Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls – University

Heidelberg

Presented by M.Sc, Sofia Rucli

Born in Cividale del Friuli, Italy

Oral examination: 05.10.2023

Functional study of O-GlcNAc modification of chromatin associated proteins

Referees: Prof. Dr. Robert Russell

Dr. James Hackett

Acknowledgements

It seems yesterday I have travelled to EMBL Rome as a PhD candidate. The time has passed very fast, but these years have been incredibly impactful for my professional and personal growth. At my PhD interviews, a friend and at that time PhD student told me “PhD is very tough”, and indeed it was, but I would do it again a million times. These last years at EMBL were crucial for my scientific and personal growth. I had the opportunity to meet an incredibly rich and precious diversity of people, to take part of such a big diversity of activities and to develop so many new and different skills, which will accompany me for the rest of my life. All of this would not have been possible without the contribution of the people I have met.

First, I would like to sincerely thank my supervisor Mathieu. Thank you for giving me the opportunity to be part of your lab and to work together. Thank you for your help from the first day when you were pipetting with me at the bench to teach me the basic protocols to start approaching my PhD. As my experience was growing, you have always been up to help me and support me during my everyday life in the lab. Your mentorship and your guidance have been crucial for me during these years. All the scientific discussions we had shaped me as a scientist and reinforced my critical thinking and my open-mindedness. Your guidance has been crucial to deal with the difficult situations as well as to encourage me with my project and my self-confidence. I am grateful to have worked with you, always with a team-up spirit. My outcome would not have never been the same without your mentorship.

I would like to sincerely thank the members of my doctoral examination committee, Prof. Dr. Robert Russell, Dr. Jamie Hackett, Dr. Arnaud Krebs and Prof. Dr. Michael Knop for taking the time to read my thesis and for your contribution to this work. I would also like to thank Dr. Alexander Aulehla for his guidance and insights during my TAC meetings.

Dear Jamie, I am very grateful for your support during these years. You have been a great mentor for me, always available to suggest me new strategies and providing me with solutions and ideas to solve my experimental problems. The exchange with your lab has been extremely important for my scientific improvement.

I would like to thank Ana for her great contribution to this project, for her availability to talk about science and not only, and for being such an inspiring example for me.

Thanks to Nicolas, with whom I have shared the moments of excitement and disappointment of this project. Thanks for your great contribution, for the brainstorming and the never-ending analysis for this project, and for your encouragements for my future career.

My PhD would not have been the same without my lab mates. Aga, Sara, Monika, Katell, Basilia, Francesco, Rahul, Severin, Aileen, Yulia, Dasha, all of you contributed to improve my scientific work with advice, new ideas, experimental tricks and practical help, with your support and team-up spirit in the everyday life in the lab. I always had the feeling I could work in a team with all of you. I am grateful to have shared the science with you, and for all the non-scientific moments spent together.
I would also like to thank Evangelia and Jeanne, you are two great persons, it was a great pleasure to supervise you, working together was very inspiring for me.

My PhD journey started with Marina and Irene, great friends and scientists. Thank you for always being by my side during these years and for all the moments we have shared together. I would have never been able to become the person I am without you.
I would like to thank the EMBL community for making the environment such a great place where to work. Thanks, Valentina, Yasmin, Tom, Tristan, Pino, Adriana, Alessandro, Pietro, Flavio, Mathilde, Anastasiia, Kristjan, Marzia, Cristina and all the other people of the EMBL Rome community for accompanying me these years.
I am also thankful to all the Facility members that supervised me in very exiting experiments and for the contribution in making my experiments work. Your expertise had a significant impact on my project. Thanks to the rest of the EMBL Staff who contributed to the outcome of this work.

I would like to thank Matteo for his wisdom and for his constant and decisive support during this adventure. Your help and inspiration had a crucial impact on my approach to science and on my personal growth.
Thanks also to Beatrice and Laura for always being up to listen, to support me and to give a decisive spark to my days.

I would like to finally thank my parents for giving me the opportunity to study, to develop as a scientist and for allowing me to be the person I am. I would not be here without your support during all these years.

"Curiosity is the key to problem solving."

- Galileo Galilei

Abstract

O-GlcNAcylation is a reversible posttranslational modification (PTM) found on thousands of nuclear and cytoplasmic proteins catalyzed by OGT and OGA enzymes. Because of the embryonic and cellular lethality resulting from *Ogt* mutation, the mammalian function of O-GlcNAc modification is still unknown. Despite the extensive studies since 1987, when O-GlcNAcylation was identified for the first time, the role of O-GlcNAc and OGT in regulating gene expression remains still enigmatic.

With my PhD project I aim at studying the role of O-GlcNAc modification in mouse chromatin by combining genomic and proteomic approaches. Pan-GlcNAc chromatin profiling revealed that O-GlcNAc proteins densely occupy promoter regions of pluripotent and neuronal differentiated cells. Unbiased bioinformatic screening for co-occupancy with public ChIP-seq data sets revealed that O-GlcNAc strongly co-localizes with RNA Polymerase II (RNA Pol II) at promoters. RNA Pol II was previously shown to be modified by O-GlcNAc modification; while it is well described how RNA Pol II catalyzes the DNA-directed mRNA synthesis of protein-coding genes, the mechanism by which O-GlcNAc regulates its activity is still unknown. O-GlcNAc nuclear perturbation followed by super-resolution imaging revealed a novel role for O-GlcNAc modification in regulating RNA Pol II localization at nuclear transcription factories. To gain deeper insight into O-GlcNAc mechanism in modulating RNA Pol II activity, I have established new tools to perturb and detect O-GlcNAc modification selectively on the C-terminal domain (CTD) of RNA Pol II.

To investigate how O-GlcNAc regulates transcription at promoter regions, we have engineered and validated a novel inducible perturbation system to deplete O-GlcNAc at CpG dense promoters in pluripotent and differentiated cells. Our transcriptomics analysis revealed that targeted O-GlcNAc depletion at CpG-dense promoters causes the perturbation of several metabolic and mitochondrial genes, as well as a significant upregulation in ribosomal subunits expression. qPCR analysis of rRNA demonstrated that O-GlcNAc perturbation at CpG-rich sites leads to a strong deregulation of rRNA expression. Overall, I identified a novel function for O-GlcNAc in the regulation of ribosome biogenesis. This finding is consistent with my proteomic data obtained from the characterization of the OGT interactome, which revealed that nuclear OGT strongly interacts with a large number of nucleolar and ribosomal subunits.

Preliminary mass spectrometry analysis of the nuclear O-GlcNAc proteome showed that the majority of nuclear O-GlcNAc proteins are factors involved in RNA Pol I and III transcriptional regulation, the two RNA Polymerases that control rRNA transcription.

Taken together, my results demonstrate how O-GlcNAc chromatin proteins are not uniformly distributed all over the genome, but instead occupy predominantly gene promoters: part of these promoters are transcriptionally silenced, while the majority of them are active genes. Chromatin enrichment analysis revealed that O-GlcNAc highly correlates with RNA Pol II occupancy, suggesting that O-GlcNAc might regulate the expression of RNA Pol II-rich genes. Finally, by transcriptomics and proteomics approaches I have described a novel mechanism of nuclear O-GlcNAcylation: my analysis demonstrated how nuclear OGT regulates, *via* O-GlcNAc modification, ribosomal biogenesis by controlling the expression of ribosomal proteins and of rRNA. These results bring new avenue into the possible mechanism of OGT in adapting the cellular metabolic state in response to the intracellular O-GlcNAc levels.

Zusammenfassung

O-GlcNAcylierung ist eine reversible posttranslationale Modifikation (PTM), die auf Tausenden von nukleären und zytoplasmatischen Proteinen zu finden ist und von den Enzymen OGT und OGA katalysiert wird. Aufgrund der embryonalen und zellulären Letalität, die aus einer *Ogt*-Mutation resultiert, ist die Funktion der O-GlcNAc-Modifikation bei Säugetieren noch unerforscht. Trotz umfangreicher Studien seit 1987, als die O-GlcNAcylierung zum ersten Mal identifiziert wurde, bleibt die Rolle von O-GlcNAc und OGT bei der Regulierung der Genexpression nach wie vor ungeklärt. Im Rahmen meines PhD Projektes möchte ich die Rolle der O-GlcNAc-Modifikation im Chromatin der Maus untersuchen, indem ich genomische und proteomische Ansätze kombiniere. Die Analyse des Pan-GlcNAc-Chromatinprofils ergab, dass O-GlcNAc-Proteine die Promotorregionen von pluripotenten und neuronal differenzierten Zellen dicht besetzen. Ein unvoreingenommenes bioinformatisches Screening von kooperativen Bindungsereignissen 17526 publizierter ChIP-seq-Datensätze ergab, dass O-GlcNAc stark mit RNA Polymerase II (RNA Pol II) an Promotoren kolokalisiert. Es wurde zuvor gezeigt, dass RNA Pol II durch O-GlcNAc-Modifikation modifiziert wird. Während gut dokumentiert ist, wie RNA Pol II die DNA-gesteuerte mRNA-Synthese von proteinkodierenden Genen katalysiert, ist der Mechanismus, durch den O-GlcNAc seine Aktivität reguliert, noch unbekannt. Perturbation von O-GlcNAc im Zellkern, gefolgt von Superresolution-Imaging, hat eine neue Rolle für O-GlcNAc-Modifikationen bei der Regulierung der RNA Pol II-Lokalisierung in nuklearen Transkriptionszentren aufgezeigt. Um einen detaillierteren Einblick in den O-GlcNAc-Mechanismus bei der Modulation der RNA Pol II-Aktivität zu erhalten, habe ich neue Methoden entwickelt, um die O-GlcNAc-Modifikation selektiv an der C-terminalen Domäne (CTD) von RNA Pol II zu stören und dies nachzuweisen.

Um zu untersuchen, wie O-GlcNAc die Transkription an Promotorregionen reguliert, haben wir ein neuartiges induzierbares Störungssystem entwickelt und validiert, um O-GlcNAc an CpG-dichten Promotoren in pluripotenten und differenzierten Zellen zu verringern. Unsere Transkriptomik-Analyse ergab, dass die gezielte Eliminierung von O-GlcNAc an CpG-dichten Promotoren zu einer Störung mehrerer metabolischer und mitochondrialer Gene sowie zu einer signifikanten Hochregulierung der Expression ribosomaler Untereinheiten führt. qPCR-Analyse von rRNA zeigte, dass eine O-GlcNAc-Störung an CpG-reichen Stellen zu einer starken Deregulierung der rRNA-Expression führt. Dadurch habe ich eine neue Funktion für O-GlcNAc bei der Regulierung der Ribosomenbiogenese identifiziert. Dieser Fund stimmt mit meinen proteomischen Daten überein, die ich durch die Charakterisierung des OGT-Interaktoms erhalten habe. Diese zeigten, dass das nukleäre OGT stark mit einer großen

Anzahl von nukleolaren und ribosomalen Untereinheiten interagiert. Vorläufige massenspektrometrische Analysen des nuklearen O-GlcNAc-Proteoms zeigten, dass die Mehrzahl der nuklearen O-GlcNAc-Proteine an der Regulation der RNA Pol I- und III-Transkription beteiligt sind, den beiden RNA-Polymerasen, die die rRNA-Transkription steuern.

Zusammengenommen zeigen meine Ergebnisse, dass O-GlcNAc-Proteine im Chromatin nicht gleichmäßig über das gesamte Genom verteilt sind, sondern überwiegend Genpromotoren besetzen: Ein Teil dieser Promotoren ist transkriptionell stummgeschalten, während die sich Mehrheit davon in aktiven Genregionen befinden. Chromatin-Anreicherungsanalysen zeigten, dass O-GlcNAc stark mit der RNA Pol II-Besetzung korreliert, was darauf hindeutet, dass O-GlcNAc die Expression von RNA Pol II-reichen Genen regulieren könnte. Abschließend habe ich mit Hilfe von Transkriptomik- und Proteomik-Ansätzen einen neuen Mechanismus der nukleären O-GlcNAcylierung beschrieben: Meine Analyse hat gezeigt, wie das nukleäre OGT über die O-GlcNAc-Modifikation die ribosomale Biogenese reguliert, indem es ribosomale Proteine und die rRNA-Expression kontrolliert. Diese Ergebnisse eröffnen neue Möglichkeiten für den möglichen Mechanismus von OGT bei der Anpassung des zellulären Stoffwechselzustands als Reaktion auf den intrazellulären O-GlcNAc-Gehalt.

Table of Contents

Abbreviations

5caC	5-carboxylxytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ATAC-seq	Assay for Transposase-Accessible Chromatin with sequencing
bFGF	basic fibroblast growth factor
C. elegans	*Caenorhabditis elegans*
CGI	CpG island
ChIP	Chromatin immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeat
CTD	C-terminal domain
CUT&RUN	Cleavage under targets and release using nuclease
DAPI	4,6-diamino-phenylindole
ddPCR	Digital droplet-based PCR
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DOX	Doxycycline
Drosophila	*Drosophila Melanogaster*
E. Coli	*Escherichia Coli*
EGFP	Enhanced green fluorescent protein
ER	Endoplasmic reticulum
ESC	Embryonic stem cell
ETD	Electron transfer dissociation
GEEF	Gene Editing & Embryology Facility
gRNA	Guide RNA
FBS	Fetal bovine serum

FDR	False discovery rate
IAP	Intracisternal A particle
ICM	Inner cell mass
ICR	Imprinting control region
IF	Immunofluorescence
LIF	Leukemia inhibitor factor
LTR	Long terminal repeat
MEF	Mouse embryonic fibroblast
NGS	Next generation sequencing
NLS	Nuclear localization sequence
NPC	Neuronal progenitor cell / Nuclear pore complex
OGA	O-GlcNAcase
O-GlcNAc	*O*-linked β-*N*-acetylglucosamine
OGT	O-GlcNAc transferase
PEPCore	Protein Expression and Purification Core Facility
PTM	Post translational modification
qPCR	Quantitative PCR
rDNA	Ribosomal DNA
rRNA	Ribosomal RNA
RNA Pol I	RNA Polymerase I
RNA Pol II	RNA Polymerase II
RNA Pol III	RNA Polymerase III
RNA	Ribonucleic acid
RNP	Ribonucleoprotein particle
RPL	Ribosomal protein large
RPS	Ribosomal protein small

SAG	Smoothened agonist
siRNA	Small interfering RNA
STED	Stimulated emission depletion
PcG	Polycomb group
PIC	Pre-initiation complex
PR-DUB	Polycomb Repressive Deubiquitinase
PRC	Polycomb repressive complex
PRE	Polycomb response elements
TES	Transcription end site
TET	Ten-Eleven Translocation
TF	Transcription factor
TPR	Tetratricopeptide
TSS	Transcription star site
WB	Western Blot
WGA	Wheat germ agglutinin
WT	Wild type

1. Introduction

1.1 From DNA sequence to protein biosynthesis

About 1% of the entire genome is constituted by protein-coding genes which serve as genetic information to produce proteins, which are macromolecular polypeptides made up of amino acid residues linked by peptide bonds. Proteins they are essential to ensure the cellular structure, function and regulation of the signaling pathways, from unicellular to multicellular organisms.

The information encoded in the DNA sequence of protein-coding genes is first transferred into the mRNA molecule by the enzyme RNA Polymerase II and the transcriptional machinery, with the process called RNA transcription. The information encoded in the mRNA is subsequently converted into the amino acid protein sequence by ribosomes, with the process known as protein translation. Each group of three mRNA nucleotides form a so-called codon that represents the coding sequence of a specific amino acid; a sequence of specific amino acids characterizes each specific protein. This entire process defines the central dogma of molecular biology, stated for the first time by Francis Crick in 1957 (published in 1958 in Nature as “On Protein Synthesis”) (Borsook, 1958), which describes how the flow of the genetic information is interpreted in a biological system.

1.1.1 Ribosomes

Protein translation is achieved by specialized macromolecules called ribosomes, which use the mRNA template to synthetize polypeptide chains. Ribosomes are bound to the endoplasmic reticulum (ER) and to the nuclear envelope (Fig. 1A), or can be found as free ribosomes in the cytosol. In eukaryotic cells, ribosomes are constituted by two large ribonucleoprotein particles (RNPs): the large 60S subunit and the small 40S subunit (Falvey and Staehelin, 1970) (Fig. 1B). The 60S and 40S subunits, which together form the 80S subunit, assemble with the mRNA molecule when ready to be translated. The 40S subunit is responsible for the decoding of the mRNA information, while the 60S

catalyzes the formation of the amino acid peptide bonds (Klinge et al., 2011; Lafontaine and Tollervey, 2001; Rabl et al., 2011).

Ribonucleoprotein particles are composed by ~60% of ribosomal RNA (rRNA) and by ~40% of ribosomal protein subunits; these two components assemble together to form the mature ribosomes.

1.1.2 Ribosomal RNA (rRNA)

The rRNA is a type of non-coding RNA which makes up about 80% of the total cellular RNA. The rRNA is also defined as ribozyme due to its capacity to catalyze biochemical reactions. The ribosomal RNA is transcribed from the ribosomal DNA (rDNA), which in human is constituted by a series of DNA tandem repeats of different copy numbers located in five clusters of chromosomes 13, 14, 15, 21 and 22 (Henderson et al., 1972; McStay, 2016). The number of rDNA repeats varies based on the species, but the phylogenetic interrelation is still not well understood (Gibbons et al., 2015; Nurk et al., 2022; Stults et al., 2008). In eukaryotes, the rDNA is organized in 3D chromatin clusters that confine in the nucleolus organizer region (NOR) (Stults et al., 2008). This chromatin organization makes the nucleolus the primary site of ribosome biogenesis (see paragraph 1.1.4 for nucleolus definition).

In eukaryotes, the rRNA is characterized by four types of RNA units: the 18S, 5.8S, 28S and 5S units. The 18S, 5.8S and 28S units are transcribed by RNA Pol I from a common transcript: the pre-RNA is initially constituted by the 18S, 5.8S and 28S units comprised within the 5' and 3' external transcribed spacers and interspersed by two internal ITS1 and ITS2 transcribed spacers (Goodfellow and Zomerdijk, 2013). The pre-rRNA is processed into two intermediates steps, the 47S and 45S pre-RNA, and is subsequently cleaved and covalently modified by snoRNAs to generate the final 18S, 5.8S and 28S rRNA units. The 5S rRNA unit is instead transcribed by RNA Pol III from repeated gene copies of human chromosome 1 (Pelletier et al., 2018) (Fig. 1C).

The rRNA can assume different structural conformations due to the base-pairing events within its sequence. When processed and folded, the rRNA assembles with the ribosomal subunits to form the mature ribosomes.

1.1.3 Ribosomal subunits (RPs)

The second constituent of ribosomes is characterized by ribosomal proteins, which is a large group of proteins classified between small ribosomal subunits (RPSs) and large ribosomal subunits (RPLs). Ribosomal protein genes are among the most highly expressed genes in the cell and their expression is regulated by RNA Pol II. In eukaryotes, up to 80 ribosomal proteins were identified so far, which show a high conservation across evolution. The overall number of RPs varies depending on the complexity of the species, with a RPs core component which seems to be conserved across evolution (Jansen et al., 1991; Yonath and Franceschi, 1998). In bacteria, RPs transcription is regulated by a unique regulon which control the transcription of all RPs in response to nutrients availability. In eukaryotes however, ribosomal protein regulation is more convoluted as RPs genes are encoded by independent genes which are interspersed across several chromosomes (Brown et al., 2008; Chaker-Margot, 2018; Eichler and Craig, 1994; Laferté et al., 2006). Because of their heterogeneous localization, RP genes regulation of higher organisms is more complex and was demonstrated to be also controlled by the TOR and PKA pathways in response to nutrient availability (Mayer and Grummt, 2006; Powers and Walter, 1999). Interestingly, multiple RP genes share common promoter and translation motifs (Liao et al., 2008; Mayer and Grummt, 2006), even though RPs expression seems to be driven by different mechanisms which regulate independent RP gene clusters (Ghulam et al., 2019).

1.1.4 The nucleolus regulates the biogenesis of ribosomes

Ribosome biogenesis is the process that comprises the transcription, modification and processing of the rRNA, the synthesis of RPs and the final assembly of rRNA and RPs to generate mature ribosomes. This process is one of the most dynamic and energy-consuming processes of mammalian cells, and it is precisely coordinated in response to growth factors, nutrient availability and proliferation stimuli of the cell (Fromont-Racine et al., 2003; Kressler et al., 2010; Moss, 2004; Warner, 1999). Specifically in eukaryotes, ribosome biogenesis is a structured process which reflects the complexity of mammalian ribosomes and composition of nucleolar proteome (Andersen et al., 2005; Andersen et al., 2002; Couté et al., 2006; Huh et al., 2003; Scherl et al., 2002).

The synthesis of ribosomes occurs in the nucleolus, which is the largest subnuclear structure assembled around arrays of ribosomal genes and enriched by nucleolar-

specific proteins (Fig. 1A). The nucleolus is a membrane-less organelle made of different compartments: the fibrillar center (FC), the dense fibrillar component (DFC) and the granular component (GC) (Trumtel et al., 2000).

Ribosome biogenesis is regulated by the coordinated activity of RNA Pol I, RNA Pol II, RNA Pol III, small nucleolar RNAs (snoRNAs), transcription factors and nucleolar proteins (Hori et al., 2023). The pre-rRNA 47S is synthetized between the FC and the DFC compartments by RNA Pol I, while the 5S rRNA in the nucleoplasm by RNA Pol III. The 5S subunit is then assembled in the GC compartment together with the RPs, which are transcribed by RNA Pol II. Mature ribosomes form the mature functional 80S subunit and are finally exported to the cytoplasm where they exert their protein synthesis function.

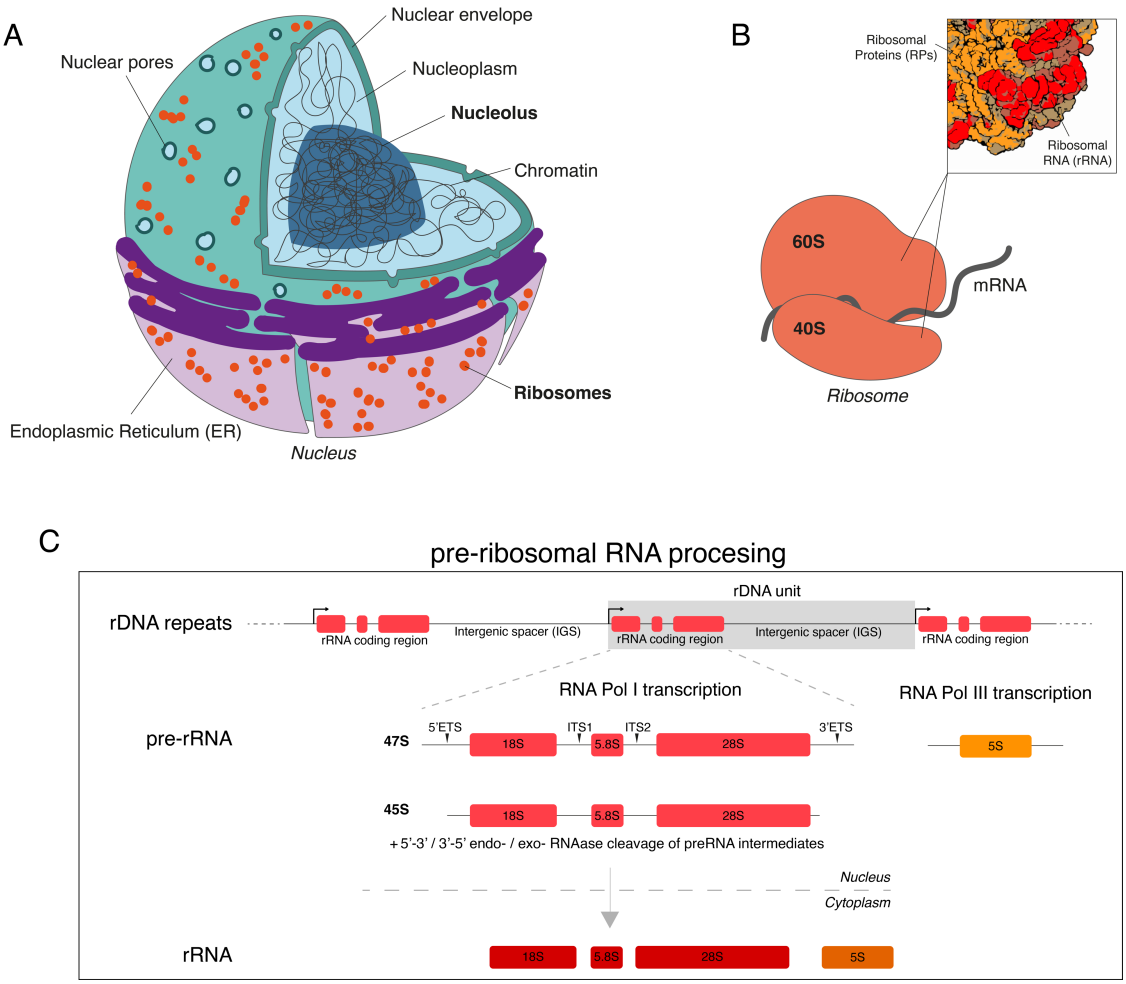

Fig. 1 (A) A comprehensive diagram of the nucleus, with a focus on the nucleolus and on ribosomes on the nuclear envelope and on the endoplasmic reticulum (ER). **(B)** Representation of eukaryotic ribosomes, which are constituted by the 60S and 40S subunits. The assembly of the ribosomal protein (RPs) and of the ribosomal RNA (rRNA) make up the 60S and 40S subunits. **(C)** The rRNA is processed by several intermediate steps. The rDNA is characterized by several repeats constituted by a rRNA coding region and by an intergenic spacer (IGS). The 18S, 5.8S and 28S RNA subunits derive from the long 47S pre-rRNA transcripts transcribed by RNA Pol I. The 18S, 5.8S and 28S sequences are flanked by the 5'-ETS and 3'-ETS external spacers and by the ITS1 and ITS2 internal spacers. These spacers are removed in different steps during the pre-rRNA processing by endo- and exo- RNAases. The 5S rRNA is transcribed by RNA Pol III from a different genomic locus. The mature rRNA transcripts are exported to the cytoplasm, where together with the ribosomal proteins, they constitute the mature ribosomes.

1.2 Post-translational modifications (PTMs)

To adapt to environmental changes, cells must interpret the extracellular signals to trigger precise molecular responses. This complex process is achieved through cell signaling, which refers to all intracellular processes that mediate the communication between the extracellular environment and tailor specific intracellular molecular responses. Regulatory protein-protein interactions are essential to coordinate the responses to extracellular stimuli. Furthermore, protein interactions and functions are modulated also by post-translational modifications (PTMs), which are chemical groups that can be added to protein chains after protein biosynthesis. PTMs can be reversible (covalent) or irreversible and their turnover is essential for the cell signaling regulation as it was shown to affect protein dynamics and protein chemical properties (Huang et al., 2019).

PTMs signaling is regulated by protein writers, enzymes that add the chemical groups on mature proteins, protein readers, which recognize the specific PTMs, and by protein erasers, enzymes that catalyze the removal of the protein chemical groups. The catalysis of PTMs occurs in the endoplasmic reticulum (ER) and in the Golgi apparatus only after the protein is fully synthetized in the cytoplasm by ribosomes, and occurs on specific functional groups that serve as nucleophiles, such as the hydroxyl groups of serine, threonines and tyrosynes, on the amines of lysine, on arginine and histidine, on the thiolate anion of cysteine, on carboxylates of aspartate and glutamate. PTMs can additionally occur on the N-term and C-term amino acid side chains of proteins. So far, more than 400 different types of PTMs were identified, the best characterized are phosphorylation, acetylation, methylation, ubiquitination, SUMOylation and glycosylation (Ramazi and Zahiri, 2021).

The PTMs dynamic turnover is a mechanism that the cell adopts to respond rapidly to the environmental stimuli. The presence of several PTMs generate a dynamic code that function as a second layer of coding information that is translated to specific biological functions.

A well-studied example is represented by the PTMs of histone tails. Histones (H2A, H2B, H3, H4) are proteins that assemble to form nucleosomes, the core subunits of chromatin indispensable for the DNA packing. Histones are characterized by N- and C- terminal tails that are subjected to a wide range of PTMs (methylation, acetylation, phosphorylation, ubiquitylation etc.); histone modifications recruit specific chromatin remodelers and transcriptional activators/repressors which modulates gene expression

by regulating chromatin compaction and cell signaling in specific developmental and environmental contexts (Soshnev et al., 2016).

Protein modifications can also affect the function of other modifications. This phenomenon is very frequent on histone tails, where the presence of specific PTMs affects the binding of other protein readers, as well as the catalysis of other protein modifications. Moreover, it is widely described that PTMs can be mutually exclusive, as for example in the case of O-GlcNAcylation and phosphorylation, that can compete for the same serine or threonine residue (Wang et al., 2007).

Below I introduce the post-translational modifications that are most relevant for this study: methylation, phosphorylation and O-GlcNAcylation, the main focus of my PhD project.

1.2.1 DNA methylation and histone tail methylation

Methylation is one of the chemical modifications most extensively characterized so far. Methylation can occur in two different contexts: on the DNA, where a methyl group can be added on the fifth carbon of cytosine (5-methylcytosine (5mC)) (Doskočil and Šorm, 1962), and on proteins such as histone tails, where a methyl group can be instead added on arginines (R), lysisnes (K) and histidines (H) protein residues (Byvoet et al., 1972; Fischle et al., 2008; Murray, 1964).

Methylation is a very conserved modification across evolution (Feng et al., 2010). DNA methylation in mammals occurs on symmetrical CpG dinucleotides (Zemach et al., 2010), where its deposition and maintenance is regulated by three enzymes: the DNA Methyltransferase 1 (DNMT1) is the canonical enzymes responsible for the maintenance of DNA methylation during DNA replication (Bestor et al., 1988; Bestor, 2000), while the DNA methyltransferases 3A (DNMT3A) and 3B (DNMT3B) are responsible for the de novo DNA methylation at CpG sites (Jurkowska et al., 2011). Active DNA demethylation is catalyzed by TET methylcytosine dioxygenases, a family of enzymes that oxidize the 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylxytosine (5caC) which serve as intermediates for DNA dilution during replication (as only m5C can be maintained upon cell division) (Hashimoto et al., 2012; Ito et al., 2011; Tahiliani et al., 2009).

Mammalian genomes are depleted in CpGs except for CpG island (CGIs), genomic regions of ~1 kb that include two thirds of mammalian promoters (Edwards et al., 2017). 70-80% of these CpG sites are methylated in any somatic cell (Li and Zhang, 2014).

CpG methylation patterns undergo two waves of reprogramming during embryogenesis, after fertilization and after germline cell specification (Monk et al., 1987; Sanford et al., 1987). DNA methylation at CpG rich promoters causes strong and heritable transcriptional silencing and was described to act as a repressor in several epigenetic mechanisms, including X-inactivation (Grant et al., 1992), genomic imprinting (Bourc'his et al., 2001), in the control of germline-specific genes (Mohammad et al., 2011) and repression of retrotransposons (Walsh et al., 1998). DNA methylation is also found at gene bodies where it is positively correlated with transcription, however in this context its function remains unclear (Bender et al., 1999; Varley et al., 2013).

Methylation can also modify histone tails, where it contributes with other PTMs in the control of gene expression. Histone methylation can be found as mono-methylation (me1), di-methylation (me2) or tri-methylation (me3) on the amine group of lysines (K) and as mono-methylation (me1) on guanidinyl group of arginines (R) and histidines (H) (Borun et al., 1972; Gershey et al., 1969; Hempel et al., 1968; Murray, 1964; Paik and Kim, 1967). Among these, the most extensively characterized histone methylation sites are H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20. The turnover of histone methylation is regulated by several histone methyltransferases (HMTs), which use the acetyl-CoA and S-adenosylmethionine (SAM) intermediate metabolites as donors for histone acetylation and methylation respectively, and by several histone demethylases (HDMs). The recognition of methylated histones is controlled by proteins with a methyl-binding domain and with a PHD domain (Jain et al., 2020; Taverna et al., 2007). The genomic patterns of histone modifications are associated with specific gene expression profiles, from development to differentiation, during the cell cycle and after specific cellular stress responses (Greer and Shi, 2012; Wang et al., 2022b).

1.2.2 Phosphorylation

Protein phosphorylation is the most abundant post-translational modification in the cell, with more than 200,000 human phosphorylation sites mapped so far (Ochoa et al., 2020). Because the phosphate group is charged and hydrophilic, it affects the protein conformation and protein interactions properties.

Protein phosphorylation is a reversible modification catalyzed by protein kinases, enzymes that transfer a γ-phosphate from ATP to Serine, Threonine and Tyrosine residues. Protein phosphatases are instead enzymes that use an H_2O molecule to cleave

the phosphate group from the protein amino acid chains. The high frequency of phosphorylation sites reflects the abundancy of protein kinases, one of the largest protein families in eukaryotes with 538 human kinases identified so far (Zhang et al., 2021a). Protein phosphorylation regulates nearly all cellular processes, including cellular growth, metabolism, differentiation, division, immunity and trafficking (Manning et al., 2002).

So far, almost 25,000 proteins were found to be phosphorylated (Li et al., 2009), including many proteins involved in transcriptional regulation. One example is the phosphorylation of Ser and Thr residues of the C-terminal domain of RNA Polymerase II (Eick and Geyer, 2013). In a similar fashion as lysine methylation, phosphorylation can also modify histone tails and contribute to the PTM histone code. Moreover, protein phosphorylation influences protein-protein interactions.

Several studies have shown that phosphorylation and O-GlcNAcylation are interconnected by an extensive cross-talk: modulation of phospho-kinases induces an alteration of the O-GlcNAcylation sites on several proteins, and overexpression of the O-GlcNAc transferase OGT has as well an effect on the protein phosphorylation levels (Wang et al., 2008; Wang et al., 2007). In fact some Ser and Thr residues can be modified by both phosphorylation and O-GlcNAcylation (Fig. 2A), including Ser and Thr on the C-terminal domain of RNA Pol II, implying an antagonism between the two chemical groups (Cheng and Hart, 2001; Comer and Hart, 2001; Du et al., 2001; Kelly et al., 1993). Interestingly, phosphorylation and O-GlcNAcylation were found to be mutually exclusive also on proximal sites of several proteins, implying a steric/charge competition (Dias et al., 2009; Housley et al., 2008; Housley et al., 2009; Slawson et al., 2008; Yang et al., 2006).

1.2.3 O-GlcNAcylation (O-GlcNAc)

The covalent modification of proteins by a sugar is a PTM known as glycosylation. Protein glycosylation incudes a vast range of different glycosylation types, such as N-linked glycans (when linked to nitrogen atoms of amino acid residues), O-linked glycans (when instead linked to oxygen atoms of amino acid residues), glycosaminoglycans, glycosphingolipids, phosphorylated glycans, glycosylphosphatidylinositol (GPI) anchors and C-mannosylation of tryptophan residues.

Glycosylation protein features differ depending on the cellular localization of the protein substrate. Numerous proteins of the extracellular matrix can be covalently modified by

simple or more structured glycan chains, which are formed by the combination of a variety of monosaccharides. Glycosylation of extracellular proteins occurs in the endoplasmic reticulum (ER) and in the Golgi apparatus after protein synthesis, and it is regulated by glycosyltransferases and glycosidases. After glycosylation, the glycoproteins are exported to the extracellular surface where they are permanently modified (Reily et al., 2019). Glycans of the plasma membrane have disparate functions which are essential for the integration of the extracellular cellular signaling with the intracellular responses.

For long time protein glycosylation was thought to occur only on extracellular proteins, until 1984, when Torres and Hart identified a previously unknown form of intracellular glycosylation. In their study, they described how a large number of intracellular proteins can be modified by O-GlcNAcylation (Torres and Hart, 1984).

O-GlcNAc is a monosaccharide linked to the hydroxyl group of serines (S) and threonines (T) of nuclear and cytoplasmic proteins (Hart et al., 2007; Wells et al., 2003) (Fig. 2A). Contrary to extracellular glycosylation, intracellular O-GlcNAc occurs only in single monosaccharide residues and can be reversibly added and removed to proteins in response to specific cellular stimuli. Nuclear and cytoplasmic O-GlcNAc turnover is regulated by one single pair of enzymes: the O-GlcNAc transferase (OGT) and the O-GlcNAcase (OGA) (see paragraph 1.3) (Fig. 2A).

Intracellular O-GlcNAcylation levels are modulated by the metabolic status of the cell. 2-3% of the glucose intake, together with other metabolic molecules like amino acids (glutamine), lipids (acetyl-CoA) and nucleotides (UTP) are integrated by the hexosamine biosynthetic pathway (HBP) to generate the uridine diphosphate intermediate UDP-GlcNAc, which serves as a donor substrate for O-GlcNAc synthesis (Bond and Hanover, 2013) (Fig. 2B). Due to its direct connection to the HBP, O-GlcNAc is considered as a metabolic sensor of the cell.

Serines and threonines residues can in principle be modified by both O-GlcNAcylation and phosphorylation. The dynamic cross-talk between O-GlcNAcylation and phosphorylation has been widely documented (Wang et al., 2012; Wang et al., 2008; Wang et al., 2007) (see paragraph 1.2.2).

Proteins previously shown to be O-GlcNAcylated are involved in disparate biological processes, including metabolic signaling, stress response, structural regulation, transcription and translation (Wulff-Fuentes et al., 2021). However, due to the technical limitations of O-GlcNAc detection, glycoprotein dataset still remains incomplete.

Conversely to other protein modifications, the mapping O-GlcNAc proteome is particularly challenging due to the intrinsic chemical properties of the N-acetylglucosamine moiety. O-GlcNAc is an uncharged and substoichiometric modification and the conventional techniques are often not sensitive enough for its detection. Moreover, because O-GlcNAc is a reversible modification, it can be rapidly lost during the experimental procedures. Finally, the O-GlcNAc moiety is very labile and is not resistant to the mass spectrometer ionization, therefore it cannot be detected with the standard MS/MS approaches (Gambetta and Müller, 2015; Hart et al., 2007).

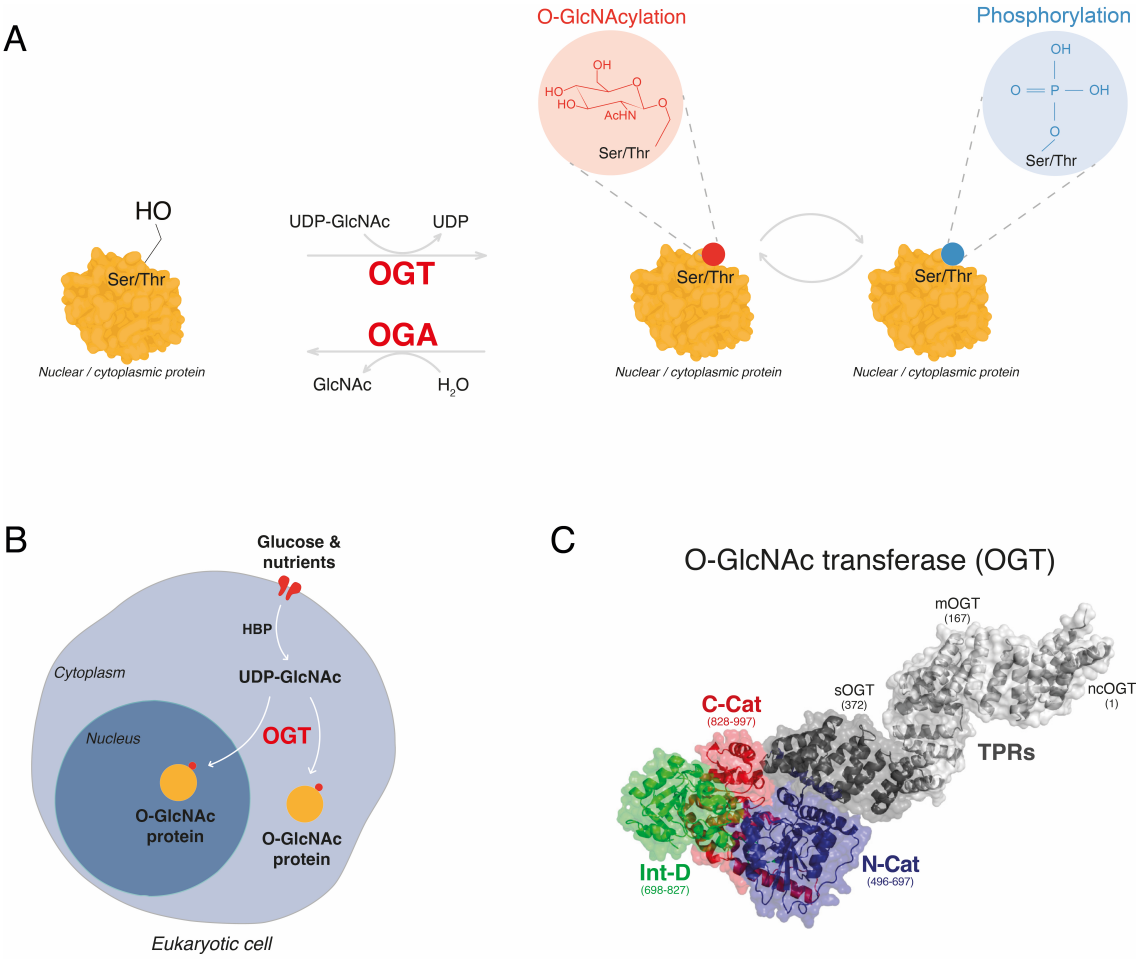

Fig. 2 (A) O-GlcNAc is added on Serine and/or Threonine residues of nuclear and cytoplasmic proteins by OGT. The same residues can be also phosphorylated, suggesting an interplay between these two modifications. OGA catalyzes the hydrolysis of GlcNAc. **(B)** The O-GlcNAc cycle. O-GlcNAc is synthetized from UDP-GlcNAc, which derives from the metabolic intermediates generated from the glucose intake and processed by the hexosamine biosynthetic pathway (HBP). **(C)** Model of human OGT constructed from the 11.5-TPR structure and the hOGT4.5 structure. The dark-gray TPRs are derived from the hOGT4.5 structure, whereas the light-gray TPRs are from the 11.5-TPR structure. The C-Cat, N-Cat and Int-D constitute the catalytic domain. hOGT structure figure adapted from (Janetzko and Walker, 2014).

1.3 OGA and OGT structure, isoforms and function

Albeit O-GlcNAcylation was detected on thousands of nuclear and cytoplasmic proteins, its turnover is regulated by one single pair of enzymes: O-GlcNAc transferase (OGT) catalyzes the O-GlcNAc addition to Ser/Thr hydroxyl moieties, while O-GlcNAcase (OGA) regulates its removal. Interestingly, OGT and OGA can also be O-GlcNAcylated, suggesting a cross-talk between the two enzymes (Nagel and Ball, 2014).

OGA, also known as meningioma expressed antigen 5 (MGEA5), has two main isoforms: the nucleocytoplasmic isoform (ncOGA) and the splicing variant short isoform (sOGA). The ncOGA is characterized by an N-terminal O-GlcNAc hydrolase domain and by a C-terminal histone acetyltransferase-like (HAT-like) domain, while the sOGA isoform, which was localized prevalently in the endoplasmic reticulum, lacks the HAT-like domain (Radermacher et al., 2014; Ruan et al., 2013). In fly, OGA is not essential for viability, while mice *Oga* loss-of-function mutants die shortly after birth (Radermacher et al., 2014).

