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. Sara Formichetti

born in Rome, Italy

Oral examination: 26.09.23

The role of a nuclear glycosylation in the early mammalian embryo

Referees: Prof. Dr. Ingrid Lohmann

Dr. Jamie Hackett

Acknowledgements

What a journey. If I think about the person I was five years ago and the person I am now, I realize that this period made an incredible impact on me, both as a scientist and as a person. I know this is an overheard sentence, but I really mean it. In these five years, I experienced so much excitement - not only for my scientific results, but for the scientific results of colleagues, for seminars, conversations, experiences - and so many challenges. Some challenges were exciting as well, others led me to moments of great struggle, even long ones, when I doubted I would have continued this PhD. However, I am thankful also to those, because they made me stronger. Most importantly, I could face these difficult moments thanks to the people I have had around.

Indeed, the best part of this journey were the people I was so lucky to meet at EMBL Rome, which is a very special place. I would like to thank everyone who teached me something, or sparked a brainstorming, or listened to me, or even "just" made me a little happier during intense days. But five years is a lot of time, thus I will be able to name only a few of you in this restricted space.

Firstly, I thank my reviewers. Ingrid, for trusting me when accepting to dig into the field of O-GlcNAc and mammalian development and in advance for your time spent among these pages. Michaela, for her enthusiasm in being part of this final run. Alexander, for the inspiring discussions around biophysics and for your pure passion for interesting biological questions. Jamie, thank you for listening and for your extremely useful feedback whenever I needed an impartial and brilliant eye to judge scientific decisions, along my whole PhD.

I am extremely thankful to my supervisor, Matthieu. In my fifth year, I cannot feel luckier to have such a supportive mentor. Thank you Matthieu for always trusting me, for always being available to listen to my (even long) flow of potential ideas and genuinely happy to discuss with me. Thank you for believing in me when I was not, it was essential to overcome specific challenges. Importantly, thank you for making me feel I could be honest also in more difficult conversations. Last but not least, I truly appreciated your support in these last months of searching for a postdoc lab, the hours you have been spending brainstorming with me about the

best options for my career.
In the last couple of years, I was particularly lucky to have a second (informal) exceptional supervisor, Ana. Dear Ana, what you gave me the opportunity to learn from you and the extraordinary Na around embryos, together with the passion you have for your research, changed my scientific path. Thank you for always being encouraging but also for challenging questionable hypotheses. Thanks to you as well for always being happy to discuss with me! And thank you for the opportunity to work with Urvashi, that talented big-hearted student. Urvashi, I am grateful I could meet you as a person and team up with you in our embryonic O-GlcNAc adventure.

Then, my lab. Aga, if I grew professionally during these years, you had a big contribution. Moreover, thank you for everything you teached me and for all our cross-implantation collaborations. It was great to work with you. Sofia, I cannot imagine this journey without my O-GlcNAcylated mate with whom sharing complaints and excitements (e.g. do the moments in front of the WB imaging machine remind you of something?). Thank you for your willingness to help in the lab and for being there, one chair away, to listen to me. Monika, it was me and you five years ago, at the start. Thanks for driving away with your irony my fear of making a mess in the wet lab, and thanks for sharing emotions with me at jazz concerts. Katell, thanks for your empathy in understanding my weak point - the bench - and not only encouraging me, but also providing wise hands when mine were not enough. Francesco, I wish you never change in your willingness to help and teach, because it is truly special. I am grateful for your presence in the lab, with your excitement for DevOps but (even if you don't admit it) also so much for biology. Basilia, your reassuring smile and your wisdom is bringing a real plus to our team. Aileen, it is stimulating to see your fresh enthusiasm, also to be in my country. Pino, I include you in the lab, not only because without your daily dedicated work to my numerous mice I would not be at this point, but also for caring about teaching me not to fear any aspect of the mouse work - together with the rest of the exceptional group of EMBL Rome's animal caretakers. Thank you to the two other students that I was lucky to tutor: Alya, for your dedication and resilience in a very challenging moment of the project; Julia, for challenging me hence helping me learn how to supervise.

These years at EMBL Rome gave me friendships. Valentina, Adriana, Alessandro, Irene, Marzia, Cristina, it has been special to know I have you crazy group of people around to both discuss rigorous science but especially enjoy life outside science! Bobby and Jasmina, you

made me more joyful to come to the institute this last year. Bobby, your positivity is contagious. Yasmin, I will not say you are my Louise because I want us to live a wonderful long life, but that is the concept. Knowing you are there, to care about me, to talk and laugh and share our selves has been bringing me happiness since Monterotondo times.

These five years 100% confirmed the feeling I had when I accepted the position: EMBL is exceptional. It gave me the opportunity to organize international seminars, meet great scientists for lunch, take inspiring courses and talk to inspiring people (including my fellows!), both in a more formal setting but also in front of beer(s). It gave me incredible support both to grow personally and scientifically but also in my work, thanks to the facilities and especially the people running them, like Alvaro, Neil, Vladimir. Finally, I thank Tommaso and Rossana, because the experiences with high school students they organize with so much passion were one of the best parts of my PhD.

As in scientific literature, the end contains the remarkable statement. I thank the people that are the structure to my wings: my family. Nonno, sei il mio modello di pazienza. I could never imagine a stronger support than the one I have from you, Fabiola Giorgio e Giada. If I fly, it is because you give me stability, and teached me not to be afraid of anything.

Abstract

O-GlcNAcylation (O-GlcNAc) is a post-translational modification found on serine and threonine residues of a variety of nuclear and cytosolic mammalian proteins. In spite of the diversity of its targets, this modification is catalyzed by one single enzyme known as O-GlcNAc transferase (OGT), which is highly conserved in animals. In mammals, a maternal functional copy of the *Ogt* gene is essential for embryonic development to postimplantation stages, indicating an important function of OGT in mammalian early development. An ever-increasing number of *in vitro* studies report functions of OGT and O-GlcNAc on many cellular pathways, spanning from the cell cycle to the regulation of transcription. Nuclear factors reported to bear O-GlcNAc moieties include RNA Polymerase II and the transcription factors OCT4 and SOX2, which are master regulators of pluripotency. These findings, together with the early lethality caused by the lack of maternal OGT, raise the intriguing possibility that OGT and O-GlcNAc might be involved in transcriptional regulation during early mammalian development. Furthermore, the donor substrate for O-GlcNAcylation is UDP-GlcNAc, the end product of one of the metabolic routes of intracellular glucose. This could in principle make O-GlcNAc levels responsive to changes in the environmental conditions or intracellular metabolic demands, hence a potential mediator of cellular adaptation to such changes.

Due to the difficulty of disrupting mammalian *Ogt* using classical genetic approaches, these intriguing hypotheses have never been tested *in vivo.* With my study, I probed the effect of O-GlcNAc perturbation on the developing mammalian early embryo for the first time. To this end, I developed three different functional strategies using the mouse as a mammalian model organism.

The first method is a degron system for a fast and inducible degradation of the endogenous OGT protein. This approach proved to be very efficacious in primary mouse embryonic fibroblasts (MEFs) *in vitro*, thus I used it to get a first insight into OGT-mediated transcriptional regulation. The transcriptome of OGT-depleted MEFs showed a low-magnitude but widespread change in gene expression, including a significant downregulation of the mesenchymal differentiation pathway and genes associated with translation. Nonetheless, the *Ogt*-degron system was very poorly effective when applied to the preimplantation embryo grown *ex vivo,* based on both O-GlcNAc staining and transcriptomics analyses.

The second approach was the creation of a collection of *Ogt* hypomorphic mouse alleles, predicted to have a range of severity of phenotypes based on the catalytic activity of the corresponding OGT mutant proteins previously measured *in vitro*. I found that the severity of phenotypes of the different *Ogt* mutant mouse lines mirrored the level of *in vitro* activity reduction. In addition, unexpectedly, one of these mouse alleles showed a maternal effect phenotype: the genotype of the mother impacted the early embryo independently of the embryo's genotype.

Finally, I employed the overexpression of a recombinant OGA (the enzyme which removes O-GlcNAc) to remove the O-GlcNAc modification itself from the nucleus of the mouse zygote. By injecting the mRNA of OGA in the zygote, O-GlcNAc was reduced to undetectable levels from the early 2-cell stage, before the activation of the embryonic genome. This allowed me to specifically test the role of O-GlcNAc in genome regulation during preimplantation. I discovered that nuclear O-GlcNAc is largely indispensable for embryonic genome activation, but that O-GlcNAc depletion leads to a subtle preimplantation developmental delay visible in the morula only at the transcriptome level. When embryos developed from O-GlcNAc-depleted zygotes were re-implanted into a foster mother and dissected postimplantation, they were developmentally delayed. Based on my transcriptome data, I propose possible molecular mechanisms that could explain this phenotype. One is the effect of O-GlcNAc perturbation on the cell cycle, another is the modulation of the rate of protein synthesis by O-GlcNAc.

In sum, my PhD work brings novel insights into the biological function of the still enigmatic O-GlcNAc modification in the early mammalian embryo.

Abstrakt

Die O-GlcNAcylierung (O-GlcNAc) ist eine posttranslationale Modifikation, die an Serin- und Threoninresten einer Vielzahl von nukleären und zytosolischen Säugetierproteinen zu finden ist. Trotz der Vielfalt ihrer Ziele wird diese Modifikation durch ein einziges Enzym katalysiert, die O-GlcNAc-Transferase (OGT), die bei Tieren sehr konserviert ist. Bei Säugetieren ist eine funktionsfähige mütterliche Kopie des Ogt-Gens für die Embryonalentwicklung bis zum Postimplantationsstadium unerlässlich, was auf eine wichtige Funktion der OGT in der frühen Entwicklung von Säugetieren hinweist. Eine ständig wachsende Zahl von In-vitro-Studien berichtet über Funktionen von OGT und O-GlcNAc in vielen zellulären Prozessen, vom Zellzyklus bis zur Regulierung der Transkription. Zu den Kernfaktoren, von denen berichtet wird, dass sie O-GlcNAc-Anteile tragen, gehören die RNA-Polymerase II und die Transkriptionsfaktoren OCT4 und SOX2, die die Pluripotenz maßgeblich regulieren. Diese Ergebnisse und die frühe Letalität, die durch das Fehlen von OGT im Mutterleib verursacht wird, lassen die faszinierende Möglichkeit aufkommen, dass OGT und O-GlcNAc an der Transkriptionsregulation während der frühen Säugetierentwicklung beteiligt sein könnten. Außerdem ist das Donorsubstrat für die O-GlcNAcylierung UDP-GlcNAc, das Endprodukt eines der Stoffwechselwege der intrazellulären Glukose. Dies könnte grundsätzlich dazu führen, dass der O-GlcNAc-Gehalt auf Veränderungen der Umweltbedingungen oder des intrazellulären Stoffwechsels reagiert und somit ein potenzieller Vermittler der zellulären Anpassung an solche Veränderungen ist.
Da es schwierig ist, Ogt bei Säugetieren mit klassischen genetischen Methoden zu stören, wurden diese faszinierenden Hypothesen noch nie in vivo getestet. Mit meiner Studie habe ich zum ersten Mal die Auswirkungen einer Störung von O-GlcNAc auf den sich entwickelnden frühen Säugetierembryo untersucht. Zu diesem Zweck habe ich drei verschiedene funktionelle Strategien entwickelt, wobei ich die Maus als Säugetiermodellorganismus verwendet habe.

Die erste Methode ist ein Degron-System für einen schnellen und induzierbaren Abbau des endogenen OGT-Proteins. Dieser Ansatz erwies sich in primären embryonalen Fibroblasten der Maus (MEFs) in vitro als sehr wirksam, so dass ich ihn nutzte, um einen ersten Einblick in die OGT-vermittelte Transkriptionsregulation zu erhalten. Das Transkriptom von OGT-depletierten MEFs zeigte eine geringfügige, aber weit verbreitete Veränderung der Genexpression, einschließlich einer signifikanten Herabregulierung des mesenchymalen

Differenzierungsweges und von Genen, die mit der Translation in Verbindung stehen. Dennoch war das Ogt-Degron-System bei der Anwendung auf den ex vivo gezüchteten Präimplantationsembryo nur sehr wenig wirksam, wie sowohl die O-GlcNAc-Färbung als auch die Transkriptomik-Analysen zeigten.

Der zweite Ansatz bestand in der Schaffung einer Sammlung hypomorpher Ogt-Mausallele, für die auf der Grundlage der zuvor in vitro gemessenen katalytischen Aktivität der entsprechenden OGT-Mutantenproteine ein unterschiedlicher Schweregrad des Phänotyps vorhergesagt wurde. Ich fand heraus, dass der Schweregrad der Phänotypen der verschiedenen Ogt-Mauslinien den Grad der in vitro-Aktivitätsminderung widerspiegelte. Darüber hinaus zeigte eine dieser Mauslinien überraschenderweise einen Phänotyp mit maternalem Effekt: Der Genotyp der Mutter wirkte sich auf den frühen Embryo unabhängig vom Genotyp des Embryos aus.

Schließlich verwendete ich die Überexpression eines rekombinanten OGA (des Enzyms, das O-GlcNAc entfernt), um die O-GlcNAc-Modifikation selbst aus dem Zellkern der Maus-Zygote zu entfernen. Durch Injektion der mRNA von OGA in die Zygote wurde O-GlcNAc ab dem frühen 2-Zell-Stadium, also vor der Aktivierung des embryonalen Genoms, auf nicht nachweisbare Mengen reduziert. Dies ermöglichte es mir, die Rolle von O-GlcNAc bei der Genomregulierung während der Präimplantation gezielt zu testen. Ich entdeckte, dass das nukleäre O-GlcNAc für die Aktivierung des embryonalen Genoms weitgehend unverzichtbar ist, dass aber ein Mangel an O-GlcNAc zu einer subtilen Entwicklungsverzögerung vor der Präimplantation führt, die in der Morula nur auf der Ebene des Transkriptoms sichtbar ist. Wenn Embryonen, die aus O-GlcNAc-depletierten Zygoten entwickelt wurden, in eine Pflegemutter reimplantiert und nach der Einpflanzung seziert wurden, waren sie entwicklungsverzögert. Auf der Grundlage meiner Transkriptomdaten schlage ich mögliche molekulare Mechanismen vor, die diesen Phänotyp erklären könnten. Einer davon ist die Auswirkung der O-GlcNAc-Störung auf den Zellzyklus, ein anderer ist die Modulation der Proteinsynthese durch O-GlcNAc.
Insgesamt bringt meine Doktorarbeit neue Erkenntnisse über die biologische Funktion der noch rätselhaften O-GlcNAc-Modifikation im frühen Säugetierembryo.

Table of contents

Abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AID	auxin-inducible degron
ALL	allantois.
ANM	amnion
ARP2/3	actin-nucleating complex
Asn	asparagine
ATAC-seq	assay for transposase accessible chromatin-sequencing
ATP	adenosine triphosphate
AVE	anterior visceral endoderm
BC	blastocoele
BMP	bone morphogenetic protein
BP	Biological Process
CB	chorioallantoic branching
ChEA	ChIP enrichment analysis
CID	collisional-induced dissociation
CTD	C-terminal domain
DAPI	4',6-Diamidino-2-Phenylindole, Dihydrochloride
DEGs	differentially expressed genes
DHS	DNaseI hypersensitivity
DMEM	Dulbecco's modified Eagle medium
DNAme	DNA methylation
DUX4	double homeobox 4
DVE	distal visceral endoderm
ECM	extracellular matrix
ECT	ectoderm
EGA	embryonic genome activation
EMT	epithelial-mesenchymal transition
EOGT	EGF domain specific O-Linked N-Acetylglucosamine transferase
EPI	embryonic epiblast
EPI	epiblast
EPS	early prenatal stress
ER	endoplasmic reticulum
ERVL	endogenous retrovirus type L
ERα	estrogen receptor α
ETD	electron transfer dissociation

E[N]	embryonic day N
ExE	extraembryonic ectoderm
FGF	fibroblast growth factor
FG-Nups	FG nucleoporins
FRAP	fluorescence recovery after photobleaching
GAGs	glycosaminoglycans
GalNAc	N-acetylgalactosamine
GalT	galactosyltransferase
GEEF	Gene Editing and Embryology Facility
GlcNAc	N-acetylglucosamine
GLT	germline transmission
GREB1	growth regulator estrogen receptor 1
GSEA	gene set enrichment analysis
GSH	glutathione
GT-B	glycosyltransferase-B
H3K4me3	trimethylation of histone 3 at lysine 4
HBP	hexosamine biosynthetic pathway
HCD	higher-energy collisional dissociation
hCG	human chorionic gonadotropin
HIF1-α	hypoxia-inducible factor 1α
IACUC	Institutional Animal Care and Use Committee
ICM	inner cell mass
ICR	imprinting control region
Int-D	intervening domain
IVF	in vitro fertilization
iXCI	imprinted X chromosome inactivation
KAP1	KRAB-associated protein 1
KO	knock-out
KRAB	Krüppel-associated box
LC-MS/MS	chromatography coupled to tandem mass spectrometry
LOF	loss-of-function
LTR	long terminal repeat
MEFs	mouse embryonic fibroblasts
MEK	mitogen-activated protein kinase
MES	mesoderm
mESCs	mouse embryonic stem cells
mOGT	mitochondrial OGT
MSigDB	Molecular Signatures Database

mtDNA	mitochondrial DNA
MuERV-L	(MERVL) murine endogenous retrovirus type L
MZT	maternal-to-zygotic transition
N	node
NADPH	nicotinamide adenine dinucleotide phosphate
ncOGT	nuclear-cytoplasmic OGT
NES	normalized enrichment score
Neu5Ac	N-acetylneuraminic acid
NF	neural fold
NLBs	nucleolar-like-bodies
NPC	nuclear pore complex
NUP62	nuclear pore protein p62
OGA	O-GlcNAcase
O-GlcNAc	O-GlcNAcylation
OGT	O-GlcNAc transferase
OGT-CDG	congenital disorder of glycosylation associated to OGT
OxPhos	oxidative phosphorylation
PAC	pro amniotic cavity
PCA	principal component analysis
PcG	Polycomb group
Ph	Polyhomeotyc
PMSG	pregnant mare serum gonadotropin
PN	pronuclear
PolII	RNA Polymerase II
POU5F1	POU domain, class 5, transcription factor 1
PPP	pentose phosphate pathway
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PrE	primitive endoderm
PS	primitive streak
PSD	postsynaptic density
PTM	post-translational modification
rDNA	ribosomal DNA
rXCI	random X chromosome inactivation
scRNA-Seq	cingle cell RNA-sequencing
SEC	SECRET AGENT
SNP	single nucleotide polymorphism
sOGT	short OGT

SPY	SPINDLY
SRE	sterol regulatory element
STOP	stop codon
TE	transposable elements (ambiguity clarified in text)
TE	trophectoderm (ambiguity clarified in text)
TET enzymes	ten-eleven translocation enzymes
TGF-β	transforming growth factor beta
TPR	N-terminus tetratricopeptide repeat
UDP-GlcNAc	uridine diphosphate N-acetylglucosamine
VE	visceral endoderm
WGA	wheat germ agglutinin
WT	wild type
XCI	X chromosome inactivation
XIC	X inactivation center
ZFP57	KRAB-containing zinc-finger protein 57
ZP	zona pellucida
αKG	α-ketoglutarate
5hmC	5-hydroxymethylcytosine

Introduction

Preamble

My PhD study aims at understanding the role of the post-translational modification named O-GlcNAcylation during *in-vivo* mammalian early development, with a focus on its potential transcriptional regulatory functions. In order to guide the reader to the profound understanding of why asking this specific biological question is relevant, I divided the Introduction in four main sections. Firstly, I will put the O-GlcNAc modification in the context of the other cellular glycosylations and describe what makes it unique in the world of glycobiology. Then, I will put O-GlcNAc in the context of evolution, which suggests an important role for this modification in animal development, that evolved as essential in mammals. Next, I will start focusing on mammals. In section three I will present an overview of early development of the mouse - the model I use throughout the study. The fourth and last section of the Introduction will review the scientific literature available reporting functions of O-GlcNAc in mammalian transcriptional regulation. While these studies are mostly performed using *in vitro* systems, they already implicate key developmental players. With my study, I investigate the biology of the O-GlcNAc modification *in vivo* and at molecular level for the first time.

