by APC mutation followed by KRAS activation and suppression of p53 activity. Other alterations associated to carcinoma development were identified such as TGFBR and PIK3CA (Armaghany et al., 2012).

I.2.10.2. CpG island methylator phenotype (CIMP)

DNA methylation, whether it is global or localized to the promoter region, is common epigenetic alteration frequently observed in cancers and considered as a subgroup of CRC (about 30% of cases). This epigenetic modification is proposed to be the main tumorigenesis driver in CRC, over genetic alteration, to inactivate tumor suppressor genes. Many hypermethylated genes are of unknown function yet, whereas other are tumor suppressor genes (p16 coded by *CDKN2A*, etc) (Lao and Grady, 2011). In CRC context, hypermethylation of *CXCL12*, a chemokine ligand gene, can promote the metastatic ability of colon cancer cell lines (Lao and Grady, 2011). CpG islands hypermethylation targets miss match repair genes as well, such as *MLH1*, thus suppressing transcription (Ahuja et al., 1997). Recently, inactivation of tumor suppressor genes *PAMR1* in breast cancer (Lo et al., 2015) and *WIF1* in colorectal cancer (Zhu et al., 2018) as a result of promoter hypermethylation has been demonstrated.

I.2.10.3. Microsatellite Instability (MSI)

Microsatellite Instability is another type of genomic instability and it occurs in 15-20% of sporadic CRC. Microsatellites are DNA sequences with one to five base pairs tandem repeats. MSI or high frequency or replicative error within these sequences lead to the generation of longer or shorter alleles (Abdel-Rahman and Peltomäki, 2004). This insertion or deletion mutations lead to frame shift and dysregulation of tumor associated and suppressor genes. Cancers can be classified based on the number of MSI within the 5 standard microsatellite markers: two mononucleotide (BAT26 and BAT25) and three dinucleotide (D2S123, D5S346, and D17S250) repeats, according to the Bethesda Guidelines (Boland et al., 1998). Several loci of microsatellite coding regions have been identified to be mutated in different genes within colorectal cancer context (Mori et al., 2001). These genes are listed in Table 4. Among which, *TGFBR2*, encoding a kinase receptor involved in transduction of the *TGFB1/2/3* signal from the cell surface to the cytoplasm to inhibit cellular proliferation, is the most commonly affected gene. *BAX* is well known mutated pro-apoptotic gene in CRC as well (Boland and Goel, 2010). Several other genes affected by MSI were then identified that encoded regulators of cell proliferation (*GRB1*, *TCF-4*, *WISP3*, *activin receptor-2*, *insulin-like growth factor-2 receptor*, *axin-2*, and *CDX*), the cell cycle or apoptosis (*BAX*, *caspase-5*, *RIZ*, *BCL-10*, *PTEN*, *hG4-1*, and *FAS*), and DNA repair (*MBD-4*, *BLM*, *CHK1*, *MLH3*, *RAD50*, *MSH3*, and *MSH6*) (Duval and Hamelin, 2002).

Table 4: The genetic targets of MSI in CRC. According to (Boland and Goel, 2010)

Microsatellite length	Gene
A10	<i>AIM2</i>
	<i>CASPASE-5</i>
	<i>MBD-4</i>
	<i>OGT</i>
	<i>SEC63 (also, A9)</i>
	<i>TGFβ,R2</i>
A9	<i>BLM</i>
	<i>CHK1</i>
	<i>GRB-14</i>
	<i>MLH3</i>
	<i>RAD50</i>
	<i>RHAMM</i>
	<i>RIZ (also, A8)</i>
	<i>TCF-4</i>
A8	<i>WISP3</i>
	<i>ACVRII</i>
	<i>APAF</i>
	<i>BCL-10</i>
	<i>hG4-1</i>
A6	<i>MSH3</i>
	<i>PTEN (2 A6's)</i>
T10	<i>OGT</i>
T9	<i>KIAA0971</i>
	<i>NADH-UOB</i>
G8	<i>BAX</i>
	<i>IGF2R</i>
C9	<i>SLC23A1</i>
C8	<i>MSH6</i>
G7	<i>AXIN-2 (A6, A6, C6)</i>
	<i>CDX2</i>
T7	<i>FAS</i>

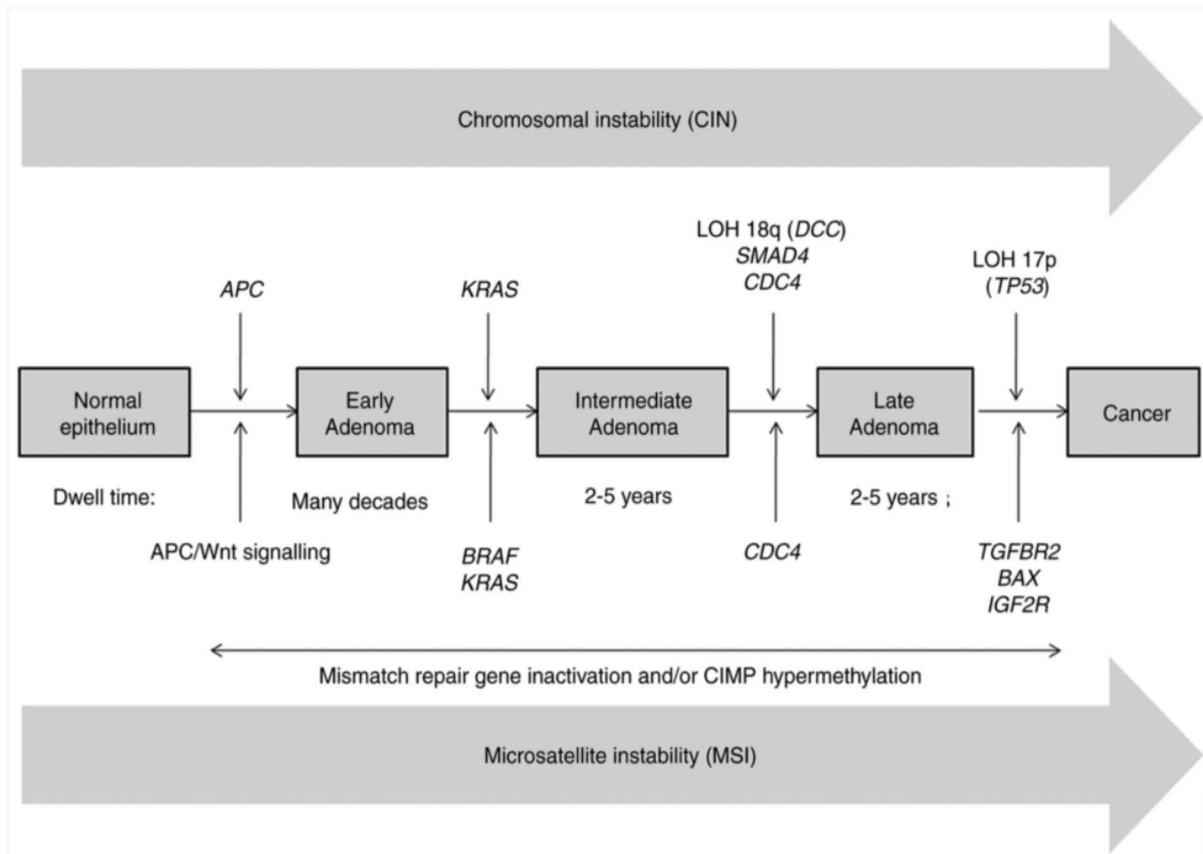


Figure 8: General representation of carcinogenesis process or adeno-carcinoma sequence.

Adeno-carcinoma transformation as a result of three pathways: CIN, CIMP, and MSI. The scheme is adapted from (Nguyen and Duong, 2018)

Tumorigenesis initiation and progression is not limited to the latter three mentioned pathways. Glycosylation is well known biomarker of tumor progression. Incomplete glycosylation due to dysregulation of one of the actors of glycosylation mechanism aids in the progression of oncogenesis. This mechanism is detailed in the next chapter.

Chapter II. Glycosylation

II.1. Different glycosylation modifications

Protein glycosylation is a post-translational modification (PTM) by which sugar moieties are attached to a specific peptide consensus sequence leading to the formation of glycoprotein. Different glycosidic linkages, including *N*-, *O*-, and *C*-linked glycosylation, and less frequent *S*-glycosylation, define this PTM. All glycosylations are built, via glycosyltransferases in the presence of a donor substrate, from a single monosaccharide (fucose, mannose, galactose, *N*-acetylglucosamine, *N*-acetyl galactosamine, etc.) which can subsequently be extended by other sugars to give different glycan structures. Besides glycosyltransferases, glycosidases are crucial to complement the glycosylation process. Glycosidases are responsible for the hydrolysis of specific sugars and monosaccharide precursors during glycans maturation (Kötzler et al., 2014). The glycosylation process has been well preserved throughout evolution, as it is found in archaea, bacteria, and eukaryotes (Dell et al., 2010). In humans, over 70% of human proteins are known to be glycosylated (An et al., 2009). Glycosylation is essential to monitor protein folding, stability, transport, and activity of the protein

With regard to protein modification, glycosylation is primarily of two types: *N*-linked glycosylation with the attachment of the reducing end of the glycan to an Asn residue, and *O*-linked glycosylation with the attachment of the reducing end of the glycan to a hydroxyl side-chain-containing amino acid, most commonly Ser or Thr and less commonly hydroxylysine or hydroxyproline.

II.2. N-glycosylation

N-glycosylation is known to be abundant and highly conserved during evolution, especially in eukaryotes. *N*-glycosylation is defined as the covalent attachment of an oligosaccharide to the nitrogen of an asparagine residue's amide group (*N*-glycosidic linkage). This sugar moiety would be added only to a specific consensus sequence, within the peptide sequence, where it is in most of the cases (96.5%) as follows: \underline{N} -X-(S/T) (X is any amino acid except proline). In rare cases, this sequence could be *N*-X-C (1.3%), *N*-X-V, or *N*-G (Zielinska et al., 2010). Other non-canonical consensus sequences could be present too.

Glycosyltransferases involved in the synthesis of the dolichol oligosaccharide precursor of *N*-glycans are primarily found in the cytosol and the lumen of the ER whereas those involved in the maturation of *N*-glycans of glycoproteins are found in the Golgi apparatus. *N*-glycosylation is a two-step process: the formation/creation of a lipid-linked oligosaccharide (LLO) and its transfer to selected asparagine residues on the consensus sequence of the polypeptide chains. The assembly of lipid-linked oligosaccharide (LLO) occurs on both sides of the Endoplasmic Reticulum (ER) membrane, via various glycosyltransferases, with the involvement of a lipidic base, dolichol. The LLO is comprised of three major carbohydrate components: two *N*-acetylglucosamine (GlcNAc), nine mannose (Man), and three glucose (Glc) residues (GlcNAc₂Man₉Glc₃). The formed LLO is then translocated via oligosaccharyltransferase (OST) to the asparagine of the polypeptide chain, thus forming the *N*-glycan linkage. Thanks to the glycosidases, some sugars are then trimmed by hydrolysis followed by the addition of others, in ER or Golgi apparatus, thus forming complex *N*-glycans (Bieberich, 2014). Nevertheless, all *N*-glycans retain the main structure composed of two *N*-acetylglucosamine and three mannose residues.

After the glycoprotein passes through the Golgi apparatus, it is modified by different glycosyltransferases, at the base of the pentasaccharide core (GlcNAc₂Man₃), resulting in

three main families of *N*-glycans (Figure 9): Oligo-mannosidic or paucimannose, hybrid and complex (Stanley et al., 2022). Oligomannosidic (or high mannose) *N*-glycans bear only mannose antennae extension. The antennae are initiated from GlcNAc and extend the core in complex *N*-glycans, whereas mannose extends through Man α 1-6 arm and one or two GlcNAc-initiated antennae extend the Man α 1-3 arm. Paucimannose *N*-glycans are simple *N*-glycans with a core that could be modified by fucose or xylose. *N*-glycans are crucial for the glycoprotein biological function, folding, secretion, and modulation of signaling pathways, but also its involvement in immunological processes (such as activation of T lymphocytes, etc)... (Varki, 2017).

In vitro, the three types of *N*-glycans can be removed by the bacterial enzyme peptide-N-glycosidase F (PNGase F) or PNGase A (from almonds). (Stanley et al., 2022)

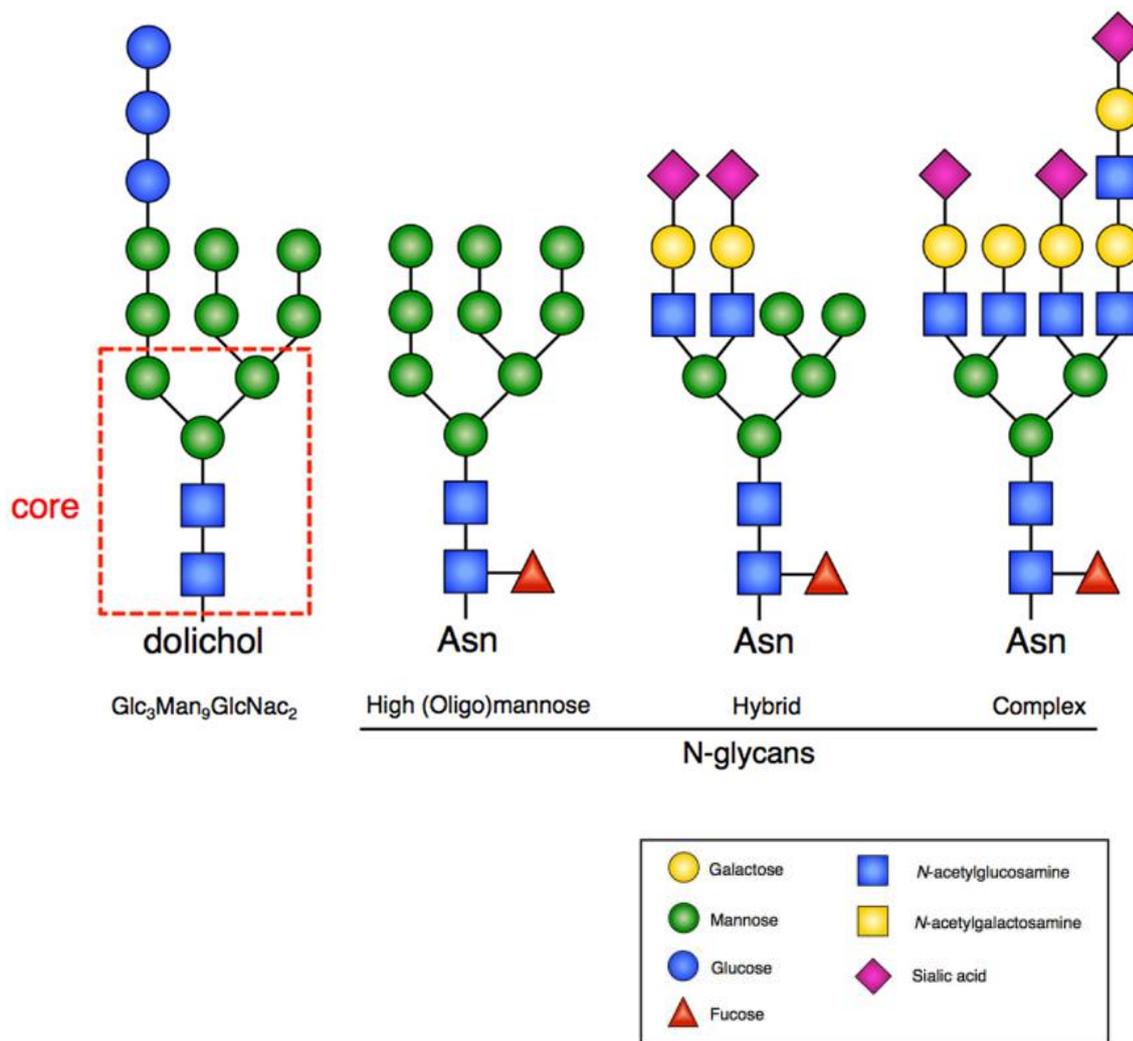


Figure 9: The three main N-glycans families

Representation showing the different forms of N-glycans with increasing complexity from left to right. The scheme is adapted from (Lyons et al., 2015).

II.3. O-glycosylation

Murine-like O-glycosylation is a posttranslational modification characterized by the conjugation of an ose on the hydroxyl side chain of serine or threonine amino acids, and less frequently to hydroxylysine or hydroxyproline (usually in plants). The most common type of O-glycosylation was O-acetylgalactosamine (O-GalNAc) which occurs in 80% of secreted proteins. However, further research recognized other classes of O-glycosylation initiated by the O-linkage of other sugar, other than acetylgalactosamine: mannose, fucose, glucose, galactose, N-acetylglucosamine, and xylose (Wells and Feizi, 2019). The biosynthesis of O-linked glycosylation either starts in the ER and continues in the Golgi apparatus, or takes place exclusively in the Golgi Apparatus.

II.3.1. Different types of O-glycans

Other than O-glycosylation (O-GlcNAc), O-glycosylation can be either protein-specific (O-Man), or domain-specific ((O-Fuc, O-Glc, extracellular O-GlcNAc, and O-Gal) or Intracellular glycosylation modifying nuclear, cytosolic, and mitochondrial glycoproteins (Wells and Feizi, 2019) (Figure 10). These O-linked monosaccharides are capable of being elongated by different oses forming more complex glycans.

The simplest form of mucin O-glycans is formed of attached GalNAc to the serine or threonine and named Tn (Thomsen Nouvelle). This O-GalNAc can be extended by various sugar moieties forming different “core” structures that are counted to be 8 cores (Brockhausen et al., 2009) (Figure 11). An example: Gal β 1-3GalNAc- is the most common O-GalNAc and is referred to as core 1.

O-mannosylation is known to be conserved in bacteria and humans, it is mediated by protein O-mannosyltransferases (POMT). Dysregulation of these enzymes can lead to muscular dystrophy (MD). Dystrophin-Glycoprotein Complex (DGC) is a multimeric transmembrane protein found in skeletal muscle. It plays a role in maintaining the structural stability of sarcolemma during muscular contraction. Dystroglycan (α DG) is an integral membrane component of DGC where mannosylation is crucial for its activity. Defects in POMT1/2 leads to damage of muscle fibers, resulting in MD (Barresi and Campbell, 2006). Neurological disorders can be a result of these enzymes dysregulation as well. (Larsen et al., 2019).

O-glucosylation is less familiar than most O-linked glycosylation, it refers to the attachment of O-glucose to Epidermal growth factors repeats (EGF-like) of proteins, especially NOTCH. It is mediated by O-glucosyltransferases, including POGLUT1 known to add O-glucose on serine of consensus sequence C¹-X-S-X-(P/A)-C² of EGF-like domain. O-glucosylation is crucial for the regulation of notch trafficking (Yu and Takeuchi, 2019). Recently it has been demonstrated that ER-resident POGLUT2/3 mediate the addition of O-Glc between C3 and C4 of the EGF-like domain consensus sequence (Takeuchi et al., 2018) (Pennarubia et al., 2021).

Another O-linked glycosylation restricted to the EGF-like domain of glycoprotein is O-N-acetylglucosamine mediated by EOGT and OGT that differ by their cellular localization and protein targets. O-GlcNAc mediated by endoplasmic reticulum EGF-domain specific O-GlcNAc transferase (EOGT). In mammals, EOGT enhances Notch signaling pathway mediated by a Delta-like ligand by targeting EGF-like domain consensus sequence of secreted and membrane proteins implicated in this pathway (Ogawa and Okajima, 2019). On the other hand, OGT modifies nuclear and cytoplasmic proteins with O-GlcNAc (Kreppel et al., 1997).

O-Galactose glycan (O-Gal) is initiated in ER and catalyzed by galactosyltransferase on hydroxylysine residues on collagen-like domain. It can be extended by glucose (Hennet, 2019).

We are especially interested in O-fucosylation. Fucose can also be added to the hydroxyl group of Serine or Threonine in a covalent O-linkage manner. It was first demonstrated in 1975 in glycopeptide isolated from human urine (Hallgren et al., 1975). O-fucose is added on consensus sequences of Epidermal Growth Factor-like (EGF-like) repeats and Thrombospondin Type 1 Repeats (TSRs) in properly folded proteins, catalyzed by Protein O-Fucosyltransferase 1 (POFUT1 or FUT12) and Protein O-Fucosyltransferase 2 (POFUT2 or FUT13) (Hofsteenge et al., 2001) (Chen et al., 2012) respectively (Holdener and Haltiwanger, 2019). By that POFUT1 and POFUT2 mediate the anchorage directly to the polypeptide. On the other hand, among 13 FUTs discovered in humans, there are Golgi apparatus localized-Fucosyltransferases that modify N-linked glycans. FUT1 (H enzyme) and FUT2 (secretor enzyme) are α -1,2-fucosyltransferases responsible for the formation of ABH and Lewis blood group antigens. FUTs3–7 together with FUTs9–11 have α -1,3-fucosyltransferase activities. FUT3 and FUT5 have α -1,4-fucosyltransferase activity. FUT8 is an α -1,6-fucosyltransferase that is responsible for the N-glycan core fucosylation by adding fucose to asparagine-linked GlcNAc moieties. (Shan et al., 2019).

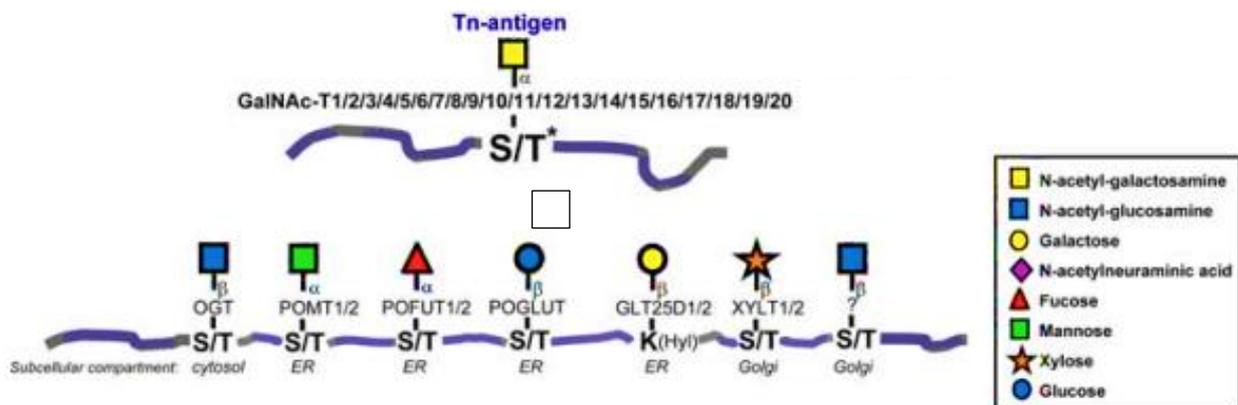


Figure 10: Different O-glycosylations in mammals

Various types of protein O-glycosylations in mammals with their initiating enzymes. The scheme is adapted from (Bennett et al., 2012).

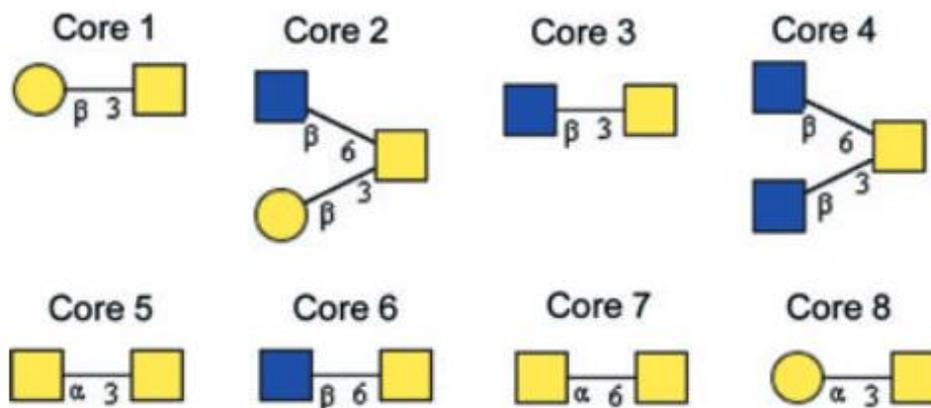


Figure 11: Core structures of O-glycans

The eight different core structures of O-glycans attached to Serine or Threonine hydroxyl group of glycoproteins. These cores are then processed and extended. The figure is adapted from (Schedin-Weiss et al., 2014)

II.3.2. O-fucosylation mediated by POFUT1 and POFUT2

The consensus sequence for O-fucosylation by POFUT1 and 2 are not identical. POFUT2 is able to recognize and target the two types of TSR domains (TSR1 and TSR2) or types 1 and 2 despite the different positions where the three conserved disulfide bridges are formed. For TSR1, the disulfide bonds occur between C¹-C⁵, C²-C⁶, and C³-C⁴ while these bonds are between the cysteines C¹-C⁴, C²-C⁵, C³-C⁶ for type 2 of TSRs (Leonhard-Melief and Haltiwanger, 2010). As for POFUT1, O-fucosylation is established by POFUT2 on a specific consensus sequence of TSR: C¹-X-X-(S/T)-C² for TSR I, C²-X-X-(S/T)-C³ for TSR II (Holdener and Haltiwanger, 2019). For POFUT1, O-fucose is anchored to serine and threonine of the following consensus sequence C²-XXXX-(S/T)-C³, where C² and C³ are the 2nd and 3rd conserved cysteine of EGF-like domain. The fact that POFUT1 and POFUT2 only modify properly folded modules and are ER localized has led to the hypothesis that both enzymes participate in quality control (Vasudevan and Haltiwanger 2014). The anchored O-fucose to EGF repeats of TSRs is capable of being elongated with different glycosyltransferases (Figure 12). The crystal structures of both enzymes POFUT1 and POFUT2 are similar in a way and differ in another (Figure 13). Both enzymes harbor a huge substrate binding cavity for the binding of GDP-fucose as well as EGF and TSR in case of POFUT1 and POFUT2, respectively. Interestingly, POFUT2 has a second substrate binding cavity, by that POFUT2 has the advantage of binding and modifying two adjacent TSRs simultaneously. (Vasudevan and Haltiwanger, 2014)

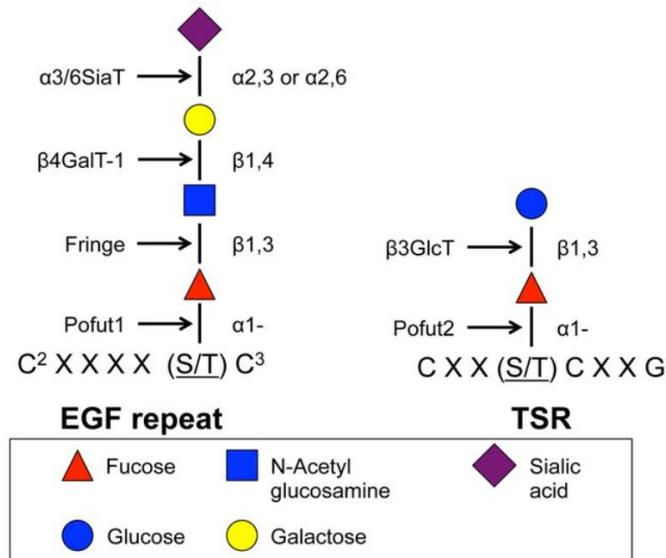


Figure 12: O-fucosylation of EGF repeat and TSR by POFUT1 and POFUT2 respectively.

The O-fucose on TSRs is elongated by β 3-glycosyltransferase (β 3GlcT). O-Fucose on EGF repeats is extended by Fringe, β 4-galactosyltransferase 1 (β 4GalT-1) and α 2-3/6-sialyltransferase (α 3/6SiaT). According to (Vasudevan and Haltiwanger, 2014)

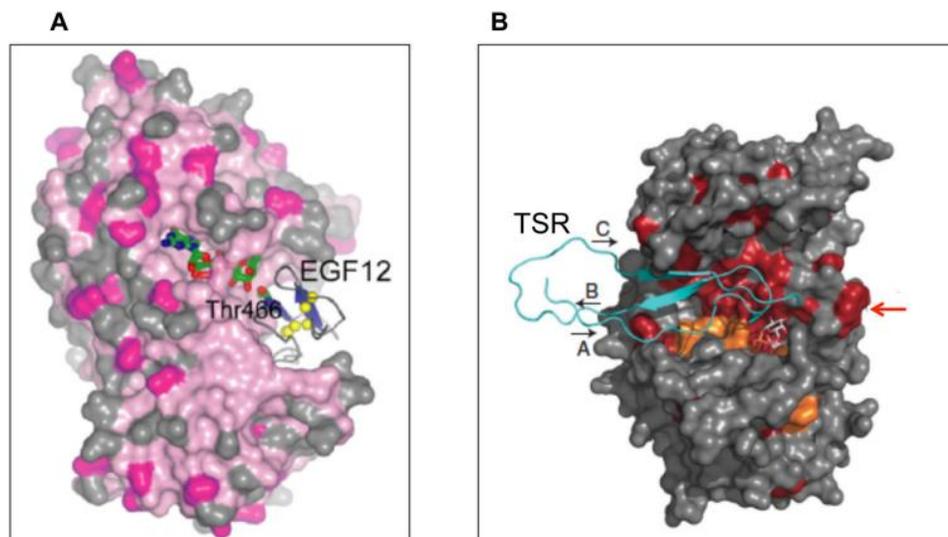


Figure 13: Crystal structures of POFUT1 and POFUT2 binding their respective substrates

A: POFUT1 has two Rossmann-like folds large enough to hold GDP-fucose at first and EGF as well. The modification site on EGF12 (Thr466) is indicated, demonstrating it is in close proximity to the active site. **B:** POFUT2 has a large substrate binding pocket to hold both GDP-fucose and TSR. It is characterized by having another pocket (red arrow) allowing binding to another adjacent TSR. According to (Vasudevan and Haltiwanger, 2014)

II.3.3. POFUT1 mediated O-Fucosylation

II.3.3.1. Structure of POFUT1

POFUT1, or OFUT1, is an ER-resident protein encoded by the 30778 bp *POFUT* gene located on 20q11.21. Five transcripts arise as a result of alternative splicing, identified by the ENSEMBL database, where only the first two of them encode for proteins. The first transcript encodes for the canonical isoform (isoform 1) of POFUT1 (388 amino acids). The second transcript encodes for a truncated isoform 2 of POFUT (194 amino acids) at the C-terminal with undetermined function yet ("The Status, Quality, and Expansion of the NIH Full-Length cDNA Project," 2004).

POFUT1 crystalized structure was first obtained in 2011 (Lira-Navarrete et al., 2011). This protein structure is conserved among CePOFUT1 (*Caenorhabditis elegans* POFUT1) and higher eukaryotes POFUT1, with about 41% of identity (Lira-Navarrete et al., 2011).

POFUT1 is formed of two main domains *N*-terminal and *C*-terminal domains. The *N*-terminal domain contains 9 α -helices and 8 β -strands. The *C*-terminal domain contains 10 α -helices and 5 β -strands. Each of these domains adopts a Rossman folding composed of a β sheet surrounded by α helices on either side giving POFUT1 signature of GT-B fold. This structure is characterized by a formed pocket in which the acceptor substrate can be housed. This conformation is conserved between species (Lira-Navarrete et al., 2011).

POFUT1 possess *N*-terminal signal peptide, three binding domains to GDP-fucose (donor substrate), two *N*-glycosylation sites, and an ER-retention KDEL-like sequence. The three conserved motifs (I, II, and III) of POFUT1, presenting a good correlation to α 1,2 and α 1,6 fucosyltransferases, are involved in acceptor substrate recognition (motif I) as well as recognition and binding to donor substrate (motifs I and II) (Mollicone et al., 2009). In addition to DxD peptide sequence that has a crucial role for the enzyme catalytic activity, the *N*-glycosylation sites (N^{62} and N^{60} in humans) are crucial for the enzyme activity in humans. The first *N*-glycosylation site is highly conserved among species (*Spodoptera frugiperda*, *Homo sapiens*, *Mus musculus*...). The second site is less conserved among species and found to be absent in *Drosophila* POFUT1 and CePOFUT1.

II.3.3.2. EGF like domains

EGF like domains are conserved protein domains during the evolution, found in many secreted proteins or in the extracellular domain of membrane proteins such as NOTCH. EGF-like domains are composed of 30-40 amino acid sequence stabilized by three conserved disulfide bridges formed between C¹-C³, C²-C⁴, and C⁵-C⁶. It is consisted of major *N*-terminal and minor *C*-terminal domain formed by two beta sheets. EGF like domains are classified mainly into two large groups or types: human EGF like domain (hEGF) and complement C1r-like EGF (cEGF) that differ by the number of residues residing between cysteins 5 and 6. hEGFs have between 8 and 9 residues, while that cEGFs have more than 9 residues (Li et al., 2017). Rare EGF like domains laminin and integrin like domains characterized by having an additional interdomain disulfide bridge (Wouters et al., 2005).

A hEGF-like domain can be modified by POFUT1 in addition to other modifications by POGLUTs and EOGT. Focusing on *O*-fucosylation, the interaction between POFUT1 and hEGF-like domain has been studied *in vitro*. The *O*-fucosylation of NOTCH receptor, WIF1 and POFUT1 is highlighted in the following part.

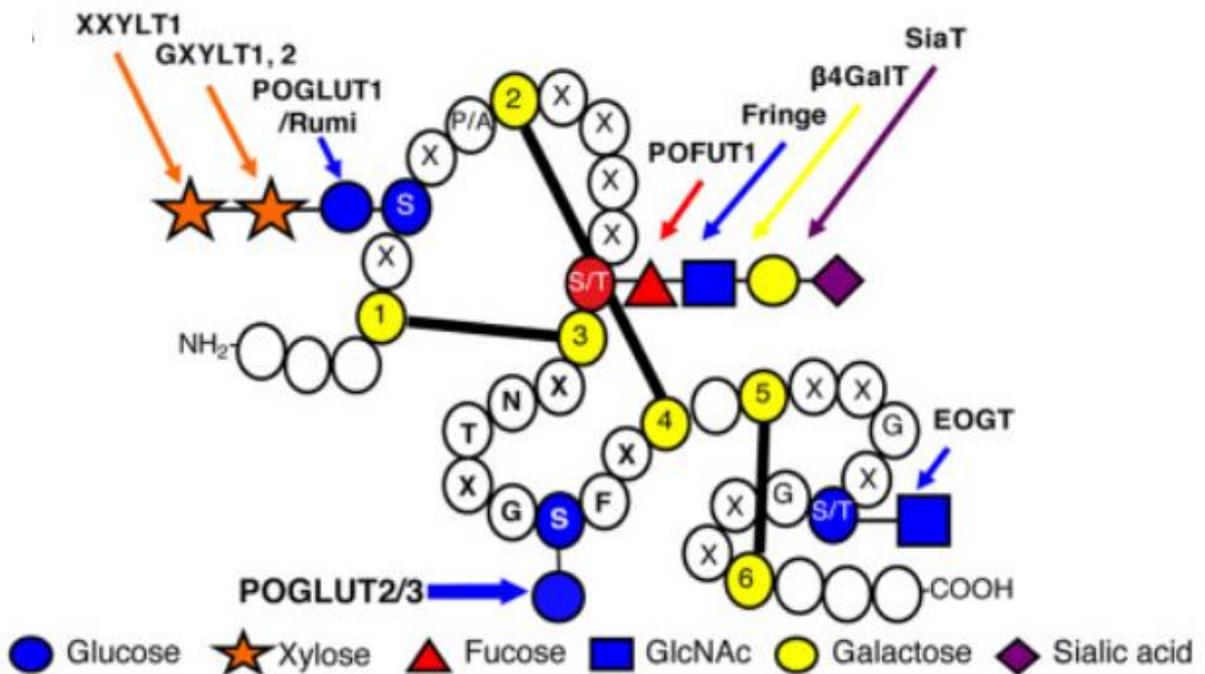


Figure 14: Representation of a single EGF-like domain modified by various glycosyltransferases

Each circle represents amino acid of EGF like domain. The yellow circles are conserved cysteine residues forming disulfide bridges. O-fucosylation site in red, O-Glc and O-GlcNAc in blue. The enzymes implicated are indicated on the top of the arrows. (S) Serine; (T) Threonine; (G) Glycine; (W) Tryptophan; (X) any amino acid, (a) any aromatic amino acid. The figure is adapted from (Schneider et al., 2017).

II.3.3.3. Malfunction of POFUT1

POFUT1 is essential for regulation of defined signaling pathways. Homozygous deletion of POFUT1 in mice is lethal in the early stages of development with a severe defects In somitogenesis, vasculogenesis, cardiogenesis, and neurogenesis reflecting its crucial function in notch signaling pathway of mammals (Shi and Stanley, 2003). This reflects the indirect effect of POFUT1 through the O-fucosylation of core members of this pathway or other implicated ones. Consistently, homozygous missense mutation of *pofut1* generated severe defects in development, cardiogenesis, vasculogenesis mild liver failure, and osteopenia (Takeuchi et al., 2018). An alteration of the expression of POFUT1 is associated with various diseases and malignancy. In humans, the knockout of POFUT1 in human keratinocytes led to the decrease of KRT5 expression, thus Dowling-Degos Disease, a rare autosomal dominant genodermatosis that leads to the loss of skin pigmentation (Li et al., 2013). This suggests that POFUT1 is essential for the synthesis of melanin in cells as well as melanocytes and keratinocytes differentiation, besides POGLUT1 and others (Stephan et al., 2021)

Chapter III. POFUT1- target proteins

III.1. Membrane and secreted glycoproteins O-fucosylated by POFUT1

About 87 human proteins bore one or more EGF-like domains with O-fucose consensus sequence C²-X-X-X-X-(S/I)-C³ to be modified by POFUT1. Schneider *et al.*, identified POFUT1 target proteins (Table 5); however, modifications are not confirmed for all of them. Of these proteins, 13 have been shown to carry an O-fucose (in blue and orange) and participate in the functional activity for 2 of them (orange) (Schneider *et al.*, 2017).

Table 5: Table listing the potential human target protein of POFUT1

Gene	Protein	Gene	Protein
AGRIN	Agrin	LTBP2	latent-transforming growth factor beta-binding protein 2
ATRAID	All-trans retinoic acid-induced differentiation	MEGF6	Multiple epidermal growth factor-like domains protein 6
CELSR1	Cadherin EGF LAG seven-pass G-type receptor 1	MEGF8	Multiple epidermal growth factor-like domains protein 8
CELSR2	Cadherin EGF LAG seven-pass G-type receptor 2	MEGF10	Multiple epidermal growth factor-like domains protein 10
CELSR3	Cadherin EGF LAG seven-pass G-type receptor 3	MEGF11	Multiple epidermal growth factor-like domains protein 11
CD93	Complement component C1q receptor	MMRN1	Multimerin-1
CD97	CD97 antigen	NCAN	Neurocan core protein
CFCL	Cryptic protein	NELL1	Protein kinase C-binding protein NELL1
CFCLB	Cryptic family protein 1B	NID2	Nidogen-2
CNTNAP5	Contactin-associated protein-like 5	NOTCH1	Neurogenic locus notch homolog protein 1
CRB1	Protein crumbs homolog 1	NOTCH2	Neurogenic locus notch homolog protein 1
CRB2	Protein crumbs homolog 2	NOTCH2NL	Notch homolog 2 N-terminal-like protein
CSPG2	Versican core protein	NOTCH3	Neurogenic locus notch homolog protein 3
CUBN	Cubilin	NOTCH4	Neurogenic locus notch homolog protein 4
DLK1	Protein delta homolog 1	PEAR1	Platelet endothelial aggregation receptor 1
DLK2	Protein delta homolog 2	PGBM	Basement membrane-specific heparan sulfate proteoglycan core protein
DLL1	Delta-like protein 1	PGCB	Brevican core protein
DLL3	Delta-like protein 2	PROC	Vitamin K-dependent protein C
DLL4	Delta-like protein 4	PROZ	Vitamin K-dependent protein Z
DNER	Delta and Notch-like epidermal growth factor-related receptor	RAMP	Regeneration-associated muscle protease homolog
EDIL3	EGF-like repeat and discoidin I-like domain-containing protein 3	RELN	Reelin
EGF	Pro-epidermal growth factor	SLIT1	Slit homolog 1 protein
EGFL7	Epidermal growth factor-like protein 7	SLIT2	Slit homolog 2 protein
EGFLAM	Pikachurin	SLIT3	Slit homolog 3 protein
EMR1	Adhesion G protein-couplet receptor E1	SNED1	Sushi, nidogen and EGF-like domain-containing protein 1
EMR2	Adhesion G protein-couplet receptor E2	SREC2	Scavenger receptor class F member 2
EYS	Protein eyes shut homolog	STAB1	Stabilin-1
FA7	Coagulation factor VII	STAB2	Stabilin-2
FA9	Coagulation factor IX	SUSD1	Sushi domain-containing protein 1
FA12	Coagulation factor XII	SVEP1	Polydrom protein
FAT1	Protocadherin Fat 1	TEN1	Teneurin-1
FAT2	Protocadherin Fat 2	TEN2	Teneurin-2
FAT3	Protocadherin Fat 3	TEN4	Teneurin-4
FAT4	Protocadherin Fat 4	TIE1	Tyrosine-protein kinase receptor Tie-1
FBLN1	Fibulin-1	TPA	Tissue-type plasminogen activator
FBLN7	Fibulin-7	TSP3	Thrombospondin-3
FBN2	Fibrillin-2	UMOD	Uromodulin
FBN3	Fibrillin-3	UMODL1	Uromodulin-like 1
HABP2	Hyaluronan-binding protein 2	UROK	Urokinase-type plasminogen activator
HGFAC	Hepatocyte growth factor activator	VASN	Vasorin
JAG1	Protein Jagged-1	VWDE	von Willebrand factor D and EGF domain-containing protein
JAG2	Protein Jagged-2	VWA2	AMACO
LRP1	Prolow-density lipoprotein receptor-related protein 1	WIF1	Wnt inhibitory factor 1
LRP1B	Low-density lipoprotein receptor-related protein 1B		

This table has been updated in our Laboratory listing target proteins and their number of EGF-like domains targeted by glycosyltransferases: POGlut1/2/3, POFUT1, EOGT of *Mus musculus* and *Homo sapiens*. Supplementary data (Pennarubia *et al.*, 2021).

Among the above listed target protein, O-fucosylation of Notch receptor, WIF1 and PAMR1 will be highlighted.

III.2. NOTCH receptors

III.2.1. Notch signaling pathway

The Notch signaling pathway is an ancient conserved evolutionary mechanism of cell-cell interaction in multicellular organisms. The notch pathway plays a role in cell fate determination and differentiation during development as well as regulation of homeostasis and cell proliferation in adult tissue. Notch is trans-activated by ligands expressed on adjacent cell surfaces namely: Delta/Serrate in *Drosophila* and Delta-like/jagged in Vertebrates. Receptor-bound ligand undergoes conformational changes and is endocytosed by ligand expressing cells. This induces proteolytic cleavage and the release of the receptor intracellular domain that acts as a transcription factor/regulator in the nucleus after activation of the intracellular protein cascade. This ligand-dependent mechanism is referred to as the canonical Notch signaling pathway. Another ligand-independent mechanism is also present and referred to as noncanonical Notch signaling pathway. Here we will focus on the canonical pathway, detailed in Figure 15 (Steinbuck and Winandy, 2018).

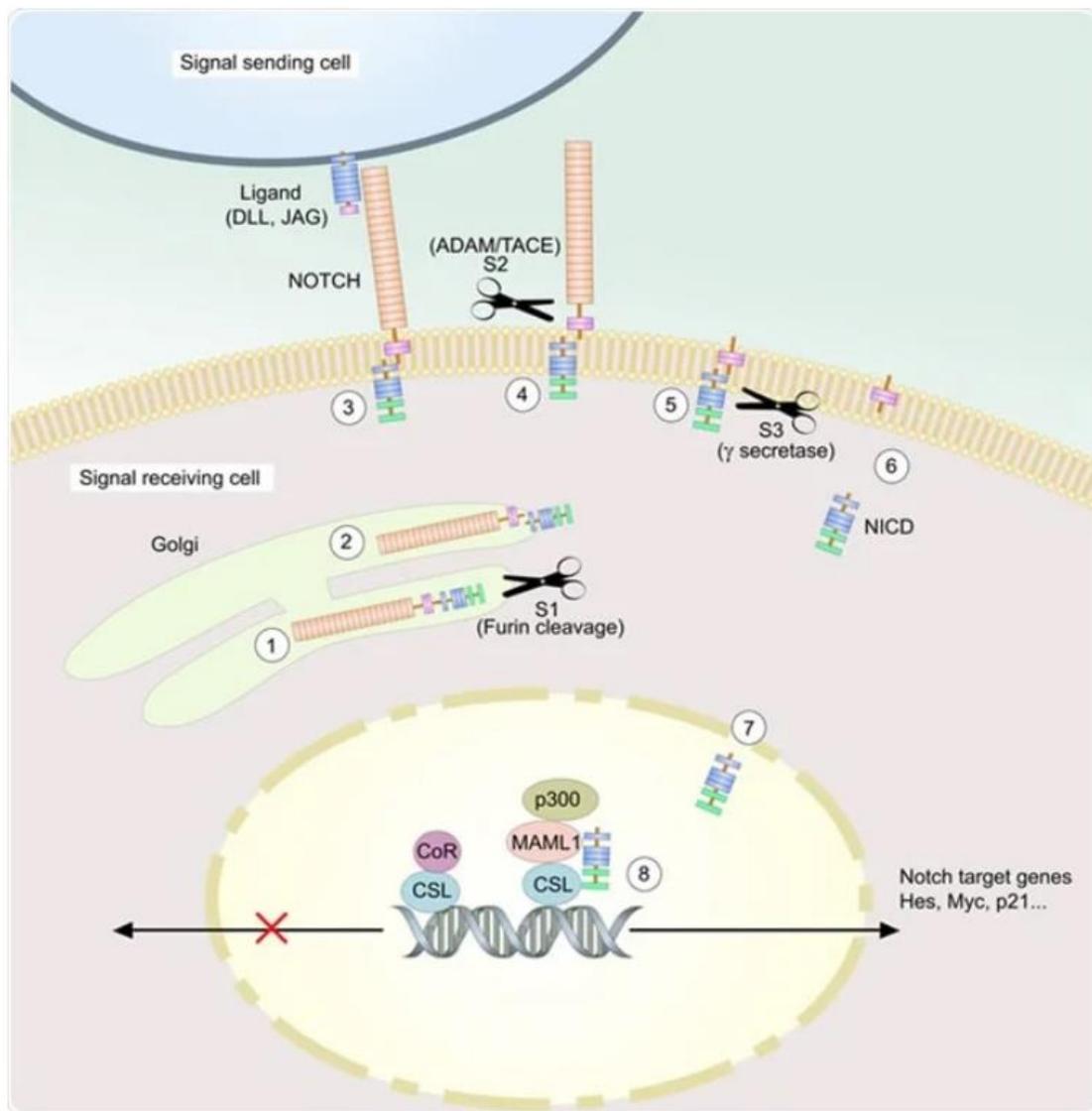


Figure 15: Notch signaling pathway

(1) After O-glycosylation of NOTCH, mainly O-fucosylation in the ER, it is translocated to Golgi Apparatus and lengthened by other glycosyltransferases. (2) It is cleaved in the Golgi apparatus by Furin-like protease generating heterodimeric mature receptor that is then translocated to the plasma membrane. (3) The receptor on the signal receiving cell interacts with the ligand (DLL or JAG) of the signal sending cell leading to receptor conformational changes and exposing its intracellular part to the action of ADAM/TACE protease (4). (5) The intracellular part is exposed to another cleavage by gamma-secretase and the release of Notch Intracellular Domain (NICD) (6). (7) The NICD is transported to the nucleus and interacts with other transcription factors (CSL, MAML1, p300...) (8). Transcription of Notch target genes is activated (Hes, Myc, p21...). The figure is adapted from <https://www.news-medical.net/life-sciences/Notch-Signaling-Pathway.aspx>.