The first crystal structure of near full-length human OGT was solved in 2011 (Lazarus et al., 2011). OGT is constituted by two main domains connected by a linker region: the C-terminal multidomain catalytic region and the N-terminal domain of several tetratricopeptide (TPR) repeats. The C-term catalytic domain is responsible for the glycosylation of the target peptides (Lazarus et al., 2013; Pravata et al., 2020) and it is characterized by three sub-domains, the amino (N)-terminal domain (N-Cat), the carboxyl (C)-terminal domain (C-Cat) and the intervening domain (Int-D) (Fig. 2C). Instead, the N-term TPR domain, was shown to regulate the protein-protein interaction of several OGT substrates (Levine et al., 2018), and to induce the protein cleavage of HCFC-1(Lazarus et al., 2013).
Four main OGT isoforms were identified in human and mouse. The most abundant OGT isoform is a 1046 aa protein (~116 kDa) present in both the nucleus and the cytoplasm, thus classified as nuclear-cytoplasmic OGT isoform (ncOGT). A second ncOGT isoform of 1036 aa (~116 kDa) was also identified, which is generated by a splicing variant of the same transcript (Kreppel et al., 1997; Love et al., 2003; Nolte and Müller, 2002). In addition, a shorter isoform (sOGT) of 665 aa (78 kDa) was also described at the transcript and protein level (Kreppel et al., 1997; Love et al., 2003; Shafi et al., 2000). Finally, a 920 aa (103 kDa) isoform containing a mitochondrial targeting sequence was

localized specifically in the mitochondria (Love et al., 2003). Whether the sOGT and the mOGT isoforms are transcribed from alternative coding sequence (CDS) start sites or originate from alternative splicing of the ncOGT isoform is still not understood. All three ncOGT, sOGT and mOGT isoforms are characterized by an identical C-term catalytic domain, however they differ by the length of their N-term TPR domain.

Ogt is a X-linked gene that is essential for mammalian cell viability. In mouse, *Ogt* mutagenesis was shown to be lethal at the cellular level (Shafi et al., 2000), and both zygotic and gametic mutants fails to develop beyond embryonic day E5.5 (O'Donnell et al., 2004). Due to the severe phenotype of *Ogt* deletion in mammals, it has not been possible yet to obtain an *Ogt* loss-of-function mouse allele.
In *Drosophila*, OGT is encoded by the *super sex combs* (*sxc*) gene and *sxc*-null larvae arrest their development at the pupal stage (Maria Cristina Gambetta, 2009). Fly's OGT was originally classified as a member of the Polycomb group (PcG) gene family due to the homeotic transformation phenotype resulting from its mutation, a characteristic Polycomb mutant phenotype. Polycomb proteins are essential transcriptional regulators which control the repression of *Drosophila* homeobox (HOX) genes during larvae embryonic development. The Polycomb Repressive Complexes (PRCs) requires a functional OGT to silence transcription: the lack of O-GlcNAcylation on S/T residues of the Polycomb protein Polyhomeotic (Ph) causes Ph protein misfolding and failed gene silencing (Gambetta and Müller, 2014). Importantly, there is no credible evidence that mammalian OGT associates to the Polycomb group protein complexes. Intriguingly, in mammals the Polycomb repressive complex 2 was shown to be required to maintain the physiological OGT and nuclear O-GlcNAc levels, however impaired levels of OGT does not have any measurable effect on Polycomb activity (Myers et al., 2011). While the modification of RNA Pol II by GlcNAc indicates a function in gene expression, it remains unclear how the O-GlcNAc modification regulates transcription; this long-standing enigma is the overarching question of my PhD thesis.

1.4 The role of O-GlcNAcylation in evolution

Intracellular O-GlcNAcylation has been documented in many species, from fungi to metazoans and plants. O-GlcNAc function was so far investigated by loss-of-function studies in mouse, fly, zebrafish and worm, while in humans O-GlcNAc disfunction was investigated in the context of intellectual disability and developmental delays.

OGT and OGA enzymes are highly conserved during evolution (Lubas et al., 1997). In *C. elegans* instead, where OGT has a lower sequence conservation, it has a role in the modulation of macronutrient storage and insulin pathway (Love and Hanover, 2005) and impaired OGT levels are responsible for metabolic disfunctions. OGT-null *C. elegans* mutants are viable and fertile (Love et al., 2010a) and decreased intracellular O-GlcNAc levels induce deregulated insulin-like signaling as well as altered lipid and carbohydrate storage. Due to the mild phenotype caused by loss of OGT in *C.elegans*, the function of O-GlcNAcylation in worm still remains an open question.

In mouse, fly and zebrafish *Ogt* has evolved to be an essential gene and is in fact necessary for the animal viability at different stages of embryonic development (Maria Cristina Gambetta, 2009; O'Donnell et al., 2004; Webster et al., 2009).

In zebrafish, embryos lacking functional OGT exhibit impaired epiboly movement, strong morphological phenotype and overall increased rates of cell death (Webster et al., 2009).

More extensive *Ogt* loss-of-function studies were carried out in *Drosophila* where mutant *Ogt* larvae arrest development at the end of embryogenesis and display a strong transformation of the body plan due to improper repression of developmental regulators. This phenotype resembles the morphogenetic PcG mutant phenotype, whereby HOX genes are expressed outside their expression domains. OGT was described to be also important for *Drosophila* glucose-insulin homeostasis and circadian rhythm (Kim et al., 2012; Sekine et al., 2010).

In mammals, *Ogt* loss-of-function studies were not possible due to *Ogt* requirement for embryonic survival and mammalian cell proliferation. To investigate the mechanisms behind *Ogt* lethality in early development, the authors of a recent study developed a new system to delete *Ogt* using *Ogt-floxed Mx1Cre* mice; a CRISPR-Cas9 screen for small-guide RNAs (sgRNAs) that rescued *Ogt*-induced proliferation arrest unveiled a hyperactivation of mTOR pathway, increased proteasomal activity and reduced mitochondrial fitness after conditional *Ogt's* loss (Li et al., 2023).

In human, impaired levels of O-GlcNAc are responsible for neurological and metabolic diseases, as Alzheimer, type II diabetes and cancer (Slawson et al., 2010; Yang and

Qian, 2017), suggesting a cardinal role for protein O-GlcNAcylation in cellular homeostasis. So far, several human OGT mutations have been mapped and all of them are associated with X-linked intellectual disability (Pravata et al., 2020; Pravata et al., 2019; Willems et al., 2017).

All together, these studies demonstrate that OGT is a conserved enzyme across evolution; the severe phenotype of *Ogt* null mutants in mammals suggests that OGT might have evolved in complex eukaryotes to regulate essential functions for the cellular viability.

1.5 Evidence of O-GlcNAcylation in transcriptional regulation

The O-GlcNAc modification was mapped on thousands of proteins, many of which are transcription factors and transcriptional regulators, implicating a role for this modification in regulating transcription (Wulff-Fuentes et al., 2021). In line with these data, genetic studies in fly showed that O-GlcNAc has an essential function in gene silencing during development (Gambetta and Müller, 2014). In fact, genome-wide profiling experiments of nuclear O-GlcNAc proteins in mammalian cell types identified O-GlcNAc modification to be enriched at transcriptional regulatory sites (Deplus et al., 2013). Moreover, biochemical purification of OGT complexes identified several transcriptional regulators to stably interact with OGT. Below I will report the most relevant genomic results that describe a role for O-GlcNAc in regulating transcription.

The first study of nuclear O-GlcNAcylation revealed a wide distribution of glycosylated proteins on *Drosophila* polytene chromosomes (Kelly and Hart, 1989). More recent genomic analyses of O-GlcNAc proteins on chromatin evidenced a correlation between O-GlcNAc and Polycomb proteins at Polycomb Response Elements (PREs), specific nucleosome-depleted regions near the TSSs (Maria Cristina Gambetta, 2009; Oktaba et al., 2008). Supporting results were characterized by genomic studies in *C. elegans*, where O-GlcNAc proteins were likewise profiled upstream the transcription start site (TSS) of multiple genes (Love et al., 2010a). These data indicate that O-GlcNAc might be directly involved in transcriptional regulation. In mammals, however, the nuclear function of O-GlcNAc modification is still ambiguous: OGT and O-GlcNAc chromatin profiling datasets in mouse cells are scarce, discordant and lack consensus on the genomic loci features regulated by glycosylated proteins. Despite OGT binds transiently to its substrates, two independent datasets have profiled OGT at TSS of CpG-rich

promoters of actively transcribed genes, where it colocalizes with TET2, TET3 and HCFC1 (Deplus et al., 2013; Vella et al., 2013) (for TETs and HCFC1 explanation see paragraphs 1.6.1 and 1.6.2).

1.6 The OGT interactome

More direct evidence indicating a direct role for O-GlcNAc in transcriptional regulation derives from immunoprecipitation experiments of OGT complexes from mammalian cells. The protein factors that were identified to strongly interact with OGT are HCFC1, TET1, TET2 and SIN3A.

- Immunoprecipitation experiment of flag-tagged biotinylated OGT from nuclear extract identified HCFC1, SIN3A, TET1 and TET2 to co-precipitate with OGT (Vella et al., 2013). In line with these results, immunoprecipitation experiments of TET complex from mammalian cells identified OGT among the most stable TETs interactors.
- In agreement with HCFC1 co-immunoprecipitation with OGT, HCFC1 complex immunoprecipitation identified OGT as a strong HCFC1 partner. Deeper investigations suggest that the interaction of OGT with TETs and HCFC1 occurs in two distinct complexes (Wysocka et al., 2003), and it has been proposed that TETs mediate the O-GlcNAcylation of several OGT substrates, including HCFC1 (Chen et al., 2012; Deplus et al., 2013).
- In addition, OGT was independently co-purified with BRCA1 associated protein 1 (BAP1) in FLAG-BAP1 complex immunoprecipitation. Interestingly, HCFC1 was also co-purified together with OGT in the same protein complex.

I will hereafter introduce the function of the main OGT interactors, relevant for the understanding of this study, and outline their role in complex with OGT.

1.6.1 The TET family dioxygenases

The Ten-Eleven Translocation (TET) family is composed by TET1, TET2 and TET3 proteins which are Fe^{2+} and 2-oxoglutarate-dependent dioxygenases. The TET proteins catalytic activity is the oxidation of the DNA 5-methylcytosine (5mC) to 5-

hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Tahiliani et al., 2009). TETs are very conserved enzymes and are present in all metazoans, including Drosophila (Skvortsova et al., 2022) which is devoid of 5mC indicating an 5mC independent function.

TET1 and TET3 proteins are characterized by an N-terminal CXXC domain which recognizes non-methylated CpG rich sequences (Lee et al., 2001), and by a C-terminal catalytic domain. For TET2 instead, no DNA binding domain was identified and its recruitment to DNA is mediated through its binding to CXXC4 (Ko et al., 2013). TET1 and TET3 proteins are prevalently enriched at CpG rich promoters (Ko et al., 2013; Xu et al., 2011). TET1 knockdown was shown to induce gene expression activation of TET1 enriched genes, while TET2 seems to rather be a positive regulator of gene expression as its knockdown induces transcriptional repression (Chen et al., 2012; Williams et al., 2011; Wu et al., 2011; Xu et al., 2011).

TET proteins are active during embryonic development and play a minor role in regulating gene expression during DNA methylation waves of embryonic development (Dai et al., 2016). Moreover, TETs share catalytic activity but have differences in their expression level and in their recruitment mechanisms. TET1 and TET2 are highly expressed in the blastocyst inner cell mass (ICM); while TET1 gets gradually downregulated, TET2 remains highly expressed. TET3, instead, is highly expressed in mouse oocyte, in the early preimplantation embryo and during the differentiation of the three germ layers, but no in the ICM (Dawlaty et al., 2013).

Genetic knockout mouse models were instrumental to investigate the TET's function in mammals: while *Tet3* knockout leads to neonatal lethality (Gu et al., 2011; Kang et al., 2015), *Tet1* and *Tet2* knockouts do not show any developmental phenotype (Dawlaty et al., 2013; Moran-Crusio et al., 2011; Quivoron et al., 2011), indicating that TET1 and TET2 cannot compensate the loss of TET3. Interestingly, simultaneous loss of *Tet1* and *Tet2* has a severe phenotype on newborn pups (Dawlaty et al., 2013).

Genome profiling studies of TETs and OGT have shown that TET proteins recruit OGT at promoters, as TETs depletion leads to a decrease in OGT localization (Vella et al., 2013). The model resulting from these experiments is a recruitment of the OGT-TET1 complex to CpG-rich promoter through TET1's CXXC motif that binds to unmethylated CpG dinucleotides. OGT instead does not seem to have any role in TET chromatin localization, even though alterations of ~20% of 5hmC levels were detected upon loss of OGT. Despite the biological function of TETs and OGT not being completely understood, these studies suggest that TETs and OGT could be essential for their reciprocal activity

and that TETs might be important for OGT chromatin localization (Chen et al., 2012; Vella et al., 2013).

1.6.2 The Host Cell Factor 1 (HCFC1)

HCFC1 is a transcriptional coregulator involved in the control of cell cycle (Mahajan and Wilson, 2000) which was firstly described as a host-cell factor for human herpes simplex virus infection (Wysocka and Herr, 2003). In vertebrates, HCFC1 is synthetized as a long precursor protein characterized by a central region of six series of 26 threonine-rich amino acid repeats known as $HCFC1_{PRO}$ repeats. During the protein maturation process, this region gets cleaved and generates two non-covalently associated N-term and C-term subunits. HCFC1 promotes the cell cycle progression via these N-term and C-term subunits which regulate the exit from mitosis and the passage through the G(1) phase of the cell (Julien, 2003).

HCFC1 is described as the most abundant OGT protein partner; the nature of OGT and HCFC1 complex was resolved by crystallography studies which shown that their interaction occurs via the threonine-rich HCFC1 stretch and the asparagine rich-region of OGT TPR (Lazarus et al., 2013). Further studies have determined that HCFC1 is not only a strong OGT partner, but it is also a substrate that is highly O-GlcNAcylated at the level of its cleavage site (Capotosti et al., 2011; Wilson et al., 1993). OGT is also essential for the proteolytic maturation of HCFC1; it was hypothesized that OGT could either induce the cleavage itself, albeit no protease-like site was identified, or could act as a co-protease in collaboration with the $HCFC1_{PRO}$ subunit. In HeLa cells, a strong phenotype of binucleated cells was described when the HCFC1 proteolytic repeats were mutated (Capotosti et al., 2011).

In *Drosophila*, HCFC1 is also O-GlcNAcylated but it gets cleaved by the protease TASPASE1. Interestingly, impaired O-GlcNAcylation levels induce *Drosophila's* HCFC1 to form large molecular aggregates, indicating that HCFC1 O-GlcNAcylation is important for its stability (Gambetta and Müller, 2014).

Because nutrient intake affects cell proliferation through the cell cycle, it has been proposed that OGT could sense the cellular glucose levels and respond as a sensor of the cellular metabolic state by activating HCFC1 and regulating the cell cycle progression (Love et al., 2010b).

1.6.3 The Polycomb Repressive Complex (PRC) and the Polycomb Repressive Deubiquitinase (PR-DUB) complex

Polycomb group (PcG) proteins are a group of chromatin modifiers described for the first time in *Drosophila*, when in 1947 the first Polycomb-related phenotype was reported (Lewis and Mislove, 1947). PcG proteins are conserved in all animals and are distinguished in two main multi-protein complexes: Polycomb repressive complex 1 (PRC1) and Polycomb repressive complex 2 (PRC2). Both PRC1 and PRC2 are associated with transcriptional repression, however their mechanism of silencing is different:

- The Polycom Repressive Complex 1 (PRC1) has a E3 ubiquitin ligase activity and is responsible for the mono-ubiquitinylation of H2A Lys118 (*Drosophila*) and Lys119 (mammals). PRC1 can be distinguished in the canonical and non-canonical complexes; both are characterized by the subunit RING1A/RING1B, responsible for its catalytic activity.
- The Polycomb Repressive Complex 2 (PRC2) has a methyltransferase activity and is responsible for the mono/di/tri-methylation of H3 Lys27. The PRC2 complex is composed by the following subunits: the Embryonic Ectoderm Development (EED), the Suppressor of Zeste 12 (SUZ12), the Jarid2/Jumonji (Jarid2) and the Enhancer of Zeste Homologue 2 (EZH2) or its paralogue EZH1. PcG proteins are highly expressed during embryonic development and have a fundamental role in keeping temporally restricted the transcriptional activation of developmental genes. PRC1 and PRC2 are active at silent CpG rich promoters, which are characterized by H3K27me3, H2AK119ub and PcG proteins (Ku et al., 2008; Mikkelsen et al., 2007). Despite H3K27me3 and H2AK119ub have catalytic-independent functions (Blackledge et al., 2020; Eskeland et al., 2010; Francis et al., 2004; Hansen et al., 2008; Pengelly et al., 2013; Pengelly et al., 2015; Tamburri et al., 2020), these two histone marks are both necessary to repress gene expression of developmental genes and safeguard the cell identity. As mentioned in paragraph 1.4, the most relevant genetic evidence for *Ogt* function in eukaryotes were obtained from knockout studies in *Drosophila*, where the loss of *Ogt* results in a specific Polycomb phenotype. In fly, *Ogt* is required for Polycomb repression through the O-GlcNAcylation of the PRC1 core subunit Polyhomeotic (Ph) domain, essential to prevent Ph to form high molecular weight aggregates. Ph's O-GlcNAc occurs in the low-complexity S/T Ph stretch; notably,

mutation of this sequence reproduces the *Ogt* null mutant. O-GlcNAc chromatin signal is significantly enriched at Polycomb responsive Elements (PRE), which are regulatory loci necessary to mediate Polycomb target gene silencing. Interestingly, PREs do not seem to exist in mammals, where the Polycomb binding domain is characterized by hyper-conserved CpG islands (Bauer et al., 2016). These findings demonstrate that, in fly, the main function of *Ogt* is to regulate Polycomb repression (Gambetta and Müller, 2014; Gambetta et al., 2009; Maria Cristina Gambetta, 2009).

- The Polycomb Repressive Deubiquitinase (PR-DUB) complex consists of the deubiquitinase Breast cancer type 1 susceptibility protein (BRCA1)-Associated Protein-1 (BAP1) and of one of the three regulatory ASX-Like proteins (ASXL1-3). The C-terminal domain of BAP1 is a hydroxylase (UCH) domain (Conway et al., 2021) and its main function is to constrain pervasive H2AK119ub1 throughout the genome (Fursova et al., 2021). PR-DUB prevalently binds to gene promoters and distal regulatory elements; the deubiquitination activity of the PR-DUB complex prevents the diffuse accumulation of H2AK119ub1 and H3K27me3, implying an antagonistic role between PR-DUB and PRC1. Nonetheless, the function of PR-DUB and its reciprocal interplay with PRC1/PRC2 seems to be more complex, as PR-DUB was shown to promote and at the same time limit Polycomb activity; BAP1 is in fact essential to maintain the PRC2 and H3K27me3 stability at their genomic loci (Conway et al., 2021).
  Two independent studies of the Polycomb protein interactors have identified OGT as a stable partner of the PR-DUB complex (Conway et al., 2021; Hauri et al., 2016), however the function of this interaction remains unknown.

1.7 RNA Polymerase II and its role transcriptional regulation

RNA Polymerase II (RNA Pol II) is a 12-subunit multiprotein complex that catalyzes the transcription of DNA into the mRNA precursors of protein-coding-genes, non-coding RNAs, small nuclear RNA (snRNA), small nucleolar RNAs (snoRNAs) and microRNAs. The largest subunit of RNA Pol II is RBP1, a DNA-directed RNA polymerase encoded by the gene *Polr2a*. RBP1 is characterized by an N-terminal domain, responsible for the catalytic function of RNA Pol II, and by a long protruding unstructured C-terminal domain (CTD) (Fig. 3A).

In mouse and human, the CTD sequence consists of 52 imperfect tandem repeats of the heptapeptide $Y_1S_2P_3T_4S_5P_6S_7$. The CTD sequence undergoes a dynamic cycle of several PTMs that precisely coordinate RNA Pol II activity and the transcriptional progression (Eick and Geyer, 2013). Phosphorylation is the best characterized modification of RNA Pol II CTD: phosphorylation of Serines at positions 2 (Ser2), 5 (Ser5) and 7 (Ser7), Tyrosine at position 1 (Tyr1) and Threonine at position 4 (Thr4) orchestrate transcription by promoting and/or inhibiting the recruitment of specific transcription factors, histone modifiers and chromatin remodeling complexes (Fig. 3B) (Akhtar et al., 2009; Chapman et al., 2005; Glover-Cutter et al., 2009; Hsin, 2011; Liao et al., 1995; Phatnani and Greenleaf, 2006; Sun et al., 1998).

Transcription of active genes occurs in controlled sequence of steps that coordinate the mRNA production. RNA Pol II is recruited to gene promoters in a hypo-phosphorylated state; transcription of the mRNA molecule starts with phosphorylation of Ser5 by the TFIIH transcription factor (Komarnitsky et al., 2000) and by the recruitment of the 5' RNA capping machinery. At transcription initiation RNA Pol II is still unstable and can abort within the first 2-10 nucleotides.

Promoter's escape is followed by RNA Pol II entrance in the transcription pause site, which occurs between 10 and 50 nucleotides from the transcription start site (TSS) (Brookes and Pombo, 2009) where RNA Pol II remains stably associated with the nascent RNA. Promoter proximal pausing is an important check-point during transcription progression as it represents a mechanism that higher eukaryotes have adopted to tune transcription with a precise timing in response to developmental and environmental stimuli. Here, Ser5 is hyper-phosphorylated and RNA Pol II is assembled in complex with the pre-initiation complex (PIC), which comprises the general transcription factors necessary to initiate transcription. RNA Pol II is predominantly distributed at the pause site in the proximity of the TSS, at the 5' of the gene body

sequence: the ratio of promoter to gene body of RNA Pol II density is defined with the pausing index.

The entrance into the productive elongation phase requires the phosphorylation of Ser2 by the factor P-TEFb and the recruitment of the elongation-specific (P-TEFb)-containing elongation complexes. Ser2P promotes chromatin remodeling by recruiting H3K36 histone methyltransferases (HMTs), which generate a chromatin environment compatible with transcriptional elongation (Krogan et al., 2003; Li et al., 2005). SerP2-mediated transcriptional elongation is also characterized by the presence of splicing and polyadenylation factors and by H3K36me3.

Serine at position 7 can be also phosphorylated by TFIIH, however its regulatory function is less understood. Ser7P was mapped at promoter and coding regions (Chapman et al., 2007) and was shown to be essential for small nuclear RNAs (Egloff et al., 2007) and for early stages of transcription (Akhtar et al., 2009).

When transcription of the mRNA molecule is completed, RNA Pol II is released from the DNA template. RNA Pol II release is preceded by a series of events that coordinates the release of the RNA transcript and of the transcription complexes (Richard and Manley, 2009).

At transcription termination, a series of termination factors are recruited on RNA Pol II. The CTD modification that defines transcriptional termination is the hyperphosphorylation of Ser2 by the Ctk1 kinase. Threonine 4 participates as well in coordinating the transcriptional progression and termination: Thr4 gets phosphorylated by the Polo-like kinase (Plk3) downstream of the poly-A site, while no Thr4P was detected at the TSS and at the gene body. The function of Thr4P is still not clear, and was proposed to either inhibit the binding of factors of the unphosphorylated Thr4, or to trigger the binding of other essential factors (Hintermair et al., 2012).

Tyr1 was also described to be phosphorylated in human and yeast during transcription (Baskaran et al., 1993; Mayer et al., 2012). TyrP correlates with SerP2 and was shown to be a negative regulator of transcription termination as it counteracts the binding of the termination factors (Mayer et al., 2012).

In most cases, transcription of mRNA-coding genes generates stable mRNA transcript that will be processed by the ribosomes for protein translation in the cytoplasm. In some cases however, non-coding RNAs (ncRNAs) can be also produced and degraded at the step of transcription termination.

1.8 RNA Pol II is a glycoprotein

In 1993, William G. Kelly et al. described for the first time the presence of O-GlcNAc modification on the CTD of RNA Polymerase II. The authors demonstrated the existence of two different states of RNA Pol II, the phosphorylated and the O-GlcNAcylated forms, and showed these two modifications to be mutually exclusive (Kelly et al., 1993).
Using Edman degradation in calf thymus, O-GlcNAcylation was detected on Thr4 and Ser5 residues (Fig. 3A), while no O-GlcNAc was detected on the elongation-specific phosphorylation form of RNA Pol II. Additional studies based on Serine-to-Alanine substitution on human CTD reported Ser5 and Ser7 to be necessary for OGT activity (Ranuncolo et al., 2012), while Thr4 function for OGT activity has not been tested yet. Notably, the authors did not verify the Ser5 and Ser7 O-GlcNAc state prior to alanine substitution. Moreover, while in Kelly *et al*. study the mapping was performed *in vivo* on non-modified CTD, Ser-to-Ala substitution was performed *in vitro*, and OGT activity was tested by incubating GST-CTD with OGT and UDP-GlcNAc. More recent *in vitro* studies revealed how inhibition of OGT activity prevents RNA Pol II recruitment to the promoter region (Lewis et al., 2016) and that *in vitro* inhibition of OGT and OGA interferes with the preinitiation complex (PIC) assembly, dispensable for transcription initiation (Ranuncolo et al., 2012).
Despite these investigations, the role of O-GlcNAc in transcriptional control and its functional relationship with phosphorylation still remains unresolved.

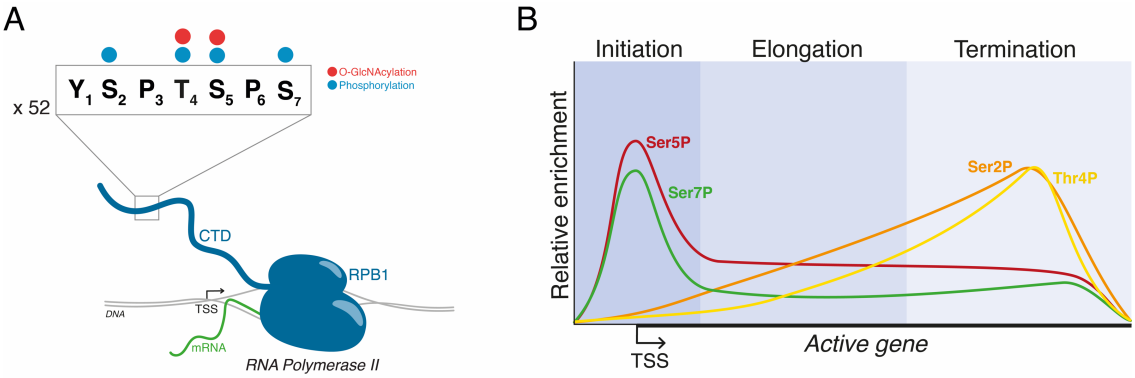

Fig. 3 (A) RNA Polymerase II transcribes the coding genes from the TSS to generate the mRNA molecule. RNA Pol II is characterized by the large RPB1 subunit and by a long protruding and flexible C-terminal domain (CTD) composed by 52 repeats of the $Y_1S_2P_3T_4S_5P_6S_7$ heptapeptide sequence. The CTD was described to be phosphorylated and O-GlcNAcylated on Ser and Thr residues. **(B)** The CTD undergoes a dynamic cycle of Serines and Threonines phosphorylation which fine-tune the transcriptional progression.

2. PhD project aims

My PhD project aims to bring new insights into the role of O-GlcNAcylation in regulating mammalian gene expression. To this end, I used mouse embryonic stem cells (mESCs) and neuronal progenitor cells (NPCs) to probe chromatin O-GlcNAcylation dynamics during mammalian embryonic development by means of genomic and proteomic approaches.

1. *Genomic approach*. To characterize the patterns of O-GlcNAc modified protein bound to the genome, I profiled the occupancy of O-GlcNAc proteins using chromatin profiling approaches combined by next-generation sequencing (NGS). The pan-GlcNAc profiling will reveal the genome-wide DNA binding sites of O-GlcNAcylated proteins.

2. *Proteomic approach*. To characterize the identity of chromatin GlcNAc proteins, I performed biochemical isolation of nuclear O-GlcNAcylated proteins followed by MS/MS to characterize the O-GlcNAc proteome and identify the specific O-GlcNAcylated residues. This approach will provide a complete dataset of nuclear O-GlcNAc proteins and will reveal the protein sequence domains mostly enriched by GlcNAc modification.
   In parallel, to characterize the main cellular pathways regulated by O-GlcNAc modification and by OGT enzyme, I investigated the OGT interactome by native protein complex immunoprecipitation followed by quantitative mass spectrometry.

3. *Functional approach*. To specifically remove O-GlcNAc at promoters without interfering with OGT's functions, I targeted a bacterial homologue of OGA, namely BtGH84, by phusion with a CXXC-PHD domain. This system will enable me to investigate how perturbation of O-GlcNAc at proteins bound to CpG-rich promoters affect their transcription.

3. Material and Methods

3.1 Nucleic acids quantification

Nucleic acid concentration was quantified using the Qubit DNA Kit and the RNA Assay Kit (Thermo Fisher Scientific), Broad Range or High Sensitivity depending on the input material amount, using a Qubit 3 Fluorometer (Invitrogen) according to manufacturer's instructions.

3.2 Gel electrophoresis

DNA molecules were size separated with horizontal electrophoresis by agarose gels at a concentration between 0.8-3% in 1X TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.3) or 1X TBE buffer diluted in H_2O from 10X TBE buffer (0.89M Trizma®, 0.02M EDTA, pH 8.3 with boric acid) supplemented with EtBr (Sigma Aldrich #E1510). Samples were loaded with 6X loading dye diluted to 1X (Biotech Rabbit #BR0800301) and 100 bp ladder (New England Laboratories #N3231L) or 1 kb ladder (New England Laboratories #N3232L) were used for fragment size estimation. Gel imaging was performed using ChemiDoc imaging system (Bio-Rad).

3.3 Plasmid construction and cloning

pETM14 plasmid vector used to clone pA/G MNase was kingly provided by Jamie Hackett lab. Purified pETM14 vector was digested with HindIII restriction enzyme (New England Laboratories #R3104L) according to manufacturer's instructions. Protein A/G (pA/G) and protein L (pL) sequence were ordered as gBlocks from Integrated DNA technologies (IDT) and assembled with the digested pETM14 vector using the NEBuilder HiFi DNA Assembly Master Mix (New England Laboratories #E2621L). Purified plasmids from transformed bacteria were sent to PEPCore EMBL Heidelberg Facility for protein expression and purification from Bl21 DE3 competent cells.

gp41-BtGH84-3XNLS-SPOT and *gp41-BtGH84*^{D242A}*-3XNLS-SPOT* plasmids were engineered using the *p2Lox* vector obtained from *p2Lox-BtGH84-TRIM28* (Boulard et al., 2020) plasmid by enzymatic digestion with NheI-HF (#R3131S) and NotI-HF (#R3189S) restriction enzymes according to manufacturer's instructions, and the PCR amplificons of *3XNLS-BtGH84* and *3XNLS-BtGH84*^{D242A} sequences from *2Plox*-3X*NLS-BtGH84* and *2Plox-3XNLS-BtGH84*^{D242A} plasmids. p2Lox vector and *BtGH84*/*BtGH84*^{D242A} amplified sequences were assembled using the NEBuilder HiFi DNA Assembly reaction (New England Laboratories #E2621). The primers for cell lines engineering and for PCR genotyping are listed in Table 1.

3.4 Bacteria transformation

Chemically competent DH5a *Escherichia Coli* (*E.coli*) bacteria were transformed with 2 µl of HiFi reaction product and incubated on ice for 30 min. Heat shocked reaction was performed for 30 seconds at 42°C and, immediately after, bacteria were returned back to ice for 2 min and resuspended in 950 µl of room temperature S.O.C rescue medium. Bacteria were incubated at 37°C for 1h for recovering under agitation and plated on L-agar plates with the appropriate antibiotic selection (30 µg/mL Kanamycin or 100 µg/mL Ampicillin). Bacteria were grown overnight at 37°C.

3.5 Bacteria culture and plasmid purification

Transformed and selected bacterial colonies were inoculated under agitation at 37°C for 8h in Luria-Bertani (LB) culture media with the appropriate antibiotic selection. 100 µl of bacterial culture was inoculated in 200 mL of LB under appropriate antibiotic selection at 37°C from 12h to maximum 16h. Plasmid extraction and purification was performed with either MidiPrep (Qiagen #27106) or MaxiPrep (Qiagen #12863) Kits. Correct plasmid assembly and sequence of screened colonies were verified by proper restriction enzyme digestion and by Sanger sequencing.

3.6 Generation of conditional gene expression cell lines with *A2loxP-Cre* system

The cDNA encoding for O-GlcNAc hydrolase (*Oga*, *Btgh84*) from Bacteroides and for catalytically inactive ("dead") *Btgh84*^{D242A} (Boulard et al., 2020) fused to Nuclear Localization Sequence (NLS) and CXXC domain were designed and engineered by Yulia Ermakova by cloning the cDNA of each construct in the exchange cassette vector *2loxP* as previously described (Iacovino et al., 2011). To engineer the *2Plox*-3X*NLS-BtGH84*, *2Plox-3XNLS-BtGH84*^{D242A}, *2Plox*-CXXC-*BtGH84* and *2Plox-CXXC-BtGH84*^{D242A} conditional gene expression cell lines, *A2loxP-Cre* cells were grown in T2i on gelatin-coated plates and *Cre* expression was induced by exposing the cells to 1 *μ*g/ml of Doxycycline (Sigma-Aldrich #3447) for 24h. *A2loxP-Cre* cells were nucleofected using the P3 Primary Cell 4D X LONZA Kit (LONZA #V4XP-3024) with the specific cloned plasmid and exposed to 350 *μ*g/ml G418 selection (Sigma Andrich #G8168-50ML) for 10 days. Positive clones were selected after single clone isolation and genotyping for the *knock-in* sequence.

3.7 Mouse ES cells culture

Murine E14 or A9 wild type (WT) embryonic stem (ES) cell lines and the engineered ES cell lines were maintained in humidified atmosphere at 37°C and 5% CO_2 and grown on gelatin-coated plates in different media: t2i/L media consisted in NDiff (N2B27) (Takara #Y40002), titrated 2i (0.2 µM PD0325901 and 3 µM CHIR 99021, Sigma-Aldrich), 1,000 U/ml leukaemia inhibitory factor (LIF), 1% FBS (Gibco), 1% penicillin streptavidin (Thermo Fisher Scientific #15140122); 2i/L media consisted in 50% NDiff (N2B27) (Takara #Y40002), 50% DMEM/F-12 (Gibco #11320033), 1% FBS (Gibco), 1 mM Sodium Pyruvate (Sigma-Aldrich #S8636), 0.1 mM Non-Essential AAs (Sigma-Aldrich #M7145), 2 mM L-Glutamine (Sigma-Aldrich #G7513), 1% penicillin streptavidin (Thermo Fisher Scientific #15140122), 0.1 mM β-Mercaptoethanol (Thermo Fisher Scientific #31350010), full 2i (1 µM PD0325901 and 3 µM CHIR 99021, Sigma-Aldrich), 1,000 U/ml leukaemia inhibitory factor (LIF); KO DMEM (Thermo Fisher Scientific #10829018) or DMEM (Thermo Fisher Scientific #10564011), 15% FBS (Gibco), 0.1 mM Non-Essential AAs (Sigma #M7145), 2 mM L-Glutamine (Sigma-Aldrich #G7513), 1% penicillin streptavidin (Thermo Fisher Scientific #15140122), 0.1 mM β-Mercaptoethanol

(Thermo Fisher Scientific #31350010), 1,000 U/ml leukaemia inhibitory factor (LIF), titrated 2i (0.2 µM PD0325901 and 3 µM CHIR 99021, Sigma-Aldrich). Medium was changed daily and cells were passaged every 2-3 days via dissociation in TrypLE Express Enzyme (Thermo Fisher Scientific #12604039). Mycoplasma-free status of the cell cultures was verified periodically.

Human colon adenocarcinoma DLD-1 cells expressing OsTIR and with a cassette encoding mini-AID (mAID) and fluorescent protein mClover (mAID+mClover) at the initiation site of the endogenous *Rpb1* gene locus (POLR2A)(Nagashima et al., 2019) were grown in RPMI 1640 medium (Thermo Fisher Scientific #A1049101) supplemented with 10% FBS (Gibco).

3.8 mES cells differentiation to neural progenitors (NPCs)

Differentiation of mES cells to neuronal progenitor cells (NPCs) was performed as previously described (Petracovici and Bonasio, 2021). To obtain NPCs, ESCs were seeded in Basal Medium supplemented with 2i/LIF at a density of 20,000 cells/cm^2 on gelatinized-coated plates and cultured for 12h to allow for cell attachment. Differentiation was induced by replacing the medium with Basal Medium supplemented with 10 ng/mL of recombinant human bFGF (R&D #233-FB); cells were maintained for 72h with daily medium exchange. Medium was then replaced with Basal Medium supplemented with 500 nM SAG (Sigma #566661); cells were maintained for 48h with daily medium exchange until NPCs collection.

3.9 CRISPR-Cas9 mediated genome editing

To generate the mES *RNA Pol II*$^{HA\text{-}SPOT}$, *RNA Pol II*$^{HA\text{-}MoonTag}$, *Ogt*$^{N\text{-}dTAG}$ and *Ogt* $^{N\text{-}FLAG,\ C\text{-}HA}$ cell lines, *A2loxP-Cre* or A9 mES cell line were engineered using the Crispr/CAS9 technology by the Gene Editing and Embryology Facility (GEEF) in EMBL Rome. The guide RNA with the interested *knock-in* sequence was designed with homology arms for the interested endogenous gene and nucleofected in mES cells grown in T2i medium on gelatin-coated plates, together with the crRNA, tracrRNA and purified Cas9 (PEPCore Facility EMBL Heidelberg). Single clones were isolated, expanded in T2i medium and screened by genotyping for positive clones for the interested *knock-in* sequence.

3.10 *Ogt* siRNA and Reverse RNA transfection

A9 and E14 WT ES cell lines were grown in Serum/LIF and 2i media without penicillin streptavidin on a 6 cm plate surface. Before cell transfection, 5 μl of Lipofectamine2000 (Invitrogen #11668-019) were mixed with 250 μl of Opti MEM reduced Serum Medium (Gibco #51985-034) and incubated for 5 min at RT. Meanwhile, 10 μl of 10 μM OGT siRNA (Santa Cruz Biotechnology #sc-40781) were mixed with 250 μl of Opti-MEM and then combined with the Lipofectamine2000-Opti-MEM mix to make a transfection mix and incubated for 20 min at RT. Cells were detached with TrypLE and counted to plate 300 x10^5 cells/well. 500 μl of transfection mixture were mixed with the cell suspension and the solution was transferred on a pre-gelatinized 6 cm plate. In parallel, a negative control was prepared following the same conditions but using a non-targeting siRNA (Santa Cruz Biotechnology #sc-37007), and a negative control reaction without siRNA was also processed. Growing medium was changed 12h after the experiment and the cells collected 48h after siRNA transfection.

3.11 Protein extraction and western blotting

Cells were detached with TrypLE (Thermo Fisher Scientific #12604039), washed with PBS 1X and pelleted for 5 min at 1,000 RPM. For nuclear proteins, nuclei were extracted by resuspending the cell pellet in buffer A (10 mM HEPES pH 7.65, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 1X cOmplete Mini EDTA-free protease inhibitors (Roche)) then incubated for 15 min at 4°C under gentle rotation. Nuclei were released from cells with the Dounce homonizer, and pelleted at 250 g for 5 min at 4°C. Nuclei were washed in buffer N (15 mM HEPES pH 7.65, 10 mM $MgCl_2$, 0.5 mM DTT, 250 mM sucrose, 1X cOmplete Mini EDTA-free protease inhibitors (Roche)) and pelleted at 2,800 g for 10 min at 4°C.

Cells or nuclei were lysed by resuspending the pellets in RIPA Buffer (150 mM NaCl, 1% IGEPAL, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM tris, pH 8.0) containing 1X cOmplete Mini EDTA-free protease inhibitors (Roche) and incubated 5 min on ice. Genomic DNA was digested using Universal Pierce Nuclease 250 U (Thermo Fisher Scientific #88702) at 37° C for 5 min and insoluble chromatin removed by centrifugating the samples for 1 min at 4°C at 18,000 g. Protein concentration was determined using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific #23227). Equal amount of

proteins was loaded on 4-20% Tris-Glycine gel (NuPAGE), or 4-12% Bis-Tris gel (NuPAGE) for protein detection. Protein dry transfer was performed on 0.2 μm nitrocellulose membranes (BioRad #1704159) using a Trans-Blot Turbo Transfer System system (Bio-Rad). Membranes were blocked for 1h in Saturating Buffer (5% BSA, 0.1% Tween-20, PBS 1X) and incubated overnight at 4°C in 1st Antibody Buffer (5% BSA, 0.1% Tween-20, PBS 1X) with primary antibodies (Table 3), washed three times for 5 min at RT in Washing Buffer I (0.5% Triton X-100, 0.5 M NaCl, PBS 1X), one time for 10 min at RT in Washing Buffer II (0.5 M NaCl, PBS 1X), one time for 15 min at RT in PBS 1X and incubated with HRP-conjugated secondary antibodies for 1h at room temperature. Membranes were washed again as previously described and signal was revealed using the ECL-Prime Western Blot System (Sigma #RPN2232) on an Amersham ImageQuant 800 system.

3.12 OGT native complexes immunoprecipitation (IP)

Immunoprecipitation of native OGT complexes from fractionated cytoplasm, nuclear and insoluble chromatin fractions was performed from WT A9 and *Ogt*$^{N\text{-}FLAG,C\text{-}HA}$ mES cell lines grown in DMEM Serum/LIF T2i medium. For each experiment, 100 x10^6 cells were harvested, washed with 20 mL PBS 1X and collected at 1,500 g for 5 min. Native nuclear protein-complexes extraction was performed using the Dignam et al method (Dignam et al., 1983): cell pellets were resuspended in 5 mL of Buffer A (10 mM HEPES pH 7.65, 1 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 1X cOmplete Mini EDTA-free Protease Inhibitors (Roche)) and incubated on ice for 15 min. Nuclei were extracted using the Dounce homogenizer and pelleted at 300 g for 10 min at 4°C. Cytoplasmic and nuclear fractions were subsequently separated; cytoplasm was centrifuged at 16,000 g for 5 min at 4°C to isolate the insoluble cytoplasmic fraction, salt concentration was adjusted to 250 mM by slowly adding 5 M NaCl. The extract was incubated for 5 min at RT with 1 μl of Pierce Universal Nuclease 250 U to digest DNA and RNA (Thermo Fisher Scientific #88702). Nuclear fraction was washed twice with Buffer N (10 mM HEPES pH 7.65, 1 mM $MgCl_2$, 0.5 mM DTT, 1X cOmplete Mini EDTA-free Protease Inhibitors (Roche), 250 mM sucrose), centrifuged at 2,800 g for 10 min at 4°C and resuspended in 500 μl of Buffer B supplemented with 250 mM NaCl (20 mM HEPES pH 7.65, 25% Glycerol, 5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 1X cOmplete Mini EDTA-free protease Inhibitors (Roche), 0.05% IGEPAL, 250mM NaCl). Salt concentration was increased to 320 mM with 5 M

NaCl and incubated for 3h at 4°C under gentle rotation to release the protein complexes through nuclear pores. Soluble nuclear extract was isolated by pelleting the sample at 13,000 g for 10 min at 4°C. Salt concentration of the soluble nuclear fraction was reduced to 250 mM with Buffer B without salt (20 mM HEPES pH 7.65, 25 % Glycerol, 5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 DTT, 0.05% IGEPAL, 1X cOmplete Mini EDTA-Free Protease Inhibitors (Roche)). Insoluble nuclear fraction was resuspended in Buffer B without salt and insoluble chromatin was sonicated using Covaris 220 system. 50 µl of Pierce™ Anti-HA Magnetic Beads or Anti-FLAG® M2 Affinity Gel were used for each respective sample: Anti-HA magnetic beads/Anti-FLAG affinity gel were washed with Beads Washing Buffer (10 mM HEPES pH 7.65, 0.01% IGEPAL, 0.01% BSA) and resuspended in Buffer B supplemented with 250 mM NaCl. Anti-HA Magnetic Beads or Anti-FLAG affinity gel were incubated with protein extracts for 2h at 4°C, washed 5 times for 5 min with 1 mL of 250 mM NaCl, eluted twice in 100 µl of SDS-Elution Buffer (50 mM HEPES pH 7.65, 5% Glycerol, 1% SDS, 1 mM DTT, 50 mM NaCl) and boiled at 95 °C for 5 min. Immunoprecipitated protein samples were analysed by total protein staining using SYPRO Ruby Protein Gel Staining (Thermo Fisher Scientific #S12000) and by Western Blot analysis.