1) O-GlcNAc: history of a unique protein glycosylation

In the cell, proteins are decorated by a wide variety of post-translational modifications (PTMs). Some of these modifications are structural requirements for protein function, such as the ramified glycosylations which are abundant on membrane proteins. Other modifications are highly dynamic and enable the cell to respond to the environment in a specific manner and adapt to external stimuli. The best studied example of dynamic post-translational modification is phosphorylation, which is governed in the human cell by over 500 specific kinases and about 200 phosphatases (Manning et al., 2002; Sacco et al., 2012). Another example is intracellular mono-O-GlcNAcylation (henceforth O-GlcNAc). O-GlcNAc consists of a single N-acetylglucosamine (GlcNAc) moiety covalently attached in O-linkage to serine (Ser) and threonine (Thr) protein residues (Fig.1a). To date, this is the main monoglycosylation that has been found in animals on nuclear and cytosolic proteins. Before the discovery of O-GlcNacylation, glycosylation was thought to be exclusive to the extracellular compartment.

1.1 Extracellular glycans

There are two main types of protein glycosylation, depending on the linkage with the protein residue: N-glycans are covalently attached to proteins at asparagine (Asn), O-glycans are linked to the hydroxyl of Ser or Thr residues. Before O-GlcNAc was brought to light, all known O-linked and N-linked glycans were composed of many monosaccharides, linked together in a variety of combinations forming a range of branched structures. Those glycans have been consistently selected by evolution to be at the interface between the cell and the extracellular environment, probably because of their hydrophilicity and mobility. In bacteria, Archaea and fungi they have an essential function in the cell wall. In eukaryotes, they reside on the plasma membrane or on secreted proteins and function in the extracellular matrix, with roles in cell-cell communication, binding of signaling ligands and organ morphogenesis. In eukaryotes, proteins facing the extracellular space pass through the endoplasmic reticulum (ER) and Golgi apparatus, where most of the glycosyltransferases are found. Therefore, most ramified glycosylations never face the eukaryotic cytosolic space (Varki et al., 2022).

1.2 Mono-O-GlcNAcylation

O-GlcNAc was unexpectedly discovered in 1983 on the plasma membrane of mouse lymphocytes and macrophages using purified bovine milk galactosyltransferase and a radiolabeled donor substrate (UDP-[3H]galactose). This enzyme attaches galactose to almost any terminal N-acetylglucosamine residue. To the surprise of the authors of this study, downstream biochemical analysis of the resulting labeled glycans revealed that they were mostly consisting of a single O-linked GlcNAc moiety (Torres and Hart, 1984). In a follow-up study, the same group investigated the subcellular localization of O-GlcNAc in rat liver and found the modification in all organelles, including the cytosol and especially in the nucleus (Holt and Hart, 1986). Later on, other types of glycosylations have been found in these compartments, but O-GlcNAc is the only monoglycosylation known to face the cytosolic and nuclear space in animals. Some recent reports suggest that select intracellular proteins of mammalian cells are modified by N-acetylgalactosamine (GalNAc) (Varki et al., 2022), but to the best of my knowledge neither the responsible glycosyltransferases nor evidence for potential functions have been reported. The other known eukaryotic intracellular monoglycosylation is O-fucosylation, which is found on tens of nucleocytoplasmic proteins in plants, protists and maybe bacteria (Varki et al., 2022).

Since the 1980s, an increasing number of proteins have been found to bear O-GlcNAc. As of May 2023, over 5000 human proteins have been reported to be O-GlcNAcylated (Wulff-Fuentes et al., 2021a).

1.3 The O-GlcNAc cycle

Shortly after O-GlcNAc discovery, the enzymes responsible for the turnover of this modification were discovered. The enzyme which catalyzes the addition of a single GlcNAc moiety on serine and threonine protein residues is the O-GlcNAc transferase (OGT) (Kreppel et al., 1997; Lubas et al., 1997a) and the enzyme responsible for its removal is the O-GlcNAcase (OGA) (Gao et al., 2001). The donor substrate used by OGT is UDP-GlcNAc (Haltiwanger et al., 1990), the end product of a metabolic pathway called the hexosamine biosynthetic pathway (HBP), which utilizes around 2-5% of intracellular glucose (Marshall et al., 1991). OGT and OGA are responsible for the unique dynamics of O-GlcNAc compared to the low turnover of ramified glycans. In their study, Wang et al. developed a method based on the combination of metabolic labeling of UDP-GlcNAc with ^{13}C, mass spectrometry and mathematical modeling to accurately quantify the turnover rate of 20 O-GlcNAc peptides from 14 proteins extracted from HeLa cells (Wang et al., 2016). The small number of quantified peptides is due to the exclusion of all proteins with more than one O-GlcNAcylated residue (~20% of total identified O-GlcNAc proteins) and of all proteins for which they did not manage to quantify the half-life in a separate metabolic labeling experiment. They found O-GlcNAc turnover varying from 0.02 h^{-1} to 1.6 h^{-1}. Although faster compared to branched glycans, this turnover is on average quite slow if compared to the same variable obtained for protein phosphorylation (0.06 h^{-1}-2040 h^{-1}) (Molden et al., 2014) and histone acetylation (0.3-1 h^{-1}) (Evertts et al., 2013; Zheng et al., 2013).

1.4 UDP-GlcNAc: not only one destiny

OGT is not the only enzyme using UDP-GlcNAc as a donor substrate (Fig.1b).

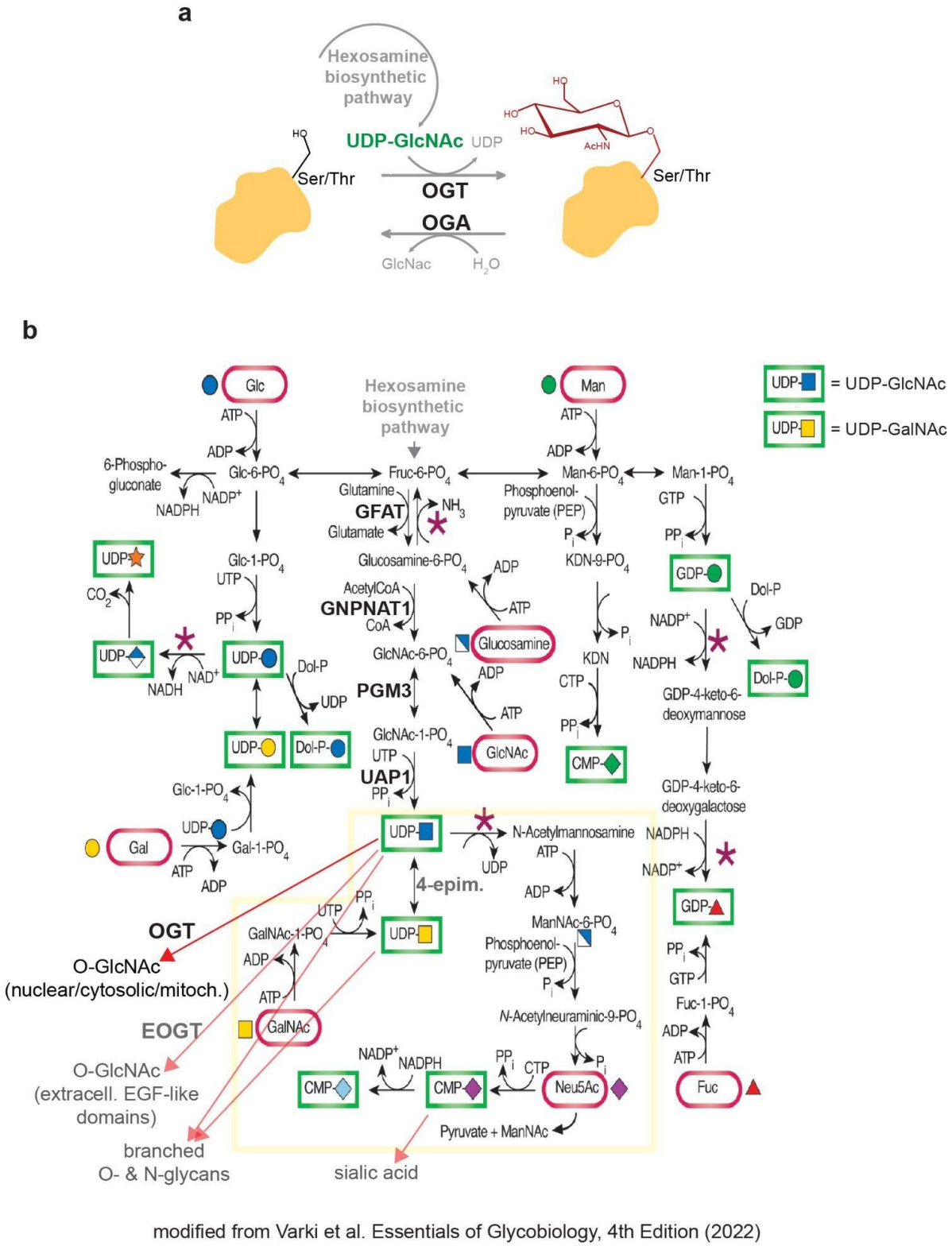
1.4.1 Incorporation to branched glycans

UDP-GlcNAc is the activated form of the modified monosaccharide GlcNAc. First of all, UDP-GlcNAc is directly incorporated as a GlcNAc moiety into branched oligosaccharides by other glycosyltransferases (Varki et al., 2022). Secondly, UDP-GlcNAc can be isomerized by enzyme 4-epimerase to UDP-GalNAc (Kingsley et al., 1986), which is the substrate for the addition of GalNAc moieties to branched O-glycans and N-glycans (Varki et al., 2022). Among others, the

LDL receptor on the cell membrane is heavily dependent on O-GalNAc for its function (Kingsley et al., 1986). Another class of branched glycans containing alternating GlcNAc and GalNAc are glycosaminoglycans (GAGs), which include hyaluronan, dermatan sulfate, keratan sulfate, chondroitin sulfate, heparin and heparan sulfate. GAGs are found on secreted proteins that are components of the extracellular matrix (ECM) of connective tissues and are also important for e.g. neurite outgrowth in the nervous system (Varki et al., 2022). Moreover, UDP-GlcNAc is an intermediate for the synthesis of N-acetylneuraminic acid (Neu5Ac, also known as sialic acid), which in vertebrates include mucins, found on mucus membranes to prevent infections and functioning as receptors for some pathogens (Varki et al., 2022).

1.4.2 OGT-independent O-GlcNAc

Many years after the discovery of OGT, two other enzymes have been found that in animals catalyze the addition of a single O-linked GlcNAc. EOGT, which has been discovered in *Drosophila,* is responsible for O-GlcNAcylation of a specific consensus sequence found on EGF-like domains of a few membrane proteins (Sakaidani et al., 2011). In contrast with OGT, which localizes in the cytosol, in the nucleus and probably in the mitochondria, EOGT localizes to the endoplasmic reticulum and only catalyzes extracellular O-GlcNAcylation. One of the main EOGT's targets in fly is a protein known as Dumpy, whose O-GlcNAcylation is required for proper cell-matrix interactions during fly embryogenesis (Sakaidani et al., 2011). EOGT is conserved in mammals, where it modifies several membrane proteins containing EGF-like domains, including components of NOTCH signaling. However, in mammalian cells EOGT-dependent O-GlcNAc seems to be in the form of elongated glycans(Ogawa and Okajima, 2019). The second OGT-independent mono-O-GlcNAcylation is found in the intracellular compartment and has been discovered very recently. It consists in the specific O-GlcNAcylation of the estrogen receptor α (ERα) by the glycosyltransferase GREB1 and leads to ERα stabilization (Shin et al., 2021). GREB1 is an early response gene in the estrogen receptor-activated pathway, thus GREB1-dependent O-GlcNAcylation is an inducible form of O-GlcNAcylation (Shin et al., 2021). To my knowledge, it is not known whether OGA is able to remove GREB1-dependent O-GlcNAc.



modified from Varki et al. Essentials of Glycobiology, 4th Edition (2022)

Figure 1: Metabolic origin and the many destinies of UDP-GlcNAc.
(a) The O-GlcNAc cycle.
(b) The hexosamine biosynthetic pathway and the different destinies of its end product UDP-GlcNAc. Adapted from (Varki et al., 2022).

From now on, when discussing about O-GlcNAc, I will always refer to the OGT-dependent mono-O-GlcNAcylation of serine and threonine residues.

1.5 Challenges of O-GlcNAc mapping

Despite its discovery in 1983, the unambiguous and comprehensive identification of cellular O-GlcNAcylated proteins lags behind. This is due to technical challenges in detecting O-GlcNAc in proteomics studies.

The first issue is the sub-stoichiometry of the O-GlcNAcylated form of most O-GlcNAc targets. This implies that a method for enriching the O-GlcNAc-modified fraction of the proteome is necessary before any mass spectrometry analysis. To this aim, several strategies have been employed. O-GlcNac metabolic labeling techniques exploit the endogenous glycan biosynthetic enzymes for the incorporation of analogues of monosaccharides containing a bioorthogonal functional group (e.g. the azide or alkyne). These unnatural monosaccharides are added to the culture media, they get incorporated during cellular biosynthesis and they can then be chemically detected using reactions which link them to desired probes such as biotin ("click chemistry") (Hao et al., 2019). This approach is limited to cell culture. A similar approach uses cell lysates as an input, by adding a recombinant galactosyltransferase (GalT1(Y298L)) to selectively label O-GlcNAcylated sites on proteins with modified monosaccharides suitable to click chemistry (Khidekel et al., 2003). A third strategy, which can identify native O-GlcNAc sites, is lectin weak affinity chromatography (Myers et al., 2011), an inexpensive method which is however difficult to apply to low-input samples. Above I described the variety of cellular O-glycosylations besides O-GlcNAc. This implies that each of these O-GlcNAc-enrichment methods has the risk of enriching for other types of glycosylations as well. Nonetheless, all these methods have contributed to the ever-increasing knowledge of the O-GlcNAc proteomes.

The second pitfall of O-GlcNAc proteomics is the labile nature of the O-GlcNAc bond compared to other PTMs such as phosphorylation. Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) techniques include a fragmentation of the primary ion for better identification of the peptide. The most typical fragmentation techniques, namely collisional-induced dissociation (CID) and higher-energy collisional dissociation (HCD), cause the loss of the GlcNAc moiety. More recently, an alternative fragmentation method called electron transfer dissociation (ETD) has been developed that specifically fragments the peptide backbone and preserves PTMs (Syka et al., 2004). Being a gentler fragmentation process than collisional methods, ETD results in fewer peptides identification (Burt et al., 2022). However,

combining ETD with CD led several groups to compile always better lists of the O-GlcNAcylated proteins in different cell types (Burt et al., 2022).

Because of the limitations described above, the identity of many O-GlcNAc proteins and of the precise modified residues of known O-GlcNAcylated proteins remains to be determined. In addition, the challenges of O-GlcNAc proteomics underline the importance of validating the O-GlcNAc site of a protein with functional experiments.

Finally, the existence of EOGT-dependent OGT-independent mono-GlcNAcylation of EGF-like domains on plasma membrane should be kept in mind in studies which aim at understanding the role of OGT through the analysis of the O-GlcNAc proteome. This observation reinforces the necessity to develop functional strategies perturbing OGT or specifically removing O-GlcNAc from nuclear and cytosolic proteins, which should not significantly affect EOGT-dependent O-GlcNAc (Sakaidani et al., 2011).

2) O-GlcNAc enzymes: a conserved role in development

OGT and OGA are the only known regulators of O-GlcNAc cycling, they are present across the whole animal kingdom and in most organisms they are encoded in one single *Ogt* and one single *Oga* gene. In animals, the only known exception is zebrafish, which contains two *Ogt* genes (Sohn and Do, 2005).

The OGT protein is highly conserved across metazoans (Lubas et al., 1997b). It is composed of two main regions: a C-terminal catalytic region and a long N-terminal tetratricopeptide repeat (TPR) region (Joiner et al., 2019) (Fig.3a,c). The catalytic domain is characterized by two Rossmann-folded catalytic lobes which classify OGT as a glycosyltransferase-B (GT-B) (Lairson et al., 2008) (Fig.3b). However, differently from other GT-B glycosyltransferases, OGT possesses a 120-amino acid insertion (termed the intervening domain (Int-D)) between its two catalytic lobes (Lazarus et al., 2011). The function for this insertion remains mysterious. Concerning the N-terminus OGT domain, each TPR repeat is a 34 amino acids helix-turn-helix motif that is known for mediating protein-protein interactions (Allan and Ratajczak, 2011). Mutations in the TPR domain of human OGT are responsible for very similar clinical features than mutations in the catalytic domain (Pravata et al., 2020).

Interestingly, higher plants have two O-glycosyltransferases: SPINDLY (SPY), which is an O-fucose transferase (Zentella et al., 2017), and SECRET AGENT (SEC), which is an O-GlcNAc transferase like OGT (Hartweck et al., 2002). SPY and SEC have multiple overlapping and specific roles in regulating plant growth and development, including transcription and the circadian clock (Wang et al., 2020). These two proteins are very similar between them and like animal OGT they are made of an N-terminus tetratricopeptide repeat (TPR) region for protein-protein interactions a C-terminal catalytic domain with a bilobed GT-B fold (Hartweck et al., 2002).

In line with its conservation, genetic studies in evolutionary distant animals have found *Ogt* to be essential for embryonic development. In fly, the lack of a functional embryonic *Ogt* gene (also known as *super sex combs* (*sxc*)) results in lethality at the pupal stage of development (Gambetta et al., 2009a; Sinclair et al., 2009); in the absence of a maternal functional *Ogt* gene, lethality already arises at late embryonic stages or early larval stages (Ingham, 1984). In mammals, a functional maternal copy of *Ogt* is essential for embryonic development postimplantation (O'Donnell et al., 2004a). The requirement of maternal *Ogt* for development in both of these organisms is noteworthy because it might imply that *Ogt* plays a critical role in

the first embryonic stages, when the embryo relies on the maternal payload of macromolecules (see section 4.1.2).

While *Ogt* is required for development in flies and in mice , differences emerge through evolution: *sxc*/*Ogt* is dispensable for cell viability in fly (Gambetta et al., 2009b), while mammalian *Ogt* is essential for the viability of mESCs (Shafi et al., 2000) and all somatic cell types analyzed so far: in T lymphocytes, loss of *Ogt* led to apoptosis, while *Ogt*-null fibroblasts failed to undergo more than three cell divisions and went to senescence and death (O'Donnell et al., 2004b).

Consistently with the lower level of conservation of *C. elegans*, OGT compared to flies and mammals, *C. elegans ogt-1*-null mutants develop into viable and morphologically normal adults. Interestingly (see section 4.7.1), their defects are related to misregulation of insulin-like signaling, stress and aging pathways (Dona C Love et al., 2010).

The lack of OGA has less disruptive consequences both in flies and mice, yet it affects development. Flies lacking OGA are fully viable and fertile, but present larger adult size, a semi-penetrant oogenesis defect, mild neuronal and behavioral phenotypes and increased intestinal stem cell proliferation (Czajewski and Aalten, 2023). In mice, the lack of a catalytically functional OGA leads to pre/perinatal lethality (Muha et al., 2021; Yang et al., 2012), reduced intrauterine growth already at embryonic day (E) 14.5 (Yang et al., 2012) and defects in kidney, brain, liver and stomach at E18.5 (Muha et al., 2021). *Oga*-null pups had a reduced glucose blood level and glycogen storage (Keembiyehetty et al., 2015).

3) Mouse embryonic development

Murine gestation period lasts for about 20 days. The moment when the embryo implants into the uterus of the mother (embryonic day 4.5 in the mouse) can be used to divide this period into preimplantation and postimplantation development.

Disclaimer

I wish to give my apologies in advance for this introduction being far from an exhaustive description of mouse embryogenesis. *Ogt*-null mice die around the time of implantation and the experiments presented in this study span mammalian development from fertilization to embryonic day 7 (E7) post-fertilization. For this reason, I will concentrate my introduction on this developmental window (Fig.2a), with a focus on transcriptional and epigenetic dynamics, because these are the processes for which I mostly investigate a potential role of O-GlcNAc.