III.2.2. Structure of Notch 1

Notch receptors are single-pass transmembrane proteins, belong to transmembrane family type-I, and consist of two domains: Notch extracellular and intracellular domains. Mammals possess four paralogs of Notch receptors: Notch1, Notch2, Notch3, and Notch4.

Notch glycosylated extracellular domain (NECD) or subunit consists of 29 to 36 EGF-like tandem repeats depending on the paralog/type of notch receptor. These EGF-like domains are responsible for the interaction with the ligand. For example, in *Drosophila*, EGF-like domains 11 and 12 mediate ligand binding to the receptor. The majority of EGF-like domains are involved in the binding of calcium thus participating in the binding capacity of the receptor with its ligand (Cordle et al., 2008). EGF-like repeats are followed by a negative regulatory region (NRR) consisting of three LNR (Lin-12/Notch repeat) repeats and a hydrophobic region crucial for heterodimerization (HD) of the receptor. LNR is rich in cysteine and mediates Ca^{2+} noncovalent binding of the extracellular and intracellular domains. NRR regions prevent ADAM metalloproteases (ADAM17/TACE, ADAM10) to access the S2 cleavage site (Sanchez-Irizarry et al., 2004). The transmembrane domain (TMD) ends with a "translocation stop" signal composed of 3 to 4 residues Arginine/Lysine. This domain also contains the S3 cleavage site targeted by γ -secretase leading to the release of the intracellular domain of NOTCH (NICD) (Fortini, 2009). The NICD consists of a RAM sequence of 12 to 20 amino acids. Near the RAM region, there are seven ankyrin repeats (ANK) limited by two Nuclear Localization Sequence (NLS), where both are involved in the interaction with CSL (CBF1, Suppressor of Hairless, Lag-1)(Yuan et al., 2012), followed by NOTCH cytokine response (NCR) region and a transactivation domain (TAD). Finally, NICD has a PEST sequence at the C-terminus that is associated with ubiquitin-mediated degradation (Mao and Ito, 2017). For ligands of Notch receptors, they are grouped into two families the Delta-like family (DII1, 3, and 4) and the Jagged family (JAG1 and 2). (Figure 16)

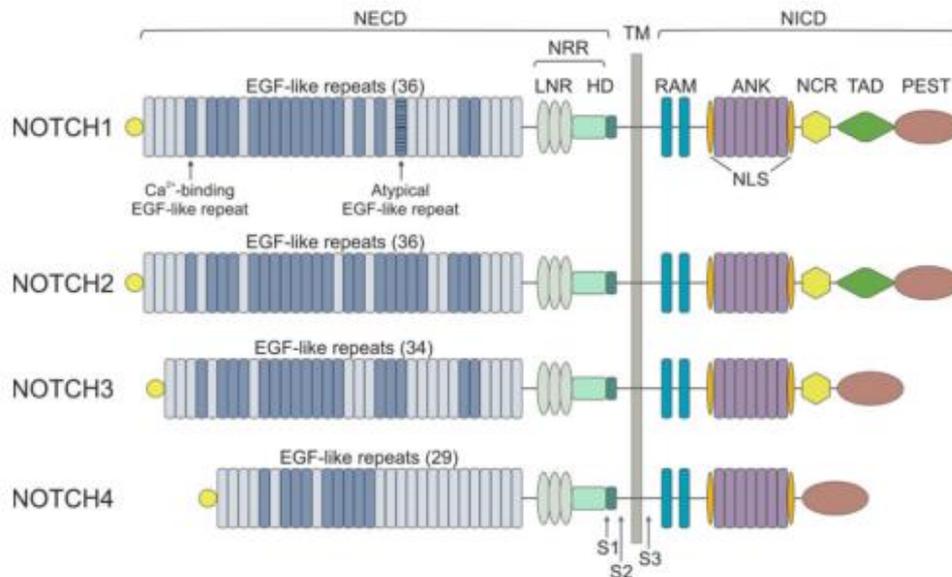


Figure 16: Structure of the four NOTCH receptor paralogs. According to (Arruga et al., 2018).

III.2.3. Notch and O-fucosylation

Notch is the most O-fucosylated protein among other POFUT1 target proteins (Table 5). Thus, POFUT1 normal expression is crucial for normal notch activity. Dysregulation of POFUT1 expression, whether silencing or overexpression, affects the notch signaling pathway. Okamura and Saga showed that the knockout of *POFUT1* is lethal for mice since it affects the notch signaling pathway among others (Okamura and Saga, 2008). On the other hand, Notch signaling is enhanced by the overexpression of POFUT1 and promotes tumorigenesis in cancers such as colorectal cancer (Deschuyter et al., 2020).

The extracellular domain of Notch1 possesses EGF-like domains, many of which (20 repeats) contain a consensus sequence of O-fucosylation and 17 of them have been shown to be effectively occupied by O-fucose (Stahl et al., 2008) including three conserved O-fucosylated EGF-like domains among species: namely EGF 12, 26, and 27 known to play a major role in NOTCH interactions (Rampal et al., 2005). Mutations in the O-fucosylation sites of any of the latter EGFs mediated dysregulation of Delta-1 and Jagged-1 mediated Notch signaling pathways, whereas mutation with less conserved sites (EGFs 9, 16, 20, 24, and 30) resulted in no remarkable effect. Mutation of mouse EGF12 induced the activity of Notch signaling while that on EGF26 ameliorated this pathway (Rampal et al., 2005).

In addition, an extension of O-fucose by other monosaccharide could be crucial for notch activity (Moloney et al., 2000). An example, the lengthening of O-fucose on EGF12 threonine 466 of NOTCH by Lunatic Fringe increases the affinity of Jagged 1 and Delta1 ligands for NOTCH receptors (Taylor et al., 2014).

III.3. The Wnt signaling factor 1 (or WIF1)

III.3.1. Wnt signaling pathway

The Wnt family belongs to secreted glycolipoproteins essential for the activation of canonical Wnt/ beta-catenin pathways as well as non-canonical Wnt pathways. Wnt/beta-catenin pathway is fundamental for directing cell proliferation, cell polarity during embryonic development, and tissue homeostasis (Logan and Nusse, 2004). Wnt ligands are known to be conserved among species. Mammals possess 19 Wnt ligands that bind to Wnt receptors Frizzled and LRP5/6 (MacDonald et al., 2009). This process activates the intracellular protein cascade leading to the accumulation of cytosolic beta-catenin. Beta-catenin acts as a co-activator of the transcription factor TCF, thus initiating gene transcription. Wnt signaling antagonists are required to modulate and regulate its activity throughout a lifetime. Among secreted Wnt antagonists: WIF1 (Wnt Inhibitory Factor)(Hsieh et al., 1999) and sFRP (secreted Frizzled Related Protein) (Bovolenta et al., 2008). Deregulation of any of the Wnt signaling pathway proteins, including WIF1, has a positive correlation with diseases and cancers (references). Epigenetic promoter methylation of *Wif1*, leading to silencing of its transcription and concomitant up-regulation of Wnt signaling, is a common feature during cancer progression. WIF1 is also considered one of the POFUT1 target proteins (Table 5)

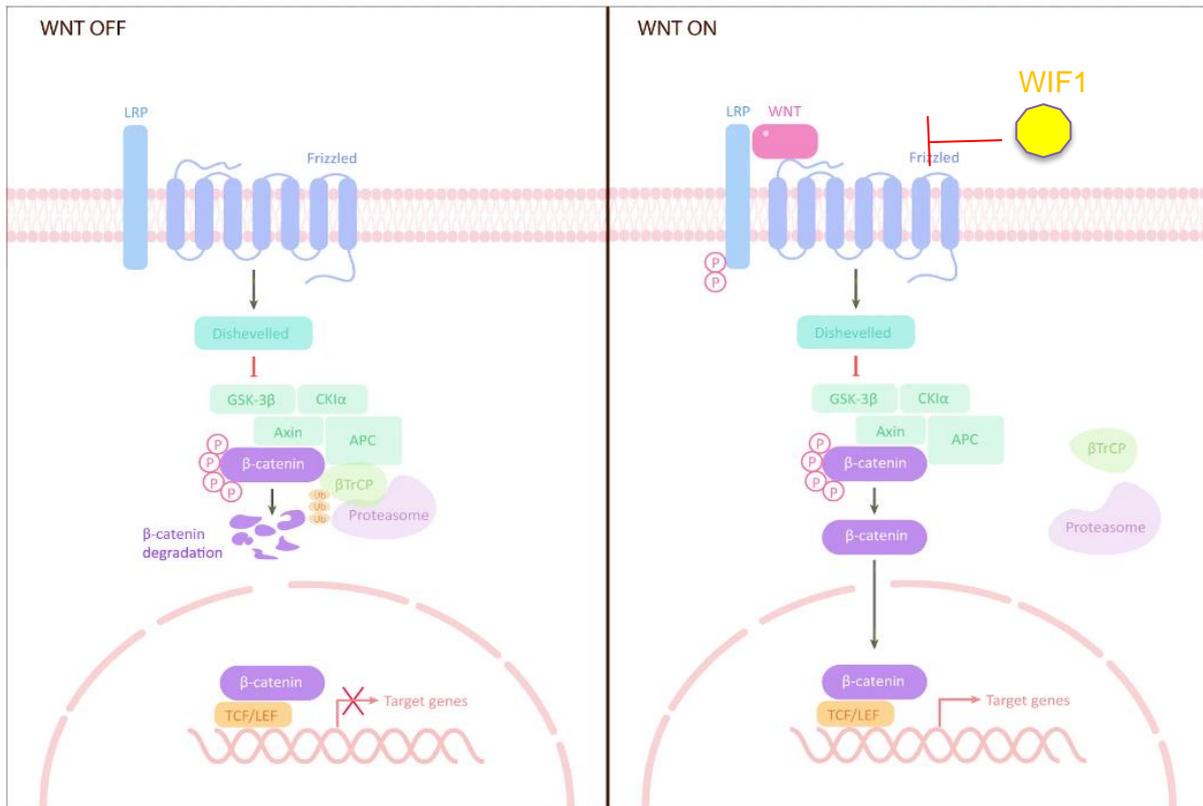


Figure 17: Canonical Wnt/ beta-catenin pathway

Inactivation of Wnt/beta-catenin signaling pathway (on the left): In the absence of Wnt signaling, β -catenin is degraded by protein complexes, including AXIN, APC, serine/threonine kinase GSK-3, and CK1 and E3 ubiquitin ligase β -TrCP. Activation of Wnt pathway (on the right): Wnt ligand binds to its receptor and activates Wnt pathway by inducing AXIN binding to LRP (phosphorylated lipoprotein receptor-related protein). Beta catenin is then stabilized, translocated to the nucleus and regulates transcription of target genes after binding with TCF/LEF. WIF1 is the antagonist of Wnt. GSK-3 glycogen synthase kinase-3, AXIN axis inhibition protein, CK1 casein kinase 1, APC adenomatous polyposis coli, TCF T cell factor, LEF lymphocyte enhancer factor-1. The scheme is adapted from (Liu et al., 2022).

III.3.2. Structure of WIF1

Human WIF1 (hWIF1) is constructed of 379 amino acids constituting the following domains: N- terminal signal peptide, a conserved WIF domain (WD), 5 EGF-like domains, and a hydrophilic C-terminal end (Figure 18A). Through its WD domain, WIF1 can bind to Wnts 3a, 4, 5a, 7a, 9a, and 11 (Surmann-Schmitt et al., 2009). It was demonstrated that the binding of the EGF-like domains of WIF1 with glypican Heparan Sulfate ProteoGlycans (HSPG) is crucial for WIF1-mediated Wnt activity inhibition (Avanesov et al., 2012). In *Drosophila*, Shifted (Shf) is the ortholog of hWIF1. However, Shf is known to regulate Hedgehog signaling in the wing imaginal disc (Figure 18).

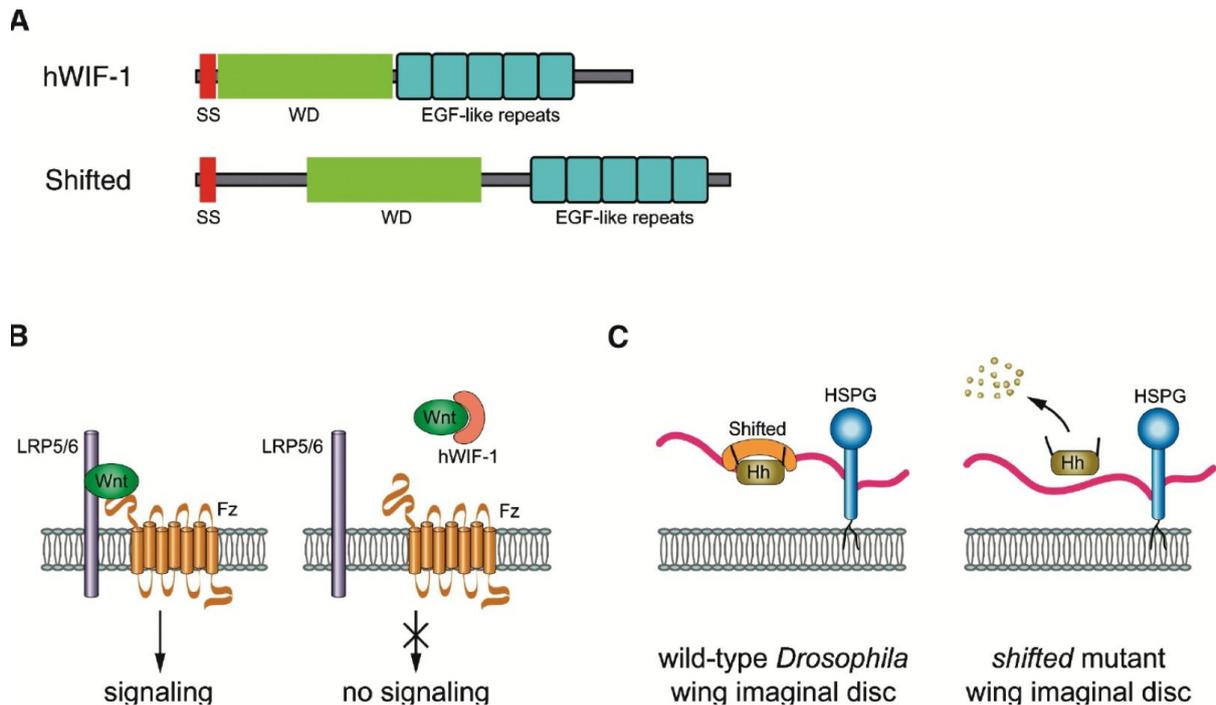


Figure 18: Structure of WIF1 and Shifted

(A) Common domain structure of human WIF-1 (hWIF-1) and Shifted. Both proteins contain a signal sequence (SS), WIF domain (WD), and five EGF-like repeats. WD of hWIF-1 is sufficient for its function, whereas both WD and EGF-like repeats are essential for the activity of Shifted. **(B)** hWIF-1 antagonizes Wnt and inhibits it from binding to its receptors (Frizzled (Fz) and LRP5/6). **(C)** Shifted enhances Hh-HSPG interaction. The figure is adapted from (Han and Lin, 2005).

III.3.3. POFUT1-mediated O-fucosylation of mouse WIF1 by and its role in WIF1 secretion

Among the five EGF-like domains of WIF1 of different species, two conserved O-fucosylation consensus sites are located on EGF III (C²XNGGIC³) and EGF V (C²GX(H/Y)G(S/T)C³). Only EGFIII of mouse WIF bears O-fucose, where the formation of steric clashes by amino acids (H³¹⁷ and Y⁷⁸) side chains prevents EGFV O-fucosylation (Pennarubia et al., 2020). No data showed the effect of WIF1 O-fucosylation on its activity and its consequences, whereas only Pennarubia demonstrated that recombinant mouse WIF1 secretion was dependent on the occupation of the O-fucosylation site of EGFIII (Pennarubia et al., 2020). However, the presence of O-fucose on C²-C³ loop of EGFV was not detected. This is suggested to be due the steric clash as a result of presence of arginine at position C⁵⁺¹ in the consensus sequence. Similar to that found in gnathostomes, the presence of H or Y at position C²⁺³ of the sequence C²GX(H/Y)G(S/T)C³ inhibited POFUT1 binding (Li et al., 2017). Assuming that hWIF1 has the same O-fucosylation site and effect as mouse WIF1, then the overexpression of POFUT1, as shown in cancers (Wan et al., 2017) (Chabanais et al., 2018), could lead to modification of WIF1 with O-fucose with high efficiency (100% molecules modified). Then, WIF1 could be highly secreted and active, thus efficiently inhibiting Wnt signaling pathways and degradation of cytosolic beta-catenin.

III.4. PAMR1

III.4.1. Structure and isoforms of PAMR1

PAMR1 or Peptidase domain containing Associated with Muscle Regeneration 1, is a multi-domain secreted protein formed of 5 domains: N-terminal Cubilin domain or CUB (Complement C1r/C1s, Uegf, Bmp1), a unique EGF-like domain (Epidermal growth factor-like domains or ELD), two SUSHI domains (SUSHI I and SUSHI II) and C-terminal Peptidase S1 domain (Figure 19).

There are eight human PAMR1 transcripts coding for different isoforms (ENSEMBL database), some of which are well described, and others are predicted. By referring to the UniProt database, three human PAMR1 isoforms were described and obtained by alternative splicing (Figure 19):

Isoform 1 (Iso 1) is considered the canonical form with a non-mature 720 amino acid (aa) sequence of molecular weight (MW) 80198.80 Da. After cleavage of the signal peptide (21 aa), Iso 1 does not have more than 699 amino acid residues and its theoretical MW is 77849.95 Da.

Isoform 2 (Iso 2) differs from Iso1 by the presence of 17 additional aa at position 274 (737 aa) following the EGF-like domain. Iso 2 is therefore composed of 716 aa after cleavage of its signal peptide and its theoretical MW is 79593.93 Da.

Isoform 3 differs from isoform1 by lacking a CUB domain (Iso1 Δ CUB), it is of 609 aa and 67541 Da of MW.

Other described or potential isoforms exist, Iso 1 without signal peptide (Iso1 Δ SP) or with only the EGF-like domain called ELD (ISO 1 ELD). For Iso 2, there would also be an isoform without signal peptide (Iso2 Δ SP). Finally, there would be an isoform with only the Sushi2 and peptidase S1 domains (Iso Δ CES1).

Mouse PAMR1 shares the same domains as the canonical isoform of human PAMR1. Mouse PAMR1 is 90.3% identical and 93.75% similar to human PAMR1 isoform 1. Pennarubia et al. have demonstrated recently the O-glycosylation of mouse PAMR1 by O-fucose, O-glucose and O-GlcNAc. (Figure 20).

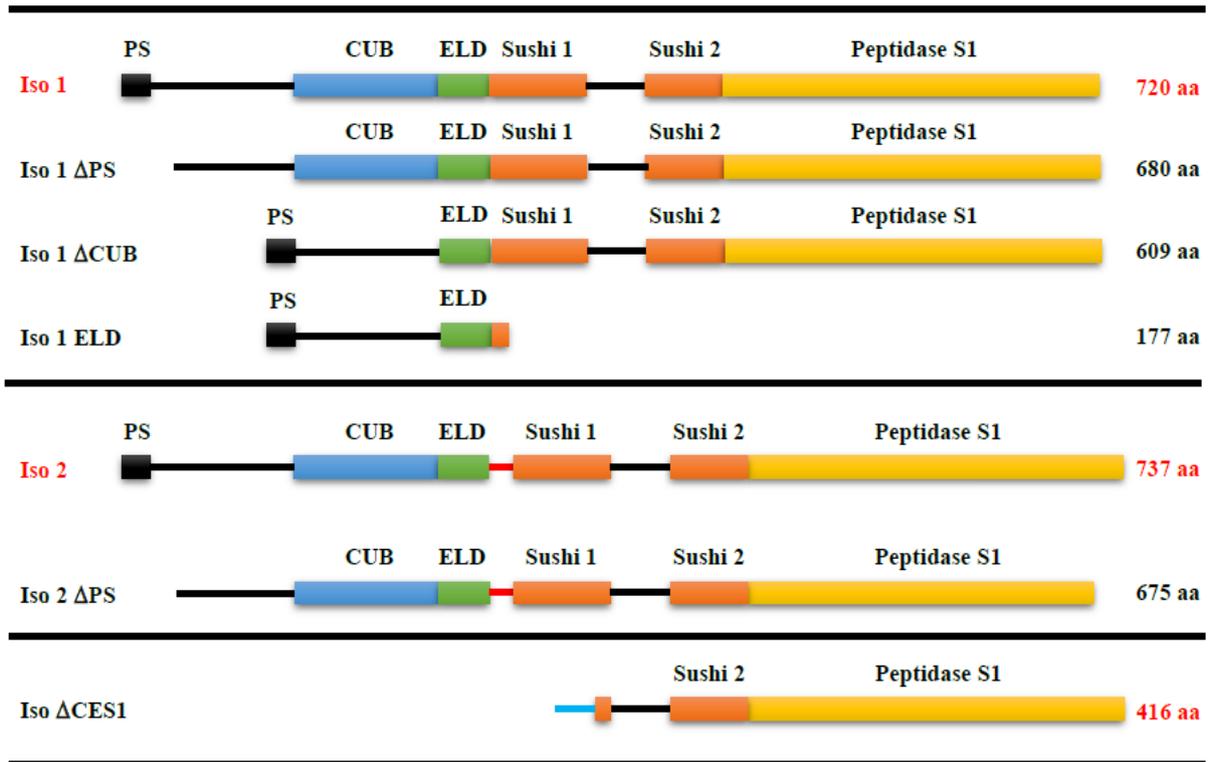


Figure 19: Scheme representing the different isoforms of human PAMR1

Based on UniProt (www.uniprot.org/) and Ensembl (www.ensembl.org/) databases.

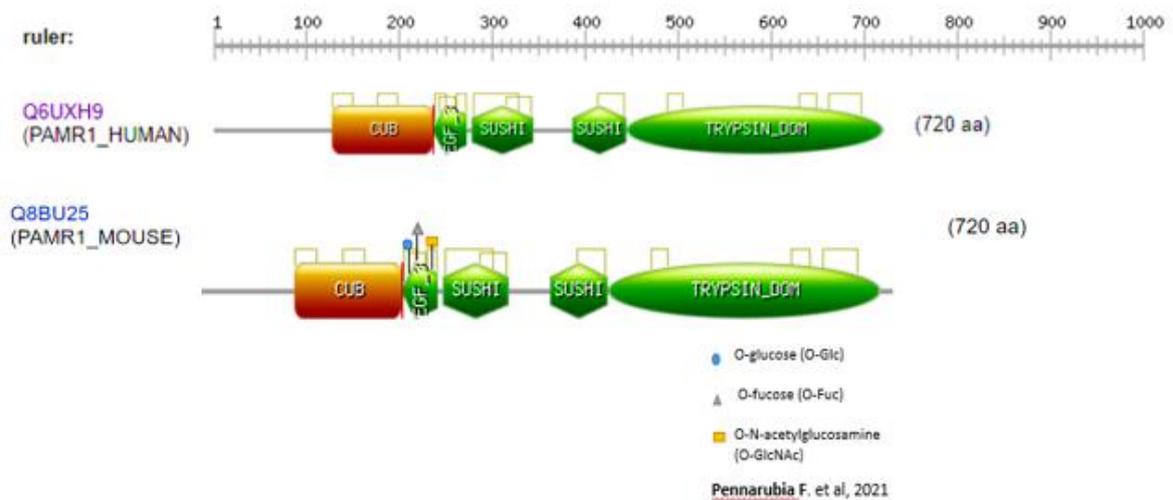


Figure 20: Representation of different domains of human and mouse PAMR1

Based on PROSITE database (<https://prosite.expasy.org/>).

III.4.2. O-glycosylation of PAMR1 EGF-like domain

Knowing that 16 mouse and 18 human proteins bore one EGF-like domain, or more, with the three glycosylation consensus sequences corresponding to O-Glc, O-Fuc, and O-GlcNAc modifications, PAMR1 was identified as one of them (Pennarubia et al., 2021). The single EGF-like domain of PAMR1 belongs to hEGF like domains. At first, Alfaro shows that the C⁵-C⁶ consensus site in mouse PAMR1 (from the brain) can be occupied by O-GlcNAc at position T²⁶⁷ (Alfaro et al., 2012). Pennarubia was able to identify the ability of isolated EGF-like domain of mouse PAMR1 to be modified *in vitro* by the three recombinant glycosyltransferases: POGLUT, POFUT1, and EOGT enzymes at their corresponding consensus sites. The anchorage of the three sugars, at the identified site, simultaneously explains the distinct positions of the consensus site (C¹-C², C²-C³, and C⁵-C⁶), with no steric clashes that prevent one or more of these sites to be accessible- to the enzyme. Previous results were confirmed using full-length recombinant mouse PAMR1 protein that possesses the triple O-glycosylation consensus sites with at least EGF-like domains modified by O-Glc1 and O-Fuc (and/or O-GlcNAc) (Pennarubia et al., 2021). These sugar modifications can be also elongated by other saccharides on full-length PAMR1 to form different glycoforms. O-GlcNAc can be elongated by two sugars (HexNAc + Hex), whereas O-Glc1 is predicted and not confirmed to be elongated with xylose in PAMR1. On the contrary, no PAMR1 O-fucosylglycans were detected by MS-MS. This could be explained by the low quantity of these glycoforms or initially, PAMR1 O-fucose cannot be elongated by GlcNAc as in some cases of NOTCH1 (Kakuda and Haltiwanger, 2017). Up to now, the role of O-glycans modification in PAMR1 is unknown. It could be suggested that this modification aid in PAMR1 folding, secretion, activity, stability or even protein-protein interactions.

III.4.3. Biological role of PAMR1

As its name suggests, the Peptidase domain containing Associated with Muscle Regeneration 1 (PAMR1), is particularly expressed during muscle regeneration (Nakayama et al., 2004a). PAMR1 transcript is highly expressed in the skeletal muscle and the surrounding area of muscle regeneration, confirming its role in the muscle fiber repair mechanism. PAMR1 was shown to be significantly suppressed in muscle cells derived from patients with Duchenne Muscular Dystrophy (DMD). The authors speculate that this under-expression could lead to DMD progression due to the low efficiency of muscle regeneration. However, the exact mechanism of action of PAMR1 in the muscle context remains under investigation.

Related to the cancer context, the PAMR1 transcript was shown to be downregulated among other phagocytosis-promoting genes in prostate cancer (Banerjee et al., 2019). The transformation of macrophages into tumor-associated macrophages (TAM) in prostate cancer is due to the secretion of regulatory molecules by cancer cells. This results in a decrease in transcripts of several genes involved in phagocytosis and of tumor suppressors genes in macrophages, including PAMR1, and an increase in proto-oncogenes transcripts. This suggests that PAMR1 could exert a tumor suppressor role in prostate cancer by enhancing the phagocytosis mechanism of macrophages against cancer cells.

PAMR1 is differentially expressed based on the type of cancer. Also, its exact role is to be defined. However, by comparing it to other proteins sharing the same/similar domains, we can have an idea or predict its role in cancer and its mechanism of action. The expression of PAMR1 in cancers as well as its predicted functions are described later (Chapter IV).

III.5. Expression of recombinant glycoproteins such as POFUT1 and its target proteins for biochemical analyses and structure-function studies

The expression, production and purification of recombinant glycoproteins allows their biochemical characterization (structure, stability, solubility, function, presence of post-translational modifications...) and also structure-function studies. There are various protein expression systems. The choice of one of them will depend on the properties of protein of interest and the considered use for the produced recombinant protein. Among the different expression systems available, procaryotes such as *E.coli* are not very relevant to produce recombinant human glycoproteins (POFUT1, PAMR1, WIF1...) contrary to mammalian expression systems or eventually some insect, yeast, and plants expression systems. Each has its respective advantages or drawbacks concerning cost of production, ease of use, and ability to perform some post-translational modifications. Relying on one of these systems will depend on its efficiency to produce stable, soluble and functional glycoproteins of interest at a low cost and a high yield. These recombinant glycoproteins are then often purified before their use for *in vitro* and *in vivo* experiments or in human healthcare (vaccines, drugs, or antibodies).

III.5.1. Expression of glycoproteins in mammalian expression systems

The most relevant expression system to produce recombinant human glycoproteins at low scale could be the mammalian one. Transient transfection of mammalian cells became the main advantage to relying on it as a means of protein expression system in terms of ease and speed, mainly for secreted and cell glycoproteins. The short time frame for the production of recombinant proteins using various host mammalian cells allowing intrinsic genetic stability are the key features for selecting this system. In addition, this system is easy for construction of the transfection vector (plasmid-based, adenoviral or retroviral expression vectors) as well as for the suitability of much simultaneous processing (Khan, 2013).

Among mammalian cells, Chinese hamster Ovary cells (CHO cells) are widely used for glycoproteins expression and production at the "lab" and industrial scales. The advantage of using these mammalian cells such as CHO cells is their ability to produce recombinant glycoproteins with *N*- and *O*-glycans close to those found in human glycoproteins and to perform most of posttranslational modifications as found in humans. In spite of all the advantages of using this system, it still has some limitations. Protein expression using mammalian expression system is considered highly expensive and often subjected to potential contamination (bacteria, fungi and mycoplasma), in addition to its low efficiency of large production of recombinant proteins compared to other systems. These limitations could be solved by another expression system used in the lab, namely the baculovirus-insect cell system.

III.5.2. Expression of glycoproteins in the baculovirus insect cell system

In the early 1980, the baculovirus-insect cell expression system was used as protein expression vector (Smith et al., 1983) and now became one of the mostly used eukaryotic expression system (Scholz and Suppmann, 2017). Heterologous proteins are produced by this system used in basic and applied research, thanks to the two strong promoters of polyhedrin *polH* and *p10* genes of *Autographa californica* multicapsid Nuclear Polyhedrosis Virus (AcMNPV). These latter drive high efficiency of protein production, besides the high proliferation capacity of lepidopteran cells (*Spodoptera frugiperda* (Sf21 and Sf9) and *Trichoplusia ni* (Hi5)).

These cells are capable of performing *N*-glycosylation of the recombinant glycoproteins at the same sites as those in mammalian cells (Harrison and Jarvis, 2006). The only difference is that insect cells produce simpler paucimannose-type *N*-glycans compared to complex-type *N*-glycans (with frequent terminal sialylation) found in mammals. However, different strategies can be applied to “mammalianize” or “humanize” this system. These structural differences in *N*-glycans could affect half-life of glycoproteins but in general are not responsible of major differences in folding, trafficking and *in vitro* bioactivities of these glycoproteins. The *O*-glycosylation pathway in many lepidopteran insect cell lines produces glycoproteins containing GalNAc α -O-Ser/Thr and a subpopulation of these structures is further processed to produce the Gal β 1,3GalNAc α -O-Ser/Thr core-1 structure (März et al., 1995). These structures are simpler than those found in mammals, exhibiting terminal sialic acids and sometimes bi-antennary mucin-type *O*-glycans. Finally, rat NOTCH1 EGF11-13 expressed in the baculovirus insect cell system using Hi-Five (Invitrogen) cells from *Trichoplusia ni* exhibited both *O*-Glucose (on EGF11,12 and 13) and *O*-Fucose (on EGF12 only) at expected sites and involved in interaction with Dll4, demonstrated that such insect cells can express functional POFUT1, POGLUT1 and POGLUT2/3 (Luca et al., 2015). In addition, this latter study showed elongation of *O*-fucose with GlcNAc revealing a Fringe activity (GlcNAc transferase) in Hi-Five cells.

To conclude, this system is highly efficient, able to perform main post-translational modifications (even with some differences with mammalian systems) and less expensive but it is time-consuming regarding obtaining recombinant baculoviruses to allow expression of the recombinant proteins in insect cells.

Chapter IV. PAMR1 in cancers

IV.1. Expression of PAMR1

Little is known about the expression of PAMR1 in normal and malignant/cancerous tissues, as well as its role in different tissues/cell lines. PAMR1 is a low-tissue-specific protein. It is known to be overexpressed in cervical and endocervical tissues at its mRNA level. It is recorded to be highly present in the serum and gallbladder at its protein level (Gene card database). The analysis of Lo et al., to 27 normal tissues, showed the highest expression of PAMR1 transcriptome is found in brain, bladder, aorta, and colon tissues, with lower expression in breast and skeletal tissues of both isoforms 1 and 2. On the other hand, no data is available concerning PAMR1 expression in embryonic and stem cells.

Concerning PAMR1 expression in cancers, *in silico* analysis of RNA Seq comparing the RNA expression of different genes between normal and tumoral tissue is reported in the Firebrowse database. PAMR1 is downregulated in tumoral tissues of several cancers including COAD (Colon Adenocarcinoma), READ (Rectal Adenocarcinoma), COADREAD (Colorectal Adenocarcinoma), CESC (Cervical and Endocervical carcinoma), confirming earlier research on cervical (Yang et al., 2021), breast (Lo et al., 2015), hepatocellular (Yin et al., 2016) and prostate (Makoukji et al., 2016) cancers as well as for cutaneous squamous cell carcinoma. PAMR1 could be, as well, overexpressed in some/exceptional tumoral tissues such as the case of KIRC (Kidney renal clear cell carcinoma), PCPG (Pheochromocytoma, and Paraganglioma), and SKCM (Skin Cutaneous Melanoma), and also meningioma as previously reported. On the other hand, data is missing for PAMR1's expression in the following cancers: ACC (Adrenocortical Carcinoma), DLBC (Lymphoid Neoplasm Diffuse Large B-cell Lymphoma), LAML (Acute Myeloid Leukemia), LGG (Brain Lower Grade Glioma), MESO (Mesothelioma), OV (Ovarian Serous Cystadenocarcinoma), TGCT (Testicular Germ cell Tumors), UCS (Uterine Carcinosarcoma), and UVM (Uveal Melanoma). All these differences of PAMR1 expression according to the considered cancer type could be correlated to potentially different roles of this glycoprotein in normal different cell types before their malign transformation

IV.2. PAMR1 in breast and cervical cancer

Recent studies showed the downregulation of PAMR1 in breast and cervical cancers. However, based on database analysis, it was reported in 2016 that PAMR1 is suppressed in the Lebanese population and overexpressed in the western population in case of breast cancer (Makoukji et al., 2016). This could be due to several unidentified factors. Many tumor suppressor genes such as those in breast cancer (APC, BRCA1, p16, p21, TIMP3, but also PAMR1) have multiple CpG islands in their promoter regions. Due to promoter hypermethylation in breast cancer, PAMR1 is down-expressed through epigenetic silencing. PAMR1 expression can be restored through the use of demethylating agents, such as 5-aza-2, deoxycytidine (Lo et al., 2015). In breast cancer cell lines, the overexpression of PAMR1 isoforms 1 and 2 resulted in a notable inhibition of cell proliferation. As a result, as in the instance of cervical cancer (Yang et al., 2021), PAMR1 is regarded as a tumor suppressor gene. Yang et al. showed that PAMR1 expression suppression through RNA interference increased cervical cancer cell proliferation, migration, and invasion. This confirms PAMR1 as a putative tumor suppressor in breast and cervical cancers.

IV.3. Mechanism of action of PAMR1 in cancer

PAMR1 is a multi-domain protein that belongs to the family of S1 peptidases, which are predominately serine proteases and possesses a C-terminal "trypsin-like" domain. These proteolytic enzymes interact using a catalytic triad made up of a residue nucleophile (serine), a base similar to histidine, and an acid-like aspartate to cleave peptide bonds. Human PAMR1 exhibits a threonine in place of the conserved serine of the catalytic triad at position 665. Thus C-terminal Peptidase S1 domain of PAMR1 is probably inactive since the threonine residue is known to be substantially less nucleophilic than a serine residue. However, no evidence in the literature supports this hypothesis.

PAMR1 bears a CUB domain preceded by an EGF-like domain (EGF-LD). The CUB domain is among the sequence of many extracellular proteins, such as PAMR1 or proteins associated with the plasma membrane. It was demonstrated that the SCUBE2 protein, a protein with anti-tumor activity toward breast cancer cells, specifically bears a CUB domain linked to EGF-LD domains, similar to those found in PAMR1 (Cheng et al., 2009). Similarly, to SCUBE2, PAMR1 protein might interact with membrane proteins such as E-cadherin and proteins involved in intercellular adhesion *via* its CUB/ELD domains. The putative interactions of PAMR1 with protein partners have not yet been subjected to any investigation.

IV.4. Signaling pathways linked to PAMR1

Cancer's molecular alterations are intricate and induce alteration in numerous signaling pathways. Recently, it was revealed by Yang et al. that PAMR1 could be involved in the suppression of MYC and mTOR signaling pathways (Yang et al., 2021). MYC is a proto-oncogene activated among others by the MAPK pathway, which plays a role in favoring proliferation, migration, apoptotic resistance, and angiogenesis. MYC has many properties: it can regulate the transcription of other genes and stabilize mRNA and proteins (Kress et al., 2015). When it is positively deregulated, it will activate the transcription of target genes which will have a favorable effect on tumorigenesis and tumor progression (Lourenco et al., 2021). mTOR, a serine/threonine kinase protein, is activated by the PI3K/AKT pathway involved in the same biological processes as the MAPK pathway (Yang et al., 2021). mTOR is known also to activate the metastatic cascade in cancers. SIN1 and MLST8 are two subunits of mTORC1 and mTORC2 that promote cell migration and invasion. The regulation of ULK1 by PAMR1, a mTORC1 negative regulator, as well as SIN1 and MLST8, could suppress cell migration and invasion in cervical cancer.

By analogy with the SCUBE2 protein (Lin et al., 2011), PAMR1 could interact with a surface membrane protein such as E-cadherin and affect signaling pathways as mentioned above. However, the protein partners of PAMR1 are not known and remain to be discovered depending on the tumor context.

Chapter V. Objectives and Experimental Approaches

Early screening of colorectal cancer, which is becoming prominent cancer worldwide, allows more targeted treatment, enhances the overall survival rate, and ameliorates its devastating progressive effects. Despite the various blood-based biomarkers, there is no defined, up to now, ideally specific and sensitive biomarker for colorectal cancer. Finding new biomarkers necessitates understanding the molecular mechanisms leading to the occurrence and development of CRC. Several proteomic analyses revealed dysregulation of many proteins in CRC context, whether by having upregulated or downregulated expression levels. Among of them, many proteins belong to glycoproteome. In addition to variations in the expression levels of these glycoproteins in cancer, the nature of their glycans can also vary in a cancerous cell.

In our laboratory, we were interested in studying some dysregulated glycosyltransferases in CRC. Among these enzymes of glycosylation, the protein O-fucosyltransferase 1 (POFUT1). Chabanais et al. demonstrated in 2018 the overexpression of POFUT1 in colorectal cancer, as early as stage I, as the case in other cancers such as gliomas (Kroes et al., 2007) and oral carcinomas (Yokota et al., 2013). This suggested the ability of POFUT1 to be considered as a new early biomarker of CRC (Chabanais et al., 2018). POFUT1 overexpression can lead to hyper-O-fucosylation of its potentially targeted proteins such as NOTCH receptors thus potentially affecting their biological effects, function, activity, or even stability. Among one hundred of POFUT1 target glycoproteins, PAMR1 which has a single EGF-like domain (Pennarubia et al., 2021) was already shown to be involved in breast cancer in 2015 (Lo et al., 2015).

PAMR1, a secreted glycoprotein, is known to be dysregulated in cancer and was recently considered as a tumor suppressor. Indeed, in most of the cases, PAMR1 was shown to be downregulated as a consequence of epigenetic alteration by promoter hypermethylation as recently confirmed in cases of breast cancer (Lo et al., 2015) and cervical cancer (Yang et al., 2021). In addition, PAMR1 suppression was accompanied by enhancement of cell proliferation, migration, and invasion in the previously mentioned cancers. In the context of CRC, PAMR1's expression and role/effect on biological activities were unknown yet. Taking into account that PAMR1 possesses a single EGF-like domain harboring different O-linked monosaccharides (O-Glc, O-Fuc, and O-GlcNAc) (Pennarubia et al., 2021), potentially involved in protein-protein interactions, the O-glycosylation status of PAMR1 especially its O-fucosylation should be determined.

The **main goal of my thesis** is to determine whether PAMR1 exerts a tumor suppressive activity in CRC, as in the case of other cancers. First, an *in silico* analysis was carried out to determine the expression level of PAMR1 in colorectal cancer patients compared to normal cases using RNASeq data found in the Firebowse database. Then, the quantity of PAMR1 was determined in CRC tissue samples as well as CRC cell lines at the RNA and protein levels. The CRC cell lines used are HCT116, HT29, and SW620 reflecting the three different stages of CRC in comparison to normal CCD841CoN colon cell line. Finally, two main experimental approaches were carried out to study the role of PAMR1 in these CRC cell lines, focusing, specifically, on cell proliferation and migration.

- The first approach was to increase the quantity of PAMR1 in growth medium of CRC cells by adding "home-made" recombinant PAMR1. Murine PAMR1 was used for this exogenous

treatment instead of its human isoform 1 counterpart, which was much less efficiently produced by stable CHO cell lines. Since both mature proteins share a high percentage of identity (90,56%) and similarity (94.13%), we assumed that they could exert the same action.

- The second approach was to overexpress human PAMR1 isoform 1 in CRC cell lines, whether by stable or transient transfection. HeLa cervical cancer cells, known to be sensitive to a dysregulation of PAMR1 quantity (Yang et al., 2021), were also transfected and studied in parallel.

To perform exogenous treatments of CRC cell lines (first approach mentioned above), we tried before to produce recombinant PAMR1 in the most suitable expression system. For this purpose, we used CHO mammalian cells and Sf9 insect cells to produce PAMR1 focusing on production, protein stability and O-fucosylation of the protein of interest. Due to very low-level expression of recombinant human PAMR1 in stable CHO cell lines, we tried to produce this recombinant glycoprotein in another eukaryotic expression system to achieve a best production allowing exogenous treatments. Among the different expression systems, we focused on the baculovirus insect cell expression system, known for its high efficiency of production of recombinant glycoproteins compared to mammalian expression systems.

A **second part of my work** concerns the biochemical characterization of *Spodoptera frugiperda* POFUT1 (SfPOFUT1) and determination of its ability to add O-fucose to any EGF-like domain, including the single one of human PAMR1. As previously described (Pennarubia et al., 2018), *in vitro* POFUT1 O-fucosylation assays were performed with GDP-azido-fucose and a blot analysis with streptavidin was performed after copper-catalyzed azide-alkyne cycloaddition (CuAAC) of biotin alkyne (or click chemistry). In this work, the efficiencies of mouse and *Spodoptera frugiperda* POFUT1 to add *in vitro* O-fucose to isolated EGF-like domains of different proteins (NOTCH1 EGF26, WIF1 EGF3 and PAMR1 EGF) were compared.

VI.1. PAMR1 and Colorectal Cancer

PAMR1 negatively impacts cell proliferation and migration of Human Colon Cancer HT29 Cell Line

Layla Haymour¹, Alain Chaunavel², Mona Diab Assaf³, Abderrahman Maftah^{1, #} and Sébastien Legardinier^{1, #, §}

¹ University of Limoges, PEIRENE UR22722, Glycosylation and Cell Differentiation, F-87060 Limoges, France.

² Department of Pathology, Limoges University Hospital, 87042 Limoges, France.

³ Molecular Cancer and Pharmaceutical Biology Laboratory, Faculty of Sciences II, Lebanese University Fanar, 1500 Beirut, Lebanon.

A. Maftah and S. Legardinier are considered co-last authors and contributed equally to this work

§ To whom correspondence should be addressed

BioRxiv (The preprint server for Biology). Posted September 8, 2022.

DOI: 10.1101/2022.09.07.506931

Our laboratory is interested in the dysregulation of protein O-fucosyltransferase 1 (POFUT1) and its target proteins in the context of colorectal cancer. An overexpression of POFUT1 was seen in CRC as early as stage 1, mainly due to chromosomal amplification (Chabanais et al., 2018) and was positively associated with colorectal tumor progression through activation of Notch signaling pathway (Chabanais et al., 2018) and (Du et al., 2018). POFUT1 might be a potential novel biomarker for CRC diagnosis.