3.13 MS/MS analysis of OGT IP

Sample preparation

Reduction of disulphide bridges in cysteine containing proteins was performed with dithiothreitol (56°C, 30 min, 10 mM in 50 mM HEPES, pH 8.5). Reduced cysteines were alkylated with 2-chloroacetamide (RT, in the dark, 30 min, 20 mM in 50 mM HEPES, pH 8.5). Samples were prepared using the SP3 protocol (Hughes et al., 2014; Hughes et al., 2019) trypsin (sequencing grade, Promega) was added in an enzyme to protein ratio 1:50 for overnight digestion at 37°C. Next day, peptide recovery in HEPES buffer by collecting supernatant on magnet and combining with second elution wash of beads with HEPES buffer.

Peptides were labelled with TMT6plex (Dayon et al., 2008) Isobaric Label Reagent (ThermoFisher) according the manufacturer's instructions. In short, 0.8mg reagent was dissolved in 42 µl acetonitrile (100%) and 8 µl of stock was added and incubated for 1h RT. Followed by quenching the reaction with 5% hydroxylamine for 15min at RT, samples were combined. For further sample clean up an OASIS® HLB µElution Plate

(Waters) was used. Offline high pH reverse phase fractionation was carried out on an Agilent 1200 Infinity high-performance liquid chromatography system, equipped with a Gemini C18 column (3 µm, 110 Å, 100 x 1.0 mm, Phenomenex). The solvent system consists of 20 mM ammonium formate (pH 10.0) (A) and 100% acetonitrile as mobile phase (B). The separation was carried out at a flow rate of 0.1 mL/min using the following linear gradient: 100% A for 2 min., from 100% A to 35% B in 59 min., to 85% B in a further 1 min, and held at 85% B for an additional 15 min, before returning to 100% A and re-equilibration for 13 min. Thirty-two fractions were collected along with the LC separation that were subsequently pooled into 6 fractions. The first and the two last fractions were discarded. Pooled fractions were dried under vacuum centrifugation, reconstituted in 10 µL 1% formic acid, 4% acetonitrile and then stored at -80 °C until LC-MS analysis.

Liquid Chromatography with tandem mass spectrometry (LC-MS-MS)

An UltiMate 3000 RSLC nano LC system (Dionex) ((fitted with a μ-Precolumn C18 PepMap 100, 5μm, 300 μm i.d. x 5 mm, 100 Å and an analytical column (nanoEase™ M/Z HSS T3 column 75 μm x 250 mm C18, 1.8 μm, 100 Å, Waters)) was coupled to the QExactive Plus (Thermo Fisher) via a Pico-Tip Emitter 360 μm OD x 20 μm ID; 10 μm tip (CoAnn Technologies). A spray voltage of 2.2 kV was applied. Trapping was carried out with a constant flow of 0.05% trifluoroacetic acid in water at 30 μL/min for 6 minutes. Peptides were eluted via the analytical column, solvent A (0.1% formic acid in water) with a constant flow of 0.3 μL/min, with increasing percentage of solvent B (0.1% formic acid in acetonitrile) from 2% to 4% in 4 min, from 4% to 8% in 2 min, then 8% to 28% for a further 37 min, in another 9 min from 28%-40%, and finally 40%-80% for 3 min followed by re-equilibration back to 2% B in 5 min. The outlet of the analytical column was coupled directly to an Orbitrap QExactive™ plus Mass Spectrometer (Thermo) using the Nanospray Flex™ ion source in positive ion mode.

The capillary temperature was set at 275°C. Full mass scan was acquired with mass range 375-1200 m/z in profile mode with resolution of 70000. The filling time was set at maximum of 100 ms with a limitation of 3x106 ions. Data dependent acquisition (DDA) was performed with the resolution of the Orbitrap set to 17500, with a fill time of 50 ms and a limitation of $2x10^5$ ions. A normalized collision energy of 32 was applied. Dynamic exclusion time of 20 s was used. The peptide match algorithm was set to ‘preferred’ and charge exclusion ‘unassigned’, charge states 1, 5 - 8 were excluded. MS2 data was acquired in profile mode.

3.14 CUT&RUN-seq

Cleavage under targets and release using nuclease (CUT&RUN) was performed as previously described on live nuclei (Skene and Henikoff, 2017). E14 mES cells were cultured in T2i and 1 x10^6 cells for each sample were collected at 600 g for 3 min at 4°C. Nuclei were extracted by resuspending the cell pellet in 1 mL of Nuclear Extraction Buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 0.5 mM Spermidine, 0.1% Triton X-100, 20% Glycerol, 1X cOmplete Mini EDTA-Free Protease Inhibitors (Roche)). 25 µl of Concanavalin A magnetic brads (Bangs Laboratories #BP531) previously washes twice in Binding Buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 1 mM $CaCl_2$) were gently added to each nuclei sample and incubated for 10 min at RT under gentle rotation. The bead-bound nuclei were isolated on a magnetic stand and beads were blocked with 1 mL of Blocking Buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 0.1% BSA, 2 mM EDTA, 1X cOmplete Mini EDTA-Free Protease Inhibitors (Roche)) for 5 min at RT. The bead-bound nuclei were isolated on the magnetic stand, washed with 1 mL of Wash Buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 0.1% BSA, 1X cOmplete Mini EDTA-Free Protease Inhibitors (Roche)) and resuspended in 300 µl of Antibody Buffer supplemented with 1 µg of antibody. For the nuclei samples permeabilized with Triton X-100, after the first wash with Wash Buffer, nuclei were incubated for 5 min with 1 mL of Wash Buffer supplemented with 0,1%, 0,2% or 0,5% of Triton X-100, and subsequently washed twice with Wash Buffer. Samples were incubated under gentle rotation overnight at 4°C, washed twice with 1 mL of Wash Buffer and resuspended in 300 µl of Wash Buffer supplemented with 700 ng/ml of pA/G MNase. The bead-bound nuclei were incubated for 1h at 4°C under gentle rotation and washed twice with Wash Buffer. The samples were resuspended in 50 µl of Wash Buffer and placed on iced water to pre-cool to 0°C. pA/G MNase targeted digestion was initiated by adding 2 µl of $CaCl_2$, samples were mixed by flicking and incubated for 30 min on iced water at 0°C. Digestion reaction was stopped by adding 50 µl of 2X Stop Buffer (200 mM NaCl, 20 mM EDTA, 4 mM EGTA, 1% IGEPAL, 1 mM $MnCl_2$). Samples were incubated at 37°C for 15 min to release the CUT&RUN fragments from the insoluble nuclear chromatin and centrifuged at 16,000 g for 5 min at 4°C. Supernatants were collected after magnetic beads separation on magnetic stand and transferred to new tubes. 2 µl of 10% SDS and 2.5 µl of 20 mg/ml of Protein K were added and samples were incubated at 70°C for 10 min. DNA fragments were purified and size selected using SPRIselect magnetic beads (Beckman Coulter #B23318) following the manufacturer's protocol for

single selection to purify fragments higher than 100 bp. DNA fragments were eluted in 30 μl of 0.1 M TE. Samples concentrations were verified using Qubit III and libraries prepared with NEBNext Ultra II DNA Library Prep Kit for Illumina (E7645S) following the manufacturer's protocol. Libraries were size selected from Nusieve 3:1 Agarose gel to isolate fragments from 250 bp to 460 bp and DNA purified using Monarch® DNA Gel Extraction Kit Protocol (New England Laboratories #T1020). Libraries quality was evaluated using Tape Station DNA HS D1000 Kit on a Tape Station system (Agilent 4150) and single-end sequenced (SE75) with Illumina NextSeq 500 platform at the EMBL Genomics Core Facility.

3.15 Chromatin immunoprecipitation and sequencing

For O-GlcNAc chromatin profiling, E14 and *A2loxP-Cre* cell lines in pluripotent and differentiated states were used. For O-GlcNAc chromatin profiling at ICRs, ♀$^{\mathbf{Cast}}$/♂$^{\mathrm{CD1}}$ and ♀$^{\mathbf{CD1}}$/♂$^{\mathbf{Cast}}$ XY F1 hybrid cells derived from the breeding between *Mus Castaneus* and *Mus Musculus* CD1 strains were used. Cells were detached with TrypLE, washed with PBS 1X and collected to a final number of 30 x10^6 for each replicate. Cells were resuspended in 45 mL PBS 1X and 3.5 mL of Fixing Solution (50 mM HEPES-KOH pH 7.5, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0) and were fixed with 1.1% formaldehyde (Sigma-Aldrich #252549) for 10 min at RT. Fixing reaction was quenched with 125 mM Glycine and cells were washed with PBS 1X at 2,000 g for 6 min. Cells were lysed in Lysis Buffer 1 (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8, 10% Glycerol, 0.5 % IGEPAL, 0.25% Triton X-100, cOmplete Mini EDTA-Free Protease Inhibitors (Roche)), and washed in Lysis Buffer 2 (10 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM EDTA pH 8, 0.5 mM EGTA pH 8, protease inhibitors). Soluble chromatin was sheared by sonication in Sonication Buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 1 mM EGTA pH 8.0, 1% Triton X-100, 0.1% $C_{24}H_{40}O_4$, 0.1% SDS, cOmplete Mini EDTA-Free Protease Inhibitors (Roche)) in 1 mL Covaris tubes (Covaris #520081) using a Covaris 220 system to a chromatin average size of 200-250 bp with the following program: peak power 150.0, duty factor 25.0, cycles/burst 200. Soluble chromatin was separated by collecting the supernatant after centrifugation at 18,000 g for 2 min at 4°C. 60 µg or 100 µg of chromatin were immunoprecipitated overnight at 4°C with 10 µg of antibody. Spike-in internal control was also included: 120 ng (1/500 ratio) of Drosophila chromatin (Active motif #53083) and 3 µg of anti-H2Av antibody

(Active motif #61686) were added together with mouse chromatin. Samples were incubated with 50 µl of Dynabeads protein G magnetic beads (Life Technologies), previously blocked with 0.5% BSA for 4h at 4°C. Bead-bound chromatin was washed once with 1 mL of Sonication Buffer, twice with 1 mL of Sonication Buffer supplemented with 500 mM NaCl, twice with 1 mL of Sonication Buffer supplemented with 1 M NaCl and once with 1 mL of LiCl Wash Buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% $C_{24}H_{40}O_4$). Beads-bound chromatin was eluted twice for 15 min at 65°C in Elution Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) and de-crosslinked overnight at 65°C. Chromatin was treated with 0.2 mg/mL of RNaseA at 37°C for 1h and with 0.2 mg/mL of Proteinase K at 55°C for 30 min. DNA was isolated by phenol/chloroform extraction followed by ethanol precipitation and resuspended in 30 µl of 0.1 M TE. Samples concentrations were verified using Qubit III and libraries prepared with NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Laboratories #E7645S) following the manufacturer's protocol. Libraries were size selected from Nusieve 3:1 Agarose gel to isolate fragments from 250 bp to 460 bp and DNA purified using Monarch® DNA Gel Extraction Kit Protocol (New England Laboratories #T1020). Libraries quality was evaluated using Tape Station DNA HS D1000 Kit on a Tape Station system (Agilent 4150) and single-end sequenced (SE75) with Illumina NextSeq 500 platform at the EMBL Genomics Core Facility.

For OGT double-fixation chromatin profiling, *Ogt*$^{AID/Myc}$ and E14 cell lines were used. Cells were grown in T2i medium on 3X 15 cm plates for 48 h. Succinimidyl succinate (EGS, Thermo Fisher Scientific #21565) was diluted in PBS 1X to a final concentration of 1.5 mM. Growing medium was removed and cells were covered with 15 mL of EGS solution and incubated for 30 min at RT. After incubation timing, 2 mL of 16% formaldehyde were added to the cells already in EGS solution, to a final concertation of 1.88%, and incubated at RT for 10 min. Fixation reaction was stopped by addition of 1 mL of 2 M Glycine and cells were incubated at RT for 5 min. Cells were washed twice with ice-cold PBS 1X, scrape-collected and centrifuged for 5 min at 500 g. Cells were lysed in 6 mL of Nuclear Lysis Buffer (0.5% Triton X-100, 0.1 M Sucrose, 5 mM $MgCl_2$, 1 mM EDTA, 10 mM tris-HCl pH 8.0, 1X cOmplete Mini EDTA-Free Protease Inhibitors) for 10 min on ice and dounce-homogenized 20X first with loose pestle A followed by 20X with tight pestle B. Nuclei were collected by centrifugation at 2000 RPM for 15 min at 4°C and resuspended in 900 µl of Lysis Buffer (1 mM EDTA, 0.5 mM EGTA, 10 mM tris pH 8.0, 0.5% N-Lauroylsarcosine, 1X cOmplete Mini EDTA-Free Protease Inhibitors).

Chromatin was sheared by sonication in 1 mL Covaris tubes (Covaris #520081) using a Covaris 220 system to a chromatin average size of 300-400 bp with the following program: peak power 140.0, duty factor 5.0, cycles/burst 200 repeated for 5 cycles. Chromatin was centrifuged for 15 min at 20,000 RPM at 4°C to remove unsonicated chromatin, supernatant was collected and diluted in 3.6 mL of Dilution Buffer (1.25% Triton, 0.125% sodium deoxycholate, 6 mM EDTA, 10 mM Tris-HCl pH 8.0, 1X cOmplete Mini EDTA-Free Protease Inhibitors). 50 µl of Dynabeads protein G magnetic beads were prepared as previously described and incubated with 6 µg of primary antibody (Table 3). The primary antibody-coated magnetic beads were added to chromatin and samples were incubated overnight at 4°C under gentle rotation. Beads were isolated on magnetic stand and washed with 1.5 mL of LiCl Buffer (0.5 M LiCl, 1% IGEPAL, 1.1 & sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.1) and High Salt Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 1 M NaCl, 0.1% sodium deoxycholate), each one five times for 10 at 4°C minutes under gentle rotation. Beads-bound chromatin was washed once with 1.5 mL of 1X TE and eluted in 48 µl of Proteinase K Digestion Buffer (20 mM HEPES, 1 mM EDTA, 0.5% SDS, 0.8 mg/mL Proteinase K) for 30 min at 56°C. Beads were removed using the magnetic stand and eluted chromatin was supplemented with 2 µl of 20 mg/mL Proteinase K (Biotech Rabbit # BR1100901) and 3 µl of 5 M NaCl. Chromatin was de-crosslinked for 2h at 56°C and for 4h at 68°C. DNA was purified using SPRIselect magnetic beads following manufacturer's instructions. Samples concentrations were verified using Qubit III and libraries prepared with NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Laboratories #E7645S) following the manufacturer's protocol. Libraries were size selected from Nusieve 3:1 Agarose gel to isolate fragments from 250 bp to 460 bp and DNA purified using Monarch® DNA Gel Extraction Kit Protocol (New England Laboratories #T1020). Libraries quality was evaluated using Tape Station DNA HS D1000 Kit (Agilent 4150) and single-end sequenced (SE75) with Illumina NextSeq 500 platform at the EMBL Genomics Core Facility.

3.16 Total and nuclear RNA-seq

For nuclear RNA extraction, fresh ESCs and NPCs were washes 3 times with PBS 1X and collected at 200 g for 5 min. Cell pellets were resuspended with 1 mL of HS Lysis Buffer (50 mM KCl, 10 mM $MgSO_4 \cdot 7H_2O$, 5 mM HEPES, 0.05% IGEPAL, 3 mM DTT, 1X

cOmplete Mini EDTA-Free protease Inhibitors (Roche)) and nuclei pelleted at 1,000 g for 5 min at 4°C. Nuclei were washed three times with 1 mL of RSB Buffer (10 mM NaCl, 10 mM tris pH 8.0, 3 mM $MgCl_2$) and pelleted at 1,000 g for 5 min at 4°C. For total RNA extraction, fresh ESCs or NPCs were washed twice with 1X PBS and collected as cell pellets. Total and nuclear RNA were extracted using the Monarch RNA MiniPrep T2010 kit following the manufacturer's guidelines. RNA was quantified using Qubit III and quality checked with Tape Station RNA HS D1000 Kit on a Tape Station system (Agilent 4150) to ensure RNA integrity number (RIN) >8.5. Messenger RNA libraries and RNA sequencing were performed at the EMBL Genomics Core Facility, with pair-ends (PE40) sequencing mode with Illumina NextSeq 500 platform.

3.17 ATAC-seq

Prior to harvesting, cells were treated with DNase at a final concentration of 200 U/ml for 30 min at 37°C to digest possible degraded DNA released from dead cells. Cells were washed with PBS 1X, detached with TrypLE and 50,000 cells for each condition were pelleted at 500 g for 5 min at 4°C. Cell pellet was resuspended in 50 µl of cold ATAC-Resuspension Buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) supplemented with 0.1% IGEPAL, 0.1% Tween-20 and 0.01% Digitonin and incubated on ice for 3 min. Lysis was washed out with 1 mL of cold ATAC-Resuspension Buffer supplemented with 0.1% Tween-20 and cells pelleted at 500 g for 10 min at 4°C. Cell pellet was resuspended in 50 µl of transposition mixture (25 µl 2xTD buffer, 2.5 µl transposase (Illumina Tagment DNA Enzyme and Buffer Kit #20034197), 16.5 µl 1X PBS, 0.5 µl 1% digitonin, 0.5 µl 10% Tween 20, 5 µl H2O) and incubated at 37°C for 30 min in a thermomixer shaking at 1,000 RPM. Reaction clean-up was performed with DNA Clean and Concentratot-5 Kit (Zymo #D4014) and purified DNA was eluted in 21 µl of Elution Buffer. The purified DNA was amplified by PCR amplification using a unique combination of dual barcoded primers P5 and P7 from the Nextera XT Index Kit (Illumina #15055293), with the Q5 Hot Start High-Fidelity polymerase (NEB #M0494S) and the following cycling conditions: 72°C for 5 min, 98°C for 30 sec, 5 cycles of 98°C for 10 sec, 63°C for 30 sec, 72°C for 1 min. 10% of the pre-amplified mixture was used to run a qPCR and determine the number of additional cycles needed to amplify the DNA. qPCR was run using SYgreen Blue Mix (PCRbio #PB20.15-05) and the previously used P5 and P7 primers in a LightCycler® (Roche) thermal cycler. After qPCR amplification, the amplification

profiles were manually assessed by plotting linear Rn versus cycle based on previously optimized workflow analysis (Buenrostro et al., 2015) in order to perform an additional number of cycles equal to one-third of the maximum fluorescent intensity of the plot reported by the authors. The remaining DNA material was used to run the additional calculated cycles without the addition of other reagents. The final PCR reaction was purified using the DNA Clean and Concentratot-5 Kit (Zymo #D4014) and final DNA was eluted in 20 µl of H_2O. Samples concentrations were verified using Qubit III and size selected from Nusieve 3:1 Agarose gel to isolate fragments from 150 bp to 800 bp and DNA purified using Monarch® DNA Gel Extraction Kit Protocol (NewEngland Laboratories #T1020). DNA quality was evaluated using Tape Station DNA HS D1000 Kit on a Tape Station system (Agilent 4150) and pair-end sequenced (PE40) with Illumina NextSeq 500 platform at the EMBL Genomics Core Facility.

3.18 Reverse transcription

500 nm of purified RNA from *2Plox*-CXXC-*BtGH84* cell line before and after Dox induction were used to synthetize cDNA using the cDNA Synthesis Kit (BioRabbit #BR0400404). In brief, RNA was incubated with 1mM of dNTPs, 1 U/µl of RNase Inhibitor, 1.25 mM of Hexamer Primers, 1X cDNA synthesis buffer and 10 U/µl RevertUP™ II Reverse Transcriptase and incubated for 10 minutes at 30°C followed by an incubation of 55°C for 60 minutes. Enzymes were inactivated at 99°C for 5 min. To check for gDNA contaminations, a control reaction with the same mix but no RevertUP™ II Reverse Transcriptase (no RT) was also prepared in parallel for each RNA sample.

3.19 Real-time quantitative PCR

Assessment of positive and negative loci from CUT&RUN experiment was achieved by real-time quantitative qPCR. A mix of 2x qPCRBIO SyGreen Blue Mix and 400 nM of primers designed at the interested genomic loci was prepared and mixed with extracted DNA diluted 1/10. The reaction was performed with the following program: 95°C for 3 min, 95°C for 10 seconds and 60°C for 20 seconds for 40 cycles. The reaction was implemented in a LightCycler® (Roche) thermal cycler.

Analysis of pre-rRNA transcription was performed on cDNA obtained from reverse transcription of *2Plox*-CXXC-*BtGH84* RNA samples. The reaction was performed using the same program of CUT&RUN qPCR. The used primers are listed in Table 2. The relative quantification of each gene to β-Actin was calculated with the $2^{-\Delta\Delta}C_T$ method, using t-test to assess statistical significance.

3.20 RNA Pol II Immunoprecipitation (IP)

Immunoprecipitation of RNA Pol II was performed from *RNA Pol II-HA-SPOT* cell line grown in KO DMEM Serum/LIF T2i. For each IP, 100 x10^6 cells were detached with TrypLE, washed twice with PBS 1X and collected at 1,500 g for 5 min. Cell pellet was resuspended in 7 mL of Buffer A (10 mM HEPES pH 7.65, 1 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 1X cOmplete Mini EDTA-Free Protease Inhibitors) and incubated on ice for 15 min. Nuclei were extracted using the Dounce homogenizer, volume was equilibrated to 10 mL with buffer A and nuclei were collected by centrifugation at 300 g for 10 min at 4°C. The supernatant, which corresponds to the cytoplasmic fraction, was discarded and nuclear pellet was washed with 10 mL of Buffer N (40 mM HEPES pH7.65, 1 mM $MgCl_2$, 0.5 mM DTT, 250 mM Sucrose, 1X cOmplete Mini EDTA-Free Protease Inhibitors) and centrifuged at 1,000 g for 10 min at 4°C. Nuclear pellet was resuspended in 400 µl of Buffer N and nuclei were lysed by adding 400 µl of 2X Lysis Buffer (0.5% SDS). Genomic DNA was digested by adding 8 µl of Universal Pierce Nuclease and by incubating the samples for 10 min at 37°C. Proteins were denaturated at 95°C for 2 min and insoluble debris removed by collecting the supernatant after centrifugation at 13,000 g for 5 min at 4°C. Protein concentration was estimated using the BCA Protein Assay Kit as previously described and equal amount of nuclear protein extract was used for each IP. 100 µl of either Pierce Anti-HA Magnetic beads (Thermo Fisher Scientific #88836) or Spot-Trap Magnetic patricles M-270 (Chromotek #etd-20) were washed twice with 1 mL of beads Washing Buffer (10 mM HEPES pH 7.65, 1 mM $CaCl_2$, 0.01% NP-40, 0.1% Bovine Serum Albumine (BSA)) and resuspeded in 50 µl of Washing Buffer (20 mM HEPES pH 7.65, 250 mM NaCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 0.5% IGEPAL, 1X cOmplete Mini EDTA-Free Protease Inhibitors). 3 volumes of 3/4X Binding Buffer (66 mM HEPES pH 7.65, 333 mM NaCl, 1.4 mM $CaCl_2$, 2 mM $MgCl_2$, 1X cOmplete EDTA-Free Mini Protease Inhibitors) and 50 µl of magnetic beads were added to the nuclear extract and sample incubated overnight at 4°C. Bead-bounded proteins were washed 4 times for 10

min with 1 mL of Washing Buffer and eluted twice for 5 min at 95 °C in 50 µl of Elution Buffer (20 mM HEPES pH 7.65, 500 mM NaCl, 1% SDS, 1% IGEPAL, 0.5% Sodium deoxycholate, 0.5 mM DTT). Immunoprecipitated protein samples were analysed by total protein staining using SYPRO Ruby Protein Gel Staining (Thermo Fishe Scientific #S12000) and by Western Blot analysis.

3.21 Immunofluorescence of RNA Pol II

A9 WT and *RNA Pol II-HA-SPOT* cell lines were plated on glass coverslips (170 µm thickness) at a density of 3 x10^4/0.8 cm^2. For mESCs, cells were plated on the glass coverslips 12h before the experiment, while for NPCs the cells were plated on glass coverslips when starting the differentiation protocol. Cells were fixed for 20 min at RT with 4% of formaldehyde solution diluted in PBS 1X and permeabilized for 5 min at RT with 0.3% Triton X-100 solution diluted in PBS 1X. Cells were washed three times for 5 min with 1X PBS and blocked for 30 min at RT with PBT solution (1X PBS, 0.1% Tween-20, 5% BSA). Blocking solution was removed and cells were incubated for 1h at RT with primary antibody diluted 1:1000 in PBT solution. Primary antibody solution was removed, cells washed three times with PBS 1X for 5 min at RT and incubated 1h at RT with secondary anti-mouse antibody labeled with STAR RED diluted 1:200 in PBT solution by covering the coverslips with aluminium foil. Secondary antibody solution was removed, cells were washed twice in PBS 1X for 5 min at RT and incubated for 5 min at RT in PBS 1X supplemented with 1:1000 DAPI solution (Thermo Fisher Scientific #D1306). Cells were washed once with PBS 1X, excess of solution was drained and coverslips were properly mounted with Abberior Mount Solid Antifade (Abberior #MM-2013-2X15ML). Imaging by stimulated emission depletion (STED) microscopy was performed using a STEDYCON mounted on an upright Zeiss microscope in confocal or STED modes. Samples were imaged with a Zeiss 100×1.46 NA objective, 20 nm pixel size, 5 μs pixel dwell time, 15-line accumulations and a pinhole of 64 μm. STED laser powers were 100% of the 640 nm and 775 nm 51.73% for the STAR RED channel, whereas for the DAPI channel was 1.8% of the 405 nm laser.

3.22 Generation of RNA Pol II CTD O-GlcNAc T4 mouse monoclonal antibody

Mouse monoclonal antibody against O-GlcNAcylated threonine at the position four of the heptamere repeat of RNA Polymerase II was produced in collaboration with Jackson Laboratory (JAX) Protein Sciences Service. The 13-mere peptide containing two repeats with modified threonines $Y_1S_2P_3T_4$(O-GlcNAc)$S_5P_6S_7$-$Y_1S_2P_3T_4$(O-GlcNAc)$S_5P_6S_7$ conjugated with Diphteria toxoid were used to raise antibodies against O-GlcNAcylated RNA Polymerase II. Cohorts of two Balb\cJ or JD-1 mice from JAX Laboratories were pre-bled before immunization to provide a negative control serum sample. Mice were immunized intra-peritoneally with immunogens at 10 μg of immunogen per animal per immunization round. Immunogens were mixed with alum adjuvant and CpG oligos to stimulate TLR-9 receptors and the production of IgG2 isotypes. Half of each immunogen sample was heat denatured at 95°C for 5 min. Two weeks after the initial immunization, an identical booster immunization was administrated to each animal in the cohort, and two weeks later a third identical round of immunization was also administrated. One animal was randomly selected for fusion and, three weeks after the previous immunization, the animal received a final pre-fusion booster with intra-peritoneal injection in the same manner as previously described. Four days after the pre-fusion boost, the mouse was euthanized, the spleen removed under sterile conditions and rendered into single cell suspension by mechanical disruption in order to obtain splenocytes. The splenocytes were treated to lyse red blood cells and the isolated white cells counted and mixed with pre-cultured SP2 myeloma cells at a 5:1 ratio. Fusion was carried out using a Tecan Evo 200 workstation, in a custom sterile enclosure, by measured dispensing of polyethylene glycol and measured dilution of PEG over time with agitation in a small-radius orbital shaker integrated into the robotic platform's work deck. Post-fusion slurried were plated out in 20X 96-well plates, 10 of which were seeded at a half density. The remaining quarter of the fusion was cryo-preserved in N_2. Cells were selected using azaserine-hypoxanthine to kill unfused myeloma cells, while unfused splenocytes naturally die after several days. Seeding plates at this density ensures that most cell lines produced are monoclonal from the moment they are seeded. Post-fusion plates were returned to humid CO_2 incubator for two weeks to grow and secrete antibody into the selection medium. After two weeks, 50 μl of the 1920 supernatants in the 96-well plates were arrayed into five 384-well plates using the Multi Channel Arm (MCA) of the Tecan Evo 200. The 384-well plates were used as source

plates to generate microarrays using an Arrayjet Super Marathon microarray printer. A microarray screening was performed: culture supernatants were spotted on to pre-blocked aminosilane slides coated with the antigen and allowed to bind. The microarrays were washed and probed with isotype specific anti-mouse secondary antibodies to assay for presence of IgG isotypes bound to the antigen. Each supernatant was screened against multiple antigenic targets to exclude antibodies that are reactive to the unmodified peptide target $Y_1S_2P_3T_4S_5P_6S_7$-$Y_1S_2P_3T_4S_5P_6S_7$. Antibodies binding to the targets were identified and the corresponding clones in the fusion plates were transferred to three wells of a 24 well plate at serial dilutions for the expansion process. When at confluency, the cells were transferred in a T25 culture flask and allowed to grow until ready for cryopreservation. An aliquot of cells was retained, added back to the culture medium and allowed to continue growing to provide a higher concentration of antibody in the culture supernatant. Finally, supernatants were harvested and filtered through a 0.22 μm filter. The clones were tested by Western Blot analysis on nuclear protein extract, 2D3 positive clone was rescreened by direct ELISA and subsequently expanded and purified. 2D3 clone was tested for mycoplasma and weaned from serum-containing medium to serum free medium by successive passages into serum that was progressively diluted with serum-free medium. Cells were introduced into a Celline 1000 bioreactor flask/hollow fiber bioreactor and cell density monitored using a hemocytometer/GlucCell glucose monitoring system. Cells and supernatants were harvested at appropriate intervals, centrifuged to remove cells and filtered through a 0.22 μm membrane to remove debris. The antibody was isotyped and the culture supernatant was circulated over protein-A/G sepharose. The antibody was eluted using low pH glycine into neutralizing buffer. Concentration of the antibody was measured and an ELISA performed for quality control (QC).

3.23 Isolation of nuclear proteins from ESCs and NPCs for ETD MS/MS

A9 WT mESCs and NPCs grown as previously described were detached with TrypLE from 2X 225 cm^2 flasks, washed with PBS 1X and pelleted for 5 min at 1,000 RPM. Nuclei and nuclear proteins were extracted as previously described (paragraph 2.11). Prior to ETD MS/MS analysis, efficiency of differentiation was verified by Western Blot analysis.

3.24 WGA-IP of nuclear proteins

Nuclear protein samples were prepared as previously described (paragraph 3.11). 200 µl of Promega High Magnetic® Streptavidin magnetic beads (Promega #V7820) were washed twice with 2 mL of Beads Washing Buffer, resuspended in 2 mL of Beads Washing Buffer supplemented with 500 µg of Biotin- conjugate WGA and incubated for 2h at RT on a rocker. WGA-bound beads were washed with 1 mL of Beads Washing Buffer and resuspended in 50 µl of WGA Washing Buffer. Two negative controls were also used: to saturate WGA binding sites and ensure the specificity of O-GlcNAc protein binding, 200 µl of Promega High Magnetic® Streptavidin magnetic beads were incubated in parallel with 1.1 M of free O-GlcNAc powder (Sigma Aldrich #A8625). Purified bacterial BtGH84 protein was also used in parallel as negative control to treat the nuclear protein extract for O-GlcNAc global removal. For BtGH84 treatment, 1/2 of the IP sample was eluted at 95°C in 20 mM HEPES pH 7.65, to avoid BtGH84 inactivation by buffer detergents. 4 mg/ml of purified bacterial BtGH84 were diluted in 20 mM HEPES pH 7.65 to a final concentration of 50 ng/µl and 2 µl were added to the WGA-IP sample to have a final enzyme quantity of 100 ng. WGA-IP was treated with bacterial BtGH84 for 4h at 20°C. WGA-IP sample and negative control samples were analyzed by Western Blot analysis.

3.25 Proteomics analysis of ESC and NPC O-GlcNAc proteome

3.25.1 Mass spectrometry analysis of ESC and NPC proteome

For sample preparation, 300 µg of protein lysates were precipitated by methanol-chloroform precipitation. Briefly, samples were brought to 100 µL with H_2O. 400 µL MeOH were added and vortexed, followed by 100 µL of chloroform and vortexed. Lastly, 300 µL were added and vortexed. Samples were spun at 10000 g for 5 minutes. The supernatant was aspirated and the protein precipitates were washed 3 times with 400 µL MeOH and finally air dried. Dried protein pellets were brought to 50 µL with 8 M urea (50 mM Tris pH 8.0, 75 mM NaCl). Samples were reduced with 10 mM of TCEP (tris(2-carboxyethyl)phosphine) solution for 30 minutes at RT in the dark and alkylated with 5 mM IAA (Indole-3-Acetic Acid) for 30 minutes at RT in the dark. Protein samples were diluted to 2 M urea with 50 mM Tris pH 8.0 and digested with 6 mIU of Lys-C for 2 hours at RT, shaken at 800 rpm. Afterwards, samples were digested with 6 µg trypsin overnight

at RT, shaken at 800 rpm. Digestion was quenched with 1% FA and spun down at max speed to remove debris.

Digested peptides were desalted using SepPAK columns and dried down in vacuum centrifuge. Peptides were quantified with BCA assay and brought up in 1 μg/ μL of 50 mM HEPES pH 8.5. 150 μg peptides in each sample were labeled with 300 μg TMTpro channel as followed:

Sample	TMTpro channel
ESC 1	127N
ESC 2	127C
ESC 3	128N
ESC 4	128C
NPC 1	129N
NPC 2	129C
NPC 3	130N

The mixture of peptides and TMT reagents were incubated at RT for 1h shaking at 1000 RPM. TMT labeling efficiency was checked to be at least 98%. Samples were quenched with 0.5% hydroxylamine for 15 min at RT, shaking at 1000 RPM. All TMT-labeled samples were pooled together and the mixture was desalted using SepPAK column. 20 μg of TMT-labeled peptides were saved for Stage tip bSDB fractionation as previously described (Papanastasiou et al., 2019). Briefly, 200 μL pipette tip was packed with two punches of sulfonated diphenylbenzene. The Stage tip was conditioned 1x with 100 μL MeOH, washed 1x with 100 μL 50% ACN/ 0.1% FA, and equilibrated 4x with 0.1% FA. The peptide was loaded at 2000 g until the entire sample passed through. A pH switch was performed using 100 μL 20 mM NH_4HCO_2 pH 10 and was considered fraction 0. Then loaded peptides were eluted in 10 fractions of 20 mM NH_4HCO_2 with ACN concentrations of 2.5, 5, 7.5, 10, 15, 20, 25, 30, 35, 40 and washed with 90% ACN. Fractionated peptides were transferred to HPLC vials, dried down and brought to 9 μL of 2% ACN/ 0.1% FA and 4/9 μL was injected for LC-MS/MS analysis.

3.25.2 O-GlcNAc immunoprecipitation

Remaining peptides were dried down in vacuum centrifuge and brought to 1 mL of chilled immunoaffinity buffer (IAP buffer). 20 μL of O-GlcNAc noncovalently linked to protein-A coated agarose bead slurry was washed twice with chilled IAP buffer. Peptide was incubated with anti-OGlcNAc antibodies by rotation for 2 hours at 4^{o}C. The flow-through

was collected and anti-OGlcNAc antibodies conjugated beads were washed 1X with chilled IAP buffer and 2x with chilled PBS. Samples were eluted in 75 μL of 0.15% TFA aqueous solution with light agitation for 10 min, repeated twice. Sample was centrifuged for 1 min and eluates were transferred to equilibrated Stage tip. Stage tip was packed with two punches of Empore 2215-C18 Octadecyl 47 mm disks to a 200 μL pipette tip. Stage tip was conditioned 1X with 100 μL MeOH, washed 1x with 100 μL 50% ACN/ 0.1% FA, and equilibrated 4x with 0.1% FA. Sample was loaded onto Stage tip after equilibration, washed twice with 100 μL 0.1% FA and eluted in 60 μL 50% ACN/ 0.1% FA. Eluate was transferred to HPLC vials, dried down and brought to 12 μL of 2% ACN/ 0.1% FA. 4/12 μL was injected for LC-MS/MS analysis.

3.25.3 LC-MS/MS analysis

Chromatography was performed using a UHPLC at a flow rate of 200 nL/min. Peptides were separated on a 75 μm i.d. PicoFrit column (New Objective, Woburn, MA) packed with 1.9 μm AQ-C18 material (Dr. Maisch, Germany) to 33-cm in length over 110 min run. The online gradient started from 10 % buffer B (80% (v/v) MeCN/0.1% (v/v) FA) at 1 min to 30 % B in 84 mins, followed by an increase to 72 % B by minute 94, then to 90 % B by min 95, and finally to 60% B by min 101 until the end of the run. Mass spectrometry was performed on a Thermo Scientific Eclipse mass spectrometer. A precursor scan is performed from 350 to 2000 m/z at 60000 resolution with a duty cycle of 2s. Higher energy collisional dissociation (HCD) was performed at 32% and at a resolution of 50000. Precursor isolation width was set to 0.7 m/z and the maximum MS2 injection time was 105 msecs for an automatic gain control of 1e5. Dynamic exclusion was set to 45 s and only charge states of two to five were selected for MS2 scans.

3.26 Bioinformatics analysis

3.26.1 OGT's native complexes analysis

MS/MS data were analyzed in collaboration with the EMBL Proteomic Core Facility. The raw output files of IsobarQuant (protein.txt – files) were processed using the R programming language (ISBN 3-900051-07-0). Contaminants were filtered out and only proteins that were quantified with at least two unique peptides were considered for the analysis. Log2 transformed raw TMT reporter ion intensities ('signal_sum' columns) were first cleaned for batch effects using limma (PMID: 25605792) and further normalized using VSN (variance stabilization normalization - PMID: 12169536). Proteins were tested for differential expression using the limma package. The replicate information was added as a factor in the design matrix given as an argument to the 'lmFit' function of limma. A protein was annotated as a hit with a false discovery rate (fdr) smaller 5% and a fold-change of at least 50%.

Protein-protein interaction network analysis was performed using STRING (v11.5) database with *Mus Musculus* gene symbols, with physical network type and minimum required interaction score ≥0.7 between the hits of each fraction. Clustering analysis was performed on STRING using Markov clustering option with an inflation value of 2.5. Protein interaction networks were visualized on Cytoscape (v3.9.1) using a yFiles Circular Layout. Edge transparency represent the score value, which indicates the confidence of interactions, while the nodes weights depend on the degree value, which indicates the number of edges that connect each node. Each color of node group represents a specific cluster annotated by Gene Ontology based on Biological Process.

3.26.2 Chromatin immunoprecipitation and CUT&RUN analysis

The following process was performed on Galaxy (Community, 2022). Raw reads were quality checked with FastQC tool (v0.11.9), adapter-trimmed with TrimGalore (https://github.com/FelixKrueger/TrimGalore) (v0.4.3.1, -phred33 --quality 20 --stringency 1 -e 0.1 --length 20), aligned to the mouse genome (mm10/GRCm38) using Bowtie2 (Langmead et al., 2009) (v2.3.4.3, single-end settings --sensitive --no-unal) and only valid primary alignments were kept with Samtools view (https://samtools.github.io/hts-specs/) (v1.8, -h -b -q 20 -F 0x800). Reads that did not align on the major chromosomes were filtered out with Samtools view (v1.8, -h -b input.bam 'chr1' 'chr2' 'chr3' 'chr4' 'chr5' 'chr6' 'chr7' 'chr8' 'chr9' 'chr10' 'chr11' 'chr12'

'chr13' 'chr14' 'chr15' 'chr16' 'chr17' 'chr18' 'chr19' 'chrX' 'chrY'). Aligned records in BAM dataset were examined to discard duplicates with Picard tools (http://broadinstitute.github.io/picard/) (v2.18.2,MarkDuplicates REMOVE_DUPLICATES='true'ASSUME_SORTED='true'DUPLICATE_SCORING_STRATEGY='SUM_OF_BASE_QUALITIES' OPTICAL_DUPLICATE_PIXEL_DISTANCE='100' VALIDATION_STRINGENCY='LENIENT'). Bigwig files were created with Deeptools (v3.0.2.0, bamCoverage --bam in.bam --outFileName out.bam --outFileFormat 'bigwig' --binSize 50 --normalizeUsing RPGC --effectiveGenomeSize 2308125349 --scaleFactor 1.0 --extendReads 150 --minMappingQuality '1'). Bigwigs were loaded on IGV 2.12.3 for data visualization.

3.26.3 RNA-seq analysis

The following process was performed on Galaxy (Community, 2022). Raw reads were quality checked with FastQC (v0.11.9) and trimmed using Trim Galore (v0.4.3.1, -phred33 --quality 20 --stringency 1 -e 0.1--length 20). Processed reads were mapped to the mouse genome (mm10/GRCm38) genome using RNA STAR (v2.6.0b-2, 100 --sjdbOverhang, 255 --outSAMmapqUnique, 1 --outFilterMultimapScoreRange, 10 --outFilterMultimapNmax, 10--outFilterMismatchmax, 0.3 --outFilterMismatchNoverLmax, 1 --outFilterMismatchNoverReadLmax, 0.66 --outFilterScoreMinOverLread, -1--outSAMmultNmax, 50--outBAMsortingBinsN) and reads covering exons of each gene were then counted with featureCounts tool (v2.0.1) using default parameters. BAM files were used to generate bigwig files with RPKM, CPM and TPM normalization using Deeptools (v3.0.2.0).
DEGs were determined with DESeq2 (v2.11.40.6) using the featureCounts tables as input (FDR <0.05, log(fold-change) < 0.1).