3.1 Overview

3.1.1 Preimplantation development

Murine embryonic development begins when the haploid oocyte is fertilized by the haploid sperm. During its very first days of life, the embryo will be sustained by the biomolecules accumulated in the oocyte during its maturation. In females, the first phases of oogenesis are completed before birth. At birth, the oocytes stored in the ovary have already undergone meiotic DNA replication and recombination and are arrested in meiotic prophase. Periodically, some of these initiate a prolonged growth period when they accumulate all RNA and protein molecules necessary to sustain the first embryonic stages, also thanks to the exchange of macromolecular precursors with the surrounding somatic cells. Once every estrus cycle, a peak in gonadotropins triggers a release from the meiotic block: the oocyte nucleus breaks down, the meiotic spindle forms and half of the chromosomes are segregated into a small non-functional cell called polar body. The remaining chromosomes are captured by a second spindle and the oocyte stays arrested at this phase (meiosis II, hence MII oocyte) until the time of fertilization. This last step of maturation takes around 12 h in mice. In the lab, it can be artificially triggered with PMSG and hCG hormonal injections of the female 62 h and 14 h before the planned time of fertilization, respectively. When the sperm binds to the zona pellucida (ZP) - a glycoprotein matrix that surrounds the egg - and its membrane fuses with the egg's plasma membrane, a signaling cascade starts. This ultimately leads to a rise in free Ca^{2+}, which induces the

completion of meiosis II of the oocyte and the release of a second polar body (Clift and Schuh, 2013). Then, embryonic development starts (Fig.2b). Upon fertilization, the nuclear envelope of the sperm is rapidly degraded and nuclear membranes deriving from maternal material contribute to the formation of the two "pronuclei" containing the two separate parental haploid genomes (Usui et al., 1997). In the next ~12 h (in the mouse) the two pronuclei grow in diameter and migrate towards each other thanks to actin nucleation from the cell membrane and dynein-dependent movement along microtubules (Scheffler et al., 2021). This period is usually classified into 5 pronuclear (PN) stages, based on the size and relative location of the pronuclei respectively to each other (Fig.2c). Upon nuclear envelope breakdown, the two parental genomes undergo the first mitosis and a 2-cell stage embryo is formed. The latter is a particularly long stage (~ 24 h) when the crucial event of embryonic genome activation (EGA, to which all section 3.2 will be dedicated) takes place. In parallel, starting from the zygote, maternal transcripts are gradually degraded. At the transcriptome level, this is visible as a gradual exchange of maternal mRNAs with embryonic ones from the early 2-cell to the late 2-cell embryo (Svoboda, 2018). The next mitotic divisions - asynchronous among blastomeres in mice - are faster (~12 h each cycle) and they are not accompanied by embryonic growth, thus they are called cleavages, which result in smaller single blastomeres. At the 8-cell stage, the blastomeres start to adhere to each other and undergo compaction into the so-called morula. During the two subsequent cleavages, two distinct tissues emerge for the first time in the embryo: an external layer of cells which will give rise to the extraembryonic tissues, the trophectoderm (TE), and an internal layer of cells called the inner cell mass (ICM), which will give rise to the embryo proper. Concomitantly, from the 32-cell stage, a fluid-filled cavity (blastocoele) begins to form inside the embryo, which grows as the embryo matures into a blastocyst. In addition, at the mid-to-late blastocyst stage, cleavage ceases (Aiken et al., 2004) as cells start to replenish their cytosol and organelles before division. Furthermore, the cells of the ICM rapidly diverge into the epiblast and the extraembryonic hypoblast (also called primitive endoderm) (Schrode et al., 2014). Importantly, from the ICM onwards, the embryonic cells are no longer totipotent, because they cannot contribute to the extraembryonic lineages. The cells of the ICM and the blastocyst epiblast are called pluripotent as they can give rise to the three primary embryonic germ layers, namely ectoderm, endoderm and mesoderm.

By the time the blastocyst forms, the embryo has migrated into the uterus. The blastocyst hatches from the zona pellucida and implants into the uterine wall, which marks the beginning of postimplantation development.

3.1.2 Postimplantation development

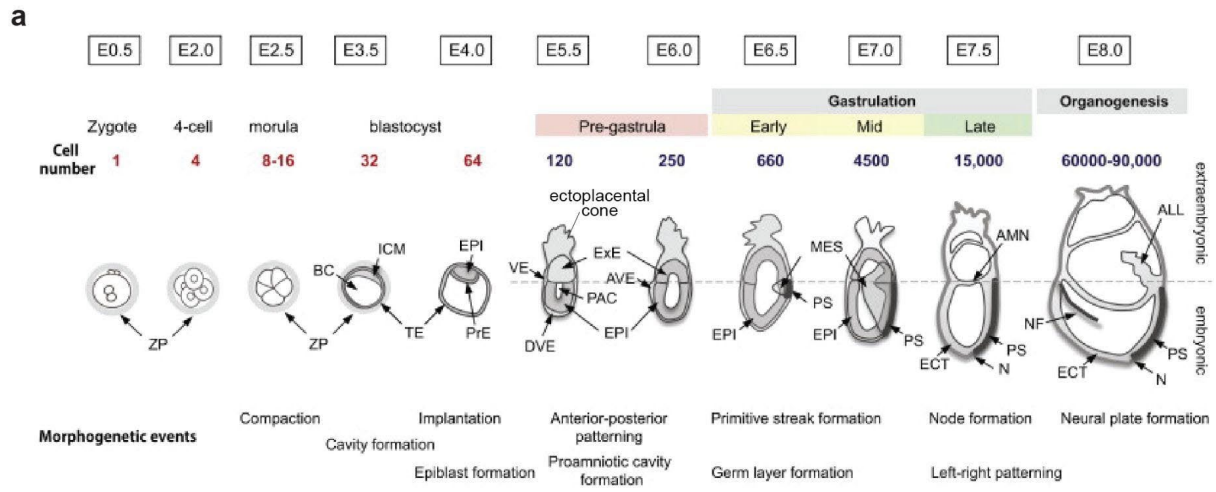
Within one day from implantation, the blastocyst has developed into an elongated structure that is made up of the embryonic epiblast (EPI), deriving from the blastocyst epiblast, and the extraembryonic tissues. At E5, the latter consists of the extraembryonic ectoderm (ExE), which derives from the trophectoderm and is connected to the uterus through the ectoplacental cone, and the visceral endoderm (VE), deriving from the primitive endoderm.

Around E6 the embryo starts gastrulation, a process that refers to the formation of the three primary germ layers: ectoderm, mesoderm and endoderm. Gastrulation initiates with local cellular ingression in a posterior region of the epiblast, called the primitive streak (PS), from which the mesoderm and endoderm will form (Fig.2d). At the site of ingression, the epiblast cells undergo epithelial-mesenchymal transition (EMT), consisting at the cellular level of reconfiguration of the cytoskeleton and resulting in changes in cell shape, deconstruction of the intercellular junctional complexes and formation of microfilament-mediated cell-matrix interactions. At the molecular level, among other events, this involves downregulation of E-cadherin expression and Rho activity (Tam and Loebel, 2007).

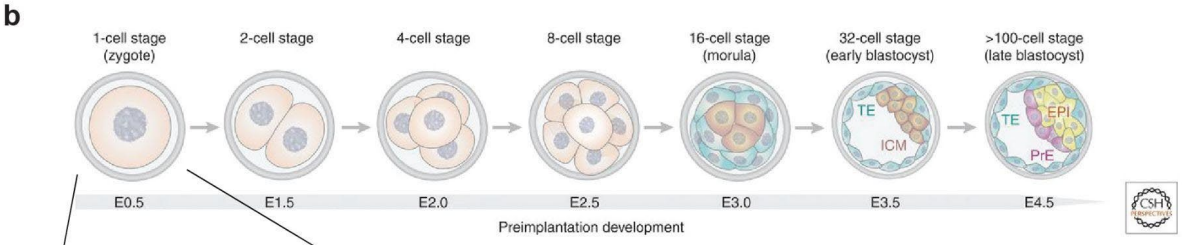
Embryos that show signs of initiation of gastrulation (which are found at E6.5-E7) - formation of the primitive streak and emergence of nascent mesoderm - contain an average of 1000 cells. Studies where the number of preimplantation blastomeres was experimentally manipulated, leading to a reduced number of ICM cells, suggest that a certain number of cells is required to initiate gastrulation, hence smaller embryos have a delay in gastrulation onset (Kojima et al., 2014).

Over the past years, models for *in vitro* the study of gastrulation have been developed, which start from derivatives of the blastocyst (embryonic stem cells or ESCs, described in section 3.4) that are induced to differentiate. These models showed that cell fate specification can in principle happen independently of the specification of the body axes. *In vitro*, exposure to the BMP, Nodal and Wnt signaling pathways is seemingly sufficient to lead to the formation of the three primary germ layers in 2D structures also called as 2D gastruloids. The same pathways are involved in mammalian gastrulation *in vivo*, but in this setting the constant communication between the embryonic and extraembryonic tissues at each step results in a net of feedback mechanisms between pathway effectors, which gives rise to cell specification in parallel with the establishment of embryonic polarity and tissue patterning (Sheng et al., 2021).

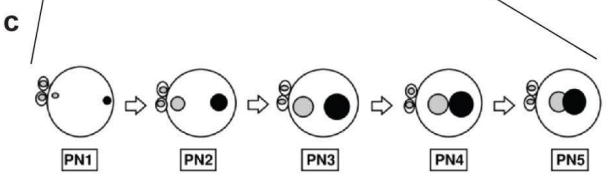
Another concept in embryonic patterning is that asymmetry in gene expression always arises before any tissue morphogenetic event is observed. Mixed populations of *Nanog* and *Gata6* expressing cells are already observed in the blastocyst ICM; *Gata6* is required for primitive endoderm formation (Schrode et al., 2014). In the elongated implanted embryo before gastrulation, the regionalization of the visceral endoderm orchestrated by Wnt and Nodal signaling precedes and is necessary for the emergence of the primitive streak at the right location. Then, cellular migration starting with primitive streak formation is a combination of active cell-autonomous migration and the bulk movement of the tissue. Active migration is undertaken both by cells of the visceral endoderm and by the cells which form the PS itself. In the VE, WASF1 - among other factors - mediates changes in the cytoskeleton by activating the actin-nucleating complex (ARP2/3). In the PS cells, fibroblast growth factor (FGF) signaling - *Fgf1* and *Fgf8* in particular - are required for epithelial-to-mesenchymal transition and migration. Then, various levels of Wnt and Nodal signaling experienced along the PS are responsible for the formation of different mesoderm and endoderm structures (Tam and Loebel, 2007).



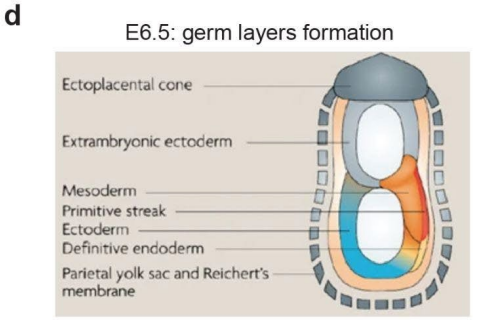
from Kojima et al. Seminars in Cell & Dev Biol (2014)



from White and Platcha Cold Spring Harb. Perspect. Biol. (2020)



from Adenot et al. Development (1997)



from Tam et al. Nature Reviews Genetics (2007)

Figure 2: Mouse early embryonic development.

(a) Overview of mouse embryogenesis, adapted from (Kojima et al., 2014). ZP: zona pellucida, ICM: inner cell mass, TE: trophectoderm, BC: blastocoele, EPI: epiblast, PrE: primitive endoderm, ExE: extraembryonic ectoderm, VE: visceral endoderm, DVE: distal VE, AVE: anterior VE, PAC: pro amniotic cavity, PS: primitive streak, MES: mesoderm, ECT: ectoderm, N: node, ANM: amnion, NF: neural fold, ALL: allantois.

(b) More detailed schematic of preimplantation stages, adapted from (Kojima et al., 2014; White and Plachta, 2019).

(c) Relative distance and diameter of the zygotic pronuclei according to the definition of pronuclear (PN) stages which is used in this study. Adapted from (Adenot et al., 1997). (d) More detailed schematic of the E6.5 embryo at the moment of formation of the three germ layers, adapted from (Tam and Loebel, 2007).

3.2 Embryonic genome activation

During the course of preimplantation development, the embryo takes charge of its own gene expression. This process, named maternal-to-embryonic transition (or maternal-to-zygotic transition, MZT), consists of a series of events mechanistically coordinated, which include the degradation of the maternal transcripts, the activation of the embryonic genome and the complete reorganization of the nuclear structure. All of this is additionally coordinated with DNA synthesis and cell division.

In vertebrates, the timing of embryonic genome activation (EGA) differs between fast-developing species such as frogs and zebrafish, characterized by very fast cell divisions following fertilization, and slow-dividing species, such as mammals, which evolved a slow initial rate of cleavage (Jukam et al., 2017). Based on the organisms for which this information is available, such a distinction seems to coincide with the split between anamniotes and amniotes, in other words with the possibility to grow the embryo while protected by the mother. In frogs and zebrafish, EGA starts after the embryo has already undergone many cell divisions, at the mid-blastula transition. In mice and humans, the major wave of EGA is observed at the 2- and 8-cell stage, respectively. The molecular mechanism defining EGA timing might not be conserved among fast-developing and slow-developing species. In frogs and fish, the fast increase in nuclear/cytosolic ratio (N/C) could be at least partially responsible for the activation of a bulk of genes. The N/C increase would titrate the concentration of maternal repressors of genome activation, such as histone proteins (Jukam et al., 2017; Newport and Kirschner, 1982; Prioleau et al., 1994). In mammals, this model is attractive as well, although reducing the cytosol of 1-cell embryos did not change the timing of RNA synthesis (Lee et al., 2001). Yet, we cannot exclude that a more sensitive technique could unveil that the N/C regulate at least a subset of mammalian genes. A non-exclusive mechanism for the start of EGA is the accumulation of a “pioneer” transcription factor (TF) targeting the regulatory loci of early EGA genes. A master regulator of EGA has been found for the invertebrate *Drosophila* embryos, where the TF named Zelda is necessary for the activation of essential early EGA genes (Liang et al., 2008) and capable of displacing the nucleosome barrier at EGA enhancers (Sun et al., 2015). Zelda has no clear ortholog in fish, frogs and mammals. In zebrafish, the transcription factors Nanog, SoxB1 and Pou5f3 (Oct4) are required to initiate 75% of the major phase of zygotic gene activation (Lee et al., 2013). Nanog and Oct4 are conserved in mammals, but they

do not seem to be required for mouse EGA (Wu et al., 2013), where in fact they are known to have an essential role in later stages, namely at pluripotency (see section 3.4).

Both in mouse and human, a remarkable feature of EGA (2- and 8-cell embryo, respectively) is the expression of the species-specific type of endogenous retrovirus type L (ERVL) transposable elements (Hendrickson et al., 2017). ERVL is a class of retrotransposons which contain Long Terminal Repeat (LTR) sequences at their promoter. Retrotransposons are genomic elements that, when expressed at the RNA level, can be retrotranscribed to DNA and then be integrated at new genomic loci. Because of their autonomous mode of replication, transposable elements have proliferated in genomes of sexual species, representing roughly 40% of the mammalian DNA sequence. The jumps of transposable elements to new genomic locations can threaten genome integrity, therefore the mammalian cell keeps these elements mostly silenced. However, transposable elements can also be a source of evolutionary innovations, for example by providing new regulatory sequences (Franke et al., 2017; Senft and Macfarlan, 2021). For example, LTRs were found enriched among species-specific enhancers in human placenta and mouse trophoblast stem cells (Chuong et al., 2013; Sun et al., 2021).
Some classes of transposable elements are highly expressed at specific developmental stages. In mice, high MuERV-L/MERVL (the mouse-specific ERVL) expression is associated with the 2-cell stage (Peaston et al., 2004). Both full-length MERVL and MERVL "solo LTRs" (isolated LTRs that lack the rest of the MERVL sequence) peak in expression at EGA (Franke et al., 2017). It has been proposed that the full-length transcripts - or their transcription *per se* - is necessary for EGA (Sakashita et al., 2023), an intriguing scenario that awaits further experimental evidence. In addition, several MERVL LTRs were co-opted at the 5' region of protein-coding genes, hence representing a potential binding platform for mammalian EGA regulators (Macfarlan et al., 2012). The transcription factor DUX4 in humans and its ortholog DUX in mice were found to bind to MuERV-L promoters of EGA genes and to be responsible for their activation (Hendrickson et al., 2017; Iaco et al., 2017). However, DUX is not the only necessary factor for mammalian EGA, as some *Dux*$^{-/-}$ embryos were found to be viable (Iaco et al., 2019; Macfarlan et al., 2012). Most likely, it is a combination of transcription factors to be responsible for the bulk of mammalian EGA transcription, again in coordination with all other nuclear events that are happening at the same time at that stage. Also, DUX itself is zygotically expressed by unknown mechanisms. A possibility is that DUX is produced by the first minor wave of transcription of the embryonic genome ("minor wave" of EGA), already happening at the 1-cell stage in both mouse and human (Aoki et al., 1997; Asami et al., 2022). In mice, this

minor wave was detected at the paternal pronucleus during the S and G2 cell cycle phase (Aoki et al., 1997) and the most accepted vision to date is that it is an unproductive transcription. In fact, it includes intergenic regions and single-copy genes that are left unspliced and non-polyadenylated (Abe et al., 2015). However, some studies suggest that a role for minor EGA in paving the way for major EGA cannot be excluded (Hendrickson et al., 2017; Iaco et al., 2017; Liu et al., 2020). Minor EGA transcription could be necessary to remodel chromatin for major EGA (Liu et al., 2020) or it could lead to the expression of master regulators of major EGA. *Dux*, for example, is an intron-less gene hence does not require splicing to be functional.

Recently, the profiling of RNA Polymerase II (PolII) chromatin localization from the zygote to the blastocyst stage has been reported (Liu et al., 2020). This showed that in the PN5 1-cell stage embryo PolII is comparably enriched at promoters of oocyte-specific genes, EGA genes and developmental genes (expressed from peri-implantation), while at the early 2-cell stage PolII is found specifically enriched at EGA genes (Liu et al., 2020). The cues that cause the focusing of PolII to these loci remain to be discovered. Minor EGA transcription is a possibility, another non-exclusive hypothesis is the remodeling of histone modifications' landscape. Indeed, a significant event occurring after fertilization that is certainly coupled with EGA - although with a still unclear functional relationship - is the reprogramming of the epigenome, which is the focus of the next section.

Finally, for other cellular processes which have been historically less considered a role in EGA is starting to emerge. For instance, the maturation of the nuclear pore complex (NPC) is required for efficient EGA in zebrafish, by regulating the import of maternal TFs (Shen et al., 2022). Curiously, the intrinsically-disordered subunits of the NPC, called FG nucleoporins (FG-Nups) because of the phenylalanine-glycine repeats, are highly O-GlcNAcylated in mice (Snow et al., 1987) and the modification appears as soon as these subunits are incorporated in the complex at the time of EGA (Shen et al., 2022). FG-Nups are important for NPC selectivity in the process of nuclear import/export of biomolecules (Aramburu and Lemke, 2017).

The picture emerging from this body of studies is that mammalian EGA is far from being understood. The many unknowns leave space for unexplored mechanisms, such as the role of PTMs in regulating TF activity or transcription itself in this process.

3.3 Epigenetic dynamics in the early embryo

3.3.1 Preimplantation epigenetic dynamics

Following the fertilization of the oocyte, the transition from germ cells to totipotency and then pluripotency is characterized by dramatic changes in the nuclear architecture, first in the two pronuclei and then in the blastomeres. The nucleus is a complex organelle, with subcompartments associated to specific functions, such as the nucleoli, which is the nuclear compartment where rRNA transcription occurs, and the peripheral nuclear lamina, which constitutes the primary scaffold for the nuclear envelope. Various studies have suggested that the position of genes relative to different subnuclear compartments can influence gene expression (Crosetto and Bienko, 2020). A subnuclear compartment that shows remarkable changes during the first embryonic stages is the nucleolus. The nucleolus forms around the clusters of tandemly repeated ribosomal RNA genes (rDNA clusters). In the somatic cells, it has a tripartite architecture corresponding to a functional compartmentalization of the two processes of rRNA transcription and rRNA processing to form preribosomes. However, in the mature oocyte and until the early 2-cell stage, rDNA repeats cluster together with heterochromatic centromeric repeats around homogenous structures very different from the somatic nucleoli, which are called nucleolar-like-bodies (NLBs). In the oocyte, NLBs are transcriptionally silent. Transcription of embryonic rDNA is detected again in the zygote and gradually increases in the 2-cell stage. In parallel, rDNA gradually reorganizes into somatic-like nucleoli while pericentromeric heterochromatin forms various clusters around the nucleus, visible in somatic cells as intense spots with DAPI staining (Borsos and Torres-Padilla, 2016). Noteworthy, translation-related genes, including both ribosomal proteins and translation factors, are among the most highly dynamically regulated ones between the early and late 2-cell stage.