Among 100 target human proteins of POFUT1, comprising many membrane proteins such as NOTCH receptors 1-4 and ligands of Delta-like and Jagged families, secreted glycoproteins were also found such as PAMR1, our protein of interest. In addition of the presence of a single EGF-like domain bearing O-fucose (Pennarubia et al., 2021), PAMR1 is a multi-domain protein which possess a cubilin domain (CUB). The association of these two types of proteindomains, namely CUB and EGF-like domains, was also found in SCUBE2 and associated with the tumor suppressor activity of this secreted protein in breast cancer (Cheng et al., 2009). At the beginning of my thesis, the light was shed on PAMR1 in cancer where it was demonstrated to be down-expressed in breast cancer (Lo et al., 2015) and more recently in cervical cancer (Yang et al., 2021). In this latter study, PAMR1 was considered as a putative tumor suppressor: it was shown to exert anti-proliferation, anti-migration, and anti-invasion roles in cervical cancer cells (HeLa and Me180 cell lines). Nevertheless, up to now, there is no clear idea about the mechanism of action of PAMR1 in cancer.

The paper below deals with investigation of the expression of PAMR1 in CRC as well as its role in proliferation and migration of CRC cell lines. The expression of PAMR1 was first assessed *in silico* using public data available in databases, *ex vivo* using CRC tissue samples from patients and *in vitro* with CRC cell lines at both RNA and protein levels. Upon realizing the suppression of PAMR1 in CRC, we looked forward to reversing its expression based on two complementary experimental approaches. The first approach was to increase the quantity of PAMR1 through exogenous treatment of CRC cell lines by adding “homemade” recombinant PAMR1 (purified or concentrated from supernatants of stable CHO cells) to growth medium of CRC cell lines. The second approach was to overexpress untagged human PAMR1 in CRC cell lines after transient or stable transfections of constructs harboring cDNA encoding the canonical isoform 1 of human PAMR1. The effects of both exogenous treatments and PAMR1 overexpression were determined for different relevant cell properties, such as viability, cell proliferation and migration. In parallel to CRC cell lines, HeLa cervical carcinoma cells, which were known to respond to a dysregulation of PAMR1 expression (Yang et al., 2021) were treated in the same experimental conditions to validate the relevance of our treatments.

Using both approaches, the effects of an increase of PAMR1 quantity on cell proliferation and migration were mainly assessed for the CRC cell line HT29. The increase of PAMR1 quantity was correlated with a reduction of HT29 cell proliferation and migration activities, suggesting the anti-proliferative and potential tumor suppressor effect of PAMR1 in CRC. As in breast and cervical cancers, PAMR1 might be considered as tumor suppressor protein in CRC. Tremendous efforts should be done to elucidate its mechanism of action.

PAMR1 negatively impacts cell proliferation and migration of Human Colon Cancer HT29 Cell Line

Layla Haymour¹, Alain Chaunavel², Mona Diab Assaf³, Abderrahman Maftah^{1, #} and Sébastien Legardinier^{1, #, §}

¹ University of Limoges, PEIRENE UR22722, Glycosylation and Cell Differentiation, F-87060 Limoges, France.

² Department of Pathology, Limoges University Hospital, 87042 Limoges, France.

³ Molecular Cancer and Pharmaceutical Biology Laboratory, Faculty of Sciences II, Lebanese University Fanar, 1500 Beirut, Lebanon.

A. Maftah and S. Legardinier are considered co-last authors and contributed equally to this work

§ To whom correspondence should be addressed: Tel: +33555457792; Fax: +33555457653; e-mail: sebastien.legardinier@unilim.fr

Running head: PAMR1 reduces HT29 proliferation and migration

Keywords: Colorectal cancer, PAMR1, Proliferation, Migration, Biomarker

ABSTRACT

Colorectal cancer (CRC) is becoming one of the most prevalent cancers worldwide. Among cancers, it ranks the third place in terms of incidence and the second in terms of mortality. Even though immunological test allows fast and easy diagnostic method, there is no specific and reliable methods for early detection of CRC. Despite different treatments, high risk of re-occurrence is associated with advanced and metastatic CRC stages. An exhaustive knowledge on specific biomarkers or molecular actors involved in CRC could help to eradicate tumors or limit cancer recurrence. In this study, we focused on PAMR1 (Peptidase Domain Containing Associated with Muscle Regeneration 1), which is already considered as a tumor suppressor in breast and cervical cancers. *In silico* analysis of RNASeq data showed that PAMR1 was significantly downregulated in CRC tissues compared to their adjacent normal ones, as well as in cervical cancer. Our analysis showed that this downregulation, probably due to promoter hypermethylation, such as in breast cancer tissues, appeared in the four cancer stages as early as the first stage. In consistency with *in silico* analyses, the expression of PAMR1 was found to be lower at the transcript and protein levels in CRC tissue samples compared to normal ones, as well as in different CRC cell lines (HCT116, HT29, and SW620) compared to normal colon cell line (CCD841CoN). To understand the role of PAMR1 in CRC cancer, recombinant purified PAMR1 or concentrated secretome from CHO overexpressing PAMR1 were used to exogenously treat CRC cell lines with a focus on HT-29 cells as well as Hela cervical cancer cell line known to be sensitive to PAMR1. Transient or stable transfections were also performed to determine the impact of PAMR1 overexpression in HT29 and/or HeLa cells. In this study, we finally showed that presence of PAMR1 could reduce both cell proliferation and cell migration with a positive correlation between these biological effects and PAMR1's quantity. This implies that PAMR1 expresses anti-proliferative and anti-migrative effects in CRC. Further studies to be done in order to confirm the tumor suppressive role of PAMR1 in CRC.

INTRODUCTION

Cancer is becoming the most leading cause of death worldwide. Cancer diagnosis and patients' treatment were impacted negatively with the Coronavirus disease 2019 (Covid-19) pandemic in Europe (Neamțiu et al., 2022) besides other factors that enhance its prevalence worldwide. According to the latest Global Cancer Observatory Statistics in 2020 (Globocan 2020), colorectal cancer (CRC) was classified the third cancer in terms of its incidence (1,931,590 cases), after lung and breast cancers, as well as the second leading cause of cancer death (935,173 cases), after lung cancer. In Europe, 4,398,443 new CRC cases (out of 19292789 cancer cases) were estimated in 2020. Although the risk of developing CRC is more pronounced after the age of 50 years (Byrne, 2017), environmental factors (Diergaard et al., 2007) and genetic hereditary factors, such as Familial Adenomatous Polyps (Jasperson et al., 2010) and APC gene mutation (Valle, 2014), can be also incriminated in its occurrence and development. Screening of CRC can be assayed by stool-based, imaging and endoscopic tests (Hadjipetrou et al., 2017); as well as detecting tumor biomarkers, Carcinoembryonic Antigen (CEA) or Carbohydrate Antigen 19-9 (CA19-9), that are more or less specific for CRC. Despite various treatment methods of this malignancy, especially surgery in early stages, high mortality rate is associated with more advanced/metastatic stages. Have more knowledge on different actors, such as oncogenes and tumor suppressor genes, involved in CRC is still a challenge. Especially the finding of an early specific biomarker of CRC is a crucial issue for its early diagnosis, more targeted treatment, and high survival rate.

Peptidase Domain Containing Associated with Muscle Regeneration 1 (PAMR1) is a multi-domain secreted glycoprotein, formed of five main domains: CUB domain (Complement C1r/C1s, Uegf, Bmp1), one EGF-like domain (Epidermal Growth factor – like domain), two SUSHI domains (SUSHI 1 and SUSHI 2), and a trypsin-like peptidase S1 domain. PAMR1 was first shown to be downregulated in Duchenne Muscular Dystrophy (DMD) (Nakayama et al., 2004), with no clear idea about its regenerative mechanism of action. It was also reported to be suppressed in some cancers including breast cancer (Lo et al., 2015), cervical cancer (Yang et al., 2021) and gynecologic cancer (Yu et al., 2021). Studies focusing on PAMR1 in breast cancer turned out to show that PAMR1 is downregulated by means of epigenetic silencing due to its promoter's hypermethylation (Lo et al., 2015). Recovering PAMR1's expression by an epigenetic drug was shown to inhibit DNA methylation such as 5-aza-2'-deoxycytidine, leading to diminishing the invasion and migration of breast cancer cells (Lo et al., 2015). In this last study, PAMR1 was

considered for the first time as a tumor suppressor. This role in cancer was recently confirmed by Yang et al. showing that PAMR1's knockdown promoted proliferation, migration, and invasion of cervical cancer cells such as HeLa and Me180 cells (Yang et al., 2021). Despite these findings on PAMR1's biological roles, the mechanism of action, including proteins partners, of this secreted glycoprotein is still unknown in skeletal muscle cells as well as in cancer cells. However, the presence of CUB and EGF-like domains in PAMR1 suggests its involvement in protein-protein interactions with other secreted proteins or cell-surface membrane proteins. Indeed, EGF-like domains are small protein domains (30-40 residues) stabilized by three disulfide bonds (Wouters et al., 2005) and known to regulate protein interactions such as those between Notch receptors and their Jagged and Delta-like ligands (Rand et al., 1997). In addition, The secreted protein SCUBE2 (secreted Signal Peptidase CUB-EGF domain containing protein 2), which exerts a tumor suppressor activity in breast cancer (Cheng et al., 2009), has similarly to PAMR1 a CUB domain and a multi-repeat region composed of 9 EGF-like domains, both involved in its anti-tumor effect (Cheng et al., 2009)(Lin et al., 2013). In addition to this CUB/EGF-like domains combination, the presence of *O*-fucose, known to modulate NOTCH-ligands interactions (Okajima et al., 2003)(Luther and Haltiwanger, 2009), on mouse PAMR1 was recently demonstrated in our lab (Pennarubia et al., 2020). However, the contribution of *O*-fucose in the function of PAMR1 has not yet been determined.

In spite of the absence of a clear view of PAMR1's mechanism of action in cancer, its effects on different signaling pathways related to cell proliferation and cell survival were investigated. Indeed, Yang et. al illustrated ability of PAMR1 to suppress MYC and mTORC1 signaling pathways in cervical cancer (Yang et al., 2021). However, the role of PAMR1 and its mechanism of action may be different depending on the type of cancer where PAMR1 exerts a tumor suppressor role.

Starting with preliminary *in silico* analysis showing a reduced amount of *PAMR1* transcripts in colorectal cancer tissues from patients than in normal controls, we were interested to investigate whether PAMR1 exhibited a tumor suppressor activity in CRC such as in breast and cervical cancers. In addition of public data obtained by RNASeq methods, we confirmed PAMR1 down expression (qPCR, western blots) in tissue samples from CRC patients and in three different CRC cell lines (HCT116, HT29, SW620) *versus* a normal colon cell line (CCD841CoN). To study the potential role of PAMR1 in colorectal cancer, two main strategies were carried out, namely

exogenous treatments of cancer cell lines with recombinant PAMR1 stably produced in mammalian CHO cells and transient (or stable) overexpression of the canonical isoform 1 of human PAMR1 in HT-29 cells as well as in cervical cancer HeLa cells. Since the production yield of human PAMR1 in stable CHO cells was too low for exogenous treatments, its murine counterpart exhibiting 90.3% identity with the mature human isoform 1 was used. To assess the relevance of the use of recombinant mouse PAMR1, cervical cancer HeLa cells were treated in parallel. The effects of PAMR1 treatment on cell viability, proliferation and migration were analyzed in both HT-29 and HeLa cancer cell lines.

MATERIALS AND METHODS

Database Analysis. RNA Seq data were extracted from the FireBrowse database (www.firebrowse.org), which examines various types of cancer by comparing tumor samples to normal ones. In this study, we focused on Colon Adenocarcinoma (COAD), Rectal Adenocarcinoma (READ), Colorectal Adenocarcinoma (COADREAD), and Cervical and Endocervical Cancers (CESC). PAMR1 expression levels were fused from COAD/READ/COADREAD or CESC.uncv2.mRNAs_normalized_log2.txt found in COAD or READ or COADREAD or CESC.mRNAs_Preprocess.level file. RNA Seq data were analyzed with PAST4 software.

Clinical specimen. A panel of cancerous colorectal tissue samples with their corresponding cancer-adjacent tissues were retrieved from the archive of the CRB Limousin - CHU of Limoges. Ethics approval (CRB-CESSION-2021-008) was obtained from the “Comité médico-scientifique de la tumorotheque de l’Hôpital Dupuytren”, the bioethics committee of CHU of Limoges. Tissue samples used were classified either by colorectal cancer stages (I-IV) or by T classification of TNM staging. The clinicopathological information of the patients was also available.

Cell Lines. Three human colorectal cancer cell lines were used in this study, namely HCT116 (ATCC CCL-247), HT29 (ATCC HTB38) and SW620 (ATCC CCL-227). Only one human cervical cancer cell line was also used, HeLa cells (ATCC CCL-2). HCT116, HT29, HeLa or derived-stable cell lines were cultured in DMEM growth medium (Gibco, ThermoFisher Scientific). However, the SW620 cell line was grown in RPMI growth medium (Gibco, ThermoFisher Scientific). The normal colorectal cell line CCD814CoN (ATCC CRL-1790), grown

in EMEM growth medium (ATCC), was used with less than 15 passages. Flp-InTM CHO cells (Thermo Fisher Scientific, Waltham, MA, USA) stably overexpressing mouse PAMR1, previously obtained (Pennarubia et al., 2020), were cultured in F12 growth medium. All culture media were supplied with 10 % Fetal Bovine Serum (FBS) (S1810 biowest, South America) and 0.5 % penicillin/streptomycin antibiotics (100 U/mL penicillin and 100 µg/mL streptomycin) (Gibco, USA). The cells were maintained at 37°C in a humidified atmosphere with 5 % CO₂.

Plasmid constructs. In order to produce the secreted forms of recombinant human PAMR1 (isoforms 1 & 2) in Flp-InTM CHO- cells, we used the modified pSec-NtermHis6 vector containing secretory signal peptide (IgK Leader) as found originally in commercial vector pSecTag/FRT/V5-His-TOPO^R vector (Thermo Fisher Scientific, Waltham, MA, USA), but fused to six histidine residues (His6) and followed by Kpn I and BamH I cloning sites, as previously described (Pennarubia et al., 2018). This modified pSec-NtermHis6 vector was digested by Kpn I and BamH I and using the same strategy of prehybridized overlapping oligonucleotides as in the previous study (Pennarubia et al., 2018), a new cassette was inserted containing the sequence of V5 epitope, downstream of His6 tag, and new cloning sites Hind III and Xho I to generate new vector referred to as pSecPSHisV5. The cDNA sequences of human PAMR1 isoform 1 (NP_001001991.1) and isoform 2 (NM_001001991.3) without the signal peptides, were cloned between Hind III and Xho I restriction sites downstream of the sequence encoding N-terminal His6 and V5 tags. Resulting constructs named pSecPSHisV5-hPAM1 and pSecPSHisV5-hPAM2 harbored the sequence of human PAMR1 isoform 1 and human PAMR1 isoform 2, respectively. After nucleotide sequence verification, each plasmid construct was subjected to a cotransfection with pOG44 vector (Thermo Fisher Scientific, Waltham, MA, USA) expressing the Flp recombinase to produce stably transfected Flp-InTM CHO cells (Thermo Fisher Scientific, Waltham, MA, USA). In order to overexpress human PAMR1 isoform 1 in HT29 cells, the commercial pcDNA3.1(+) (Thermo Fisher Scientific, Waltham, MA, USA) was used. The cDNA sequence of human PAMR1 isoform 1 (NP_001001991.1) was amplified by PCR from HEK total cDNAs and inserted downstream of the cytomegalovirus promoter of the vector using Hind III and Xho I cloning sites. The obtained recombinant vector was named pcDNA3.1-hPAMR1. The nucleotide sequence was verified before cells transfection.

Cell culture and transfection. Recombinant human HisV5-PAMR1 isoforms 1 and 2 were produced by stable transfection of Flp-In CHO cells. Flp-In CHO cells were co-transfected with 1 μ g of either pSecPSHisV5-hPAM1 or pSecPSHisV5-hPAM2 construct and 4 μ L of the transfectant X-tremeGENETM DNA Transfection Reagent (Sigma-Aldrich, Saint Louis, MO, USA) according to the manufacturer's protocol. The selection started 24h post-transfection by Hygromycin B (ThermoFisher Scientific, Waltham, MA, USA) of final concentration 500 μ g/mL in F-12 medium. The recombinant PAMR1 produced by hygromycin-resistant cells was assessed by Western blot. Stable HT-29 cells overexpressing untagged human PAMR1 isoform 1 as well as the Mock cells were co-transfected with pcDNA3.1-hPAMR1 construct and empty pcDNA3.1 vector pcDNA3.1(+) (Thermo Fisher Scientific, Waltham, MA, USA) respectively, with transfectant X-tremeGENETM DNA Transfection Reagent (Sigma-Aldrich, Saint Louis, MO, USA) according to the manufacturer's protocol. The selection of transfected cells started 24h post transfection by changing DMEM growth medium containing 750 μ g/mL of Geneticin (G-418). Different clones of Geneticin-resistant pool of cells overexpressing PAMR1 were selected and amplified. The level of PAMR1 expression in the Pool, Clones and Mock cells was assessed by qPCR.

Real-Time quantitative PCR (qPCR). Total RNA from tissue samples (after being grinded in Liquid Nitrogen) and cell lines was extracted using RNeasy mini kit (QIAGEN) and reverse transcribed using High-Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific) according to the manufacturers' protocol. qPCR was performed using TaqMan gene expression Master Mix (ThermoFisher Scientific, Lithuania).

Protein production and purification. Recombinant PAMR1 protein was produced from stably transfected CHO cells. Post cell seeding by 24h, the cells with >80 % confluency were washed with PBS 1X and cultured in fresh warm F-12 medium supplemented with 10 % FBS for 96h (optimal protein production with least degradation). The supernatant was then collected and proteins were precipitated in ammonium sulfate to reach 50 % saturation at RT and then centrifuged at 10,000g RT for 15 min. The precipitated proteins were purified based on nickel affinity purification by AKTA Prime Plus automated purification system (GE Healthcare). The sample passed over a 1 mL Nickel-Sepharose (HisTrap HP) affinity column at a flow rate of 1 mL/min according to a pre-recorded purification program. Using Buffer B (25 mM Tris-HCl,

500 mM NaCl, 500 mM Imidazole, pH 7.5), the sample was eluted by different imidazole concentrations. Different eluted fractions were established. The fractions that correspond to the peaks were tested by Coomassie blue staining and Western blot. The purest fractions containing PAMR1 were selected and concentrated in 10K Amicon (Sigma Aldrich, Ireland) by repeated centrifugation steps of 4500 G each at 4°C during 45 min each.

Secretome concentration. Stable CHO-mPAMR1 cells that were already produced in the lab (Pennarubia et al., 2018), were seeded in 20 cm² Petri dishes to become confluent after 24h. The culture media were aspirated and cells were washed twice with PBS 1X before adding 1 % FBS F-12 medium. After 48 h of incubation at 37°C, the supernatant was collected and centrifuged at 2500 g for 5 minutes to discard any floating dead cells. The supernatant was then concentrated 20 fold using Amicon 3K using several rounds of centrifugation, each was done at 4500 g, 4°C, for 45 minutes.

Protein extraction and Western blot. Tissue samples were grinded in liquid nitrogen. Lysis buffer named RIPA (50 mM Tris-HCl, 150 mM NaCl, 1 % Triton X-100 (v/v), 0.5 % sodium deoxycholate (w/v), 0.1 % sodium dodecyl sulfate (v/v), pH 8) containing a cocktail of protease and phosphatase inhibitors (Roche Applied Science, Mannheim, Germany) was used to extract total proteins from tissue samples and CRC cell lines by its incubation with cell pellets for 1 h at 4°C. The lysates were centrifuged at 14,000 g for 15 min at 4°C. The concentration of proteins, in the supernatant, was quantified using BCA protein Assay Kit (Thermofisher scientific, USA). Proteins were resolved by SDS-PAGE in 8 % polyacrylamide gel at 24 mA. Then, proteins were blotted on 0.45 µm nitrocellulose membrane for 2 h at 50 mA. The membranes were blocked by TBS-Tween20 0.1 % (50 mM Tris, 150 mM NaCl, pH 7.6, 0.1% Tween-20 (v/v)) supplemented in 5 % Bovine Serum Albumin (BSA) (Sigma Aldrich, USA) or 5 % half-fat milk for 1h at room temperature. The membranes were then incubated with sheep anti-PAMR1 antibody (AF6517, R&D systems) diluted at 1:1000 in TBST 0.1 % supplemented with 2.5 % BSA or with Goat anti-GAPDH antibody (AF5718, R&D systems) diluted at in TBST-0.1 % supplemented with 2.5 % milk or anti-V5 HRP (Thermofisher scientific) overnight at 4°C. After washing the membrane with TBST-0.1 % three times, the corresponding secondary antibodies were added at 1:1000 for 1 h at room temperature. The membranes were revealed after adding the chemiluminescent substrate using an Amersham Imager 600 device (GE Healthcare, Uppsala, Sweden).

Cell Viability Assay. Colorectal cancer cell lines were seeded to a cell density of 100,000 cells per well in 96-well plates. The cells were incubated for 24h in humidified condition at 37°C and 5% CO₂. Different concentrations of purified recombinant PAMR1, diluted in growth medium, were added to cells, which were then incubated for 24 h, 48 h and 72 h. At each time point, 10 µL of cell counting kit (CCK8) (WST-8 CCK8, ab228554, abcam) was added in each well. The results were revealed by spectrometer at a wavelength of 460 nm 1 h post incubation with CCK8.

Cell proliferation Assay. Cell lines were seeded at seeding density 500,000 cells per well in six-well plates. After 24 h, the medium was changed by adding different concentrations of purified PAMR1 or 20 folds concentrated stable CHO-mouse PAMR1 secretome. The cells were then incubated for 24 h, 48 h and 72 h. At each time point, cells were detached and added to the cells of the supernatant. Total cells were counted after being stained by Trypan blue using Malassez chambers.

Cell Migration Assay. Cell migration assay was performed using two well-silicon inserts (Culture, Insert 2 well, Ibidi, Germany). Cells were put in the wells after being trypsinized and resuspended with growth medium supplemented with 10 % FBS. After the cells reached confluency, the inserts were removed, the cells were cultured in growth medium supplemented with 1% FBS. The closure of the cell-free gap was visualized daily and measured using Image J software.

Statistical Analyses. All the experiments were performed independently at least three times. t-Student test found in GraphPad Prism 7 (GraphPad Software Inc, San Diego, CA, USA) was used to perform the statistical comparison. Results were considered as statistically significant if the p-value was less than 0.05.

RESULTS AND DISCUSSION

Low PAMR1 expression in colorectal cancer (CRC) and in many other cancers

Based on RNA Seq public data available in FireBrowse database (<http://firebrowse.org/>), *in silico* analysis comparing mRNA expression encoding many proteins in tumoral *versus* normal tissues can be done. For PAMR1, a down-expression was seen in colorectal cancer (COADREAD) and in many other cancers (Figure 1A), confirming previous findings for cervical (Yang et al., 2021),

breast (Lo et al., 2015) and hepatocellular (Yin et al., 2016) cancers and also for cutaneous squamous cell carcinoma (Wei et al., 2018).

In this study, we especially focused on rectal adenocarcinoma (READ) (normal tissues = 10 samples vs tumoral tissues = 166 samples) (Figure 1B) and colon adenocarcinoma (COAD) (normal tissues = 41 samples vs tumoral tissues = 457 samples) (Figure 1C) or on all compiled data available for colorectal adenocarcinoma (COADREAD) (normal tissues = 51 samples vs tumoral tissues = 623 samples) (Figure 1D). We were also interested in cervical cancer (CESC) (normal tissues = 3 samples vs tumoral tissues = 304 samples) (Figure 1E) exhibiting low quantity of PAMR1 and for which a tumor suppressor role was recently suggested (Yang et al., 2021). PAMR1 mRNA quantity was dramatically and significantly reduced in all these CRC tumoral tissues compared to their adjacent normal ones. Interestingly, this downregulation arose as early as stage I (or T stage according to TNM staging) and this low amount of PAMR1 mRNA was found in all the other stages analyzed. The stage-independent decrease of PAMR1 expression in cancer raises a question if PAMR1 could be considered as an early biomarker of colorectal cancer, such as in cervical cancer (Yang et al., 2021).

Strikingly, PAMR1 could be overexpressed in a few tumoral tissues such as the case of KIRC (Kidney renal clear cell carcinoma) and PCPG (Pheochromocytoma and Paraganglioma). However, due to missing data on normal tissues, we unfortunately could not evaluate the expression level of PAMR1 in the following cancers: ACC (Adrenocortical Carcinoma), DLBC (Lymphoid Neoplasm Diffuse Large B-cell Lymphoma), LAML (Acute Myeloid Leukemia), LGG (Brain Lower Grade Glioma), MESO (Mesothelioma), OV (Ovarian Serous Cystadenocarcinoma), TGCT (Testicular Germ cell Tumors), UCS (Uterine Carcinosarcoma), and UVM (Uveal Melanoma).

These very important differences in the expression of PAMR1 according to the type of cancer suggest that the role of this protein could differ according the cell type.

Reduced expression of PAMR1 in CRC tissue samples from patients *versus* normal ones

In collaboration with CRB Limousin – CHU of Limoges, the expression of PAMR1 was analyzed at the transcript and protein levels in a panel of specimen collected from CRC patients, classified by pathological stages (Figure 2). According to qPCR results, PAMR1 expression was dramatically reduced in tumoral tissues from CRC patients compared to their adjacent non-cancerous ones (Figure 2A), confirming RNA seq data obtained from the FireBrowse database as

shown in Figure 1. Only for the stage T1 of tumoral samples, a non-significant downward trend at transcript level was seen. Total proteins were also extracted from normal and CRC tissue samples followed by Western Blot using the same anti-PAMR1 antibody as in a previous study (Lo et al., 2015). The global analysis of three different sets of tissue samples showed a downward trend of PAMR1 quantity in tumoral samples for the four stages (Stages I-IV), as early as stage I (Figure 2B). However, PAMR1 was very difficult to quantify at the protein level, due to its low expression (even in healthy tissues), its instability and its propensity for degradation.

To conclude, all of these results showed at least a downward trend of PAMR1 expression in colorectal cancer at all stages. PAMR1 might be considered as an early biomarker of colorectal cancer as in cervical and breast cancers. This finding could help to limit the number of CRC patients diagnosed with aggressive stages (III and IV), correlated with poor prognosis and short overall survival time (Kuo et al., 2003)(Mukai et al., 2018).

PAMR1 expression was significantly reduced in different colorectal cancer cell lines

The *in vitro* study was mainly based on the use of three colorectal cancer cell lines available in the lab, namely HCT116, HT-29 and SW620 cells, which represent colorectal cancer pathological stages I, II, and III, respectively. According to old Duke's staging system, they are classified as Duke's A, B, and C stages (Dukes, 1932) (Akkoca et al., 2014), respectively. Normal colon cell line CCD841CoN was used as normal control cells. In addition of these CRC cell lines, HeLa cells, representing the cervical cancer, were chosen in our study due to recent findings showing their sensitivity to a dysregulation of PAMR1 expression (Yang et al., 2021).

After extraction of total RNA from all the cell lines mentioned above, PAMR1 transcripts were specifically quantified by RT-qPCR using Taqman technology. As shown in Figure 3, the quantity of PAMR1 transcripts was found to be extremely lower (CT values above 35) in the three colorectal cell lines than in normal colon cell line CCD841CoN. If considering these high CT values, we can assume that PAMR1 expression was totally abolished in the three CRC cell lines as it was the case for HeLa cells studied here, consistent with the previous study on cervical cancer (Yang et al., 2021). The failure of detection of PAMR1 protein signal by Western blot from the crude secretome or intracellular proteins of these cell lines (data not shown) confirms qPCR results. Surprisingly, we were also not able to detect PAMR1 in crude secretome of CCD841CoN, probably due to low sensitivity of the antibody used or a too low secretion of PAMR1 in spite of its good expression at the transcript level.

In compatibility with data from FireBrowse database, we thus ascertained the downregulated expression of PAMR1 in three colorectal cell lines of different stages and in HeLa cells, representing the cervical cancer.

Production of recombinant PAMR1 for exogenous treatments of cancer cell lines

In a first approach, we wanted to produce recombinant human PAMR1 in order to carry out exogenous treatments of the three available CRC lines. These exogenous treatments required to produce human PAMR1 in an expression system and to purify it in the view of its addition to culture medium of non-modified cancer cell lines. As in our previous study (Pennarubia et al., 2020), stable CHO cell lines were generated to produce secreted forms of both isoforms 1 and 2 for human PAMR1, with N-terminal Histidine and V5 tags. Unfortunately, the expression level of both isoforms was much lower than for mouse HisV5-PAMR1, even undetectable for the canonical isoform 1 of human PAMR1 (data not shown). Thus, we chose the baculovirus insect cell system to express human PAMR1 isoforms 1 and 2. However, human PAMR1 expressed in the baculovirus-insect cell expression system was not very stable, prone to form aggregates and subjected to degradation both during its production and its purification. For all these reasons, only small quantities of the recombinant human PAMR1 were produced (data not shown) but did not allow to perform exogenous treatments of cancer cell lines with doses up to 5 µg/mL. Taking into consideration that mouse recombinant PAMR1 exhibits 90.3 % identity with the isoform 1 of human PAMR1 and whose level rate of production in stable CHO cells was much better, as previously shown (Pennarubia et al., 2020), we chose to rely on mouse PAMR1 to perform exogenous treatments. As shown in Figure 4A, significant differences of production were not seen in complete or serum-free medium for recombinant mouse PAMR1, detected by Anti-V5-HRP antibody around 95 kDa. However, a specific signal also appeared at about 50 kDa at 96h, probably due to partial protein degradation. Recombinant mouse PAMR1 was thus produced in complete medium and harvested after 96 h maximum followed by its purification on nickel affinity column (Ni-NTA). The analysis of the eluted fraction was done by Coomassie blue staining and Western blot (Figure 4B). By comparison to Western blot, the most enriched fractions in purified monomer of PAMR1 were fractions 9 and 10. Bands with high MW (up to 130 kDa) were seen and could correspond to protein aggregates that were formed either by interaction of PAMR1 with other protein partners or due to PAMR1-PAMR1 dimer/complex formation. These aggregates were not detected by the anti-V5 antibody, contrary to monomeric recombinant mouse PAMR1 which

appeared at the expected size. The most enriched elution fractions in protein of interest (Figure 4B) were then pooled and concentrated before protein quantification using the BCA method.

HT-29 was the most sensitive cell line to exogenous treatment with recombinant PAMR1

Different concentrations of purified recombinant mouse PAMR1 (0, 1, 2.5, and 5 $\mu\text{g/mL}$) were added to culture medium of the three colorectal cancer cell lines, namely HCT116, HT29, and SW620 cells. To limit quantities of purified PAMR1 used, we chose to determine cell viability by using Cell Counting Assay 8 (CCK8) at different time points (0 h, 24 h, 48 h, and 72 h).

Exogenous treatment with purified mouse PAMR1 had no real effect on HCT116 and SW620 cells viability despite different doses and at different time points. However, the percentage of cell viability of HT29 decreased 48h and 72h post-treatment with the highest dose of PAMR1, namely 5 $\mu\text{g/mL}$. This suggests that PAMR1 either increased cell mortality and/or decreased cell proliferation. Compared to previous studies, it was tempting to think that PAMR1 exerted an anti-proliferative role with respect to HT29 (Figure 5). Cell viability assays were repeated several times, but with different preparations of purified mouse PAMR1. Each time, the results obtained ensure the sensitivity of HT29 to PAMR1. However, this effect was seen at different concentrations of purified recombinant PAMR1 depending on its purity in different preparations. This could be explained by the fact that, after concentration of elution fractions, purified mouse PAMR1 did not exhibit the same purity between preparations. Nevertheless, taking into account all of these results, we therefore chose to stably overexpress PAMR1 isoform 1 only in the HT-29 line, which exhibited a response to PAMR1 treatment.

Stable overexpression of human PAMR1 did not affect HT29 cell proliferation and migration

The priority of transfection went to HT29 that was found to be the most sensitive CRC cell line to exogenous treatment with recombinant PAMR1, compared to HCT116 and SW620. HT29 cells were first stably transfected with pCDNA3.1[hPAMR1 isoform1] or empty pCDNA3.1 vector. Hygromycin-resistant cells were selected and amplified to be analyzed by both RT-qPCR and Western blot (Figure 6). As expected, mock cells, represented hygromycin-resistant cells obtained after integration of empty vector, did not express PAMR1 at the transcript level (Figure 6A) and protein level using the anti-PAMR1 antibody (Figure 6B). However, the quantity of mRNA increased significantly for the pool and clone 2 (Figure 6A). The pool corresponded to a combination of several clones overexpressing different levels of PAMR1 whereas “Clone 2” from

the pool was selected for its highest overexpression of human PAMR1 among all selected hygromycin-resistant clones. This reflects the overall low expression of PAMR1 in the Pool, with different clones exhibiting a relatively low expression level. The anti-PAMR1 antibody used for Western blot allowed the specific detection of 20 X concentrated overexpressed human PAMR1 isoform 1 at the expected size for the pool and clone 2 (Figure 6B). However, a strong specific band was also detected by anti-PAMR1 antibody at a higher apparent molecular weight around 120 kDa for clone 2. This suggests that beyond a certain concentration in the culture medium, PAMR1 was unstable and probably prone to form protein aggregates as seen for mouse PAMR1 purified after its ammonium sulfate precipitation (Figure 4B). This could also result from protein concentration of the secretome by ultrafiltration.

The biological effects of PAMR1 overexpression were assessed for all stably transfected cell lines and for non-modified HT29. No significant change was observed for cell proliferation (Figure 7A) and cell viability (Figure 7B) in stable cell lines overexpressing PAMR1 (Pool, clone 2) compared to WT HT29 or to Mock. For cell migration, the gap closure was almost seen at 144h for WT HT29 but not for the Cl2 and Pool stable cell lines, where the gap remained unclosed (Figure 7C). Thus, the quantification of gap closure showed a significant difference between stable cell lines and HT29 but not between cells overexpressing PAMR1 and Mock (Figure 7D). All these results could be explained by the insufficient quantity of PAMR1 expressed by stable cell lines compared to previous studies (Yang et al., 2021).

The increase of PAMR1 quantity significantly reduced cell proliferation of HT29 cells

Since HT29 cell line was shown to be sensitive to recombinant mouse PAMR1 when added in culture medium but not to *in cellulo* human PAMR1 stable overexpression, we decided to combine transfection and treatment to increase the quantity of PAMR1 in the culture medium. Same experiments were done in parallel with HeLa cells, which are cervical cancer cell lines sensitive to PAMR1 dysregulation (Yang et al., 2021).

In a first approach, HT29 and HeLa cells were transiently transfected with the expression plasmid pCDNA3.1 bearing cDNA for human PAMR1 isoform 1 referred to as pCDNA-[hPAM iso1]. A second approach consisted on the overexpression of recombinant mouse PAMR1 in a stable CHO cell line already available in the lab (Pennarubia et al., 2020). The concentrated secretome of stable CHO cells was used instead of purified mouse PAMR1 to avoid protein degradation occurring during purification steps.

HeLa and HT29 were transiently transfected by pCDNA3.1-hPAMR1 iso1 construct, followed by detection of PAMR1 transcript by qPCR. The results showed a pronounced increase in PAMR1 expression at RNA level compared to non-transiently transfected cells with pCDNA3.1 (Figure 8). HT29 and HeLa cells were transiently transfected by pCDNA3.1[hPAM iso1] and/or either treated with concentrated recombinant PAMR1 (or with growth medium only). Then, cell proliferation was assessed after 72h of incubation by cell counting assay (Figure 9). The presence of PAMR1 whether being overexpressed by transient transfection or exogenous treatment provoked a significant decrease of cell number for HT29 (Figure 9A) but “slightly” reduced the cell proliferation of HeLa cells (Figure 9B). Consequently, the exogenous treatment of transfected cells by recombinant PAMR1 significantly reduced cells proliferation of both HeLa and HT29 cells. So, the combination of the two approaches highly increased the quantity of PAMR1, reflecting more efficient effect, thus more pronounced reduction of cell proliferation.

Cell migration of transiently transfected HT29 and HeLa cells with their control (non-transfected cells) was assessed by wound-healing assay (Figures 9C and 9D). The gap closure started after 48h for non-transfected HeLa and HT29 cells and became more pronounced after 144h hours, whereas less cell migration was observed for the transfected ones. The results obtained took more than 48h to be visualized and were not significant reflecting that PAMR1 expression level is low and not sufficient to exert a pronounced effect within a short time duration.

To conclude, a significant increase of PAMR1 amount in the secretome of colorectal cancer cell lines, as well as in cervical cancer ones, significantly diminished cell proliferation but only a downward trend was observed for cell migration of both cell lines. These biological effects seem to be dependent of the level expression or quantity of PAMR1. This confirms the induction of HeLa and Me180 cells proliferation, migration and invasion as a result of PAMR1 knockdown. On the other hand, a huge PAMR1 overexpression in these cell lines reduced these effects.(Yang et al., 2021).

Silencing PAMR1 expression in CRC might be due to promoter hypermethylation

Molecular events leading to PAMR1 downregulation in colorectal cancer as early as in stage I is unidentified yet; however, recovering its expression could be through the use of drug treatments as the case in other cancers (Lo et al., 2015) and with other suppressed genes (Zhu et al., 2018). Recovering the tumor suppressor effect of PAMR1 in breast cancer was through treatment with demethylation agent, 5-aza²'deoxycytine2 (Lo et al., 2015). Since PAMR1 was inactivated due to

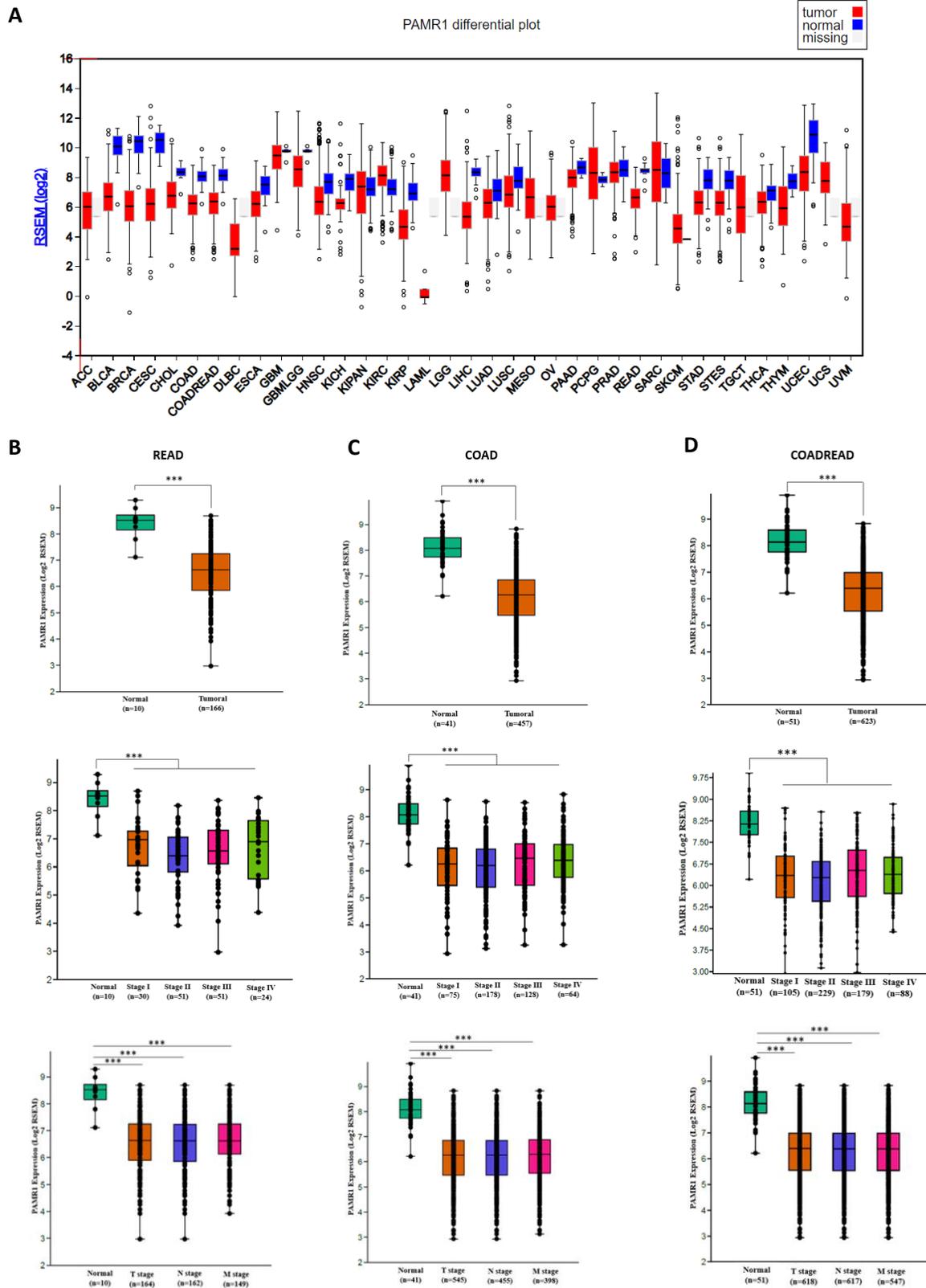
epigenetic silencing through promoter hypermethylation in breast cancer cells, 5-aza2'deoxyctine² led to PAMR1 re-expression, thus reduction of cancer cell growth. To assess whether PAMR1 downexpression in CRC is due to its epigenetic silencing, treatments with demethylation agents could be performed. However, this treatment is not specific of PAMR1 and other tumor suppressor genes might be re-expressed following the use of this non-specific demethylating agent.

The precise mechanism action of PAMR1 remains to be elucidated

The mechanism of action of PAMR1 is still unknown but PAMR1, which is a secreted multi-domain protein, could interact with one or several proteins expressed on the cell surface or present in the extracellular space. We can hypothesize that its protein partners, potentially different according to the cell type, could modulate its action. In the physiological state, PAMR1 could participate to the maintenance of a normal proliferation rate of different cell types. However, the suppression of its expression, by epigenetic inactivation or by other molecular events, in cancer cells undoubtedly participates to their increased proliferation. PAMR1 was thus recently considered as a tumor suppressor gene (Yang et al., 2021).

In conclusion, we confirmed the down expression of PAMR1 in colorectal cancer. The overexpression of PAMR1 is crucial for reduction of cell proliferation and migration of colorectal cancer cells. By that, PAMR1 could be predicted as an early biomarker and tumor suppressor of colorectal cancer. However, its mechanism of action is to be investigated.

FIGURES



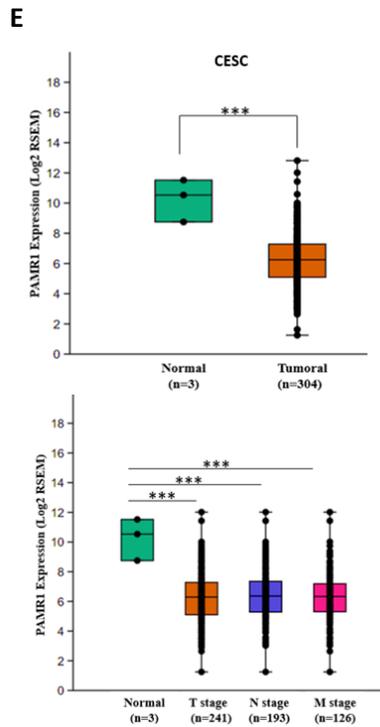
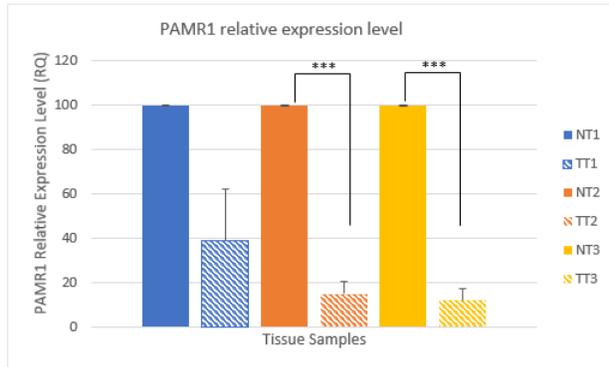


Figure 1: PAMR1 is suppressed in most cancers including colorectal and cervical cancers in stage-independent manner. (A) Expression of PAMR1 in normal and tumoral cancer samples from TCGA Firebrowse database (<http://firebrowse.org/>). PAMR1's expression is downregulated in Cervical and Endocervical Cancers (CESC), Colon Adenocarcinoma (COAD), Colorectal Adenocarcinoma (COADREAD) and Rectal Adenocarcinoma (READ) tumoral samples compared to normal ones. (B, C, D and E) RNA Seq data Analysis showed a significant PAMR1 mRNA down expression in tumoral samples compared to normal samples in rectal, colon, colorectal adenocarcinomas and cervical cancer respectively. This downregulation is pronounced in all pathological stages (Stage I, Stage II, Stage III, Stage IV) and TNM stages, as early as stage I and T stage. The box plots represent the mean of Log₂ RSEM ± SEM. The RNA Seq database analysis was carried out by PAST software. *p < 0.05, **p < 0.01, ***p < 0.001.