3.26.4 ATAC-seq analysis

Raw reads were trimmed with TrimGalore (v0.4.3.1, reads >20bp, phred20, 1 overlap with adapter sequence, 0.1 max allowed error rate) and output files aligned to the mouse genome (mm10/GRCm38) with Bowtie2 (v2.3.4.3, paired-end settings, fragment lenght 0-2000, --fr, allow mate dovetailing, --sensitive). Only primary alignments were kept with

Samtools view (https://samtools.github.io/hts-specs/) (v1.1.2, MapQ >=20, Read is mapped in a proper pair, Supplementary alignment) and duplicate reads removed with Picard tools (http://broadinstitute.github.io/picard/) (v2.7.1.1, MarkDuplicates, max offset between two duplicate clusters 100, Lenient validation stringency, SUM_OF_BASE_QUALITIES). Only major chromosomes were selected. The mapped and filtered sequences were analyzed by MACS2 call peaks (v2.1.1.20160309.5, extension size 75, peak detection based on q-value, min FDR (q-value) cutoff for peak detection 0.1) and BigWig files generated with the Deeptools function bamCoverage (v2.4.1.0, bin size in bases 1, normalize coverage to 1X, scaling factor 1.0). Bigwigs were loaded on IGV 2.12.3 for data visualization.

3.26.5 STED images analysis

High-resolution images were analyzed with Fiji (v2.3.0). Using the images acquired with the DAPI channel, nuclei were isolated from the cell area by modulating the median and threshold values and by generating a binary file. Nuclear SPOT antibody signal acquired with the STAR RED channel was analyzed using the RS-FISH plugin (Bahry et al., 2022) for spot detection. Number of spots, spots mean and max intensity and spots density parameters on the nuclear area were analyzed on Prism (v 9.3.1) using unpaired t test for statistical analysis.

Table 1 List of primers used for cloning and PCR genotyping

Plasmid name	Primer name	Sequence (5'-3')
P164_pETM14	pETM14F_Fw	CGATCCCGCGAAATTAATAC
	pETM14F_Rv	ATGCTAGTTATTGCTCAGCG
gp41-BtGH84-3XNLS-SPOT	EGFP-gp41vectorTail_Fw	AGGGGGGAGTGTGAGCAAGGGCGAGGAGCT
	gp41vector-EGFPTail_Rv	CCTTGCTCACACTCCCCCCTCCACCACTGC
	gp41vector-EGFPTail_Fw	GCTGTACAAGACTGGAGACCCAAAGAAAAA
	EGFP-gp41vectorTail_Rv	GGTCTCCAGTCTTGTACAGCTCGTCCATGC
2Plox-CXXC-BtGH84	EGFP-CXXCvectorTail_Fw	GGAAGCATCAGTGAGCAAGGGCGAGGAGCT
	CXXCvector-EGFPTail_Rv	CCTTGCTCACTGATGCTTCCACTTTCCTCT
	CXXCvector-EGFPTail_Fw	GCTGTACAAGGCGGCCAGTCCAATCGTGAG
	EGFP-CXXCvectorTail_Rv	GACTGGCCGCCTTGTACAGCTCGTCCATGC
2Plox-3XNLS-BtGH84	BtGH84_Fw	ATTGACCAGACGTCTAACCA
	BtGH84_Rv	GCCGGACAACAATTCCTTAA

Table 2 List of primers for qPCR analysis

Gene target	Primer direction	Sequence (5'-3')
5'ETS	Forward	TGTTTCACTTTGGTCGTGTCTC
	Reverse	TCGACGCTTACAAGAAACAGC
47S	Forward	GGTGTCCAAGTGTTCATG
	Reverse	CAAGCGAGATAGGAATGTCTTAC
ITS1	Forward	TCTCGTTTCGTTCCTGCTGG
	Reverse	GATCCACCGCTAAGAGTCGTATC
18S	Forward	GCAATTATTCCCCATGAACG
	Reverse	GGCCTCACTAAACCATCCAA
28S	Forward	TCATCAGACCCCAGAAAAGG
	Reverse	GATTCGGCAGGTGAGTTGTT
β-Actin	Forward	AAATCGTGCGTGACATCAAA
	Reverse	TCTCCAGGGAGGAAGAGGAT

Table 3 List of used antibodies

Antibody	Use	Company	Catalogue number
BtGH84	Western Blot	Thermo Fisher Scientific	PA5-67426
DNMT3B	Western Blot	Abcam	ab2851
FLAG	Western Blot	Sigma	F3165
H3	Western Blot	Abcam	Ab1791
HA	Western Blot	Abcam	ab9110
HA	ChIP-seq	Santa Cruz	901515
HDAC2	Western Blot	Thermo Fisher Scientific	A300-705A-M
Myc	ChIP-seq, CUT&RUN	BioLegend	9E10
O-GlcNAc	Western Blot, ChIP-seq, CUT&RUN	Thermo Fisher Scientific	HGAC85
OCT3/4	Western Blot	R&D Systems	AF1759
OGT	Western Blot, ChIP-seq, CUT&RUN	Abcam	ab177941
RNA Pol II	Western Blot	Santa Cruz	sc-55492
SPOT	Immunofluorescence	Chromoteck	28A5
β-Actin	Western Blot	Sigma Aldrich	A5441
H3K27me3	CUT&RUN	Millipore	07-449
β-tubulin	Western Blot	Sigma Aldrich	T4026

4. Results

4.1 Uncovering the OGT's protein interactomes in the nucleus and in the cytoplasm

The O-GlcNAc transferase (OGT) is an evolutionary conserved enzyme essential for mouse development. OGT is present in both cytoplasm and nucleus of cells, where it modifies its protein targets by adding an N-acetylglucosamine moiety to specific threonines and serine residues. In addition to its enzymatic activity, OGT was reported to have alternative roles: the interaction of OGT with HCFC1 regulates HCFC1 proteolysis and ensures HCFC1 transcriptional activity and the cell cycle progression. Several other nuclear and cytoplasmic factors were also found to interact with OGT (Conway et al., 2021; Deplus et al., 2013; Vella et al., 2013; Wysocka et al., 2003). An example is represented by TET proteins; however, the functional output of the OGT-TET interaction is still not fully understood.
The severe phenotype cause by *Ogt*-null mutant raises the question whether mammalian OGT could have other essential functions in addition to its O-GlcNAc transferase activity. The characterization of the OGT interactome enables to gain insights into the molecular mechanisms of OGT function and elucidate which are the molecular pathways regulated by the O-GlcNAc transferase enzyme.

To investigate the biochemical features of OGT's complexes, we have designed a novel $OGT^{FLAG/HA}$ mES cell line where *Ogt* was endogenously tagged at its C-term with *3X FLAG* tag sequences, and at its N-term with *2X HA* tag sequences. Due to its shorter TPR domain, the short OGT isoform (sOGT) encodes for the HA tag only, making the FLAG tag sequence specific to the long nuclear-cytoplasmic OGT (ncOGT) isoform. The HA tag is instead present in both short and long isoforms at the conserved catalytic OGT C-term domain (Fig. 4A). The $OGT^{FLAG/HA}$ mES cell line represents a physiological model system to determine the cellular localization of the OGT isoforms, and to study the characteristics of the OGT interactors in each cellular compartment.
To study the OGT's interactome, I have first validated the $OGT^{FLAG/HA}$ system by testing the expression of $OGT^{FLAG/HA}$ protein after cellular fraction (Fig. 4B). OGT expression

levels are not affected by the FLAG/HA tag *knock-in* epitope tags, which are comparable to the WT OGT control. The long ncOGT isoform (116 kDa) was detected by HA and FLAG tags in both nuclear and cytoplasmic compartments, in accordance with the previously annotated localization of the long ncOGT transcript (Kreppel et al., 1997), where it shows equal expression levels. In comparison to WT OGT, $OGT^{FLAG/HA}$ is characterized by a molecular weight shift due to the presence of the FLAG and HA tags (1 kDa and 1.1 kDA for single FLAG and HA tag respectively). The long ncOGT isoform revealed to be by far the most abundant OGT isoform in mESCs. The short ~78 kDa OGT isoform was also detected by HA tag in both nuclear and cytoplasmic fractions (Kreppel et al., 1997, Shafi et al., 2000), where it constitutes only a very small fraction of the total OGT. The sOGT isoform showed a smaller molecular weight shift comparing to the long OGT isoform, in accordance to the presence of the C-term HA tag but not the FLAG tag. Compared to HA tag, the FLAG tag resulted in higher background (as shown by signal in the wild type control), probably due to the lower specificity of the anti-FLAG antibody. Additional OGT bands at lower molecular weight were also detected, however it is was possible to assess whether these bands originated from the proteolytic cleavage of OGT occurred during the experimental procedure, thus they will not be considered for this study.

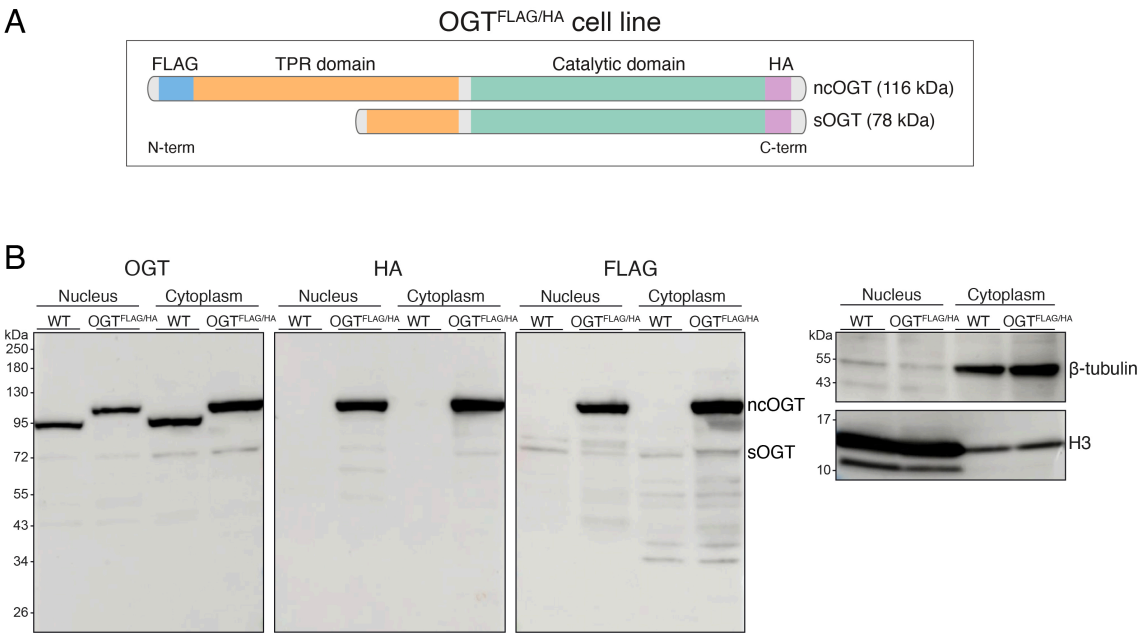

Fig. 4 (A) Schematic representation of the OGT$^{FLAG/HA}$ mECS cell line. **(B)** Western Blot analysis of nuclear and cytoplasmic OGT expression in WT and OGT$^{FLAG/HA}$ cell lines. Samples were blotted with OGT, HA and FLAG antibodies. Purity of cellular fractionation was checked with H3 (ab1791) and β-tubulin (T4026) as nuclear and cytoplasmic markers respectively.

To identify the interacting partners of OGT, Dr. Yulia Ermakova, former PostDoc in the lab, has performed the native pull-down of OGT protein complexes from insoluble chromatin, nuclear extract and cytoplasmic cellular compartments in OGT$^{FLAG/HA}$ mESCs. In contrast to formaldehyde crosslinking based protein immunoprecipitation (IP), native-IP allows to achieve higher specificity and lower signal background by avoiding non-specific protein interactions (O'Neill LP, Methods 2003). Due to the higher binding affinity of HA tag over FLAG tag, and due to the experimental advantages of using HA-magnetic beads instead of FLAG-agarose beads, we have decided to focus on the anti-HA OGT pull-down only.

The efficiency and specificity of OGT immunoprecipitation (IP) was verified for each fraction by total protein Ruby staining and by immunodetection of OGT using WB (Fig. 5A). Purified OGT's native protein complexes from insoluble chromatin, nuclear extract and cytoplasmic fractions were analysed by quantitative Mass Spectrometry (MS/MS). In collaboration with the EMBL ProteomicCore Facility, I have implemented the analysis of the OGT pull-down MS/MS data. The OGT-IP analysis identified with high confidence (FDR > 0.05) 44, 108 and 145 OGT-interactors from the insoluble chromatin, nuclear extract and cytosol compartments respectively (Fig. 5B first panel), while no significant proteins were detected from the WT negative control.

Analysis of the OGT's protein interactors identified several factors previously characterized as stable partners of OGT, namely HCFC1, TET1 and TET2 (Vella et al., 2013): all three factors were identified among the strongest OGT interactors in all subcellular fractions, validating the efficiency of our native OGT-IP protocol (Fig. 5B, second panel). BAP1, the core component of the Polycomb Repressive Deubiquitinase (PR-DUB) complex (Ge et al., 2023; Thomas et al., 2023) was also detected among the OGT partners in all analysed cellular compartments, in the cytoplasm together with its binding partner ASXL1. Our data are consistent with previous results which identified OGT as a PR-DUB subunit in mammalian cells (Conway et al., 2021).

The nuclear pore complex (NPC) protein NUP62L was detected among the top hits in all three cellular fractions. The nuclear pore proteins are well known to be highly modified by O-GlcNAc residues which were proposed to control the protein trafficking between the cytosolic and nuclear compartments (Yoo and Mitchison, 2021).

In the cytosolic fraction, several proteins of the mitochondrial respiratory chain complex were purified together with OGT. A functional interaction between OGT and the mitochondria was already suggested (Hanover et al., 2003), and some studies have attributed it to the mitochondrial splice variant of OGT (mOGT) (Bond and Hanover, 2013; Love et al., 2003; Lubas et al., 1997). A recent study has characterized a new function for OGT in maintaining the mitochondrial fitness by restricting the proteasomal activity and by supervising the homeostasis of the intracellular amino acid levels (Li et al., 2023). Of note, several mitochondrial proteins were also described to be O-GlcNAcylated (Zhao et al., 2016). The interaction of OGT with mitochondrial proteins supports the view that OGT could be directly involved in regulating the mitochondrial homeostasis.

The effectiveness of the native immunoprecipitation method is validated by the unbiased identification of OGT as one of the most enriched protein in all fractions (Fig. 5C). Among the OGT-IP top hits, PROSER1 was identified as a new strong OGT interactor in all analysed compartments (Fig. 5C). *Proser1* encodes for a conserved protein containing a proline and serine rich region with uncharacterized function. Interestingly, PROSER1 was found to interact with UTX in complex with TET2 and OGT, where it was shown to mediate TET2-OGT interaction (Wang et al., 2022a). To assess the significance of OGT-PROSER1 interaction, and to investigate whether PROSER1 acts as an OGT cofactor or as an OGT substrate, I have decided to generate a *Proser1* KO cell line by CRISPR-Cas9 knock-out with two sgRNAs against *Proser1* exon 4. Using *Proser1* KO cells, I will

assess the effects of PROSER1 loss on OGT activity and on global O-GlcNAc levels (the experiment will be implemented after the submission of my PhD thesis).

Surprisingly, unbiased proteomic analysis revealed a prominent enrichment of both large and small ribosomal proteins (RPLs) and (RPSs) in the nuclear extract fraction. Importantly, the nuclear fractions were treated with benzonase prior immunoprecipitation which excludes indirect binding through nucleic acid bridges. O-GlcNAcylation of ribosomal proteins was previously described in rat pancreatic tissue and in human osteosarcoma cells (Ohn et al., 2008; Zeidan et al., 2010), however no previous data have described a direct interaction between OGT and ribosomal proteins. Ribosomal biogenesis is the process that regulates ribosome production in response to cellular growth, cellular division and nutrient availability. This process takes place in the nucleolus, where rRNA is synthetized and assembled with ribosomal proteins by over 200 nucleolar proteins (Burger and Eick, 2013). To assess whether OGT interaction is specific to RPs or whether it is extended to other nucleolar proteins, I have overlapped the OGT-IP hits with a published human nucleolar proteome Mass Spec dataset (Andersen et al., 2005). Surprisingly, almost half of OGT interactors in the insoluble chromatin and nuclear extract compartments resulted to be nucleolar proteins (Fig. 5B, third panel).

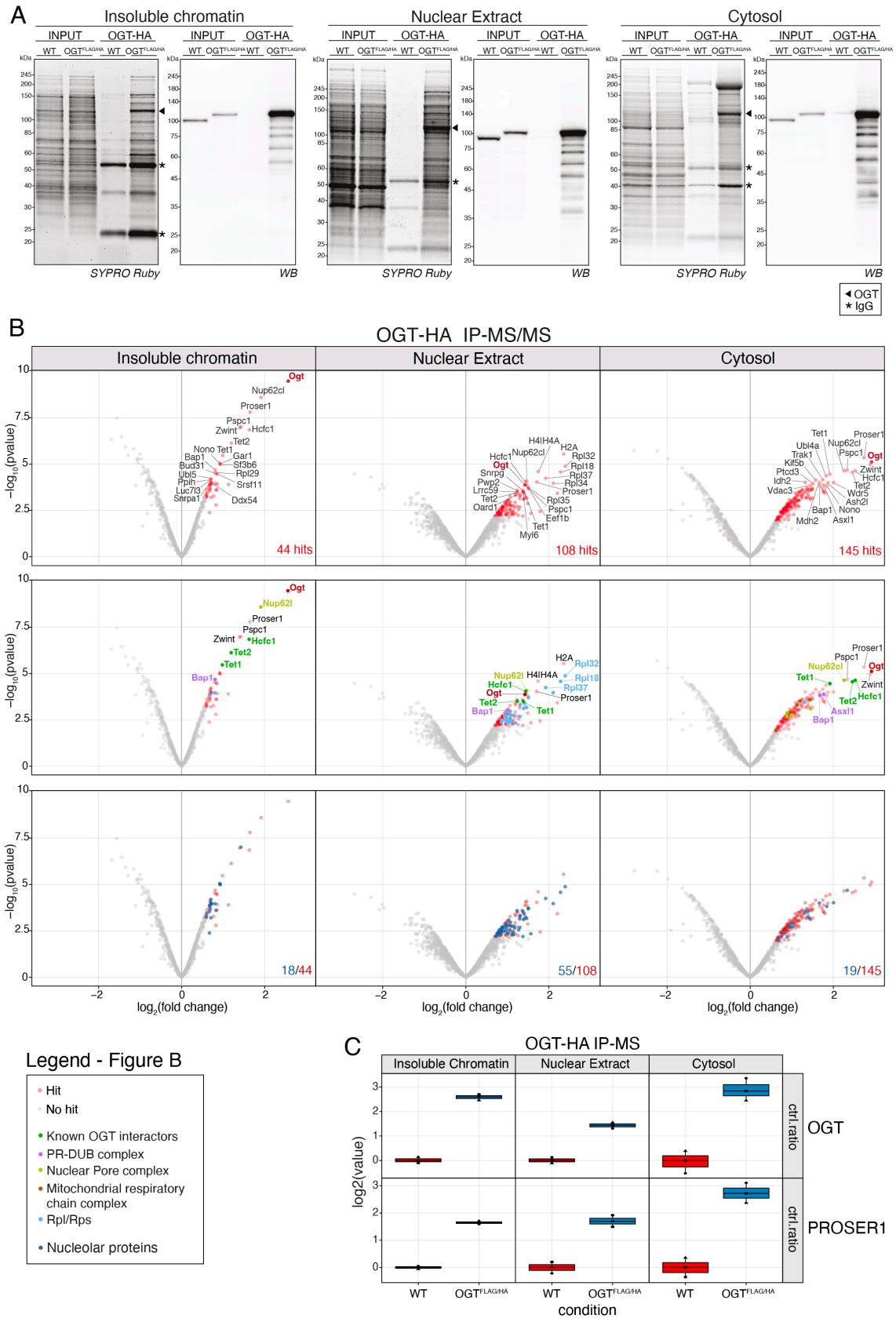

Fig. 5 (A) SYPRO Ruby staining and Western Blot analysis of HA- OGT IP from insoluble chromatin, nuclear extract and cytosolic fractions. For cytosolic and nuclear fractions, 3 μg of IP INPUT protein material were loaded, for nuclear insoluble fraction 1 μg was loaded. 7% of HA- OGT IP protein material was loaded for

Western Blotting. OGT WB membranes were probed with ab177941 OGT antibody. WT cells were used as a control. Experiment implemented by Yulia Ermakova. **(B)** Volcano plots of MS/MS limma analysis of HA-OGT IP showing the 20 top hits (first panel), protein complexes (second panel) and nucleolar proteins (third panel). **(C)** OGT and PROSER1 enrichment in insoluble chromatin, nuclear extract and cytosolic fractions.

To investigate whether OGT interactors form specific protein networks, I have queried the OGT protein partners for protein-protein interactions using STRING database (Szklarczyk et al., 2019). The analysis highlighted distinct protein networks specific for each cellular compartment.

In the insoluble chromatin compartment, the majority of OGT partners separate into two main protein networks: the mRNA splicing and the chromatin organization and histone modification protein pathway, with Mcrs1 and Kansl3 both components of the histone acetyltransferase complex (Fig. 6A).

In line with the previous analysis (Fig. 5B), most of OGT interactors of the nuclear fraction are part in the rRNA binding and RNA processing pathways (Fig. 6B). A less numerous group of proteins is instead characterized by chromatin proteins, among which CXXC1 protein was also identified (also known as CFP1). CXXC1 is a transcriptional activator which binds specifically to non-methylated CpG motifs through its CXXC domain and stimulates trimethylation at H3K4 (Clouaire et al., 2012; Lee and Skalnik, 2005). Intriguingly, TET proteins (TET1 and TET3 have a CXXC domain) and CXXC1 bind to non-methylated rich CpG sequences. These results suggest that OGT might cooperate with TETs and CXXC1 in influencing chromatin composition and perhaps regulating gene expression at non-methylated CpG-dense promoters.

In the cytoplasm, OGT interactors cluster into two distinct protein networks: the histone modification proteins group and the predominant cluster constituted by mitochondrial proteins (Fig. 6C). It is intriguing to notice that OGT interacts with histone modification proteins, TETs, PR-DUB complex, HCFC1 and CXXC1 in both nuclear and cytoplasmic compartments: this indicates that OGT might be essential to ensure the stability of these protein complexes and/or to maintain their O-GlcNAc state not only in the nuclear compartment where these proteins are mainly active, but during their entire protein lifetime. Interestingly, two novel components of the histone methyltransferase complex that methylates H3K4, namely ASH2L and SETD1A, were also identified to interact with OGT.

Taken together, the OGT-IP results validated the previously reported OGT's partners and revealed novel factors that form stable protein complexes with OGT. Among these,

an uncharacterized protein, PROSER1, is one of the most abundant OGT's partners; whether PROSER1 coordinates the protein-protein interaction of OGT with TET1 and TET2 or whether it has other OGT-specific functions still remains to be elucidated. It is of particular interest that several chromatin modifiers were found to interact with OGT not only in the insoluble chromatin and nuclear compartments, but also in the cytoplasm, suggesting a role for OGT in coordinating gene expression at different cellular levels.
The O-GlcNAcylation state of OGT partners will be further investigated by ETD Mass Spectrometry (chapter 4.11).

Protein network analysis of OGT protein complexes highlighted three major OGT partner networks: 1) on chromatin, OGT interacts mainly with chromatin modifiers; 2) in the cytoplasm, the predominant group of proteins that form a complex with OGT are mitochondrial proteins; 3) in the nucleus instead, OGT interacts with a large number of ribosomal and nucleolar proteins, suggesting that the prevailing function of nuclear OGT might be the regulation of ribosomal biogenesis and ribosomal homeostasis. The direct role of OGT and O-GlcNAc in ribosomal production was further investigated and will be elaborated in 4.8.2 and 4.9.

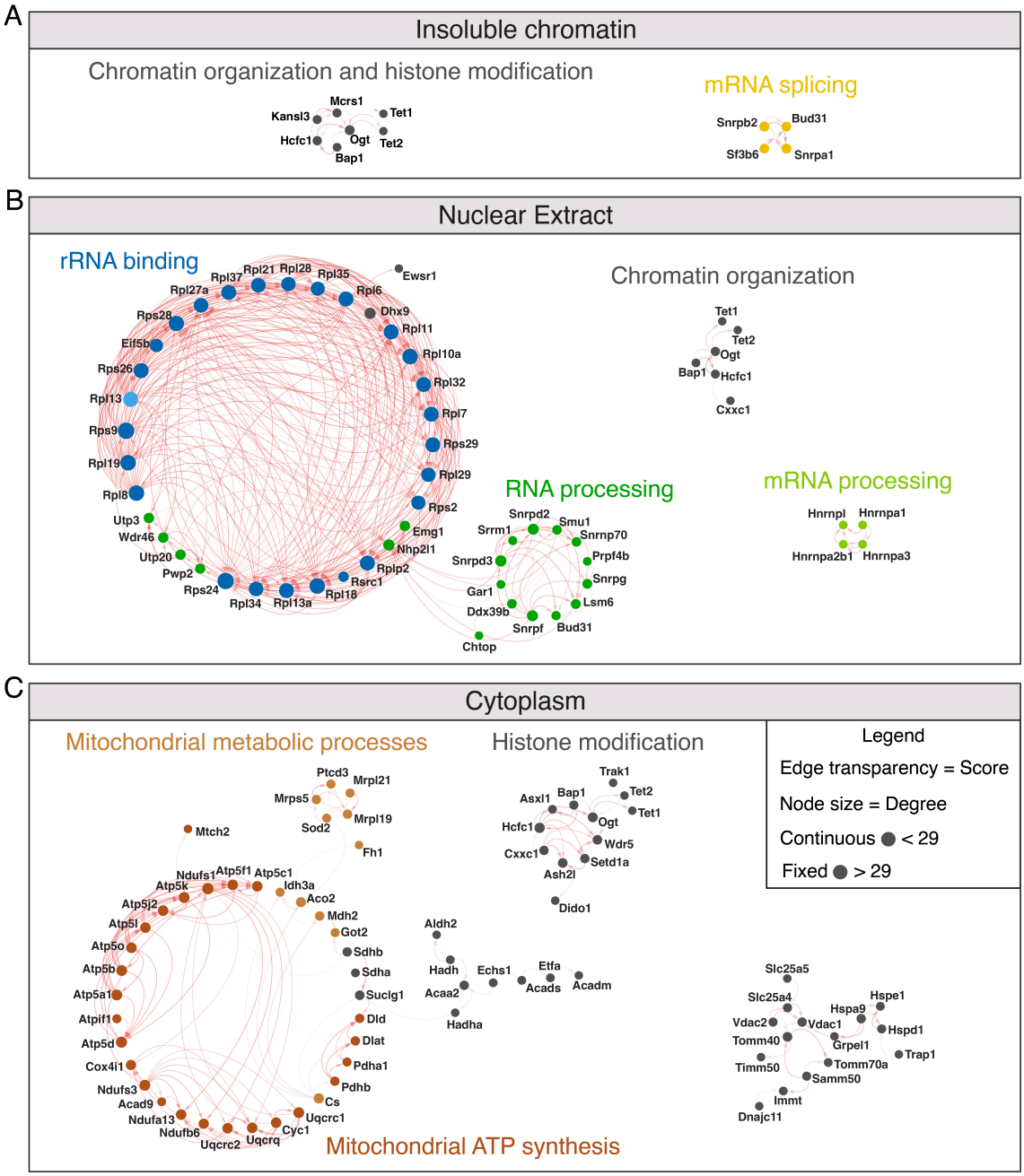

Fig. 6 STRING clustering analysis of the 44, 108 and 145 identified OGT interactors in insoluble chromatin **(A)**, nuclear extract **(B)** and cytosolic **(C)** compartments respectively. Functional interactions based on STRING database are shown with red lines between proteins. The darker red lines depict interactions between clustered proteins uncovered by Markov clustering (MCL).

4.2 Chromatin profiling of O-GlcNAcylated proteins

The investigation of OGT interactome identified several proteins involved in chromatin organization and histone modifications. In line with my results, several proteins implicated in transcriptional regulation has been mapped as O-GlcNAcylated proteins (Gambetta and Müller, 2015). These results indicate that OGT and O-GlcNAc modification have a direct role in chromatin regulation. A first step towards the understanding of the role of O-GlcNAcylation in regulating gene expression is the identification of the genomic compartments occupied by O-GlcNAc proteins. Towards this goal, I profiled O-GlcNAc on chromatin using antibodies that binds the O-GlcNAc group independently from the identity of the modified protein, enabling the characterization of the loci prevalently occupied by O-GlcNAcylated proteins. Of note, previously published pan-O-GlcNAc chromatin profiling experiments are scarce and often discordant (Deplus et al., 2013; Gambetta et al., 2009). Furthermore, little is known about the chromatin O-GlcNAc dynamic during cellular differentiation. In fact, glucose metabolism has a crucial role during the process of cellular differentiation and proliferation (Conaghan et al., 1993; Miyazawa and Aulehla, 2018) and metabolic intermediates of the glucose pathway can serve as signalling molecules to regulate gene expression during stem cell differentiation (Harvey et al., 2019; Lu and Craig, 2012). A prerequisite to understand the role of O-GlcNAc in transcriptional regulation during mammalian embryonic development is to establish a reproducible cellular differentiation system where the profiling of O-GlcNAc modification can be implemented.

4.2.1 Establishment of a cellular differentiation system for the investigation of O-GlcNAc role in transcription

Mouse embryonic stem cells (ESCs) are pluripotent cells isolated from the inner cell mass (ICM) of the mouse blastocyst which self-renew and can differentiate into the ectoderm, endoderm, and mesoderm germ layers (Rossant and Tam, 2009). Mouse ESCs are a well-established model system to study the first stages of embryonic development. Because of their capacity to differentiate into cellular derivatives, ESCs also represent an excellent system to induce *in vitro* differentiation of tissue-specific cell types.

To characterize the O-GlcNAc dynamics during pluripotency progression, I implemented a differentiation system to mimic the in-culture mESCs transition to a morphologically

and transcriptomically distinguishable cell type, and compare the O-GlcNAc state between the two cell stages. NPCs are neuronal progenitor cells of the central nervous system (CNS) capable of dividing a limited number of times and with the ability to further differentiate in specific neuronal and glial cell types. The ESCs to NPCs differentiation can be obtained with a rapid and reproducible protocol; in only 5 days, ESCs differentiate in mature NPCs with significant morphological and gene expression changes. NPCs differentiation is induced by replacing the leukemia inhibitor factor (LIF) and the 2i components with the basic fibroblast growth factor (bFGF) for 72h, followed by replacement of bFGF with the smoothened agonist (SAG) for 48h (Fig. 7A) (Petracovici and Bonasio, 2021).

The efficiency of NPC differentiation was validated according to three measurements: microscopy inspection of cell morphology (Fig. 7B), quantification of markers using WB (Fig. 7C) and RNA-seq analysis (Fig. 7D). After 5 days, NPCs become morphologically distinguishable from ECSs due to the typical spiny processes formation (Fig. 7A), while at the transcriptomic level the cells undergo a simultaneous downregulation of pluripotency genes and an upregulation of neuronal genes (Fig. 7D). Interestingly, gene expression analysis of ESCs and NPCs showed an upregulation of *Ogt* and *Oga* genes from ESCs to NPCs, suggesting possible differences in O-GlcNAcylation between the two cell types (Fig. 7E). After validating the differentiation protocol, I went ahead and used the ESCs and NPCs cellular systems as a differentiation model to explore the O-GlcNAc dynamics and the O-GlcNAc function in transcriptional regulation during cellular differentiation.

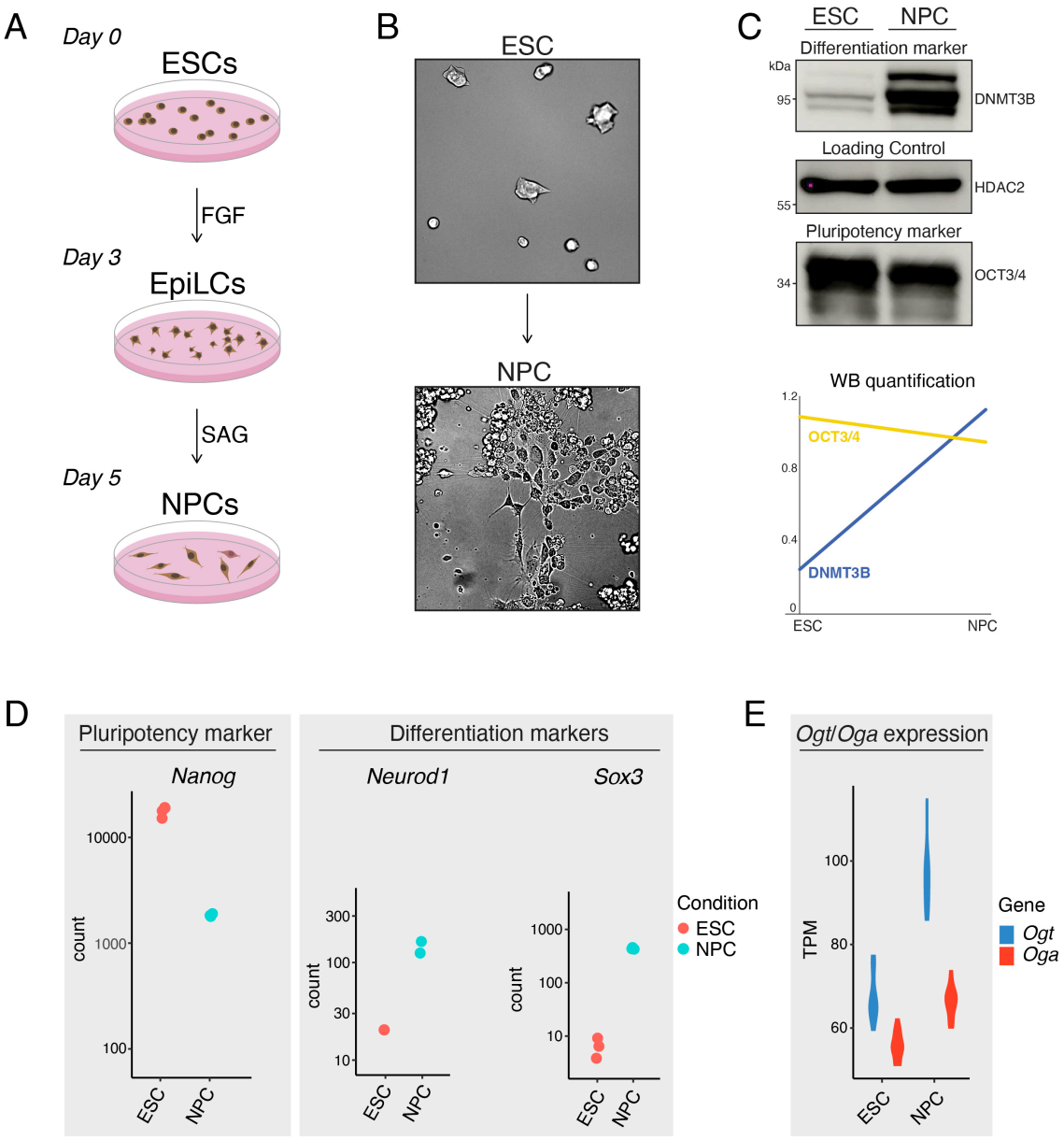

Fig. 7 (A) Schematic representation of ESCs to NPCs differentiation protocol. **(B)** Brightfield images of cells at ESC and NPC stages performed by Yulia Ermakova. **(C)** WB analysis and quantification of ESCs to NPCs differentiation using DNMT3B (ab2851) and OCT3/4 (AF1759) markers. **(D)** Dot plot of RNA-seq analysis of ESCs and NPCs for pluripotency and differentiation markers. Analysis performed by Sara Formichetti. **(E)** Dot plot of RNA-seq analysis of *Ogt* and *Oga* expression in ESCs and NPCs. Analysis performed by Sara Formichetti.

4.3 Chromatin profiling of O-GlcNAcylated proteins by ChIP-seq

To establish the first epigenomic map of O-GlcNAc proteins in mammalian cells, I have performed chromatin immunoprecipitation (ChIP) followed by NGS sequencing of O-GlcNAc proteins with a pan O-GlcNAc antibody which was previously shown to be specific for O-GlcNAc modification, but not tested yet in mESCs (Maria Cristina Gambetta, 2009). Bioinformatic analysis of O-GlcNAc ChIP-seq dataset identified 46718, 31803 and 22650 peaks for the three Chip-seq replicates and revealed three major genomic compartments occupied by O-GlcNAc proteins, namely LTR sequences, imprinting control regions (ICRs) and gene promoters (Fig. 8).

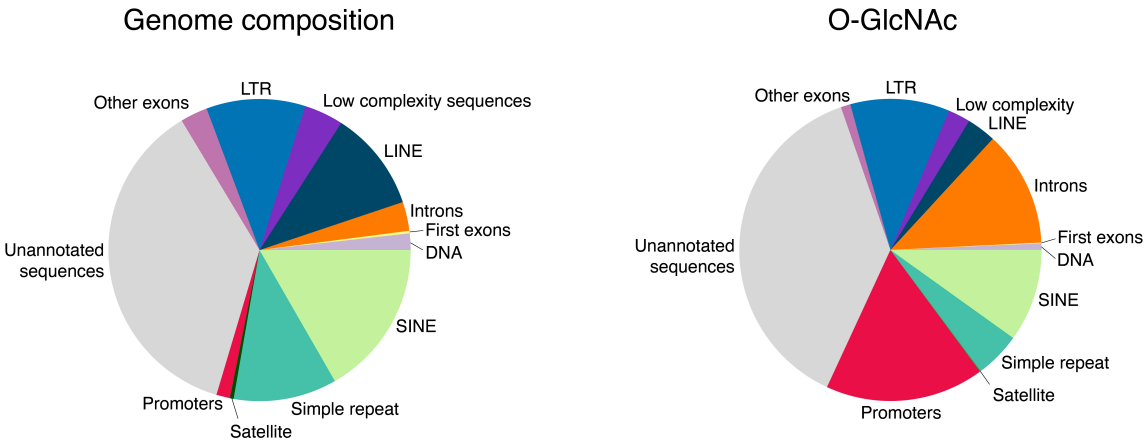

Fig. 8 Pie-chart showing the O-GlcNAc ChIP-seq signal distribution in different genome compartments (right) in comparison with the mouse genome (left). O-GlcNAc ChIP-seq was performed in triplicates using O-GlcNAc HGAC85 antibody on E14 WT mESCs cultured in T2i conditions. The bioinformatic analysis were performed by Nicolas Descostes.

4.3.1 Long Terminal Repeats (LTR) sequences of retrotransposons

Retrotransposons are heterogenous classes of integrated retroviral sequences that can change their genomic location by a copy/paste transposition mechanism. Mammalian cells utilize cytosine methylation to selectively repress retrotransposons, however it remains unknown how methyl-cytosines represses transcription. The analysis of O-GlcNAc ChIP-seq data unveiled an occupancy of O-GlcNAc proteins at long terminal repeats (LTR) of intracisternal A-particle (IAP), the most active type of retrotransposons in the mouse genome (Rebollo et al., 2020) (Fig. 9). Retrotransposon sequence is composed by genes encoding for GAG and POL proteins (reverse transcriptase necessary for their autonomous retrotransposition) flanked by long terminal repeat (LTR)

sequences. O-GlcNAc signal revealed to be prevalently enriched at the 5' LTR, which contains the U3 promoter region of the IAP element, and at the GAG encoding sequence (Fig. 9A). Interestingly the peak of O-GlcNAc in the *Gag* region coincides with the SHIN region that was shown to be sufficient to promote heterochromatinization (Sadic et al., 2015). Comparative analysis with available TRIM28 ChIP-seq data (Elsässer et al., 2015) revealed that the LTRs occupied by O-GlcNAc proteins are also occupied by TRIM28, the core repressor of proviruses. All classes of LTRs examined showed an enrichment of O-GlcNAc proteins and TRIM28. On the contrary, DNA transposons, which do not have the capacity of being transcribed, are not enriched by O-GlcNAc proteins (Fig. 9B, Fig. 9C). These results, together with the evidences of IAP reactivation after targeted removal of O-GlcNAc from their promoter regions (Boulard et al., 2020), highlight a new role of O-GlcNAc in repressing methylated retrotransposons. I published these results as the second author in Boulard et al., PNAS 2020 (Boulard et al., 2020).

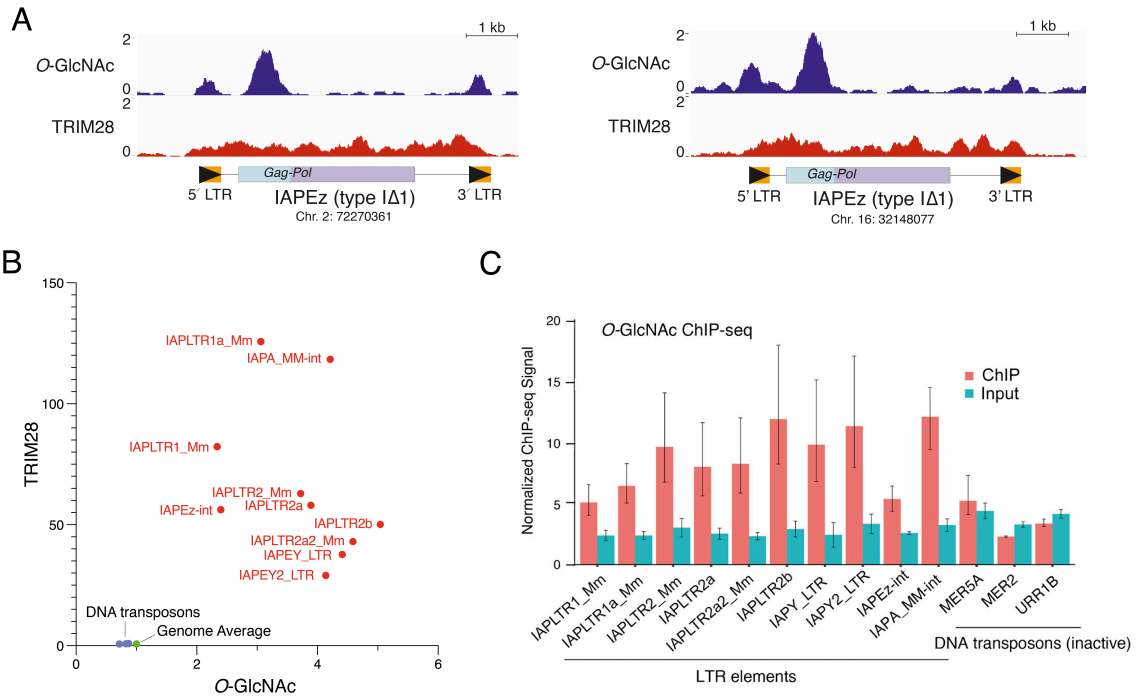

Fig. 9 (A) Genome browser view of O-GlcNAc ChIP-seq data showing the enrichment of O-GlcNAc proteins at the 5' LTR promoter region and the colocalization of O-GlcNAc and TRIM28 at IAPEz retrotransposons. IAP class I is a 7-kb full-length element containing three ORFs (e.g. Gag, Prt, and Pol); class IΔ1 has one single ORF that encodes a novel GAG-POL fusion protein. **(B)** O-GlcNAc and TRIM28 co-occupancy at several classes of IAP LTR elements. DNA transposons and genome average are considered as reference. **(C)** Bar plot showing the occupancy of O-GlcNAc proteins at different classes of IAP LTR elements comparing to three classes of DNA transposons, which are inactive and not enriched in O-GlcNAc proteins. The bioinformatic analysis were performed by John R. Edwards (Washington University, MO, USA).