Nowadays, the term “epigenome” is used with ambiguous meanings (Greally, 2018). To avoid confusion, with this term I will specifically refer to the two main types of chemical modifications of chromatin (the complex formed by the DNA molecule wrapped around histone proteins): methylation of the DNA molecule on the fifth carbon of cytosine (5mC), mostly occurring at CpG dinucleotides, and histone post-translational modifications, namely the addition of various chemical groups to histone tails.

The DAPI staining of DNA of the zygote at different PN stages reveals a different pattern of fluorescence signal between the two pronuclei, already suggesting that they are undergoing different chromatin dynamics. Previous studies have shown that the oocyte and the sperm are two completely different cells in terms of their epigenome. An extensive remodeling of their chromatin starting at fertilization almost completely resets these differences.

The mature sperm has its genome extremely condensed in a crystal-like structure due to the protamines packaging instead of histones. Upon fertilization, protamines are promptly removed, hence the paternal genome is "a blank page" in terms of chromatin. Protamines are replaced by maternal histones, including the non-canonical histone variant H3.3, that at this stage is mostly incorporated in the paternal pronucleus. H3.3 is a replication-independent histone variant which is associated with transcriptional activation in somatic cells (Ahmad and Henikoff, 2002). The incorporation of H3.3 in the paternal pronucleus is required for the first embryonic cleavage (Lin et al., 2014); the modification of H3.3 tail at lysine 27 is essential for heterochromatin formation in the paternal pronucleus and developmental progression (Santenard et al., 2010). Overall in the zygote, the H3.3 variant is evenly distributed across the genome and it acquires a more somatic enrichment at active genomic loci during embryonic cleavages, when the replication-dependent canonical variant H3.1 and H3.2 are incorporated (Ishiuchi et al., 2021).

The oocyte has a unique pattern of histone marks compared to somatic cells: trimethylation of histone 3 at lysine 4 (H3K4me3), usually present at promoters of actively transcribed genes, spans large >10 kb domains in the oocytes, which do not necessarily correlate with gene promoters. The broad H3K4me3 domains get restricted to the promoters of EGA genes at the 2-cell stage (Eckersley-Maslin et al., 2018; Liu et al., 2020). The H3K27me3 mark, usually associated with transcriptional repression, is also present at broad domains in the oocytes (not correlating with the H3K4me3 domains), which are erased during the first cleavages and found at the promoter of developmental gene from peri-implantation (Zheng et al., 2016). At gastrulation, H3K27me3 is necessary for the temporal and spatial regulation of key developmental genes.

Besides the dynamics of histone marks, preimplantation development is also characterized by changes in chromatin mobility and accessibility. DNaseI hypersensitivity (DHS) mapping and assay for transposase accessible chromatin-sequencing (ATAC-seq) of preimplantation embryos showed that before EGA the 2-cell embryo has a noisy and weak chromatin accessibility profile, with large domains of accessible chromatin, while after EGA it acquires

the commonly observed peaks of accessibility at promoters (Eckersley-Maslin et al., 2018). The measure of histone mobility in preimplantation embryos with experiments of fluorescence recovery after photobleaching (FRAP) shows a high mobility of canonical histones during totipotency, which decreases as the cell becomes pluripotent but decreases even more with differentiation (Bošković et al., 2014). In fact, pluripotent cells of the blastocyst epiblast and their *in vitro* counterpart ESCs still display a remarkably decondensed chromatin pattern with low levels of compact heterochromatin (Ahmed et al., 2010), which fits their high developmental plasticity.

DNA methylation (DNAme) is generally associated with long-term transcriptional silencing (Greenberg and Bourc'his, 2019). Historically, its developmental significance has been associated with an essential role in repressing transposable elements (Walsh et al., 1998) and regulating the monoallelic expression of imprinted genes (Li et al., 1993) (see hereafter).
DNAme is catalyzed in mammals by three DNA methyltransferases: DNMT3a and DNMT3b are responsible for *de novo* DNAme, while DNMT1 can recognize hemi-methylated DNA sites and it is necessary for maintenance of DNAme across mitotic division (Greenberg and Bourc'his, 2019). A third de-novo DNA methyltransferase, namely DNMT3c, has been more recently discovered in rodents and is responsible for silencing young and still potentially active hence harmful transposable elements in the male germline (Barau et al., 2016). As far as the removal of DNAme is concerned, it can either occur passively during DNA replication, if the DNAme maintenance machinery is absent, or actively, thanks to the activity of ten-eleven translocation (TET) enzymes. TETs catalyze the hydroxylation of 5-methylcytosine to 5-hydroxymethylcytosine (5hmC), which is the first step of 5mC re-conversion to C as 5hmC is not maintained upon cell division (Tahiliani et al., 2009). TET3 is highly expressed only in the oocyte and zygote (Gu et al., 2011; Inoue and Zhang, 2011), while TET1 is the most highly expressed in the post-EGA preimplantation embryo (Williams et al., 2011). Notably, TET proteins have been reported to form stable complexes with OGT, but the functional relevance of this interaction remains unclear (Vella et al., 2013a).

Compared to somatic cells, the oocyte and the sperm have lower and higher levels of DNAme, respectively (Greenberg and Bourc'his, 2019). As seen for H3K4me3, the oocyte has also a non-canonical pattern of DNAme, which spans actively transcribed regions, leaving intergenic or transcriptionally inactive regions with low levels of methylation (Smallwood et al., 2011).

Parental DNAme is almost fully erased from the zygote to the blastocyst stage, except for important exceptions such as the IAP class of LTR-containing retrotransposons (Dahlet et al., 2020; Lane et al., 2003) and imprinted genes (Edwards et al., 2017). Those are a subset of about one hundred genes that instead of being expressed from both alleles - as the majority of genes - are expressed from either the maternal or paternal allele in a parent-of-origin fashion. This monoallelic expression derives from the differential methylation of a regulatory sequence, called imprinting control region (ICR), between the two alleles in female and male germ cells (Ferguson-Smith, 2011). The somatic differential allelic methylation is inherited from the methylation status of that sequence in the oocyte and sperm. It follows that DNAme at ICRs has to be maintained during the genome-wide demethylation wave of the preimplantation embryo. The resistance of ICRs from preimplantation demethylation is encoded in their sequence, which is bound by the Krüppel-associated box (KRAB)-containing zinc-finger protein 57 (ZFP57), which recruits KRAB-associated protein 1 (KAP1, also known as TRIM28) and other silencing factors (Li et al., 2008; Quenneville et al., 2011).

DNAme levels start decreasing after fertilization and the dynamics of DNAme loss is different in the two pronuclei. DNA methylation on the maternally inherited genome is predominantly passively lost, mainly through the exclusion of DNMT1 from the maternal pronucleus due to the cytoplasmic retention of DNMT1 recruiter UHRF1 by STELLA (also known as DPPA3) (Greenberg, 2021; Li et al., 2018). The paternal genome undergoes a faster and seemingly active demethylation, from which the maternal genome is protected, probably through binding of STELLA at dimethylated histone H3 lysine 9 (H3K9me2) (Nakamura et al., 2012). H3K9me2 is not found at this stage on the paternal genome (Santos et al., 2005).
The active DNAme loss at the paternal genome was initially thought to be driven by TET3 (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011). But better time-resolved immunofluorescence staining in the zygote showed that 5hmC levels begin to rise in the paternal genome only from PN3 onwards, when the main drop in 5mC level has just occurred (Amouroux et al., 2016). Other candidate mechanisms for paternal DNA demethylation have been suggested but not demonstrated yet (Eckersley-Maslin et al., 2018). Therefore, the full understanding behind the mechanism of paternal genome demethylation is still missing.

Besides preimplantation stages, genome-wide DNA demethylation also occurs in primordial germ cells during their differentiation towards female and male germ cells. In this case, also

ICRs lose DNAme and acquire the specific methylation status required by the maternal and paternal genome at each ICR.

3.3.2 Peri- and postimplantation epigenetic dynamics

DNAme levels start to rise again from the blastocyst onwards and reach somatic-tissue levels at E6.5. Accordingly, the level of *de novo* DNMT enzymes reaches a peak at peri-implantation and is found to decrease again from E6.5 onwards (Fig.S5c). In the pre-gastrula embryo, a notable difference is found in extraembryonic tissues (ExE and VE), which are both hypomethylated relative to the epiblast. Such hypomethylation includes retrotransposons and persists in the differentiated placenta (Smith et al., 2017).

About 80% of CpG dinucleotides are methylated in somatic cells; a notable exception are gene promoters, which for the high majority are devoid of DNAme (Edwards et al., 2017), with few exceptions such as gametic genes, which are repressed in somatic cells by promoter DNA methylation (Dahlet et al., 2020). As a consequence, most gene promoters do not undergo the waves of demethylation and remethylation described above. However, the advent of technologies capable of base-resolution mapping of 5mC revealed a small but detectable percentage of regulatory regions that are methylated in a tissue-specific manner in adult tissues (Hon et al., 2013). This indicates that during tissue differentiation at postimplantation stages they are subjected to temporal de-novo DNA methylation and demethylation waves. In mice, the majority of these regions correspond to enhancers or other regulatory elements (Hon et al., 2013), raising the possibility that their methylation status might influence the binding of methylation-sensitive transcription factors which have a role in lineage specification. Interestingly, blastocyst deficient in all TET proteins generate E7.5 embryos which initiate gastrulation, form the endoderm but an abnormal nascent mesoderm and fail to generate ectoderm and mature mesoderm lineages (Cheng et al., 2022).

At peri-implantation, epigenetic regulators play significant roles in cell differentiation. The best studied examples are Polycomb group (PcG) proteins. PcG proteins form two distinct complexes: Polycomb repressive complex 1 and 2 (PRC1 and PRC2), responsible for the deposition of H2AK119 ubiquitination and H3K27me3, respectively. H2AK119 is recognized by PRC2 and facilitates the nucleation of H3K27me3 and vice versa, although the two modifications can also be found independently (Schuettengruber et al., 2007). The chromatin domains formed by Polycomb are essential for maintaining transcriptional repression of

developmental genes during peri- and postimplantation development. For example, embryos which lack a functional embryonic *Eed* or *Ezh2* gene (PRC2) die during gastrulation stages (Faust et al., 1998; O'Carroll et al., 2001). EZH1/2 is PRC2's methyltransferase while EED is the PRC2 subunit responsible for binding H3K27me, hence inducing H3K27me3 spreading. Single cell RNA-sequencing (scRNA-Seq) of $Eed^{-/-}$ embryos showed that they are unable to produce neural ectoderm and bias the primitive streak towards posterior lineages such as the extraembryonic mesoderm and primordial germ cells (Grosswendt et al., 2020). *Eed*-mutants ectopically express developmental genes such as genes characterizing primordial germ cells and *Hox* genes, a class of genes with a conserved role across bilateralians in specifying the anterior-posterior body axis (Grosswendt et al., 2020).

3.4 Establishment of the pluripotent state

A specific set of transcription factors have been implicated in orchestrating the differentiation of the blastocysts's trophectoderm (TE) and pluripotent ICM, respectively. This first embryonic lineage specification event starts at a much earlier stage. The earliest TFs to show lineage-specific expression patterns are CDX2 and SOX2 (Strumpf et al., 2005; Wicklow et al., 2014). *Sox2* starts to be expressed at the 4-cell stage and at the 16-cell stage its expression is restricted to the inner cells of the morula. *Cdx2* expression starts at the 8-cell stage and is restricted to the outer cells in the 16-cell embryo (White et al., 2016). The mechanism initially leading to these molecular differences between the inner and outer cells of the 16-cell embryo is only partially understood. The Hippo signaling pathway is involved, which senses a change in cell polarity - the difference between the apical and basal cell membrane - and in turn regulates the expression of *Sox2* (White and Plachta, 2019). However, the recent application of live imaging to early embryos revealed that the cells of the 4-cell embryo - which has not established cell polarity yet - are already heterogeneous in terms of SOX2 binding to the DNA (White et al., 2016). Therefore, a complete picture of the mechanism establishing the first two embryonic lineages is still missing.

SOX2 and CDX2 are joined by another set of factors to further sustain and maintain differentiation towards the two lineages. The network of pluripotency-associated genes characterizing the ICM is also composed of OCT4 also known as POU5F1 (POU domain, class 5, transcription factor 1) and NANOG (Chambers et al., 2003; Loh et al., 2006; Mitsui et al., 2003). Loss of OCT4 does not prevent blastocyst formation, but the ICM fails to maintain pluripotency and expresses TE markers (Nichols et al., 1998).

Pluripotency is a transient transcriptional and epigenetic state, which drew a lot of attention because it encodes a unique differentiation potential while self proliferating in the undifferentiated state. The study of this state *in vivo* has always encountered technical challenges due to the low amount of cells in the ICM. A big leap forward in the field was the establishment in the 1980s of the culture conditions to grow cells derived from the mouse blastocyst, called mouse embryonic stem cells (mESCs), and maintain them undifferentiated (Evans and Kaufman, 1981). However, in the next three decades it became clear that in these culture conditions mESCs were closer to the postimplantation epiblast. The pluripotency of cells of the postimplantation epiblast is defined as "primed" pluripotency, while the blastocyst's ICM and epiblast cells are "naive" pluripotent. Both preimplantation and postimplantation pluripotent cells can give rise to the three primary germ layers, but they are different in many regards, including the level of DNA methylation (preimplantation pluripotent cells are hypomethylated), expression of different subset of genes of the pluripotency network, growth requirements and metabolism (Weinberger et al., 2016). The culture conditions which make mESCs more similar to the epiblast of the blastocyst were identified almost 30 years later and consist in the addition of mitogen-activated protein kinase (MEK) and GSK3β inhibition plus leukaemia inhibitory factor (2i/LIF) (Ying et al., 2008).
These culture conditions can make mESCs genetically unstable (Yagi et al., 2017) and sometimes a titration of these inhibitors is used. However, with the consequence of epigenetic changes introducing differences with the *in vivo* counterpart (Hackett et al., 2013), which needs to be taken into consideration when studying developmental epigenetics using this model.

As mentioned in the previous section, pluripotency is also functionally associated with a particularly open and dynamic chromatin. Of note for this study, this permissive pluripotent chromatin state is also tuned by cellular growth pathways, especially the level of translation (Bulut-Karslioglu et al., 2018). A proposed mechanism is that the short half-life found for some epigenetic regulators makes them susceptible to faster changes when tuning the translational machinery (Bulut-Karslioglu et al., 2018). This introduces the notion of coupling between metabolism and gene expression regulation during early development.

3.5 Metabolic dynamics of the early embryo

In mammalian cells, ATP production relies on the processes of glycolysis and oxidative phosphorylation (OxPhos). Glycolysis converts the 6-carbon glucose molecule into two 3-carbon molecules of pyruvate to generate 2 ATPs. In the absence of oxygen, pyruvate is

predominantly reduced to lactate; in aerobic conditions, pyruvate can be further converted to water and CO_2 by OxPhos. OxPhos is the chain of reactions generating 33 ATP molecules via the electron transport chain in the mitochondrial membranes (Nelson and Cox, 2021).
Not only the extracellular conditions, but also intracellular cues can induce changes in cellular metabolic choices. For example, a high proliferation state, such as that of cancer cells, often correlates with the shift of metabolism towards the glycolytic pathway, even under aerobic conditions (the so-called "Warburg effect"). The most likely explanation is that the glycolytic flux still produces sufficient ATP and at the same time the first intermediate glucose-6-phosphate can be used for the pentose phosphate pathway (PPP). PPP generates ribose-5-phosphate for nucleotides and NADPH-reducing power for the nucleotide and lipid biosynthesis. This is true also in the opposite direction: stem cells primarily depend on glycolysis and their differentiation is accompanied by a metabolic switch to OxPhos. The reverse happens upon cellular reprogramming to pluripotency. Notably, the metabolic switch during reprogramming is observed before the expression of pluripotency markers, suggesting that metabolism is itself a contributor to changes in developmental states (Miyazawa and Aulehla, 2018).

During mammalian early development, metabolic pathways have to adapt to the changing needs of the embryo going from a zygote, to a blastomere of a cleaving embryo, to a cell of a compacted morula and then to a bigger and more complex pre-gastrula. After gastrulation, at the chorioallantoic branching (CB) stage (E8/E9), the embryo starts to exchange gas and nutrients with the maternal circulation. Decades of research have established that both the nutrient requirements of the embryo and the balance between metabolic fluxes are very dynamic during this developmental window. In detail, before the 8-cell stage the embryo mostly uses pyruvate, glutamine and citrate as substrates for energy production. Glucose import is low and an increase in intracellular glucose can even lead to developmental arrest at the 2-cell stage (Dumollard et al., 2009). Furthermore, during preimplantation the embryo mostly relies on ATP production in the mitochondria and glycolysis is almost undetected (Dumollard et al., 2009). Accordingly, the expression of glycolytic enzymes is low while the expression of enzymes participating in the OxPhos pathway is upregulated at EGA and gradually increases during the next cell divisions (Malkowska et al., 2022). However, the higher mitochondrial respiration needs to be balanced with the extremely high sensitivity of the cleavage-stage embryo to oxidative stress. Therefore, especially in these first stages, a high level of antioxidants is required for developmental progression (Dumollard et al., 2009). Antioxidants include NADPH

and glutathione (GSH), both produced during maturation of the oocytes and needed for the first embryonic stages. Interestingly, reducing oocytes GSH impairs sperm nuclear decondensation (Perreault et al., 1988).

At implantation, the combined effect of the decrease in oxygen levels together with the need for high proliferation rate and biomass production should favor the Warburg effect: the increase in glycolysis and PPP and reduction in OxPhos flux. Indeed, this was what was initially observed *in vitro* when comparing ESCs (the counterpart of the blastocyst epiblast) with EpiSCs (the counterpart of the E6.5 epiblast): ESCs have a more bivalent metabolism, while EpiSCs rely on glucose and glycolysis (Zhou et al., 2012). The mechanism found in EpiSCs was the stabilization of hypoxia-inducible factor 1α (HIF1-α), which activates glycolytic genes and suppresses mitochondrial OxPhos genes (Zhou et al., 2012). This was mediated by the Activin/Nodal signaling pathway, which is required to maintain EpiSC in a pluripotent state, preventing their spontaneous differentiation (Camus et al., 2006). The measure of metabolic flux at peri-implantation stages *in vivo* has always been difficult. A recent study exploited the more easily accessible transcriptomics datasets of mouse embryos from preimplantation to E6.5 (including peri-implantation E4.5 and E5.5) to measure the expression of glycolytic and OxPhos enzymes. In line with *in vitro* model, implantation was accompanied by an increase in expression of glycolytic enzymes while the expression of OxPhos enzymes starts to decrease (Malkowska et al., 2022). Finally, at ~E8, with the establishment of embryonic circulation, OxPhos genes raise again, however with differences between tissues which account for their specific energetic needs. For example, the production of lactate stays high in some tissues such as the neural tube, the somite and the head mesenchyme, where glucose is also redirected to PPP (Miyazawa et al., 2017).

Metabolic fluxes are not only sources of ATP. They also produce the substrates for epigenetic modifications. For example, TET demethylates depend on α-ketoglutarate (αKG, an intermediate of the mitochondrial TCA cycle, the mitochondrial pathway producing NADH necessary for the electron transport chain). Naive ESCs maintain high the αKG to succinate ratio, maybe to facilitate DNA and histone demethylation (Carey et al., 2015). In cleavage-stage embryos, which require pyruvate as a carbon source to develop further than the 2-cell stage, only a fraction of pyruvate is oxidized for energy production and the reason is unknown (Miyazawa and Aulehla, 2018). Recently, it was shown that pyruvate is required for EGA (Nagaraj et al., 2017). In the same study, the mitochondrial enzymes which are responsible for

the production of acetyl-CoA (the entry point of the TCA cycle, deriving from pyruvate) and αKG were found to transiently localize to the nucleus in the 2-cell embryo (Nagaraj et al., 2017).