A



B

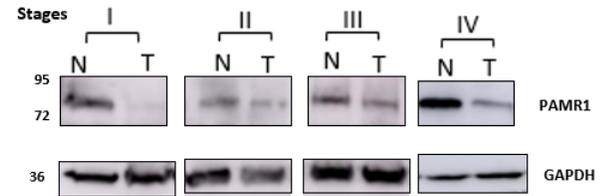


Figure 2: PAMR1 expression in CRC tissue specimen. (A) Bar graph showing PAMR1 relative expression level (ratio \pm SEM for PAMR1/HSPA8) at transcriptomic level of three different sets of tissue samples classified by T stages (T1, T2, T3) of TNM classification. (B) Western blot analysis for PAMR1 expression in tumoral tissues from patients compared to their adjacent normal ones in the four CRC pathological stages. N: Normal. T: Tumoral. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

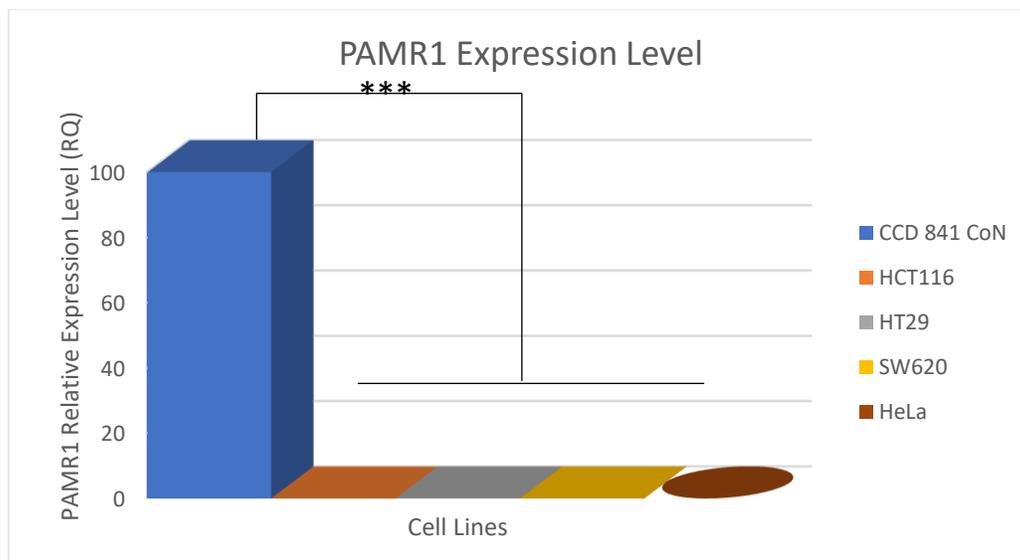


Figure 3: PAMR1 expression level in colorectal and cervical cell lines. PAMR1 transcripts are significantly downregulated in colorectal cancer cell lines (HCT116, HT29, and SW620), compared to normal colon CCD841CoN cells. Similarly, to CRC cell lines, cervical cancer HeLa cells also exhibited very low PAMR1 expression with high CT values around 36 (normal cervix cell lines not available). The histogram represents mean \pm SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

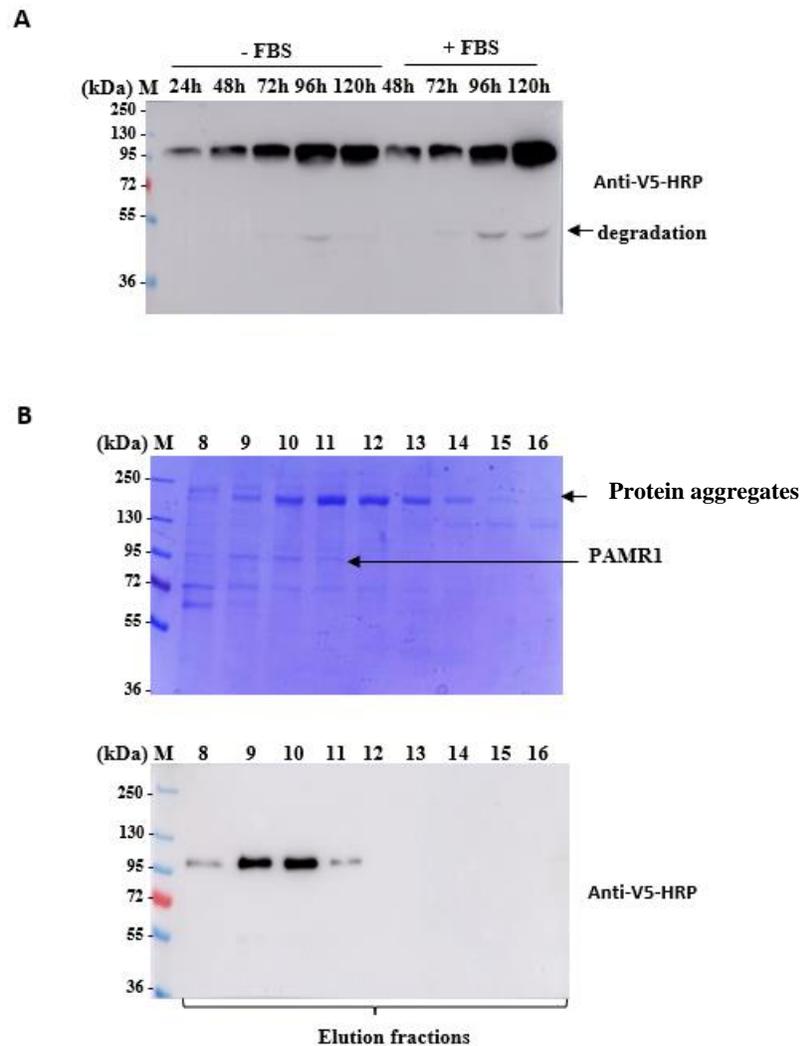


Figure 4: Purification of recombinant mouse PAMR1 produced by stable Flp-InTM CHO cells. (A) Time course of production of recombinant mouse HisV5-PAMR1 in stable CHO cells in the presence and absence of FBS in Growth DMEM medium. (B) SDS-PAGE with Coomassie blue staining of polyacrylamide gel and Western blot analyses using anti-V5-HRP antibody of elution fractions following nickel-affinity purification using imidazole gradient of mouse HisV5-PAMR1, produced in secretome of stable CHO cells.

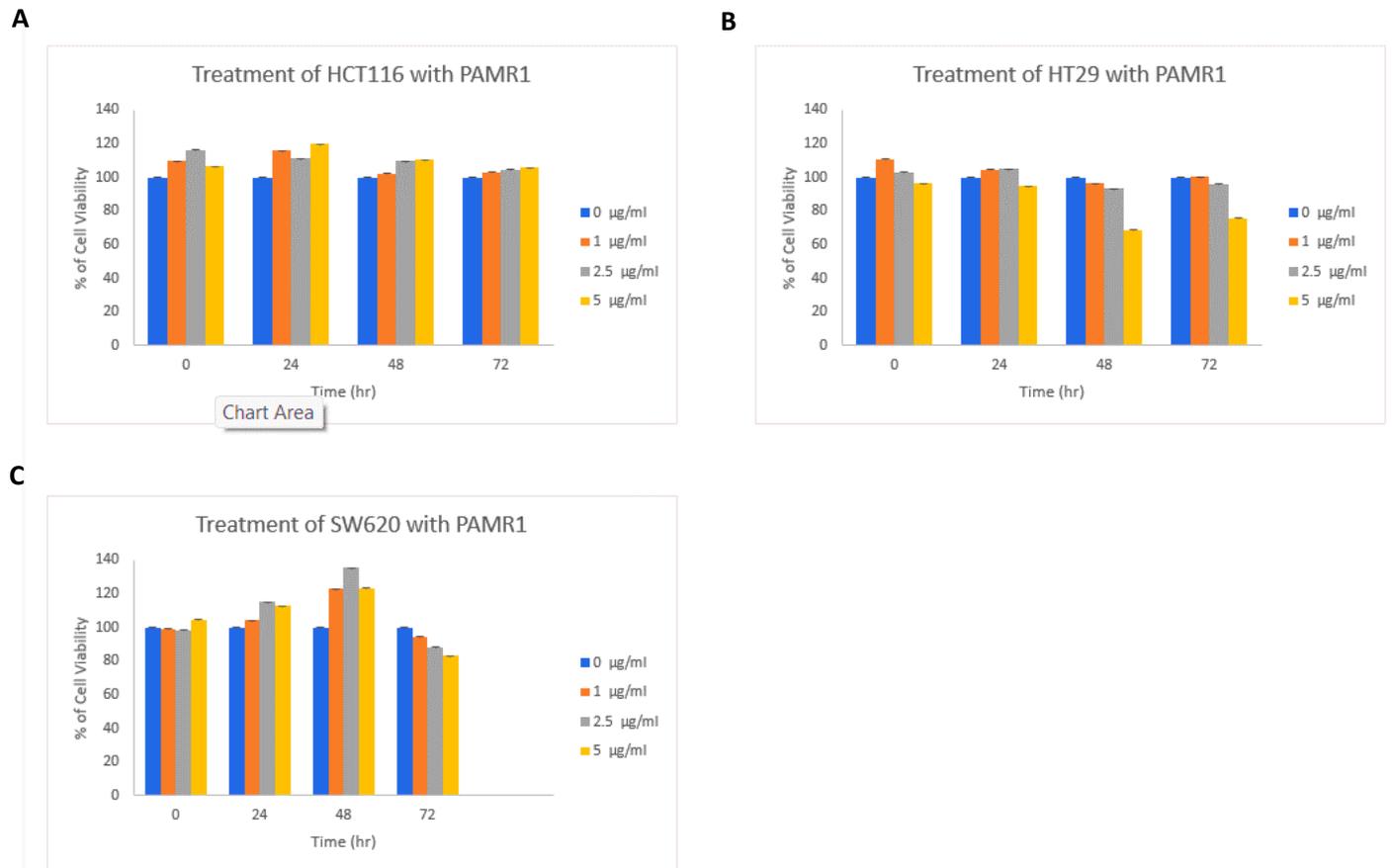
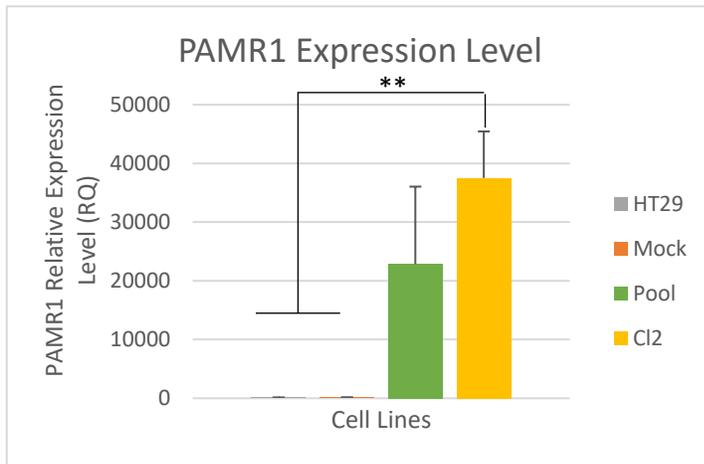


Figure 5: Cell viability assay (CCK8) for CRC cell lines. Representative histograms showing the percentage of cell viability \pm SEM for HCT116 (A), HT29 (B) and SW620 (C) cell lines, exogenously treated with different doses of purified recombinant PAMR1 (0, 1, 2.5 and 5 μ g/ml), added at time point 0 h. Then, cells were incubated for 24, 48 or 72h before absorbance measurement of metabolized CCK8.

A



B

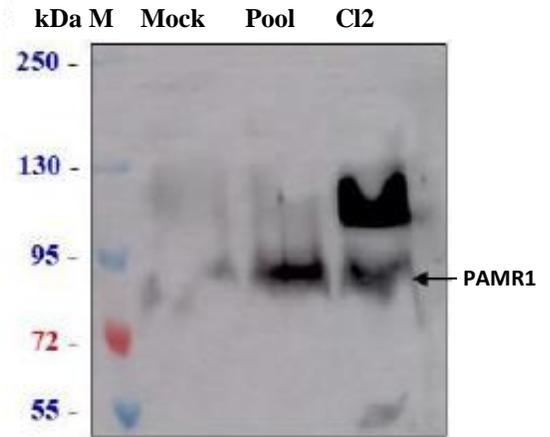
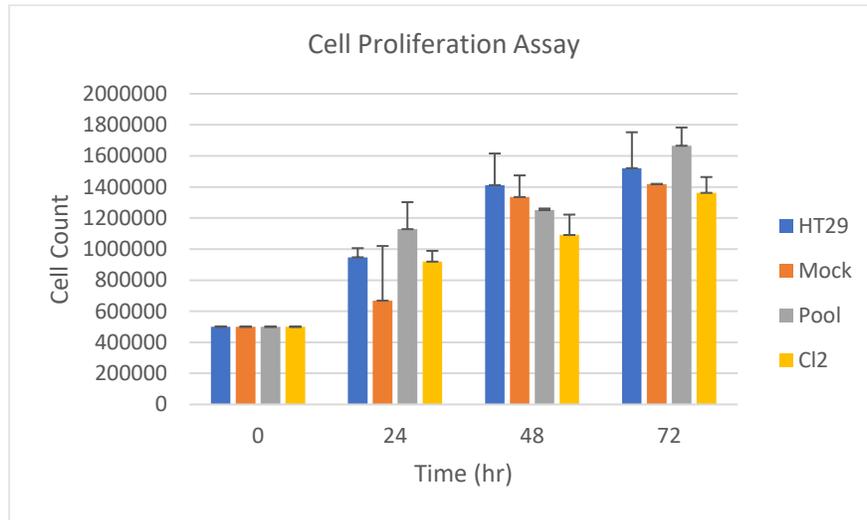
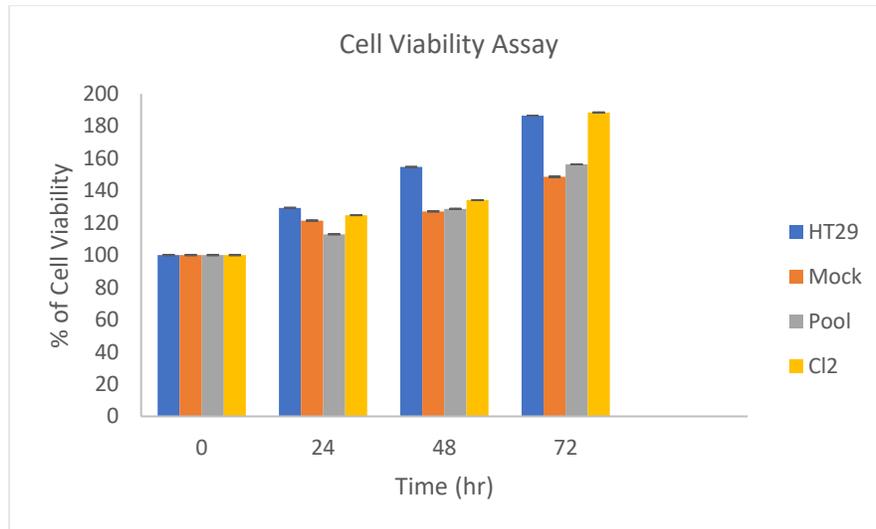


Figure 6: PAMR1 expression level in colorectal cancer HT29 cell line and derived stable cell lines. (A) PAMR1 expression at transcriptomic level in HT29 in comparison to stably transfected cell lines (Mock, Pool, and CI2). The mRNA data of each cell line is normalized to its corresponding GAPDH mRNA level. The bar graph represents mean \pm SEM. * $p < 0.01$, ** $p < 0.05$, *** $p < 0.001$. (B) Western blot analysis for PAMR1 protein expression in the concentrated secretome of Mock, Pool and Clone 2 (CI2) using anti-PAMR1 antibody.

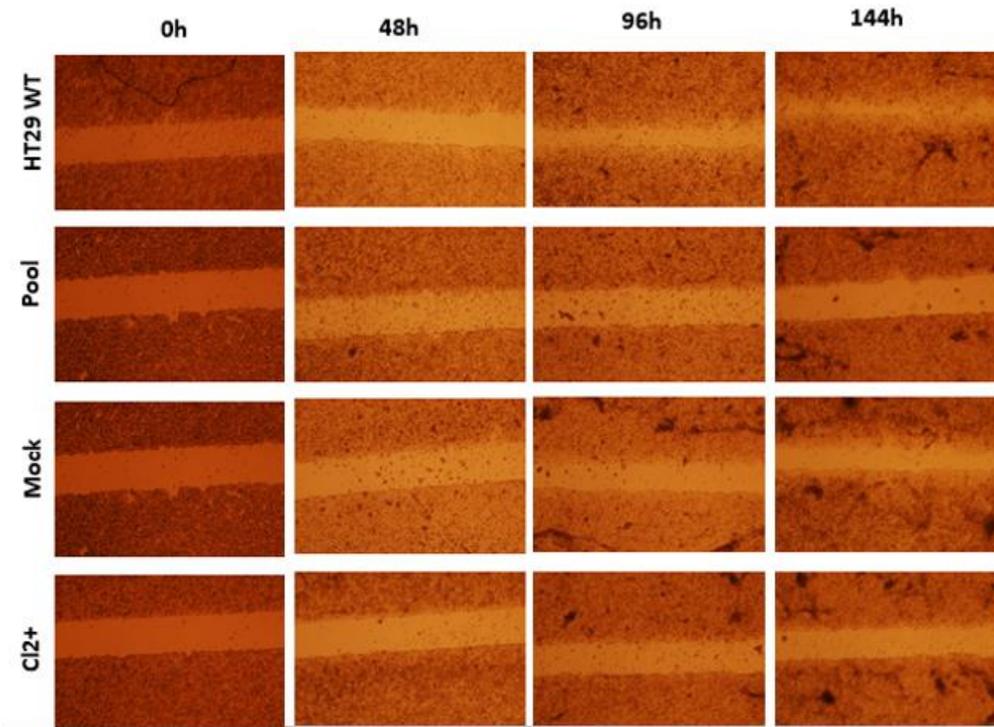
A



B



C



D

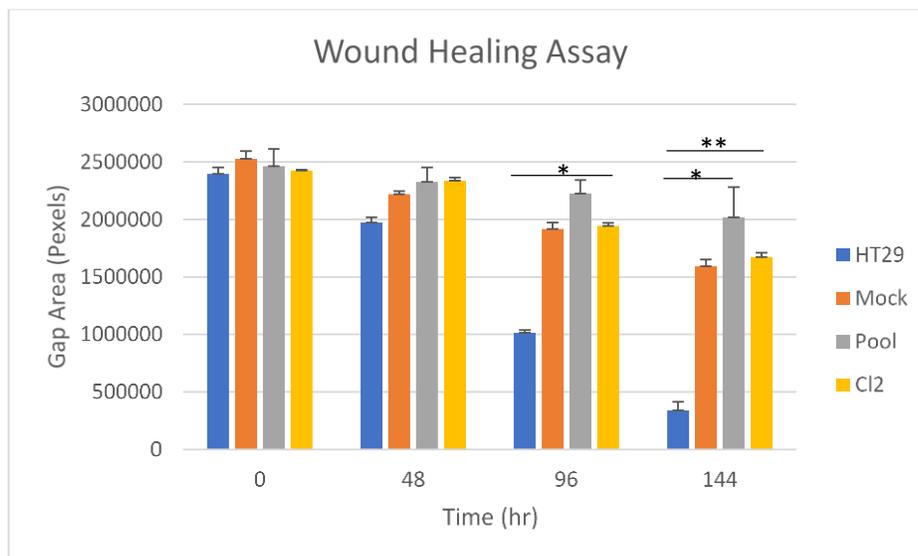


Figure 7: PAMR1 overexpression has no effect on cell viability, proliferation, and migration of stable HT29 hygromycin B-resistant cell lines. Cell proliferation (A) and cell viability (B) assays for HT29 stable cell lines (Mock, Pool, Cl2) compared to non-transfected HT29 cells, respectively. (C) Wound healing assays for HT29 and same stably transformed cell lines at different time points. (D) The bar graphs represent the mean of gap closure \pm SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

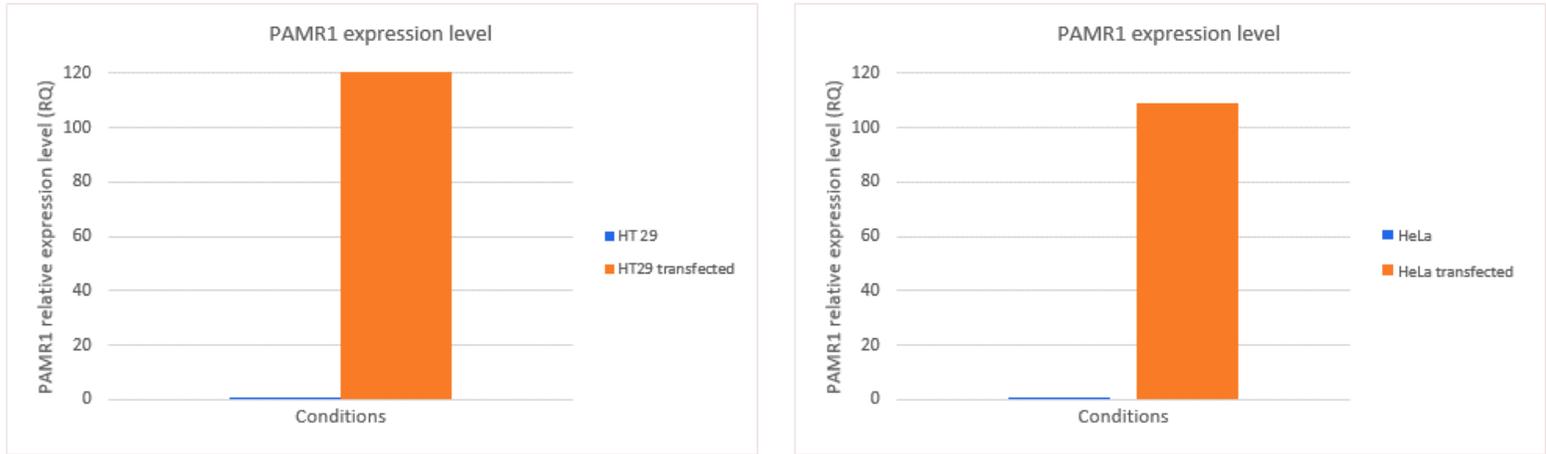
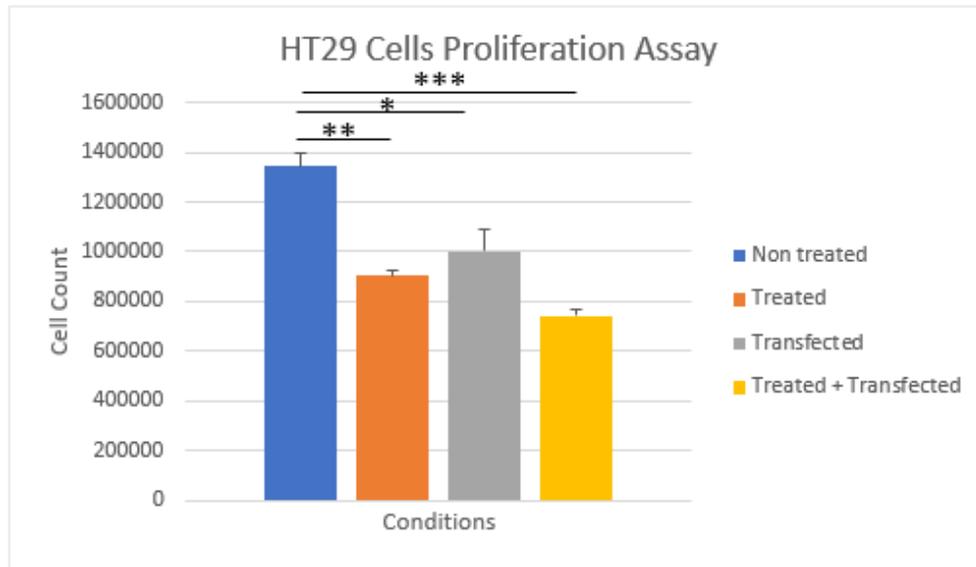
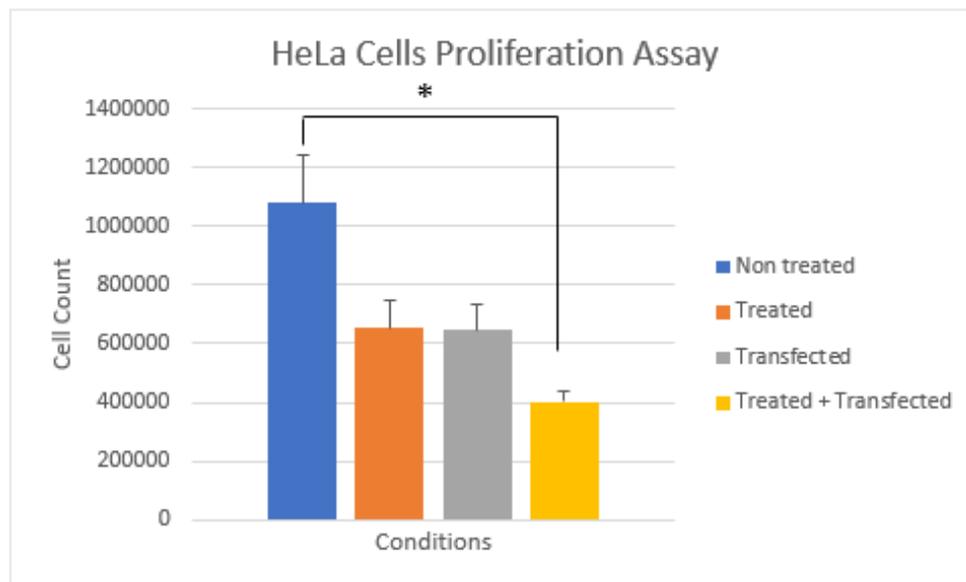


Figure 8: Overexpression of PAMR1 in transiently transfected cells. PAMR1 is overexpressed in transiently transfected cells by pCDNA- [hPAM iso 1] to a ratio 1:4 with respect to DNA transfection reagent. HT29 cells (left panel) and HeLa cells (right panel) compared to non-transfected ones.

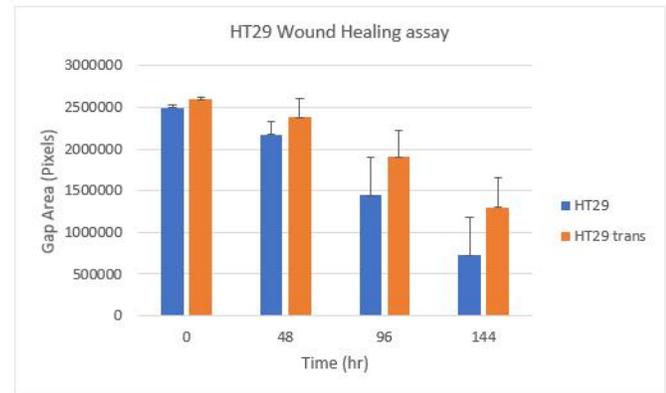
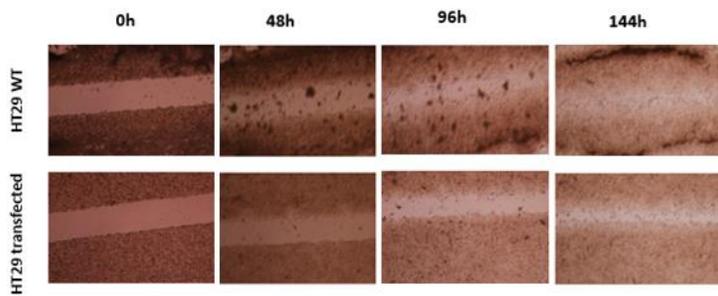
A



B



C



D

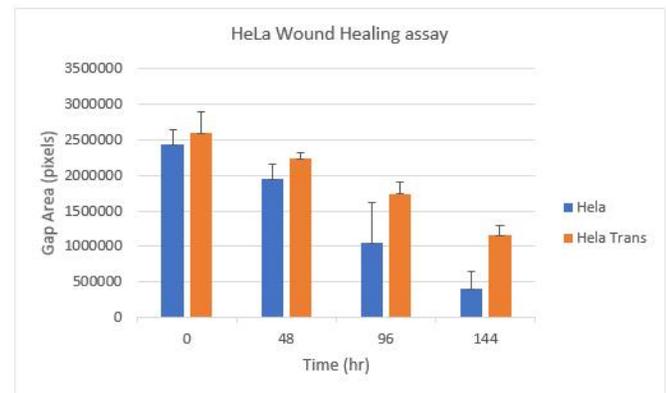
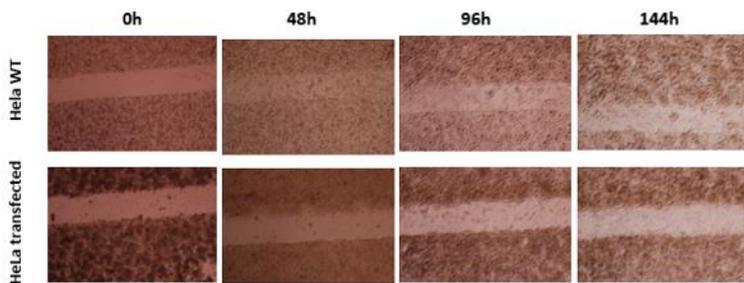


Figure 9: PAMR1 reduces HT29 and HeLa cells proliferation and migration. Cell proliferation assay for HT29 (A) and HeLa cells (B) in different conditions. Non treated: WT cells exogenously treated with concentrated CHO-mPAMR1 secretome. “Treated” corresponds to exogenously treated HT29 with concentrated CHO-mPAMR1 secretome. “Transfected” corresponds to transiently transfected cells by pCDNA3.1-hPAMR1. “Transfected + Treated” means that transfected cells were exogenously treated by concentrated CHO-mPAMR1 secretome. The bar graph represents mean \pm SEM. * $p < 0.05$, ** $p < 0.001$. Wound healing assays for non-transfected HT29 and transiently transfected HT29 (C) and HeLa (D) cells at different time points. The bar graphs represent the mean of gap closure \pm SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

REFERENCES

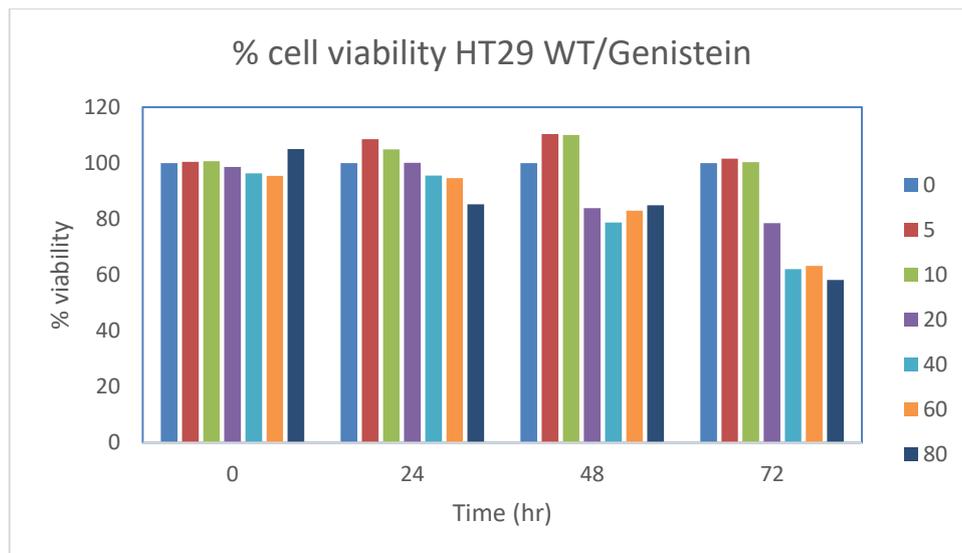
- Akkoca, A.N., Yanık, S., Ozdemir, Z.T., Cihan, F.G., Sayar, S., Cincin, T.G., Cam, A., Ozer, C., 2014. TNM and Modified Dukes staging along with the demographic characteristics of patients with colorectal carcinoma. *Int. J. Clin. Exp. Med.* 7, 2828–2835.
- Byrne, R.M., 2017. Colorectal polyposis and inherited colorectal cancer syndromes. *Ann. Gastroenterol.* <https://doi.org/10.20524/aog.2017.0218>
- Cheng, C.-J., Lin, Y.-C., Tsai, M.-T., Chen, C.-S., Hsieh, M.-C., Chen, C.-L., Yang, R.-B., 2009. SCUBE2 Suppresses Breast Tumor Cell Proliferation and Confers a Favorable Prognosis in Invasive Breast Cancer. *Cancer Res.* 69, 3634–3641. <https://doi.org/10.1158/0008-5472.CAN-08-3615>
- Diergaarde, B., Braam, H., Vasen, H.F., Nagengast, F.M., van Muijen, G.N.P., Kok, F.J., Kampman, E., 2007. Environmental Factors and Colorectal Tumor Risk in Individuals With Hereditary Nonpolyposis Colorectal Cancer. *Clin. Gastroenterol. Hepatol.* 5, 736–742.e1. <https://doi.org/10.1016/j.cgh.2007.02.019>
- Dukes, C.E., 1932. The classification of cancer of the rectum. *J. Pathol. Bacteriol.* 35, 323–332. <https://doi.org/10.1002/path.1700350303>
- Hadjipetrou, A., Anyfantakis, D., Galanakis, C.G., Kastanakis, M., Kastanakis, S., 2017. Colorectal cancer, screening and primary care: A mini literature review. *World J. Gastroenterol.* 23, 6049–6058. <https://doi.org/10.3748/wjg.v23.i33.6049>
- Jasperson, K.W., Tuohy, T.M., Neklason, D.W., Burt, R.W., 2010. Hereditary and Familial Colon Cancer. *Gastroenterology* 138, 2044–2058. <https://doi.org/10.1053/j.gastro.2010.01.054>
- Kuo, L.-J., Leu, S.-Y., Liu, M.-C., Jian, J.J.-M., Cheng, S.H., Chen, C.-M., 2003. How aggressive should we be in patients with stage iv colorectal cancer? *Dis. Colon Rectum* 46, 1646–1652. <https://doi.org/10.1007/BF02660770>
- Lin, Y.-C., Lee, Y.-C., Li, L.-H., Cheng, C.-J., Yang, R.-B., 2013. Tumor suppressor *SCUBE2* inhibits breast-cancer cell migration and invasion through the reversal of epithelial-mesenchymal transition. *J. Cell Sci.* jcs.132779. <https://doi.org/10.1242/jcs.132779>
- Lo, P.H.Y., Tanikawa, C., Katagiri, T., Nakamura, Y., Matsuda, K., 2015. Identification of novel epigenetically inactivated gene PAMR1 in breast carcinoma. *Oncol. Rep.* 33, 267–273. <https://doi.org/10.3892/or.2014.3581>
- Luther, K.B., Haltiwanger, R.S., 2009. Role of unusual O-glycans in intercellular signaling. *Int. J. Biochem. Cell Biol.* 41, 1011–1024. <https://doi.org/10.1016/j.biocel.2008.10.001>
- Mukai, T., Uehara, K., Aiba, T., Nakamura, H., Ebata, T., Nagino, M., 2018. Outcomes of stage IV patients with colorectal cancer treated in a single institution: What is the key to the long-term survival? *J. Anus Rectum Colon* 2, 16–24. <https://doi.org/10.23922/jarc.2017-021>
- Nakayama, Y., Nara, N., Kawakita, Y., Takeshima, Y., Arakawa, M., Katoh, M., Morita, S., Iwatsuki, K., Tanaka, K., Okamoto, S., Kitamura, T., Seki, N., Matsuda, R., Matsuo, M., Saito, K., Hara, T., 2004. Cloning of cDNA Encoding a Regeneration-Associated Muscle Protease Whose Expression Is Attenuated in Cell Lines Derived from Duchenne Muscular Dystrophy Patients. *Am. J. Pathol.* 164, 1773–1782. [https://doi.org/10.1016/S0002-9440\(10\)63735-2](https://doi.org/10.1016/S0002-9440(10)63735-2)

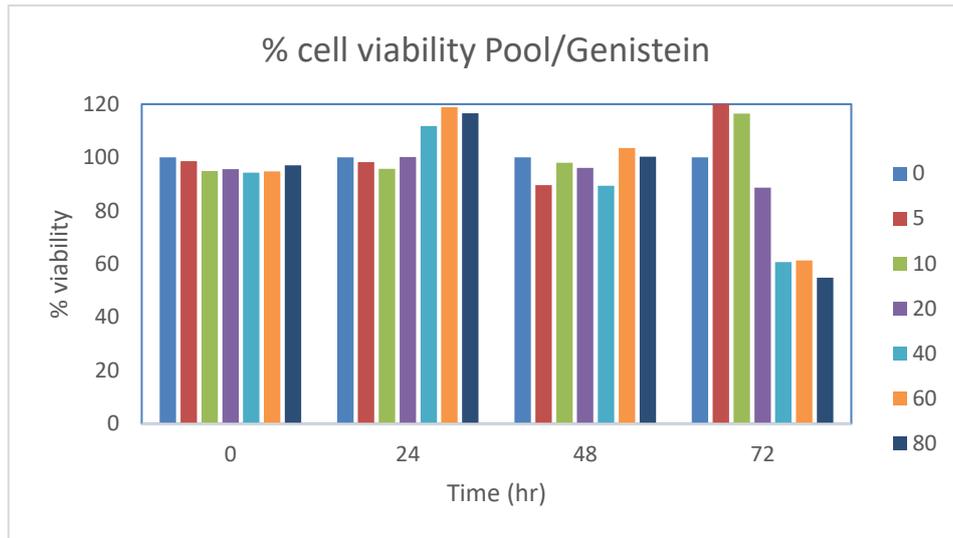
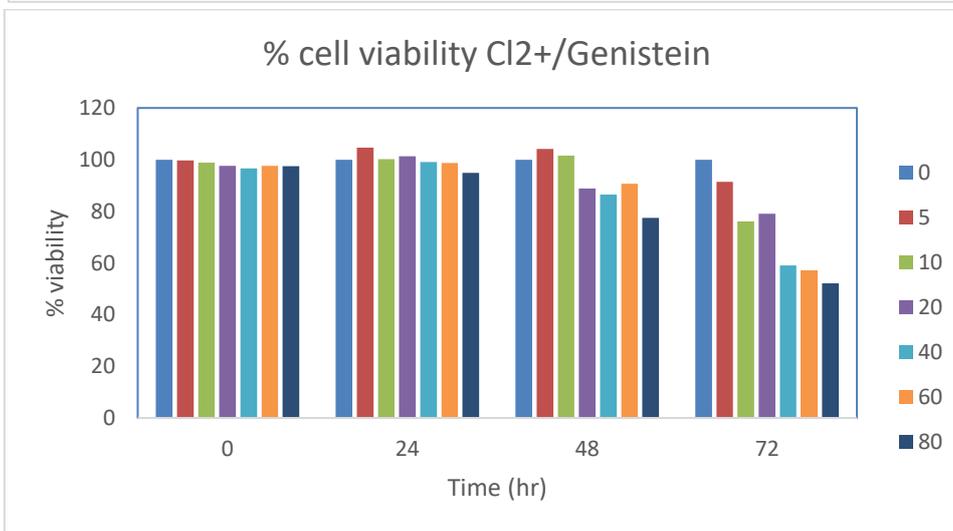
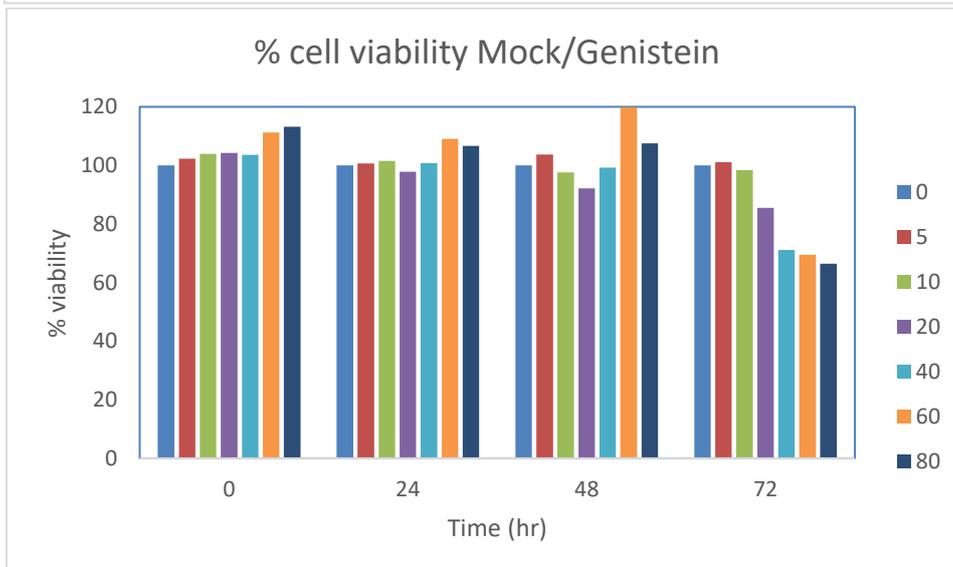
- Neamțiu, L., Martos, C., Giusti, F., Negrão Carvalho, R., Randi, G., Dimitrova, N., Flego, M., Dyba, T., Bettio, M., Gavin, A., Visser, O., the ENCR Steering Committee, Gavin, A., Visser, O., Sánchez, M.J., Eden, M., Stracci, F., Šekerija, M., Trojanowski, M., Bray, F., van Eycken, E., Miranda, A., Storm, H., 2022. Impact of the first wave of the COVID-19 pandemic on cancer registration and cancer care: a European survey. *Eur. J. Public Health* 32, 311–315. <https://doi.org/10.1093/eurpub/ckab214>
- Okajima, T., Xu, A., Irvine, K.D., 2003. Modulation of notch-ligand binding by protein O-fucosyltransferase 1 and fringe. *J. Biol. Chem.* 278, 42340–42345. <https://doi.org/10.1074/jbc.M308687200>
- Pennarubia, F., Germot, A., Pinault, E., Maftah, A., Legardinier, S., 2020. The single EGF-like domain of mouse PAMR1 is modified by O-Glucose, O-Fucose and O-GlcNAc. *Glycobiology*. <https://doi.org/10.1093/glycob/cwaa051>
- Pennarubia, F., Pinault, E., Maftah, A., Legardinier, S., 2018. In vitro acellular method to reveal O-fucosylation on EGF-like domains. *Glycobiology* 29, 192–198. <https://doi.org/10.1093/glycob/cwy106>
- Rand, M.D., Lindblom, A., Carlson, J., Villoutreix, B.O., Stenflo, J., 1997. Calcium binding to tandem repeats of EGF-like modules. Expression and characterization of the EGF-like modules of human Notch-1 implicated in receptor-ligand interactions. *Protein Sci.* 6, 2059–2071. <https://doi.org/10.1002/pro.5560061002>
- Valle, L., 2014. Genetic predisposition to colorectal cancer: Where we stand and future perspectives. *World J. Gastroenterol.* 20, 9828. <https://doi.org/10.3748/wjg.v20.i29.9828>
- Wei, W., Chen, Y., Xu, J., Zhou, Y., Bai, X., Yang, M., Zhu, J., 2018. Identification of Biomarker for Cutaneous Squamous Cell Carcinoma Using Microarray Data Analysis. *J. Cancer* 9, 400–406. <https://doi.org/10.7150/jca.21381>
- Wouters, M.A., Rigoutsos, I., Chu, C.K., Feng, L.L., Sparrow, D.B., Dunwoodie, S.L., 2005. Evolution of distinct EGF domains with specific functions. *Protein Sci.* 14, 1091–1103. <https://doi.org/10.1110/ps.041207005>
- Yang, R., Ma, M., Yu, S., Li, X., Zhang, J., Wu, S., 2021. High Expression of PAMR1 Predicts Favorable Prognosis and Inhibits Proliferation, Invasion, and Migration in Cervical Cancer. *Front. Oncol.* 11, 742017. <https://doi.org/10.3389/fonc.2021.742017>
- Yin, F., Shu, L., Liu, X., Li, T., Peng, T., Nan, Y., Li, S., Zeng, X., Qiu, X., 2016. Microarray-based identification of genes associated with cancer progression and prognosis in hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* 35, 127. <https://doi.org/10.1186/s13046-016-0403-2>
- Yu, S.-H., Cai, J.-H., Chen, D.-L., Liao, S.-H., Lin, Y.-Z., Chung, Y.-T., Tsai, J.J.P., Wang, C.C.N., 2021. LASSO and Bioinformatics Analysis in the Identification of Key Genes for Prognostic Genes of Gynecologic Cancer. *J. Pers. Med.* 11, 1177. <https://doi.org/10.3390/jpm11111177>
- Zhu, J., Ren, J., Tang, L., 2018. Genistein inhibits invasion and migration of colon cancer cells by recovering WIF1 expression. *Mol. Med. Rep.* <https://doi.org/10.3892/mmr.2018.8760>

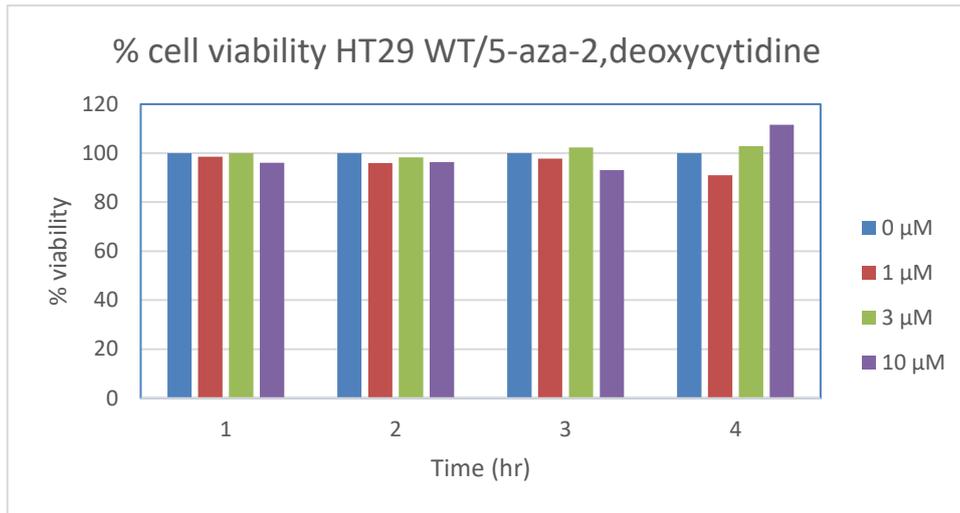
Supplementary Data:

We were able to identify and confirm the downregulation of PAMR1 in CRC. However, the mechanism of PAMR1 inactivation in CRC is unknown yet. Taking into consideration that hypermethylation of the CpG islands of the tumor suppressor gene's promoter region is the most driving cause of its inactivation in cancer, as the case in breast cancer (Yang et al., 2021), we wondered whether PAMR1's suppression in CRC is due to promoter hypermethylation as well. We investigated the effect of demethylation drugs such Genistein and 5-aza-2'-deoxycytidine on HT29 CRC cell line as well as stable HT29 cells overexpressing PAMR1 (namely Pool and Cl2) in comparison to the control stable HT29 (transfected with empty vector), Mock cell. Since treatment by 5-aza-2'-deoxycytidine recovered PAMR1 expression in breast cancer cells, MCF-7 cells (available in our lab) were used as a control upon CRC – 5-aza-2'-deoxycytidine treatment. These cells were treated with different doses of the above mentioned drugs (Genistein: 0, 5, 10, 20, 40, 60 and 80 $\mu\text{mol.L}^{-1}$. 5-aza-2'-deoxycytidine: 0, 1, 3 and 10 μM), at different time points (0, 24h, 48h, 72h) followed by cell viability assay using CCK8 (Cell Counting Kit 8). Unfortunately, treatment of CRC cells with demethylation drugs showed no effect on cell viability. This experiment was done only twice with the same result outcome. These preliminary results are discussed later in the General Discussion.