4.3.2 Imprinting control regions (ICRs)

Genomic imprinting refers to the parental-origin mono-allelic specific expression of ~200 mammalian genes (Ferguson-Smith, 2011). Imprinting genes are often clustered around a single imprinting control region (ICR) which regulates their monoallelic expression. The repressed allele is characterized by allele-specific-methylation, either on the maternal allele (paternal imprinting), or on the paternal allele (maternal imprinting). The allele-specific methylation of ICRs orchestrates the recruitment of proteins and/or ncRNAs that further regulate the mono-allelic gene expression. O-GlcNAc ChIP-seq revealed an enrichment of O-GlcNAc proteins at the majority of ICRs of imprinting gene clusters. As for LTR sequences, O-GlcNAc enrichment overlaps with TRIM28 occupancy (Fig. 10A, Fig. 10B); in fact, TRIM28 is the only known trans-acting repressor of imprinted genes (Alexander et al., 2015; Quenneville et al., 2011). Based on these results, we hypothesized that O-GlcNAc at ICRs could regulate genomic imprinting in a methylation-dependent manner. To assess whether O-GlcNAc is enriched on the maternal and/or paternal allele at ICRs, I have performed a new O-GlcNAc ChIP-seq experiment using F1-hybrid mESCs. The cells were derived from the first generation (F1) of *Mus-Castaneus*/*Mus Musculus* CD1 mice crossing. Due to the presence of multiple SNPs between the two strains, the DNA sequence of maternal and paternal origin can be discriminated after NGS sequencing. The data were analyzed in collaboration with Sara Formichetti using AllelomePRO, a pipeline specifically designed to infer allele-specific genomic features from high-throughput sequencing data (Quenneville et al., 2011). The replication of the O-GlcNAc ChIP-seq with other cells confirmed that most ICRs are enriched by O-GlcNAc proteins (Fig. 10C). However only 3 ICRs (*Peg13*, *Peg3* and *H13*) had informative SNPS; at these three ICRs, O-GlcNAc enrichment showed to be biased towards the unmethylated allele (Fig. 10D). This result indicates that O-GlcNAcylated proteins may be involved in the cis-regulation of the transcribed allele rather than in the silencing of the methylated allele. To assess whether O-GlcNAc perturbation alters the methylated allele repression, together with Sara Formichetti, we quantified the allelic expression of two imprinted genes (*Snrpn* and *Impact*) using a droplet-based PCR (ddPCR) assay. We used *CAST/EiJ (CAST)* x *FVB/NCrlEmr (FVB)* F1 hybrid MEFs, transgenically modified by the auxin-inducible destabilizing domain (AID) fused to OGT protein, and by *OsTIR* gene (OGT-degron system). The addition of the plant hormone Auxin to the cell media induces the Auxin-binding receptor TIR1 to target the AID-protein for proteasomal degradation. Our degron system successfully depleted OGT and O-GlcNAc levels in MEFs (data not shown). Previous RNA-seq experiment performed in

the lab on *OgtNterAID/Y; Gt(ROSA)26OsTIR* male MEFs showed an up regulation of the maternally imprinted genes *Impact*, *Snrpn*, *Peg10* and *Nnat* imprinting genes upon OGT degradation (data not shown). We have performed ddPCR assay in *OgtNterAID/Y; Gt(ROSA)26OsTIR* male MEFs for these genes to test whether O-GlcNAc regulates their monoallelic expression. The allelic quantification of mRNAs showed no reactivation of the silenced allele upon induction of OGT depletion (Fig. 10E, here shown for *Impact* gene). While it is not possible to exclude that residual O-GlcNAc after OGT degradation may be sufficient to fulfil its repressive function, our results do not support the hypothesis of a role for GlcNAc in imprinting silencing.

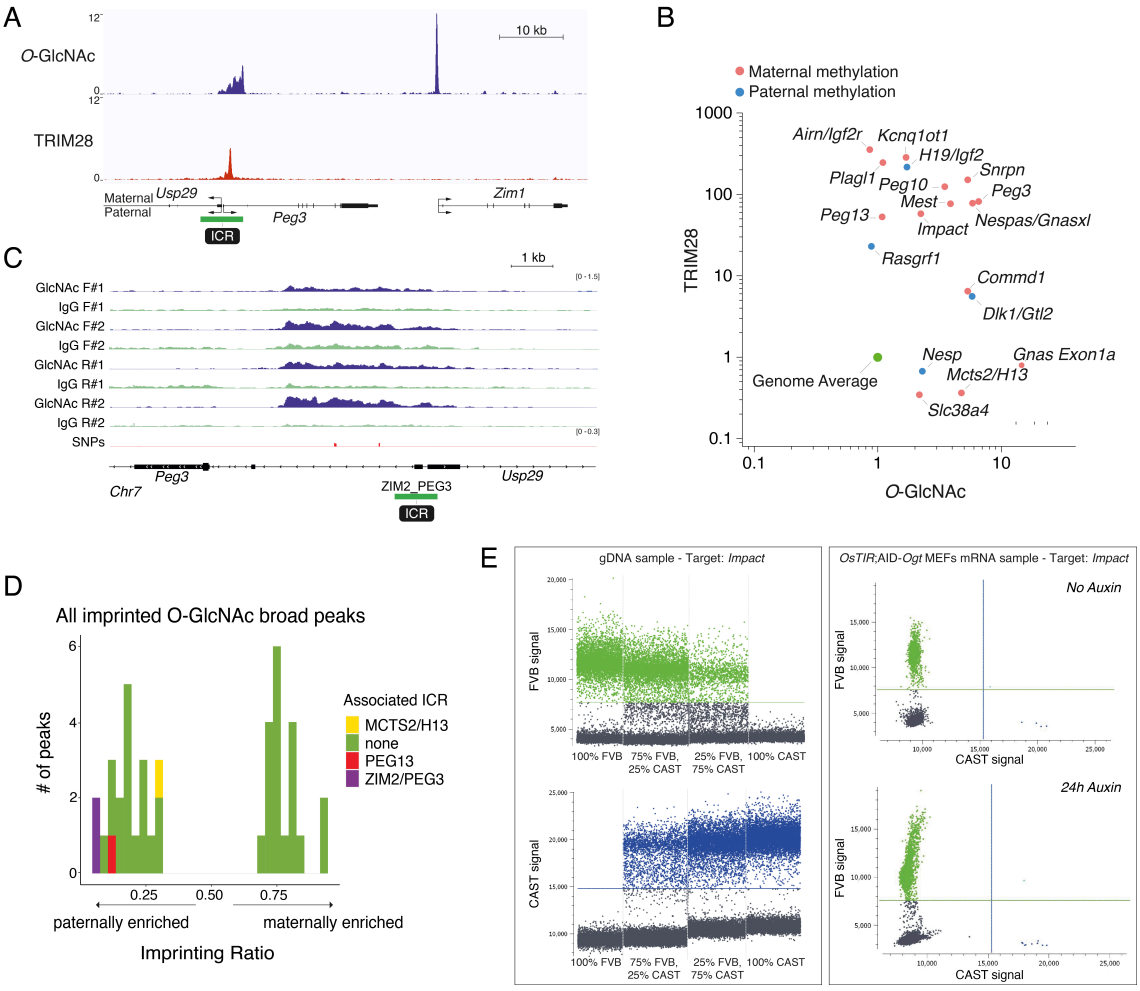

Fig. 10 (A) Genome browser view of O-GlcNAc ChIP-seq data (Boulard et al., 2020) and TRIM28 ChIP-seq data (Elsässer et al., 2015) showing that O-GlcNAc proteins and TRIM28 co-occupy at ICR of the *Peg3* imprinting gene (maternally methylated). **(B)** Correlation analysis showing that TRIM28 and O-GlcNAc are both enriched at the large majority of ICRs. Analysis performed by John R. Edwards. **(C)** Genome browser view of ChIP-seq data from Cast/CD1 F1 hybrid cells confirming the binding of glycosylated proteins at *Peg3* ICR. O-GlcNAc ChIP-seq was performed on mESCs cultured in T2i conditions in duplicates using two clones for both forward (F) and reverse (R) crossings. O-GlcNAc HGAC85 antibody and IgG control were used. Analysis performed by Sara Formichetti. **(D)** Histogram of Imprinting Ratio (IR) values for all broad O-GlcNAc peaks (found with MACS2) which were defined as imprinted (i.e. showing a significant allelic-specific enrichment of the O-GlcNAc signal at FDR < 1%) by the AllelomePRO pipeline in *CAST/CD1* F1 hybrid mESCs. High IR indicates enrichment on the maternal allele, low IR on the paternal allele. Among all germline differentially methylated Imprinted Control Region (ICRs) from Proudhon et al. (Proudhon et al., 2012), only three contained SNPs with enough read coverage to compute an allelic score; they all showed imprinted O-GlcNAc enrichment on the paternally (expressed) allele and are colored differently in the plot. **(E)** ddPCR assay to measure the expression levels of *Impact* from *CAST/EiJ (CAST)* and *FVB/NCrlEmr (FVB)* alleles. In the left panel, proof of principle of allele discrimination using CAST and FVB genomic DNAs mixed at different proportions. In the right panel, the assay is applied to the cDNA of *OgtNterAID/Y; Gt(ROSA)26OsTIR* male MEFs, untreated or treated for 24h with Auxin, for *Impact* gene. In all plots, each dot represents one reaction partition. Experiment and analysis performed by Sara Formichetti.

4.3.3 Gene promoters

Gene promoters are the type of sequences with the most significative enrichment of O-GlcNAc modification. Comparison of GlcNAc levels in the A chromatin compartment (euchromatin) versus B chromatin compartment (heterochromatin) confirmed that GlcNAc is more abundant in euchromatin, which includes promoters (Fig. 11A). To determine whether O-GlcNAc functions as a transcriptional activator or repressor at cellular genes, GlcNAc-rich promoters were subclassified based on their expression levels. Our analysis shows no clear-cut correlation between promoter activity and O-GlcNAc occupancy; however, O-GlcNAc tends to occupy active promoters (Fig. 11B). The complex relationship between O-GlcNAc levels and transcription can be explained by the heterogeneous nature of GlcNAc proteins, which can be part of both activator and repressor complexes.

Collectively, our evidences for O-GlcNAc function at IAPs, ICRs and promoters genomic compartments show no univocal role for O-GlcNAcylation in positively or negatively regulating transcription (Boulard et al., 2020) (Fig. 10E, Fig. 11B). Because GlcNAc-modified proteins are highly enriched at Polycomb response elements (PRE) in *Drosophila* (Gambetta and Müller, 2014), we inspected whether mouse O-GlcNAcylated proteins also correlate with Polycomb Repressive Complex (PRC) subunits. The data show that while O-GlcNAcylated proteins are present at Polycomb target genes (e.g. developmental genes poised for activation in specific lineages), the O-GlcNAc peaks do not overlap with that of the Polycomb repressive complexes (PRC) (Fig. 11C). Therefore, it is very unlikely that PRC subunits could be significantly modified by O-GlcNAc in the mouse. Of note, it was previously shown that EZH2, a PRC2 subunit, interacts with OGT and that EZH2 O-GlcNAcylation modulates EZH2-directed H3K27me3 deposition (Chu et al., 2014). Moreover, *Eed–/–* and *Suz12–/–* affect nuclear OGT and O-GlcNAc levels in ESCs, while *Ogt* knock-down does not interfere with *Ezh2* expression and H3K27me3 deposition (Myers et al., 2011). Overall, these results suggest that the feedback regulatory mechanisms between OGT and PRCs in mammals affect only a specific subset of genes, however the specific molecular mechanisms remain still enigmatic.

To gain insight into the role of O-GlcNAc at promoters, we have performed a meta-analysis to identify overlapping chromatin features (e.g. histone modification and TF binding) with O-GlcNAc peaks: unbiased bioinformatic studies can establish the binding co-occurrence between transcription factors (TFs) and O-GlcNAc at selected DNA sequences. We have here performed unbiased screening analysis between O-GlcNAc ChIP-seq dataset and other 17526 public ChIP-seq datasets from ChIP-Atlas database

(Oki et al., 2018). Our analysis identified 8 factors which chromatin occupancy significantly correlates with O-GlcNAc enrichment at promoters, and at LTRs, SINEs and LINEs retrotransposon classes (Fig. 11D). The identified factors are listed below:

1. RNA Pol II, a DNA-dependent RNA polymerase responsible for transcription of nuclear genes. As discussed in section 1.8, RNA Pol II was previously reported to be modified by O-GlcNAc. It is perhaps not surprising that RNA Pol II is the top hit considering that it is made by a repeat of 52 hepta-peptides that can be in principle all modified and therefore amplify the O-GlcNAc ChIP-seq signal.
2. DR1, a TATA box-binding protein with transcriptional repressor activity.
3. FBXL10, a histone demethylase of H3 K4 and K36 which occupies CpG-dense promoters via its CXXC domain. FBXL10 protects Polycomb-occupied promoters against ectopic de novo methylation (Boulard et al., 2015).
4. KMT2B, also known as MLL2, is a histone methyltransferase that methylates K4 of histone H3, a mark associated with transcription activation. KMT2B also occupies CpG-dense promoters via its CXXC domain.
5. SIN3A is a transcriptional repressor that acts cooperatively with OGT to repress transcription in parallel with histone deacetylation. Interestingly, SIN3A is known to be a glycoprotein (Myers et al., 2011) and to interact with OGT (Vella et al., 2013).
6. FOXH1, a transcriptional activator which forms a transcriptionally active complex with SMAD2 and SMAD4.
7. SMC1, a protein involved in chromosome cohesion during cell cycle and DNA repair.
8. TBP, a general transcription factor that functions at the core of the DNA-binding multiprotein factor TFIID.

These identified factors are candidates which are potentially modified by O-GlcNAc, however these results do not provide direct evidences of their O-GlcNAc state. It is possible that some candidates are not O-GlcNAcylated but bind to the same genomic loci with other O-GlcNAc factors. Therefore, the O-GlcNAc state of each candidate will be further validated (4.11). Also, these data do not exclude the presence of additional uncharacterized O-GlcNAcylated factors at the analyzed genomic compartments.

Of note, RNA Polymerase II most highly correlates with O-GlcNAc at promoter regions (Fig. 11D). Deeper inspection revealed that, genome-wide, all genes enriched by RNA Pol II at the transcription start sites (TSSs) are also enriched by O-GlcNAcylation (Fig. 11E). The localization of the O-GlcNAc signal at the TSS suggest that O-GlcNAcylation could be involved in the pre-initiation complex assembly, RNA Pol II pausing or initiation of elongation. For these reasons, I have decided to further investigate the O-GlcNAc function on RNA Pol II CTD (4.10).

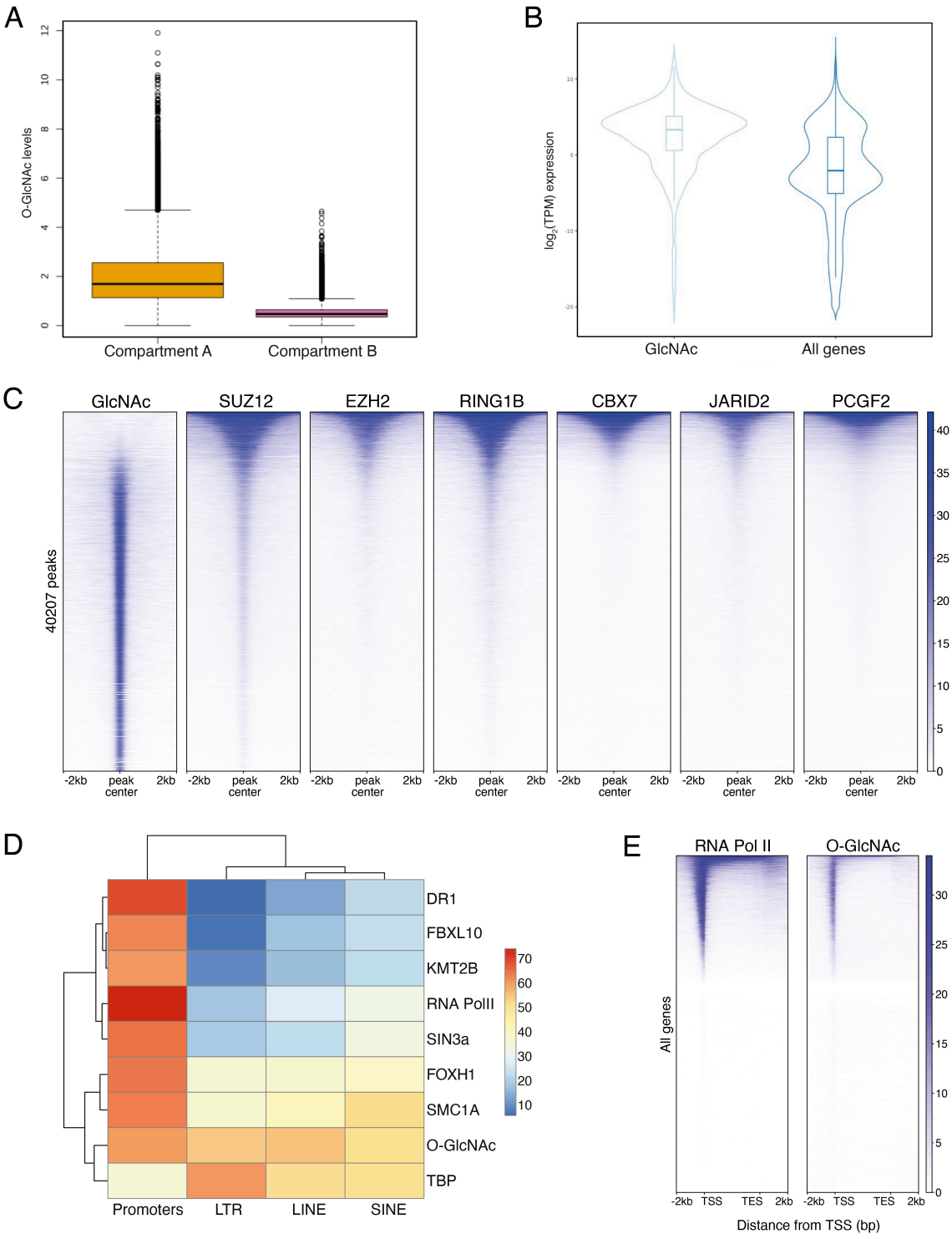

Fig. 11 (A) Box plot showing O-GlcNAc enrichment at compartments A and B. **(B)** Violin plot showing RNA-seq expression levels of O-GlcNAc enriched genes in correlation with all genes. **(C)** Heatmap showing O-GlcNAc enrichment in correlation with SUZ12, EZH2, JARID2 (PRC2 complex), RING1B, CBX7, and PCGF2 (PRC1 complex) Polycomb subunits at all genes. **(D)** Heatmap illustrating the percentage of overlap between O-GlcNAc with 8 top candidates at promoters, LTR, LINE and SINE, found by screening of 17526 datasets of pluripotent stem cells. The results were limited to ESC. **(E)** Heatmap showing the correlation between O-GlcNAc and RNA Pol II enrichment at transcription start site (TSS) of all genes. The bioinformatic analyses were performed by Nicolas Descostes.

4.4 Chromatin profiling of O-GlcNAcylation in ESCs and NPCs

In the previous chapter I have shown that ChIP-seq is effective to characterize the O-GlcNAc genomic map in mESCs. Therefore, together with Jeanne Couturier, Master student in the lab that I have directly supervised for 6 months, I have decided to optimize the ChIP-seq experimental conditions and replicate the experiment in both ESCs and NPCs to investigate the O-GlcNAc dynamics during cellular differentiation.

To accurately compare the differences between the two cell types and normalize the results for the experimental biases, I have included in each sample a spike-in internal control consisting of *Drosophila* chromatin. Moreover, because the previously used chromatin input control revealed to be inaccurate for ChIP-seq data normalization, I have included an IgG negative control to normalize the signal for non-specific binding of the antibody. Finally, to reduce the over amplification of the DNA during library preparation, I have optimized the number of PCR cycles for each sample to maintain the libraries complexity before NGS sequencing.

The analysis of the optimized O-GlcNAc ChIP-seq showed a predominant occupancy of O-GlcNAc modification at gene promoters in both ESCs and NPCs (Fig. 12), validating the previous ChIP-seq results. These results indicate that gene promoters represent the main genomic compartment regulated by O-GlcNAc modification.

Due to the stringent experimental conditions, the global O-GlcNAc signal was considerably reduced in comparison to the first O-GlcNAc ChIP-seq: two ChIP replicates were selected for each cell type, with 444 and 425 peaks for ESCs, and 640 and 794 for NPCs. The comparison of O-GlcNAc enrichment between the two cell types did not highlight any major difference, demonstrating that O-GlcNAc on chromatin proteins remains overall stable during differentiation of pluripotent stem cells to neuronal progenitor cells.

The variability of the O-GlcNAc ChIP-seq signal (in comparison with the previous ChIP-seq) at the other genomic compartments might be explained by several reasons. The genome-wide chromatin profiling of post-translational modifications (PTMs) is a challenging experiment because it is highly dependent on the efficiency of the antibody and on the chromatin context. Moreover, the abundancy and accessibility of the O-GlcNAc sites on the target proteins also affect the antibody target site recognition. For instance, the pull-down efficiency of highly O-GlcNAcylated protein complexes is higher than the pull-down of rare TFs modified by a single O-GlcNAc site; at the step of DNA amplification, the PCR efficiency is thus biased towards the DNA fragments that are more

enriched, while the less abundant DNA fragments will require more cycles to be amplified. Finally, the affinity of pan-PTM antibodies depends on the nature of the peptide and of the O-GlcNAc site used to generate the antibody. For all these reasons, the readout of a PTM ChIP-seq can be often of difficult interpretation and the library preparation is one of the most critical steps to be optimized to preserve the quality of the data.

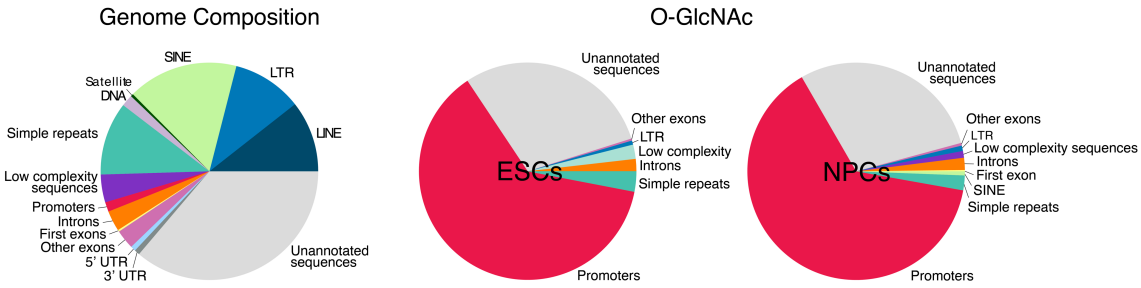

Fig. 12 Pie-chart showing the O-GlcNAc ChIP-seq signal distribution in different genome compartments (right) in comparison with the mouse genome (left). O-GlcNAc ChIP-seq was performed in triplicates using O-GlcNAc HGAC85 antibody on A2loxP mESCs. The experiment was performed in collaboration with Jeanne Couturier, the bioinformatic analysis was performed by Nicolas Descostes.

Overall, my results indicate that ChIP-seq is an efficient approach to characterize the genomic distribution of O-GlcNAc modification in mESCs. With two independent experiments I have demonstrated that O-GlcNAcylated proteins prevalently occupy gene promoters, in accordance with the previous published works (Deplus et al., 2013; Maria Cristina Gambetta, 2009). However, because the ChIP-seq results showed to be sensitive to the experimental conditions, I have decided to test the newly developed Cleavage Under Targets and Release Using Nuclease (CUT&RUN) technique for O-GlcNAc chromatin profiling on fresh cells.

4.5 Chromatin profiling of O-GlcNAcylation by CUT&RUN

CUT&RUN-seq (Cleavage Under Targets and Release Using Nuclease followed by sequencing) is a recently developed approach for the genome-wide profiling of TFs and histone marks (Skene and Henikoff, 2017). Briefly, the interested protein factor or histone modification is profiled in intact cells using a specific antibody which is subsequently recognized by protein A/G (pA/G) recombinantly fused to Micrococcal Nuclease protein (MNase). Binding of the MNase to the factor of interest results in very high local concentration of MNase at the binding site. The DNA regions that flank the target factor are cleaved by MNase and released from the cells through nuclear pores (if nuclei only are used) or through the digitonin-permeabilized plasma membrane (if whole cells are used). The purified DNA can be then analysed by quantitative real time PCR (qPCR) or next generation sequencing (NGS) (Skene and Henikoff, 2017).
CUT&RUN has many advantages over ChIP-seq: 1) it requires a lower input material (up to a single cell, typically 10,000 cells); 2) the profiling does not require prior chromatin crosslinking and sonication, significantly reducing the experimental background; 3) the low input material and decreased background allows the reduction of the sequencing coverage and the overall experimental costs; 4) the experimental procedure is very fast and can be completed in only few days and parallelized.

I have thus optimized the CUT&RUN protocol specifically for O-GlcNAc detection. Because of the O-GlcNAc nature and O-GlcNAc cellular distribution, I hypothesized that CUT&RUN approach could present the following limitations: 1) the dense layer of glycoproteins on the plasma membrane might prevent the antibody from entering the cell surface; 2) the highly O-GlcNAcylated nuclear pore proteins (Xu et al., 2022) could trap the antibody before its entrance through the nuclear membrane barrier; 3) due to its instability, O-GlcNAc modification might be lost during sample processing on unfixed chromatin. To overcome these possible limitations, I have optimized the protocol by implementing the following steps: 1) to avoid the cross-reactivity of the O-GlcNAc antibody with the glycoproteins of the plasma membrane, I have performed CUT&RUN on mESCs nuclei only; 2) because the pA/G-MNase and the antibody diffuse through the nuclear pores, I have further permeabilized the nuclear envelope by testing different concentrations of Triton X-100; 3) to increase the O-GlcNAc modification stability, I have tested the nuclei formaldehyde fixation condition (Zheng and Gehring, 2019).

To implement the CUT&RUN optimization protocol, I have designed a new plasmid and cloned the pA/G-MNase plasmid which was expressed in bacteria and purified by the PEPCore EMBL Facility. The efficiency of pA/G-MNase endonuclease activity was validated by plasmid digestion at different pA/G-MNase concentrations (Fig. 13A).

CUT&RUN experiment was implemented in triplicates in mESCs nuclei using the protocol from Skene and Henikoff, eLife 2017 (Skene and Henikoff, 2017). In parallel, on separate samples I have tested additional conditions: the nuclear permeabilization with 0.1%, 0.2% and 0.5% of Triton X-100 and the chromatin fixation. Nuclear permeabilization resulted in higher efficiency of peak detection comparing to the standard CUT&RUN protocol, where 7337, 5217 and 1687 peaks were detected respectively from 0.1%, 0.2% and 0.5% conditions (in comparison to 1529, 1457 and 1462 peaks from the replicates with unpermeabilized nuclei). CUT&RUN samples permeabilized with 0.1% and 0.2% of Triton X-100 showed the highest peaks overlap and also a good overlap with O-GlcNAc ChIP-seq performed under stringent conditions (Fig. 13B). The samples processed with formaldehyde fixation did not show any DNA enrichment at the library preparation step, thus they were not processed for NGS sequencing.

CUT&RUN profiling analysis revealed a predominant enrichment of O-GlcNAcylation at promoter regions, supporting the previous results obtained from the two independent O-GlcNAc ChIP-seq experiments. Additionally, enhancer and intron compartments showed an enrichment of O-GlcNAc modification, a result that did not emerge from previous ChIP-seq profiling (Fig. 13C, Fig. 13E). Unbiased screening analysis between O-GlcNAc CUT&RUN dataset and other 10740 public ChIP-seq datasets from ChIP-Atlas database (Oki et al., 2018) identified several chromatin binding proteins and histone marks to overlap with O-GlcNAc peaks (Fig. 13D). Among the identified proteins, RNA Pol II resulted to be again one of the top hits that most highly correlates with O-GlcNAc at promoter regions. Analysis of O-GlcNAc distribution between the TSS and the TES (Transcription End Site) revealed that all promoters enriched by RNA Pol II are also enriched by O-GlcNAc, however their occupancy relative to the TSS does not completely overlap (Fig. 13E). While O-GlcNAc signal is mainly detected upstream the TSS, RNA Pol II enrichment is characterized by a bimodal distribution, where a small peak upstream the TSS is followed by a higher peak from the TSS towards the gene body. These data might suggest that RNA Pol II colocalize only upstream the TSS with O-GlcNAc -modified proteins, where O-GlcNAc might be involved in the regulation of RNA Pol II transcription pre-initiation step.

Because gene promoters and enhancers showed the highest O-GlcNAc enrichment distribution, we asked whether O-GlcNAc occupancy correlates with gene expression. Analysis of GlcNAc rich promoter expression levels showed a significant correlation between O-GlcNAc occupancy and gene expression, in line with the previous results (Fig. 13D). Scatterplot representation of each O-GlcNAc gene in correlation with expression data evidenced two main groups of O-GlcNAc promoters: a smaller group of silenced promoters, and a larger cluster of active promoters. These results indicate that O-GlcNAc proteins do not have a unique function in activating or repressing transcription, as it was also demonstrated for IAP elements and imprinting genes; nonetheless, the majority of O-GlcNAc-occupied promoters are actively transcribed. These results are also evident in Fig. 13D which highlights a correlation of O-GlcNAc occupancy with both activating and repressing chromatin marks.

All together, these results demonstrate that the optimized version of CUT&RUN protocol enables to profile O-GlcNAc patterns with higher details than ChIP-seq (e.g enhancers were only detectable with CUT&RUN). CUT&RUN experiment validated gene promoters to be the genomic compartment most prevalently occupied by O-GlcNAc proteins and identified enhancers as new genomic compartments regulated by O-GlcNAcylation. In agreement with OGT's interaction with proteins involved in H3K4 methylation, O-GlcNAc proteins primarily occupy actively transcribed promoters of cellular genes.

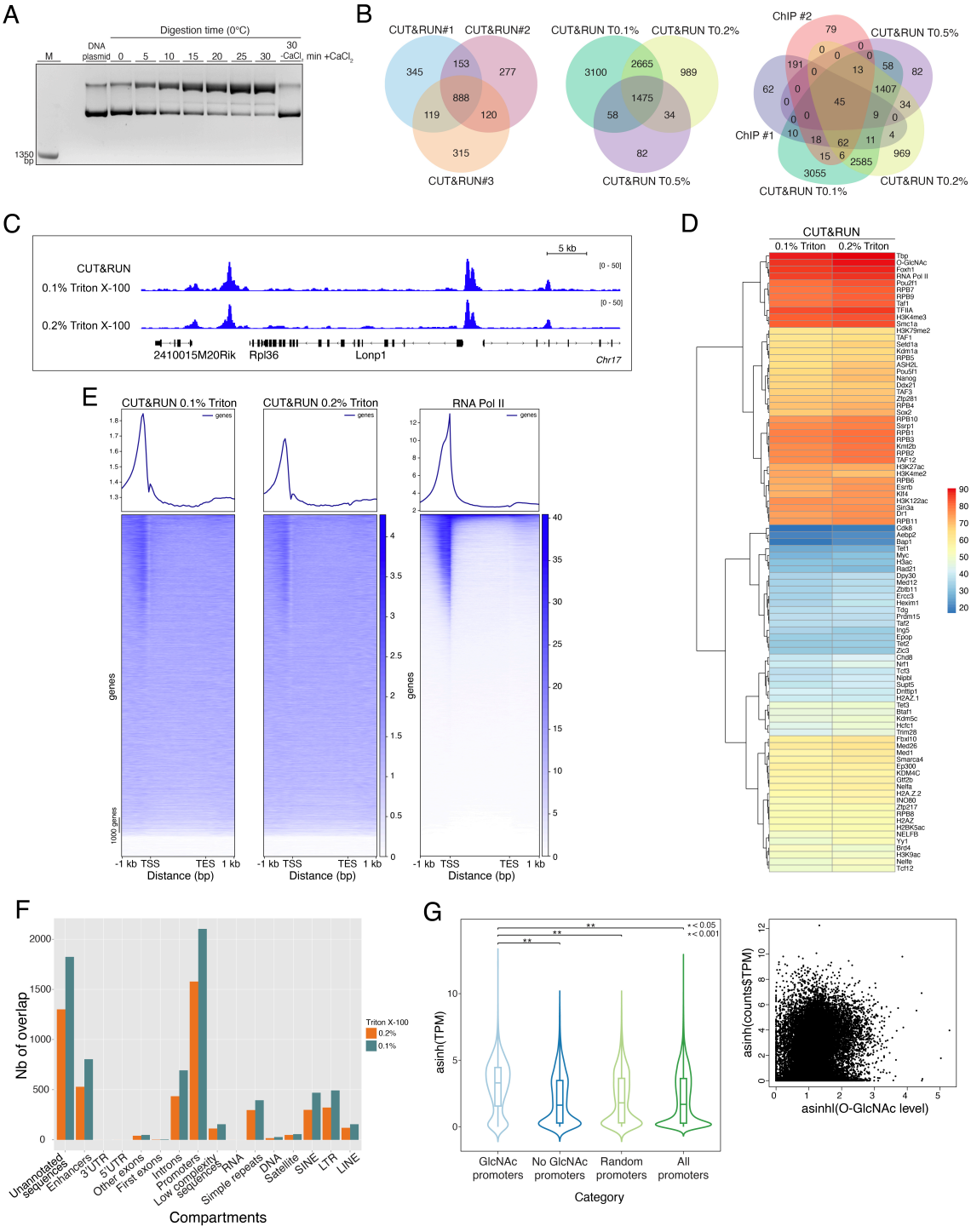

Fig. 13 (A) Time points of pA/G-MNase activity test by plasmid digestion. **(B)** Venn diagrams of peak numbers overlap between O-GlcNAc CUT&RUN experiments and O-GlcNAc ChIP-seq experiments (introduced in paragraph 4.4). **(C)** Examples of O-GlcNAc enrichment at gene promoters by genome browser view. **(D)** Heatmap illustrating the percentage of overlap between O-GlcNAc with other proteins and histone marks candidates at promoters. The results were limited to ESC. **(E)** Heatmap showing the correlation between O-GlcNAc and RNA Pol II enrichment at transcription start site (TSS) and transcription end site (TES) of all genes. **(F)** Histogram of O-GlcNAc CUT&RUN signal distribution at different genomic compartments for 0.1% and 0.2% Triton X-100 CUT&RUN conditions. **(G)** Violin plot and scatter plot showing RNA-seq expression levels of O-GlcNAc enriched genes in correlation with promoters with no GlcNAc signal, random promoters and all promoters. The analysis was performed with mESCs RNA-seq expression data from Gretarsson & Hackett Nature Structural & Molecular Biology 2020 (Gretarsson and Hackett, 2020).

Statistics analyses were performed with Mann-Whitney-Wilcoxon two-sided test. The bioinformatic analyses were performed by Nicolas Descostes.

4.6 Chromatin profiling of OGT

To gain further insights into OGT's function on chromatin, I attempted to profile its genomic occupancy. Previous works have reported an enrichment of OGT at gene promoters, precisely at their TSSs, where it highly overlaps with TET1 and H3K4me3 (Vella et al., 2013). Despite these results, the OGT chromatin occupancy remains still understudied comparing to other nuclear factors. I have performed two parallel chromatin profiling experiments:

1. OGT chromatin profiling by ChIP-seq with double fixation treatment: because OGT is an enzyme, the kinetic of interaction with its protein substrates is very fast. To improve the efficiency of OGT pull-down, I have optimized an OGT ChIP-seq protocol where, before the immunoprecipitation step, the chromatin was fixed twice with two different chemicals.
2. OGT chromatin profiling by CUT&RUN: because O-GlcNAc profiling by CUT&RUN significantly improved the ChIP-seq data quality, I have decided to test the CUT&RUN experiment also for OGT profiling.

OGT was profiled using an OGT^{myc} mES cell line, where 2X *myc* tags sequences were endogenously knocked-in at the *Ogt* C-term. OGT ChIP-seq was performed in triplicates following a previously published protocol (Żylicz et al., 2019), where chromatin is first fixed with formaldehyde and subsequently with succinimidyl succinate (EGS). For OGT CUT&RUN, the experiment was performed in triplicates as previously described (4.5). The immunoprecipitation step of the two experiments was performed using a MYC antibody in OGT^{myc} and WT cells. Moreover, because the efficiency of MYC antibody has not been previously validated by CUT&RUN, the experiment was performed in parallel also with an OGT antibody.
Bioinformatic analysis of OGT ChIP-seq and OGT CUT&RUN datasets showed no OGT enrichment at any chromatin locus (Fig. 14, for simplicity, only one replicate for each experiment is shown). Unfortunately, from these results it was not possible to conclude whether the immunoprecipitation was not successful because OGT does not stably bind

to the DNA, or because of technical limitations. Due to the complexity of the experiment, I have decided to focus only on the profiling of O-GlcNAc by CUT&RUN.

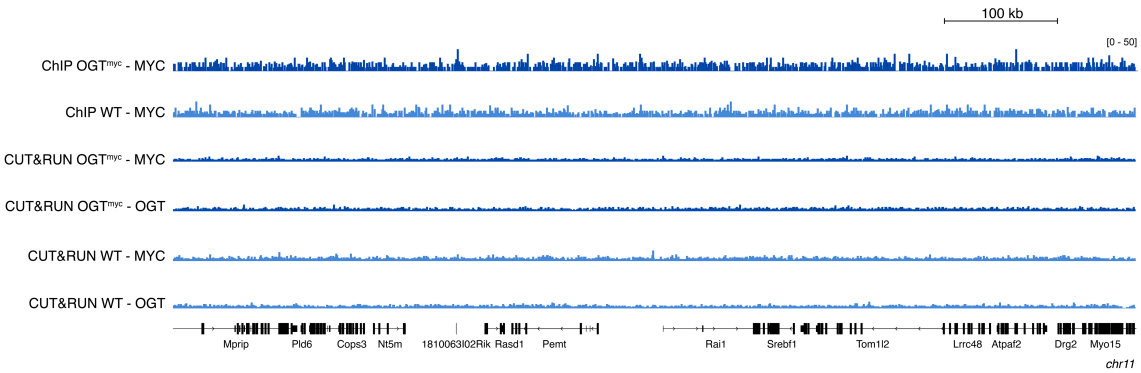

Fig. 14 Genome browser view of OGT ChIP-seq and CUT&RUN data profiled with MYC antibody (9E10) for ChIP-seq, and with both MYC and OGT (ab177941) antibody for CUT&RUN. The tracks show no OGT chromatin enrichment. The browser view is a representative example of the rest of genome.

4.7 Creation of genetically-encoded tools for inducible O-GlcNAc depletion

Little is known about the functional mechanism of O-GlcNAc modification in fine-tuning transcription. The OGT-IP and CUT&RUN results strongly suggest a direct role of O-GlcNAc modification in regulating gene expression, with several chromatin modifiers interacting with OGT and with 2103 promoters covered by O-GlcNAc peaks. Gene expression analyses of O-GlcNAc promoters and of O-GlcNAc at imprinting genes (Fig. 10D) indicate a correlation of O-GlcNAcylation with active transcription, even though O-GlcNAc seems to also be enriched at some repressed promoters (Fig. 9, Fig. 13D).

4.7.1 OGT depletion by siRNA *knock-down*

Ogt's loss-of-function mutation is lethal in both embryo and mESCs, while OGT protein inhibitors have toxic side effects and are often inefficient at non-lethal dose (Hong Zhang, 1992; Trapannone et al., 2016). To overcome these limitations, I implemented alternative strategies to induce OGT's depletion and investigate the resulting effects on gene expression. The first tool I have tested is based on *knock-down* induced by a synthetic small interfering RNA (siRNA). siRNA knock-out approaches have been developed based on the RNA interference conserved pathway used by eukaryotic cells to induce

mRNA molecules degradation by short complementary RNAs. In cells, the RNAi pathway gets initiated by the enzyme DICER which cleaves the long double-stranded RNA (dsRNA) molecules into short double-stranded fragments of 21 to 23 nucleotide siRNAs. The RNA-Induced Silencing Complex (RISC) incorporates the siRNA, which is subsequently unwound into two single-stranded RNAs (ssRNAs) and cleaved by the protein ARGNONAUTE2 (Ago2). The RISC complex, containing the antisense strand of the siRNA, is the active complex responsible for the targeted mRNA degradation (Whitehead et al., 2009). The RNAi eukaryotic pathway was adapted for the first time in 2001 for siRNA -directed sequence-specific gene *knock-down* in mammalian cell lines (Sayda M. Elbashir, 2001).

To induce the *Ogt* knock-down, a synthetic siRNA against *Ogt* mRNA was co-transfected into ESCs for 48h in different growing media, while an unspecific siRNA was used as negative control. siOGT efficiency was validated by WB analysis which revealed a lower level of OGT protein, while no decrease in OGT level was detectable in the negative control and in non-transfected cells (Fig. 15A). Despite these results, the transfected cells showed only partial depletion of OGT, indicating that siRNA-mediated KD is not an efficient approach to induce complete protein degradation and study the effects on gene expression. Therefore, I have decided to implement an alternative OGT degradation approach (4.7.2).

4.7.2 OGT knock-out by dTAG degron system

To obtain the fast and complete degradation of a protein of interest, a novel degradation tag (dTAG) genetic system was recently developed (Nabet et al., 2018). In the dTAG approach, the target protein is fused to a $FKBP12^{F36V}$ E3-binding heterobifunctional degrader; upon addition of the chemical dTAG molecule, the $FKBP12^{F36V}$-fusion complex and the ubiquitin proteasome machinery assemble to a ternary complex which induces the rapid degradation of the fusion-protein complex. The dTAG system leads to the protein degradation within minutes to hours, representing an ideal solution to overcome the OGT KO cellular lethality.
An $Ogt^{Nterm\text{-}dTAG}$ cell line was created by the EMBL Rome GEEF Facility, where the dTAG sequence was fused at the *Ogt* N-term of endogenous *Ogt*. Male ES cells were targeted in order to tag the only *Ogt* allele present, *Ogt* being X-linked. The $Ogt^{Nterm\text{-}dTAG}$ cell line was tested at different time points after dTAG-13 addition for OGT degradation.

However, Western Blot analysis showed no detectable OGT's degradation during the time course treatment (Fig. 15B). Different concentrations of dTAG were tested, with no effect on protein degradation (data not shown). The lack of OGT degradation after dTAg-13 addition could be due to the inaccessible dTAG recognition site caused by an improper protein folding of the FKBP12^{F36V} subunit at the N-term of OGT. Therefore, in collaboration with the GEEF Facility, I have decided to generate a novel cell line where *Ogt* was endogenously tagged at its C-term. The *Ogt*$^{Cterm\text{-}dTAG}$ cell line is currently under production and will be validated for further gene-expression analysis in OGT-depleted cells.