4) O-GlcNAc in mammals

4.1 The genetics of mammalian O-GlcNAc enzymes points to a key role in early development

The genomic location of *Oga* and *Ogt* genes in mammals may link them to early development. The *Oga* gene resides on chromosome 10 in humans and on chromosome 19 in mice. In both cases (and even in *Drosophila*) it is part of a highly conserved vertebrate cluster of genes, the *NK* cluster (Dona C. Love et al., 2010). The latter is one of the three large homeobox gene clusters in vertebrates, the other two being the *Hox* cluster and the *Parahox* cluster. *Hox*, *ParaHox* and *NK* clusters are proposed to originate from a common proto-cluster and function primarily in specifying the mesoderm, ectoderm and endoderm tissues, respectively (Garcia-Fernàndez, 2005). The homeobox genes they contain are transcription factors bearing a 60 amino acid domain called homeodomain. The first cluster of homeobox genes to be discovered was the *Hox* cluster, which is present in all bilateralians and which has been studied extensively by developmental biologists. The genes contained in the *Hox* cluster are expressed in a spatially collinear fashion compared to their genomic position during development, a phenomenon which leads to the specification of the anterior-posterior (A-P) body axis. Studies mostly in *Drosophila* uncovered a fundamental role of PcG proteins for Hox genes spatially regulated repression (Schwartz and Pirrotta, 2007). Mutation in PcG genes lead to the ectopic expression of some *Hox* genes in the incorrect spatial location and consequent fly body transformations (Jürgens, 1985; Lewis, 1978). The role of Polycomb in the temporal expression of *Hox* genes is conserved in mammals, where the presence of H3K27me3 - deposited by PRC2 - at a specific *Hox* gene temporally and spatially correlates with the window of expression of that gene (Soshnikova and Duboule, 2009).

The *Ogt* gene was mapped to the X-chromosome both in mice (XqD) and humans (Xq13) (Shafi et al., 2000). Interestingly, in both species it is located ~2Mb away from the X inactivation center (XIC). XIC is the locus demonstrated to be necessary and sufficient for X chromosome inactivation (XCI), the fundamental developmental process that in mammalian females leads to the silencing of genes residing on one copy of the X chromosome and is used to compensate for the presence of two X chromosomes. This process requires the transcription of two long non-coding RNAs, *XIST* and *TSIX*, from the XIC locus. *XIST* and *TSIX* then orchestrate a series of silencing events, including epigenetic modifications - also requiring Polycomb complexes -

and changes in the 3D structure of chromatin. XCI happens in two fashions during early mammalian development. The first event of XCI starts at the 4-cell stage where it affects specifically the paternal X chromosome (Xp) in each cell (Okamoto et al., 2004). This is called imprinted X inactivation (iXCI) and it is maintained in the trophectoderm of the blastocyst and in the extraembryonic tissues that will derive from the TE, which include the placenta. The second event of XCI happens in the ICM of the blastocyst. In fact, the ICM temporarily re-activates Xp to then inactivate one of the two X chromosomes. In the ICM, XCI happens in a random fashion in each cell, for this reason it is called random XCI (rXCI) (Okamoto et al., 2004). rXCI is mitotically maintained in each somatic cell of the adult except for female germ cells.

A certain percentage of genes (called "escapees") escape silencing of one of both types of XCI. With variations among tissues and individuals, ~25% X-linked genes have been found to escape XCI at least in one cell type in humans. Escaping genes are expressed at higher doses in females; this suggests that they might have female-specific functions. Escaping genes might also be responsible for some of the phenotypes observed in individuals with X chromosome aneuploidies, such as XO individuals, which develop the Turner syndrome (Berletch et al., 2011).

4.1.1 A higher dose of OGT in female extraembryonic tissues

Notably, *Ogt* escapes XCI in the extraembryonic ectoderm (Andergassen et al., 2021). This leads to the sexually dimorphic expression of OGT protein in the mammalian placenta, together with only a handful of other genes (Howerton et al., 2013). The evolutionary reason for this is unknown. A role for O-GlcNAc in iXCI in this tissue cannot be excluded. This hypothesis is conceivable also because of the reported function of *Ogt* in *Drosophila* Polycomb complex, which is involved in X chromosome silencing during XCI. In fly, *Ogt* loss-of-function causes no other obvious developmental defects apart from the ectopic expression of *Hox* genes, the characteristic phenotype of Polycomb mutants. For this reason, fly *Ogt* was initially identified as a PcG gene called *super sex combs* (*sxc*) (Ingham, 1984). More recently, the investigation of the mechanistic details revealed that O-GlcNAcylation of *Drosophila* PRC1 core subunit Polyhomeotyc (Ph) prevents its aggregation and ensures its proper function (Gambetta and Müller, 2014a).

The lower dose of OGT in male placentas (and almost double dose in female placentas) could be associated to sex-specific phenotypes starting from a perturbation to this tissue during

development. Thus far, one experimentally-tested example has been reported. In this study, the significantly lower levels of O-GlcNAcylation found in male placentas (of both mice and humans) were further reduced in case of experimentally-induced early prenatal stress (EPS), together with a reduction in *Ogt* expression (Howerton and Bale, 2014). Mice with a placenta-specific knockout of *Ogt* showed lower body weights and elevated hypothalamic-pituitary-adrenal stress axis responsiveness, phenotypes previously reported for EPS males. When looking at gene expression changes in the hypothalamus of the placental *Ogt*-KO males, some gene pathways were shared with EPS wild type male placentas and they were all involved in mitochondrial functions (Howerton and Bale, 2014). Finally, the same authors analyzed the placenta of females heterozygous for a trophoblast-specific *Ogt*-KO allele and found an increase in gene expression similarity with wild type male placentas, together with a seemingly lower H3K27me3 level, again as in male placentas (Nugent et al., 2018). However, H3K27me3 is the only histone mark they analyzed.

4.1.2 Maternal phenotype of *Ogt* loss-of-function

In mice, oocyte- and sperm- specific Cre-Lox-mediated *Ogt*-knockouts showed that a loss-of-function of the maternal *Ogt* allele is sufficient to cause lethality around implantation (E4-5). Instead, when the *Ogt*-null allele is inherited from the sperm, heterozygous female embryos develop into viable and fertile adults. However, female heterozygous are unable to transmit the *Ogt*-null allele to their progeny (O'Donnell et al., 2004a). This phenotype specifically dependent on the maternal allele has important implications. It suggests that *Ogt* has an essential role in at least one of the two following processes: i. oocyte maturation or ii. preimplantation development, when the embryo mostly relies on the OGT mRNA and protein molecules provided by the egg. Since the lethality of maternal *Ogt* loss-of-function affects both female and male embryos, a defect in X chromosome inactivation, although possible, is not enough to explain the maternal allele-dependent phenotype.

4.2 Mammals produce different OGT protein isoforms

The mammalian *Ogt* gene encodes for several alternative protein isoforms, which differ in the length of their TPR domain (Fig.3c): nuclear-cytoplasmic OGT (ncOGT, 12 TPRs, 116 kDa), mitochondrial OGT (mOGT, 9 TPRs, 103 kDa) and short OGT (sOGT, 2 TPRs, 78 kDa) (Hanover et al., 2003; Love et al., 2003). The three isoforms have different substrate selectivity when overexpressed in *E.coli* (Lazarus et al., 2006). Using both western blot of endogenous OGT and OGT overexpression fused to tags, ncOGT has been localized to the nuclear and

cytosolic cellular fraction, while mOGT has been found in the mitochondrial fraction and a putative mitochondrial targeting signal has been mapped at its N-terminus (Love et al., 2003).

In spite of the detection of the three isoforms at the protein level, the transcriptional or post-transcriptional mechanism leading to the production of mOGT and sOGT remains mysterious. Hanover *et al.* suggested that mOGT starts from an ATG codon which is present in intron 4 of the *ncOgt* mRNA isoform (Hanover et al., 2003). The use of this alternative ATG would require that the most N-terminal ATG - that used for the synthesis of ncOGT and beared by its exon 1 - is skipped. To my knowledge, such an alternative transcript with coding sequence starting from intron 4 does not have experimental evidence. In addition, this putative transcript is not present in Ensemble (www.ensembl.org) among those annotated for the last mouse (GRCm39) or human (GRCh38) genome assembly.

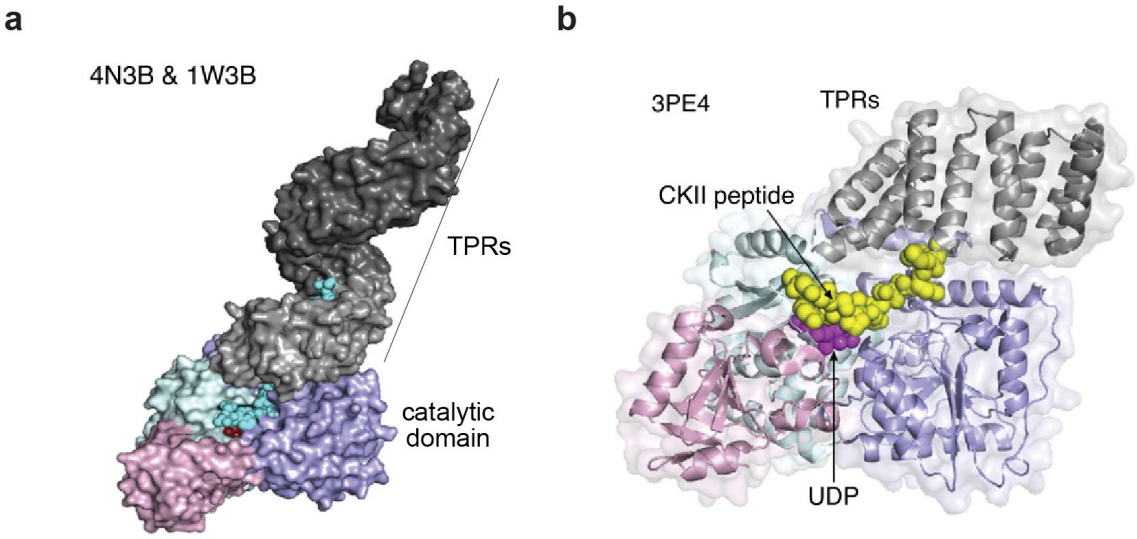
4.3 Structure of human ncOGT and OGA and target selection

Solving the crystal structure of ncOGT was essential to get insights into the mechanism of its catalysis and the rules for the recognition of the protein targets. The first crystal structure obtained for human OGT contained only the TPR region (PDB 1W3B) (Jínek et al., 2004) (Fig.3b) and it showed an elongated, right-handed superhelix. Based on the ~40° bending of one monomer, this region is likely to have a high degree of conformational flexibility. In 2011, the first structure of the catalytic region of human OGT bound to a substrate was reported (PDB 3PE4) (Lazarus et al., 2011) (Fig.3b), followed by others (such as 4N3B in Fig.3b, where it is combined with 1W3B using an overlapping region). From these structural data, it emerged that the catalytic site of OGT makes almost no contacts with the side chains of the substrate. Only a few features have been found that can partially predict OGT substrate selectivity, for example an extended conformation of the peptide at the catalytic site and contacts with a ladder of asparagines which span the entire length of OGT TPR domain (Joiner et al., 2019). In essence, structural studies suggest that OGT has a broad tolerance in target selectivity, which could explain how this one enzyme can modify such a diversity of proteins. In order to confer selectivity in specific contexts, a fraction of targets could be directed to the active site through the interaction with adaptor proteins that in turn interact with the TPR domains, but structural and biochemical evidence for such a mediated target recognition have yet to be provided.

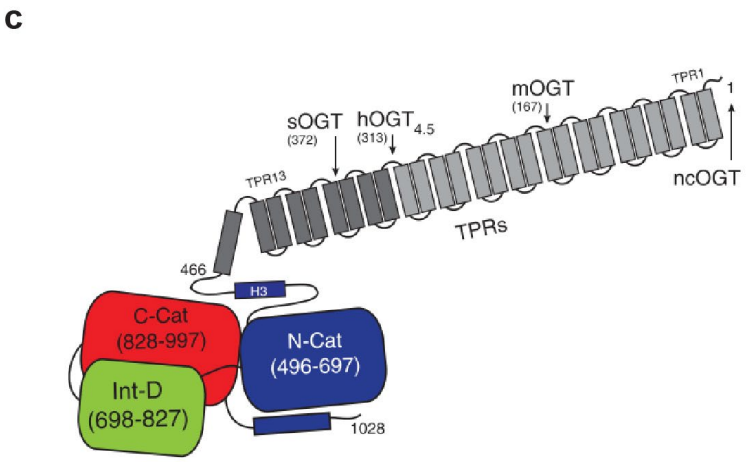
The requirement for an extended conformation of the peptide substrate is difficult to reconcile with O-GlcNAcylation sites mapped to many ordered or inaccessible regions of proteins. One possibility is that O-GlcNAc is added co-translationally, a hypothesis which is now supported for at least two OGT targets, SP1 and nuclear pore protein p62 (NUP62) (Zhu et al., 2015). The extended conformation requirement also implies that disordered domains are potentially more favorable OGT targets. In this regard, it is worth mentioning the presence of O-GlcNAc on TAU, an intrinsically disordered protein which is renowned for improperly aggregating and forming toxic amyloid structures in Alzhaimer disease. O-GlcNAcylation of TAU has been reported to compete with phosphorylation and protect against TAU's aggregation (Liu et al., 2004). This is supported by hyperphosphorylation of TAU in neurons without a functional *Ogt (O'Donnell et al., 2004a)*. A role for O-GlcNAc in preventing cellular aggregates is further supported by another study, where a synthetically O-GlcNAcylated form of α-SYNUCLEIN has been shown to be sufficient - even at substoichiometric concentrations - to block the toxic effect of this protein when added to cells in culture (Marotta et al., 2015). Furthermore, O-GlcNAc is present on the three heat shock proteins with anti-amyloidogenic activity in neurodegenerative disease, where it enhances their chaperone activity (Balana et al., 2021). Noteworthy, this happens only when specific sites are O-GlcNAcylated (instead of all sites on which O-GlcNAcylated was mapped), underlying the importance of maintaining a certain intracellular O-GlcNAc level in order to achieve the correct O-GlcNAc stoichiometry on proteins. The studies mentioned above suggest that - via different effects on protein function, interactions or structure depending on the target - O-GlcNacylation protects against the formation of aggregates, at least in the context of neurodegenerative diseases. As discussed above, the loss of O-GlcNAcylation in *Drosophila* has been proposed to cause the aggregation of Polycomb protein Polyhomeotic (Ph) (Gambetta and Müller, 2014a). Finally, substoichiometric O-GlcNAcylation of SynGAP, one of the most abundant proteins in postsynaptic densities (PSD) of neurons, has been shown to modulate the formation of PSD by controlling the size of liquid-liquid phase separating droplets that SynGAP forms with other PSD components (Lv et al., 2022).

Human OGA is a multidomain protein with an N-terminal domain similar to glycoside hydrolase family 84 (GH84), a stalk domain, a C-terminal pseudo histone acetyltransferase (HAT) domain and several low-complexity regions (Joiner et al., 2019). To be precise, mammalian OGA exists as two major spliced isoforms differing by the presence (OGA-L) or absence (OGA-S) of the HAT-like domain. Nucleocytoplasmic OGA-L is the main isoform

while the function of OGA-S, which seems to accumulate to lipid droplets (Keembiyehetty et al., 2011), is less understood. The structure of hOGA-L (hOGA from now on) was solved in 2017, both in apo form and together with different substrates, and showed that OGA is present in crystals as a dimer, as well as hOGT. However, hOGT dimeric form does not seem to be the functional one in the cell, while the stable dimeric form of hOGA in solution has been confirmed by biochemical studies and it is supported by the evolutionary conservation of residues on the dimeric interface (Joiner et al., 2019). As for hOGT, also the structure of hOGA revealed some principles of target recognition. Firstly, the GlcNAc moieties engage in significant interactions with residues at the active site, conferring absolute selectivity to OGA for hydrolyzing β-O-GlcNAc substrates over either α-O-GlcNAc or β-O-GalNAc substrates (Joiner et al., 2019). Furthermore, OGA makes some interactions with the side chains of the target peptide, which seem to enhance the binding affinity (Li et al., 2017). Therefore, the rate of O-GlcNAc hydrolysis might vary on the different O-GlcNAcylated proteins.



modified from Joiner et al. Current Opinion in Structural Biology (2019)



modified from Janetzko et al.
Journal of Biological Chemistry (2014)

Figure 3: Structure of OGT.

(a-b) Structures of human OGT, adapted from (Joiner et al., 2019).

(a) Complete model of the structure of ncOGT, obtained by combining PDB 4N3B (containing the catalytic domain and the last 4 TPR repeats) and PDB 1W3B (TPRs 1-9) using an overlapping region.

(b) Ribbon representation of the catalytic domain of ncOGT while binding a target peptide and UDP (PDB 3PE4).

(c) Schematic of human OGT adapted from (Janetzko and Walker, 2014), showing the differences in the length of the TPR region of the three isoforms nuclear-cytoplasmic OGT (ncOGT), mitochondrial OGT (mOGT) and short OGT (sOGT). Int-D: intervening domain. N-Cat and C-Cat: N- and C-terminal lobes of the catalytic domain. $hOGT_{4.5}$: end of structure 3PE4.

4.4 The phenotype of *Ogt* mutations in humans is a neurodevelopmental disorder

The loss-of-function of the *Ogt* gene is not compatible with human life, but a dozen male individuals with several (7 at the time of writing this manuscript) different single amino acid substitutions of *Ogt* have been associated with human pathologies. These mutations affect either the catalytic domain or the TPR region. Interestingly, independently of the site of the mutation, these individuals share a spectrum of clinical features, which are recollected to X-linked intellectual disability (XLID; OMIM #300997). Intellectual disability is a highly heterogeneous early-onset neurodevelopmental syndrome affecting ~1% of the population in the developed world and characterized by ID, developmental delay, facial dysmorphia and other body abnormalities (Schalock et al., 2021). In many cases, etiology is unknown, but 40% of cases with a genetic component have been found to be monogenic. A significant proportion of genes associated with ID (~18%) are X-linked genes, which are known to be highly expressed in brain tissues (Nguyen and Disteche, 2006). All OGT-XLID patients share a decreased intellectual ability (IQ<<70) and mental and physical developmental delay, including intrauterine growth retardation, low birth weight, short stature, drooling and compromised language skills; anatomical brain anomalies are very common, as well as some body dysmorphologies, hypotonia, epilepsy and genital defects; half of the patients had behavioral problems (Pravata et al., 2020). The commonalities of OGT-XLID patients led to the definition of a novel Congenital Disorder of Glycosylation (OGT-CDG).

The specific cellular alterations leading to the symptomatology of OGT-CDG are yet to be understood. The overall level of O-GlcNAc is not detectably altered in these patients, due to a compensatory decrease in OGA protein expression (Pravata et al., 2019a; Vaidyanathan et al., 2017a; Willems et al., 2017a). However, the hypo-GlcNAcylation of some OGT targets could be enough for a disruption of specific cell functions in specific cell types. Based on the variety of OGT-CDG manifestations, it is difficult to predict the identity of such targets. Mitchell *et al.* bioinformatically screened all O-GlcNAcylated proteins deriving from available O-GlcNAc-omic studies based on i. sequence conservation of the O-GlcNAcylation site, ii. structural features emerged as favorable for targeting by OGT, iii. clinical data reporting these proteins as mutated in ID/developmental delay (DD) syndromes, iv. literature. These led to 22 protein candidates, which were involved in diverse processes including Ras/MAPK signaling,

translational repression, cytoskeletal dynamics and chromatin remodeling (Mitchell et al., 2022).