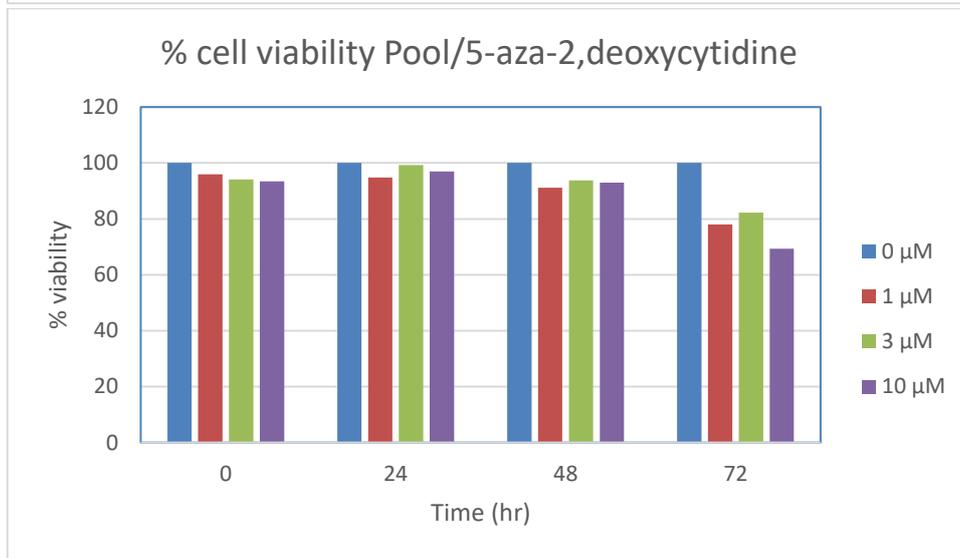
A



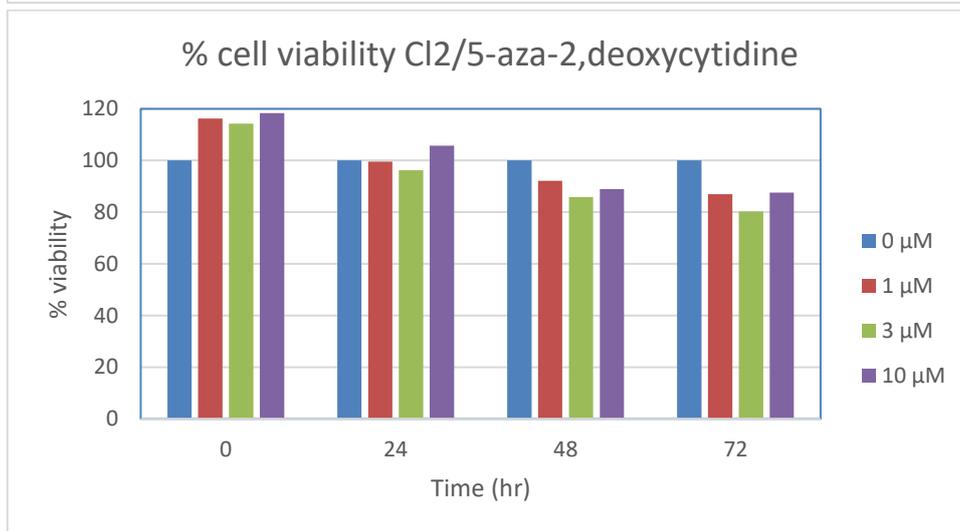
B**C****D****E**



F



G



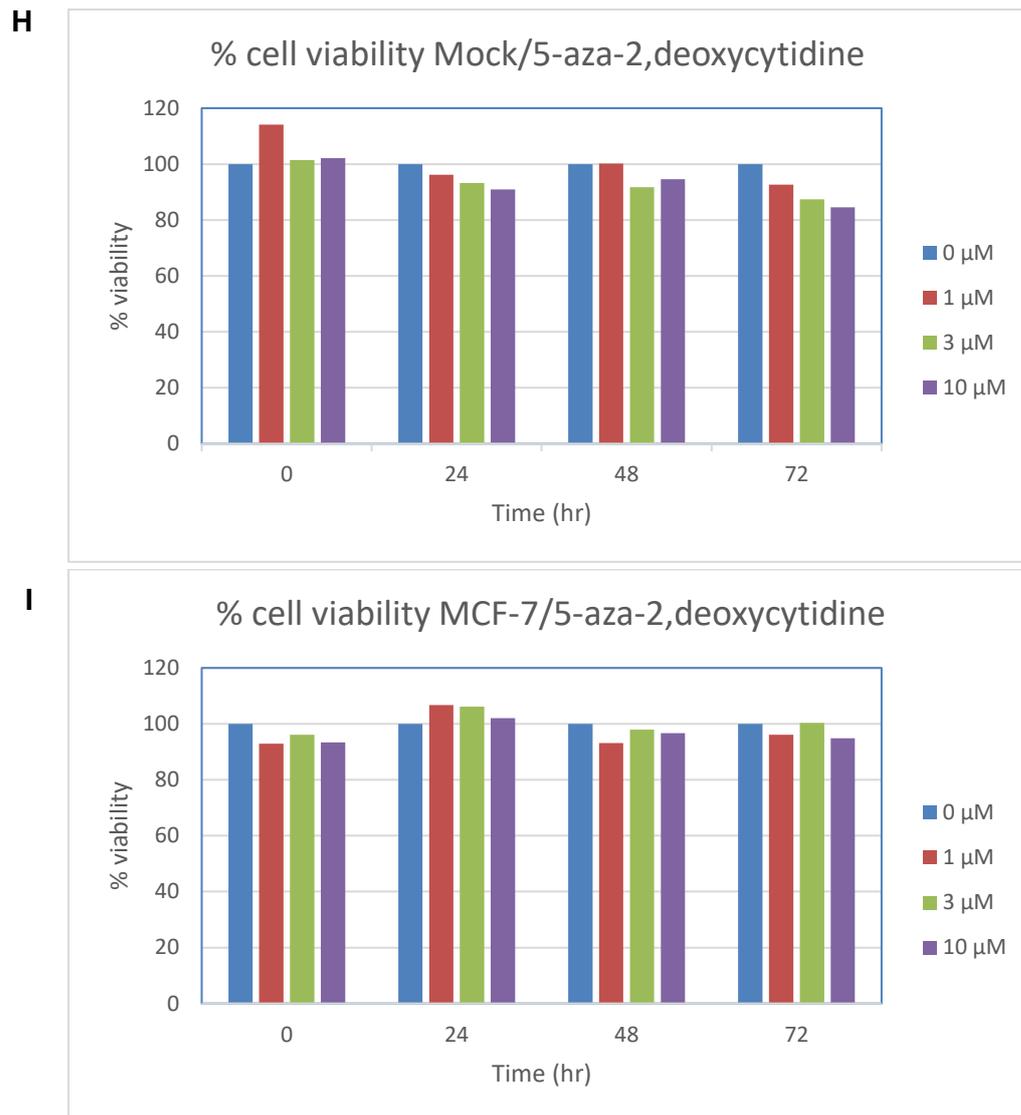


Figure S 1: Cell Viability assay of HT29 WT and stable HT29 cells (Pool, CI2, and Mock) treated with demethylation drugs

Representative histograms showing the percentage of cell viability of HT29 WT (A), Pool (B), CI2 (C), Mock (D) treated with Genistein (0, 5, 10, 20, 40, 60, 80 $\mu\text{mol.L}^{-1}$), added at time 0h. Representative histogram showing the percentage of cell viability of HT29 WT (E), Pool (F), CI2 (G), Mock (H), and MCF-7 (I) treated with 5-aza-2'-deoxycytidine (0, 1, 30, 10 μM), added at time 0h. The absorbance of metabolized CCK8 (460nm) was measured after incubation 0h, 24h, 48h, and 72h.

It was impossible to perform exogenous treatments of CRC cell lines with purified human PAMR1 isoforms 1 and 2 (instead of mouse PAMR1 used in the paper) produced by stable CHO cell lines because of a too low expression of human PAMR1 and thus too low amount of the recombinant protein in the secretome of CHO cells. We thus chose to produce both isoforms of human PAMR1 in the baculovirus insect cell system, known to provide high-level of expression of recombinant proteins and at a low cost. In spite of a good expression of human PAMR1 in complete growth medium, the production in serum-free medium combined with nickel-affinity purification led to protein degradation and aggregation of PAMR1.

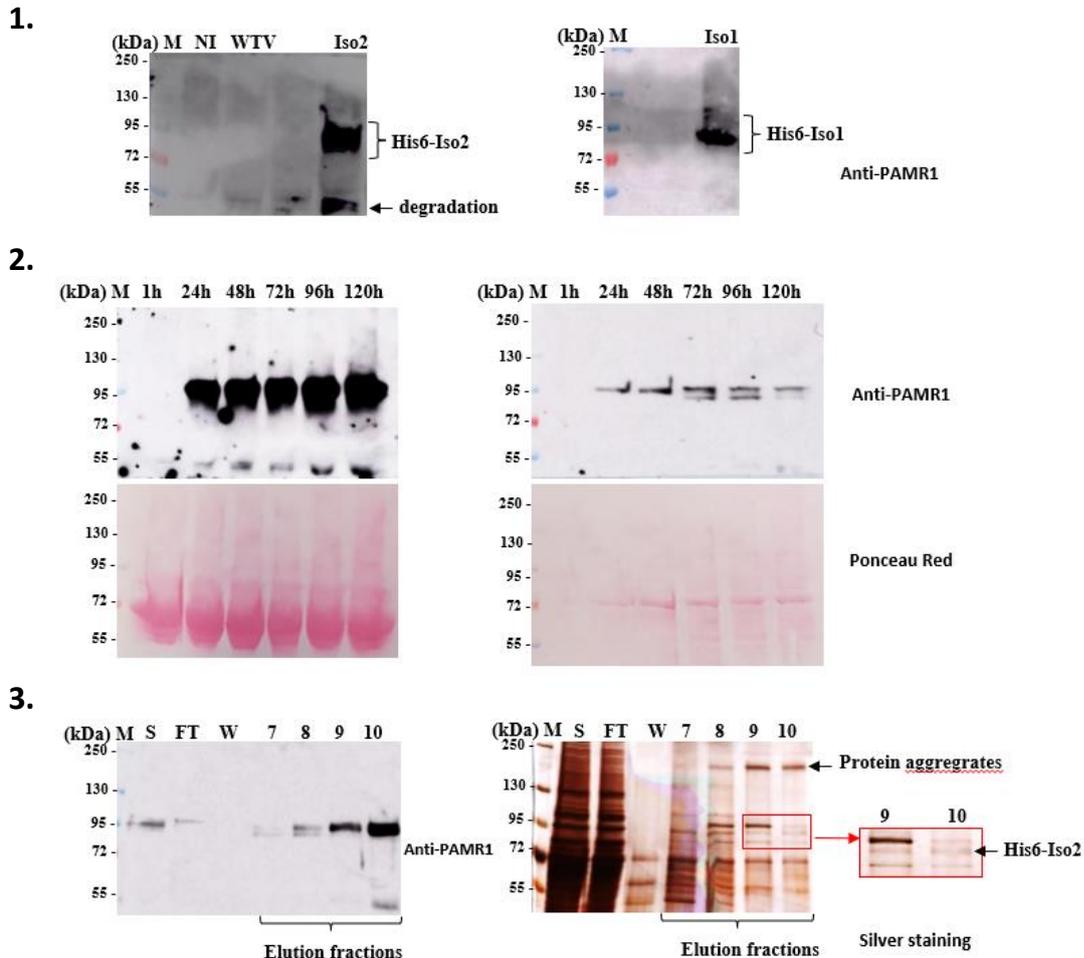


Figure S 2: Purification of recombinant human PAMR1 His6-Iso1 and His6-Iso2 in the baculovirus insect cell expression system

S2.1: Western blot for detection of His6-hPAMR1 iso 1 (right panel) and Iso 2 (left panel) produced by the baculovirus-insect cell system. NI: Non-Infected *Sf9* cells. WT V: WT baculoVirus. S2.2: The kinetics of production of His6-Iso2 by baculovirus-infected *Sf9* cells in the complete medium (left panel) and serum-free medium (right panel) by Western blot and Ponceau red staining. S2.3: Detection by Western blot (left panel) and silver nitrate staining (right panel) of His-Iso2 in elution fractions after nickel affinity purification of the protein from secretome of infected *Sf9* insect cells by recombinant baculovirus. Fractions 9 and 10 are the purest fractions. All fractions contain contaminating proteins. Higher MW band was seen probably as a result of protein aggregation.

Table S 1: The expected molecular weights (MW) of recombinant human PAMR1 isoforms 1 and 2 with *N*-terminal histidine tags produced by baculovirus-infected *Sf9* insect cells (named His6-Iso1 and His6-Iso2) and canonical isoform 1 of human PAMR1 overexpressed in HT29 colorectal cancer cell line after cleavage of signal peptide. Taking into consideration predictions of *N*-glycosylation and *O*-GalNAc glycosylation sites using online bioinformatic tools.

Proteins	After signal peptide cleavage (aa)	MW (Da)	N-Glycosylation				Mucine type O-glycosylation		Estimated MW
			Consensus sequences	Predicted sites	Theoretical MW	Mass of N-glycans	Predicted sites	Mass of O-glycans	
His6-Iso1	707	78915	5	4	1 kDa per N-glycan	4 kDa	18	4-7,2 kDa	90,1 kDa
His6-iso2	724	80659	5	4	1 kDa per N-glycan	4 kDa	19	4,2-7,2 kDa	91,8 kDa
Iso1 in HT29 and CHO cells (mammalian cells)	699	77850	5	4	2,5 kDa per N-glycan	10 kDa	17	3,8-12 kDa	100 kDa

The estimated molecular mass of paucimannose-type *N*-glycans ± fucose (on chitobiose) in *Sf9* cells (GlcNAc₂Man₃) is about 1 kDa, whereas complex *N*-glycans in mammalian cells, namely bi-antennary complex-type *N*-glycans ± fucose (GlcNAc₂Man₃GlcNAc₂Gal₂Sia₂), is about 2.5 kDa.

Mucin-like *O*-glycans in *Sf9* cells are predominantly *O*-GalNAc ± Gal type whereas in mammalian cells their structure is of *O*-GalNAc ± Gal ± Sia type or more complex (bi-antennary, etc.).

Assuming that all predicted glycosylation sites are occupied, the molecular weights of His6-Iso1 and His6-Iso2 in *Sf9* cells could be 90.1 kDa and 91.8 kDa respectively. The molecular weight of iso1 overexpressed in clone 2 of stable HT29 is about 100 kDa or even more.

Bioinformatic tools used:

- ProtParam (<https://web.expasy.org/protparam/>)
- NetNGlyc (<https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0>)
- NetOGlyc (<https://services.healthtech.dtu.dk/service.php?NetOGlyc-4.0>)

Molar masses of monosaccharides: Sialic acid (Sia): 309 g/mol, Fucose (Fuc): 164 g/mol, Galactose (Gal): 180 g/mol, Glucose (Glc): 180 g/mol, Mannose (Man): 180 g/mol, N-Acetylglucosamine (GlcNAc): 221 g/mol, N-Acetylgalactosamine (GalNAc): 221 g/mol.

POFUT1-mediated O-fucosylation of glycoproteins expressed in the baculovirus Sf9 insect cell expression system

Layla Haymour¹, Florian Pennarubia², Claire Colombel Le Faou¹, Emilie Pinault³, Agnès Germet¹, Abderrahman Maftah¹ and Sébastien Legardinier^{1, §}

¹ Univ. Limoges, LABCiS, UR 22722, F-87000 Limoges, France. ² Complex Carbohydrate Research Center, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA. ³ University of Limoges, BISCEM, US 42 Inserm – UAR 2015 CNRS, Mass Spectrometry Platform, F-87025 Limoges, France

§ To whom correspondence should be addressed

In preparation

To perform exogenous treatments of CRC cell lines, we first chose to produce the secreted glycoprotein human PAMR1 (isoforms 1 and 2) in stable CHO cell lines, generated after targeted insertion of transgene using the Flp-In™ System (Thermo Fisher Scientific). This Flp-In System enabled the generation of stable mammalian expression CHO cell lines, after a site-specific recombination to allow integration of the gene of interest into a specific site in the genome of CHO cells. Contrary to recombinant mouse PAMR1, which was previously efficiently produced in the secretome of stable CHO cells (Pennarubia et al., 2021), both isoforms 1 and 2 of human PAMR1 were much more less expressed by stable CHO cell lines that I obtained. We thus decided to produce human PAMR1 in the Baculovirus insect cell expression system, known to allow high yield of production of recombinant proteins.

Human PAMR1 isoforms 1 and 2 were efficiently produced and secreted in the supernatant of *Sf9* insect cells infected with recombinant baculoviruses but the production of the protein in serum-free medium followed by its nickel-affinity purification led to protein aggregation and proteolytic degradation of PAMR1 (See Supplementary data). To understand the origin of this protein instability, we wondered if human PAMR1 was correctly glycosylated and especially modified with O-fucose when expressed in *Sf9* insect cells. Indeed, its murine PAMR1 counterpart produced in CHO cells and more stable was previously shown to be modified with O-fucose within its single EGF-like domain (Pennarubia et al., 2021).

In the following paper, we focused on the biochemical characterization of *Spodoptera frugiperda* POFUT1 (*Sf*POFUT1) produced in *Sf9* insect cells, namely on conservation and occupation of its N-glycosylation sites as well as its O-fucosyltransferase activity. More precisely, we determined its ability to add *in vitro* O-fucose to EGF-like domains of three POFUT1 target proteins, namely EGF26 of NOTCH1, EGF3 of WIF1 and the single EGF of our protein of interest PAMR1, as previously carried out in our lab with mouse POFUT1 produced by CHO cells (Pennarubia et al., 2018) (Pennarubia et al., 2020) (Pennarubia et al., 2021). For this purpose, glycosyltransferase assays were carried out with the purified *Sf*POFUT1, GDP-azido-fucose and isolated target EGF-like domains purified from *E.coli*. Copper-catalyzed azide-alkyne cycloaddition (CuAAC), also called click chemistry, was then used to covalently bind alkyne-biotin to azido O-fucose transferred to EGF-like domains. A blotting technique using streptavidin was finally used to reveal the modification by O-fucose.

Contrary to recombinant mouse POFUT1 from CHO cell lines, which was able to specifically add *in vitro* O-fucose to isolated EGF-like domains of the three POFUT1 target proteins, *Sf*POFUT1 was totally unable to add O-fucose to the single EGF-like domain of PAMR1. Unfortunately, we were unable to demonstrate the presence or not of O-fucose on full-length PAMR1 isoforms produced in the baculovirus insect cell system due to their propension to proteolytic degradation during purification steps and protein concentration. However, by using targeted mass spectrometry (MRM-MS), we were able to show that full-length human WIF1 was as expected modified by O-fucose on its EGF3 when expressed in *Sf9* insect cells and with a high efficiency.

Overall results confirmed ability of *Sf9* POFUT1 to specifically modify some POFUT1 target proteins with O-fucose but this ability seems to depend on the protein of interest.

POFUT1-mediated *O*-fucosylation of glycoproteins expressed in the baculovirus *Sf9* insect cell expression system

Layla Haymour¹, Florian Pennarubia², Claire Colombel Le Faou¹, Emilie Pinault³, Agnès Germot¹, Abderrahman Maftah¹ and Sébastien Legardinier^{1, §}

¹University of Limoges, LABCiS, UR 22722, F-87000 Limoges, France. ²Complex Carbohydrate Research Center, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA. ³University of Limoges, BISCEM, US 42 Inserm – UAR 2015 CNRS, Mass Spectrometry Platform, F-87025 Limoges, France

§ To whom correspondence should be addressed: Tel: +33555457792; Fax: +33555457653; e-mail: sebastien.legardinier@unilim.fr

Running head: POFUT1-mediated *O*-fucosylation in *Sf9* insect cells

Keywords: Baculovirus/ EGF-like domain/ Heterologous proteins/ *O*-fucosylation/ POFUT1

Supplementary Data Included: Figures S1-S4 and Tables I-II

ABSTRACT

Around 100 mouse or human proteins including Notch receptors and their Jagged and Delta-like ligands are proteins potentially targeted by the protein *O*-fucosyltransferase 1 (POFUT1). Such glycoproteins, having at least one EGF domain prone to be modified by *O*-fucose, were largely expressed in *Drosophila* cells and some in mammalian expression systems but till now very few have been expressed in the baculovirus-insect cell expression system.

In this paper, recombinant *Spodoptera frugiperda* POFUT1 (*Sf*HisV5Po), produced and purified from baculovirus-infected *Sf9* insect cells, was biochemically characterized and compared to mouse POFUT1 (*Mm*HisV5Po) for its ability to add *in vitro* *O*-fucose to recombinant isolated mouse EGF domains of POFUT1 target proteins (NOTCH1, WIF1, PAMR1). *In vitro* POFUT1-mediated *O*-fucosylation experiments, followed by click chemistry and blot analyses, showed that recombinant purified *Sf*HisV5Po was able to add *in vitro* *O*-fucose to mouse NOTCH1 EGF26 and WIF1 EGF3, similarly to *Mm*HisV5Po produced by stable CHO cells; however, with less efficiency. Consistent with this latter result, full-length human WIF1 expressed in *Sf9* insect cells was also found with *O*-fucose on its EGF3, as proved by mass spectrometry. However, *Sf*HisV5Po was unexpectedly shown to be incapable of modifying the single EGF domain of mouse PAMR1 with *O*-fucose, contrary to *Mm*HisV5Po.

All these results confirm that *Sf9* insect cells can indeed perform POFUT1-mediated *O*-fucosylation of some EGF-containing glycoproteins, but POFUT1 efficiency can depend on the nature and properties of the recombinant protein of interest. This could be marked as a difference compared to mammalian expression system.

INTRODUCTION

The baculovirus-insect cell system is a powerful tool for the production of recombinant foreign proteins (Jarvis 2009), allowing to perform biochemical analyses and structure-function studies. Many post-translational modifications such as protein glycosylation (*N*- and *O*-glycosylation) occur in lepidopteran cells but noticeable differences exist with those found in mammalian cells (Altmann, Staudacher et al. 1999). *N*-glycans are mostly complex-type and sialylated in mammalian cells, whereas paucimannose-type *N*-glycans with a pentasaccharide structure (Man₃GlcNAc₂) are often found in insect cells such as in *Sf9* cells derived from the Lepidoptera *Spodoptera frugiperda* (Jarvis and Finn 1995, Tomiya, Narang et al. 2004, Legardinier, Klett et al. 2005). In addition, *N*-glycans of membrane glycoproteins in lepidopteran cells can bear alpha-linked fucose at the asparagine-linked GlcNAc residue (Kubelka, Altmann et al. 1994, Tomiya, Narang et al. 2004), as it was shown since a long time for recombinant glycoproteins expressed in the baculovirus-insect cell system (Wathen, Aeed et al. 1991). Indeed, MB-0503 cells from *Mamestra brassicae* were shown to be able to transfer fucose into alpha 1-3 and alpha 1-6 linkage to the innermost GlcNAc of *N*-glycans but only alpha 1-6-fucoses were detected with BM-N cells from *Bombyx mori* and *Sf9* lepidopteran cells (Staudacher, Kubelka et al. 1992). The cDNA of the gene encoding FUT8 allowing anchorage of alpha1-6-fucose to the innermost GlcNAc of *N*-glycans was indeed cloned from the *Sf9* lepidopteran cell line (Juliant, Harduin-Lepers et al. 2014).

Concerning *O*-glycosylation of insect cells, several studies reported the presence of mucin-type *O*-*N*-acetylgalactosamine (GalNAc) glycans (Lopez, Tetaert et al. 1999, Nakamura, Katano et al. 2004), which are also the most abundant *O*-glycosylations found in mammalian cells (Brockhausen, Wandall et al. 2022). This *O*-GalNAc can be extended with galactose (Gal) to form the *O*-linked disaccharide Galβ1-3GalNAc as in *Sf9* cells (Legardinier, Klett et al. 2005) and *Drosophila melanogaster* S2 cells (Schwientek, Mandel et al. 2007). Unlike mammals, this latter study reported that there was no sialyltransferase activity in *Sf9* and *Mb* cells as well as in *Trichoplusia ni* cells. Therefore, no terminal sialic acid on mucin-type *O*-glycans produced in these three lepidopteran cell lines.

In *Drosophila melanogaster*, many *O*-linked monosaccharides are found such as mucin-type *O*-GalNAc and *O*-Mannose (*O*-Man) (Zhang and Ten Hagen 2019) but also *O*-glucose (*O*-Glc), *O*-fucose (*O*-Fuc)

and *O*-N-acetylglucosamine (*O*-GlcNAc) usually found in EGF-like domain (EGF) of membrane proteins such as NOTCH receptors (Harvey, Rana et al. 2016). However, few studies reported the characterization of EGF-containing glycoproteins and potentially modified by such *O*-linked monosaccharides in the baculovirus-insect cell system using lepidopteran cells. Nevertheless, the X-ray structure of rat NOTCH1 EGF11-13, expressed in the baculovirus insect cell system using Hi-Five (Invitrogen) cells from *Trichoplusia ni*, exhibited *O*-Glucose in the loops C¹-C² or C³-C⁴ of different EGFs (EGF11, 12 or 13) and *O*-Fucose in the loop C²-C³ of EGF12 at the different expected sites. The presence of these *O*-linked monosaccharides revealed that such insect cells can express functional POFUT1, POGLUT1 (Luca et al., 2015), as well as POGLUT2/3 (Takeuchi, Schneider et al. 2018). In addition, this study of Luca et al. showed the elongation of *O*-fucose with GlcNAc also revealing a Fringe activity (GlcNAc transferase) in these *Trichoplusia ni*-derived cells.

In *Drosophila melanogaster* and mammals, many studies showed that the homologous Notch receptors (one in fly and 4 in mammals), which contain 29-36 EGF repeats in their extracellular part, were highly modified with *O*-fucose, following the action of the protein *O*-fucosyltransferase 1 (POFUT1). POFUT1-mediated *O*-fucosylation is known to modulate interactions of NOTCH receptors with the DSL (Delta, Serrate, LAG-2) family ligands (Okajima, Xu et al. 2003, Luther and Haltiwanger 2009). In human and mouse, there are, for each, about 100 target proteins of POFUT1, including WIF1 and PAMR1. We recently showed that WIF1, an inhibitor of the Wnt signaling pathway, exhibited on its EGF3 an *O*-fucose required for its optimal secretion (Pennarubia, Pinault et al. 2020). We also showed that PAMR1, another secreted multidomain protein considered as tumor suppressor in breast cancer (Lo, Tanikawa et al. 2015), was also modified with *O*-fucose on its unique EGF (Pennarubia, Germot et al. 2021).

In this work, a secreted form of *Spodoptera frugiperda* POFUT1 with *N*-terminal polyhistidine and V5 tags (named *SfHisV5Po*) was first purified from the supernatant of baculovirus-infected *Sf9* cells and characterized for its *N*-glycosylation and its *O*-fucosyltransferase activity. According to our results of PNGase F deglycosylation, lectin blot and mass spectrometry, recombinant *SfHisV5Po* only exhibited only one *N*-glycan at the highly evolutionary conserved across Metazoa NRT site, whereas recombinant mouse POFUT1 (*MmHisV5Po*) from stable CHO cells exhibited an additional second N-glycan at N¹⁶⁵

position. We then decided to compare the ability and efficiency of *SfHisV5Po* and *MmHisV5Po* to modify EGFs of different glycoproteins with *O*-fucose. Using copper-catalyzed azide-alkyne cycloaddition (CuAAC) referred to as click chemistry, we demonstrated that purified *SfHisV5Po* was able to specifically add *in vitro* *O*-fucose to isolated EGFs of NOTCH1 EGF26 and WIF1 EGF3, but with less efficiency compared to *MmHisV5Po*. This result was confirmed by using multiple reaction monitoring-mass spectrometry (MRM-MS) for recombinant full-length human WIF1 (HisWIF1) bearing an *O*-fucose on its EGF3 when expressed in baculovirus-infected *Sf9* insect cells. However, *SfHisV5Po* was unexpectedly unable to add *in vitro* *O*-fucose to the single EGF domain of PAMR1 contrary to *MmHisV5Po*.

RESULTS

Spodoptera frugiperda POFUT1 carries two potential *N*-glycosylation sites

We chose to align the primary sequence of POFUT1 from *Spodoptera frugiperda* (*Sf*) with those of POFUT1, whose X-ray structures were known, namely those of *Homo sapiens* (McMillan, Zimmerman et al. 2017), *Mus musculus* (Li, Han et al. 2017) and the nematod *Caenorhabditis elegans* (*Ce*) (Lira-Navarrete, Valero-Gonzalez et al. 2011). Regions involved in binding to GDP-fucose (called substrate binding) as well as the key residues of active site involved in binding to EGFs (M⁴⁶, G⁴⁷, N⁵¹, Y⁷⁸, N¹⁵¹), as previously defined for mouse POFUT1 (Li, Han et al. 2017), were highly conserved among these species (Figure 1). The first *N*-glycosylation site NRT found at position 67 in mice was also found in *Sf*POFUT1 but not in *Ce*POFUT1 whereas the second *N*-glycosylation site NKS at position 165 in mice was not present in both *Sf*POFUT1 and *Ce*POFUT1. However, we can notice that POFUT1 from *Spodoptera frugiperda* bears a second potential *N*-glycosylation site, namely NMS site at position 219. When compared at the level of Protostomes, the NRT site was highly conserved (Figure 2A). 311 database extracted homologous sequences of POFUT1 over 348 sequences encompassing the largest possible taxonomic diversity (supplemental Table I) shared this potential *N*-glycosylation site (or a NRS site for two dipteran species), *i.e.* an overall conservation of 89.4% (especially 97.8% for Hexapoda and 65.7% for Nematoda). Concerning the NMS site, it was found in 55.9% of lepidopteran species, with a heterogeneous distribution among families. In Noctuidae, taxon of *Spodoptera*, 6 species over 8 shared

this potential *N*-glycosylation site, whereas in Pieridae, only one over 7. In protostome POFUT1, the number of potential *N*-glycosylation sites varied from 0 to 5, in most cases 1 to 2 sites, except for Collembola and Platyhelminthes where it is 4 and 5 respectively (Figure 2B).

The analysis of the X-ray structure of murine POFUT1 clearly showed that the GDP-fucose was deeply buried in the active site (Figure 3A). Conserved residues involved in binding to GDP-fucose (in orange) were close to those involved in binding to acceptor substrates (in red), namely EGFs comprising a potential *O*-fucosylation site. The presence of GlcNAc residues on N⁶⁷ and N¹⁶⁵ showed that murine POFUT1 carried as predicted two *N*-glycans. The X-ray structure of *Spodoptera frugiperda* POFUT1 was not known but an automatic structural model was obtained by using COACH-D server (Figure 3B). It allowed us to obtain a structural model for *Sf*POFUT1 by comparison with the X-ray structure of *Ce*POFUT1 (PDB 3ZY5) (Lira-Navarrete, Valero-Gonzalez et al. 2011). To position the donor substrate, namely the GDP-fucose, in the cavity of the active site as seen in X-ray structure of *Mm*POFUT1 (Figure 3A), a docking was performed between the known structure of GDP-fucose and the structural model for *Sf*POFUT1 (Figure 3B). The GDP-fucose appeared to be correctly positioned and oriented in the active site when compared to murine POFUT1 as well as the conserved regions required for substrate binding. The location of the *N*-glycosylation site N⁵³RT was also conserved at the back side of the enzyme. On the other hand, the N²¹⁹MS site was seen on the same side as the EGF binding site. Due to a relative distance from the active site, the potential occupation of this site by an *N*-glycan is unlikely to prevent EGFs of target glycoproteins from binding to the active site.

Recombinant *Sf*HisV5Po only carries one *N*-glycan with terminal alpha-linked mannoses

In order to characterize the *N*-glycosylation of *Spodoptera frugiperda* POFUT1, a recombinant counterpart named *Sf*HisV5Po, with N-terminal histidine and V5 tags, was produced as a secreted form in culture media of baculovirus-infected *Sf9* insect cells. Then, *Sf*HisV5Po was purified and compared to recombinant mouse POFUT1 (*Mm*HisV5Po) purified from the culture media of stable CHO cells and also used in our previous study (Pennarubia, Pinault et al. 2018).

The analysis of 1 µg of each purified enzyme on a Coomassie blue-stained polyacrylamide gel showed that *Sf*HisV5Po appeared at an apparent molecular weight (MW) of approximately 45 kDa, very close

to its predicted MW of 43601 Da (Figure 4A). Since *N*-glycans produced by *Sf9* cells are generally small (about 1 kDa per *N*-glycan) with the pentasaccharide structure GlcNAc2-Man3 (\pm fucose), we could not determine if one or both potential *N*-glycosylation sites were occupied or not. On the contrary, the purified recombinant *MmHisV5Po* from mammalian CHO cells, whose predicted MW was 43434 Da, appeared at an apparent MW of about 50 kDa, consistently with occupation of its two *N*-glycosylation sites by complex-type *N*-glycans (2-3 kDa per *N*-glycan). Using anti-V5 antibody, Western blot analysis of *SfHisV5Po*, treated or not with PNGase F, revealed a slightly lower apparent MW for deglycosylated *SfHisV5Po* showing that *SfHisV5Po* was indeed *N*-glycosylated (Figure 4B). The analysis on a silver nitrate-stained polyacrylamide gel of both recombinant purified enzymes, treated or not by PNGase F, confirmed the slightly lower MW for deglycosylated *SfHisV5Po* consistently with occupation of only one *N*-glycosylation site (Figure 4C, upper panel). On the other hand, the treatment of *MmHisV5Po* with PNGase F led to a decreased apparent MW of about 5 kDa, related to the efficient elimination of its two *N*-glycans. Consistent with their predicted MW based on amino-acid sequences (43601 Da for *SfHisV5Po* versus 43434 Da for *MmHisV5Po*), the two recombinant enzymes exhibited the same apparent MW after PNGase F treatment. Prior to analyses by lectin blot, commercial purified control glycoproteins, namely Carboxypeptidase Y (CPY) and Fetuin (F), were also subjected to PNGase F treatment in the same conditions and appeared at expected lower MW after PNGase F treatments, showing the efficient elimination of their *N*-glycans. An analysis of all these purified proteins by lectin blot was then performed using GNA lectin, recognizing terminal alpha-linked mannoses (Figure 4C, lower panel). Before PNGase F treatment, CPY, which bears high-mannose *N*-glycans and thus used as a positive control glycoprotein for GNA, strongly reacted with GNA whereas Fetuin (F), bearing complex type *N*-glycans and used as a negative control, was not recognized at all. As expected, the purified *SfHisV5Po* produced by *Sf9* insect cells was also recognized by the lectin GNA, consistent with detection of terminal alpha-linked mannoses on *N*-glycans, such as those found in paucimannose structures. However, *MmHisV5Po* from CHO cells was as expected not recognized by GNA, due to presence of complex-type *N*-glycans as found in mammalian cells. Finally, CPY and *SfHisV5Po* were no longer detected by GNA after PNGase F treatment, thus consistent with their total deglycosylation by PNGase F in our experiments.

Finally, the MS-MS analysis of tryptic peptides from the digestion of purified *SfHisV5Po* first confirmed the identity of the purified protein (Figure 4E). In addition, the detection of the trypsin digested peptide containing the potential *N*-glycosylation site \underline{N}^{219} MS (NMSGGAFLGIHLRNGQDWVK) meant that this site was therefore mainly not occupied by an *N*-glycan. On the contrary, the one carrying the \underline{N}^{53} RT conserved *N*-glycosylation site remained not detected. All these results demonstrated that *SfHisV5Po* was indeed a glycoprotein bearing an *N*-glycan with terminal mannoses at position N⁵³.

Recombinant human HisWIF1 expressed in *Sf9* cells was modified with *O*-fucose on its EGF3

We previously demonstrated that mouse WIF1 produced in CHO cells exhibited an *O*-fucose on its EGF3 (Pennarubia, Pinault et al. 2020), we thus wondered if it was the case for its human counterpart when expressed in the baculovirus expression system. Recombinant full-length human WIF1, produced with an *N*-terminal histidine tag (HisWIF1) in the secretome of infected *Sf9* insect cells, was then purified and analyzed by MRM-MS after thermolysin digestion. Figure 5A showed potential cleavage sites of thermolysin for human WIF1 EGF3. Resulting peptides were analyzed by micro-LC MRM-MS. This digestion generated a peptide of interest with the sequence FNGGT²⁵⁵CFYPGKC for human HisWIF1 EGF3 after a missed cleavage by thermolysin (Figure 5B). The non-modified peptide (Figure 5B, left panel) was almost not detected for EGF3 peptide with peak intensity of about 30 whereas EGF3 peptide modified with *O*-fucose (Figure 5B, right panel and supplemental Figure 1) was clearly mainly detected with a much higher peak intensity around 2200.

Consistent with previous results concerning the Notch1-DLL4 complex produced in Hi-Five insect cells (Luca, Jude et al. 2015), the baculovirus express system allows the addition of *O*-fucose to different recombinant mammal proteins, such as human WIF1 produced here in *Sf9* insect cells. Interestingly, the expression level of endogenous POFUT1 in *Sf9* insect cells was sufficient for addition of *O*-fucose to almost 100 % of HisWIF1 molecules produced and overexpressed after baculovirus infection.

Recombinant purified *SfHisV5Po* efficiently added *O*-fucose to mouse WIF1 EGF3 and NOTCH1 EGF26 but not to the single EGF of mouse PAMR1

As in our previous work with mouse POFUT1 (Pennarubia, Pinault et al. 2018), the ability of *SfHisV5Po* to transfer *in vitro* *O*-fucose was determined by using purified isolated EGFs of different POFUT1 target

glycoproteins containing a potential *O*-fucosylation site with the consensus sequence C₂XXXX(S/T)C₃ (Shao, Moloney et al. 2003). For this purpose, *SfHisV5Po* was incubated with an isolated EGF (WT and T/A mutated) as an acceptor substrate and GDP-azido-fucose as a donor substrate. After incubation, copper-assisted azide-alkyne cycloaddition (or CuAAC) referred to as click chemistry was performed to link covalently biotin-alkyne to azido-fucose, specifically transferred by the enzyme (Figure 6A).

The *in vitro* ability of recombinant *MmHisV5Po* from CHO cells, to specifically transfer *O*-fucose to purified isolated EGFs was previously shown for mouse NOTCH1 EGF26 (Pennarubia, Pinault et al. 2018), mouse WIF1 EGF3 (Pennarubia, Pinault et al. 2020) and the unique EGF of mouse PAMR1 (Pennarubia, Germot et al. 2021). New experiments were thus performed with same purified EGFs produced in *E. Coli* and the recombinant *SfHisV5Po* purified in this study from secretome of baculovirus-infected Sf9 insect cells (Figure 6B). Before performing *in vitro* *O*-fucosylation experiments, purified WT and mutated EGFs were analyzed on Coomassie blue-stained polyacrylamide gels (Figure 6B, upper panels). Equal quantity and equivalent purity were seen between WT and T/A mutated EGFs for each murine glycoprotein. After 20 hours incubation of WT or T/A mutated EGFs for each glycoprotein of interest with purified *SfHisV5Po* and GDP-azido-fucose, click chemistry was performed followed by SDS-PAGE and blotting technique using streptavidin-HRP to reveal biotinylation of transferred *O*-fucose. Strong specific signals at expected sizes were obtained for the WT EGF of mouse NOTCH1 and WIF1 but not for mutated EGFs counterparts (Figure 6B, middle panels). This demonstrates that azido-fucose was specifically and efficiently *in vitro* transferred to mouse NOTCH1 EGF26 and WIF1 EGF3 in our experimental conditions. For PAMR1 EGF, no clear specific signal was obtained for its WT EGF even after long time exposure to chemiluminescent substrate indicating that this EGF was not recognized by recombinant *SfHisV5Po*. Western blot using the anti-V5 antibody, recognizing the transferred *SfHisV5Po* used in *in vitro* *O*-fucosyltransferase assays, revealed that similar quantities were indeed used for incubation with WT and T/A mutated EGFs of each mouse glycoprotein (Figure 6B, bottom panels).

To check that preparations of purified EGFs were similar to those used in our previous studies, *O*-fucosyltransferase assays were repeated by incubating newly purified *MmPOFUT1* with the same EGFs preparations as those used for *SfHisV5Po* over 20 h incubation time (Figure 6C). As expected

according to our previous studies, recombinant *MmHisV5Po* specifically and efficiently transferred *O*-fucose to mouse NOTCH1 EGF26, WIF1 EGF3 and even to mouse PAMR1 EGF. However, only the single EGF of mouse PAMR1 remained unmodified after only 1h incubation time with *MmHisV5Po*, showing much less efficiency of fucose transfer to this EGF compared to mouse NOTCH1 EGF26 and WIF1 EGF3 (Supplemental Figure 2). These results thus highlighted different efficiencies of both recombinant enzymes towards PAMR1 EGF, probably due to some structural differences between mouse and lepidopteran POFUT1 that affect EGF binding and/or fucose transfer mechanism. Moreover, contrary to NOTCH1 and WIF1 where orthologous sequences were found in Protostomes, no ortholog was identified for PAMR1.

Finally, we can also notice that 1h incubation time was not sufficient to reveal the *O*-fucosyltransferase activity of *SfHisV5Po*, highlighting a less efficiency compared to *MmHisV5Po* (Supplemental Figure 2).

DISCUSSION

In this paper, we first focused on the characterization of *Spodoptera frugiperda* POFUT1 and its recombinant counterpart expressed in baculovirus-infected *Sf9* insect cells. Similarly to *CePOFUT1* (Lira-Navarrete, Valero-Gonzalez et al. 2011), *SfPOFUT1* is significantly divergent from its human and murine counterparts (43.63 % and 42.16 %, respectively) but the regions and key amino acid residues required for binding of the donor substrate GDP-fucose and the EGFs-containing glycoproteins, are highly conserved. This is also the case of the *N*-glycosylation site $\underline{N}^{53}\text{RT}$ in *SfPOFUT1*, conserved in animals. The $\underline{M}\text{RT}$ site modified by *N*-glycan had been shown to be essential for the activity of bovine POFUT1 (Loriol, Audfray et al. 2007). However, *SfPOFUT1* as other lepidopteran counterparts, exhibited a second potential *N*-glycosylation site $\underline{N}\text{MS}$ at position 219, which was not conserved elsewhere (except for one convergence in Coleoptera). The presence of *N*-glycan at the same face as the active site of the enzyme, might affect overall conformation and/or activity of POFUT1, but this is not the case for recombinant *SfHisV5Po* produced in secretome of baculovirus-infected *Sf9* cells. Indeed, we showed that this site, was not actually occupied by an *N*-glycan.

We then focused on the ability of the *Sf9* insect cells to modify recombinant glycoproteins with *O*-fucose. Recombinant human HisWIF1 was shown to be fully modified on its EGF3 by endogenous POFUT1 expressed by *Sf9* cells as shown by MRM-MS analyses. Thus, although we cannot compare quantitatively our previous results (Pennarubia, Pinault et al. 2020) to this present study due to the fact that the POFUT1 target protein (mouse WIF1) and the detected peptide were different (FNGGT²⁵⁵C³ vs FNGGT²⁵⁵C³FYPGKC⁴), both eukaryotic expression systems based on the use of CHO mammalian cells and *Sf9* insect cells showed high efficiency in endogenously modifying an *O*-fucosylable protein, namely WIF1. Consistently with this result, we demonstrated that recombinant purified *SfHisV5Po* was able to add *in vitro* *O*-fucose to isolated EGF3 of mouse WIF1 and even to mouse NOTCH1 EGF26, but with a lower efficiency than recombinant *MmHisV5Po* from CHO cells. Surprisingly, only PAMR1 EGF did not appear to be modified by *SfHisV5Po* contrary to *MmHisV5Po*. All these results highlight the limitations of the baculovirus insect cell system for the study of *O*-fucosylation of recombinant mammalian glycoproteins.

Interestingly, the reduction of the *O*-fucosylation time by *MmHisV5Po* to 1h removed all signal for the EGF of PAMR1 (but not of other EGFs). This shows a lower affinity of *MmHisV5Po* to the EGF of PAMR1 compared to *SfHisV5Po*. Due to the low efficiency of *SfHisV5Po*, it seems unable to modify the latter unique EGF. The low affinity of these two Pofut1 to EGF of PAMR1 could be due to the presence of an aspartate residue (D) at the position C²⁺³ in the *O*-fucosylation consensus sequence (C²FHDGTC³) which can lead to steric clash as it has been shown for the EGF12 of NOTCH1 (C²QNDATC³) (Li, Han et al. 2017). Differences in POFUT1 *in vitro* *O*-fucosylation capacity were seen for *SfHisV5Po* according to the isolated EGF of interest, the same might also be observed *in cellulo* with different full-length glycoproteins. Unfortunately, it was not possible to purify full-length human recombinant PAMR1 expressed in the baculovirus-insect cell system due to its protein instability and a strong propensity for degradation (data not shown).