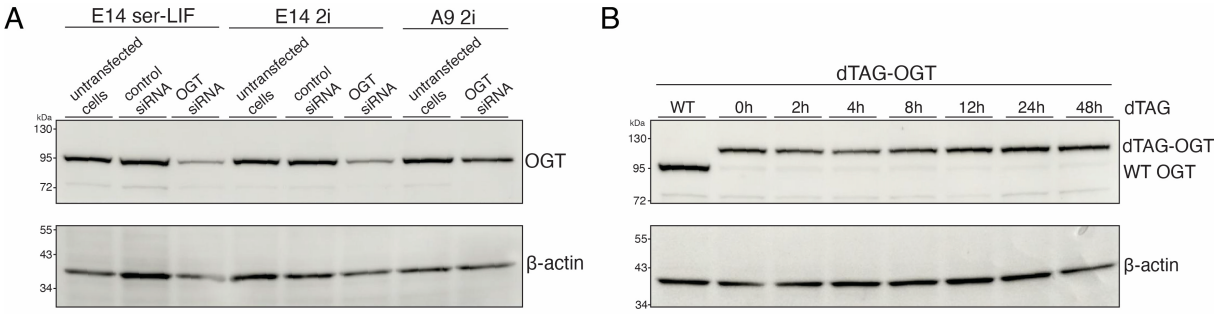

Fig. 15 (A) Western Blot analysis of total protein extraction from siOGT experiment in different growing media. Untransfected cells and transfected cells with an unspecific siRNA were used as a control. β-actin was used as a protein loading control. **(B)** Western Blot analysis of total protein extraction from dTAG-OGT experiment at different timepoints of OGT degradation with dTAG-13. β-actin (A5441) was used as a protein loading control.

4.8 Genetically-encoded effectors for inducible depletion of nuclear O-GlcNAcylation

Because the siRNA and dTAG tools demonstrated to be inefficient for OGT degradation, I implemented alternative genetic tools to perturb the nuclear O-GlcNAc state and investigate the effects on gene expression.

In collaboration with Dr. Yulia Ermakova, we have engineered two mES cell lines based on the tetracycline-controlled Tet-On system. The Tet-On/Tet-Off inducible gene expression system was developed by Prof. Dr. H. Bujard and colleagues at the University of Heidelberg and first described for *in vivo* use in 1992 by M. Grossen and colleagues (Gossen and Bujard, 1992). The Tet-On/Tet-Off system is widely used for conditional expression approaches of (trans)genes because it allows to selectively modulate the expression of a gene of interest under the control of a Tet-On/Tet-Off promoter: only in presence of the antibiotic Doxycycline, the gene expression is respectively switched on/off (Gossen and Bujard, 1992; T. Das et al., 2016).

The mechanism of Tet-On/Tet-Off system is based on regulatory elements that control the activity of the bacterial tetracycline-resistance operon. In natural circumstances, bacteria have different mechanisms to control the tetracycline resistance, and one of them is based on the Tet Operator (TetO). The tetracycline-responsive repressor protein TetR controls the expression of protein TetA, essential to mediate the antibiotic efflux from the cell membrane. TetR and TetA expression are regulated by the P_{tet} promoter and by two superimposed TetO operator sequences. TetR protein forms a dimer which suppress the P_{tet} promoter, thus TetA and TetR expression, through the TetO elements. In presence of antibiotics, TetR dimer is subjected to a conformational change that prevents TetO binding allowing the expression of TetR and TetA. To introduce different effectors in ES cells under the control of a Tet-ON promoter, we took advanrage of the cassette-exchange system previously developed in the Kyba lab (Iacovino et al., 2011).

To induce targeted nuclear O-GlcNAc depletion, we have decided to transgenically express in mESCs the enzyme *BtGH84*, a bacterial *B. thetaiotaomicron* VPI-5482 form of *Oga* which was shown to efficiently process *O*-GlcNAc-modified eukaryotic proteins (Dennis et al., 2006). In our system, we have fused the sequence of *BtGH84* under the control of a Tet-On promoter, allowing the direct control of *BtGH84* expression by Doxycycline. To investigate the O-GlcNAc perturbation effects on gene expression, we have engineered two different Tet-On cell lines: the NLS-BtGH84 cell line for the global

nuclear de-GlcNAcylation, and the CXXC-BtGH84 cell line for the targeted CpG-rich promoter de-GlcNAcylation.

4.8.1 Global nuclear O-GlcNAc perturbation with the *NLS-BtGH84* cell line

To study the impact of global nuclear O-GlcNAc perturbation, we have engineered a novel mES cell line where *BtGH84* sequence was flanked by the Nuclear Localization Sequence (NLS) (Fig. 16A): upon addition of Doxycycline (Dox), *BtGH84* is expressed and targeted selectively to the cell nucleus. An analogue cell line with a $BtGH84^{D242A}$ mutant, unable to bind or hydrolyze O-GlcNAc (Dennis et al., 2006), was designed as negative control. The nuclear localization of BtGH84 was verified by immunofluorescence (IF) (Fig. 16B). The BtGH84 activity was tested by O-GlcNAc WB: BtGH84 expression induces a global de-GlcNAcylation of nuclear proteins with the strongest effect at 48h of *BtGH84* expression (Fig. 16C). A small O-GlcNAc depletion was detected upon $BtGH84^{D242}$ expression after 12 hours onwards, showing that the point mutation has some residual enzymatic activity. Despite the strong global de-GlcNAcylation induced by *BtGH84* expression, nuclear O-GlcNAc levels were still detectable even after 48h of Dox treatment. These results might be due to the endogenous OGT activity, which can probably partially compensate BtGH84's removal of O-GlcNAc. Another reason might be that BtGH84 protein levels might not be enough to obtain a 100% nuclear O-GlcNAc depletion. Despite the relative efficiency of this system, I have decided to exploit the *NLS-BtGH84* cell line system to start investigating the effects of nuclear O-GlcNAc perturbation on gene expression.

4.8.1.1 Investigation of O-GlcNAc role in transcription by *NLS-BtGH84* cell line

To investigate the role of O-GlcNAcylation in controlling gene expression, I have differentiated the *NLS-BtGH84/NLS-*$BtGH84^{242A}$ mESCs cell lines into NPCs and performed RNA-seq on the two cellular states before and after Dox treatment. After 48h of *BtGH84* expression in ESCs, 258 genes were upregulated and 341 downregulated (with a logFC value >0), while after NPC differentiation in presence of Dox, 1238 were upregulated and 1491 downregulated (Fig. 16D). Unexpectedly, $BtGH84^{242A}$ expression also induced a perturbation of gene expression, with 1928 upregulated and 1791 downregulated genes in ECSs, and with 2336 upregulated and 2381 downregulated

genes in NPCs. The strong transcriptional effects of BtGH84^{242A} could be explained by a partial retained activity of the inactive BtGH84^{242A} enzyme, as seen in Fig. 16C. It is possible that BtGH84^{242A} residual activity could cause a milder O-GlcNAc perturbation and delayed response in the cellular adaptation. On the contrary, the lower number of deregulated genes after 48h of active BtGH84 expression suggests that the cellular response to the O-GlcNAc perturbation might be much faster, thus masking the rapid and subtle effects on transcription. However, the data does not enable to conclude on the unpredicted effect of BtGH84^{242A}. Because my RNA-seq results could not be fully interpreted, I have decided to investigate whether O-GlcNAc perturbation could have an effect on chromatin accessibility by performing ATAC-seq analysis (4.8.1.2).

4.8.1.2 Investigation of O-GlcNAc role in chromatin accessibility by *NLS-BtGH84* cell line

To test the effects of nuclear O-GlcNAc depletion on chromatin accessibility, I have performed the Assay for Transposase-Accessible Chromatin with sequencing (ATAC-seq) in ESCs and NPCs before and after Dox treatment. As for RNA-seq, in the ATAC-seq analysis I have compared the differentially accessible regions between active *NLS-BtGH84* and the inactive mutant *NLS-BtGH84^{242A}* cell lines. In ESCs, O-GlcNAc nuclear perturbation showed no significant difference in chromatin accessible regions between active and inactive BtGH84^{242A}. Instead, *BtGH84* expression in NPCs resulted in mild effects on chromatin compaction (Fig. 16E). Detailed inspection of the affected chromatin loci did not highlight any gene cluster enrichment. Moreover, several of these chromatin loci resulted to encode for mitochondrial chromosome; it is well known that a major drawback of ATAC-seq technique is the contamination sequencing reads derived from mitochondrial DNA (Montefiori et al., 2017). Therefore, the majority of the differential chromatin accessible regions from treated NPCs could not be considered for further analysis.

All together, these results shows that a reduction of nuclear O-GlcNAc levels does not have a major role in modulating chromatin accessibility at pluripotency stage. Moderate changes in chromatin compaction were identified only in NPCs, however no specific gene function was identified.

Collectively, the data show that the expression of BtGH84 in *NLS-BtGH84* cell line rapidly perturbs the global nuclear O-GlcNAcylation levels: the *NLS-BtGH84* cell line represents the first genetic tool for the inducible perturbation of nuclear O-GlcNAcylation in mESCs which 1) overcomes the limitations of the endogenous *Ogt* knock-out and 2) allows to investigate the nuclear O-GlcNAcylation function independently from OGT function. Nonetheless, my experiments have highlighted two major limitations of the NLS-BtGH84 system: 1) the global nuclear O-GlcNAc levels were only partially affected by BtGH84 expression and 2) inactive BtGH84^{242A} has demonstrated to not be the expected negative control for my transcriptomics analysis.

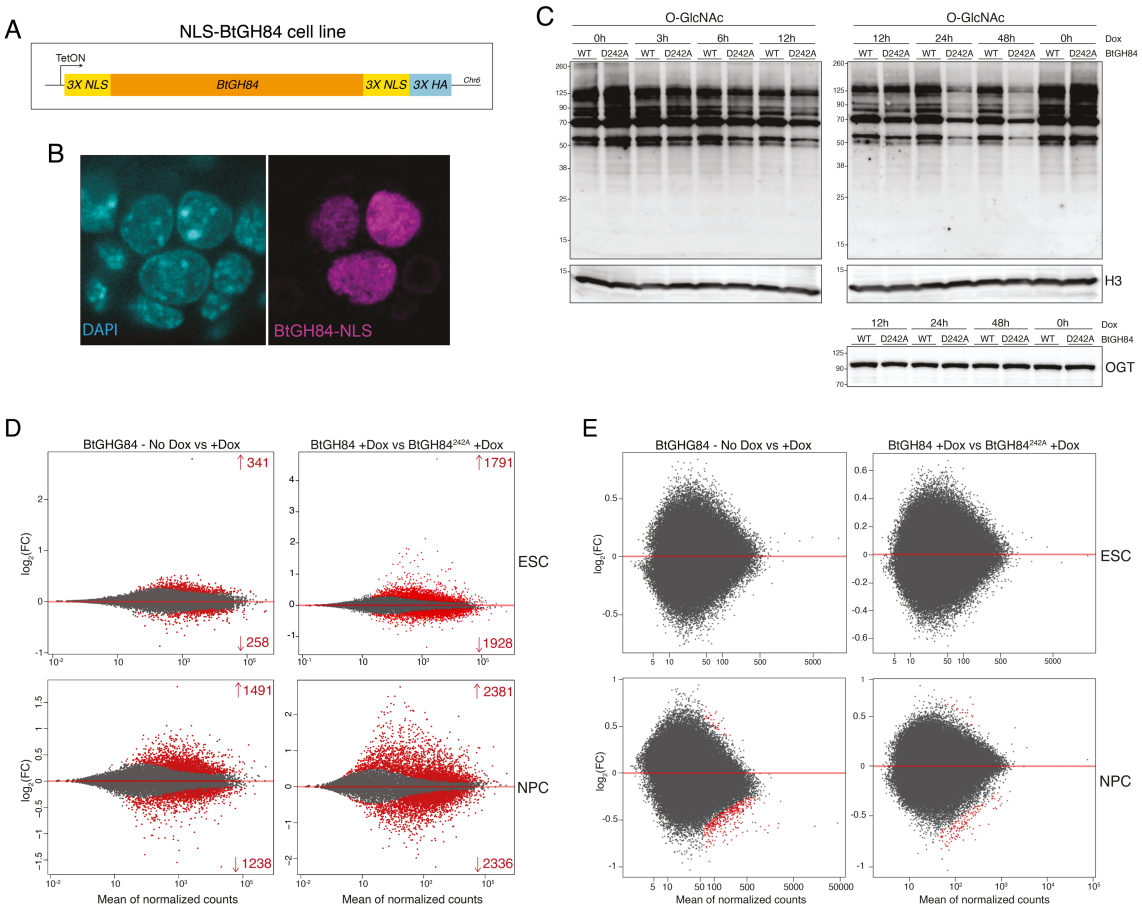

Fig. 16 (A) Representation of the transgenic cell line expressing bacterial OGA BtGH84 fused to a localization peptide (NLS) and regulated by Tet-On system. **(B)** Immunofluorescence (IF) analysis of nuclear BtGH84 targeting in mESCs. DNA is visualized with DAPI, BtGH84 is visualized with HA antibody (abcam 9110). Experiment performed by Yulia Ermakova. **(C)** WB analysis of global O-GlcNAc depletion in ESCs nuclei at 0h, 3h, 6h, 12h, 24h, 48h after BtGH84 induction. WGA-HRP was here used to detect GlcNAc levels. OGT levels were verified with OGT ab177941 antibody, while H3 antibody was used as loading control. Experiment performed by Yulia Ermakova. **(D)** MA plot of RNA-seq analysis of NLS-BtGH84 and NLS-BtGH84^{242A} cell lines before and after Dox treatment in ESCs and NPCs. Threshold of log_2(FC)>0. **(E)** MA plot of ATAC-seq analysis of NLS-BtGH84 cell line before and after Dox treatment in ESCs and NPCs. Bioinformatic analysis performed by Nicolas Descostes.

4.8.2 Investigation of targeted O-GlcNAc removal at proteins bound to CpG-rich promoters

Chromatin profiling of O-GlcNAc modification revealed a significant enrichment of O-GlcNAc proteins at gene promoters and I have found that OGT interacts with a number of CXXC containing proteins that bind to CpG-dense promoters. In vertebrate genomes, approximately 70% of the annotated gene promoters are associated with CpG islands, which are CpG-dense loci that are kept unmethylated (Deaton and Bird, 2011). To target O-GlcNAc at CpG-dense promoters, I generated in collaboration with Dr. Yulia Ermakova, an inducible cell line whereby *BtGH84* sequence was fused to the CXXC domain of KDM2A (Xu et al., 2018). The CXXC domain is a zinc-finger domain of ~50-70 amino acids which binds to unmethylated CpG dinucleotides (Long et al., 2013) which form clusters at promoters (Deaton and Bird, 2011). The unmethylated state of CpG-dense promoter is conserved in all cell types regardless of the transcriptional activity (Bestor et al., 2015). Upon induction with Doxycycline, *BtGH84* is expressed and targeted specifically to unmethylated CpG promoters by the CXXC domain (Fig. 17A). CXXC-BtGH84 protein expression was validated by WB analysis (Fig. 17B), while CXXC-BtGH84 localization at CpG sites was tested by ChIP-seq using an antibody against the HA epitope that is fused to BtGH84 (Fig. 17C). Because inactive BtGH84^{242A} demonstrated to not be a suitable control for transcriptomic analysis, I have decided to focus only on active BtGH84 and compare the untreated and treated conditions (expression of the effector vs no expression in the same cells).

CXXC-BtGH84 ECSs were differentiated into NPCs and transcriptomic analysis by RNA-seq were performed in the two cell states in presence and absence of Doxycycline. *CXXC-BtGH84* expression in ESCs resulted in a total of 140 differentially expressed genes (DEGs), with the upregulation of 19 genes and downregulation of 121 genes (Fig. 17D). Gene network analysis of the differentially expressed genes did not reveal any cluster enrichment. In NPCs instead, a larger group of DEGs was identified (1093 genes), with 594 upregulated and 499 downregulated genes (Fig. 17D). Among the upregulated NPC genes, 51 DEGs classified into the small and large ribosomal proteins group (*Rps/Rpl*). These data demonstrate that O-GlcNAc depletion at CpG-dense promoters has an effect on ribosomal protein expression, which is detectable only after 5 days of NPC differentiation. Gene pathway analysis of NPC DEGs identified three major protein networks: a large cluster characterized by genes encoding for rRNA binding proteins, a second cluster of genes encoding for chromatin organization and histone modification proteins, and a third cluster of genes encoding for mitochondrial

metabolic proteins. These data are remarkably consistent with the proteomic results I obtained from the characterization of the OGT interactome, uncovering the main cellular pathways regulated by OGT and O-GlcNAc modification.

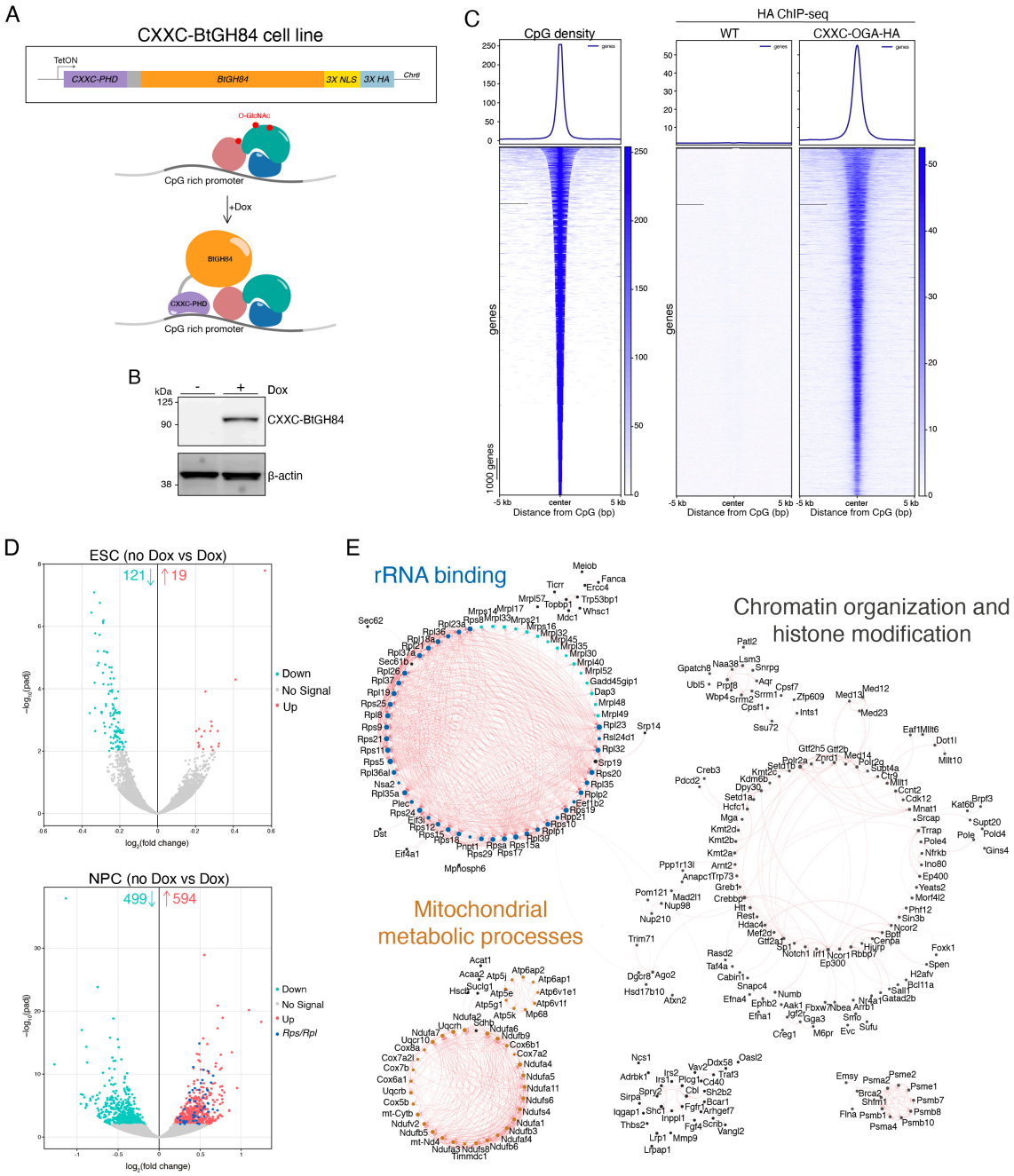

Fig. 17 (A) Representation of the transgenic cell line expressing bacterial OGA BtGH84 fused to CXXC domain and regulated by Tet-On system. **(B)** Western Blot analysis of CXXC-BtGHG84 expression. BtGH84 was detected with Mgea5 (OGA) PA5-67426 antibody, β-actin was used as loading control. Experiment performed by Yulia Ermakova. **(C)** Heatmap showing genome-wide CpG density, O-GlcNAc distribution at CpG sites (from CUT&RUN experiment in presence of 0.1% Triton) and CXXC-BtGH84 profiling by HA ChIP-seq. **(D)** Volcano plot of RNA-seq analysis of CXXC-BtGH84 cell line before and after Dox treatment in ESCs and NPCs. The threshold of pAdj is <0.01. **(E)** STRING clustering analysis of NPC DEGs from RNA-seq. Functional interactions based on STRING database are shown with red lines between proteins. The darker red lines depict interactions between clustered proteins uncovered by Markov clustering (MCL).

4.9 Investigation of O-GlcNAc role in rRNA production

The proteomic analysis of OGT's native complexes together with the transcriptomic analysis by targeted removal of O-GlcNAc at CpG-dense promoters highlighted a novel function for O-GlcNAcylation in ribosomal biogenesis regulation. Two findings of mine support this hypothesis: firstly, OGT was unexpectedly found to interact with a large number of nucleolar and ribosomal proteins, and O-GlcNAc perturbation at CpG-rich promoters resulted in an upregulation of both small and large ribosomal proteins genes. Next, I investigated whether O-GlcNAc could regulate rRNA transcription. To this end, because rDNA is CpG dense, I used the CXXC-BtGH84 strategy. Inspection of O-GlcNAc signal from CUT&RUN experiment at rDNA sequence revealed an enrichment of O-GlcNAc proteins at the 5'ETS, 18S, 5.8S and 28S rDNA sequences. These results indicate that O-GlcNAc could be directly involved in regulating rRNA expression (Fig. 18A).

Conventional RNA-seq library preparation is based on the enrichment of RNAs with poly(A) tails, such as messenger RNAs (mRNAs) and long non-coding RNAs (lncRNAs). This process leads to ribosomal RNA (rRNA) depletion, which constitute almost 80% of the total RNA. Therefore, rRNA expression could not be investigated from my CXXC-BtGH84 RNA-seq experiment. To study the effects on rRNA expression after O-GlcNAc perturbation at CpG sites, I performed a qPCR for pre-rRNA and processed rRNA using the total RNA collected from treated and untreated *CXXC-BtGH84* cells. The qPCR was implemented using primers for the 5'ETS, 47S, ITS1 pre-rRNA regions and for the 18S and 28S processed rRNA regions.

Expression of pre-rRNA in ESCs was significantly reduced of ~35% after O-GlcNAc depletion at CpGs, while processed rRNA showed an effect only for the 18S subunit (Fig. 18B) probably due to the short time point of O-GlcNAc perturbation and to the already high presence of processed rRNA in the cell. In NPCs instead, no significant effect on rRNA expression was detected on pre-rRNA and processed rRNA after CXXC-BtGH84 induction.

All together, these results demonstrate that O-GlcNAc perturbation at CpG-rich promoters triggers a rapid effect on rDNA transcription which is detectable in ESCs at the level of pre-rRNA, while no effects are detected at later stages in NPCs owning the huge pool of rRNA present in the cell. The effects on ribosomal protein expression after O-GlcNAc perturbation are instead only secondary and are detected in NPCs (Fig. 17D). These data show that O-GlcNAc regulates rRNA production, which levels drastically

decrease upon impairment of the CpG O-GlcNAc state. I speculate that; the upregulation of ribosomal protein expression in NPCs could be the results of a cellular compensation after the rRNA levels get impaired as a consequence of O-GlcNAc depletion.

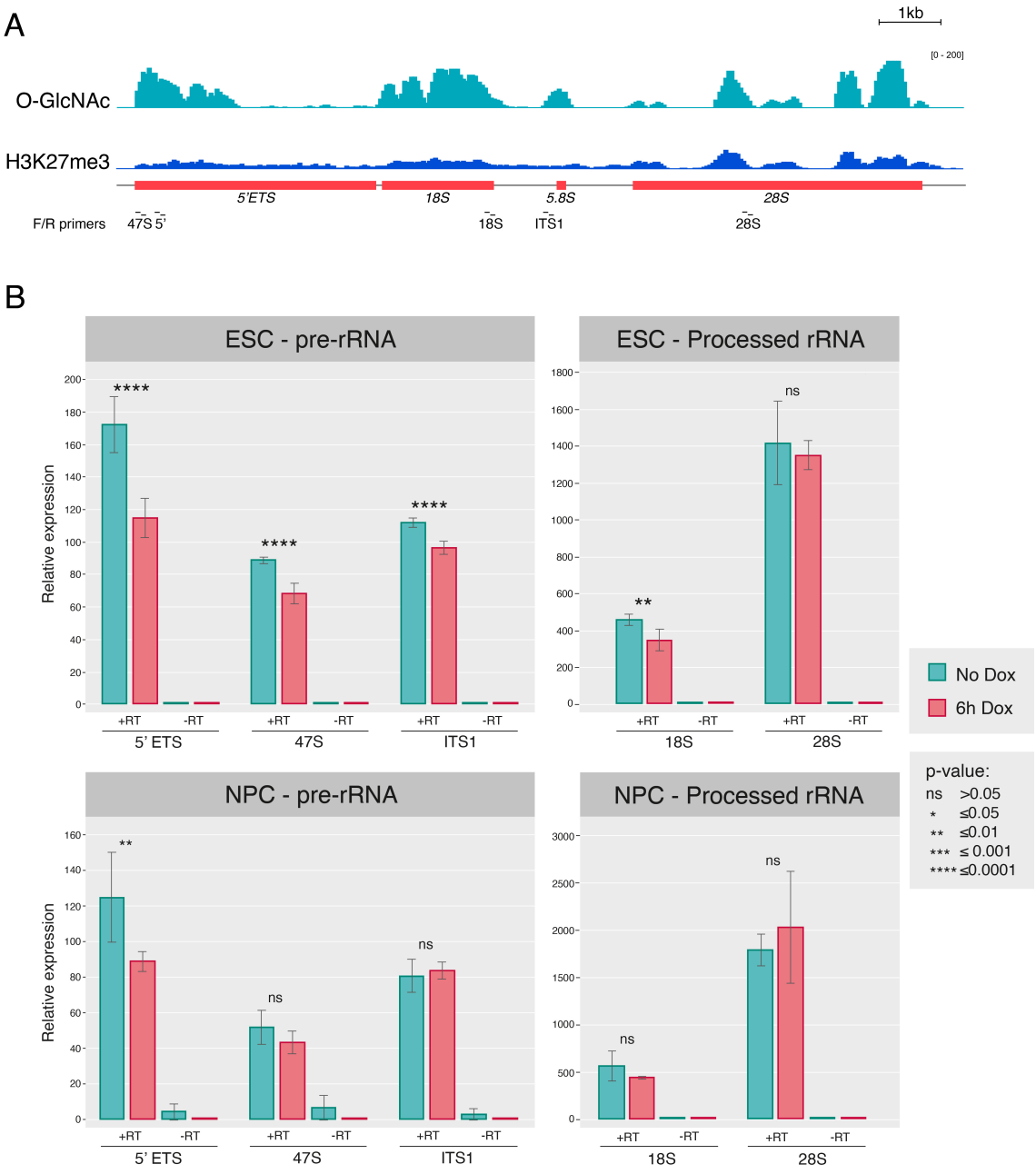

Fig. 18 (A) Genome browser view of CUT&RUN signal for O-GlcNAc at rDNA encoding for 5'ETS, 18S, 5.8S and 28S rRNA. H3K27me3 was used as positive control. Forward and reverse primers used for rRNA qPCR are indicated at the bottom of the reference genome track. **(B)** qPCR analysis of pre-rRNA (5'ETS, 47S and ITS1 regions) and processed rRNA (18S and 28S regions) in *CXXC-BtGH84* ESC and NPC systems. Statistical analysis of rRNA expression was performed using a t-test.

4.10 Characterization of O-GlcNAc role on RNA Pol II

My chromatin profiling experiments revealed a predominant O-GlcNAc enrichment at gene promoters which highly overlaps with RNA Pol II chromatin occupancy (Fig. 13E). These results support previous works which demonstrated that RNA Pol II is modified by O-GlcNAcylation on its C-terminal domain (CTD) (F. I. Comer, 2001; Hart, 1989) (1.8). Because the specific mechanism of RNA Pol II O-GlcNAc still remains unresolved, these data prompted me to investigate whether RNA Pol II CTD O-GlcNAc might have a role in regulating the transcriptional progression in combination with phosphorylation. Therefore, I created novel genetic and detection tools to study the function of O-GlcNAcylation on the RNA Polymerase II,.

4.10.1 Investigation of RNA Pol II O-GlcNAc levels

To investigate the O-GlcNAc state of RNA Pol II, I have created a new Tet-On inducible cell line where RNA Pol II *Polr2a* gene was endogenously tagged at the C-terminus of the two *Polr2a* alleles by the *HA* and *SPOT* epitope tags bridged by a T2A cleavage site, in *NLS-BtGH84* background (Fig. 19A). The C-term position of the tag sequence is aimed to not interfere with the functional activity of the protein. Because the binding of commercially available RNA Pol II antibodies is likely influenced by the CTD's post-translational modifications, the tagging of endogenous *Polr2a* C-term allows to unbiasedly detect RNA Pol II protein, independently of its PTMs.

The correct targeting of RNA Pol II tag sequence was verified by Sanger sequencing (data not shown), while the presence of the *knock-in* tags was tested by WB analysis (Fig. 19B). Moreover, the nuclear localization of the HA signal was validated in ESCs by immunofluorescence (IF) analysis (Fig. 19C). To analyse the efficiency of BtGH84-induced de-GlcNAcylation on RNA Pol II, I have implemented an RNA Pol II immunoprecipitation protocol (IP) using SPOT-tagged magnetic beads from denaturated nuclear protein extract of BtGH84-RNA Pol II$^{HA/SPOT}$ cell line at different time points of *BtGH84* induction. RNA Pol II -IP efficiency was validated by RNA Pol II WB using input as control, while the efficiency of O-GlcNAc depletion was verified by O-GlcNAc WB (Fig. 19D). To demonstrate the specificity of the IP, RNA Pol II was purified in parallel from *NLS-BtGH84* background cell line: the experiment showed no enrichment of RNA Pol II protein in the negative control fraction (Fig. 19D). The quantification of WB O-GlcNAc signal demonstrated that BtGH84 treatment induces a progressive depletion of O-

GlcNAc levels on RNA Pol II, with a strong effect already at 24h after *BtGH84* expression. Intriguingly, comparison of total RNA Pol II protein levels with RNA Pol II O-GlcNAc levels suggests that O-GlcNAc RNA Pol II represents only a small fraction of the total amount of nuclear RNA Pol II.

Overall, my results validated that RNA Pol II is a glycoprotein and demonstrated that *NLS-BtGH84* cell line represents a valuable tool to investigate the effects of O-GlcNAc perturbation on RNA Polymerase II.

4.10.2 RNA Pol II localization at nuclear transcription factories after nuclear O-GlcNAc perturbation

Transcription of active genes occurs within specific spatially organized chromatin sites known at transcription factories, whereby protein clustering ensures the local concentration of molecules which drive the entire transcriptional process. RNA Pol II molecules condensate at these transcription factories together with the transcriptional machinery factors, forming DNA and protein condensates that are dispensable to ensure a successful transcriptional output (Ghamari et al., 2013).

To investigate how O-GlcNAcylation regulates RNA Pol II I activity, I have tested whether CTD O-GlcNAc regulates RNA Pol II spatial localization and transcriptional foci formation. To address this question, I have tracked RNA Pol II clusters at transcription factories in ESCs and NPCs and compared RNA Pol II localization at steady-state and after O-GlcNAc perturbation. Transcription factories imagining was performed with the stimulated emission depletion (STED) microscopy technique: among the super-resolution microscopy (SRM) approaches, STED microscopy allows to circumvent Abbe's optical resolution limit, providing spatial resolution below the limit imposed by the diffraction of light. With STED microscopy, imaging can be performed at a resolution down to 20 nm, making it suitable for single molecule localization analysis.

To assess whether RNA Pol II molecules clustering is regulated by O-GlcNAc modification, I have quantified the RNA Pol II transcription factories by SPOT tag and calculated the distribution of intensity of RNA Pol II signal, which represents the average of RNA Pol II foci signal intensity. In ESCs, RNA Pol II distribution of intensity was not affected by *BtGH84* expression. In NPCs instead, RNA Pol II localization at transcription factories showed a significantly higher distribution of intensity after O-GlcNAc perturbation (Fig. 19E). These results indicate that O-GlcNAc modification has a direct role in regulating RNA Pol II spatial localization. This effect is specific for neuronal

progenitor cells (NPCs), where RNA Pol II molecules cluster in denser foci when O-GlcNAc is perturbed.

Of note, it was previously shown that some protein kinases regulate CTD hyperphosphorylation and phase separation to promote transcriptional elongation (Lu et al., 2018). Moreover, recent results demonstrated that O-GlcNAc on SynGAP T1306 protein drives liquid-liquid phase separation and the interaction between SynGAP and PSD-95 neuronal protein (Lv et al., 2022). It might be speculated a similar role for O-GlcNAc on RNA Pol II where, together with phosphorylation, it might regulate RNA Pol II condensates formation at transcribed genes possibly by phase-phase separation.

Because the *NLS-BtGH84* system is not specific for RNA Pol II, I will validate these results using a novel *gp41-BtGH84-NLS^{SPOT}/RNA Pol II$^{HA/gp41}$* cell line where O-GlcNAc will be perturbed selectively on RNA Pol II (4.10.3). The future experiments with this novel tool will allow to assess the specific effects of O-GlcNAc perturbation on the RNA Pol II CTD.

4.10.3 Generation of RNA Pol II- Moon tag cell line

The *NLS-BtGH84/RNA Pol II$^{HA/SPOT}$* system allowed me to 1) validate that RNA Pol II is modified by O-GlcNAc; 2) demonstrate that nuclear O-GlcNAc perturbation by BtGH84 affects RNA Pol II O-GlcNAc state; 3) report that O-GlcNAc depletion affects RNA Pol II clustering at NPCs nuclear transcription factories.

To improve the precision of O-GlcNAc perturbation on RNA Pol II specifically, I have decided to generate a novel inducible cell line system where BtGH84 will be inducibly targeted specifically on the CTD of RNA Pol II by a recognition tag and a nanobody (Fig. 19F). Using the Tet-On exchange cassette system, *BtGH8*4 sequences was fused with *gp41* nanobody, a single-domain llama antibody consisting of 123 amino acids that was shown to in vitro bind its gp41 15 amino acids peptide with an affinity of ~30 nM (Lutje Hulsik et al., 2013). In parallel, the endogenous *Polr2a* gene was tagged by the GEEF Rome Facility with the *gp41* tag, also known as MoonTag, at the extremity of its C-terminal domain. The MoonTag strategy was demonstrated to be an efficient system for protein localization by fluorescence labelling (Boersma et al., 2019) due to the high affinity between the Moon nanobody and its recognition peptide.

The *gp41-NLS-BtGH84^{SPOT}* construct was generated in collaboration with a Master student Jeanne Couturier I supervised. The *knock-in* sequence was validated by

genotyping and Sanger sequencing (data not shown). The same approach was used to generate a negative control cell line, where the *gp41* nanobody sequence was fused with the *enhanced green fluorescent protein (EGFP)* sequence as non-catalytic effector. After isolation of a homozygous *RNA Pol II*$^{HA/gp41}$ clone by the EMBL Rome GEEF Facility, I will transfect the cells in parallel with both *gp41-NLS-BtGH84*^{SPOT} and *gp41-NLS-EGFP*^{SPOT} constructs. After validation of the knock-in cell line, I will induce the targeted RNA Pol II CTD de-GlcNAcylation and investigate the effects on transcription in ESCs and NPCs using different strategies:

1) RNA Polymerase II foci formation will be analysed by super resolution microscopy. With this approach, I will assess the direct involvement of O-GlcNAcylation in RNA Pol II nuclear localization and in the overall transcriptional foci formation.
2) RNA Pol II will be profiled on chromatin by CUT&RUN-seq. The experiment will be performed using the HA antibody to unbiasedly target RNA Pol II. Travelling ratio will be calculated as the ratio between RNA Pol II occupancy at gene body and promoter regions (Rahl et al., 2010). This parameter will be informative to understand whether O-GlcNAc regulates a specific phase of RNA processing, as promoter recognition, pausing or elongation.
3) Transcription of coding genes is arguably the most important RNA Pol II-dependent process in eukaryotic cells. Whether O-GlcNAc is dispensable to ensure the successful formation of mRNA molecules it is still unknown. I will perform transcription analysis by nuclear mRNA-seq to assess the role of RNA Pol II O-GlcNAc in transcriptional regulation.

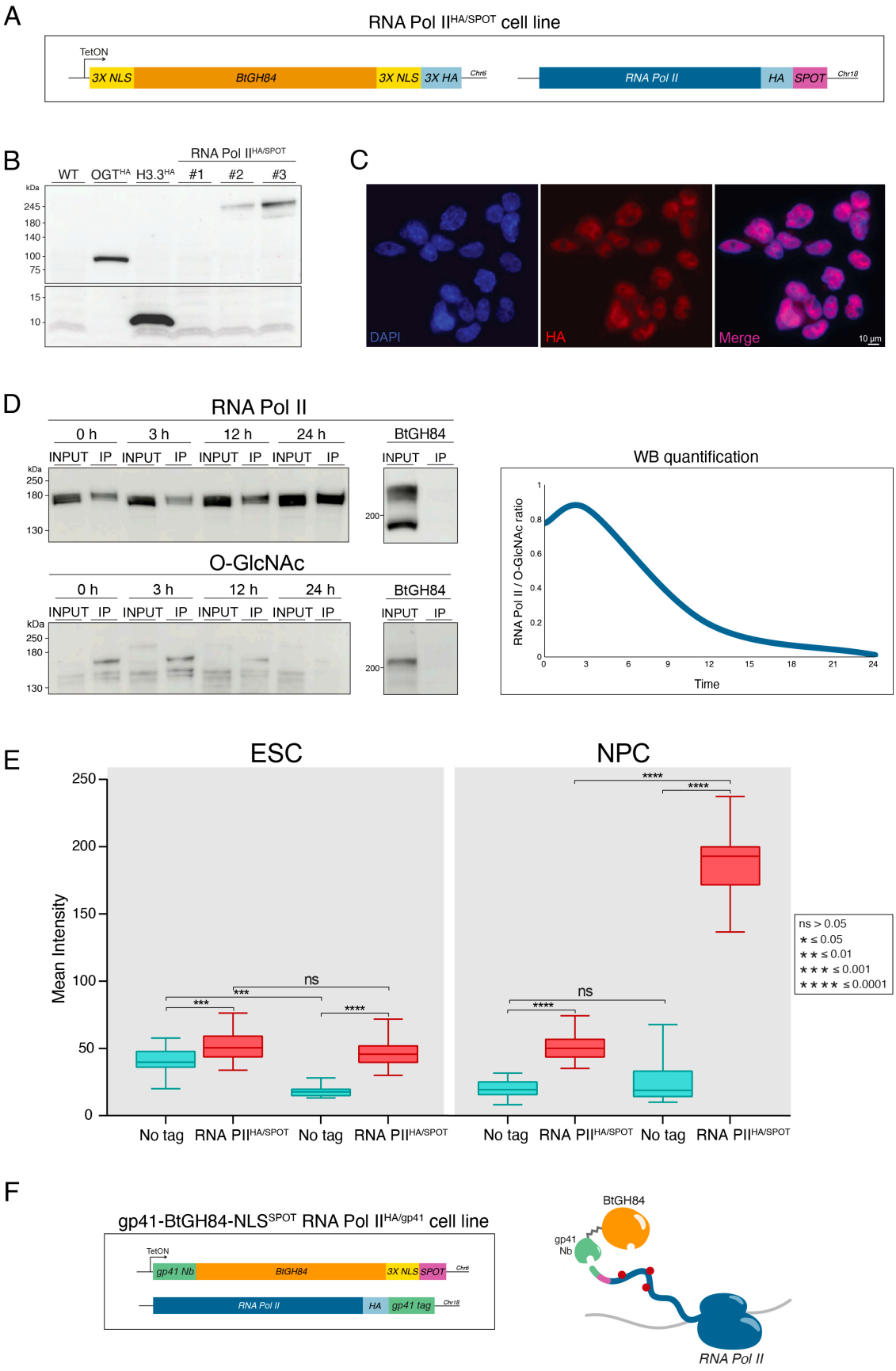

Fig. 19 (A) Representation of the transgenic cell line expressing bacterial OGA BtGH84 fused to a localization peptide (NLS) and regulated by Tet-On system. The same cell line has a homozygous tag for

the endogenous RNA Pol II gene. **(B)** RNA Pol II homozygous tagging of clone #3 was verified by HA WB (abcam 9110). **(C)** Immunofluorescence (IF) analysis of RNA Pol II-HA in ESCs. HA Biolegend 901515 HA antibody was used to visualize RNA Pol II. DNA is visualized with DAPI. Experiment performed by Rahul Sureka. **(D)** Immunoprecipitation (IP) of RNA Pol II with SPOT-tag magnetic beads at different time points of BtGH84 expression. Efficiency of RNA Pol II IP was verified by RNA Pol II WB, RNA Pol II O-GlcNAc levels were verified by O-GlcNAc WB. WB signal was quantified by RNA Pol II/O-GlcNAc ratio. INPUT samples represent the total nuclear protein fraction, IP samples represent the nuclear protein fraction after incubation with SPOT-beads. **(E)** Distribution of intensity quantification of RNA Pol II transcriptional factories before and after *BtGH84* induction in ESCs and NPCs. Statistical analysis was performed with a t-test. **(F)** Representation of *gp41-BtGH84-NLS^{SPOT}/RNA Pol II$^{HA/gp41}$* cell line and of gp41-BtGH84 targeting on RNA Pol II CTD.

4.10.4 Generation of RNA Pol II CTD O-GlcNAc T4 mouse monoclonal antibody

Independent studies have shown that the carboxyl terminal domain of RNA polymerase II is modified by O-GlcNAc (F. I. Comer, 2001; Kelly et al., 1993; Lewis et al., 2016); however, the CTD O-GlcNAc code remains completely enigmatic. One of the major drawbacks of O-GlcNAc RNA Pol II studies is the absence of antibodies directed to specific O-GlcNAc residues, essential for the direct detection and profiling of the CTD O-GlcNAc state. GlcNAcylation of single CTD amino acids was reported in previous studies, however, due to the repetitive nature of the CTD sequence that makes the mapping extremely difficult, these studies are often discordant. Interestingly, Threonine 4 was the unique residue consistently validated to be O-GlcNAcylated both *in vitro* and *in vivo* (F. I. Comer, 2001; Kelly et al., 1993).