4.5 Tight control of O-GlcNAc homeostasis in the cell

The disruptive effect of a non-functional *Ogt* or *Oga* gene on both embryonic development and cellular viability indicates that maintaining O-GlcNAc homeostasis is crucial for mammalian cell viability. Findings from several independent groups strongly support the concept of a feedback mechanism both between the level of O-GlcNAc and the expression of OGT and OGA and between OGT and OGA protein levels. This mechanism aims at maintaining a near-constant O-GlcNAcylation level in physiological conditions and also in disease. OGA and OGT expression seems to correlate between each other in several human cancers (Qian et al., 2018). In OGT-XLID patients, reduced O-GlcNAcylation because of reduced OGT protein abundance or catalytic activity increased OGA protein expression and did not lead to a significant decrease in global O-GlcNAc levels in the analyzed cell types from patients (Pravata et al., 2019a; Vaidyanathan et al., 2017a; Willems et al., 2017a), hinting at a compensatory mechanism. OGT overexpression in different cell types led to increased OGA protein level, while OGA overexpression led to increased OGT protein level (Qian et al., 2018; Z. Zhang et al., 2014). Treatment with the OGA-inhibitor TMG induces an increase in *Oga* mRNA and a decrease in *Ogt* mRNA expression (Z. Zhang et al., 2014).

The molecular mechanism responsible for this feedback loop has not been fully elucidated. Different studies showed changes of *Ogt* and *Oga* at the mRNA level as a response to OGT/OGA dysregulation (Park et al., 2017; Pravata et al., 2019b; Qian et al., 2018; Vaidyanathan et al., 2017b; Z. Zhang et al., 2014), indicating that transcriptional regulation might be involved. However, this is not always the case, as in one reported case of OGT-XLID, where OGA protein was downregulated while unaltered at the mRNA level (Willems et al., 2017b).

Some potential mechanisms of transcriptional regulation have been suggested. OGA seems to activate transcriptional activators p300 and C/EBPβ, which bind *Ogt* promoter (Qian et al., 2018); OGT is found enriched at *Oga* promoter in OGT-XLID patients, together with transcriptional repressors mSIN3A and HDAC1 (Vaidyanathan et al., 2017b).

A post-transcriptional mechanism to regulate OGT expression which involves intron retention has been described (Park et al., 2017). Intron retention is a widespread form of alternative splicing that was discovered relatively recently (Boutz et al., 2015). Transcripts containing

detained introns can have a half-life longer than an hour but they are retained in the nucleus (Boutz et al., 2015). Here, if polyadenylated, they can potentially be targets of degradation by Poly(A) binding proteins-dependent pathways or they can be reservoirs of mRNA which gets promptly spliced when needed. Another intriguing possibility is a function as non-coding RNAs. In any case, the result is a reduction in the number of mRNA molecules exported to the cytosol and translated to a functional protein. Park *et al.* found that OGT-RI (*Ogt* with detained intron 4) is abundantly present together with the cytosolic mRNA in the standard cell culture conditions of the cell lines they tested (Park et al., 2017). Upon high O-GlcNAcylation conditions, this isoform increases, as a mechanism of maintenance of O-GlcNAc homeostasis. *Ogt* intron retention seems to require a sequence element - a conserved intronic splicing silencer (ISS) - but the RNA-binding proteins targeting it and leading to this phenomenon remain to be discovered (Park et al., 2017).

4.6 The roles of mammalian O-GlcNAc

The unsuccessful identification of an O-GlcNAcylation consensus sequence so far might be less surprising if one think that - in spite of the presence of only one *Ogt* and one *Oga* gene, O-GlcNAc is found on over five thousands in humans according to the most recent O-GlcNAc database (Wulff-Fuentes et al., 2021b).

Like phosphorylation, O-GlcNAc modifies the biochemical properties of proteins. Depending on the target, the covalent addition of the O-GlcNAc moiety can impact protein activity, stability, interaction with other biomolecules or compete with other PTMs such as phosphorylation, which is also catalyzed at serine and threonines. An increasing number of studies in cell lines and adult tissues describe functions of O-GlcNAc on targets associated with a variety of cytosolic complexes, including the proteasome (Zhang et al., 2003) and the translational machinery (Li et al., 2019; Shu et al., 2021), and pathways, including glycolysis and tricarboxylic acid cycle (Ferrer et al., 2014; Nie et al., 2020; Yi et al., 2012), mitochondrial transport (Latorre-Muro et al., 2021) and insulin signaling(Yang et al., 2008). The most relevant of these studies that describe the potential link between O-GlcNAc and the regulation of cellular metabolism will be discussed in the next section.

In addition, O-GlcNAc is involved in the mammalian cell cycle. OGT is responsible for the maturation of the cell cycle factor HCF-1, through its O-GlcNAcylation followed by proteolytic cleavage (Capotosti et al., 2011). HCF-1 cleavage happens in the same catalytic site as O-

GlcNAcylation(Lazarus et al., 2013) and is the only other known catalytic function of OGT. The HCF-1$_N$ and HCF1-1$_C$ remains associated and are essential for both progression through M phase (Julien and Herr, 2003) and expression of cell cycle S phase genes. For the activation of S phase genes, HCF-1 recruits H3K4 histone methyltransferase at their promoters when they are bound by E2F transcriptional activator (Tyagi et al., 2007a).
The role of mammalian OGT for cell cycle progression could be one of the reasons for the senescence of *Ogt*-null fibroblasts (O'Donnell et al., 2004b), for the requirement of OGT for the viability of proliferating cells such as mESCs (Shafi et al., 2000) and T cells (O'Donnell et al., 2004b) as well as for the lethality of maternal *Ogt*-null embryos (O'Donnell et al., 2004b). Indeed, deletion of *Ogt* in post-mitotic neurons did not affect cell number *in vitro* and *in vivo* (Lagerlöf et al., 2016).

Finally, a large body of O-GlcNAc literature strongly suggests a pivotal function of this modification on chromatin. This literature will be reviewed in section 4.7.2.

4.6.1 O-GlcNAcylation as a potential metabolic and environmental sensor

UDP-GlcNAc, the donor substrate used by OGT, is the end product of the hexosamine biosynthetic pathway (HBP) (Haltiwanger et al., 1990), which is one of the metabolic routes of intracellular glucose. The levels of glucose and glutamine were shown to regulate the flux through the HBP (Marshall et al., 2004). However, whether an increase in the intracellular level of certain metabolites leads to higher O-GlcNAcylation is a long-standing speculation that still lacks compelling evidence. In fact, UDP-GlcNAc is also incorporated in many other types of glycosylation, as described in section 1.4. Studies using artificial adjustment of glucose in culture conditions are difficult to interpret because it is hard to define whether the resulting intracellular glucose levels are close enough to what is observed in physiology or disease. Studies in cell models which experience a physiological change in intracellular glucose can be more informative. O-GlcNAc seems to be responsive to the intracellular glucose dynamics which follows T cell activation (Swamy et al., 2016). Furthermore, hyperglycemia was found to correlate with O-GlcNAcylation and hyperactivation of CaMKII, a mechanism which was suggested to take place during diabete-induced heart failure (Erickson et al., 2013).

Because of these and other reported cases of correlation between the intracellular availability of nutrients and UDP-GlcNAc/O-GlcNAc levels, in the O-GlcNAc field there is the inclination to make the hypothesis that this modification might be a nutrient sensor. For this reason, many

groups specifically investigated the role of OGT and O-GlcNAc in metabolic cellular functions and dysfunctions. In cancer cells, an increased O-GlcNAcylation of two glycolytic enzymes - PFK1 and PGK1 - has been found to promote tumor growth by redirecting glucose flux towards the pentose phosphate pathway and lactose production, respectively, hence favoring Warburg effect (Nie et al., 2020; Yi et al., 2012) (described in section 3.5). In agreement, another study using breast cancer cells showed that downregulation of OGT causes a decrease of the glycolytic flux (Ferrer et al., 2014). The authors did not provide a direct molecular mechanism connecting *Ogt* KD to cellular phenotypes, but they described increased proteasomal degradation of HIF1-α, a TF which promotes glycolysis, and decreased activation of mTORC1 signaling (Ferrer et al., 2014). mTORC1 is one of the two complexes formed by the mammalian target of rapamycin (mTOR) kinase, which sits at a fundamental node in the signaling network controlling cell growth. mTORC1 senses the cellular nutrient and growth factor availability in order to regulate the balance between anabolism and macromolecule recycling (Valvezan and Manning, 2019). mTORC2 promotes cell proliferation and survival in response to growth factors (Zoncu et al., 2011).

OGT was also reported to be required for termination of insulin signaling through recruitment at the cell membrane and O-GlcNAcylation of proteins of the insulin-response pathway (Yang et al., 2008). Interestingly, the involvement of OGT in insulin pathway regulation is conserved in worms (Dona C Love et al., 2010) and also in flies, where knockdowns of *Ogt* and *Oga* in insulin-producing cells led to a reduction or an increase in the activity of insulin intracellular signaling, respectively, with an increase in the haemolymph carbohydrate concentration in both cases (Sekine et al., 2010). Another study found that OGT overexpression in mice muscle and fat caused hyperinsulinemia (McClain et al., 2002), potentially associating OGT dysfunction with diabete mellitus. OGT has been shown to mediate the negative effect of hyperglycemia on the mammalian embryo, both preimplantation (Pantaleon et al., 2010), in a study artificially increasing glucose level in the culture media, and in postimplantation diabetic mice (Kim et al., 2017).

Besides pathogenic conditions, the mammalian early embryo naturally undergoes a metabolic switch between aerobic respiration and glycolysis around the time of implantation (described in section 3.5), thus representing a physiologic metabolically dynamics system with the potential to be regulated by O-GlcNAc.

Although the understanding of the interplay of the O-GlcNAc modification with metabolic pathways is still in its infancy, the picture emerging is that O-GlcNAcylation could take a role in the adaptation of cellular anabolic potential to physiological and pathological changes in nutrient availability. However, the molecular mechanisms and the direction of change in O-GlcNAcylation on each target are non-trivial and might vary depending on the cell type. For example, in the hypothalamus, where OGT and overall O-GlcNAc levels are higher than in other tissues such as the muscle and the liver, these levels increased even more in response to fasting (Ruan et al., 2014). O-GlcNAc increase was shown to contribute to inhibiting the activity of the hypothalamic neurons controlling fat browning, resulting in lower heat production by the adipose tissue and better energy conservation (Ruan et al., 2014).

It is worth mentioning two molecular mechanisms that could link O-GlcNAcylation to the regulation of cell growth, because they involve fundamental cellular machineries, namely the proteasome and the ribosome. A biochemical study showed that O-GlcNAcylation can inhibit the proteasome (Zhang et al., 2003). With regard to the translational machinery, the literature shows a complex picture, since O-GlcNAcylation of specific residues has both activating and repressing consequences on the eIF4A and eIF4G translation initiation factors (Li et al., 2019). In response to stress, O-GlcNAcylation of eIF3a influences translation reinitiation (Shu et al., 2021), while other translation factors are O-GlcNAcylated during the aggregation of untranslated messenger ribonucleoproteins into stress granules (Ohn et al., 2008). The involvement of O-GlcNAc regulation in the stress response seems to be conserved in both flies and worms. In the latter, an increase in O-GlcNAc levels was shown to be required for embryonic development at higher temperatures, although the mechanism was not clarified (Radermacher et al., 2014).

4.6.2 (*in vitro*) evidence for O-GlcNAc role on mammalian chromatin

The first evidence of O-GlcNAc presence on chromosomes comes from the intense staining of *Drosophila* polytene chromosomes, both using wheat germ agglutinin (WGA), a lectin with affinity for clustered GlcNAc moieties and for sialic acids (Monsigny et al., 1980), and enzymatic labeling with bovine milk galactosyltransferase followed by hydrolysis of all N-linked oligosaccharides (such as sialic acids) (Kelly and Hart, 1989). The most clear mechanistic function of O-GlcNAc has also been reported in fly, where O-GlcNAcylation of the Polycomb protein Ph is necessary for PRC1 repression of developmental genes (Gambetta and Müller, 2014b).

According to the most updated human O-GlcNAc database, the mammalian nucleus is enriched with O-GlcNacylated proteins (Wulff-Fuentes et al., 2021b). A growing body of evidence supports a central role of O-GlcNAc in regulating mammalian transcription.

The most direct link between O-GlcNAc and transcription is the mapping of O-GlcNAc on the C-terminal domain (CTD) of RNA Polymerase II (PolII) (Kelly et al., 1993). The CTD of PolII is made of repeats of the Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (52 imperfect repeats in humans). The phosphorylation of specific residues regulates PolII initiation and elongation. The role of O-GlcNAc on the CTD is not understood, because it is difficult to perturb O-GlcNAc on intracellular PolII,f or example by mutating the O-GlcNAcylated residues, without interfering with phosphorylations occuring at the same residues. For this reason, the importance of CTD O-GlcNAcylation for PolII transcriptional cycle has been exclusively described in reconstituted *in vitro* assays (Lewis et al., 2016; Lewis and Levens, 2021; Ranuncolo et al., 2012). The result of these assays supports O-GlcNAc requirement for both PolII recruitment to promoters and transcriptional pausing, but they do not agree on the type of O-GlcNAcylated CTD residues (Lewis et al., 2016; Lewis and Levens, 2021; Ranuncolo et al., 2012).

O-GlcNAc was also mapped on many transcription factors. The first reported case was zinc finger TF Sp1, which was found to bear many O-GlcNAc moieties when purified and analyzed *in vitro* (Jackson and Tjian, 1988). Sp1 is an ubiquitously expressed TF that can regulate a wide range of processes depending on the context (O'Connor et al., 2016), thus the significance of O-GlcNAc-Sp1 might depend on the cell type and cell stimuli.
O-GlcNacylated transcription factors include the pluripotency master regulator OCT4, which seems to require this modification to induce the expression of a subset of pluripotency genes (Jang et al., 2012). The same study shows that O-GlcNAcylation of OCT4 is required for cellular reprogramming.

In spite of the role of OGT in the PRC1 complex in fly, it is unclear whether mammalian OGT has a functional link with Polycomb. In a breast cancer cell line, EZH2, the catalytic subunit of PRC2, was found to be O-GlcNAcylated on Ser75, a residue proposed to be critical for EZH2 protein stability (Chu et al., 2014). Knockdown of *Ogt* in this cell line reduced PRC2 levels and decreased bulk H3K27me3 levels by ~50% (Chu et al., 2014). However, in mESCs, *Ogt*

knockdown did not lead to a detectable reduction of EZH2 or H3K27me3 (Myers et al., 2011). PRC2 and H3K27me3 were also unperturbed in *Ogt*-null *Drosophila (Gambetta et al., 2009b)*.

Is OGT recruited to chromatin and if so, by which mechanism? Unbiased pull-downs of OGT's associated proteins found HCF1 and TET enzymes as the most prominent OGT interactors, seemingly as part of two separate complexes (Deplus et al., 2013; Vella et al., 2013a). The two complexes could contribute to target OGT to specific genomic locations.

In HeLa cells approximately 50 % of nuclear OGT is found to be stably associated with HCF1 and the latter to stabilize OGT protein levels and impact its nuclear localization (Daou et al., 2011; Ruan et al., 2012). Because HCF1 is known to interact with numerous transcriptional regulators, HCF1-OGT interaction could be required for the function of some of them, but to my knowledge evidence is scarce. One study reports O-GlcNAcylation of HCF1-associated PGC-1α, a transcriptional coactivator regulating mitochondrial biogenesis and hepatic gluconeogenesis. The authors suggest that PGC-1α O-GlcNAcylation recruits the protein deubiquitinase BAP1, which induces PGC-1α degradation (Ruan et al., 2012). It is possible that besides a few cases like the one described above, OGT and HCF1 colocalization on chromatin mostly reflects the stable association of HCF1 with OGT during proteolytic processing of HCF1 (Gambetta and Müller, 2015).

TET proteins (TET1, TET2, TET3) were found among the most enriched proteins in OGT pull-downs in both human cell lines (Deplus et al., 2013; Ito et al., 2013) and mouse embryonic stem cells (mESCs) (Q. Chen et al., 2013; Vella et al., 2013a). In these studies, OGT is not required for the recruitment of TET2 and TET3 proteins to chromatin (Q. Chen et al., 2013; Ito et al., 2013). All three TET proteins have been found to be O-GlcNAcylated (Myers et al., 2011; Shi et al., 2013; Vella et al., 2013a; Q. Zhang et al., 2014), but the functional significance is unknown. TET1, the most abundant TET in mESCs, is enriched at promoters and shown to be necessary for OGT recruitment at TSS in this cell line (Vella et al., 2013b). However, all studies performing chromatin profiling of mammalian OGT suffer from the absence of the essential *Ogt*-KO control, therefore caution is needed when interpreting these results.

Various studies reported that histones are O-GlcNAcylated (Fong et al., 2012; Fujiki et al., 2011; Hahne et al., 2012; Sakabe et al., 2010; Zhang et al., 2011). However, among all 13 O-GlcNAcylated histone residues reported, only one, threonine 32 on histone H3 (H3-T32), was identified by ETD mass spectrometry (the only mass spectrometry method which can directly detect O-GlcNAc, described in section 1.5) (Fong et al., 2012). All other groups relied on

indirect strategies to identify the O-GlcNAcylated histone residue. Furthermore, each of these studies identified a different set of residues and only two of them were common to two studies. In conclusion, the presence of O-GlcNAc on histones is dubious. Finally, a thorough biochemical investigation has failed to detect O-GlcNAc on histones (Gagnon et al., 2015).

Finally, a study by our group raised the intriguing possibility that O-GlcNAc is involved in the repression of transposable elements downstream of DNA methylation (Boulard et al., 2020), a mechanism which is only partially understood. Repressive chromatin modifiers have to be recruited at the DNAme loci in order to achieve transcriptional silencing. The identity of the components of the DNAme-dependent repressive complex is still partly unknown. KAP1/TRIM28 is required in this complex at least in the two contexts of DNAme-dependent silencing of transposable elements (Rowe et al., 2010) and DNAme-dependent monoallelic expression of imprinted genes (Alexander et al., 2015). In an unbiased biochemical assay to find DNAme-dependent TRIM28 interactors, OGT was one of the highest enriched proteins (Boulard et al., 2020). In addition, a recombinant OGA fused to dCas9 was used to target the removal of O-GlcNAcylation at LTR promoters of IAP elements and this led to their release from silencing (Boulard et al., 2020).

In sum, OGT associates with different chromatin complexes, some of which are functionally involved in transcriptional silencing (e.g. TRIM28) while others are required for transcriptional activation (e.g. PolII). Hence, the transcriptional output of the O-GlcNAc modification depends on the nature of the modified chromatin factor, on the effect of O-GlcNAc on such a factor and perhaps on the genomic and chromatin context.

Materials and Methods

1) Animal care and strains

Mice were treated in compliance with the rules and regulations of the Institutional Animal Care and Use Committee (IACUC) under protocol number 21-012_RM_MB. Mice were housed in the pathogen-free Animal Care Facility at EMBL Rome on a 12-hours light-dark cycle in temperature and humidity-controlled conditions with *ad libitum* access to food and water. For the profiling of O-GlcNAc/OGT in preimplantation embryos and for zygote microinjection experiments, female FVB/NCrl were always used, while male FVB/NCrl were used for all imaging and NGS experiments except for the blastocysts and E7 embryos Smart-Seq experiments, for which PWD/Ph sperm was used. PWD/Ph sperm was used also for the Ogt^{T931A} blastocysts Smart-Seq experiment.

2) Generation of *Ogt* mouse mutant alleles' series: Ogt^{Y851A}, Ogt^{T931A}, Ogt^{Q949N}, Ogt^{H568A}, Ogt^{NLS}, $Ogt^{NterAID\text{-}MYC\text{-}FLAG}$

Transgenic mouse production was performed by Gene Editing and Embryology Facility (GEEF) at EMBL Rome.

Ogt mutant alleles were created by CRISPR/Cas9-editing technology using FVB (Charles River) zygotes as previously described (Quadros et al., 2017). Briefly, for Ogt^{Y851A} (FVB/NCrl-Ogt$^{em1(Y851A)}$Emr) CRISPR crRNA oligo (TAGATGGGACGTCTACACCC) was annealed with tracrRNA and combined with a homology flanked ssODN donor coding for Tyrosine at amino acid position 851, substituting the wild type Alanine. The target location was exon 19 of transcript *Ogt-201*, (ENSEMBL v109); genomic coordinate: ChrX. 100719847-100719886 (GRCm39/mm39).

Similarly, for Ogt^{T931A} (FVB/NCrl-Ogt$^{em2(T931A)}$Emr) crRNA oligo (TTGTGTAATGGACACACCAC) was combined with a ssODN donor coding for Threonine at amino acid position 931 substituting Alanine, exon 20, transcript *Ogt-201*, ChrX. 100722515-100722518.