To conclude, the baculovirus-insect cell system may offer a good alternative to mammalian expression systems to produce POFUT1-targeted glycoproteins but the ability of insect cells to add *O*-fucose to recombinant proteins may depend on the nature of protein of interest. In addition to this ability of *O*-fucosylation, *Sf9* insect cells might be able to extend *O*-fucoses with GlcNAc residues through

expression of lunatic Fringe, as the case in baculovirus-infected Hi-Five insect cells to *O*-fucose of EGF12 of recombinant NOTCH1 (EGF11-13)(Luca, Jude et al. 2015). In this latter study, it was additionally demonstrated an efficient POGUT1s-mediated *O*-glucosylation, making the baculovirus expression system an efficient tool to study any EGF-containing glycoproteins.

MATERIALS AND METHODS

POFUT1 alignments and conserved potential N-glycosylation Sites. POFUT1 orthologs were retrieved from GenBank (<https://www.ncbi.nlm.nih.gov>) database using on-site tblastn and blastp facilities, with *Spodoptera frugiperda* sequence (A0A2H1WR48 in UniProt) as query. Only complete or almost complete protein sequences from species belonging to Protostomes were selected. 348 homologous POFUT1 sequences covering a large taxonomic diversity (supplemental Table I) were aligned with MUSCLE (Edgar 2004) implemented in SeaView v.4 (Gouy, Guindon et al. 2010). Multiple alignment portions, including conserved potential *N*-glycosylation sites, extracted from the entire dataset or only from Lepidoptera sequences were used to obtain logos at <https://weblogo.threeplusone.com/> (Schneider and Stephens 1990, Crooks, Hon et al. 2004). Counting number of potential *N*-glycosylation sites was performed by the search in the 348 POFUT1 sequences of the canonical motif $\underline{N}X[S/T]$ with X any amino acid except proline. Potential sites included in the signal peptide sequence predicted by SignalP 6.0 (<https://services.healthtech.dtu.dk/service.php?SignalP>) were excluded. Focuses on POFUT1 sequence conservation for a selection of species were achieved using the Multalin server (<http://multalin.toulouse.inra.fr>) (Corpet 1988).

Automatic molecular modelling. Docking experiments were made using the COACH-D server (<https://yanglab.nankai.edu.cn/COACH-D/>), based on the protein sequence of *Spodoptera frugiperda* POFUT1 and GDP-fucose as a donor substrate (in Sdf format file). The proposed model for *Sf*POFUT1 interacting with GDP-fucose as a substrate was selected according the best C-score, in PoseU, and visualized using UCSF CHIMERA (<http://www.rbvi.ucsf.edu/chimera>)(Pettersen, Goddard et al. 2004).

Construction of the baculovirus transfer vector pGmAc115T-SP-His-V5. The pGmAc115T baculovirus transfer vector contains the very late promoter and polyadenylation signal of the polyhedrin

protein (PH). A derived vector named pGmAc115T-SP-His was previously obtained by insertion of additional DNA sequence encoding the signal peptide (SP) of insect cell UDP-glucosyltransferase gene, six histidine residues (His x 6) and the new cloning site *Xba* I upstream of the cloning restriction site Asp718 I (Legardinier, Klett et al. 2005). Using the same strategy of prehybridized (A–F) overlapping primers as before, a new DNA sequence encoding V5 epitope and new cloning sites were inserted between *Xba*I and Asp718 I of pGmAc115T-SP-His to generate the new baculovirus transfer vector referred to as pGmAc115T-SP-His-V5 (Supplemental Figure 3).

Cloning of *Spodoptera frugiperda* POFUT1 and human WIF1 into baculovirus transfer vectors.

*Sf*POFUT1 (residues 20-374) cDNA, obtained from *Sf*9 insect cells and devoid of sequences encoding its endogenous signal peptide and KDEL-like motif, was cloned after PCR amplifications and sub-cloning into pGEM-T easy (Promega Corp., Madison, WI, USA) in the baculovirus transfer vector pGmAc115T-SP-His-V5 (Supplemental Figure 4). Human WIF1 (NP_009122.2, residues 29-379) cDNA from HEK was also cloned after PCR amplifications and sub-cloning into pGEM-T easy but in the previously obtained baculovirus transfer vector pGmAc115T-SP-His (Supplemental Figure 4).

Cells and viral infection. *Sf*9 insect cells were maintained at 28°C in supplemented TC-100 growth medium (Sigma-Aldrich, Saint Louis, MO, USA) containing 5% heat-inactivated fetal bovine serum, 85 IU/mL penicillin and 50 µg/mL streptomycin. *Sf*9 cells were passed when confluency was almost reached and were subcultured three times a week. Baculoviruses were propagated in *Sf*9 insect cells and recovered as previously done (Legardinier, Duonor-Cerutti et al. 2005). To prepare viral inoculi, adherent *Sf*9 cells were infected with recombinant baculoviruses at a multiplicity of infection (MOI) ranging from 5 to 10 plaque forming units (Pfu) per cell. After 45 minutes incubation with viral particles, new culture medium was added and cells were incubated at 28°C until 5 days post-infection. The viral titers were determined by plaque assay.

Cotransfection and purification of the recombinant baculoviruses. *Sf*9 cells were cotransfected with 5 µg of transfer vector and 1 µg of purified viral DNA using a lipofection method (XtremeGene HP, Sigma-Aldrich, Saint Louis, MO, USA). Viral DNA extracted from the wild-type *Autographa californica* multiple nucleocapsid polyhedrosis virus (*AcMNPV*) was thus cotransfected with polyhedrin specific baculovirus transfer vectors bearing foreign genes such as pGmAc115T-SP-His[human WIF1]

or the construct pGmAc115T-SP-HisV5[SfPOFUT1]. In each case, recombinant baculoviruses were selected by plaque assay and distinguished from the wild-type progeny by their occlusion body-negative phenotype. The screening and purification of the recombinant baculoviruses NPV-SfHisV5Po and NPV-human HisWIF1 were carried out as previously described (Legardinier, Duonor-Cerutti et al. 2005).

Production and nickel affinity purification of recombinant SfHisV5Po and human HisWIF1. Sf9 were seeded at a density of $1 \cdot 10^6$ cells/mL in 500 ml ErlenMeyer and grown at 28°C at a speed of 80 rpm during 3 or 4 days in 50 mL of serum-free medium EX-CELL 420 (ref 14420C, Sigma-Aldrich, Saint Louis, MO, USA). When cell density reached 6 or $7 \cdot 10^6$ cells/mL, cells were infected with the recombinant baculovirus NPV-SfHisV5Po or NPV-human HisWIF1. Two days post-infection, supernatants were recovered by centrifugation at 100 g for 5 min, then 1000 g for 10 min and precipitated in 80% ammonium sulfate at 4°C. After centrifugation at 10000 g for 15 min at 4°C, pellets were resuspended in binding buffer (25 mM Tris-HCl, 500 mM NaCl, 5 mM CaCl₂, 20 mM imidazole, pH 7.5) and purified on Ni-NTA column by imidazole gradient (from 20 to 500 mM imidazole) using AKTA prime system (GE Healthcare, Piscataway, NJ, USA). Recombinant purified proteins of elution fractions were then concentrated with Amicon ultra centrifugal filters 3K in a Tris-CaCl₂ buffer (25 mM Tris, 5 mM CaCl₂, pH 7.5) and quantified using a bicinchoninic acid (BCA) protein assay (Sigma-Aldrich, Saint Louis, MO, USA) with bovine serum albumin as a standard.

Production and nickel affinity purification of recombinant MmHisV5Po and recombinant EGFs. Recombinant mouse POFUT1 named MmHisV5Po was produced in stable Flp-In CHO cell line as previously described (Pennarubia, Pinault et al. 2018). Recombinant isolated WT and mutated EGFs of mouse NOTCH1, WIF1 and PAMR1 were produced in *E.coli* BL21 strain, as previously described (Pennarubia, Pinault et al. 2018, Pennarubia, Pinault et al. 2020, Pennarubia, Germot et al. 2021). All these recombinant proteins were purified on Ni-NTA column using imidazole gradient, as previously described (Pennarubia, Pinault et al. 2018). The eluted fractions were separated on SDS PAGE and blotted on nitrocellulose membrane (for detection of recombinant proteins by anti-V5-HRP (ThermoFisher scientific, Waltham, MA, USA)) and/or stained by Coomassie blue or silver nitrate. The

purest fractions were collected and concentrated by ultrafiltration in Tris-CaCl₂ buffer. Then, proteins were quantified by BCA protein dosage assay.

SDS-PAGE, gel staining or blotting. Purified proteins were separated on 12 % polyacrylamide gels (enzymes) or 17% polyacrylamide gels (EGFs), stained with Coomassie blue or silver nitrate. After glycosyltransferase assay, CuAAC (see after), SDS-PAGE and blot transfer, nitrocellulose membranes were blocked with 10% fat free milk in TBS (50 mM Tris-HCl, 150 mM NaCl, pH 7.6) - 0.1% Tween 20 (TBST) for 10 min, washed three times for 5 min in TBST and incubated with streptavidin-HRP in TBST at 25 ng/mL overnight at 4°C. After 3 new washes for 5 min per wash, membranes were revealed using enhanced chemiluminescence peroxidase substrate. Signals were visualized and quantified using Amersham Imager 600 (GE Healthcare, Buckinghamshire, UK). For Western blot, the nitrocellulose membrane was blocked with 5% fat free milk – TBST 0.1% for 1 h at room temperature and incubated with anti-V5-HRP (Thermo Fisher Scientific, Waltham, MA, USA) diluted to ratio 1/2000 in 2.5% milk-TBST 0.1% overnight at 4°C. The membrane was washed 3 times, 5 minutes each, with TBST 0.1% before being revealed as mentioned above.

PNGase F deglycosylation experiments and lectin blot. Purified proteins (1 µg used), namely *SfHisV5Po*, *MmHisV5Po* and control glycoproteins provided in DIG Glycan Differentiation Kit (Sigma-Aldrich, Saint Louis, MO, USA), were denatured by incubation with KP 10X buffer (100 mM KP, 100 mM EDTA, 5% triton X-100, 2% SDS, 10% beta-mercaptoethanol, pH 8) at 100°C for 10 min, then at 37°C for 10 min. After that, 0.06 µg of PNGase F was added to the mixture and incubated for 20 h at 37°C. Untreated or PNGase F-treated samples were separated by SDS-PAGE using 15% polyacrylamide gel and blotted on nitrocellulose membrane. For detection of *SfHisV5Po*, GNA lectin from commercial kit, used according to the manufacturer's protocol, was chosen due to its strong reactivity with terminal alpha-linked mannoses of *N*-glycans, usually found in *Sf9* insect cells as previously shown (Legardinier, Klett et al. 2005). Fetuin, bearing sialylated complex-type *N*-glycans, was thus used as a negative glycoprotein control for GNA and carboxypeptidase A, bearing only terminal alpha-linked mannoses, was used as a positive glycoprotein control for GNA.

In vitro O-fucosyltransferase reaction and click chemistry (CuAAC) experiments. *In vitro* O-fucosyltransferase reactions were carried out with 1 µg of pure *MmHisV5Po* or an equivalent quantity of *SfHisV5Po* mixed with 2 nmoles GDP-azido-fucose (R&D Systems Inc., Minneapolis, MN, USA), 3 µg isolated purified WT or T/A mutated EGF in 25 µL reaction buffer (25 mM Tris, 5 mM CaCl₂, 10 mM MnCl₂, pH 7.5) and incubated 1 h or 20 h at 37°C as previously described (Pennarubia, Pinault et al. 2018). Copper-assisted azide–alkyne cycloaddition (CuAAC) also named click chemistry was performed using 1.25 mM CuCl₂, 2.5 mM ascorbic acid and 0.125 mM alkynyl biotin (R&D Systems Inc., Minneapolis, MN, USA), directly added to the O-fucosyltransferase reaction. The mixture was incubated in the dark for 1 h at room temperature and stopped by heating for 5 min in Laemmli buffer (Laemmli 1970).

Protein digestion for mass spectrometry analyses. Human HisWIF1 samples were digested using the FASP method. Briefly, proteins were reduced in DTT 5 mM for 30 min at 56°C and alkylated in Iodoacetamide 10 mM for 30 min in the dark. After reduction/alkylation, proteins were transferred into the Amicon Ultra 10K filter (Millipore) and washed twice by centrifugation with 25 mM ammonium bicarbonate 0.5 mM CaCl₂. Digestion was performed with 1 µg of thermolysin (Sigma-Aldrich, Saint Louis, MO, USA) at 37°C for a minimum of 16 h. *SfHisV5Po* sample was digested in solution. Briefly, proteins were reduced in DTT 5 mM for 30 min at 56°C and alkylated in Iodoacetamide 10 mM for 30 min in the dark. Digestion was performed with 0.5 µg of trypsin (Promega) at 37°C for a minimum of 16 h. For all samples, peptides were cleaned using 1CC 30 mg HLB cartridge following manufacturer's protocols (Waters; Milford, MA), solubilized in loading solvent (water/ACN/TFA 98/2/0.05 (v/v)) at a concentration of 0.2 µg/µL for HisWIF1 and of 0.1 µg/µL for *SfHisV5Po*, and finally filtered on 0.22 µm spin column (Agilent Technologies, Santa Clara, CA).

MicroLC MS/MS analysis of protein tryptic digests. Resulting peptides were analysed by microLC-MS/MS using a nanoLC 425 in micro-flow mode (Eksigent, Dublin, CA) system coupled with a TripleTOF 5600+ (SCIEX, Framingham, MA). 5 µL of each sample were trapped on C18 Pepmap100 cartridge (300 µm Id x 5 mm, 5µm; Thermo Scientific), and desalting was carried out at 10 µL/min with Loading solvent for 5 minutes. The chromatographic separation was performed on a ChromXP C18

column (150 x 0.3 mm i.d., 120Å, 3 µm; SCIEX) at a flow rate of 3 µL/min. The mobile phase was a gradient of water/acetonitrile/formic acid 100/0/0.1% (A) and 5/95/0.1% (B) programmed as follows: initial, 5% B, increased to 25 % in 20 min, then increased to 95 % B in 2 min, maintained at 95 % for 4 min, and finally, decreased to 5 % B for reequilibration.

The TTOF 5600+ was operated in data-dependant acquisition (DDA) mode with Analyst 1.7TF software (SCIEX). MS and MS/MS data were continuously recorded with 1.3 s cycle time with up to 20 precursors selected for fragmentation from each MS survey scan. Precursor selection was based upon ion intensity and whether or not the precursor has been previously selected for fragmentation (dynamic exclusion). Collision energies were automatically adjusted to the charge state and m/z value of the precursor ions.

Data processing. All DDA mass spectrometer files were analyzed by Protein Pilot software (version 5.0, SCIEX) to search in the recombinant protein sequence database via the Mascot software (version 2.2, Matrix Science, UK) with the following criteria: 25 ppm and 0.05 Da tolerance respectively for precursor and fragment masses, a single or 5 missed cleavage site allowed during trypsin and thermolysin digestion respectively, and carbamidomethylation of cysteine residues as fix modification and oxidation of methionine as variable modification. Protein identification was manually validated.

Creation of the targeted method for human HisWIF1. The data were acquired in high-Resolution MRM mode: product ion scans were collected for the m/z corresponding to *O*-fucosylated or non-modified peptides during 30 min using the same parameters as previously described in the DDA method. Data were processed with MultiQuant Software 3.0.1 (SCIEX), considering the 6 most abundant fragments for each peptide with a resolution of 10000 (Supplemental Table II). When present, fragments corresponding to the peptide without the sugar moiety were added.

Funding

The work was partly funded by the project “Ligue contre le cancer 2021”.

Acknowledgements

We would like to thank Mrs Nelly Vallat for her technical help as well as Flavie Correia for contribution to experiments during her internship. We also gratefully thank Pr Martine Duonor-Cérutti for help by providing tools and protocols required for the use of non-commercial baculovirus insect cell system.

Conflict of interest statement

The authors declare no competing interests

Abbreviations

CHO, Chinese Hamster Ovary cells; CuAAC, Copper-catalyzed Azide-Alkyne Cycloaddition; EGF, Epidermal Growth Factor-Like Domain; ER, endoplasmic reticulum; Fuc, Fucose; Glc, Glucose; GlcNAc, *N*-acetylglucosamine; *Mm*, *Mus musculus*; MRM-MS, Multiple Reaction Monitoring-Mass Spectrometry; PAMR1, Peptidase domain containing Associated Muscle Regeneration 1; POFUT1, Protein *O*-fucosyltransferase 1; *Sf*, *Spodoptera frugiperda*; WIF1, Wnt Inhibitory Factor 1.

REFERENCES

- Altmann, F., E. Staudacher, I. B. Wilson and L. Marz (1999). "Insect cells as hosts for the expression of recombinant glycoproteins." *Glycoconj J* **16**(2): 109-123.
- Brockhausen, I., H. H. Wandall, K. G. T. Hagen and P. Stanley (2022). O-GalNAc Glycans. *Essentials of Glycobiology*. th, A. Varki, R. D. Cummings et al. Cold Spring Harbor (NY): 117-128.
- Corpet, F. (1988). "Multiple sequence alignment with hierarchical clustering." *Nucleic Acids Res* **16**(22): 10881-10890.
- Crooks, G. E., G. Hon, J. M. Chandonia and S. E. Brenner (2004). "WebLogo: a sequence logo generator." *Genome Res* **14**(6): 1188-1190.
- Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." *Nucleic Acids Res* **32**(5): 1792-1797.
- Gouy, M., S. Guindon and O. Gascuel (2010). "SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building." *Mol Biol Evol* **27**(2): 221-224.
- Harvey, B. M., N. A. Rana, H. Moss, J. Leonardi, H. Jafar-Nejad and R. S. Haltiwanger (2016). "Mapping Sites of O-Glycosylation and Fringe Elongation on Drosophila Notch." *J Biol Chem* **291**(31): 16348-16360.
- Jarvis, D. L. (2009). "Baculovirus-insect cell expression systems." *Methods Enzymol* **463**: 191-222.
- Jarvis, D. L. and E. E. Finn (1995). "Biochemical analysis of the N-glycosylation pathway in baculovirus-infected lepidopteran insect cells." *Virology* **212**(2): 500-511.
- Juliant, S., A. Harduin-Lepers, F. Monjaret, B. Catieau, M. L. Violet, P. Cerutti, A. Ozil and M. Duonor-Cerutti (2014). "The alpha1,6-fucosyltransferase gene (*fut8*) from the *Sf9* lepidopteran insect cell line: insights into *fut8* evolution." *PLoS One* **9**(10): e110422.
- Kubelka, V., F. Altmann, G. Kornfeld and L. Marz (1994). "Structures of the N-linked oligosaccharides of the membrane glycoproteins from three lepidopteran cell lines (*Sf*-21, IZD-Mb-0503, Bm-N)." *Arch Biochem Biophys* **308**(1): 148-157.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." *Nature* **227**(5259): 680-685.

Legardinier, S., M. Duonor-Cerutti, G. Devauchelle, Y. Combarnous and C. Cahoreau (2005). "Biological activities of recombinant equine luteinizing hormone/chorionic gonadotropin (eLH/CG) expressed in Sf9 and Mimic insect cell lines." *J Mol Endocrinol* **34**(1): 47-60.

Legardinier, S., D. Klett, J. C. Poirier, Y. Combarnous and C. Cahoreau (2005). "Mammalian-like nonsialyl complex-type N-glycosylation of equine gonadotropins in Mimic insect cells." *Glycobiology* **15**(8): 776-790.

Li, Z., K. Han, J. E. Pak, M. Satkunarajah, D. Zhou and J. M. Rini (2017). "Recognition of EGF-like domains by the Notch-modifying O-fucosyltransferase POFUT1." *Nat Chem Biol* **13**(7): 757-763.

Lira-Navarrete, E., J. Valero-Gonzalez, R. Villanueva, M. Martinez-Julvez, T. Tejero, P. Merino, S. Panjekar and R. Hurtado-Guerrero (2011). "Structural insights into the mechanism of protein O-fucosylation." *PLoS One* **6**(9): e25365.

Lo, P. H., C. Tanikawa, T. Katagiri, Y. Nakamura and K. Matsuda (2015). "Identification of novel epigenetically inactivated gene PAMR1 in breast carcinoma." *Oncol Rep* **33**(1): 267-273.

Lopez, M., D. Tetaert, S. Juliant, M. Gazon, M. Cerutti, A. Verbert and P. Delannoy (1999). "O-glycosylation potential of lepidopteran insect cell lines." *Biochim Biophys Acta* **1427**(1): 49-61.

Loriol, C., A. Audfray, F. Dupuy, A. Germot and A. Maftah (2007). "The two N-glycans present on bovine Pofut1 are differently involved in its solubility and activity." *FEBS J* **274**(5): 1202-1211.

Luca, V. C., K. M. Jude, N. W. Pierce, M. V. Nachury, S. Fischer and K. C. Garcia (2015). "Structural biology. Structural basis for Notch1 engagement of Delta-like 4." *Science* **347**(6224): 847-853.

Luther, K. B. and R. S. Haltiwanger (2009). "Role of unusual O-glycans in intercellular signaling." *Int J Biochem Cell Biol* **41**(5): 1011-1024.

McMillan, B. J., B. Zimmerman, E. D. Egan, M. Lofgren, X. Xu, A. Hesser and S. C. Blacklow (2017). "Structure of human POFUT1, its requirement in ligand-independent oncogenic Notch signaling, and functional effects of Dowling-Degos mutations." *Glycobiology* **27**(8): 777-786.

Nakamura, N., K. Katano, S. Toba and A. Kurosaka (2004). "Characterization of a novel polypeptide N-acetylgalactosaminyltransferase (dGalNAc-T3) from Drosophila." *Biol Pharm Bull* **27**(10): 1509-1514.

Okajima, T., A. Xu and K. D. Irvine (2003). "Modulation of notch-ligand binding by protein O-fucosyltransferase 1 and fringe." *J Biol Chem* **278**(43): 42340-42345.

Pennarubia, F., A. Germot, E. Pinault, A. Maftah and S. Legardinier (2021). "The single EGF-like domain of mouse PAMR1 is modified by O-Glucose, O-Fucose and O-GlcNAc." *Glycobiology* **31**(1): 55-68.

Pennarubia, F., E. Pinault, B. Al Jaam, C. E. Brun, A. Maftah, A. Germot and S. Legardinier (2020). "Mouse WIF1 Is Only Modified with O-Fucose in Its EGF-like Domain III Despite Two Evolutionarily Conserved Consensus Sites." *Biomolecules* **10**(9).

Pennarubia, F., E. Pinault, A. Maftah and S. Legardinier (2018). "In vitro acellular method to reveal O-fucosylation on EGF-like domains." *Glycobiology*.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." *J Comput Chem* **25**(13): 1605-1612.

Schneider, T. D. and R. M. Stephens (1990). "Sequence logos: a new way to display consensus sequences." *Nucleic Acids Res* **18**(20): 6097-6100.

Schwientek, T., U. Mandel, U. Roth, S. Muller and F. G. Hanisch (2007). "A serial lectin approach to the mucin-type O-glycoproteome of Drosophila melanogaster S2 cells." *Proteomics* **7**(18): 3264-3277.

Shao, L., D. J. Moloney and R. Haltiwanger (2003). "Fringe modifies O-fucose on mouse Notch1 at epidermal growth factor-like repeats within the ligand-binding site and the Abruptex region." *J Biol Chem* **278**(10): 7775-7782.

Staudacher, E., V. Kubelka and L. Marz (1992). "Distinct N-glycan fucosylation potentials of three lepidopteran cell lines." *Eur J Biochem* **207**(3): 987-993.

Takeuchi, H., M. Schneider, D. B. Williamson, A. Ito, M. Takeuchi, P. A. Handford and R. S. Haltiwanger (2018). "Two novel protein O-glucosyltransferases that modify sites distinct from POGlut1 and affect Notch trafficking and signaling." *Proc Natl Acad Sci U S A* **115**(36): E8395-E8402.

Tomiya, N., S. Narang, Y. C. Lee and M. J. Betenbaugh (2004). "Comparing N-glycan processing in mammalian cell lines to native and engineered lepidopteran insect cell lines." *Glycoconj J* **21**(6): 343-360.

Wathen, M. W., P. A. Aeed and A. P. Elhammer (1991). "Characterization of oligosaccharide structures on a chimeric respiratory syncytial virus protein expressed in insect cell line Sf9." *Biochemistry* **30**(11): 2863-2868.

Zhang, L. and K. G. Ten Hagen (2019). "O-Linked glycosylation in *Drosophila melanogaster*." *Curr Opin Struct Biol* **56**: 139-145.

LEGENDS TO FIGURES

Figure 1: Alignment of SfPOFUT1 protein sequence with those of crystallized POFUT1 enzymes.

Using Multalin server, a multiple sequence alignment for POFUT1 was obtained for the four species *Homo sapiens* (Hs), *Mus musculus* (Mm), *Caenorhabditis elegans* (Ce) and *Spodoptera frugiperda* (Sf).

The highly conserved N-glycosylation site \underline{N}^{67} RT in MmPOFUT1 was also found in SfPOFUT1. The second N-glycosylation site \underline{N}^{165} KS of MmPOFUT1, conserved in mammals, was missing in SfPOFUT1 and CePOFUT1. Surprisingly, SfPOFUT1 exhibited a second N-glycosylation site at position 219 (\underline{N} MS). Key residues of mouse POFUT1 (M⁴⁶, G⁴⁷, N⁵¹ and N¹⁵¹), located in highly conserved regions (in red) and known to interact with EGFs, were also found in SfPOFUT1. Other residues such as Y⁷⁸, known to induce a steric clash, are also point out. The conserved regions for binding to GDP-fucose are indicated with black arrows as well as the signal peptide and the KDEL-like ER (RDEF or HEEL or HIDL) retention signal. The first residue of mature POFUT1 is framed in pink for each species.

Figure 2: Conservation and proportions of N-glycosylation sites in Protostome POFUT1 sequences.

(A) The upper inset corresponds to Logos for aligned POFUT1 orthologous sequences in Protostomes, from position 50 to 59 of *Spodoptera frugiperda* sequence, where the conserved N-glycosylation site was found (\underline{N}^{53}). The height of the letters represents the amino acid relative frequency at each position. The graphical representation was obtained by comparison of 348 sequences using the web-based application WebLogo 3.7.12. The lower inset corresponds to Logos for aligned POFUT1 orthologous sequences in Lepidoptera, from position 216 to 225 of *Spodoptera frugiperda* sequence, where the second potential N-glycosylation site is found (\underline{N}^{219}). 19 sequences over 34 have the conserved NMS glycosite. The graphical representation was obtained by comparison of 34

sequences. **(B)** Proportions of potential *N*-glycosylation sites in Protostome POFUT1 sequences. The categories are distinguished according to the absence or the presence of 1 to 5 potential *N*-glycosites, which match the canonical motif $\underline{N}X[S/T]$, with X any amino acid except proline. The curve represents the percentages of sequences where the $\underline{N}RT$ site is evolutionary conserved. Numbers between brackets correspond to numbers of considered POFUT1 sequences for each taxonomic group. Ectoprocta, Neuroptera and Priapula, taxa only represented by one species each and without the $\underline{N}RT$ site, are not included. For more details, refer to supplemental table I.

Figure 3: Tridimensional structures of POFUT1. **(A)** X-ray structure of mouse POFUT1 with GDP-fucose (PDB 5KY3). The co-crystallized mouse Factor VII EGF1 mutant was here deleted. The conserved residues known to be involved, according to Li et al. (2017) in binding to GDP-fucose (in black) are shown in orange and those involved in binding to acceptor substrates, namely EGFs comprising a potential *O*-fucosylation site, are shown in red. The two occupied *N*-glycosylation sites $\underline{N}^{67}RT$ and $\underline{N}^{165}KS$ still bear the residual proximal GlcNAc of the chitobiose core of *N*-glycans. **(B)** Using COACH-D server, automatic molecular docking was performed using *Spodoptera frugiperda* POFUT1 as a template and GDP-fucose (in black) as a donor substrate. The same conserved key residues as for *Mm*POFUT1 were shown for this structural model. The location of the two *N*-glycosylation sites $\underline{N}^{53}RT$ and $\underline{N}^{219}MS$ was indicated with asparagines in blue.

Figure 4: Characterization of *N*-glycosylation of recombinant *Sf*HisV5Po from baculovirus-infected Sf9 cells. **(A)** Coomassie blue staining of polyacrylamide gel showing 1 μ g of purified recombinant *Sf*HisV5Po and *Mm*HisV5Po. **(B)** Western blot revealed by anti-V5-HRP for *Sf*HisV5Po after treatment or not with PNGase F. **(C)** Silver nitrate staining (upper panel) and Lectin Blot with GNA (lower panel) showing *N*-glycan deglycosylation of purified *Sf*HisV5Po, *Mm*HisV5Po from CHO cells and commercial purified glycoproteins, namely Fetuin (F) as negative control and Carboxypeptidase Y (CPY) as positive control, treated with PNGase F compared to non-treated ones. Asterisks indicated the presence of contaminating co-purified *N*-linked protein, well recognized by GNA. **(D)** Trypsin digestion followed by MS-MS analysis of full-length *Sf*HisV5Po.

Figure 5: MRM-MS of thermolysin-digested human HisWIF1, produced by infected Sf9 cells.

(A) The recombinant human WIF1 with its *N*-terminal polyhistidine tag is drawn with its different domains: the WIF domain (WD) and the following five EGF domains. Zoom on the amino acid sequence of its EGF3 including the potential *O*-fucosylation site T²⁵⁵ is boxed and shows the protease cleavage sites by thermolysin. (B) Full-length recombinant human HisWIF1, secreted in culture medium of *Sf9* cells, was purified, reduced, alkylated and finally digested by thermolysin. Resulting peptides were analyzed by micro-LC MRM-MS. The peptide of interest FNGGT²⁵⁵C³FYPGKC⁴ of EGF3 was almost 100 % modified with *O*-fucose

Figure 6: In vitro POFUT1-mediated O-fucosylation of recombinant isolated mouse EGFs by SfHisV5Po and MmHisV5Po, revealed by blotting technique after click chemistry.

(A) Purified isolated WT or T/A mutated at the *O*-fucosylation site were incubated separately with recombinant purified *SfHisV5Po* in the presence of GDP-azido-fucose as a donor substrate. The ability of *SfHisV5Po* to transfer *O*-fucose to EGFs was determined by blotting technique after click chemistry (CuAAC), where biotin-alkyne covalently binds to the azido-fucose if transferred to EGF by the enzyme. (B) Coomassie blue staining showing equivalent quantities (3 µg) of WT or T/A mutated recombinant isolated EGFs for mouse NOTCH1 EGF26, WIF1 EGF3 and PAMR1 EGF (upper panel). After 20 h incubation at 37°C with 1 µg of *SfHisV5Po* in the presence of GDP-azido-fucose, followed by click chemistry, blotting was performed using streptavidin-HRP to reveal the ability of the enzyme to add *O*-fucose to WT EGFs (middle panel). The same membrane was then incubated with anti-V5-HRP to reflect for each sample the quantity of enzyme used to perform *O*-fucosyltransferase assays (lower panel). (C) The same experiments as in (B) was performed with same EGFs preparations for *MmHisV5Po* from CHO cells. After incubation of WT or T/A mutated EGFs with *MmHisV5Po* in presence of GDP-azido-fucose and click chemistry, blotting technique was performed using streptavidin-HRP (upper panel). The same membrane was re-incubated with anti-V5-HRP to reveal the relative quantity of the enzyme used (lower panel).

Figure 1.

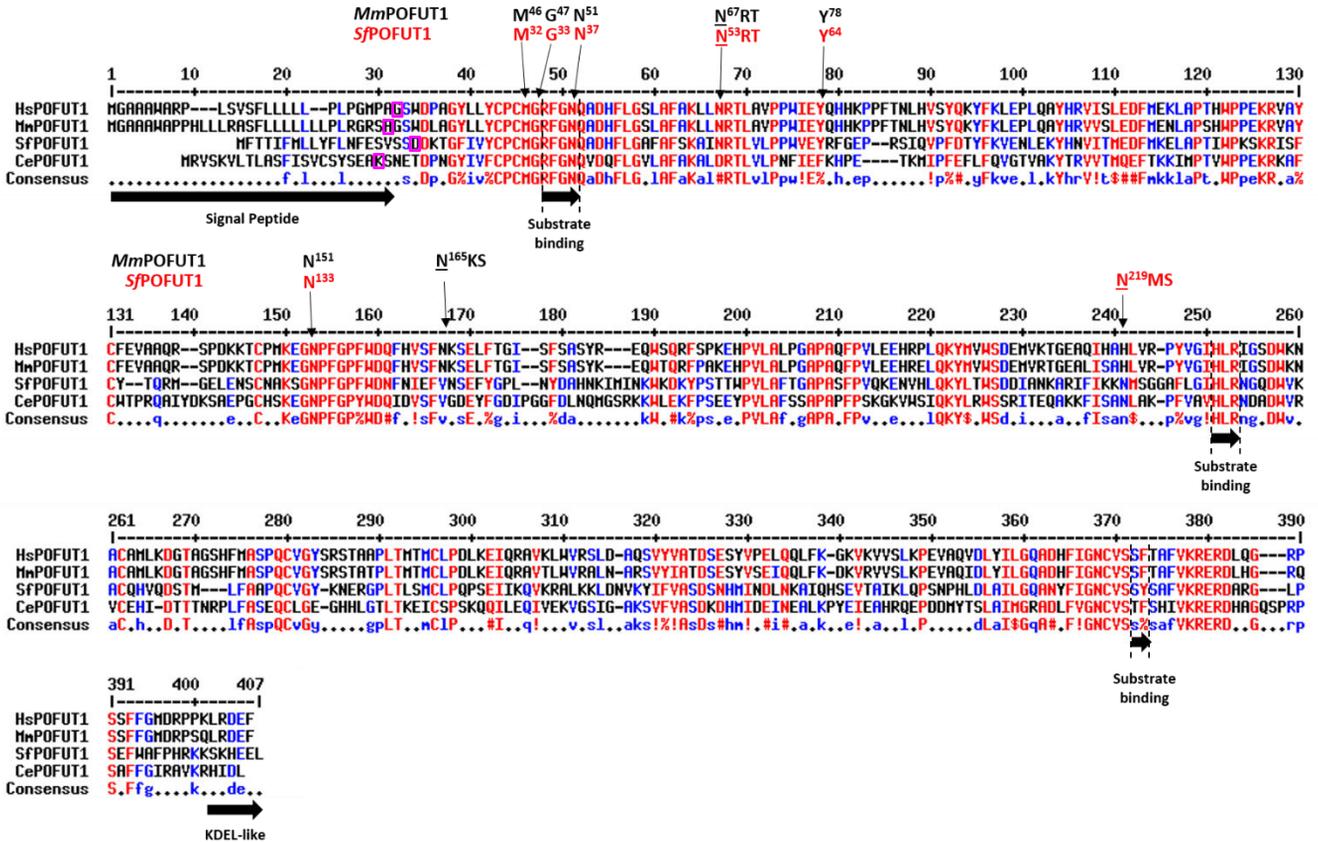
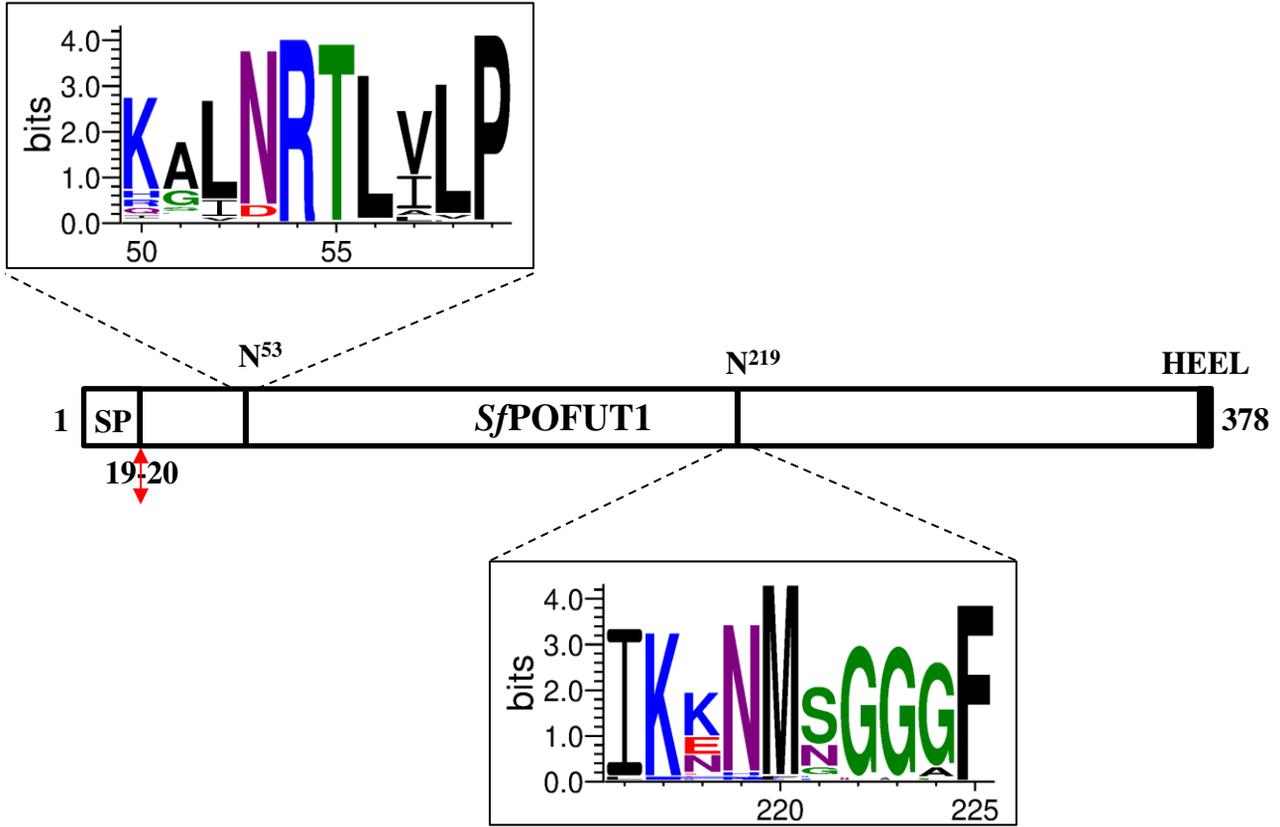


Figure 2.

A.



B.

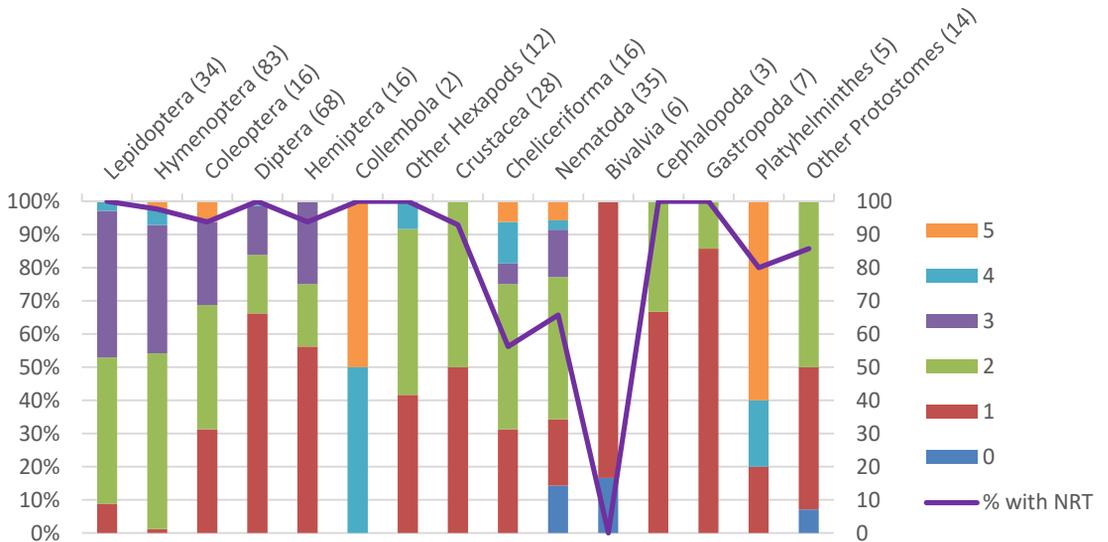
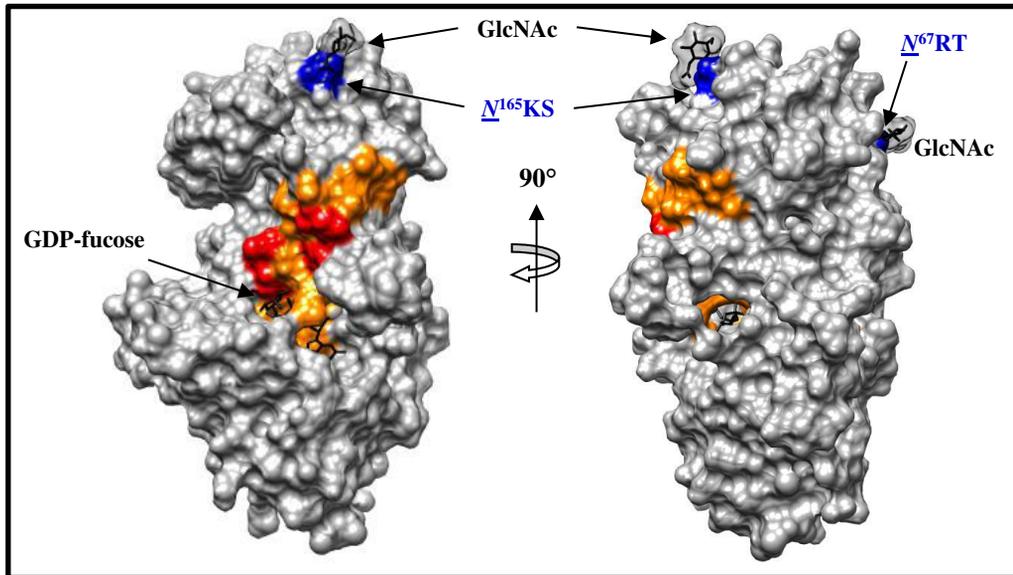


Figure 3.

A.

*Mm*POFUT1 3D structure (5KY3) without EGF



B.

*Sf*POFUT1 3D structural model and docking with GDP-fucose

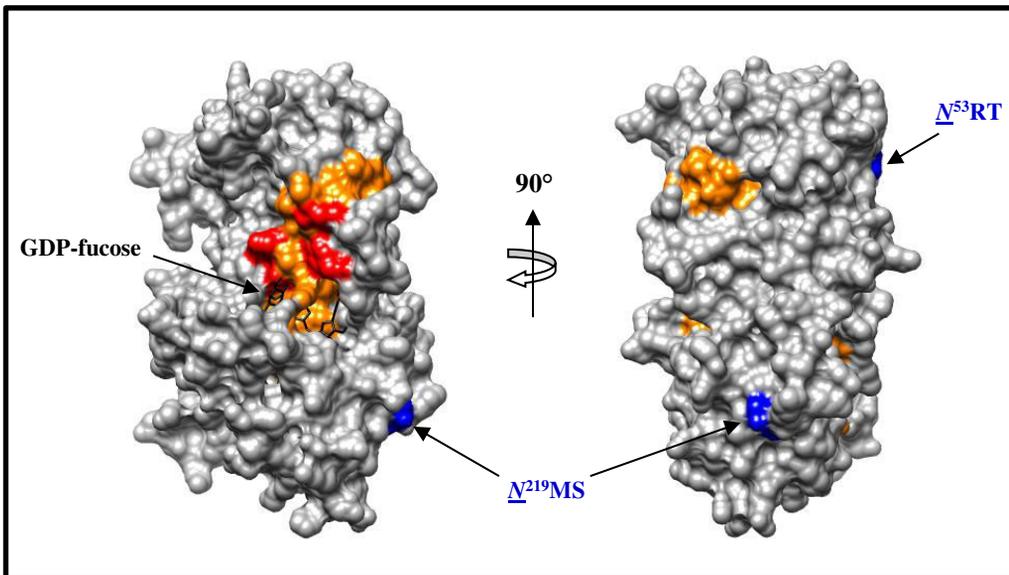
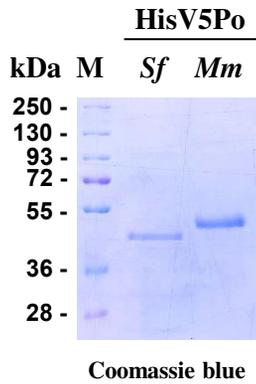
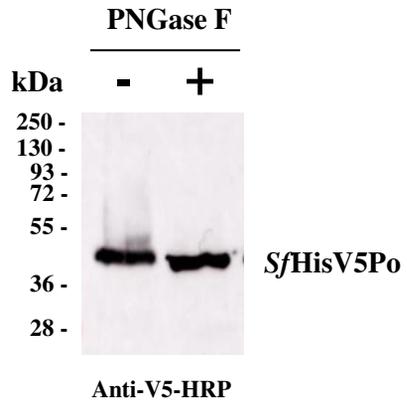


Figure 4.

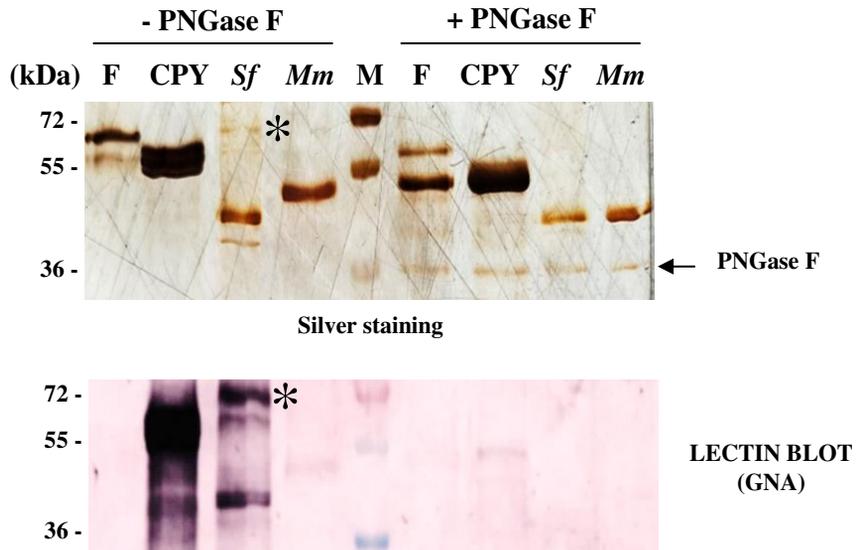
A.