To explore the chromatin occupancy dynamics of glycosylated RNA Pol II, I have collaborated with the Jackson Laboratory (JAX) Protein Sciences Service to generate the first monoclonal antibody against Threonine 4 O-GlcNAcylation on RNA Pol II CTD. The production of the antibody was performed at JAX by immunizing a mice cohort with a di-heptapeptide Y_1-S_2-P_3-T_4(GlcNAc)-S_5-P_6-S_7-Y_1-S_2-P_3-T_4(GlcNAc)-S_5-P_6, which corresponds to the conserved amino acidic sequence repeat on the CTD. The ELISA test performed by JAX highlighted 4 clones (2D3, 2H5, 4B1 and 4C4) which positively reacted with the immunization peptide only when modified by O-GlcNAc on Thr4. I have thus tested the hybridoma supernatants of the positive clones by WB analysis on nuclear protein extract of different cell types (Fig. 20A). All tested hybridomas showed high levels of unspecific signal; however, for 2D3 and 2H5 clones, a band at RNA Pol II size was detected. To validate the specificity of the interested band, I have thereafter tested 2D3 and 2H5 clones using a RNA Pol II degron human DLD-1 colorectal cancer cell line kindly

provided by Argyris Papantonis laboratory (University of Göttingen), where the RPB1 RNA Pol II subunit was tagged with a mini-AID (mAID) domain (Zhang et al., 2021b). By addition of Auxin and Doxycycline, the plant ubiquitin ligase TIR1 gets activated and induces the degradation of >80% of total RNA Pol II in only 14h (Yesbolatova et al., 2019). 2D3, 2H5 and 4B1 clones were tested before and after RNA Pol II degradation, and only 2D3 recognized the RNA Pol II band, which intensity decreases upon RNA Pol II degradation (Fig. 20B). The 2D3 antibody was purified from the hybridoma supernatant by JAX and tested again on protein nuclear extract of DLD-1-RPB1-mAID cells. Purified 2D3 antibody was efficiently validated to recognize RNA Pol II (Fig. 20C): upon RNA Pol II degradation, the intensity of the 2D3 band significantly decreased. To test whether the 2D3 antibody was specific for the O-GlcNAcylated form of RNA Pol II, the nuclear protein extract was treated with NagJ and BtGH84 purified enzymes, two bacterial forms of OGA (Dennis et al., 2006). After enzymatic treatment, no signal was detected by 2D3 antibody, showing that the signal is dependent on O-GlcNAc.

All together, these data show that the 2D3 RNA Pol II T4 O-GlcNAc antibody specifically recognizes the O-GlcNAc form of RNA Pol II, representing a direct evidence of the T4 CTD O-GlcNAcylation state and the first direct tool to detect O-GlcNAc RNA Pol II. O-GlcNAc RNA Pol II chromatin profiling will be necessary to explain the direct function of the glycosylated CTD. However, the 2D3 antibody cross reacts with a high number of proteins, hence it is not an appropriate reagent for genomic profiling studies. O-GlcNAc Thr4 antibody will instead serve as a control to prove the BtGH84-directed RNA Pol II de-GlcNAcylation in the *gp41-BtGH84-NLS^{SPOT}/RNA Pol II$^{HA/gp41}$* system.

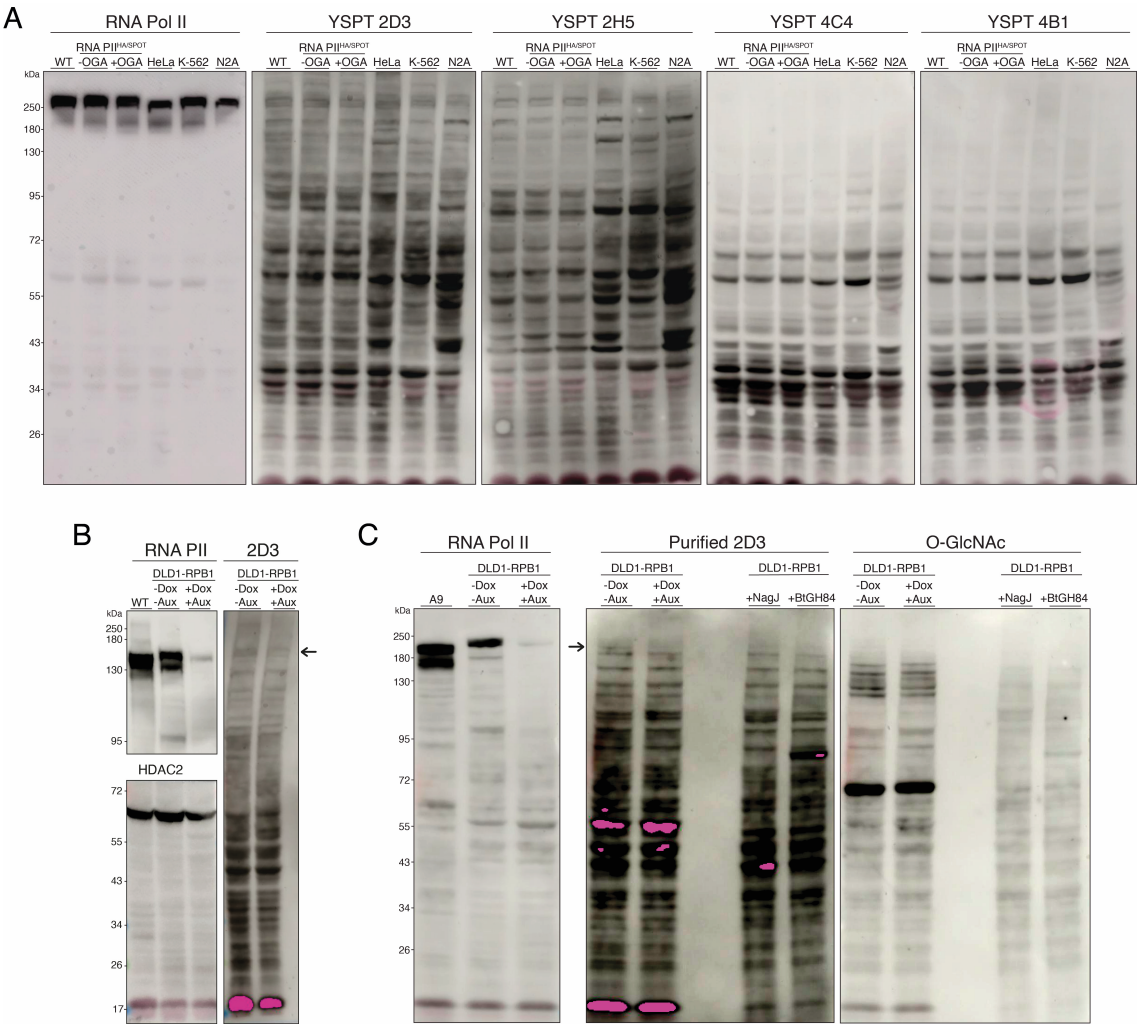

Fig. 20 (A) Western Blot analysis of nuclear protein extract from WT, *gp41-BtGH84-NLS^{SPOT}/RNA Pol II$^{HA/gp41}$*, HeLa, K-562 and N2A cells with CTD Thr4 O-GlcNAc RNA Pol II 2D3, 2H5, 4B1 and 4C4 antibody clones, **(B)** from WT and DLD1-RPB1cell line with 2D3 antibody clone, **(C)** and from DLD1-RPB1cell line with 2D3 antibody clone using NagJ and BtGH84 treatment. RNA Pol II sc-55492, HDAC2 A300-705A-M and O-GlcNAc HGAC85 antibodies were used as controls.

4.11 Identification of O-GlcNAc nuclear proteins

While genomic experiments (ChIP-seq and CUT&RUN) enabled me to describe the distribution of O-GlcNAc proteins on chromatin, biochemical tools can be used to characterize the specific identity of GlcNAc-modified proteins. Electron Transfer Dissociation (ETD) Mass Spectrometry is a MS/MS method that allows to identify the interested proteins and to characterize the post-translational modifications (PTMs) on each modified amino acid residue. This is achieved because, in ETD Mass Spectrometry, an electron reacts with the positively charged peptides causing the peptide backbone cleavage into c- and z-ions. Contrary to the standard MS/MS peptide cleavage (e.g. collision induced dissociation), this process preserves the chemical groups during peptide fragmentation, representing an optimal tool to detect labile PTMs such as O-GlcNAcylation.

To characterize the O-GlcNAc proteome, I have implemented a biochemical protocol to selectively purify O-GlcNAc proteins from mESCs nuclei. The O-GlcNAc enrichment was achieved using wheat germ agglutinin (WGA), a lectin derived from Triticum vulgaris that binds to GlcNAc. WGA is particularly useful for the O-GlcNAc studies because it is characterized by multiple O-GlcNAc binding sites (Wright, 1987; Wright, 1992) which increase the affinity for O-GlcNAc proteins. Moreover, WGA can be conjugated with HRP, functioning as a probe for O-GlcNAc protein detection by Western Blot.

WGA lectin was conjugated with biotin and WGA-GlcNAc protein complexes were purified by streptavidin magnetic beads (Fig. 21A). The efficiency of O-GlcNAc protein purification was assessed by O-GlcNAc WB on the purified extract, while the specificity of the WGA-GlcNAc binding was verified by WGA saturation with free-GlcNAc and by BtGH84 enzymatic treatment (Fig. 21B). Among the purified O-GlcNAc proteins from a denaturated nuclear extract, RNA Pol II was also identified.

The WGA-IP was analyzed by ETD MS/MS at EMBL Proteomic Core Facility and a dataset of nuclear O-GlcNAc proteins was generated. The purity of the nuclear protein extract was inspected by clustering the identified proteins based on their subcellular localization; as expected, most of the purified proteins are nuclear proteins, although some cytoplasmic proteins were detectable (Fig. 21C). Nuclear O-GlcNAcylated proteins were classified by fold enrichment using Fisher's extract and Panther GO-Slim biological process analysis was performed (Mi et al., 2019) (Fig. 21D). Interestingly, the most abundant clusters of purified O-GlcNAc proteins are involved in the regulation of RNA Polymerases I and III transcription. As previously mentioned in paragraph 1.1.2, RNA

Pol I and III regulates rRNA transcription; these results indicate that the most abundant O-GlcNAc nuclear proteins are involved in the regulation of the RNA Pol I and III -directed rRNA transcription, supporting my previous results on rDNA transcription.
Collectively, my results show that WGA-IP represents a valuable tool to purify O-GlcNAc proteins from fractionated nuclei. However, it was previously shown that WGA binding can be biased towards proteins modified by multiple O-GlcNAc sites (Gallagher et al., 1985). Therefore, I have decided to optimize the O-GlcNAc immunoprecipitation and test novel O-GlcNAc enrichment strategies by collaborating with the Myers Team at La Jolla Institute, San Diego: the first approach is based on the O-GlcNAc peptide IP by anti-O-GlcNAc monoclonal antibodies (Burt et al., 2021), while in the second approach the enzymatically locked version of BtGH84 and NagJ enzymes (purified by the EMBL PEPCore Facility) are used for the O-GlcNAc site recognition. Nuclear proteins purified from ESCs and NPCs will be enriched for O-GlcNAcylated proteins and analyzed by ETD MS/MS to obtain a O-GlcNAc proteome dataset.
Quantitative mass spectrometry analysis of ESC and NPC purified nuclear proteins revealed global changes in the total proteome between the two cell types, as expected (Fig. 21F).
Preliminary results obtained from quantitative Mass Spectrometry analysis of O-GlcNAc immunoprecipitation experiment identified 15 and 8 peptides, in ESCs and NPCs respectively, with differences in their O-GlcNAc state between the two cell types. The fold change of O-GlcNAc peptides were normalized over the fold change of their protein reference. Further optimizations of the O-GlcNAc peptide precipitation approach will be implemented, as well as ETD mass spectrometry analysis of the ESC and NPC proteome. Nonetheless, these data provide preliminary evidence of protein candidates with O-GlcNAc level variations during ESCs differentiation to NPCs. These factors represent potential candidates for future studies to untangle possible mechanisms of O-GlcNAc regulation during cellular differentiation.

One of the most intriguing unanswered questions is to understand how OGT can modify such a large group of proteins. One of the hypotheses is that OGT might recognize its substrates through a specific recognition peptide motif at their Threonines and Serines modified sites. To address this question, my aim is to explore the O-GlcNAc proteome and investigate whether O-GlcNAc proteins share specific amino acidic sequences and/or biochemical characteristics. To implement this analysis, I am collaborating with the Petsalaki Group at EMBL EBI to unbiasedly study the sequence composition of all

known O-GlcNAcylation sites. Because the characterization of the O-GlcNAc proteome by ETD MS/MS is still ongoing, we have decided to explore the available O-GlcNAc proteome datasets and query for O-GlcNAc motif enrichment. Analysis of amino acidic motifs enrichment of O-GlcNAc proteins from the PhosphoSitePlus database (Hornbeck et al., 2014) identified a high frequency of Threonine (Thr), Valine (Val) and Proline (Pro) residues between the positions +/-2, -3 from the modified O-GlcNAc site (Fig**. 21**E). Interestingly, the high frequency of Proline between positions +/- 2, -3 from the O-GlcNAc site very well resembles the amino acid sequence of the RNA Pol II CTD.

This analysis indicates that OGT modifies preferentially proteins that are rich in Thr, Val and Pro at the boundaries of the Thr and Ser O-GlcNAc modified residue. These results give a novel insight into the amino acid sequence characteristics of O-GlcNAcylated proteins.

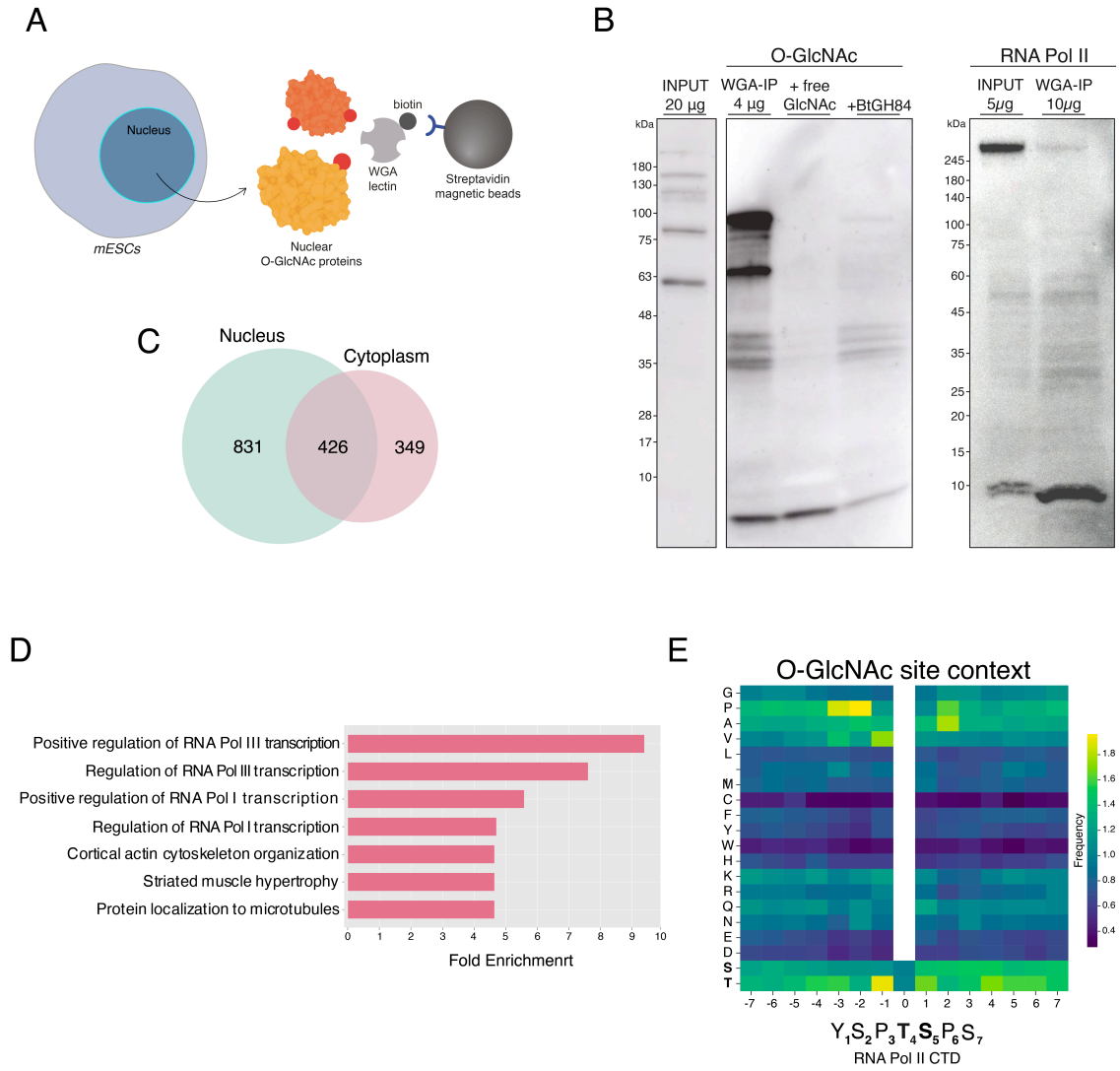
A
Nucleus
mESCs
Nuclear
O-GlcNAc proteins
biotin
WGA
lectin
Streptavidin
magnetic beads
B
O-GlcNAc
INPUT
20 µg
WGA-IP
4 µg
+ free
GlcNAc
+BtGH84
RNA Pol II
INPUT
5µg
WGA-IP
10µg
kDa
C
Nucleus
Cytoplasm
831
426
349
D
Positive regulation of RNA Pol III transcription
Regulation of RNA Pol III transcription
Positive regulation of RNA Pol I transcription
Regulation of RNA Pol I transcription
Cortical actin cytoskeleton organization
Striated muscle hypertrophy
Protein localization to microtubules
Fold Enrichmenrt
E
O-GlcNAc site context
Frequency
RNA Pol II CTD
F
Hit
No hit
ESC
NPC
adj pvlaue=0.05
logFC(NPC/ESC)
G
O-GlcNAc levels difference:
Significant
Non significant
ESC
NPC
adj pvlaue=0.05
logFC(NPC/ESC)

Fig. 21 (A) Schematic of WGA-IP approach for O-GlcNAc proteins purification. **(B)** O-GlcNAc WB analysis of INPUT and O-GlcNAc purified protein samples, also in presence of free-GlcNAc and BtGH84 purified enzyme. Anti-O-GlcNAc RL2 antibody RNA Pol II (SantaCruz sc55492 antibody) were used. WB analysis after WGA-IP demonstrate RNA Pol II to be O-GlcNAcylated. **(C)** Nuclear-cytoplasmic clustering of purified O-GlcNAc proteins. **(D)** Panther GO-Slim molecular function analysis of O-GlcNAc proteins sample. Here, only proteins with fold enrichment >4 are shown. **(E)** Analysis of amino acid motif at O-GlcNAc sites of O-GlcNAc proteins retrieved from PhosphoSitePlus. The values represent the odds ratio of a specific amino acid at a particular position around the modified S/T sites in respect to the S/T background. The analysis was performed by Alessandro Lussana from the Petsalaki Group. **(F)** Volcano plot of quantitative Mass spectrometry analysis of ESC and NPC proteome. **(G)** Volcano plot of O-GlcNAc levels of immunoprecipitated O-GlcNAc peptides in ESCs and NPCs generated from Mass spectrometry analysis.

5. Conclusions

A prerequisite to the understanding of how O-GlcNAcylation regulates mammalian gene expression is to decipher glycosylated chromatin at genomic and proteomic levels. The knowledge I gained with these approaches have enabled me to devise perturbations strategies for functional studies.

Native immunoprecipitation of OGT protein complexes identified several factors to interact with OGT in the nuclear and cytoplasmic fractions. In addition to the previously known OGT interactors, numerous new OGT partners were identified. The OGT-IP highlighted three main OGT interactor networks: 1) on chromatin, OGT interacts with chromatin modifiers, as well as mRNA splicing factors; 2) in the cytoplasm, the majority of OGT interactors are mitochondrial proteins, highlighting a role of OGT in regulating cellular metabolism through the modification of mitochondrial ATPases; 3) unexpectedly, I have also found that nuclear OGT interacts with a large group of nucleolar proteins, among which several small and large ribosomal proteins were identified.
Of note, the OGT interactome analysis do not provide evidence about the O-GlcNAc state of OGT interactors; the identified factors are not necessarily glycosylated, but instead could be OGT protein substrates which activity might be regulated by the interaction with OGT, or vice versa. The comparison between the OGT interactome and the O-GlcNAc proteome is essential to discriminate the function between OGT interactors and O-GlcNAcylated proteins.

To establish the genomic map of O-GlcNAcylated proteins occupancy, I have implemented the first chromatin profiling of O-GlcNAc modification in mouse ESCs and neuronal progenitor cells (NPCs) by optimizing O-GlcNAc ChIP-seq and by implementing CUT&RUN for O-GlcNAc chromatin profiling. My results revealed that GlcNAc proteins are not uniformly distributed across the genome, but instead occupy specific genomic loci: O-GlcNAc proteins are enriched at specific retrotransposons, at imprinting control regions (ICRs) and most predominantly cellular gene promoters.
The genomic loci bound by glycosylated proteins all function as gene regulatory sequences, however their transcriptional regulatory mechanisms are very diverse. ChIP-

seq analysis revealed that O-GlcNAc proteins occupy all classes of long terminal repeats (LTR) of intracisternal A-particle (IAP), the most active type of retrotransposons in the mouse genome. We have demonstrated that O-GlcNAc is essential to ensure transcriptional repression of IAP elements, as local chromatin de-GlcNAcylation at LTRs reactivates their transcriptional activity (Boulard et al., 2020). Notably, O-GlcNAc is not enriched at DNA transposons that are incapable of transcription, further supporting the newly identified function of O-GlcNAc in the silencing of retrotransposons. It remains to be determined whether O-GlcNAc represses other types of retrotransposons such as LINE1 and perhaps proviruses of exogenous retroviruses.

To get insight into the allele-specific O-GlcNAc proteins DNA binding at ICRs, I have performed GlcNAc ChIP-seq on ESCs F1 Cast/CD1 hybrid cells. Our data revealed a surprising enrichment of O-GlcNAc proteins on the unmethylated allele of three detected ICRs. Gene expression analysis upon OGT degradation in MEFs showed no difference in imprinting genes expression levels. I conclude that, contrary to methylated retrotransposons, O-GlcNAc at ICRs appeared to be dispensable for the silencing of the methylated imprinted allele, but is rather associated with the transcribed allele.

O-GlcNAc signal detected by CUT&RUN validated a significant enrichment of O-GlcNAc proteins at gene promoters and enhancer regions. O-GlcNAc strongly correlates with RNA Pol II occupancy, however the position of the peak relative to the TSS is different, suggesting that O-GlcNAc might regulate the transcriptional steps that precede transcription initiation, as for example PIC assembly as previously demonstrated *in vitro* (Lewis et al., 2016).
The O-GlcNAc occupancy at different types of promoters reflects its correlation with transcription: O-GlcNAc tends to occupy transcriptionally active promoters, however a modest group of transcriptionally inactive genes are also occupied by O-GlcNAc proteins. This result is in contrast with fly, whereby GlcNAc-modified proteins are highly enriched at Polycomb response elements (PRE). In mESCs, the GlcNAc signal does not overlap with that of Polycomb target genes. Thus, mammalian PRCs bear low level of GlcNAc, if any.

Additionally, I identified a correlation between O-GlcNAc and H3K4me3: proteomics analysis of the OGT interactome identified OGT to strongly interact with CXXC1, necessary for the establishment of H3K4me3, and with SETD1A, component of a histone

methyltransferase complex essential for the deposition of H3K4me3. Moreover, ChIP-seq Atlas analysis of O-GlcNAc promoters identified a strong overlap between O-GlcNAc and H3K4me3 at gene promoters. These results are in line with previous data which identified OGT to bind CpG-rich promoters of genes with high H3K4me3 that are co-bound by TET1. These data indicate that O-GlcNAc possibly regulates specific proteins, like CXXC1 and SETD1A, that are involved in H3K4me3 establishment at promoter regions. These results might explain the correlation between O-GlcNAc -rich promoters and positive gene expression, as H3K4me3 is typically associated with active transcription.

The strong correlation of O-GlcNAc with RNA Pol II occupancy prompted me to investigate the mechanism of O-GlcNAc CTD in regulating RNA Pol II activity. O-GlcNAcylation of RNA Pol II CTD was described for the first time in 1993 (Kelly et al., 1993). Our O-GlcNAc ChIP-seq data support these previous results: O-GlcNAc enrichment highly overlaps with RNA Pol II occupancy at gene promoters. To investigate the function of RNA Pol II O-GlcNAcylation, we generated an RNA Pol II-tagged knock-in allele in an ESC line where nuclear proteins can be conditionally de-GlcNAcylated. I showed that this cell line enables a strong reduction of O-GlcNAc levels on RNA Polymerase II, making our system a suitable tool to explore the effects of perturbed glycosylation on RNA Polymerase II and in the transcriptional process.
Using immunofluorescence analysis, I have demonstrated that nuclear O-GlcNAc perturbation induces a strong reorganization of RNA Pol II transcription factories: in NPCs specifically, RNA Pol II molecules cluster in denser transcription foci all over the nuclear area. These results suggest that O-GlcNAcylation might be involved in the control of protein localization, a phenomenon that reminds phase separation molecule repartitioning.

It is well established that sequential phosphorylation of CTD Ser/Thr precisely regulates each phase of transcription. GlcNAcylation/phosphorylation reciprocity might be important in orchestrating transcriptional progression. O-GlcNAc enrichment at TSS, rather than at gene body, suggests a role for GlcNAc CTD in regulating promoter-proximal pausing, or the initial steps of transcription as pre-initiation complex (PIC) formation, promoter recognition and transcription initiation. To investigate the role of O-GlcNAc RNA Pol II, I have decided to engineer a novel monoclonal antibody against CTD Thr4 O-GlcNAc. The selected hybridoma clone shows specificity towards CTD Thr4

O-GlcNAc but cross-reacts with a plethora of proteins and is therefore not usable. Nonetheless this antibody represents the first available tool to assess the O-GlcNAcylation state of Thr4 residue of the RNA Pol II C-terminal domain.

To investigate more directly the effects of nuclear O-GlcNAc depletion on gene expression, I have exploited a novel *CXXC-BtGH84* cell line to perturb O-GlcNAc modification specifically at CpG-rich promoters. The RNA-seq results demonstrated that, in neuronal progenitor cells (NPCs), O-GlcNAc depletion at CpG-dense promoters causes the deregulation of disparate chromatin organization, histone modification genes and mitochondrial metabolic genes, and an important upregulation of several ribosomal protein genes. Because native OGT-IP and the CXXC-BtGH84 RNA-seq point towards a previously unknown regulation of O-GlcNAcylation in ribosomal biogenesis, I investigated the role of O-GlcNAc in rDNA transcriptional control. The results show that rRNA expression is strongly affected by O-GlcNAc depletion at CpG-dense promoters, especially at the stage of pre-rRNA transcription. All together, these data demonstrate that O-GlcNAc perturbation triggers an early effect on rDNA transcription which presumably the cells try to compensate by inducing an upregulation of ribosomal protein expression in later stages of NPC differentiation.

To map O-GlcNAc on nuclear proteins, I have implemented a biochemical approach based on lectin affinity to selectively purify nuclear O-GlcNAc proteins for ETD MS/MS analysis. By WB analysis of the purified O-GlcNAc protein pool, I have confirmed that RNA Pol II is modified by O-GlcNAc. However, due to the repetitive nature of the CTD sequence and due to the absence of CTD protease cleavage sites, it was not possible to map O-GlcNAc on RNA pol II CTD.
Several other interesting candidates involved in epigenetic regulation were also detected by Mass Spectrometry analysis: KMT2b, DNMT1, CHD2, DNMT3A, TET2 and disparate KRAB-Zfp proteins are few examples that represents potential candidate for future studies.
Mass Spectrometry data are also instrumental to gain insight into the O-GlcNAc protein features and to define the glycoprotein sequence determinants. Preliminary results of O-GlcNAc proteome immunoprecipitation between ESCs and NPCs identified significant differences in the O-GlcNAc levels on some protein candidates. These results indicate that O-GlcNAc might play a significant role in regulating the activity of these proteins during cellular differentiation.

Amino acid motif analysis identified a high frequency of Proline, Valine and Threonine residues at the boundaries of the O-GlcNAc S/T site, albeit no specific motif has emerged.

Overall, the proteomics results indicate that O-GlcNAcylated proteins are characterized by specific biochemical features and that their function is probably affected not only by the presence or absence of O-GlcNAc modification, but also by the number of residues that get modified in a specific cellular context.

6. Discussion

O-GlcNAcylation was identified for the first time only 30 years ago and since then a lot of progresses were achieved to understand the function of the O-GlcNAc modification. With my PhD project I have acquired new knowledge regarding the occupancy patterns of O-GlcNAc proteins in the mammalian chromatin context, and about the protein complexes containing OGT. Moreover, I have uncovered a new role for OGT in ribosome biogenesis.

My data demonstrate that O-GlcNAcylation modifies proteins enriched at gene promoters in both pluripotent and differentiated cells. O-GlcNAc -rich promoters are mostly transcriptionally active, however a subgroup of repressed genes, which includes IAP elements, is also regulated by O-GlcNAcylation. These results demonstrate that O-GlcNAc does not have a unique function in regulating transcription, but instead controls the expression of different subset of active and repressed genes. This is consistent with the OGT interactome results, which show that OGT regulates diverse cellular pathways. It is possible that the function of O-GlcNAcylation depends on the biochemical features of its substrates and on the chromatin and subcellular contexts. This is in fact not surprising, as a large and diverse group of proteins (more than 5000) were so far mapped as glycosylated.
Remarkably, O-GlcNAc chromatin occupancy in mES cells is antagonistic to PRC1 and PRC2, contrary to what was previously described in fly (Maria Cristina Gambetta, 2009). Instead, I and others found that OGT binds to BAP1, core component of the PR-DUB complex which counteracts PRC1 H2AK119ub1 deposition. These results suggest that, in contrast to fly, mammalian OGT has evolved a different function in regulating Polycomb activity during development.

My proteomic study confirms that OGT forms a stable complex with TET1 and TET2, which are amongst the most highly enriched OGT's binding partners. Vella *et al.* reported that TET recruits OGT at CGI. What is the functional significance of this interaction and what is OGT's function at CGI?

O-GlcNAc chromatin occupancy is highly correlated with RNA Pol II. These results indicate that O-GlcNAc is involved, beside other mechanisms, in RNA Pol II-dependent transcription. Despite intense efforts in generating novel O-GlcNAc RNA Pol II detection tools, the repetitive CTD nature makes the characterization of the CTD O-GlcNAcylation code still unresolved.

By using an inducible perturbation system, I have described a novel O-GlcNAc function in supervising the expression of ribosomal proteins and ribosomal RNA expression. In line with these results, I have shown that OGT interacts with a conspicuous number of proteins that localize to the nucleolus, the control center of ribosome production. These findings provide new evidence of a direct role of O-GlcNAc modification in regulating ribosomal biosynthesis. Interestingly, different studies have already explored the functional link between O-GlcNAcylation and protein translation:

- In human cells, OGT was found to associate with eIF4F, a protein complex involved in translation initiation control; the eIF4F protein subunits are O-GlcNAcylated (Li et al., 2019). Interestingly, in my WGA-IP IP I have identified eIF4G1 and eIF4G2, two components of the multi-subunit protein complex eIF4F (Jaiswal et al., 2018);

- Overexpression of OGT in human cells affects ribosomal subunits homeostasis by inducing an accumulation of 60S and 80S subunits (Zeidan et al., 2010). The authors of this study demonstrated that the ribosomal protein RPS6, key component of the rapamycin (mTOR) signaling pathway, is also O-GlcNAcylated. In the past years, several studies have described a link between the mTOR signaling pathway and O-GlcNAc modification. It is interesting to notice that the mTOR pathway and O-GlcNAcylation are directly dependent on the cellular nutrients availability and that are both involved in the control on protein translation and metabolisms (Wang and Proud, 2006). Intriguingly, mTOR and O-GlcNAcylation are interconnected: two independent studies have shown that inhibition of the mTOR pathway impairs OGT expression (Park et al., 2014) and that OGA inhibition affects mTOR pathway activation (Sodi et al., 2015);

- A recent study provided a detailed inventory of the human O-GlcNAc proteins and pointed out the presence of a high abundancy of ribosomal and proteasomal subunits among the mapped glycosylated factors (Wulff-Fuentes et al., 2021);

- In budding yeast, several components of the HBP were found to be essential for the assembly of the ribonucleoprotein granules, RNA-protein assemblies that regulate the mRNA translation. Here, O-GlcNAcylation of the translational machinery is necessary to ensure the successful aggregation of the untranslated messenger ribonucleoproteins into the ribonucleoprotein granules (Ohn et al., 2008).

Despite the precise mechanism of O-GlcNAc modification in regulating gene expression has still to be fully elucidated, my results support the hypothesis that O-GlcNAc supervises the cellular response that controls ribosomal production. It can be speculated that OGT might act as a metabolic sensor that feeds back to the protein production by modulating the ribosomal biosynthesis in response to the nutrients availability.
It is noteworthy that the O-GlcNAc modification can act in a post-translational and in a co-translational manner: O-GlcNAc was in fact detected not only on mature proteins, but also on nascent polypeptide chains where it regulates their stability by protecting the nascent polypeptides by premature ubiquitin-mediated proteasomal degradation (Zhu et al., 2015). Thus it is possible that O-GlcNAc might regulate ribosome's function also at a co-translational level.

It remains to be investigated whether O-GlcNAc RNA Pol II is involved in the regulation of ribosome biogenesis, as there are no evidences which indicate that RNA Pol I nor RNA Pol III are modificed by O-GlcNAc. It is well established that pre-rRNA, further processed into mature 28S, 18S and 5.8S rRNA, is synthetized by RNA Polymerase I, while the 5S rRNA is transcribed outside the nucleoli by RNA Pol III. Thus, rRNA production is usually investigated from the angle of RNA Pol I and III, however it has been recently shown that, in human cells, RNA Pol II has a role in supervising ribosome biogenesis; antisense transcription by Pol II generates R-loops shield at the rRNA large intergenic spacers (IGSs) that are essential to prevent Pol-I-dependent sense intergenic transcripts, which can be deleterious for nucleolar condensates formation and rRNA production. Because RNA Pol II can be modified by O-GlcNAc, and because O-GlcNAc has a role in ribosome biogenesis, it is intriguing to speculate that the O-GlcNAcylation

state of the CTD might be involved in the regulation of rRNA expression. Another interesting point that remains to be elucidate is whether O-GlcNAc also modulate RNA Pol III dependent transcription of 5S rRNA and tRNAs, as my ETD mass spectrometry experiment might suggest.

O-GlcNAc modifies over 5000 human proteins and is involved in a plethora of cellular pathways. With my PhD project, I have implemented new tools to characterize the O-GlcNAc epigenomic and biochemical maps and new strategies to modulate the O-GlcNAc levels in mammalian cells. Nonetheless, effective tools to perturb intracellular OGT levels and to modulate the O-GlcNAc state at specific genomic loci are still missing. Future experiments will be necessary to complete the ongoing alternative methods: 1) a novel dTAG-OGT degron cell line is currently under production and will be exploit to understand the effects of OGT loss in gene expression and in ribosome biogenesis, as well as to elucidate why OGT is dispensable for mammalian cell viability; 2) OGT and OGA enzymes will be targeted at rDNA sequence by dCas9 to assess their role in the rRNA transcriptional control. This strategy will give new insight into the direct mechanisms of O-GlcNAc regulation in ribosome biogenesis; 3) the generation of a genetically encoded effector to directly deplete O-GlcNAcylation from RNA Pol II, in combination with the RNA Pol II CTD Thr4 antibody, represent the first tools to map and explore the biological significance of the RNA Pol II O-GlcNAc code.

Future studies will be essential to characterize the still unanswered questions on O-GlcNAc function in regulating gene expression:

1. Which is the nature of O-GlcNAc proteins at each promoter or at each class of promoters? Further analysis will have to be implemented to identify the Gene Ontology characteristics of O-GlcNAc promoters and the identity of the O-GlcNAc proteins associated with each class of O-GlcNAc-rich promoters. Moreover, the function of O-GlcNAc at enhancers has never been described so far and mass spectrometry analysis of the O-GlcNAc proteome will help to identify novel O-GlcNAc factors which potentially regulate gene expression also at enhancers;

2. How does O-GlcNAc supervise the rRNA and ribosomal subunits transcription? Does O-GlcNAc on RNA Pol II regulate the expression of ribosomal subunits? Does

O-GlcNAc instead regulate ribosome biogenesis by modulating the activity of essential transcription factors at rDNA?;

3. Is O-GlcNAc at rDNA modulated by the metabolic status of the cell? Specific experiments targeted at modulating the glucose and nutrient media supplies will be important to infer whether the extracellular environment controls ribosomal biogenesis in response to changes in the O-GlcNAc intracellular levels mediated by the glucose intake;

4. How does O-GlcNAc modification regulate gene expression, at the identified classes of promoters as well as at rDNA? O-GlcNAc occupies a numerous group of gene promoters which transcriptional activity can be regulated by different transcription factors (TFs) and chromatin remodelers.

   Several TFs were previously characterized to be glycosylated and their O-GlcNAcylation state was shown to modulate their transcriptional activity, stability and binding with DNA or other co-factors. Some examples of O-GlcNAcylated TFs are NF-kB, STAT5A, CREB, YYI, PGC-1α, p53, ER-α/β, NEURO-D1, TORC2, NFATc1, ELF-1, c-MYC, FOXO1, PDX-1, C/EBPβ, MAFA, ID2, USF, SP1 (Özcan et al., 2010). Moreover, O-GlcNAc on RNA Pol II CTD and on the Transcription Initiation Factors (TFIIs) might also cooperate, together with the O-GlcNAc TFs, to the O-GlcNAc -mediated transcriptional regulation.

   In addition, it can be speculated that O-GlcNAc might be a master regulator of metabolic pathways by supervising the transcriptional state or the protein stability of metabolic proteins, like for the mTOR pathway.

7. Bibliography

Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II. Molecular Cell *34*, 387-393.

Alexander, K.A., Wang, X., Shibata, M., Clark, A.G., and García-García, M.J. (2015). TRIM28 Controls Genomic Imprinting through Distinct Mechanisms during and after Early Genome-wide Reprogramming. Cell Reports *13*, 1194-1205.

Andersen, J.S., Lam, Y.W., Leung, A.K.L., Ong, S.-E., Lyon, C.E., Lamond, A.I., and Mann, M. (2005). Nucleolar proteome dynamics. Nature *433*, 77-83.

Andersen, J.S., Lyon, C.E., Fox, A.H., Leung, A.K.L., Lam, Y.W., Steen, H., Mann, M., and Lamond, A.I. (2002). Directed Proteomic Analysis of the Human Nucleolus. Current Biology *12*, 1-11.

Bahry, E., Breimann, L., Zouinkhi, M., Epstein, L., Kolyvanov, K., Mamrak, N., King, B., Long, X., Harrington, K.I.S., Lionnet, T.*, et al.* (2022). RS-FISH: precise, interactive, fast, and scalable FISH spot detection. Nature Methods *19*, 1563-1567.

Baskaran, R., Dahmus, M.E., and Wang, J.Y. (1993). Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. Proceedings of the National Academy of Sciences *90*, 11167-11171.

Bauer, M., Trupke, J., and Ringrose, L. (2016). The quest for mammalian Polycomb response elements: are we there yet? Chromosoma *125*, 471-496.

Bender, C.M., Gonzalgo, M.L., Gonzales, F.A., Nguyen, C.T., Robertson, K.D., and Jones, P.A. (1999). Roles of Cell Division and Gene Transcription in the Methylation of CpG Islands. Molecular and Cellular Biology *19*, 6690-6698.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. Journal of Molecular Biology *203*, 971-983.

Bestor, T.H. (2000). The DNA methyltransferases of mammals. Human Molecular Genetics *9*, 2395-2402.

Bestor, T.H., Edwards, J.R., and Boulard, M. (2015). Notes on the role of dynamic DNA methylation in mammalian development. Proceedings of the National Academy of Sciences *112*, 6796-6799.

Blackledge, N.P., Fursova, N.A., Kelley, J.R., Huseyin, M.K., Feldmann, A., and Klose, R.J. (2020). PRC1 Catalytic Activity Is Central to Polycomb System Function. Molecular Cell *77*, 857-874.e859.

Boersma, S., Khuperkar, D., Verhagen, B.M.P., Sonneveld, S., Grimm, J.B., Lavis, L.D., and Tanenbaum, M.E. (2019). Multi-Color Single-Molecule Imaging Uncovers Extensive Heterogeneity in mRNA Decoding. Cell *178*, 458-472 e419.

Bond, M.R., and Hanover, J.A. (2013). O-GlcNAc Cycling: A Link Between Metabolism and Chronic Disease. Annual Review of Nutrition *33*, 205-229.

Borsook, H. (1958). A Rate-Governing Reaction of Protein Synthesis. Nature *182*, 1006-1007.

Borun, T.W., Pearson, D., and Paik, W.K. (1972). Studies of Histone Methylation during the HeLa S-3 Cell Cycle. Journal of Biological Chemistry *247*, 4288-4298.

Boulard, M., Edwards, J.R., and Bestor, T.H. (2015). FBXL10 protects Polycomb-bound genes from hypermethylation. Nature Genetics *47*, 479-485.

Boulard, M., Rucli, S., Edwards, J.R., and Bestor, T.H. (2020). Methylation-directed glycosylation of chromatin factors represses retrotransposon promoters. Proceedings of the National Academy of Sciences *117*, 14292-14298.

Bourc'his, D., Xu, G.-L., Lin, C.-S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the Establishment of Maternal Genomic Imprints. Science *294*, 2536-2539.

Brookes, E., and Pombo, A. (2009). Modifications of RNA polymerase II are pivotal in regulating gene expression states. EMBO reports *10*, 1213-1219.

Brown, S.J., Cole, M.D., and Erives, A.J. (2008). Evolution of the holozoan ribosome biogenesis regulon. BMC Genomics *9*, 442.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Current Protocols in Molecular Biology *109*.

Burger, K., and Eick, D. (2013). Functional ribosome biogenesis is a prerequisite for p53 destabilization: impact of chemotherapy on nucleolar functions and RNA metabolism. Biol Chem *394*, 1133-1143.

Burt, R.A., Dejanovic, B., Peckham, H.J., Lee, K.A., Li, X., Ounadjela, J.R., Rao, A., Malaker, S.A., Carr, S.A., and Myers, S.A. (2021). Novel Antibodies for the Simple and Efficient Enrichment of Native O-GlcNAc Modified Peptides. Molecular & Cellular Proteomics *20*, 100167.

Byvoet, P., Shepherd, G.R., Hardin, J.M., and Noland, B.J. (1972). The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. Archives of Biochemistry and Biophysics *148*, 558-567.

Capotosti, F., Guernier, S., Lammers, F., Waridel, P., Cai, Y., Jin, J., Conaway, J.W., Conaway, R.C., and Herr, W. (2011). O-GlcNAc Transferase Catalyzes Site-Specific Proteolysis of HCF-1. Cell *144*, 376-388.

Chaker-Margot, M. (2018). Assembly of the small ribosomal subunit in yeast: mechanism and regulation. RNA *24*, 881-891.

Chapman, R.D., Conrad, M., and Eick, D. (2005). Role of the Mammalian RNA Polymerase II C-Terminal Domain (CTD) Nonconsensus Repeats in CTD Stability and Cell Proliferation. Molecular and Cellular Biology *25*, 7665-7674.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science *318*, 1780-1782.

Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2012). TET2 promotes histone O-GlcNAcylation during gene transcription. Nature *493*, 561-564.

Cheng, X., and Hart, G.W. (2001). AlternativeO-Glycosylation/O-Phosphorylation of Serine-16 in Murine Estrogen Receptor β. Journal of Biological Chemistry *276*, 10570-10575.

Chu, C.-S., Lo, P.-W., Yeh, Y.-H., Hsu, P.-H., Peng, S.-H., Teng, Y.-C., Kang, M.-L., Wong, C.-H., and Juan, L.-J. (2014). O-GlcNAcylation regulates EZH2 protein stability and function. Proceedings of the National Academy of Sciences *111*, 1355-1360.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J.-H., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes & Development *26*, 1714-1728.

Comer, F.I., and Hart, G.W. (2001). Reciprocity between O-GlcNAc and O-Phosphate on the Carboxyl Terminal Domain of RNA Polymerase II. Biochemistry *40*, 7845-7852.

Community, T.G. (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Research *50*, W345-W351.

Conaghan, J., Handyside, A.H., Winston, R.M.L., and Leese, H.J. (1993). Effects of pyruvate and glucose on the development of human preimplantation embryos in vitro. Reproduction *99*, 87-95.

Conway, E., Rossi, F., Fernandez-Perez, D., Ponzo, E., Ferrari, K.J., Zanotti, M., Manganaro, D., Rodighiero, S., Tamburri, S., and Pasini, D. (2021). BAP1 enhances Polycomb repression by counteracting widespread H2AK119ub1 deposition and chromatin condensation. Molecular Cell *81*, 3526-3541.e3528.

Couté, Y., Burgess, J.A., Diaz, J.-J., Chichester, C., Lisacek, F., Greco, A., and Sanchez, J.-C. (2006). Deciphering the human nucleolar proteome. Mass Spectrometry Reviews *25*, 215-234.

Dai, H.-Q., Wang, B.-A., Yang, L., Chen, J.-J., Zhu, G.-C., Sun, M.-L., Ge, H., Wang, R., Chapman, D.L., Tang, F.*, et al.* (2016). TET-mediated DNA demethylation controls gastrulation by regulating Lefty–Nodal signalling. Nature *538*, 528-532.

Dawlaty, M., Breiling, A., Le, T., Raddatz, G., M, Albert, Gao, Q., Benjamin, Li, Z., Xu, M.*, et al.* (2013). Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. Developmental Cell *24*, 310-323.

Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D.F., Burkhard, P.R., and Sanchez, J.-C. (2008). Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. Analytical Chemistry *80*, 2921-2931.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes & Development *25*, 1010-1022.

Dennis, R.J., Taylor, E.J., Macauley, M.S., Stubbs, K.A., Turkenburg, J.P., Hart, S.J., Black, G.N., Vocadlo, D.J., and Davies, G.J. (2006). Structure and mechanism of a bacterial β-glucosaminidase having O-GlcNAcase activity. Nature Structural & Molecular Biology *13*, 365-371.

Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Méndez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E.*, et al.* (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. The EMBO Journal *32*, 645-655.

Dias, W.B., Cheung, W.D., Wang, Z., and Hart, G.W. (2009). Regulation of Calcium/Calmodulin-dependent Kinase IV by O-GlcNAc Modification. Journal of Biological Chemistry *284*, 21327-21337.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Research *11*, 1475-1489.

Doskočil, J., and Šorm, F. (1962). Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. Biochimica et Biophysica Acta *55*, 953-959.

Du, X.L., Edelstein, D., Dimmeler, S., Ju, Q., Sui, C., and Brownlee, M. (2001). Hyperglycemia inhibits endothelial nitric oxide synthase activity by posttranslational modification at the Akt site. Journal of Clinical Investigation *108*, 1341-1348.

Edwards, J.R., Yarychkivska, O., Boulard, M., and Bestor, T.H. (2017). DNA methylation and DNA methyltransferases. Epigenetics & Chromatin *10*.

Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA Polymerase II CTD Is Specifically Required for snRNA Gene Expression. Science *318*, 1777-1779.

Eichler, D.C., and Craig, N. (1994). Processing of Eukaryotic Ribosomal RNA. In, W.E. Cohn, and K. Moldave, eds. (Progress in Nucleic Acid Research and Molecular Biology), pp. 197-239.

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev *113*, 8456-8490.

Elsässer, S.J., Noh, K.-M., Diaz, N., Allis, C.D., and Banaszynski, L.A. (2015). Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature *522*, 240-244.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A*., et al.* (2010). Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. Molecular Cell *38*, 452-464.

F. I. Comer, G.W.H. (2001). Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *40*, 7845-7852.

Falvey, A.K., and Staehelin, T. (1970). Structure and function of mammalian ribosomes. Journal of Molecular Biology *53*, 1-19.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E*., et al.* (2010). Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences *107*, 8689-8694.

Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nature Reviews Genetics *12*, 565-575.

Fischle, W., Franz, H., Jacobs, S.A., Allis, C.D., and Khorasanizadeh, S. (2008). Specificity of the Chromodomain Y Chromosome Family of Chromodomains for Lysine-methylated ARK(S/T) Motifs. Journal of Biological Chemistry *283*, 19626-19635.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574-1577.

Fromont-Racine, M., Senger, B., Saveanu, C., and Fasiolo, F. (2003). Ribosome assembly in eukaryotes. Gene *313*, 17-42.

Fursova, N.A., Turberfield, A.H., Blackledge, N.P., Findlater, E.L., Lastuvkova, A., Huseyin, M.K., Dobrinić, P., and Klose, R.J. (2021). BAP1 constrains pervasive H2AK119ub1 to control the transcriptional potential of the genome. Genes & Development *35*, 749-770.

Gallagher, J.T., Morris, A., and Dexter, T.M. (1985). Identification of two binding sites for wheat-germ agglutinin on polylactosamine-type oligosaccharides. Biochemical Journal *231*, 115-122.

Gambetta, Maria C., and Müller, J. (2014). O-GlcNAcylation Prevents Aggregation of the Polycomb Group Repressor Polyhomeotic. Developmental Cell *31*, 629-639.

Gambetta, M.C., and Müller, J. (2015). A critical perspective of the diverse roles of O-GlcNAc transferase in chromatin. Chromosoma *124*, 429-442.

Gambetta, M.C., Oktaba, K., and Muller, J. (2009). Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. Science *325*, 93-96.

Ge, W., Yu, C., Li, J., Yu, Z., Li, X., Zhang, Y., Liu, C.-P., Li, Y., Tian, C., Zhang, X*., et al.* (2023). Basis of the H2AK119 specificity of the Polycomb repressive deubiquitinase. Nature *616*, 176-182.

Gershey, E.L., Haslett, G.W., Vidali, G., and Allfrey, V.G. (1969). Chemical Studies of Histone Methylation: EVIDENCE FOR THE OCCURRENCE OF 3-METHYLHISTIDINE IN AVIAN ERYTHROCYTE HISTONE FRACTIONS. Journal of Biological Chemistry *244*, 4871-4877.

Ghamari, A., van de Corput, M.P.C., Thongjuea, S., van Cappellen, W.A., van Ijcken, W., van Haren, J., Soler, E., Eick, D., Lenhard, B., and Grosveld, F.G. (2013). In vivo live imaging of RNA polymerase II transcription factories in primary cells. Genes & Development *27*, 767-777.

Ghulam, M.M., Catala, M., and Abou Elela, S. (2019). Differential expression of duplicated ribosomal protein genes modifies ribosome composition in response to stress. Nucleic Acids Research *48*, 1954-1968.

Gibbons, J.G., Branco, A.T., Godinho, S.A., Yu, S., and Lemos, B. (2015). Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. Proceedings of the National Academy of Sciences *112*, 2485-2490.

Glover-Cutter, K., Larochelle, S.p., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. Molecular and Cellular Biology *29*, 5455-5464.

Goodfellow, S.J., and Zomerdijk, J.C.B.M. (2013). Basic Mechanisms in RNA Polymerase I Transcription of the Ribosomal RNA Genes. In (Epigenetics: Development and Disease), pp. 211-236.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proceedings of the National Academy of Sciences *89*, 5547-5551.

Grant, M., Zuccotti, M., and Monk, M. (1992). Methylation of CpG sites of two X–linked genes coincides with X–inactivation in the female mouse embryo but not in the germ line. Nature Genetics *2*, 161-166.

Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. Nature Reviews Genetics *13*, 343-357.

Gretarsson, K.H., and Hackett, J.A. (2020). Dppa2 and Dppa4 counteract de novo methylation to establish a permissive epigenome for development. Nature Structural & Molecular Biology *27*, 706-716.

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S.-G.*, et al.* (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606-610.

Hanover, J.A., Yu, S., Lubas, W.B., Shin, S.-H., Ragano-Caracciola, M., Kochran, J., and Love, D.C. (2003). Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. Archives of Biochemistry and Biophysics *409*, 287-297.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nature Cell Biology *10*, 1291-1300.

Hart, G.W., Housley, M.P., and Slawson, C. (2007). Cycling of O-linked β-N-acetylglucosamine on nucleocytoplasmic proteins. Nature *446*, 1017-1022.

Hart, W.G.K.G.W. (1989). Glycosylation of chromosomal proteins Localization of O-linked N-acetylglucosamine in Drosophila chromatin. Cell *57*, 243-251.

Harvey, A., Caretti, G., Moresi, V., Renzini, A., and Adamo, S. (2019). Interplay between Metabolites and the Epigenome in Regulating Embryonic and Adult Stem Cell Potency and Maintenance. Stem Cell Reports *13*, 573-589.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Research *40*, 4841-4849.

Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., Aebersold, R., Paro, R., Gstaiger, M., and Beisel, C. (2016). A High-Density Map for Navigating the Human Polycomb Complexome. Cell Reports *17*, 583-595.

Hempel, K., Lange, H.W., and Birkofer, L. (1968). [Epsilon-N-trimethyllysine, a new amino acid in histones]. Naturwissenschaften *55*, 37.

Henderson, A.S., Warburton, D., and Atwood, K.C. (1972). Location of Ribosomal DNA in the Human Chromosome Complement. Proceedings of the National Academy of Sciences *69*, 3394-3398.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E*., et al.* (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. The EMBO Journal *31*, 2784-2797.

Hong Zhang, G.G.a.U.T.B. (1992). Extracellular reduction of alloxan results in oxygen radicaLmediated attack on plasma and lysosomal membranes. APMIS, 317.

Hori, Y., Engel, C., and Kobayashi, T. (2023). Regulation of ribosomal RNA gene copy number, transcription and nucleolus organization in eukaryotes. Nature Reviews Molecular Cell Biology.

Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2014). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Research *43*, D512-D520.

Housley, M.P., Rodgers, J.T., Udeshi, N.D., Kelly, T.J., Shabanowitz, J., Hunt, D.F., Puigserver, P., and Hart, G.W. (2008). O-GlcNAc Regulates FoxO Activation in Response to Glucose. Journal of Biological Chemistry *283*, 16283-16292.

Housley, M.P., Udeshi, N.D., Rodgers, J.T., Shabanowitz, J., Puigserver, P., Hunt, D.F., and Hart, G.W. (2009). A PGC-1α-O-GlcNAc Transferase Complex Regulates FoxO Transcription Factor Activity in Response to Glucose. Journal of Biological Chemistry *284*, 5148-5157.

Hsin, J.-P.S., Amit
Manley, James L. (2011). RNAP II CTD Phosphorylated on Threonine-4 Is Required for Histone mRNA 3’ End Processing. Science *334* 683 - 686.

Huang, K.-Y., Lee, T.-Y., Kao, H.-J., Ma, C.-T., Lee, C.-C., Lin, T.-H., Chang, W.-C., and Huang, H.-D. (2019). dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications. Nucleic Acids Research *47*, D298-D308.

Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., Steinmetz, L.M., and Krijgsveld, J. (2014). Ultrasensitive proteome analysis using paramagnetic bead technology. Molecular Systems Biology *10*, 757.

Hughes, C.S., Moggridge, S., Müller, T., Sorensen, P.H., Morin, G.B., and Krijgsveld, J. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nature Protocols *14*, 68-85.

Huh, W.-K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686-691.

Iacovino, M., Bosnakovski, D., Fey, H., Rux, D., Bajwa, G., Mahen, E., Mitanoska, A., Xu, Z., and Kyba, M. (2011). Inducible Cassette Exchange: A Rapid and Efficient System Enabling Conditional Gene Expression in Embryonic Stem and Primary Cells. Stem Cells *29*, 1580-1588.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science *333*, 1300-1303.

Jain, K., Fraser, C.S., Marunde, M.R., Parker, M.M., Sagum, C., Burg, J.M., Hall, N., Popova, I.K., Rodriguez, K.L., Vaidya, A.*, et al.* (2020). Characterization of the plant homeodomain (PHD) reader family for their histone tail interactions. Epigenetics & Chromatin *13*.

Jaiswal, P.K., Koul, S., Shanmugam, P.S.T., and Koul, H.K. (2018). Eukaryotic Translation Initiation Factor 4 Gamma 1 (eIF4G1) is upregulated during Prostate cancer progression and modulates cell growth and metastasis. Scientific Reports *8*.

Janetzko, J., and Walker, S. (2014). The Making of a Sweet Modification: Structure and Function of O-GlcNAc Transferase. Journal of Biological Chemistry *289*, 34424-34432.

Jansen, R.P., Hurt, E.C., Kern, H., Lehtonen, H., Carmo-Fonseca, M., Lapeyre, B., and Tollervey, D. (1991). Evolutionary conservation of the human nucleolar protein fibrillarin and its functional expression in yeast. Journal of Cell Biology *113*, 715-729.

Julien, E.H., Winship (2003). Proteolytic processing is necessary to separate and ensure proper cell growth and cytokinesis functions of HCF-1. The EMBO Journal *22*, 2360-2369.

Jurkowska, R.Z., Jurkowski, T.P., and Jeltsch, A. (2011). Structure and Function of Mammalian DNA Methyltransferases. ChemBioChem *12*, 206-222.

Kang, J., Lienhard, M., Pastor, W.A., Chawla, A., Novotny, M., Tsagaratou, A., Lasken, R.S., Thompson, E.C., Surani, M.A., Koralov, S.B.*, et al.* (2015). Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. Proceedings of the National Academy of Sciences *112*, E4236-E4245.

Kelly, W.G., Dahmus, M.E., and Hart, G.W. (1993). RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. Journal of Biological Chemistry *268*, 10416-10424.

Kelly, W.G., and Hart, G.W. (1989). Glycosylation of chromosomal proteins: Localization of O-linked N-acetylglucosamine in Drosophila chromatin. Cell *57*, 243-251.

Kim, E.Y., Jeong, E.H., Park, S., Jeong, H.-J., Edery, I., and Cho, J.W. (2012). A role for O-GlcNAcylation in setting circadian clock speed. Genes & Development *26*, 490-502.

Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., and Ban, N. (2011). Crystal Structure of the Eukaryotic 60S Ribosomal Subunit in Complex with Initiation Factor 6. Science *334*, 941-948.

Ko, M., An, J., Bandukwala, H.S., Chavez, L., Äijö, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lähdesmäki, H.*, et al.* (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature *497*, 122-126.

Komarnitsky, P., Cho, E.-J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes & Development *14*, 2452-2460.

Kreppel, L.K., Blomberg, M.A., and Hart, G.W. (1997). Dynamic Glycosylation of Nuclear and Cytosolic Proteins. Journal of Biological Chemistry *272*, 9308-9315.

Kressler, D., Hurt, E., and Baβler, J. (2010). Driving ribosome assembly. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1803*, 673-683.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M.*, et al.* (2003). The Paf1 Complex Is Required for Histone H3 Methylation by COMPASS and Dot1p: Linking Transcriptional Elongation to Histone Methylation. Molecular Cell *11*, 721-729.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S.*, et al.* (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. PLoS Genetics *4*, e1000242.

Laferté, A., Favry, E., Sentenac, A., Riva, M., Carles, C., and Chédin, S. (2006). The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components. Genes & Development *20*, 2030-2040.

Lafontaine, D.L.J., and Tollervey, D. (2001). The function and synthesis of ribosomes. Nature Reviews Molecular Cell Biology *2*, 514-520.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology *10*, R25.

Lazarus, M.B., Jiang, J., Kapuria, V., Bhuiyan, T., Janetzko, J., Zandberg, W.F., Vocadlo, D.J., Herr, W., and Walker, S. (2013). HCF-1 Is Cleaved in the Active Site of O-GlcNAc Transferase. Science *342*, 1235-1239.

Lazarus, M.B., Nam, Y., Jiang, J., Sliz, P., and Walker, S. (2011). Structure of human O-GlcNAc transferase and its complex with a peptide substrate. Nature *469*, 564-567.

Lee, J.-H., and Skalnik, D.G. (2005). CpG-binding Protein (CXXC Finger Protein 1) Is a Component of the Mammalian Set1 Histone H3-Lys4 Methyltransferase Complex, the Analogue of the Yeast Set1/COMPASS Complex. Journal of Biological Chemistry *280*, 41725-41731.

Levine, Z.G., Fan, C., Melicher, M.S., Orman, M., Benjamin, T., and Walker, S. (2018). O-GlcNAc Transferase Recognizes Protein Substrates Using an Asparagine Ladder in the Tetratricopeptide Repeat (TPR) Superhelix. Journal of the American Chemical Society *140*, 3510-3513.

Lewis, B.A., Burlingame, A.L., and Myers, S.A. (2016). Human RNA Polymerase II Promoter Recruitment in Vitro Is Regulated by O-Linked N-

Acetylglucosaminyltransferase (OGT). Journal of Biological Chemistry *291*, 14056-14061.

Lewis, P., and Mislove, R. (1947). New mutants report. Drosoph Inf Serv *21*, 69.

Li, E., and Zhang, Y. (2014). DNA Methylation in Mammals. Cold Spring Harbor Perspectives in Biology *6*, a019133-a019133.

Li, H., Xing, X., Ding, G., Li, Q., Wang, C., Xie, L., Zeng, R., and Li, Y. (2009). SysPTM: A Systematic Resource for Proteomic Research on Post-translational Modifications. Molecular & Cellular Proteomics *8*, 1839-1849.

Li, M., Phatnani, H.P., Guan, Z., Sage, H., Greenleaf, A.L., and Zhou, P. (2005). Solution structure of the Set2–Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. Proceedings of the National Academy of Sciences *102*, 17636-17641.

Li, X., Yue, X., Sepulveda, H., Burt, R.A., Scott, D.A., A. Carr, S., A. Myers, S., and Rao, A. (2023). OGT controls mammalian cell viability by regulating the proteasome/mTOR/ mitochondrial axis. Proceedings of the National Academy of Sciences *120*.

Li, X., Zhu, Q., Shi, X., Cheng, Y., Li, X., Xu, H., Duan, X., Hsieh-Wilson, L.C., Chu, J., Pelletier, J.*, et al.* (2019). O-GlcNAcylation of core components of the translation initiation machinery regulates protein synthesis. Proceedings of the National Academy of Sciences *116*, 7857-7866.

Liao, S.M., Zhang, J., Jeffery, D.A., Koleske, A.J., Thompson, C.M., Chao, D.M., Viljoen, M., van Vuuren, H.J., and Young, R.A. (1995). A kinase-cyclin pair in the RNA polymerase II holoenzyme. Nature *374*, 193-196.

Liao, X.-H., Majithia, A., Huang, X., and Kimmel, A.R. (2008). Growth control via TOR kinase signaling, an intracellular sensor of amino acid and energy availability, with crosstalk potential to proline metabolism. Amino Acids *35*, 761-770.

Long, Hannah K., Blackledge, Neil P., and Klose, Robert J. (2013). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochemical Society Transactions *41*, 727-740.

Love, D.C., Ghosh, S., Mondoux, M.A., Fukushige, T., Wang, P., Wilson, M.A., Iser, W.B., Wolkow, C.A., Krause, M.W., and Hanover, J.A. (2010a). Dynamic O-GlcNAc cycling at promoters of Caenorhabditis elegans genes regulating longevity, stress, and immunity. Proceedings of the National Academy of Sciences *107*, 7413-7418.

Love, D.C., Ghosh, S., Mondoux, M.A., Fukushige, T., Wang, P., Wilson, M.A., Iser, W.B., Wolkow, C.A., Krause, M.W., and Hanover, J.A. (2010b). Dynamic O-GlcNAc cycling at promoters of Caenorhabditis elegans genes regulating longevity, stress, and immunity. Proc Natl Acad Sci U S A *107*, 7413-7418.

Love, D.C., and Hanover, J.A. (2005). The hexosamine signaling pathway: deciphering the "O-GlcNAc code". Sci STKE *2005*, re13.

Love, D.C., Kochran, J., Cathey, R.L., Shin, S.-H., and Hanover, J.A. (2003). Mitochondrial and nucleocytoplasmic targeting of O-linked GlcNAc transferase. Journal of Cell Science *116*, 647-654.

Lu, C., and Craig (2012). Metabolic Regulation of Epigenetics. Cell Metabolism *16*, 9-17.

Lu, H., Yu, D., Hansen, A.S., Ganguly, S., Liu, R., Heckert, A., Darzacq, X., and Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. Nature *558*, 318-323.

Lubas, W.A., Frank, D.W., Krause, M., and Hanover, J.A. (1997). O-Linked GlcNAc Transferase Is a Conserved Nucleocytoplasmic Protein Containing Tetratricopeptide Repeats. Journal of Biological Chemistry *272*, 9316-9324.

Lutje Hulsik, D., Liu, Y.-Y., Strokappe, N.M., Battella, S., El Khattabi, M., McCoy, L.E., Sabin, C., Hinz, A., Hock, M., Macheboeuf, P.*, et al.* (2013). A gp41 MPER-specific Llama VHH Requires a Hydrophobic CDR3 for Neutralization but not for Antigen Recognition. PLoS Pathogens *9*, e1003202.

Lv, P., Du, Y., He, C., Peng, L., Zhou, X., Wan, Y., Zeng, M., Zhou, W., Zou, P., Li, C.*, et al.* (2022). O-GlcNAcylation modulates liquid–liquid phase separation of SynGAP/PSD-95. Nature Chemistry *14*, 831-840.

Mahajan, S.S., and Wilson, A.C. (2000). Mutations in Host Cell Factor 1 Separate Its Role in Cell Proliferation from Recruitment of VP16 and LZIP. Molecular and Cellular Biology *20*, 919-928.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The Protein Kinase Complement of the Human Genome. Science *298*, 1912-1934.

Maria Cristina Gambetta, K.O. (2009). Essential Role of the Glycosyltransferase SxcOgt in Polycomb Repression. Science *325*, 93-96.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. Science *336*, 1723-1725.

Mayer, C., and Grummt, I. (2006). Ribosome biogenesis and cell growth: mTOR coordinates transcription by all three classes of nuclear RNA polymerases. Oncogene *25*, 6384-6391.

McStay, B. (2016). Nucleolar organizer regions: genomic 'dark matter' requiring illumination. Genes & Development *30*, 1598-1610.

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., and Thomas, P.D. (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nature Protocols *14*, 703-721.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Miyazawa, H., and Aulehla, A. (2018). Revisiting the role of metabolism during development. Development *145*, dev131110.

Mohammad, Goyal, P., Irina, Bilenky, M., Leung, D., Jie, Shinkai, Y., Dixie, Jones, S., Hirst, M.*, et al.* (2011). DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. Cell Stem Cell *8*, 676-687.

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development *99*, 371-382.

Montefiori, L., Hernandez, L., Zhang, Z., Gilad, Y., Ober, C., Crawford, G., Nobrega, M., and Jo Sakabe, N. (2017). Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. Scientific Reports *7*.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Maria, Vasanthakumar, A., Patel, J., Zhao, X.*, et al.* (2011). Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. Cancer Cell *20*, 11-24.

Moss, T. (2004). At the crossroads of growth control; making ribosomal RNA. Current Opinion in Genetics & Development *14*, 210-217.

Murray, K. (1964). The Occurrence of iε-N-Methyl Lysine in Histones. Biochemistry *3*, 10-15.

Myers, S.A., Panning, B., and Burlingame, A.L. (2011). Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. Proceedings of the National Academy of Sciences *108*, 9490-9495.

Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A.*, et al.* (2018). The dTAG system for immediate and target-specific protein degradation. Nature Chemical Biology *14*, 431-441.

Nagashima, R., Hibino, K., Ashwin, S.S., Babokhov, M., Fujishiro, S., Imai, R., Nozaki, T., Tamura, S., Tani, T., Kimura, H.*, et al.* (2019). Single nucleosome imaging reveals loose genome chromatin networks via active RNA polymerase II. Journal of Cell Biology *218*, 1511-1530.

Nagel, A.K., and Ball, L.E. (2014). O-GlcNAc transferase and O-GlcNAcase: achieving target substrate specificity. Amino Acids *46*, 2305-2316.

Nolte, D., and Müller, U. (2002). Human O-GlcNAc transferase (OGT): genomic structure, analysis of splice variants, fine mapping in Xq13.1. Mammalian Genome *13*, 62-64.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A.*, et al.* (2022). The complete sequence of a human genome. Science *376*, 44-53.

O'Donnell, N., Zachara, N.E., Hart, G.W., and Marth, J.D. (2004). Ogt-Dependent X-Chromosome-Linked Protein Glycosylation Is a Requisite Modification in Somatic Cell Function and Embryo Viability. Molecular and Cellular Biology *24*, 1680-1690.

Ochoa, D., Jarnuczak, A.F., Viéitez, C., Gehre, M., Soucheray, M., Mateus, A., Kleefeldt, A.A., Hill, A., Garcia-Alonso, L., Stein, F.*, et al.* (2020). The functional landscape of the human phosphoproteome. Nature Biotechnology *38*, 365-373.

Ohn, T., Kedersha, N., Hickman, T., Tisdale, S., and Anderson, P. (2008). A functional RNAi screen links O-GlcNAc modification of ribosomal proteins to stress granule and processing body assembly. Nature Cell Biology *10*, 1224-1231.

Oki, S., Ohta, T., Shioi, G., Hatanaka, H., Ogasawara, O., Okuda, Y., Kawaji, H., Nakaki, R., Sese, J., and Meno, C. (2018). ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. EMBO Rep *19*.

Oktaba, K., Gutiérrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E.M., and Müller, J. (2008). Dynamic Regulation by Polycomb Group Protein Complexes Controls Pattern Formation and the Cell Cycle in Drosophila. Developmental Cell *15*, 877-889.

Özcan, S., Andrali, S.S., and Cantrell, J.E.L. (2010). Modulation of transcription factor function by O-GlcNAc modification. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1799*, 353-364.

Paik, W.K., and Kim, S. (1967). E-N-dimethyllysine in histones. Biochemical and Biophysical Research Communications *27*, 479-483.

Papanastasiou, M., Mullahoo, J., DeRuff, K.C., Bajrami, B., Karageorgos, I., Johnston, S.E., Peckner, R., Myers, S.A., Carr, S.A., and Jaffe, J.D. (2019). Chasing Tails: Cathepsin-L Improves Structural Analysis of Histones by HX-MS. Mol Cell Proteomics *18*, 2089-2098.

Park, S., Pak, J., Jang, I., and Cho, J.W. (2014). Inhibition of mTOR affects protein stability of OGT. Biochemical and Biophysical Research Communications *453*, 208-212.

Pelletier, J., Thomas, G., and Volarević, S. (2018). Ribosome biogenesis in cancer: new players and therapeutic avenues. Nature Reviews Cancer *18*, 51-63.

Pengelly, A.R., Copur, O., Jackle, H., Herzig, A., and Muller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. Science *339*, 698-699.

Pengelly, A.R., Kalb, R., Finkl, K., and Muller, J. (2015). Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. Genes Dev *29*, 1487-1492.

Petracovici, A., and Bonasio, R. (2021). Distinct PRC2 subunits regulate maintenance and establishment of Polycomb repression during differentiation. Molecular Cell *81*, 2625-2639.e2625.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes & Development *20*, 2922-2936.

Powers, T., and Walter, P. (1999). Regulation of Ribosome Biogenesis by the Rapamycin-sensitive TOR-signaling Pathway in *Saccharomyces cerevisiae*. Molecular Biology of the Cell *10*, 987-1000.

Pravata, V.M., Gundogdu, M., Bartual, S.G., Ferenbach, A.T., Stavridis, M., Õunap, K., Pajusalu, S., Žordania, R., Wojcik, M.H., and Aalten, D.M.F. (2020). A missense mutation in the catalytic domain of O-GlcNAc transferase links perturbations in protein O-GlcNAcylation to X-linked intellectual disability. FEBS Letters *594*, 717-727.

Pravata, V.M., Muha, V., Gundogdu, M., Ferenbach, A.T., Kakade, P.S., Vandadi, V., Wilmes, A.C., Borodkin, V.S., Joss, S., Stavridis, M.P.*, et al.* (2019). Catalytic deficiency of O-GlcNAc transferase leads to X-linked intellectual disability. Proceedings of the National Academy of Sciences *116*, 14961-14970.

Proudhon, C., Duffié, R., Ajjan, S., Cowley, M., Iranzo, J., Carbajosa, G., Saadeh, H., Holland, Michelle L., Oakey, Rebecca J., Rakyan, Vardhman K.*, et al.* (2012). Protection against De Novo Methylation Is Instrumental in Maintaining Parent-of-Origin Methylation Inherited from the Gametes. Molecular Cell *47*, 909-920.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A.*, et al.* (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell *44*, 361-372.

Quivoron, C., Couronné, L., Véronique, Cécile, Plo, I., Wagner-Ballon, O., Marcio, Delhommeau, F., Arnulf, B., Stern, M.-H.*, et al.* (2011). TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. Cancer Cell *20*, 25-38.

Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A., and Ban, N. (2011). Crystal Structure of the Eukaryotic 40S Ribosomal Subunit in Complex with Initiation Factor 1. Science *331*, 730-736.

Radermacher, P.T., Myachina, F., Bosshardt, F., Pandey, R., Mariappa, D., Müller, H.A.J., and Lehner, C.F. (2014). O-GlcNAc reports ambient temperature and confers heat resistance on ectotherm development. Proceedings of the National Academy of Sciences *111*, 5592-5597.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc Regulates Transcriptional Pause Release. Cell *141*, 432-445.

Ramazi, S., and Zahiri, J. (2021). Posttranslational modifications in proteins: resources, tools and prediction methods. Database (Oxford) *2021*.

Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W., and Lewis, B.A. (2012). Evidence of the Involvement of O-GlcNAc-modified Human RNA Polymerase II CTD in Transcription in Vitro and in Vivo. Journal of Biological Chemistry *287*, 23549-23561.

Rebollo, R., Galvão-Ferrarini, M., Gagnier, L., Zhang, Y., Ferraj, A., Beck, C.R., Lorincz, M.C., and Mager, D.L. (2020). Inter-Strain Epigenomic Profiling Reveals a Candidate IAP Master Copy in C3H Mice. Viruses *12*.

Reily, C., Stewart, T.J., Renfrow, M.B., and Novak, J. (2019). Glycosylation in health and disease. Nature Reviews Nephrology *15*, 346-366.

Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. Genes & Development *23*, 1247-1269.

Rossant, J., and Tam, P.P.L. (2009). Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development *136*, 701-713.

Ruan, H.-B., Singh, J.P., Li, M.-D., Wu, J., and Yang, X. (2013). Cracking the O-GlcNAc code in metabolism. Trends in Endocrinology & Metabolism *24*, 301-309.

Sadic, D., Schmidt, K., Groh, S., Kondofersky, I., Ellwart, J., Fuchs, C., Theis, F.J., and Schotta, G. (2015). Atrx promotes heterochromatin formation at retrotransposons. EMBO reports *16*, 836-850.

Sanford, J.P., Clark, H.J., Chapman, V.M., and Rossant, J. (1987). Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. Genes & Development *1*, 1039-1046.

Sayda M. Elbashir, J.H., Winfried Lendeckel, Abdullah Yalcin, Klaus Weber & Thomas Tuschl (2001). Duplexes of 21±nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature, 494–498.

Scherl, A., Couté, Y., Déon, C., Callé, A., Kindbeiter, K., Sanchez, J.-C., Greco, A., Hochstrasser, D., and Diaz, J.-J. (2002). Functional Proteomic Analysis of Human Nucleolus. Molecular Biology of the Cell *13*, 4100-4109.

Sekine, O., Love, D.C., Rubenstein, D.S., and Hanover, J.A. (2010). Blocking O-Linked GlcNAc Cycling in Drosophila Insulin-producing Cells Perturbs Glucose-Insulin Homeostasis. Journal of Biological Chemistry *285*, 38684-38691.

Shafi, R., Iyer, S.P.N., Ellies, L.G., O'Donnell, N., Marek, K.W., Chui, D., Hart, G.W., and Marth, J.D. (2000). The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. Proceedings of the National Academy of Sciences *97*, 5735-5739.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. eLife *6*.

Skvortsova, K., Bertrand, S., Voronov, D., Duckett, P.E., Ross, S.E., Magri, M.S., Maeso, I., Weatheritt, R.J., Gómez Skarmeta, J.L., Arnone, M.I.*, et al.* (2022). Active DNA demethylation of developmental cis-regulatory regions predates vertebrate origins. Science Advances *8*, eabn2258.

Slawson, C., Copeland, R.J., and Hart, G.W. (2010). O-GlcNAc signaling: a metabolic link between diabetes and cancer? Trends in Biochemical Sciences *35*, 547-555.

Slawson, C., Lakshmanan, T., Knapp, S., and Hart, G.W. (2008). A Mitotic GlcNAcylation/Phosphorylation Signaling Complex Alters the Posttranslational State of the Cytoskeletal Protein Vimentin. Molecular Biology of the Cell *19*, 4130-4140.

Sodi, V.L., Khaku, S., Krutilina, R., Schwab, L.P., Vocadlo, D.J., Seagroves, T.N., and Reginato, M.J. (2015). mTOR/MYC Axis Regulates O-GlcNAc Transferase Expression and O-GlcNAcylation in Breast Cancer. Molecular Cancer Research *13*, 923-933.

Soshnev, A.A., Josefowicz, S.Z., and Allis, C.D. (2016). Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome. Mol Cell *62*, 681-694.

Stults, D.M., Killen, M.W., Pierce, H.H., and Pierce, A.J. (2008). Genomic architecture and inheritance of human ribosomal RNA gene clusters. Genome Research *18*, 13-18.

Sun, X., Zhang, Y., Cho, H., Rickert, P., Lees, E., Lane, W., and Reinberg, D. (1998). NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. Mol Cell *2*, 213-222.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P.*, et al.* (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res *47*, D607-D613.

T. Das, A., Tenenbaum, L., and Berkhout, B. (2016). Tet-On Systems For Doxycycline-inducible Gene Expression. Current Gene Therapy *16*, 156-167.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science *324*, 930-935.

Tamburri, S., Lavarone, E., Fernández-Pérez, D., Conway, E., Zanotti, M., Manganaro, D., and Pasini, D. (2020). Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. Molecular Cell *77*, 840-856.e845.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nature Structural & Molecular Biology *14*, 1025-1040.

Thomas, J.F., Valencia-Sánchez, M.I., Tamburri, S., Gloor, S.L., Rustichelli, S., Godínez-López, V., De Ioannes, P., Lee, R., Abini-Agbomson, S., Gretarsson, K.*, et al.* (2023).

Structural basis of histone H2A lysine 119 deubiquitination by Polycomb Repressive Deubiquitinase BAP1/ASXL1 (bioRxiv: Cold Spring Harbor Laboratory).

Torres, C.R., and Hart, G.W. (1984). Topography and polypeptide distribution of terminal N-Acetylglucosamine residues on the surfaces of intact lymphocy. J Biol Chem *259*, 3308-3317.

Trapannone, R., Rafie, K., and van Aalten, D.M. (2016). O-GlcNAc transferase inhibitors: current tools and future challenges. Biochem Soc Trans *44*, 88-93.

Trumtel, S., Léger-Silvestre, I., Gleizes, P.-E., Teulières, F., and Gas, N. (2000). Assembly and Functional Organization of the Nucleolus: Ultrastructural Analysis of Saccharomyces cerevisiae Mutants. Molecular Biology of the Cell *11*, 2175-2189.

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E.*, et al.* (2013). Dynamic DNA methylation across diverse human cell lines and tissues. Genome Research *23*, 555-567.

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K.*, et al.* (2013). Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. Molecular Cell *49*, 645-656.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nature Genetics *20*, 116-117.

Wang, S., Huang, X., Sun, D., Xin, X., Pan, Q., Peng, S., Liang, Z., Luo, C., Yang, Y., Jiang, H.*, et al.* (2012). Extensive Crosstalk between O-GlcNAcylation and Phosphorylation Regulates Akt Signaling. PLoS ONE *7*, e37427.

Wang, X., and Proud, C.G. (2006). The mTOR Pathway in the Control of Protein Synthesis. Physiology *21*, 362-369.

Wang, X., Rosikiewicz, W., Sedkov, Y., Martinez, T., Hansen, B.S., Schreiner, P., Christensen, J., Xu, B., Pruett-Miller, S.M., Helin, K.*, et al.* (2022a). PROSER1 mediates TET2 O-GlcNAcylation to regulate DNA demethylation on UTX-dependent enhancers and CpG islands. Life Science Alliance *5*, e202101228.

Wang, Z., Chivu, A.G., Choate, L.A., Rice, E.J., Miller, D.C., Chu, T., Chou, S.-P., Kingsley, N.B., Petersen, J.L., Finno, C.J.*, et al.* (2022b). Prediction of histone post-translational modification patterns based on nascent transcription data. Nature Genetics *54*, 295-305.

Wang, Z., Gucek, M., and Hart, G.W. (2008). Cross-talk between GlcNAcylation and phosphorylation: Site-specific phosphorylation dynamics in response to globally elevated O-GlcNAc. Proceedings of the National Academy of Sciences *105*, 13793-13798.

Wang, Z., Pandey, A., and Hart, G.W. (2007). Dynamic interplay between O-linked N-acetylglucosaminylation and glycogen synthase kinase-3-dependent phosphorylation. Mol Cell Proteomics *6*, 1365-1379.

Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. Trends in Biochemical Sciences *24*, 437-440.

Webster, D.M., Teo, C.F., Sun, Y., Wloga, D., Gay, S., Klonowski, K.D., Wells, L., and Dougan, S.T. (2009). O-GlcNAc modifications regulate cell survival and epiboly during zebrafish development. BMC Developmental Biology *9*.

Wells, L., Whelan, S.A., and Hart, G.W. (2003). O-GlcNAc: a regulatory post-translational modification. Biochemical and Biophysical Research Communications *302*, 435-441.

Whitehead, K.A., Langer, R., and Anderson, D.G. (2009). Knocking down barriers: advances in siRNA delivery. Nat Rev Drug Discov *8*, 129-138.

Willems, A.P., Gundogdu, M., Kempers, M.J.E., Giltay, J.C., Pfundt, R., Elferink, M., Loza, B.F., Fuijkschot, J., Ferenbach, A.T., Van Gassen, K.L.I*., et al.* (2017). Mutations in N-acetylglucosamine (O-GlcNAc) transferase in patients with X-linked intellectual disability. Journal of Biological Chemistry *292*, 12621-12631.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343-348.

Wilson, A.C., LaMarco, K., Peterson, M.G., and Herr, W. (1993). The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein. Cell *74*, 115-125.

Wright, C.S. (1987). Refinement of the crystal structure of wheat germ agglutinin isolectin 2 at 1.8Åresolution. Journal of Molecular Biology *194*, 501-529.

Wright, C.S. (1992). Crystal structure of a wheat germ agglutinin/glycophorin-sialoglycopeptide receptor complex. Structural basis for cooperative lectin-cell binding. Journal of Biological Chemistry *267*, 14345-14352.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Eve Sun, Y., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389-393.

Wulff-Fuentes, E., Berendt, R.R., Massman, L., Danner, L., Malard, F., Vora, J., Kahsay, R., and Olivier-Van Stichelen, S. (2021). The human O-GlcNAcome database and meta-analysis. Scientific Data *8*.

Wysocka, J., and Herr, W. (2003). The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. Trends Biochem Sci *28*, 294-304.

Wysocka, J., Myers, M.P., Laherty, C.D., Eisenman, R.N., and Herr, W. (2003). Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. Genes & Development *17*, 896-911.

Xu, C., Liu, K., Lei, M., Yang, A., Li, Y., Hughes, T.R., and Min, J. (2018). DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. Structure *26*, 85-95.e83.

Xu, S., Tong, M., Suttapitugsakul, S., and Wu, R. (2022). Spatial and temporal proteomics reveals the distinct distributions and dynamics of O-GlcNAcylated proteins. Cell Reports *39*, 110946.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S.*, et al.* (2011). Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. Molecular Cell *42*, 451-464.

Yang, W.H., Kim, J.E., Nam, H.W., Ju, J.W., Kim, H.S., Kim, Y.S., and Cho, J.W. (2006). Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. Nature Cell Biology *8*, 1074-1083.

Yang, X., and Qian, K. (2017). Protein O-GlcNAcylation: emerging mechanisms and functions. Nature Reviews Molecular Cell Biology *18*, 452-465.

Yesbolatova, A., Natsume, T., Hayashi, K.-i., and Kanemaki, M.T. (2019). Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole. Methods *164-165*, 73-80.

Yonath, A., and Franceschi, F. (1998). Functional universality and evolutionary diversity: insights from the structure of the ribosome. Structure *6*, 679-684.

Yoo, T.Y., and Mitchison, T.J. (2021). O-GlcNAc modification of nuclear pore complexes accelerates bidirectional transport. Journal of Cell Biology *220*.

Zeidan, Q., Wang, Z., De Maio, A., and Hart, G.W. (2010). O-GlcNAc cycling enzymes associate with the translational machinery and modify core ribosomal proteins. Mol Biol Cell *21*, 1922-1936.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. Science *328*, 916-919.

Zhang, H., Cao, X., Tang, M., Zhong, G., Si, Y., Li, H., Zhu, F., Liao, Q., Li, L., Zhao, J.*, et al.* (2021a). A subcellular map of the human kinome. eLife *10*.

Zhang, S., Ubelmesser, N., Josipovic, N., Forte, G., Slotman, J.A., Chiang, M., Gothe, H.J., Gusmao, E.G., Becker, C., Altmuller, J.*, et al.* (2021b). RNA polymerase II is required for spatial chromatin reorganization following exit from mitosis. Sci Adv *7*, eabg8205.

Zhao, L., Feng, Z., Yang, X., and Liu, J. (2016). The regulatory roles of O-GlcNAcylation in mitochondrial homeostasis and metabolic syndrome. Free Radical Research *50*, 1080-1088.

Zheng, X.-y., and Gehring, M. (2019). Low-input chromatin profiling in Arabidopsis endosperm using CUT&RUN. Plant Reproduction *32*, 63-75.

Zhu, Y., Liu, T.-W., Cecioni, S., Eskandari, R., Zandberg, W.F., and Vocadlo, D.J. (2015). O-GlcNAc occurs cotranslationally to stabilize nascent polypeptide chains. Nature Chemical Biology *11*, 319-325.

Żylicz, J.J., Bousard, A., Žumer, K., Dossin, F., Mohammad, E., da Rocha, S.T., Schwalb, B., Syx, L., Dingli, F., Loew, D.*, et al.* (2019). The Implication of Early Chromatin Changes in X Chromosome Inactivation. Cell *176*, 182-197.e123.