For Ogt^{Q949N} (FVB/NCrl-Ogt$^{em3(Q949N)}$Emr) crRNA oligo (TAGATGGGACGTCTACACCC) was combined with a ssODN donor coding for Glutamine at amino acid position 949 substituting Asparagine, exon 19, transcript *Ogt-201*, ChrX. 100719847-100719886.

For *Ogt*^{H568A} (FVB/NCrl-Ogt$^{em4(H568A)}$Emr) crRNA oligos (GCTATGTGAGTTCTGACTTC and ATGAAGTGTGGAATACGTCA) were combined with a ssODN donor coding for Histidine at amino acid position 568 substituting Alanine, exon 13, transcript *Ogt-201*, ChrX. 100713458->100713464.

For *Ogt*$^{NterAID\text{-}MYC\text{-}FLAG}$ (FVB/NCrl-Ogt$^{em6(AID)}$Emr) crRNA oligos (CTCCAGATGGCGTCTTCCGT and ACTGTCGGCCACGTTGCCCA) were combined with a ssDNA donor coding for 2xMyc and 1xFLAG tags inserted at Ogt N terminus, transcript *Ogt-201*, ChrX. 100683892.

In all cases, annealed sgRNAs were complexed with Cas9 protein and combined with their respective ssDNA donors (Cas9 protein 20 ng/ml, sgRNA 20 ng/ml, ssDNA 10 ng/ml, all IDT). These reagents were microinjected into zygote pronuclei using standard protocols (Liu and Du, 2019), and after overnight culture 2-cell embryos were surgically implanted into the oviduct of day 0.5 post-coitum pseudopregnant CD1 mice.

Founder mice were screened for by PCR, first using primers flanking the sgRNA cut sites which identifies InDels generated by NHEJ repair, and can also detect larger products implying HDR. Secondary 5' and 3' prime PCRs using the same primers in combination with template-specific primers allowed for the identification of potential founders, these PCR products were then Sanger sequenced and aligned with the *in silico* design.

Sequences of the single-stranded donor templates are provided in Table S1. *Ogt* mouse mutant allele lines were maintained on an FVB/NCrl genetic background. PCR genotyping of the mice tails was performed at weaning (P21) using primers listed in Table S2.

3) *Rosa26*^{OsTir} mouse allele

The *Rosa26*^{OsTir} mouse allele (FVB;B6J;129-Gt(ROSA)26Sor$^{tm1(OsTIR)}$Emr) was kindly donated to us by the group of Christophe Lancrin at EMBL Rome. The allele had been created by EMBL GEEF in the past using mouse embryonic stem cell targeting, which results in a mixed C57BL/6;129S background. Mice were backcrossed twice with the FVB/NCrl strain and the mice resulting from the backcrossing were used for all experiments.

4) Derivation and culture of mouse embryonic fibroblasts (MEFs)

Primary MEFs were derived from E12.5 post-coitum embryos as previously described (Jain et al., 2014). Briefly, the head and the organs were removed and gDNA was extracted using the PCRBIO Rapid Extract Lysis Kit (PCR Biosystems #PB15.11-24) for PCR genotyping with

primers in Table S3. The embryonic body was cut to small pieces using sterile forceps and pieces were transferred to 5 ml Trypsin-EDTA (0.05%) (Gibco), where tissue was homogenized through pipetting. The homogenized tissue was spun, the supernatant removed and cells resuspended in MEF media (Dulbecco's Modified Eagle Medium (DMEM), 10% fetal bovine serum (Gibco), 100U/mL Penicillin/Streptomycin (Gibco), 1x GlutaMAX (Gibco)). Cells were cultured on gelatin (0.1% gelatin in dH_2O) in 5% CO_2 at 37°C and passaged at ~80% confluency using Trypsin-EDTA (0.05%) for cell detachment and dissociation. All experiments were performed between passage 1 and 4. The three pairs of clones studied were derived from three different litters; each pair (*OsTIR*,*Ogt*^{AID} and control *OsTIR*,*Ogt*^{WT}) was coming from the same litter.

4.1 Auxin treatment of MEFs

Auxin (Sigma-Aldrich #I5148) was resuspended in ddH2O under sterile conditions to a stock concentration of 10 mM, aliquoted and stored at -20°C. For auxin treatment of MEFs, 500 uM of auxin was added to cell medium at day 0 and cells collected 24 h, 48 h and 96 h later. For the 96 h time point, the medium was changed after 48 h with fresh MEF medium supplemented with auxin 500 µM, in order to keep auxin concentration constant throughout the 4 days. At the moment of collection, 500 uM auxin was added to all reagents for cell dissociation (Trypsin-EDTA and DPBS). Cells were collected in 1.5 ml epp, resuspended in 500 µl of DPBS + 500 µM auxin and the same volume of 2x Monarch DNA/RNA Protection Reagent (component of NEB #T2010) was added before snap freezing and storage at -80°C degrees until RNA extraction.

5) mRNA-Seq library preparation and sequencing

Total RNA was extracted with Monarch Total RNA Miniprep Kit (NEB #T2010S). The integrity of the RNA was determined using the TapeStation 4150 (High Sensitivity RNA kit, Agilent Technologies #5067-5579). Library preparation was performed at the Genomics Core Facility of EMBL using the TruSeq Stranded mRNA Library Preparation kit (Illumina) according to the manufacturer instructions. All libraries were pooled and sequenced in one run (75 bp single-end mode) on the Illumina NextSeq500 sequencer.

6) Western blotting

Total protein lysates were prepared by lysing cells with RIPA buffer (150mM NaCl, 1% NP-40, 0.5% deoxycholic acid, 0.1% SDS, 50mM Tris pH 8.0 in ddH_2O) and subsequent pelleting to discard insoluble fractions. Protein concentration was quantified using the Pierce BCA Protein Quantification kit (ThermoFisher #23225). 20 μg of protein extract and 5μL of WesternSure Pre-stained Chemiluminescent Protein Ladder (Li-cor #926-98000) was loaded in a NuPAGE 4-12% or 4-20% Bis-Tris Protein Gel, 1.0 mm, 12-well (ThermoFisher #NP0335BOX) and run at 150V in an XCell Sure Lock apparatus (ThermoFisher). Proteins were transferred on a PVDF membrane in a Mini-Protean Tetra System apparatus or a Trans-Blot Turbo Transfer System (Bio-Rad). Abundant epitopes were blocked with a buffer containing 5% milk or 5% Bovine Serum Albumine (BSA) (Sigma-Aldrich #A2153) for O-GlcNAc detection and 0.1% Tween-20 in ddH_2O. Membranes were incubated overnight at 4°C (or for 1 hour at room temperature for loading controls) with primary antibody dilutions in a buffer containing 5% BSA and 0.1% Tween-20 in ddH_2O. Membranes were washed twice with 0.5% Triton X-100 and 0.5 M NaCl in ddH_2O for 5' and once 0.5 M NaCl in ddH_2O for 10', then rinsed with PBS and incubated for 1 hour at room temperature with 1:200000 secondary antibody dilutions of HRP-coupled goat anti-rabbit IgG (ThermoFisher #G-21234) or 1:100000 of HRP-coupled goat anti-mouse IgG (ThermoFisher #G-21040). Membranes were washed as done after primary antibody and incubated with GE Healthcare ECL Prime (Sigma #RPN2232) for imaging at a ChemiDoc imaging system or ImageQuant 800 (Amersham).

The following primary antibodies were used: anti-OGT (abcam #ab177941) 1:2000 (and 1:5000 only for E7.5 embryos), anti-O-GlcNAc clone RL2 (abcam #ab2739 and Merck Millipore #MABS157) 1:2000, anti-histone H3 (1:5000), anti-lamin A/C (Santa Cruz Biotechnology #sc-376248) 1:200.

7) In vitro fertilization (IVF)

In vitro fertilization was performed based on the published protocol, with minor modifications (Guan et al., 2014). Briefly, superovulation of 6-8 weeks FVB females was induced by hormonal stimulation (5 IU of PMSG and 5 IU of hCG 64 h and 16 h before collection, respectively), and cumulus-oocyte complexes were collected in KSOM media containing GSH (final concentration 10 mM; Sigma-Aldrich #G6013). Concomitantly, cauda epididymis and vasa deferentia from FVB or PWD males were dissected and sperm was gently squeezed out

into capacitation media (HTF supplemented with MBCS at final concentration of 0.75 mM) (HTF: Sigma-Aldrich #MR-070-D, MBCD: Sigma-Aldrich #C4555), and allowed to swim up for 1 hour. Sperm were subsequently counted and cumulus-oocyte complexes were inseminated with 0.2M sperm in a 200 µl fertilization drop. Four hours after sperm addition, zygotes were cleaned from the surrounding cumulus cells and sperm by 5-6 washes in KSOM, then cultured in KSOM for 2 hours prior to microinjections or for 0 to 4 hours prior to fixation for immunostaining.

8) Preimplantation embryo culture

Embryos were cultured in a standard mammalian cell incubator (37°C, 5% CO_2) in EmbryoMax KSOM Mouse Embryo Medium (Sigma-Aldrich #MR-106-D).

9) Auxin treatment of zygotes grown *ex-vivo*

For auxin treatment of preimplantation embryos, 250 µM of auxin was added to the fertilization drop (KSOM media containing GSH), to all KSOM washing drops and to the KSOM media used for culture for the next following days. In order to maintain a constant auxin concentration, embryos were moved to fresh KSOM supplemented with 250 µM auxin every second day. On the day of fixation for immunofluorescence staining or collection for Smart-Seq, M2 medium used for washing the embryos was supplemented with 250 µM auxin.

10) Cloning of the pRN3P-NLS-EGFP-Btgh-3NLS plasmid

The cDNA encoding O-GlcNAc hydrolase (OGA, *Btgh84*) from Bacteroides and fused to 3 SV40 Nuclear Localization Signals (NLSs) was amplified from the plasmid described in Boulard *et al.* (Boulard et al., 2020), fused to enhanced GFP *mgfp5* flanked by an N-terminal SV40 NLS and the whole insert cloned in the pRN3P-H3.3R8A-GFP plasmid using restriction enzymes NotI and SmaI. pRN3P-H3.3R8A-GFP is a generous gift from Maria Elena Torres-Padilla (Eid et al., 2016). The NLS-mgfp5 fragment was ordered as a dsDNA (gBlock) from Genewiz. The three fragments were assembled together using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs #E2621L) and following manufacturer instructions for three fragments. The catalytically inactive ("dead") $Btgh84^{D242A}$ (characterized in Dennis et al. (Dennis et al., 2006)) was cloned the same way. Plasmids will be available with details at Addgene (www.addgene.org) under accession number 194469 and 194470 after publication of the manuscript. Sequences of primers and gBlock used for cloning are in Table S4.

11) In vitro transcription (IVT) of NLS-EGFP-Btgh-3NLS mRNA

The plasmid was linearized using KpnI, purified and then used as a template for mRNA synthesis with mMESSAGE mMACHINE T3 (Invitrogen #AM1348). In vitro transcribed RNA was purified with RNA Clean & Concentrator Kit (Zymo Research) with elution in H_2O. The mRNA product was run on an RNA ScreenTape with TapeStation 4150 (Agilent) to verify the presence of one main band of the correct size, the concentration was measured with NanoDrop 2000 spectrophotometer (ThermoFisher) and working concentration aliquots were made by diluting the mRNA in H_2O. Aliquots were stored at -80°C until microinjection.

12) Zygote microinjection

Two hours after IVF, zygotes were transferred to an injection plate with M2 medium at 37°C, and subjected to microinjections. Zygotes were microinjected with either pRN3P-NLS-EGFP-Btgh-3NLS or pRN3P-NLS-EGFP-deadBtgh-3NLS mRNA, both at 100 ng/µl (for the 2-cell RNA-Seq and IF experiment) or 300 ng/µl (for all other experiments). RNA injections were carried out using a Femtojet (Eppendorf) microinjector at 150 hPa pressure for 0.1 seconds, with 10 hPa compensation pressure (estimated microinjection volume of 10 pl). After the microinjections, zygotes were washed 3-4 times in KSOM and placed back into culture. GFP fluorescence was verified the next day and GFP-positive (injected) 2-cell stage embryos were sorted from the non-injected ones and used for downstream microscopy, development or transcriptome analyses.

13) Preimplantation embryo collection and library preparation for single embryo Smart-Seq

Single post-EGA 2-cell embryos (28 h post-IVF), morulae (48 h post-IVF) and blastocysts (68 h post-IVF) were collected in 5 µl of 1x TCL lysis buffer (Qiagen 1031586) containing 1% (v/v) 2-mercaptoethanol (Gibco #31350010) and 0.5 U/µl of SUPERase•In™ RNase Inhibitor (ThermoFisher #AM2694). RNA was separated from the surrounding material using RNAClean XP beads (Beckman Coulter #A63987) according to the manufacturer's protocol for a 96-well plate and Small Volume Reactions. After the last ethanol wash, the RNA was eluted from RNA beads in 8 ul of H_2O, 3 ul of which were transferred in a new plate containing 1 ul of dNTPs (BiotechRabbit #BR0600204, 10 mM each) and 1 ul of 10 uM oligo-dT primer (5′–AAGCAGTGGTATCAACGCAGAGTACT30VN-3′). The plate was then sent to EMBL Genomic Core Facility, which used a modified Smart-Seq2 protocol (Picelli et al., 2014) using

SuperScript IV RT and tagmentation procedure previously described (Hennig et al., 2018) to prepare single-embryo full-length cDNA sequencing libraries. The retrotranscription reaction mix was as follows: 2 µl SSRT IV 5x buffer, 0.5 µl 100 mM DTT, 2 µl 5 M betaine, 0.1 µl 1 M MgCl2, 0.25 µl 40 U/µl RNAse inhibitor, 0.25 µl SSRT IV, 0.1 µl 100 uM TSO, 1.15 µl RNase-free H_2O; with thermal conditions: 52°C 15', 80°C 10'. cDNA was generated using 18 PCR cycles. The cDNA cleanup (0.6x SPRI ratio) was carried out omitting the ethanol wash steps and the elution volume was 13 µl of H_2O. For tagmentation, the sample input was normalized to 0.2 ng/uL. Before the final clean-up after tagmentation and PCR, 2 µl of each sample were pooled in a single tube and the pool was cleaned up using 0.7x SPRI ratio. The pool was sequenced in one run (40 bp paired-end mode) on the Illumina NextSeq500 sequencer. See Table S6 for the number of pooled embryos and average number of reads obtained per embryo at each stage.

14) Immunofluorescence staining of preimplantation embryos

MII oocytes (20 h post-hCG), zygotes (0, 2 and 4 h post-IVF), 2-cell embryos (pre-EGA: 19 h post-IVF; post-EGA: 28 h post-IVF; if different, as specified in Fig. legend), morulae (46 h post-IVF) and blastocysts (68 h post-IVF) were treated the same way, as in Bošković et al. (Bošković et al., 2012) with some modifications as follows. After two washes in M2 medium (Sigma-Aldrich #M7167), zona pellucida was removed by a brief incubation in drops of warm Acidic Tyrode's solution (Sigma-Aldrich), followed by other two M2 washes to neutralize the acid. The embryos were then washed once in PBS + 0.5% BSA, fixed in 4% PFA in PBS for 20' at 37°C, permeabilized in 0.5% Triton X-100 for 20' at 37°C and washed three times in PBS-T (0.15% Tween-20 in PBS). The epitope was then unmasked in 50 mM NH_4Cl solution in H_2O for 10' at room temperature, followed by two additional PBS-T washes and then blocking - for 3h at room temperature or overnight at 4°C - in BSA (Bovine Serum Albumin, Sigma-Aldrich A2153) 3% in PBS-T. Primary antibody incubation was performed overnight at 4°C in the blocking solution, followed by three washes in PBS-T, re-blocking for 30' at room temperature, three additional PBS-T washes, secondary antibody incubation for 1 to 2h at room temperature in the blocking solution and three final PBS-T washes. The embryos were immediately mounted on coverslips in Vectashield (Vector Laboratories H-1200) containing 4'-6-Diamidino-2-phenylindole (DAPI) or - were specified - in drops of 75% Vectashield in PBS, to preserve the 3D structure.

The primary antibodies used were: anti-OGT (abcam #ab177941), anti-O-GlcNAc clone RL2 (abcam #ab2739 and Merck Millipore #MABS157), anti-O-GlcNAc clone HGAC85 (ThermoFisher HGAC85), anti-CDX2 (abcam ab76541). Dilution of all primary antibodies was 1:200. Secondary antibodies used were: A647-conjugated goat anti-rabbit IgG (#A-21244), A488-conjugated goat anti-rabbit IgG (#A-11008), A488-conjugated goat anti-mouse IgG (#A-11001), A555-conjugated goat anti-rabbit IgG (#A-27039), A555-conjugated goat anti-mouse IgG (#A-21424) (all from ThermoFisher). Dilution of all secondary antibodies was 1:500.

Fixed immunostained samples were imaged using a Nikon AX scanning confocal (using galvanometric mirrors) or an X-light V3 Spinning disk confocal, as specified in Figure legends.

15) Quantification of microscopy images

All steps were performed using ImageJ (Schneider et al., 2012) when not differently specified. For the quantification of O-GlcNAc signal in single morulae (Fig.7b): i. brightness and contrast of the raw images were adjusted to the same scale, ii. projection of the z stacks was performed for each image using the maximum signal intensity, iii. images were denoised using Aydin (https://doi.org/10.5281/zenodo.5654826), with Classic Image Denoiser butterworth and default parameters, iv. ellipses were drawn inside each embryo and mean intensity in the area was measured for single embryos.
For the quantification of O-GlcNAc signal in 2-cell embryos (Fig.8b): i. brightness and contrast of all raw images were adjusted to the same scale, ii. the z plane showing the largest area and highest intensity of DAPI signal was selected for each 2-cell nucleus, iii. a measuring line was drawn across each nucleus and all single-nucleus measures recorded, iv. line plot was generated in R from all single-nucleus measures using function ‘geom_smooth’ with ‘gam’ method.

15) Surgical transfer of embryos for *in utero* development and E7 embryo collection and preparation for Smart-Seq

CD1 females 7-8 weeks of age in estrus were mated to vasectomized males and the day of the vaginal plug was considered to be 0.5 d of pseudopregnancy. That morning, 15-20 2-cell embryos microinjected the day before and kept in culture overnight were surgically transferred into the oviduct of each pseudopregnant female. Six days later (seven days after IVF, for which reason we consider the embryos to be E7) embryos were removed from their implantation sites and dissected in PBS +10% FBS (PAN/Biotech). Embryos were cleaned to a high level by

removing Reichardt's membrane and surrounding maternal cells with the aid of 0.01 mm Tungsten needles (F.S.T. 10130-10), then imaged with trans-illumination on Leica M205C dissection microscope with fixed objectives (10x16) and photographed. For Smart-Seq, embryos were cut with a glass needle separating the epiblast and trophectoderm halves. Both tissues were collected in 200 µl of 1x RNA/DNA protection buffer (NEB T2010S) in PBS, snap frozen and stored at -80°C until RNA extraction. RNA from single embryonic halves was extracted using Monarch Total RNA Miniprep Kit (NEB T2010S), following manufacturer's instructions for mammalian whole blood with some modifications: 5 µl of Proteinase K were added to each thawed sample, followed by incubation at room temperature for 15'. RNA was eluted with 25 µl of 1x TE buffer. For a balanced sequencing design, sex genotyping was performed using cDNA made from 2ul of each sample. cDNA was synthesized with 100 U of Maxima H Minus Reverse Transcriptase (ThermoFisher EP0752), using random hexamers and following manufacturer's instructions. cDNA genotyping was performed using primers in Table S5.

The quality of the RNA was checked on an RNA High Sensitivity ScreenTape with TapeStation 4150 (Agilent) and only good quality samples were used by EMBL GeneCore facility for Smart-Seq2 library preparation as described above and sequencing on a NextSeq-500 (40 bp paired-end mode).

16) Genotyping of mouse blastocysts from *Ogt*^{T931A} mouse line

The cDNA from single blastocysts was prepared from 3 ul of RNA extracted as in Methods section 13 using SuperScript™ IV Reverse Transcriptase (ThermoFisher #18090200) according to the manufacturer instructions and random hexamers. PCR genotyping of the resulting cDNA was performed with primers in Table S5.