B.



C.

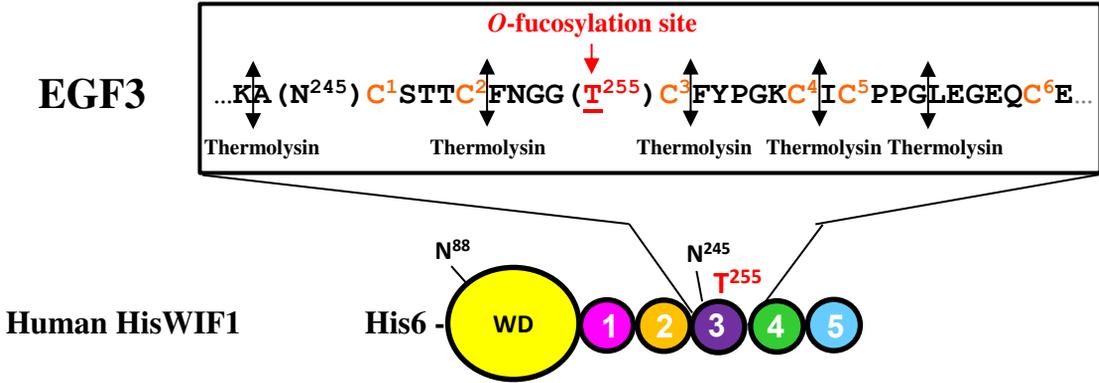


D.

1	MTILCWLALL	STLTAVNAHH	HHHLGGKPT	PNPLLGLDST	GLEDLDDKTG
51	FIVYCPMGR	FGNQADHFLG	AFAFSKAI	NR T	VLPPWVEY
101	PFDTYFKVEN	LEKYHNVITM	EDFMKELAPT	IWPKSKRISF	CYTQRMGELE
151	NSCNAKSGNP	FGPFWDNFNI	EFVNSEFYGP	LNYDAHNKIM	INKNKDKYPS
201	TTWPVLAFTG	APASFPVQKE	NVHLQKYLTV	SDDIANKARI	FIKHNMSGGA
251	FLGIHLRNGQ	DWVKACQHVQ	DSTMLFAAPQ	CVGYKNERGP	LTLSMCLPQP
301	SEIIKQVKRA	LKKLDNVKYI	FVASDSNHMI	NDLNKAIQHS	EVTAIKLPQS
351	NPHLDLAILG	QANYFIGNCV	SSYSAFVKRE	RDARGLPSEF	WAFPHRKKSK

Figure 5.

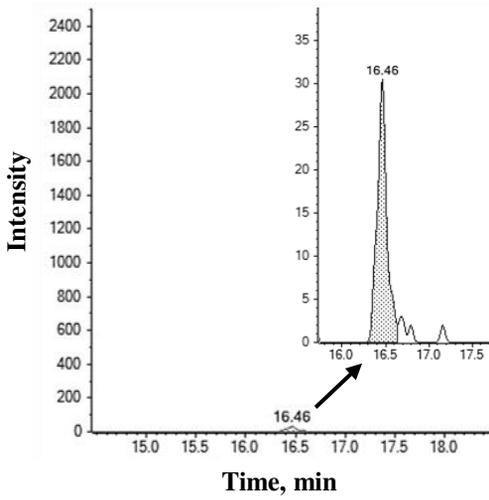
A.



B.

Non-modified EGF3 peptide from human HisWIF1

FNGG (T²⁵⁵) C³FYPGKC⁴



Modified EGF3 peptide from human HisWIF1

FNGG (T²⁵⁵) C³FYPGKC⁴ ^{O-fucose}

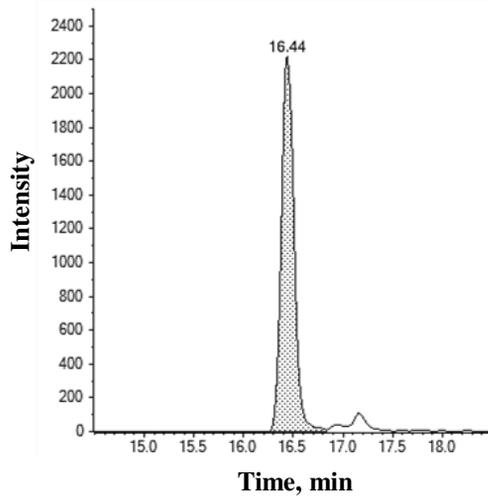
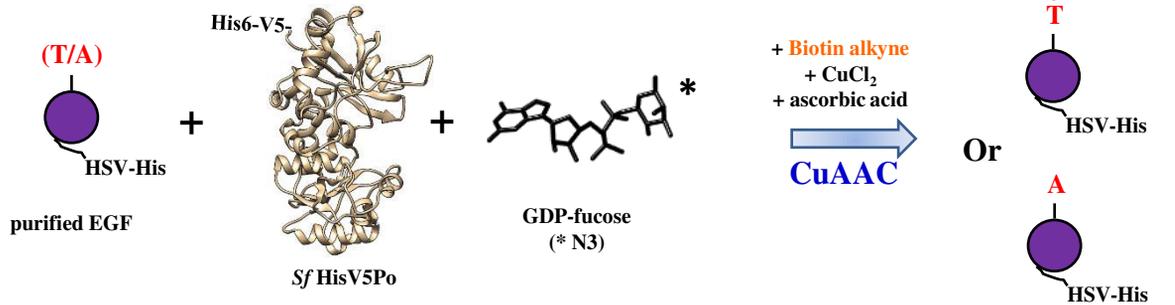


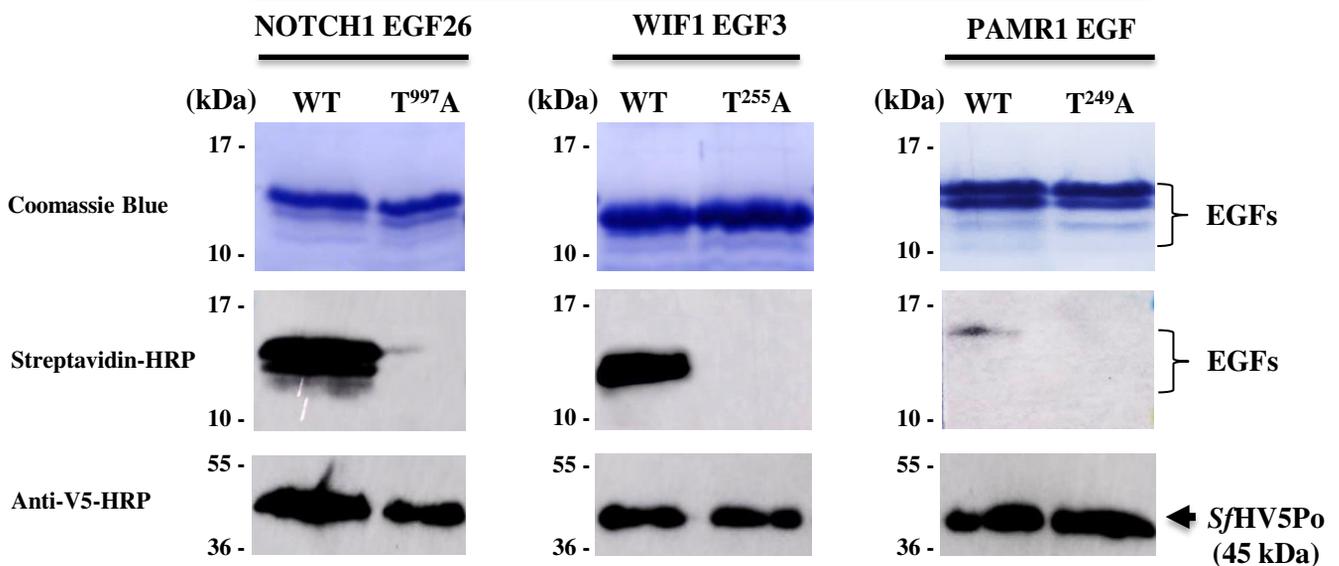
Figure 6.

A.



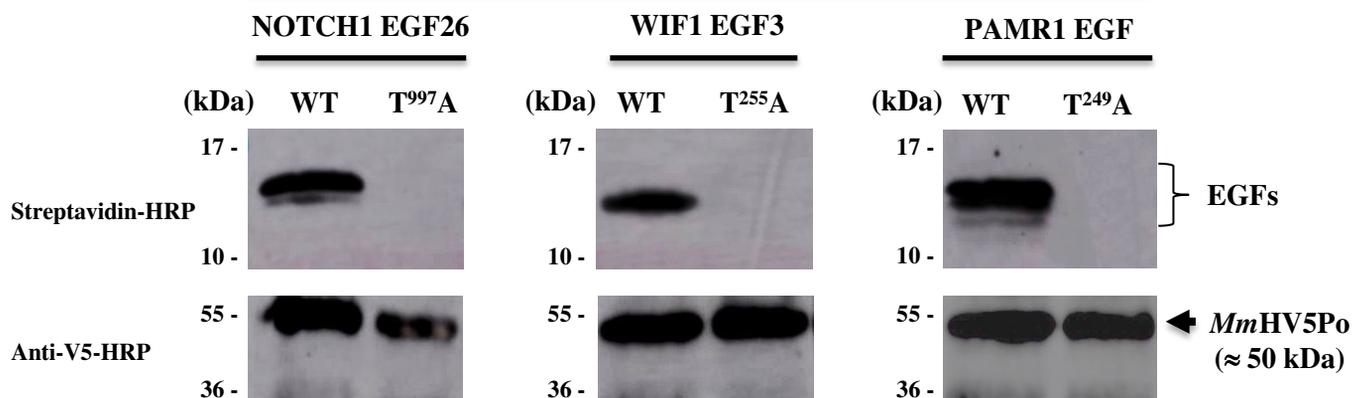
B.

20 h incubation with *Sf*HisV5Po and various acceptor substrates



C.

20 h incubation with *Mm*HisV5Po and various acceptor substrates



SUPPLEMENTARY DATA

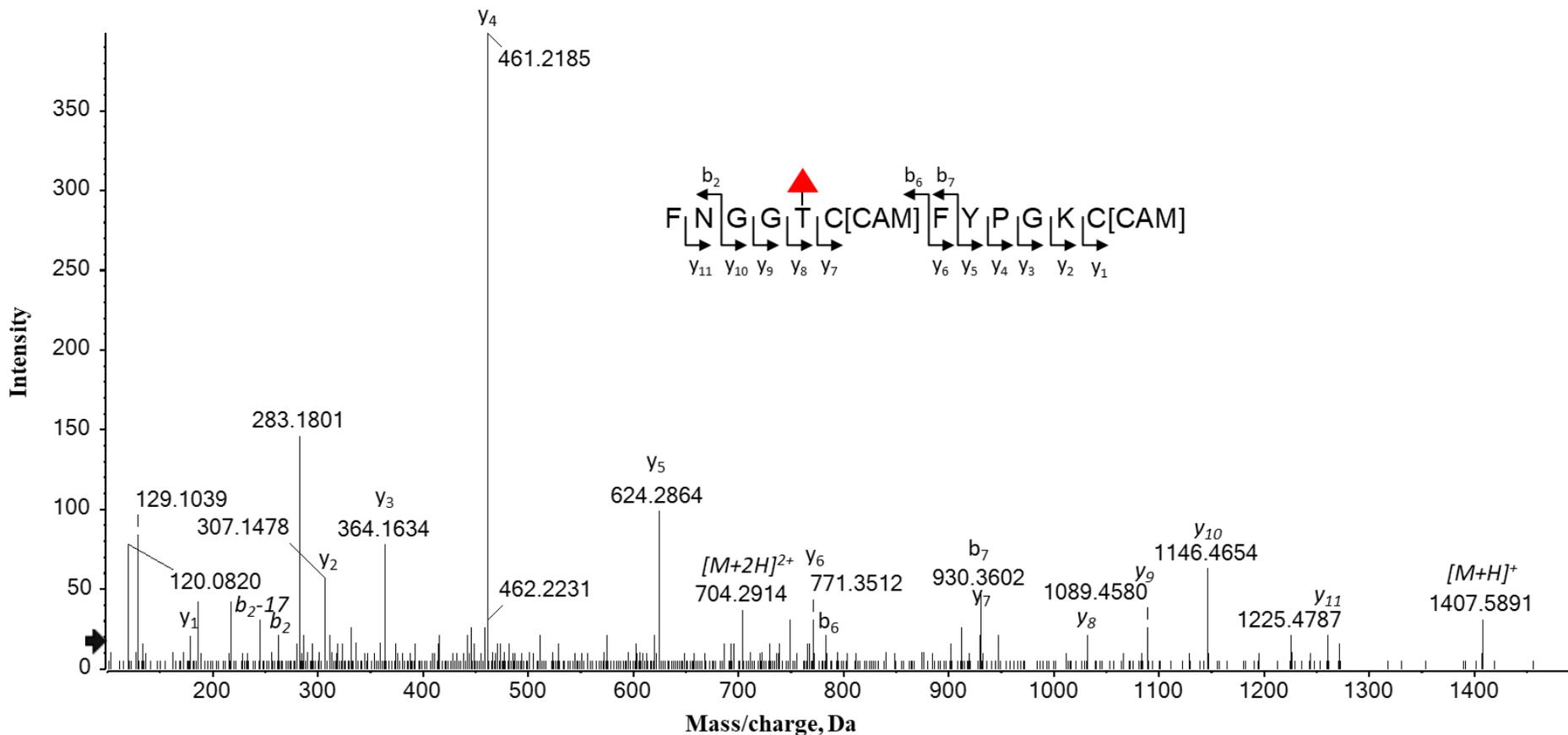


Figure 1: Representative MS/MS spectra of the *O*-fucosylated peptide from human HisWIF1 after thermolysin digestion. Spectra were acquired with a micro-LC TripleTOF system in DDA mode. Detected b/y fragments are annotated on the spectrum confirming the identified peptide sequence. The loss of fucose moiety led to the presence of non-fucosylated precursor b ions and y fragments (in italic).

SUPPLEMENTARY DATA



Figure 2: Comparison of *SfHisV5Po* and *MmHisV5Po* abilities to transfer *O*-fucose to recombinant isolated EGFs of mouse glycoproteins. Blotting technique using streptavidin-HRP of the three recombinant WT EGFs after 1 h incubation at 37°C with *SfHisV5Po* or *MmHisV5Po* in the presence of GDP-azido-fucose (upper panels). Ponceau red staining (middle panel) revealed efficient membrane transfer of all EGFs.

SUPPLEMENTARY DATA

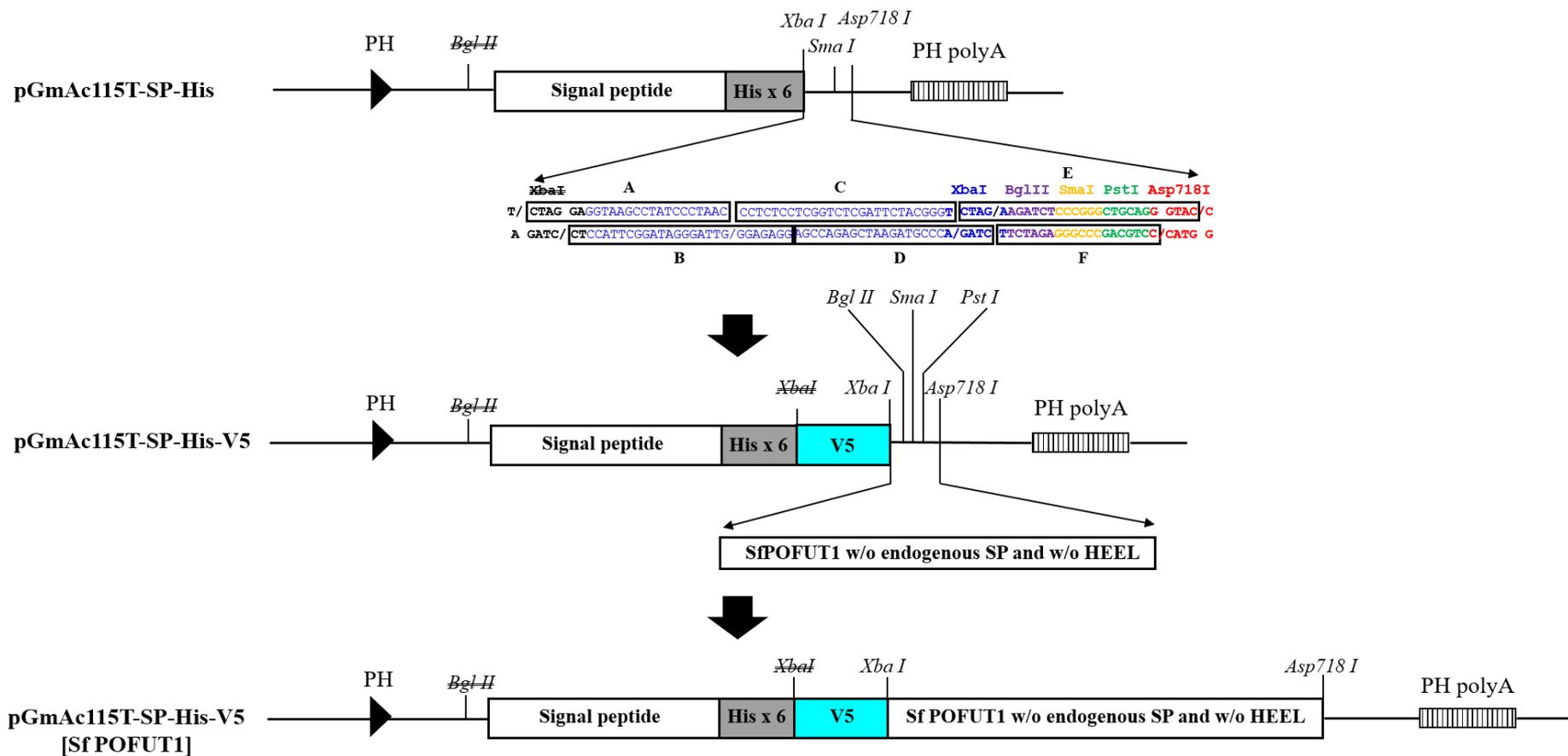


Figure 3: Cloning of *Spodoptera frugiperda* POFUT1 in the baculovirus transfer vector pGmAc115T-SP-His-V5. The pGmAc115T baculovirus transfer vector contains the very late promoter and polyadenylation signal of the polyhedrin protein (PH). A derived vector named pGmAc115T-SP-His was previously obtained by insertion of additional DNA sequence encoding the signal peptide (SP) of insect cell UDP glucosyltransferase gene, six histidine residues (His x 6) and the new cloning site Xba I upstream of the cloning restriction site Asp718 I (Legardinier, Klett et al. 2005). Using the same strategy as before, a new DNA sequence encoding V5 epitope and new cloning sites were inserted, by using prehybridized overlapping primers (A–F), between XbaI and Asp718 I of pGmAc115T-SP-His to generate the new baculovirus transfer vector referred to as pGmAc115T-SP-His-V5. SfPOFUT1 (residues 20-374) cDNA, obtained from Sf9 insect cells and devoid of sequences encoding its endogenous signal peptide and its KDEL-like motif HEEL, was cloned after PCR amplifications and sub-cloning into pGEM-T easy in the baculovirus transfer vector pGmAc115T-SP-His-V5.

SUPPLEMENTARY DATA

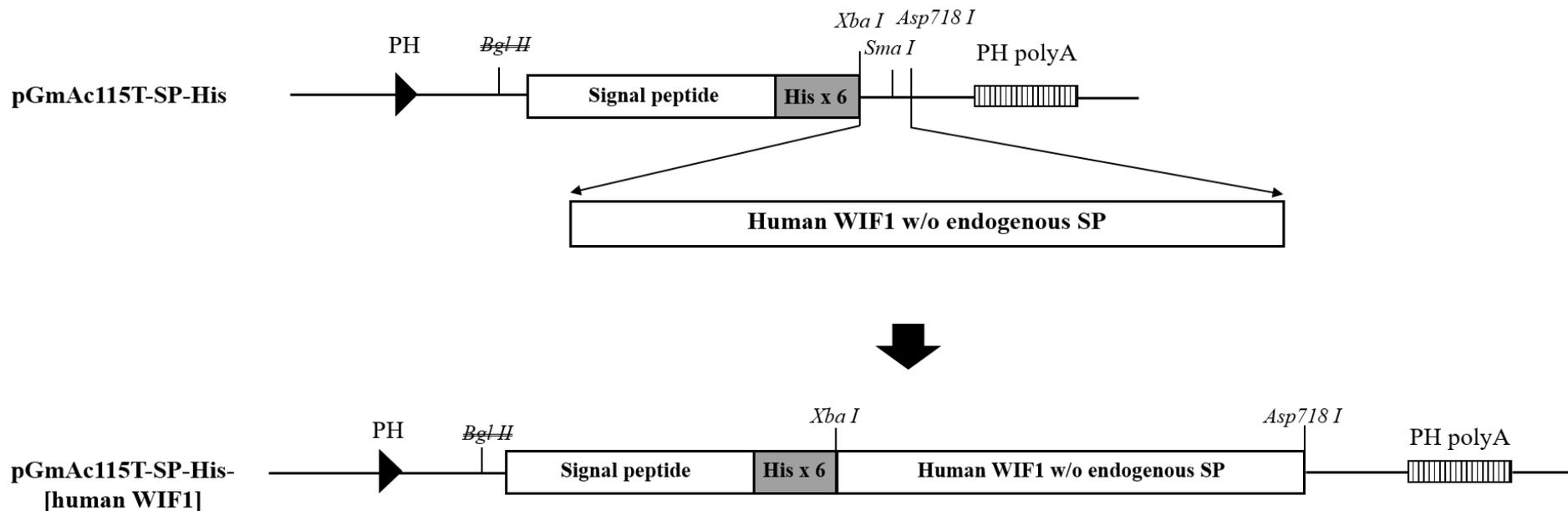


Figure 4: Cloning of human WIF1 in the baculovirus transfer vector pGmAc115T-SP-His. The baculovirus transfer vector pGmAc115T-SP-His, harboring the very late promoter and polyadenylation signal of the polyhedrin gene (PH), was previously obtained by insertion of a signal peptide (SP), six histidine residues (His x 6) and the new cloning site Xba I upstream of the cloning restriction site Asp718 I (Legardinier, Klett et al. 2005). The human WIF1 (NP_009122.2, residues 29-379) cDNA from HEK was also cloned, after PCR amplifications and sub-cloning into pGEM-T easy, in the baculovirus transfer vector pGmAc115T-SP-His by using Xba I and Asp 718 I cloning restriction sites.

SUPPLEMENTARY DATA

Targeted sequence	charge state	Non-modified peptide		Modified peptide		DP (V)	CE (V)		
		MS1	MS2	MS1	MS2				
FNGG <u>T</u> C[CAM]FY <u>P</u> GK[CAM]	+2	704.2946	y4	461.2177	777.3236	y4	461.2177	82.5	36
		not detected	y5	624.2810	detected at	y5	624.2810		
			y3	364.1649	RT 16.38 min	y3	364.1649		
			y10	1146.4707		y10-Fuc	1146.4707		
			y2	307.1435		y2	307.1435		
						b6	783.2978		

Table II. MRMHR parameters for the detection of HisWIF1 EGF3 peptide. The *O*-fucosylation site is indicated in bold and underlined. Each peptide is detected by a single charge state corresponding to a MS1 value (with optimized Declustering Potential (DP) and collision energy (CE)) recorded during the 30 min analysis. Peak detection with MultiQuant was achieved with the six most abundant fragments (MS2) with a 10,000 resolution for modified peptide. Fragment corresponding to the loss of the sugar moiety was also added for modified peptide. Quantification fragment is indicated in bold. The other MS2 were used as confirmation fragments.

Chapter VII. General Discussion

The team that worked for a long time on myogenesis and then on colorectal cancer focused on PAMR1, which represents one of the POFUT1-target proteins. There are few publications on PAMR1 (<30) in the literature. When it was discovered in 2004, it was first named RAMP (Regenerative Associated Muscle Protease) because PAMR1 is a multi-domain protein composed of a peptidase S1 domain which might exert a role in muscle regeneration (Nakayama et al., 2004b) (Hara et al., 2005). Its downexpression could even be implicated in the progression of Duchenne Muscular Dystrophy (DMD) as previously suggested by authors. Since then, the exact function of this protein in the muscle context remains to be elucidated. Finally, the involvement of PAMR1 as a putative tumor suppressor was recently established (Lo et al., 2015).

Expression of PAMR1 in cancer and mainly in CRC

As seen in databases such as Firebrowse, PAMR1 is generally downregulated in cancer including CRC, consistent with previous studies/findings in breast cancer (Lo et al., 2015), cervical cancer (Yang et al., 2021), hepatocellular (Yin et al., 2016), and cutaneous squamous cell carcinoma (Wei et al., 2018). Using these public data, our *in silico* study has globally confirmed the suppressed expression of PAMR1 in the four CRC stages, where it appeared pronounced as early as stage I. However, based on RNASeq data, PAMR1 was also seen to be overexpressed in few cancers such as Kidney renal clear cell carcinoma, skin cutaneous melanoma etc. These various expression levels among cancer types suggest that PAMR1 could exert different roles according to cell types.

Overall results with samples from CRC patients confirm this reduction of PAMR1 quantity in tumors compared to normal tissue samples. However, in rare cases, PAMR1 appeared to be overexpressed in some tissue samples from CRC patients (two samples among 32 samples). For the moment, we have no explanations for these individual differences. Anyway, the dysregulation of PAMR1 expression in cancer cells, whether by overexpression or downexpression, could help understanding its role or even its mechanism of action.

Is PAMR1 downexpression in CRC due to an epigenetic silencing?

Identifying and studying genes dysregulated in cancer allows a better understanding of the process of development of tumors within the carcinogenesis context. As known, one of the epigenetic processes driving the development of cancer is thought to be the DNA methylation of CpG islands in the promoter region of tumor suppressor genes (Patai et al., 2015), as the case of PAMR1 in breast cancer (Lo et al., 2015). Thus, demethylating drugs such as Genistein and 5-aza-2'-deoxycytidine are often used to reverse such promoter hypermethylation.

Genistein, a soy-derived isoflavone, is proposed as a potential therapeutic agent in anticancer research since many years. Besides in breast (Xie et al., 2014) and prostate cancers, Genistein can provoke DNA demethylation leading to an increased expression of tumor suppressor genes in many other cancers (esophageal, pancreatic, lymphoma...) (Javed et al., 2021), including the colorectal cancer (Zhu et al., 2018). Indeed, Genistein allowed recovered expression of the tumor suppressor gene *WIF1* in HT29 colon cancer cells by promoting its promoter demethylation, thus resulting in reduced migration and invasion of HT29 cells (Zhu et al., 2018). Nothing was mentioned in this latter study about PAMR1 expression in HT29 cells following genistein treatment. However, Lo *et al* demonstrated that the epigenetic silencing of

PAMR1 in breast cancer was in fact due to its promoter hypermethylation, rather than mutation. Indeed, PAMR1's suggested tumor suppressor activity was restored in breast cancer cells upon treatment with another demethylating agent, namely 5-aza-2'deoxyctidine (Lo et al., 2015).

Based on that, we wondered whether PAMR1 was also inactivated by promoter hypermethylation in colorectal cancer. We first confirmed that PAMR1 was indeed downexpressed in CRC cell lines available in the lab, especially in HT29 cells, compared to normal colon cells CCD841CoN (see chapter VI.I). We then investigated the effect of both demethylation agents, namely Genistein and 5-aza-2'deoxyctidine, on the expression of PAMR1 in HT29 cells and proliferation of these cells. In parallel, stable HT29 cells overexpressing PAMR1 (Cl2 and Pool) were even assayed as well as stable control cell lines (transfected with empty vector only) named control Mock (see supplemental data). The treatment of previously mentioned cell lines with Genistein or 5-aza-2'deoxyctidine drugs at different concentrations (Genistein: 0, 10, 20, 60 $\mu\text{mol}\cdot\text{L}^{-1}$. 5-aza-2'deoxyctidine: 1, 2, 3, 6, 8, 10, 12 μM) showed no remarkable effect on mRNA expression level of PAMR1 (qPCR) as well as on cell proliferation. However, a slight reduction in Pool cell proliferation was seen upon treatment with the highest dose of 5-aza-2'deoxyctidine. Taking into consideration that this demethylation drug is not PAMR1 specific, we can assume that it can exert its action on cell proliferation by inducing demethylation of other tumor suppressor genes.

The same gene or tumor suppressor gene can be silenced or dysregulated in different cancers by different mechanisms. As an example, as it has been demonstrated within our team, the activity of POFUT1 enzyme is enhanced in CRC by two different mechanisms: by DNA fragment duplications leading to the increase in the *poft1* gene copy number (Chabanais et al., 2018), or by gain-of-function mutation within *poft1* gene (Deschuyter et al., 2020). Another example, ADAM33 (A disintegrin and metalloproteinase domain 33), a transmembrane glycoprotein known to play a role in cell adhesion. Its downregulation was noted in colon cancer as a result of gene mutation (Fraile et al., 2013) whereas it was shown to be silenced by promoter hypermethylation in breast cancer and predicted as a potential tumor suppressor (Manica et al., 2017).

Then, we concluded that the down-expression of PAMR1 was not a result of promoter hypermethylation. As exposed in the introduction, other events (mutations...) or epigenetic alterations (histone deacetylation) or microsatellite instability could lead to PAMR1 downregulation. It could be interesting to discover the reasons of the reduced quantity of PAMR1 in CRC to know how to reverse this downexpression.

Means to explore the action of PAMR1 in CRC cells

Since PAMR1 was shown to be downexpressed in CRC patients and in available CRC cell lines, we chose to overexpress it to determine effects of an increased quantity on properties of tumoral cells. Since we failed to restore the expression of PAMR1 in CRC cells by using demethylation agents, only two experimental approaches were carried out, namely by producing stable cell line overexpressing PAMR1 and by exogenous treatment of CRC cells with recombinant PAMR1.

PAMR1 expression was only dysregulated in HT29 showing the lowest PAMR1 expression among available CRC cell lines in the lab. Obtained stably modified HT29 cells (Pool and Cl2) indeed allowed us to constitutively enhance the quantity of expressed PAMR1 in intracellular cell compartment as well as in secretome. A second approach was to produce recombinant PAMR1 in different expression systems for exogenous treatments by adding it to the growth

medium of CRC cell lines in culture. Finally, the most efficient approach was to increase overall quantity of PAMR1 as much as possible, by combining the exogenous treatment (with CHO-produced recombinant PAMR1) and transient transfection of CRC cell lines with constructs allowing overexpression of human PAMR1.

The large use of mammalian systems such as CHO cells to express recombinant human proteins is mainly due to an efficient quality control of protein folding and to a large range of post-translational modifications necessary for production of a functional recombinant protein. These capabilities are suitable or applicable for all types of proteins whether intracellular, secreted (such as PAMR1) or membrane integrated proteins (Almo and Love, 2014). CHO cells were chosen for production of glycoproteins with similar complex-type *N*-glycans as those found in human glycoproteins. More than 50% of therapeutic proteins are glycosylated (example : monoclonal antibodies), it's the reason why so many recombinant proteins are produced in CHO cells (Durocher and Butler, 2009).

Thanks to the advanced technologies and advantageous capacities of the mammalian expression system for production of recombinant proteins highly close to the native forms, we thus decided to create stable CHO cells to produce recombinant human PAMR1 isoforms, namely the canonical isoform 1 and its isoform 2 (having 17 additional residues between EGF-like domain and Sushi1 domain). Flp-In™ CHO cells (Thermo Fisher scientific) were used for a site-specific integration of the construct harboring PAMR1 cDNA into the genome after Flp recombinase-mediated DNA recombination.

In spite of successful generation of stable CHO cell lines harboring cDNA encoding isoforms 1 and 2 of human PAMR1 with N-terminal histidine tags, the yield of production of these recombinant glycoproteins was unfortunately very low. It did not allow to produce and purify sufficient protein quantities to perform exogenous treatments. Based on that, we looked forward to produce recombinant both isoforms of human PAMR1 in the baculovirus insect cell expression system, known for high yield of production of recombinant proteins at a low cost. Despite a good expression and secretion of both isoforms of human PAMR1 by *Sf9* insect cells infected with appropriate recombinant baculovirus, the recombinant proteins were subject to proteolytic degradation despite the addition of various protease inhibitors. In addition, their nickel-affinity purification after protein concentration accentuated their degradation and others form protein aggregates, leading to insufficient quantities of purified protein for exogenous treatments of CRC cell lines.

Finally, we relied on mouse PAMR1 produced by stable CHO cells (previously obtained in the lab (Pennarubia et al., 2021) to perform exogenous treatment-based experiments. Taking into consideration the high percentage of similarity and identity (> 90 %) between the mouse PAMR1 and the canonical isoform 1 of human PAMR1, we assumed it may exert the same role despite minimal differences between the two sequences. Indeed, exogenous treatment with concentrated secretome of CHO cells expressing mouse PAMR1 showed an effect on both HT29 and HeLa cells proliferation and migration compared to non-treated cells. The HeLa cervical cancer cell line was used to confirm the relevance of the use of mouse PAMR1, since this cell line was already known to be sensitive to PAMR1 dysregulation (Yang et al., 2021).

Origin of PAMR1 instability of recombinant human PAMR1?

Despite of the advantages of the baculovirus insect cell expression system, the production of recombinant proteins is sometimes limited by the proteolytic activity of insect cells which affects the productivity and stability of the target proteins (Dumas and Robert, 2009). Recombinant isoforms 1 and 2 of human PAMR1 named His6-Iso1 and His6-Iso2 respectively in our study

(see Supp data) were both sensitive to protein degradation, probably due to the release of some proteases by infected Sf9 insect cells. Unfortunately, the addition of inhibitors such as leupeptin (inhibitor of serine, threonine and cysteine proteases) and pepstatin A (inhibitor of aspartate proteases) in production medium showed no effect, even though they have been shown to be effective in limiting the degradation of other proteins produced in the baculovirus-insect cell system (Pyle et al., 1995). Despite of a low protein production in serum-free medium, we nevertheless chose to produce human PAMR1 with this medium to avoid that large quantities of contaminant serum albumin restrict the nickel-affinity purification of the polyhistidine tagged recombinant protein of interest. Unfortunately, protein concentration of production media by ultrafiltration or by ammonium sulfate precipitation prior to purification led to formation of protein aggregates or protein complexes and increased protein degradation. We then assumed that PAMR1 instability might be partially explained by the potential absence of O-fucosylation when expressed in the baculovirus insect cell system.

Investigation of the ability of Sf9 insect cells to add O-fucose to human PAMR1?

Most of the insect cells (including Sf9 cells) have the capability to perform most post-translational modifications, such as N-glycosylation and O-glycosylation but with simple glycan structures compared complex glycans found in mammals (Ailor and Betenbaugh, 1999). POFUT1-mediated O-fucosylation can also occur in the baculovirus-insect cell system (Luca et al., 2015) but the efficiency of fucose transfer might depend of the protein of interest and insect cells used.

Pennarubia et al, in our laboratory, demonstrated that mouse PAMR1 was modified with O-fucose within its single EGF-like domain when expressed in CHO cells (Pennarubia et al., 2021) and the presence of O-fucose could influence its stability. This opened up the question whether PAMR1 could also modified by O-fucose, when expressed in another expression system producing functional POFUT1 such as the baculovirus–insect cell system. Unfortunately, we were unable to demonstrate the presence of O-fucose for full-length human PAMR1 expressed by infected Sf9 insect cells due to protein degradation during protein concentration and purification steps. To cope with this problem, Sf9POFUT1 was first produced, purified and characterized for its ability to add *in vitro* O-fucose to the *E. coli*-produced single EGF-like domain of PAMR1 (see Paper 2). Following *in vitro* O-fucosyltransferase assay using Sf9POFUT1 and GDP-azido-fucose, click chemistry and blotting techniques were performed to show the ability of Sf9POFUT1 to specifically add O-fucose to NOCTH1 EGF26 and WIF1 EGF3, in accordance with previous reports of our lab (Pennarubia et al., 2018) (Pennarubia et al., 2021). Our experiments revealed that Sf9POFUT1 exhibited a less efficient O-fucosyltransferase activity than mouse POFUT1.

Unlike mouse POFUT1, Sf9POFUT1 strikingly failed to add O-fucose to isolated EGF-like domain of PAMR1. If the absence of O-fucose is confirmed for full-length human PAMR1, it could partially explain its instability and propension to degradation if considering that mouse PAMR1, produced by CHO cells and exhibiting more stability, was demonstrated to be modified by O-fucose.

Propension of recombinant human PAMR1 to form aggregates or protein complexes

The difference of complexity of glycans, between mammalian cells such as CHO cells and insect cells such as Sf9 cells, can explain differences of molecular weight (MW) of the expressed recombinant protein as well as differences in protein stability, in protein-protein interactions and in bioactivities. Nevertheless, the two isoforms of human PAMR1 (His6-Iso1 and His6-Iso2) expressed in baculovirus-infected Sf9 insect cells were secreted as expected

and exhibited an apparent MW of about 90 kDa. The comparison of theoretical and apparent MW highlights the presence of post-translational modifications (*O*-glycans and/or *N*-glycans) for both secreted His6-Iso1 and His6-Iso2 (supplemental data of paper 1). The apparent MW of the recombinant proteins produced in *Sf9* cells were close but slightly lower to those obtained for recombinant counterparts produced by CHO cells, indicating subtle differences of glycosylation between both expression systems.

As a secreted multi-domain protein, PAMR1 can be assumed to interact with one or more secreted (extracellular) and/or membrane proteins. Human PAMR1 isoform 1 was detected in the supernatant of stable HT29 overexpressing human PAMR1 isoform 1 (pool and Cl2) at the expected size (between 90 and 100 kDa). However, a higher MW band was frequently seen above 130 kDa, even in the case of recombinant PAMR1 expressed by *Sf9* insect cells (His6-Iso2). The higher band suggests either the formation of soluble protein aggregates of PAMR1 due to concentration by ultrafiltration or the presence of an undissociated protein complex, with other proteins, even under the denaturing and reduced conditions, which seems more unlikely.

The precise molecular mechanism of action of PAMR1 remains to be elucidated

The mechanism of action of PAMR1 is currently not known but as a secreted protein, several hypotheses can be put forward. PAMR1 could bind directly to the cell surface to exert its action, either by binding to a receptor or to another membrane protein. It could also bind to another secreted protein in the extracellular compartment to modulate its action. This is what is suggested for SCUBE2, an analogous protein which contains a CUB domain and EGF-like domains like PAMR1. In the case of SCUBE2, its *C*-terminal CUB domain interacts with BMP and antagonize it, thus suppressing breast cancer cell proliferation (Cheng et al., 2009); whereas *N*-terminal EGF-like domains have the ability to interact with E-cadherin and modulate beta catenin signaling pathway (Lin et al., 2011). This leads us to think that the antiproliferative and potential tumor suppressor role of PAMR1 could be through its interaction with proteins such as BMP and E-cadherin.

Knowing the molecular interactions of PAMR1 could help to understand how PAMR1 could affect different signaling pathways. Indeed, Yang et al. (2021), were the first to show the suppression of mTORc and MYC signaling pathways in breast cancer upon the action of PAMR1. However, up to now, there is no clear view on protein-protein interactions induced by PAMR1 and on the exact link with the latter signaling pathways or even other pathways related to regulating cell proliferation, migration, invasion, or tumorigenesis in general.

Chapter VIII. Conclusion and Perspectives

In the light of the above, we were able to validate PAMR1's downexpression in colorectal cancer, in tumors from CRC patients as well as in CRC cell lines. However, the mechanism of PAMR1 downexpression in CRC is still unknown and seems to be not dependent of promoter hypermethylation. Other modifications (histone deacetylation...) could explain an epigenetic silencing of PAMR1 expression. Phenotypes might be partially reversed by using other inhibitors of methylation or inhibitors of histone deacetylation. Finally, other events such as mutations could also be involved in the loss of PAMR1's expression in CRC.

Recovering PAMR1's expression whether by producing HT29 cells overexpressing PAMR1 or by performing exogenous treatment of CRC cells with recombinant PAMR1, or by combining the two approaches, allowed us to show PAMR1's anti-proliferative and anti-migrative role in colorectal cancer. Thus, this leads us to think that PAMR1 could be considered as one of the early tumor suppressor biomarkers of CRC, as previously shown for breast and cervical cancers. Finally, other properties of tumoral cells such as invasion, apoptosis resistance, ability to form spheroids... could also be investigated after treatment with PAMR1.

It is now established that PAMR1 plays a role in cancer, as evidenced by our study and previous studies (Lo et al., 2015) (Yang et al., 2021) but its mechanism of action is still unknown. No direct interaction of PAMR1, which is a glycoprotein found in the extracellular medium, with another protein partner has yet been shown. Protein-protein interaction studies should be conducted to identify protein partners of PAMR1. Since PAMR1's quantity is very low in CRC, experiments should be carried out either with normal colon cell lines to identify PAMR1s' protein partners in normal conditions or with CRC cell lines treated with recombinant PAMR1. Crosslinking experiments with formaldehyde or other molecules followed by mass spectrometry analyses could allow the identification of covalently bound proteins to PAMR1, corresponding to spatially close proteins and thus potential protein partners. Immunoprecipitation experiments could be performed to test the potential interaction of PAMR1 with putative protein partners such as E-cadherin. All these approaches could be also applied on breast cancer cells such as MCF7 or even cervical cancer cells such as HeLa cells. Besides protein-protein interactions studies, it would be also relevant to investigate the different players of signaling pathways (Wnt and mTOR signaling pathways...) involved in proliferation and migration. This could be done by qPCR and/or Western blot after treatment of CRC cells with recombinant PAMR1 or even within stable CRC cells overexpressing PAMR1. The overall results will be a clear view of PAMR1 mechanism of action in CRC cell lines or other cancer cell types. Moving to a higher level, *in vivo* experiments such as xenografts into nude mice could be applied. Stable HT29 cells over expressing PAMR1 (Pool or Cl2) can be introduced into these mice as well as control HT29 cells (transfected with empty vector) to investigate the tumor progression.

The study of the effect of PAMR1 in CRC by exogenous treatments with recombinant human PAMR1 (instead of mouse PAMR1 used in this study) was impossible due to the difficulties in producing large quantities of this protein, in CHO mammalian cells as well as in *Sf9* insect cells. Despite high structural identity between murine and human PAMR1 and the use of similar constructs to produce in the secretome of CHO cells recombinant PAMR1 proteins with N-terminal His and V5 tags, isoforms 1 and 2 of human PAMR1 were very low expressed by CHO cells unlike murine PAMR1.

In the baculovirus-insect cell system, it was possible to produce isoforms 1 and 2 but their production in serum-free media, followed by protein concentration (by ultrafiltration or ammonium sulfate precipitation) and Nickel-affinity purification led to their protein degradation and aggregation. So, the use of this baculovirus-cell system may not be appropriate to produce recombinant human PAMR1. Its instability and propensity to degradation may be related to the absence of O-fucose on its EGF-like domain, as suggested by our *in vitro* O-fucosyltransferase assays using purified *S₁₉POFUT1* and isolated PAMR1 EGF-like domain produced in *E. coli* (paper 2). Thus, the presence or absence of O-fucose on the single EGF-like domain of full-length human PAMR1 should be confirmed. Nevertheless, tremendous efforts must be done to improve stability of these proteins and prevent them to form aggregates or to be degraded. This will allow to produce and purify enough quantities of proteins to perform further experiments as mentioned above.

Bibliography

- Abdel-Rahman, W., Peltomäki, P., 2004. Molecular basis and diagnostics of hereditary colorectal cancers. *Annals of Medicine* 36, 379–388. <https://doi.org/10.1080/07853890410018222>
- Aboueshia, M., Hussein, M.H., Attia, A.S., Swinford, A., Miller, P., Omar, M., Toraih, E.A., Saba, N., Safah, H., Duchesne, J., Kandil, E., 2021. Cancer and COVID-19: analysis of patient outcomes. *Future Oncology* 17, 3499–3510. <https://doi.org/10.2217/fon-2021-0121>
- Ahuja, N., Mohan, A.L., Li, Q., Stolker, J.M., Herman, J.G., Hamilton, S.R., Baylin, S.B., Issa, J.P., 1997. Association between CpG island methylation and microsatellite instability in colorectal cancer. *Cancer Res* 57, 3370–3374.
- Ailor, E., Betenbaugh, M., 1999. Modifying secretion and post-translational processing in insect cells. *Current Opinion in Biotechnology* 10, 142–145. [https://doi.org/10.1016/S0958-1669\(99\)80024-X](https://doi.org/10.1016/S0958-1669(99)80024-X)
- Alfaro, J.F., Gong, C.-X., Monroe, M.E., Aldrich, J.T., Clauss, T.R.W., Purvine, S.O., Wang, Z., Camp, D.G., Shabanowitz, J., Stanley, P., Hart, G.W., Hunt, D.F., Yang, F., Smith, R.D., 2012. Tandem mass spectrometry identifies many mouse brain O -GlcNAcylated proteins including EGF domain-specific O -GlcNAc transferase targets. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7280–7285. <https://doi.org/10.1073/pnas.1200425109>
- Almo, S.C., Love, J.D., 2014. Better and faster: improvements and optimization for mammalian recombinant protein production. *Curr Opin Struct Biol* 26, 39–43. <https://doi.org/10.1016/j.sbi.2014.03.006>
- Amin, M.B., Greene, F.L., Edge, S.B., Compton, C.C., Gershenwald, J.E., Brookland, R.K., Meyer, L., Gress, D.M., Byrd, D.R., Winchester, D.P., 2017. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging: The Eighth Edition AJCC Cancer Staging Manual. *CA: A Cancer Journal for Clinicians* 67, 93–99. <https://doi.org/10.3322/caac.21388>
- An, H.J., Froehlich, J.W., Lebrilla, C.B., 2009. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr Opin Chem Biol* 13, 421–426. <https://doi.org/10.1016/j.cbpa.2009.07.022>
- Angelo, S.N., Lourenço, G.J., Magro, D.O., Nascimento, H., Oliveira, R.A., Leal, R.F., Ayризono, M. de L.S., Fagundes, J.J., Coy, C.S.R., Lima, C.S.P., 2015. Dietary risk factors for colorectal cancer in Brazil: a case control study. *Nutrition Journal* 15. <https://doi.org/10.1186/s12937-016-0139-z>
- Armaghany, T., Wilson, J.D., Chu, Q., Mills, G., 2012. Genetic alterations in colorectal cancer. *Gastrointest Cancer Res* 5, 19–27.
- Arruga, F., Vaisitti, T., Deaglio, S., 2018. The NOTCH Pathway and Its Mutations in Mature B Cell Malignancies. *Front. Oncol.* 8, 550. <https://doi.org/10.3389/fonc.2018.00550>
- Astin, M., Griffin, T., Neal, R.D., Rose, P., Hamilton, W., 2011. The diagnostic value of symptoms for colorectal cancer in primary care: a systematic review. *British Journal of General Practice* 61, e231–e243. <https://doi.org/10.3399/bjgp11X572427>
- Astler, V.B., Coller, F.A., 1954. THE PROGNOSTIC SIGNIFICANCE OF DIRECT EXTENSION OF CARCINOMA OF THE COLON AND RECTUM: *Annals of Surgery* 139, 846–852. <https://doi.org/10.1097/00000658-195406000-00015>

- Avanesov, A., Honeyager, S.M., Malicki, J., Blair, S.S., 2012. The Role of Glypicans in Wnt Inhibitory Factor-1 Activity and the Structural Basis of Wif1's Effects on Wnt and Hedgehog Signaling. *PLoS Genet* 8, e1002503. <https://doi.org/10.1371/journal.pgen.1002503>
- Baker, S.J., Fearon, E.R., Nigro, J.M., Hamilton, S.R., Preisinger, A.C., Jessup, J.M., vanTuinen, P., Ledbetter, D.H., Barker, D.F., Nakamura, Y., White, R., Vogelstein, B., 1989. Chromosome 17 Deletions and p53 Gene Mutations in Colorectal Carcinomas. *Science* 244, 217–221. <https://doi.org/10.1126/science.2649981>
- Balchen, V., Simon, K., 2016. Colorectal cancer development and advances in screening. *Clinical Interventions in Aging* Volume 11, 967–976. <https://doi.org/10.2147/CIA.S109285>
- Banerjee, H., Krauss, C., Worthington, M., Banerjee, N., Walker, R.S., Hodges, S., Chen, L., Rawat, K., Dasgupta, S., Ghosh, S., Mandal, S., 2019. Differential expression of efferocytosis and phagocytosis associated genes in tumor associated macrophages exposed to African American patient derived prostate cancer microenvironment. *J Solid Tumors* 9, 22–27. <https://doi.org/10.5430/jst.v9n2p22>
- Barresi, R., Campbell, K.P., 2006. Dystroglycan: from biosynthesis to pathogenesis of human disease. *Journal of Cell Science* 119, 199–207. <https://doi.org/10.1242/jcs.02814>
- Bennett, E.P., Mandel, U., Clausen, H., Gerken, T.A., Fritz, T.A., Tabak, L.A., 2012. Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* 22, 736–756. <https://doi.org/10.1093/glycob/cwr182>
- Bieberich, E., 2014. Synthesis, Processing, and Function of N-glycans in N-glycoproteins, in: Yu, R.K., Schengrund, C.-L. (Eds.), *Glycobiology of the Nervous System, Advances in Neurobiology*. Springer New York, New York, NY, pp. 47–70. https://doi.org/10.1007/978-1-4939-1154-7_3
- Boland, C.R., Goel, A., 2010. Microsatellite Instability in Colorectal Cancer. *Gastroenterology* 138, 2073–2087.e3. <https://doi.org/10.1053/j.gastro.2009.12.064>
- Boland, C.R., Thibodeau, S.N., Hamilton, S.R., Sidransky, D., Eshleman, J.R., Burt, R.W., Meltzer, S.J., Rodriguez-Bigas, M.A., Fodde, R., Ranzani, G.N., Srivastava, S., 1998. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* 58, 5248–5257.
- Bos, J.L., Fearon, E.R., Hamilton, S.R., Vries, M.V., van Boom, J.H., van der Eb, A.J., Vogelstein, B., 1987. Prevalence of ras gene mutations in human colorectal cancers. *Nature* 327, 293–297. <https://doi.org/10.1038/327293a0>
- Bovolenta, P., Esteve, P., Ruiz, J.M., Cisneros, E., Lopez-Rios, J., 2008. Beyond Wnt inhibition: new functions of secreted Frizzled-related proteins in development and disease. *Journal of Cell Science* 121, 737–746. <https://doi.org/10.1242/jcs.026096>
- Bray, F., Laversanne, M., Weiderpass, E., Soerjomataram, I., 2021. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer* 127, 3029–3030. <https://doi.org/10.1002/cncr.33587>
- Brockhausen, I., Schachter, H., Stanley, P., 2009. O-GalNAc Glycans, in: Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. (Eds.), *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).
- Byrne, R.M., 2017. Colorectal polyposis and inherited colorectal cancer syndromes. *Annals of Gastroenterology*. <https://doi.org/10.20524/aog.2017.0218>

- Centelles, J.J., 2012. General Aspects of Colorectal Cancer. *ISRN Oncology* 2012, 1–19. <https://doi.org/10.5402/2012/139268>
- Chabanais, J., Labrousse, F., Chaunavel, A., Germot, A., Maftah, A., 2018. POFUT1 as a Promising Novel Biomarker of Colorectal Cancer. *Cancers (Basel)* 10. <https://doi.org/10.3390/cancers10110411>
- Chan, A.T., Giovannucci, E.L., 2010. Primary Prevention of Colorectal Cancer. *Gastroenterology* 138, 2029-2043.e10. <https://doi.org/10.1053/j.gastro.2010.01.057>
- Chen, C.-I., Keusch, J.J., Klein, D., Hess, D., Hofsteenge, J., Gut, H., 2012. Structure of human POFUT2: insights into thrombospondin type 1 repeat fold and O -fucosylation: Structure of human protein O -fucosyltransferase 2. *The EMBO Journal* 31, 3183–3197. <https://doi.org/10.1038/emboj.2012.143>
- Cheng, C.-J., Lin, Y.-C., Tsai, M.-T., Chen, C.-S., Hsieh, M.-C., Chen, C.-L., Yang, R.-B., 2009. SCUBE2 Suppresses Breast Tumor Cell Proliferation and Confers a Favorable Prognosis in Invasive Breast Cancer. *Cancer Research* 69, 3634–3641. <https://doi.org/10.1158/0008-5472.CAN-08-3615>
- Cho, S., Shin, A., Park, S.K., Shin, H.-R., Chang, S.-H., Yoo, K.-Y., 2015. Alcohol Drinking, Cigarette Smoking and Risk of Colorectal Cancer in the Korean Multi-center Cancer Cohort. *Journal of Cancer Prevention* 20, 147–152. <https://doi.org/10.15430/JCP.2015.20.2.147>
- Cordle, J., RedfieldZ, C., Stacey, M., van der Merwe, P.A., Willis, A.C., Champion, B.R., Hambleton, S., Handford, P.A., 2008. Localization of the Delta-like-1-binding Site in Human Notch-1 and Its Modulation by Calcium Affinity. *Journal of Biological Chemistry* 283, 11785–11793. <https://doi.org/10.1074/jbc.M708424200>
- Corley, D.A., Kubo, A., Zhao, W., 2008. Abdominal Obesity and the Risk of Esophageal and Gastric Cardia Carcinomas. *Cancer Epidemiology Biomarkers & Prevention* 17, 352–358. <https://doi.org/10.1158/1055-9965.EPI-07-0748>
- Dell, A., Galadari, A., Sastre, F., Hitchen, P., 2010. Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. *International Journal of Microbiology* 2010, 1–14. <https://doi.org/10.1155/2010/148178>
- Deschuyter, M., Pennarubia, F., Pinault, E., Legardinier, S., Maftah, A., 2020. Functional Characterization of POFUT1 Variants Associated with Colorectal Cancer. *Cancers* 12, 1430. <https://doi.org/10.3390/cancers12061430>
- Du, Y., Li, D., Li, N., Su, C., Yang, C., Lin, C., Chen, M., Wu, R., Li, X., Hu, G., 2018. POFUT1 promotes colorectal cancer development through the activation of Notch1 signaling. *Cell Death Dis* 9, 995. <https://doi.org/10.1038/s41419-018-1055-2>
- Duffy, M.J., van Dalen, A., Haglund, C., Hansson, L., Klapdor, R., Lamerz, R., Nilsson, O., Sturgeon, C., Topolcan, O., 2003. Clinical utility of biochemical markers in colorectal cancer. *European Journal of Cancer* 39, 718–727. [https://doi.org/10.1016/S0959-8049\(02\)00811-0](https://doi.org/10.1016/S0959-8049(02)00811-0)
- Dukes, C.E., 1932. The classification of cancer of the rectum. *J. Pathol.* 35, 323–332. <https://doi.org/10.1002/path.1700350303>
- Dumas, J., Robert, B., 2009. Bioproduction de protéines thérapeutiques: Revue et perspectives. *Med Sci (Paris)* 25, 18–26. <https://doi.org/10.1051/medsci/2009252s18>
- Dunning, A.M., Healey, C.S., Pharoah, P.D., Teare, M.D., Ponder, B.A., Easton, D.F., 1999. A systematic review of genetic polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 8, 843–854.

- Durocher, Y., Butler, M., 2009. Expression systems for therapeutic glycoprotein production. *Current Opinion in Biotechnology* 20, 700–707. <https://doi.org/10.1016/j.copbio.2009.10.008>
- Duval, A., Hamelin, R., 2002. Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer Res* 62, 2447–2454.
- Fearon, E.R., Cho, K.R., Nigro, J.M., Kern, S.E., Simons, J.W., Ruppert, J.M., Hamilton, Preisinger, A.C., Thomas, G., Kinzler, K.W., Vogelstein, B., 1990. Identification of a Chromosome 18q Gene that Is Altered in Colorectal Cancers. *Science* 247, 49–56. <https://doi.org/10.1126/science.2294591>
- Finley, G.G., Schulz, N.T., Hill, S.A., Geiser, J.R., Pipas, J.M., Meisler, A.I., 1989. Expression of the myc gene family in different stages of human colorectal cancer. *Oncogene* 4, 963–971.
- Fortini, M.E., 2009. Notch Signaling: The Core Pathway and Its Posttranslational Regulation. *Developmental Cell* 16, 633–647. <https://doi.org/10.1016/j.devcel.2009.03.010>
- Fraile, J.M., Ordóñez, G.R., Quirós, P.M., Astudillo, A., Galván, J.A., Colomer, D., López-Otín, C., Freije, J.M.P., Puente, X.S., 2013. Identification of novel tumor suppressor proteases by degradome profiling of colorectal carcinomas. *Oncotarget* 4, 1919–1932. <https://doi.org/10.18632/oncotarget.1303>
- Giacomini, C.P., Leung, S.Y., Chen, X., Yuen, S.T., Kim, Y.H., Bair, E., Pollack, J.R., 2005. A Gene Expression Signature of Genetic Instability in Colon Cancer. *Cancer Research* 65, 9200–9205. <https://doi.org/10.1158/0008-5472.CAN-04-4163>
- Hadjipetrou, A., Anyfantakis, D., Galanakis, C.G., Kastanakis, M., Kastanakis, S., 2017. Colorectal cancer, screening and primary care: A mini literature review. *World Journal of Gastroenterology* 23, 6049–6058. <https://doi.org/10.3748/wjg.v23.i33.6049>
- Hajdu, S.I., 2011. A note from history: Landmarks in history of cancer, part 1. *Cancer* 117, 1097–1102. <https://doi.org/10.1002/cncr.25553>
- Hallgren, P., Lundblad, A., Svensson, S., 1975. A new type of carbohydrate-protein linkage in a glycopeptide from normal human urine. *Journal of Biological Chemistry* 250, 5312–5314. [https://doi.org/10.1016/S0021-9258\(19\)41182-4](https://doi.org/10.1016/S0021-9258(19)41182-4)
- Han, C., Lin, X., 2005. Shifted from Wnt to Hedgehog Signaling Pathways. *Molecular Cell* 17, 321–322. <https://doi.org/10.1016/j.molcel.2005.01.009>
- Hanahan, D., Weinberg, R.A., 2011. Hallmarks of Cancer: The Next Generation. *Cell* 144, 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>
- Hanahan, D., Weinberg, R.A., 2000. The Hallmarks of Cancer. *Cell* 100, 57–70. [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9)
- Hara, T., Nakayama, Y., Nara, N., 2005. [Regenerative medicine of skeletal muscle]. *Rinsho Shinkeigaku* 45, 880–882.
- Harrison, R.L., Jarvis, D.L., 2006. Protein N-Glycosylation in the Baculovirus–Insect Cell Expression System and Engineering of Insect Cells to Produce “Mammalianized” Recombinant Glycoproteins, in: *Advances in Virus Research*. Elsevier, pp. 159–191. [https://doi.org/10.1016/S0065-3527\(06\)68005-6](https://doi.org/10.1016/S0065-3527(06)68005-6)
- Hennet, T., 2019. Collagen glycosylation. *Current Opinion in Structural Biology* 56, 131–138. <https://doi.org/10.1016/j.sbi.2019.01.015>
- Hippocrates, and Emile Littré 1839 *Oeuvres complètes d'Hippocrate : traduction nouvelle avec le texte grec en regard, collationné sur les manuscrits et toutes les éditions : accompagnée d'une introduction de commentaires médicaux, de variantes et de notes*

- philologiques: suivie d'une table générale des matières. Paris: J.B. Bailliére. <http://archive.org/details/oeuvrescomplte01hippuoft>, accessed April 28, 2021, n.d.
- Hofsteenge, J., Huwiler, K.G., Macek, B., Hess, D., Lawler, J., Mosher, D.F., Peter-Katalinic, J., 2001. C-Mannosylation and O-Fucosylation of the Thrombospondin Type 1 Module. *Journal of Biological Chemistry* 276, 6485–6498. <https://doi.org/10.1074/jbc.M008073200>
- Holdener, B.C., Haltiwanger, R.S., 2019. Protein O-fucosylation: structure and function. *Current Opinion in Structural Biology* 56, 78–86. <https://doi.org/10.1016/j.sbi.2018.12.005>
- Hsieh, J.-C., Kodjabachian, L., Rebbert, M.L., Rattner, A., Smallwood, P.M., Samos, C.H., Nusse, R., Dawid, I.B., Nathans, J., 1999. A new secreted protein that binds to Wnt proteins and inhibits their activities. *Nature* 398, 431–436. <https://doi.org/10.1038/18899>
- <https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/signs-and-symptoms.html>, n.d.
- Jaspersen, K.W., Tuohy, T.M., Neklason, D.W., Burt, R.W., 2010. Hereditary and Familial Colon Cancer. *Gastroenterology* 138, 2044–2058. <https://doi.org/10.1053/j.gastro.2010.01.054>
- Javed, Z., Khan, K., Herrera-Bravo, J., Naeem, S., Iqbal, M.J., Sadia, H., Qadri, Q.R., Raza, S., Irshad, A., Akbar, A., Reiner, Ž., Al-Harrasi, A., Al-Rawahi, A., Satmbekova, D., Butnariu, M., Bagiu, I.C., Bagiu, R.V., Sharifi-Rad, J., 2021. Genistein as a regulator of signaling pathways and microRNAs in different types of cancers. *Cancer Cell Int* 21, 388. <https://doi.org/10.1186/s12935-021-02091-8>
- Kakuda, S., Haltiwanger, R.S., 2017. Deciphering the Fringe-Mediated Notch Code: Identification of Activating and Inhibiting Sites Allowing Discrimination between Ligands. *Developmental Cell* 40, 193–201. <https://doi.org/10.1016/j.devcel.2016.12.013>
- Kankanala, V.L., Mukkamalla, S.K.R., 2022. Carcinoembryonic Antigen, in: StatPearls. StatPearls Publishing, Treasure Island (FL).
- Khan, K.H., 2013. Gene Expression in Mammalian Cells and its Applications. *Advanced Pharmaceutical Bulletin*; eISSN 2251-7308. <https://doi.org/10.5681/APB.2013.042>
- Kim, E.R., 2014. Colorectal cancer in inflammatory bowel disease: The risk, pathogenesis, prevention and diagnosis. *World Journal of Gastroenterology* 20, 9872. <https://doi.org/10.3748/wjg.v20.i29.9872>
- Kötzler, M.P., Hancock, S.M., Withers, S.G., 2014. Glycosidases: Functions, Families and Folds, in: John Wiley & Sons, Ltd (Ed.), ELS. Wiley. <https://doi.org/10.1002/9780470015902.a0020548.pub2>
- Kreppel, L.K., Blomberg, M.A., Hart, G.W., 1997. Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *J Biol Chem* 272, 9308–9315. <https://doi.org/10.1074/jbc.272.14.9308>
- Kress, T.R., Sabò, A., Amati, B., 2015. MYC: connecting selective transcriptional control to global RNA production. *Nat Rev Cancer* 15, 593–607. <https://doi.org/10.1038/nrc3984>
- Kroes, R.A., Dawson, G., Moskal, J.R., 2007. Focused microarray analysis of glyco-gene expression in human glioblastomas. *J Neurochem* 103, 14–24. <https://doi.org/10.1111/j.1471-4159.2007.04780.x>

- Kuipers, E.J., Grady, W.M., Lieberman, D., Seufferlein, T., Sung, J.J., Boelens, P.G., van de Velde, C.J.H., Watanabe, T., 2015. Colorectal cancer. *Nature Reviews Disease Primers* 15065. <https://doi.org/10.1038/nrdp.2015.65>
- Lao, V.V., Grady, W.M., 2011. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* 8, 686–700. <https://doi.org/10.1038/nrgastro.2011.173>
- Larsen, I.S.B., Narimatsu, Y., Clausen, H., Joshi, H.J., Halim, A., 2019. Multiple distinct O-Mannosylation pathways in eukaryotes. *Current Opinion in Structural Biology* 56, 171–178. <https://doi.org/10.1016/j.sbi.2019.03.003>
- Lech, G., 2016. Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. *WJG* 22, 1745. <https://doi.org/10.3748/wjg.v22.i5.1745>
- Lee, T.V., Pandey, A., Jafar-Nejad, H., 2017. Xylosylation of the Notch receptor preserves the balance between its activation by trans-Delta and inhibition by cis-ligands in *Drosophila*. *PLoS Genet* 13, e1006723. <https://doi.org/10.1371/journal.pgen.1006723>
- Leonhard-Melief, C., Haltiwanger, R.S., 2010. O-Fucosylation of Thrombospondin Type 1 Repeats, in: *Methods in Enzymology*. Elsevier, pp. 401–416. [https://doi.org/10.1016/S0076-6879\(10\)80018-7](https://doi.org/10.1016/S0076-6879(10)80018-7)
- Li, M., Cheng, R., Liang, J., Yan, H., Zhang, H., Yang, L., Li, Chengrang, Jiao, Q., Lu, Z., He, J., Ji, J., Shen, Z., Li, Chunqi, Hao, F., Yu, H., Yao, Z., 2013. Mutations in POFUT1, Encoding Protein O-fucosyltransferase 1, Cause Generalized Dowling-Degos Disease. *The American Journal of Human Genetics* 92, 895–903. <https://doi.org/10.1016/j.ajhg.2013.04.022>
- Li, Z., Han, K., Pak, J.E., Satkunarajah, M., Zhou, D., Rini, J.M., 2017. Recognition of EGF-like domains by the Notch-modifying O-fucosyltransferase POFUT1. *Nat Chem Biol* 13, 757–763. <https://doi.org/10.1038/nchembio.2381>
- Lin, Y.-C., Chen, C.-C., Cheng, C.-J., Yang, R.-B., 2011. Domain and Functional Analysis of a Novel Breast Tumor Suppressor Protein, SCUBE2. *Journal of Biological Chemistry* 286, 27039–27047. <https://doi.org/10.1074/jbc.M111.244418>
- Lira-Navarrete, E., Valero-González, J., Villanueva, R., Martínez-Júlvez, M., Tejero, T., Merino, P., Panjekar, S., Hurtado-Guerrero, R., 2011. Structural Insights into the Mechanism of Protein O-Fucosylation. *PLoS ONE* 6, e25365. <https://doi.org/10.1371/journal.pone.0025365>
- Liu, J., Xiao, Q., Xiao, J., Niu, C., Li, Y., Zhang, X., Zhou, Z., Shu, G., Yin, G., 2022. Wnt/ β -catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Sig Transduct Target Ther* 7, 3. <https://doi.org/10.1038/s41392-021-00762-6>
- Lo, P.H.Y., Tanikawa, C., Katagiri, T., Nakamura, Y., Matsuda, K., 2015. Identification of novel epigenetically inactivated gene PAMR1 in breast carcinoma. *Oncol Rep* 33, 267–273. <https://doi.org/10.3892/or.2014.3581>
- Logan, C.Y., Nusse, R., 2004. THE WNT SIGNALING PATHWAY IN DEVELOPMENT AND DISEASE. *Annu. Rev. Cell Dev. Biol.* 20, 781–810. <https://doi.org/10.1146/annurev.cellbio.20.010403.113126>
- Lourenco, C., Resetca, D., Redel, C., Lin, P., MacDonald, A.S., Ciaccio, R., Kenney, T.M.G., Wei, Y., Andrews, D.W., Sunnerhagen, M., Arrowsmith, C.H., Raught, B., Penn, L.Z., 2021. MYC protein interactors in gene transcription and cancer. *Nat Rev Cancer* 21, 579–591. <https://doi.org/10.1038/s41568-021-00367-9>
- Luca, V.C., Jude, K.M., Pierce, N.W., Nachury, M.V., Fischer, S., Garcia, K.C., 2015. Structural basis for Notch1 engagement of Delta-like 4. *Science* 347, 847–853. <https://doi.org/10.1126/science.1261093>

- Lyons, J.J., Milner, J.D., Rosenzweig, S.D., 2015. Glycans Instructing Immunity: The Emerging Role of Altered Glycosylation in Clinical Immunology. *Front. Pediatr.* 3. <https://doi.org/10.3389/fped.2015.00054>
- MacDonald, B.T., Tamai, K., He, X., 2009. Wnt/ β -Catenin Signaling: Components, Mechanisms, and Diseases. *Developmental Cell* 17, 9–26. <https://doi.org/10.1016/j.devcel.2009.06.016>
- Makoukji, J., Makhoul, N.J., Khalil, M., El-Sitt, S., Aldin, E.S., Jabbour, M., Boulos, F., Gadaleta, E., Sangaralingam, A., Chelala, C., Boustany, R.-M., Tfayli, A., 2016. Gene expression profiling of breast cancer in Lebanese women. *Sci Rep* 6, 36639. <https://doi.org/10.1038/srep36639>
- Manchester, K.L., 1995. Theodor Boveri and the origin of malignant tumours. *Trends in Cell Biology* 5, 384–387. [https://doi.org/10.1016/S0962-8924\(00\)89080-7](https://doi.org/10.1016/S0962-8924(00)89080-7)
- Manica, G.C.M., Ribeiro, C.F., Oliveira, M.A.S. de, Pereira, I.T., Chequin, A., Ramos, E.A.S., Klassen, L.M.B., Sebastião, A.P.M., Alvarenga, L.M., Zanata, S.M., Noronha, L.D., Rabinovich, I., Costa, F.F., Souza, E.M., Klassen, G., 2017. Down regulation of ADAM33 as a Predictive Biomarker of Aggressive Breast Cancer. *Sci Rep* 7, 44414. <https://doi.org/10.1038/srep44414>
- Mao, H., Ito, Y., 2017. 4.19 Growth Factors and Protein-Modified Surfaces and Interfaces ☆, in: *Comprehensive Biomaterials II*. Elsevier, pp. 321–359. <https://doi.org/10.1016/B978-0-12-803581-8.10191-2>
- Mitrus, I., Bryndza, E., Sochanik, A., Szala, S., 2012. Evolving models of tumor origin and progression. *Tumor Biol.* 33, 911–917. <https://doi.org/10.1007/s13277-012-0389-0>
- Mollicone, R., Moore, S.E.H., Bovin, N., Garcia-Rosasco, M., Candelier, J.-J., Martinez-Duncker, I., Oriol, R., 2009. Activity, splice variants, conserved peptide motifs, and phylogeny of two new alpha1,3-fucosyltransferase families (FUT10 and FUT11). *J Biol Chem* 284, 4723–4738. <https://doi.org/10.1074/jbc.M809312200>
- Moloney, D.J., Panin, V.M., Johnston, S.H., Chen, J., Shao, L., Wilson, R., Wang, Y., Stanley, P., Irvine, K.D., Haltiwanger, R.S., Vogt, T.F., 2000. Fringe is a glycosyltransferase that modifies Notch. *Nature* 406, 369–375. <https://doi.org/10.1038/35019000>
- Mori, Y., Yin, J., Rashid, A., Leggett, B.A., Young, J., Simms, L., Kuehl, P.M., Langenberg, P., Meltzer, S.J., Stine, O.C., 2001. Instability typing: comprehensive identification of frameshift mutations caused by coding region microsatellite instability. *Cancer Res* 61, 6046–6049.
- Morris, E.J.A., Goldacre, R., Spata, E., Mafham, M., Finan, P.J., Shelton, J., Richards, M., Spencer, K., Emberson, J., Hollings, S., Curnow, P., Gair, D., Sebag-Montefiore, D., Cunningham, C., Rutter, M.D., Nicholson, B.D., Rashbass, J., Landray, M., Collins, R., Casadei, B., Baigent, C., 2021. Impact of the COVID-19 pandemic on the detection and management of colorectal cancer in England: a population-based study. *The Lancet Gastroenterology & Hepatology* 6, 199–208. [https://doi.org/10.1016/S2468-1253\(21\)00005-4](https://doi.org/10.1016/S2468-1253(21)00005-4)
- Nakayama, Y., Nara, N., Kawakita, Y., Takeshima, Y., Arakawa, M., Katoh, M., Morita, S., Iwatsuki, K., Tanaka, K., Okamoto, S., Kitamura, T., Seki, N., Matsuda, R., Matsuo, M., Saito, K., Hara, T., 2004a. Cloning of cDNA encoding a regeneration-associated muscle protease whose expression is attenuated in cell lines derived from Duchenne muscular dystrophy patients. *Am J Pathol* 164, 1773–1782. [https://doi.org/10.1016/S0002-9440\(10\)63735-2](https://doi.org/10.1016/S0002-9440(10)63735-2)
- Nakayama, Y., Nara, N., Kawakita, Y., Takeshima, Y., Arakawa, M., Katoh, M., Morita, S., Iwatsuki, K., Tanaka, K., Okamoto, S., Kitamura, T., Seki, N., Matsuda, R., Matsuo, M., Saito, K., Hara, T., 2004b. Cloning of cDNA Encoding a Regeneration-Associated

- Muscle Protease Whose Expression Is Attenuated in Cell Lines Derived from Duchenne Muscular Dystrophy Patients. *The American Journal of Pathology* 164, 1773–1782. [https://doi.org/10.1016/S0002-9440\(10\)63735-2](https://doi.org/10.1016/S0002-9440(10)63735-2)
- Nguyen, H., Duong, H., 2018. The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy (Review). *Oncol Lett.* <https://doi.org/10.3892/ol.2018.8679>
- Ogawa, M., Okajima, T., 2019. Structure and function of extracellular O-GlcNAc. *Current Opinion in Structural Biology* 56, 72–77. <https://doi.org/10.1016/j.sbi.2018.12.002>
- Okamoto, M., Sato, C., Kohno, Y., Mori, T., Iwama, T., Tonomura, A., Miki, Y., Utsunomiya, J., Nakamura, Y., White, R., Miyaki, M., 1990. Molecular nature of chromosome 5q loss in colorectal tumors and desmoids from patients with familial adenomatous polyposis. *Hum Genet* 85. <https://doi.org/10.1007/BF00193581>
- Okamura, Y., Saga, Y., 2008. Notch signaling is required for the maintenance of enteric neural crest progenitors. *Development* 135, 3555–3565. <https://doi.org/10.1242/dev.022319>
- Patai, Á.V., Valcz, G., Hollósi, P., Kalmár, A., Péterfia, B., Patai, Á., Wichmann, B., Spisák, S., Barták, B.K., Leiszter, K., Tóth, K., Sipos, F., Kovalszky, I., Péter, Z., Miheller, P., Tulassay, Z., Molnár, B., 2015. Comprehensive DNA Methylation Analysis Reveals a Common Ten-Gene Methylation Signature in Colorectal Adenomas and Carcinomas. *PLoS ONE* 10, e0133836. <https://doi.org/10.1371/journal.pone.0133836>
- Pennarubia, F., Germot, A., Pinault, E., Maftah, A., Legardinier, S., 2021. The single EGF-like domain of mouse PAMR1 is modified by O-Glucose, O-Fucose and O-GlcNAc. *Glycobiology.* <https://doi.org/10.1093/glycob/cwaa051>
- Pennarubia, F., Pinault, E., Al Jaam, B., Brun, C.E., Maftah, A., Germot, A., Legardinier, S., 2020. Mouse WIF1 Is Only Modified with O-Fucose in Its EGF-like Domain III Despite Two Evolutionarily Conserved Consensus Sites. *Biomolecules* 10, 1250. <https://doi.org/10.3390/biom10091250>
- Pennarubia, F., Pinault, E., Maftah, A., Legardinier, S., 2018. In vitro acellular method to reveal O -fucosylation on EGF-like domains. *Glycobiology* 29, 192–198. <https://doi.org/10.1093/glycob/cwy106>
- Perkins, G.L., Slater, E.D., Sanders, G.K., Prichard, J.G., 2003. Serum tumor markers. *Am Fam Physician* 68, 1075–1082.
- Pino, M.S., Chung, D.C., 2010. The Chromosomal Instability Pathway in Colon Cancer. *Gastroenterology* 138, 2059–2072. <https://doi.org/10.1053/j.gastro.2009.12.065>
- Pitchumoni, C.S., Broder, A., 2020. Epidemiology of colorectal cancer, in: *Colorectal Neoplasia and the Colorectal Microbiome.* Elsevier, pp. 5–33. <https://doi.org/10.1016/B978-0-12-819672-4.00002-7>
- Pyle, L.E., Barton, P., Fujiwara, Y., Mitchell, A., Fidge, N., 1995. Secretion of biologically active human proapolipoprotein A-I in a baculovirus-insect cell system: protection from degradation by protease inhibitors. *Journal of Lipid Research* 36, 2355–2361. [https://doi.org/10.1016/S0022-2275\(20\)39716-9](https://doi.org/10.1016/S0022-2275(20)39716-9)
- Rampal, R., Arboleda-Velasquez, J.F., Nita-Lazar, A., Kosik, K.S., Haltiwanger, R.S., 2005. Highly Conserved O-Fucose Sites Have Distinct Effects on Notch1 Function. *Journal of Biological Chemistry* 280, 32133–32140. <https://doi.org/10.1074/jbc.M506104200>
- Roncucci, L., Mariani, F., 2015. Prevention of colorectal cancer: How many tools do we have in our basket? *European Journal of Internal Medicine* 26, 752–756. <https://doi.org/10.1016/j.ejim.2015.08.019>
- Ruibal Morell, A., 1992. CEA serum levels in non-neoplastic disease. *Int J Biol Markers* 7, 160–166.

- Sajid, K.M., Parveen, R., Durr-e-Sabih, null, Chaouachi, K., Naeem, A., Mahmood, R., Shamim, R., 2007. Carcinoembryonic antigen (CEA) levels in hookah smokers, cigarette smokers and non-smokers. *J Pak Med Assoc* 57, 595–599.
- Sanchez-Irizarry, C., Carpenter, A.C., Weng, A.P., Pear, W.S., Aster, J.C., Blacklow, S.C., 2004. Notch Subunit Heterodimerization and Prevention of Ligand-Independent Proteolytic Activation Depend, Respectively, on a Novel Domain and the LNR Repeats. *Mol Cell Biol* 24, 9265–9273. <https://doi.org/10.1128/MCB.24.21.9265-9273.2004>
- Schedin-Weiss, S., Winblad, B., Tjernberg, L.O., 2014. The role of protein glycosylation in Alzheimer disease. *FEBS J* 281, 46–62. <https://doi.org/10.1111/febs.12590>
- Schneider, M., Al-Shareffi, E., Haltiwanger, R.S., 2017. Biological functions of fucose in mammals. *Glycobiology* 27, 601–618. <https://doi.org/10.1093/glycob/cwx034>
- Scholz, J., Suppmann, S., 2017. A new single-step protocol for rapid baculovirus-driven protein production in insect cells. *BMC Biotechnol* 17, 83. <https://doi.org/10.1186/s12896-017-0400-3>
- Sehgal, R., Sheahan, K., O’Connell, P., Hanly, A., Martin, S., Winter, D., 2014. Lynch Syndrome: An Updated Review. *Genes* 5, 497–507. <https://doi.org/10.3390/genes5030497>
- Shan, M., Yang, D., Dou, H., Zhang, L., 2019. Fucosylation in cancer biology and its clinical applications, in: *Progress in Molecular Biology and Translational Science*. Elsevier, pp. 93–119. <https://doi.org/10.1016/bs.pmbts.2019.01.002>
- Shaw, E., Farris, M.S., Stone, C.R., Derksen, J.W.G., Johnson, R., Hilsden, R.J., Friedenreich, C.M., Brenner, D.R., 2018. Effects of physical activity on colorectal cancer risk among family history and body mass index subgroups: a systematic review and meta-analysis. *BMC Cancer* 18. <https://doi.org/10.1186/s12885-017-3970-5>
- Shi, S., Stanley, P., 2003. Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5234–5239. <https://doi.org/10.1073/pnas.0831126100>
- Smith, G.E., Summers, M.D., Fraser, M.J., 1983. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol. Cell. Biol.* 3, 2156–2165. <https://doi.org/10.1128/MCB.3.12.2156>
- Stahl, M., Uemura, K., Ge, C., Shi, S., Tashima, Y., Stanley, P., 2008. Roles of Pofut1 and O-Fucose in Mammalian Notch Signaling. *Journal of Biological Chemistry* 283, 13638–13651. <https://doi.org/10.1074/jbc.M802027200>
- Stanley, P., Moremen, K.W., Lewis, N.E., Taniguchi, N., Aebi, M., 2022. N-Glycans, in: Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N.H., Prestegard, J.H., Schnaar, R.L., Seeberger, P.H. (Eds.), *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).
- Steinbuck, M.P., Winandy, S., 2018. A Review of Notch Processing With New Insights Into Ligand-Independent Notch Signaling in T-Cells. *Front. Immunol.* 9, 1230. <https://doi.org/10.3389/fimmu.2018.01230>
- Stephan, C., Kurban, M., Abbas, O., 2021. Dowling-Degos disease: a review. *Int J Dermatol* 60, 944–950. <https://doi.org/10.1111/ijd.15385>
- Stiksma, J., Grootendorst, D.C., van der Linden, P.W.G., 2014. CA 19-9 As a Marker in Addition to CEA to Monitor Colorectal Cancer. *Clinical Colorectal Cancer* 13, 239–244. <https://doi.org/10.1016/j.clcc.2014.09.004>

- Surmann-Schmitt, C., Widmann, N., Dietz, U., Saeger, B., Eitzinger, N., Nakamura, Y., Rattel, M., Latham, R., Hartmann, C., von der Mark, H., Schett, G., von der Mark, K., Stock, M., 2009. Wif-1 is expressed at cartilage-mesenchyme interfaces and impedes Wnt3a-mediated inhibition of chondrogenesis. *Journal of Cell Science* 122, 3627–3637. <https://doi.org/10.1242/jcs.048926>
- Takeuchi, H., Wong, D., Schneider, M., Freeze, H.H., Takeuchi, M., Berardinelli, S.J., Ito, A., Lee, H., Nelson, S.F., Haltiwanger, R.S., 2018. Variant in human POFUT1 reduces enzymatic activity and likely causes a recessive microcephaly, global developmental delay with cardiac and vascular features. *Glycobiology* 28, 276–283. <https://doi.org/10.1093/glycob/cwy014>
- Tariq, K., Ghias, K., 2016. Colorectal cancer carcinogenesis: a review of mechanisms. *Cancer Biol Med* 13, 120–135. <https://doi.org/10.28092/j.issn.2095-3941.2015.0103>
- Tatsuta, M., Iishi, H., Ichii, M., Baba, M., Yamamoto, R., Okuda, S., Kikuchi, K., 1989. Diagnosis of gastric cancers with fluorescein-labeled monoclonal antibodies to carcinoembryonic antigen. *Lasers Surg. Med.* 9, 422–426. <https://doi.org/10.1002/lsm.1900090415>
- Taylor, P., Takeuchi, H., Sheppard, D., Chillakuri, C., Lea, S.M., Haltiwanger, R.S., Handford, P.A., 2014. Fringe-mediated extension of O-linked fucose in the ligand-binding region of Notch1 increases binding to mammalian Notch ligands. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7290–7295. <https://doi.org/10.1073/pnas.1319683111>
- The Biology of cancer by Robert A. Weinberg, n.d.
- The Status, Quality, and Expansion of the NIH Full-Length cDNA Project: The Mammalian Gene Collection (MGC), 2004. *Genome Res.* 14, 2121–2127. <https://doi.org/10.1101/gr.2596504>
- Tiernan, J.P., Perry, S.L., Verghese, E.T., West, N.P., Yeluri, S., Jayne, D.G., Hughes, T.A., 2013. Carcinoembryonic antigen is the preferred biomarker for in vivo colorectal cancer targeting. *Br J Cancer* 108, 662–667. <https://doi.org/10.1038/bjc.2012.605>
- Varki, A., 2017. Biological roles of glycans. *Glycobiology* 27, 3–49. <https://doi.org/10.1093/glycob/cww086>
- Vasudevan, D., Haltiwanger, R.S., 2014. Novel roles for O-linked glycans in protein folding. *Glycoconj J* 31, 417–426. <https://doi.org/10.1007/s10719-014-9556-4>
- Wan, G., Tian, L., Yu, Y., Li, F., Wang, X., Li, C., Deng, S., Yu, X., Cai, X., Zuo, Z., Cao, F., 2017. Overexpression of Pofut1 and activated Notch1 may be associated with poor prognosis in breast cancer. *Biochemical and Biophysical Research Communications* 491, 104–111. <https://doi.org/10.1016/j.bbrc.2017.07.053>
- Wei, W., Chen, Y., Xu, J., Zhou, Y., Bai, X., Yang, M., Zhu, J., 2018. Identification of Biomarker for Cutaneous Squamous Cell Carcinoma Using Microarray Data Analysis. *J Cancer* 9, 400–406. <https://doi.org/10.7150/jca.21381>
- Weiss, L., 2000. Metastasis of cancer: a conceptual history from antiquity to the 1990s. *Cancer Metastasis Rev* 19, I–XI, 193–383.
- Wells, L., Feizi, T., 2019. Editorial overview: Carbohydrates: O-glycosylation. *Current Opinion in Structural Biology* 56, iii–v. <https://doi.org/10.1016/j.sbi.2019.05.010>
- Wouters, M.A., Rigoutsos, I., Chu, C.K., Feng, L.L., Sparrow, D.B., Dunwoodie, S.L., 2005. Evolution of distinct EGF domains with specific functions. *Protein Sci.* 14, 1091–1103. <https://doi.org/10.1110/ps.041207005>
- Xie, Q., Bai, Q., Zou, L.-Y., Zhang, Q.-Y., Zhou, Y., Chang, H., Yi, L., Zhu, J.-D., Mi, M.-T., 2014. Genistein inhibits DNA methylation and increases expression of tumor

- suppressor genes in human breast cancer cells. *Genes Chromosomes Cancer* 53, 422–431. <https://doi.org/10.1002/gcc.22154>
- Yang, R., Ma, M., Yu, S., Li, X., Zhang, J., Wu, S., 2021. High Expression of PAMR1 Predicts Favorable Prognosis and Inhibits Proliferation, Invasion, and Migration in Cervical Cancer. *Front. Oncol.* 11, 742017. <https://doi.org/10.3389/fonc.2021.742017>
- Yin, F., Shu, L., Liu, X., Li, T., Peng, T., Nan, Y., Li, S., Zeng, X., Qiu, X., 2016. Microarray-based identification of genes associated with cancer progression and prognosis in hepatocellular carcinoma. *J Exp Clin Cancer Res* 35, 127. <https://doi.org/10.1186/s13046-016-0403-2>
- Yokota, S., Ogawara, K., Kimura, R., Shimizu, F., Baba, T., Minakawa, Y., Higo, M., Kasamatsu, A., Endo-Sakamoto, Y., Shiiba, M., Tanzawa, H., Uzawa, K., 2013. Protein O-fucosyltransferase 1: A potential diagnostic marker and therapeutic target for human oral cancer. *International Journal of Oncology* 43, 1864–1870. <https://doi.org/10.3892/ijo.2013.2110>
- Yu, H., Takeuchi, H., 2019. Protein O-glycosylation: another essential role of glucose in biology. *Current Opinion in Structural Biology* 56, 64–71. <https://doi.org/10.1016/j.sbi.2018.12.001>
- Yuan, Z., Friedmann, D.R., VanderWielen, B.D., Collins, K.J., Kovall, R.A., 2012. Characterization of CSL (CBF-1, Su(H), Lag-1) mutants reveals differences in signaling mediated by Notch1 and Notch2. *J Biol Chem* 287, 34904–34916. <https://doi.org/10.1074/jbc.M112.403287>
- Zhu, J., Ren, J., Tang, L., 2018. Genistein inhibits invasion and migration of colon cancer cells by recovering WIF1 expression. *Mol Med Report.* <https://doi.org/10.3892/mmr.2018.8760>
- Zielinska, D.F., Gnad, F., Wiśniewski, J.R., Mann, M., 2010. Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* 141, 897–907. <https://doi.org/10.1016/j.cell.2010.04.012>

Abstract

Colorectal cancer (CRC) is becoming one of the most prevalent cancers worldwide. This necessitates better understanding the molecular mechanisms behind its occurrence and identifying biomarkers allowing CRC early detection. In this thesis work, we focused our attention on one POFUT1-target protein, namely PAMR1 (Peptidase domain-containing Associated with Muscle Regeneration 1), which is frequently inactivated in breast and cervical cancers and since considered as protein tumor suppressor. We thus wondered if PAMR1 could also exert a similar role in CRC.

Our *in silico* analysis showed a significantly reduced quantity of *PAMR1* in colorectal cancer tissues as early as stage I, indicating that PAMR1 might be an early biomarker of CRC. PAMR1 was also not detected at the protein level in the secretome of CRC cell lines, consistent with the very low transcripts levels expressed by these cells, as determined by qPCR. To study the effect of a supply or an increased expression of PAMR1 in CRC lines, two experimental approaches were carried out, namely exogenous treatments of CRC cell lines with addition of recombinant PAMR1 in growth medium and transient or stable PAMR1 overexpression in HT29 cell line. Using these two approaches, a reduction in HT29 cell proliferation and migration was shown, correlated to a potential tumor suppressor role of PAMR1 in CRC.

For the previous study, we had to use recombinant mouse PAMR1, stably produced by mammalian CHO cells, to treat the CRC lines. Indeed, we did not succeed in producing and purifying a sufficient quantity of recombinant human PAMR1, either in CHO cells or in the baculovirus-insect cell expression system. In the latter system, we showed the inability of POFUT1 from *Sf9* cells to modify the single EGF-like domain of PAMR1 with O-fucose, which could be one of the reasons for its instability and its strong propensity to degradation.

Keywords : PAMR1, colorectal cancer, Prolifération, Migration, POFUT1, *Spodoptera frugiperda*, EGF-like domain.

Résumé

Le cancer colorectal (CCR) est en train de devenir l'un des cancers les plus répandus dans le monde. Cela nécessite de mieux comprendre les mécanismes moléculaires à l'origine de son apparition et d'identifier des biomarqueurs pour une détection précoce du CCR. Dans ce travail de thèse, nous avons porté notre attention sur PAMR1 (protéine à domaine protéase associée à la régénération musculaire 1), une protéine cible de POFUT1 qui est fréquemment inactivée dans les cancers du sein et du col de l'utérus et considérée depuis comme une protéine suppresseur de tumeur. Nous nous sommes donc demandé si PAMR1 pouvait également exercer un rôle similaire dans le CCR.

Notre analyse *in silico* a montré une quantité significativement réduite de *PAMR1* dans les tissus du cancer colorectal dès le stade I, indiquant que PAMR1 pourrait être un biomarqueur précoce du CCR. PAMR1 n'a pas été détecté non plus au niveau protéique dans le sécrétome de lignées cellulaires de CCR, ce qui est cohérent avec les très faibles niveaux de transcrits exprimés par ces cellules, tels que déterminés par qPCR. Pour étudier l'effet d'un apport ou d'une expression augmentée de PAMR1 dans les lignées CCR, deux approches expérimentales ont été menées, à savoir des traitements exogènes des lignées cellulaires CCR avec ajout de PAMR1 recombinant dans le milieu de croissance et la surexpression transitoire ou stable de PAMR1 dans la lignée cellulaire HT29. Par ces deux approches, une réduction de la prolifération et de la migration des cellules HT29 a été montrée, en adéquation avec un potentiel rôle suppresseur de tumeur de PAMR1 dans le CCR.

Pour l'étude précédente, nous avons dû utiliser du PAMR1 recombinant de souris, produit de manière stable par des cellules de mammifères CHO, pour traiter les lignées CCR. En effet, nous n'avons pas réussi à produire et à purifier en quantité suffisante du PAMR1 humain recombinant, que ce soit dans les cellules CHO ou dans le système d'expression baculovirus-cellules d'insectes. Dans ce dernier système, nous avons montré l'incapacité de POFUT1 des cellules *Sf9* à O-fucosyler le domaine EGF-like unique de PAMR1, ce qui pourrait être une des raisons de son instabilité et de sa forte propension à la dégradation.

Mots-clés: PAMR1, cancer colorectal, Prolifération, Migration, POFUT1, *Spodoptera frugiperda*, domaine de type EGF.