17) Bioinformatics analysis

17.1 MEFs mRNA-Seq data analysis for single copy genes

The analysis pipeline was performed using Galaxy (Afgan et al., 2018) to obtain the table of gene counts. The quality of reads was analyzed using FastQC, then trimmed from adapters and low-quality 3'-end nucleotides using Trim Galore v0.6.3 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default parameters. Reads were mapped to Gencode vM25 (GRCm38.p6) transcript sequences using Salmon v0.8.2 (Patro et al., 2017) with default parameters for stranded single-end libraries. The gene counts

were used in a custom Rmd script as input for DESeq2 v1.34.0 (Love et al., 2014) after gene-level summarization using tximport (Soneson et al., 2016). The test used for statistical significance was the Wald test and the significance cut-off for optimizing the independent filtering was 0.05. Whenever MA-plots are shown, the log2 fold changes are shrunken using the 'ashr' method (Stephens, 2017).

17.2 Single embryo Smart-Seq data analysis for single copy genes

The analysis pipeline was performed using Galaxy (Afgan et al., 2018) to obtain the table of gene counts. Briefly, the quality of reads was analyzed using FastQC v0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were trimmed from adapters and low-quality 3'-end nucleotides using Trim Galore v0.6.3 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default parameters for paired-end libraries (-q 20 --stringency 1 -e 0.1 --length 20 –paired), before mapping them using STAR v2.7.8a (Dobin et al., 2013) and default parameters for paired-end reads to the GRCm38 mouse genome, containing the NLS-GFP-Btgh-NLS transgene in case of the zygote microinjections' datasets. Gene counts were obtained from the bam files using featureCounts (subreads v2.0.1) (Liao et al., 2014) using default parameters for paired-end reads and counting fragments instead of reads. The gene counts were used in a custom Rmd script as input for DESeq2 v1.34.0 (Love et al., 2014). The test used for statistical significance was the Wald test and the significance cut-off for optimizing the independent filtering was 0.05.

Only for datasets which were the result of more than one embryo generation and collection (microinjected 2-cell embryos, microinjected blastocysts and Ogt^{AID} blastocysts) batch effect had to be considered when testing for differential expression using DESeq function. To this aim, package RUVSeq (Risso et al., 2014) was used to compute the factors of unwanted variation, with function 'RUVs' and k=3 for microinjected 2-cell embryos and blastocysts, with function 'RUVg' and k=6 for Ogt^{AID} blastocysts. The subset of control genes to use in RUVg was found with a first run of DESeq2 differential expression analysis - separately for all untreated and all auxin-treated blastocysts: the genes with i. DESeq2-normalized mean across samples > 100, ii. adjusted p-value > 0.8 and iii. log2FC < 0.05 for each condition were selected. The 3 or 6 factors of unwanted variation computed with RUVSeq were used in 'DESeq' formula e.g. $\sim W1 + ... + WN + condition$.

Whenever MA-plots are shown, the log2 fold changes are shrunken using the 'ashr' method (Stephens, 2017). Before PCA and differential expression analysis, a few filtering steps were applied to all datasets: embryos with < 10^6 reads and outlier embryos in a scatterplot of

mitochondrial DNA (mtDNA) gene expression versus percentage of reads mapping to ribosomal DNA (rDNA) were removed as low quality samples. For blastocysts from the zygote microinjection experiment, the outliers in Fig.12c (aneuploid embryos) were additionally removed; for E7, those halves showing contamination from the complementary tissue using the expression of markers for epiblast (*Map7d3*, *Pdzd4*, *Uchl1*) and trophectoderm (*Gjb3*, *Gm9*, *Wnt6*). The exact number of embryos filtered at each step is in Table S6.

17.3 *In silico* genotyping of single embryos from RNA-Seq data

Sex assignment was done based on the DESeq2-normalized counts of chrY-mapping genes *Ddx3y* and *Eif2s3y* (Groff et al., 2019). For the blastocysts from $Ogt^{AID/WT}$ oocytes (Results section 2.1.3), the assignment of the Ogt^{AID} genotype was based on the raw sum of reads mapping to the AID-2xMyc-FLAG exogenous sequence, which was dividing the samples into two distinct populations. A sample was defined as wild type if the sum of AID-2xMyc-FLAG-mapping reads was < 72.

17.4 Analyses of gene sets enrichment

All performed in a custom Rmd script. Gene ontology (GO) over-representation test was performed on differentially expressed genes (DEGs) (adj. p-value < 0.05, any log2FC) with DESeq2-normalized mean across samples > 10 using function 'enrichGO' of R package 'clusterProfiler' v4.2.2 (Wu et al., 2021) with p-value and q-value cutoffs of 0.05 and 0.1, respectively, and default parameters, then GO level 5 was selected using function 'gofilter' and results were simplified by q-value using function 'simplify' and a similarity cutoff of 0.6.

Analysis of over-representation of Molecular Signature Database (MsigDb) hallmark gene sets was performed on DEGs with DESeq2-normalized mean across samples > 10 using function 'enricher' of R package 'clusterProfiler' v4.2.2, with p-value and q-value cutoffs of 0.05 and 0.2, respectively and default parameters.

Gene set enrichment analysis (GSEA) was performed on all genes of the dataset with DESeq2-normalized mean across samples > 10, ranked by -log10(adj. p-value)*sign(log2FC), using function 'gseGO' of R package 'clusterProfiler' v4.2.2, with p-value cutoff of 0.05 and default parameters. GSEA results were simplified based on adj. p-value after computing semantic similarity using 'mgoSim' function of R package 'GoSemSim' (Yu et al., 2010). Similarity cutoff was 0.6 for all figures showing GSEA result except Fig.13e, for which it was 0.8. For each figure, the first significant terms based on Normalized Enrichment Score (NES) are shown, ordered by GeneRatio.

17.5 Principal component analysis of single embryo Smart-Seq datasets

PCAs in Fig.7c-d and in Fig.9b were performed with function 'prcomp' on log2-transformed and DESeq2-normalized raw counts, using the first 1000 or 500 genes with the highest variance, respectively, after removing genes with DESeq2-normalized mean across samples < 10. For PCA in Fig.11a, PCA was performed with function 'prcomp' on log2-transformed, DESeq2-normalized and RUVs-adjusted counts, using the first 1000 genes with the highest variance, after removing genes with DESeq2-normalized mean across samples < 10. For Fig.9c, procedure was the same except that all DEGs (adj. p-value < 0.05, any log2FC) from the Btgh versus dBtgh and Btgh versus non-injected comparison at any stage were used. For Ogt^{AID} blastocysts (Fig.15c), PCA was performed using the ranks of the raw counts, after selecting the first 1000 genes with the highest rank.

17.6 Analysis of metabolic modules from single embryo Smart-Seq datasets

Adapted from (Malkowska et al., 2022). For computation of the single-embryo score for each metabolic module: i. for each sample, I averaged the expression of all genes belonging to that module, taken from (Malkowska et al., 2022), ii. I subtracted from this score a control score computed for that sample as follows: I bin the genes by average expression across all samples in n bins; for each bin, I select n genes; for each sample, I compute the average expression of the n genes*n bins.

17.7 ChIP Enrichment Analysis (ChEA)

ChEA was performed through the enrichment analysis web tool enrichR (https://maayanlab.cloud/Enrichr/) (E. Y. Chen et al., 2013), using ChEA 2022 dataset.

17.8 Analysis of Transposable Elements (TEs) expression

The analysis was performed using a custom Snakemake v5.9.1 pipeline (Mölder et al., 2021), available at https://github.com/boulardlab/BS_EMSeq. In summary, quality of the fastq files was checked with FastQC v0.11.8 and reads were trimmed using Trim Galore v0.6.4 with default parameters for paired-end libraries (-q 20 --stringency 1 -e 0.1 --length 20 –paired). Trimmed reads were aligned to GRCm38 mouse genome with STAR v2.7.5c (Dobin et al., 2013), with parameters recommended by Teissandier et al. (Teissandier et al., 2019) for the analysis of transcripts derived from autonomous transposable elements (TEs) in the mouse genome: --outFilterMultimapNmax 5000 --outSAMmultNmax 1 --outFilterMismatchNmax 3 -

-outMultimapperOrder Random --winAnchorMultimapNmax 5000 --alignEndsType EndToEnd --alignIntronMax 1 --alignMatesGapMax 350 --seedSearchStartLmax 30 --alignTranscriptsPerReadNmax 30000 --alignWindowsPerReadNmax 30000 --alignTranscriptsPerWindowNmax 300 --seedPerReadNmax 3000 --seedPerWindowNmax 300 --seedNoneLociPerWindow 1000. After alignment, using a custom script, reads are kept only from pairs where both mates were: - completely included into a repetitive element; - not overlapping with gene bodies of Gencode vM25. Repeat Library 20140131 (mm10, Dec 2011) was used as repetitive elements annotation, after excluding "Simple repeats" and "Low complexity repeats".

With the STAR parameters above, only one random alignment (the one with the highest alignment score) is reported for multimappers, preventing the precise quantification of single repetitive elements. Therefore, for each repName of the repetitive element annotation - often present in multiple copies in the genome - read counts were summarized using featureCounts (subreads v2.0.1) (Liao et al., 2014), with parameters -p -B -s 0 --fracOverlap 1 -M.

The downstream analysis of TEs expression was performed in a custom Rmd script. First, an analysis at family level (repFamily field in the Repeat Library annotation) was performed in order to identify samples with an important contamination from genomic DNA, since the latter can affect the quantification of TE transcripts. For this analysis, FPKM were computed for each repFamily and FPKM values of DNA transposon families were manually checked and samples with high DNA contamination were removed from the rest of the analysis. Differential Expression analysis of RNA TEs was performed at the RepName level, using DESeq2 v1.34.0 (Love et al., 2014) and adding the value of the total sum of DNA FPKM as a confounding factor to DESeq formula (~ *DNA_FPKM* + *condition*). The test used for statistical significance was the Wald test and the significance cutoff for optimising the independent filtering was 0.05. Whenever MA-plots are shown, the log2 fold changes are shrunken using the 'ashr' method (Stephens, 2017).

17.9 Analysis of differential allelic expression

The analysis pipeline was performed using a Snakemake v5.9.1 pipeline (Mölder et al., 2021). Firstly, the position and identity of all annotated Single Nucleotide Polymorphisms (SNPs) between the FVB/NCrl and PWD/Ph mouse strains was collected from GenomeMUSter, a resource of the Mouse Phenome Project (MPD, RRID:SCR_003212) (Bogue et al., 2022). The PWD genome is not included in the Mouse Genome Project, hence the accuracy of SNPs annotation highly varies compared to the better annotated FVB strain. Therefore, only SNPs

with confidence level = 1 for the FVB DNA base and ≥ 0.8 for the PWD DNA base were kept for all analyses and used to create a masked GRCm38 mouse genome with 'maskfasta' from the BEDTools suite (Quinlan and Hall, 2010). After quality check with FastQC v0.11.8 and trimming using Trim Galore v0.6.4 with default parameters (-q 20 --stringency 1 -e 0.1 --length 20 –paired), reads were mapped to the masked genome using STAR v2.7.5c (Dobin et al., 2013). Alignments parameters were the default ones, except for --alignEndsType EndToEnd and --outSAMattributes NH HI NM MD, both required for SNPsplit (see below). After keeping only uniquely aligned reads and removing duplicates with Picard Tools (http://broadinstitute.github.io/picard/) - as suggested in Castel et al. (Castel et al., 2015) - reads were assigned to either the FVB or the PWD genome using SNPsplit (https://www.bioinformatics.babraham.ac.uk/projects/SNPsplit/). FVB reads and PWD reads were used for two separate instances of featureCounts (subreads v2.0.1) (Liao et al., 2014) to obtain gene counts, with parameters -s 0 -t 'exon' -g 'gene_id' -Q 0 --minOverlap 1 --fracOverlap 0 --fracOverlapFeature 0 -p -C. featureCounts output was then used in a custom Rmd for downstream analysis of ploidy and differential allelic expression.

For the analysis of ploidy, all gene counts for the FVB and PWD genome were summed for each embryo. For differential allelic expression, autosome genes (including imprinted genes) with PWD or FVB counts < 10 in half of the samples from each condition were filtered out, then DESeq2 was used to test whether the ratio of FVB/PWD reads was changing for each gene among conditions (Btgh/dBtgh-injected and non-injected embryos). The formula used for DESeq2 Wald test was $\sim condition + condition:sample\_name + condition:genome$, where *genome* is FVB or PWD.

Objectives and hypotheses

The growing body of evidence described in the introduction raises the possibility of an important role for O-GlcNAc in transcriptional regulation during mammalian development. Furthermore, it emerges that this reversible modification could be in the position to translate changes in the metabolic intracellular demands or in the environment into an adaptation of gene expression. This idea is also consistent with the loose selectivity of OGT, which in response to those environmental changes could in principle modify a spectrum of protein targets, potentially tweaking the whole cell state.

In spite of the ever-increasing number of studies which test this hypothesis *in vitro*, the significance of O-GlcNAc during *in vivo* mammalian development has never been functionally tested. This is primarily due to the early embryonic lethality resulting from the lack of a functional *Ogt*. However, the embryonic lethality also implies the necessity for OGT activity in early development, hence it urges testing OGT's function in the developing mouse embryo *in vivo*. This was the overreaching objective of my PhD project.

To this aim, I started my study by profiling the localization and abundance of OGT and the O-GlcNAc modification during preimplantation development. Such a dataset had never been reported before (Results section 1). Then, I developed two functional strategies for the inducible and acute perturbation of the O-GlcNAc modification during *in vivo* early mouse development and dissected the resulting molecular phenotype of the embryo (Results section 2). Finally, I designed a suite of hypomorphic *Ogt* alleles to analyze the phenotype of mice with different levels of impairment of OGT's activity (Results section 3).

Before this study, there was no molecular data available in O-GlcNAc perturbed embryos. Therefore, I wanted my investigation to be centered around an unbiased exploration of the developmental and transcriptomic phenotype resulting from OGT/O-GlcNAc depletion. However, a few main non-exclusive hypotheses drove some deeper analyses: i. the potential role of O-GlcNAc on EGA, suggested by the mapping of O-GlcNAc on PolII CTD and on TFs, as well as by the peri-implantation lethality of *Ogt*-null embryos; ii. OGT's contribution to lineage specification in the blastocyst or to gastrulation; iii. OGT's function in regulating the expression of transposable elements, because of the reported repression of LTRs by OGT in

mESCs; iii. O-GlcNAc as a fundamental player of physiological adaptation of the developing embryos to metabolic changes. The latter adaptation could mechanistically be implemented through i., ii. or iii. or through an action on other pathways (or through all of these).

Results

1) OGT and O-GlcNAc profiling during mouse early development

The dynamics of protein O-GlcNAcylation in the early embryo is unknown. I performed the first spatiotemporal profiling of OGT (Fig.4a) and the O-GlcNAc modification subcellular localisation (Fig.4b) across all key preimplantation stages. In collaboration with Urvashi Chitnavis, Master Student, I generated embryos through in vitro fertilization (IVF) and collected the following stages for immunofluorescence staining: zygotes at different pronuclear (PN) stages (0-4 h post-IVF), late 2-cell stage embryo (22-26 h post-IVF), morulae (48 h post-IVF) and late blastocysts (68h post-IVF), together with MII oocytes.

Firstly, I observed an unexpected enrichment of the OGT signal in the paternal pronucleus of the zygote as compared to the maternal one, consistent with a previous report at PN3/4(Nagaraj et al., 2017) and now shown to be consistent across all PN stages. To exclude the possibility of this result being due to assisted reproductive technology artifacts, I also stained zygotes and 2-cell embryos (40 h post-hCG) obtained through natural mating and observed the same OGT localization pattern as with IVF (Fig.S1a). Surprisingly, the O-GlcNAc signal is symmetrically present on both pronuclei, pointing to a potential catalytically-independent role of OGT in the zygote. Secondly, the nuclear to cytosolic O-GlcNAc signal ratio tends to increase from the 2-cell stage onwards, prompting a further investigation of the role of this modification in the embryonic nuclei specifically.

In the zygote, the nuclear O-GlcNAc signal is the most intense at the nuclear membrane. This is likely attributable to the highly O-GlcNAcylated disordered FG nucleoporins (FG-Nups) subunits of the nuclear pore complex (NPC) (Li and Kohler, 2014; Shen et al., 2022) being particularly stained by the RL2 antibody, which has been raised against rat NPC-lamina nuclear fraction (Snow et al., 1987). However, zygotic nuclear O-GlcNAc is not exclusively found at the NPC, as a different antibody (clone HGAC85) stains perinucleolar foci in the zygote and the 2-cell stage embryo (Fig.S1b). The perinuclear signal was lost after enzymatic removal of O-GlcNAc (Fig.S1c), proving the specificity of this O-GlcNAc pattern.

From the 2-cell stage onwards, the OGT signal is roughly equivalent between the nucleus and the cytoplasm. Notably, both OGT and O-GlcNAc are undetectable on the metaphase chromosomes, both in the oocyte and in blastomeres undergoing mitosis present in morulae and blastocysts.

It needs to be considered that the anti-OGT monoclonal antibody that I used recognizes the C-terminus and thus should detect all reported OGT's isoforms. These include a nuclear-cytoplasmic one (ncOGT), a mitochondrial isoform (mOGT) and a shorter isoform (sOGT) of unknown localization and function. To gain insight into OGT isoform usage in the early embryo, I analyzed public mRNA-Seq data spanning mouse pre- and postimplantation stages (Wu et al., 2016; Zhang et al., 2017) and inspected reads coverage at the *Ogt* locus (Fig.4c). I observed that the longer transcript isoforms (ENSMUST00000119299 and ENSMUST00000044475), which encode ncOGT, are the predominant ones in the oocyte, then their levels gradually diminish up to the 4-cell stage, when they start to increase again (Fig.4c-d). The low level of full-length *Ogt* transcript at EGA might indicate that the embryonic genome initially transcribes very little amount of the full-length *ncOgt* mRNA. It also suggests that the *ncOgt* mRNA signal during the first cleavages is mainly contributed by maternally inherited mRNA. My analysis also revealed that a shorter form terminating after exon 4 is produced starting from EGA and is present at all cleavage stages until the blastocyst (Fig.4c-d). This transcript isoform could correspond to the shorter N-terminus ones annotated in Ensembl (ENSMUST00000153979, ENSMUST00000150161, ENSMUST00000155792, ENSMUST00000155713), which are predicted to be non-coding, or it could be a novel form, either non-coding or coding for a truncated protein with no catalytic domain. In any case, it is characterized by the retention of intron 4 of the full-length coding isoform.

I also performed an annotation-independent analysis of *Ogt* mRNA exon junctions in the preimplantation embryo. This analysis found only one alternative splicing event, precisely inside intron 4 and again happening only from the late 2-cell stage and disappearing from the blastocyst onwards (Fig.S1d). The position of the alternative exon in the middle of intron 4 could correspond to the 'decoy exon' which is proposed to lead to nonsense-mediated decay (NMD) of human *Ogt* mRNA upon export to the cytoplasm and translation, in response to changes in O-GlcNAc levels (Tan et al., 2020). In another study, *Ogt* with detained intron 4 (OGT-RI) has been reported to be retained in the nucleus, also as a mechanism of reduction of OGT protein level and thus maintenance of O-GlcNAc homeostasis in specific conditions (Park et al., 2017). In the post-EGA pre-blastocyst stages the high number of reads upstream of intron 5 (Fig.4c) compared to the exons downstream makes NMD or nuclear degradation of the full-length transcript unlikely and another molecular mechanism might be in place that is specific to EGA. Indeed, EGA is known to be characterized by a unique splicing regulation (Wyatt et al., 2022).

Finally, exons mapping do not support the presence in the embryo of a full-length stable transcript including intron 4, which is instead visible in mESCs (Fig.4c) and resembles the one proposed to be destined to NMD or nuclear retention (Park et al., 2017; Tan et al., 2020), nor a shorter C-terminal transcript which could encode sOGT. For the latter, I cannot exclude mechanisms of production other than alternative splicing.

In conclusion, my analysis suggests that a phenomenon of reduction of full-length catalytically active OGT levels via alternative transcript isoform production takes place at EGA. I speculate that this could be a means to maintain OGT concentration - hence O-GlcNAc stoichiometry on proteins - during the first embryonic cleavages, when cell division is not followed by an overall growth of the embryo (hence reducing cell volume and increasing intracellular concentrations).

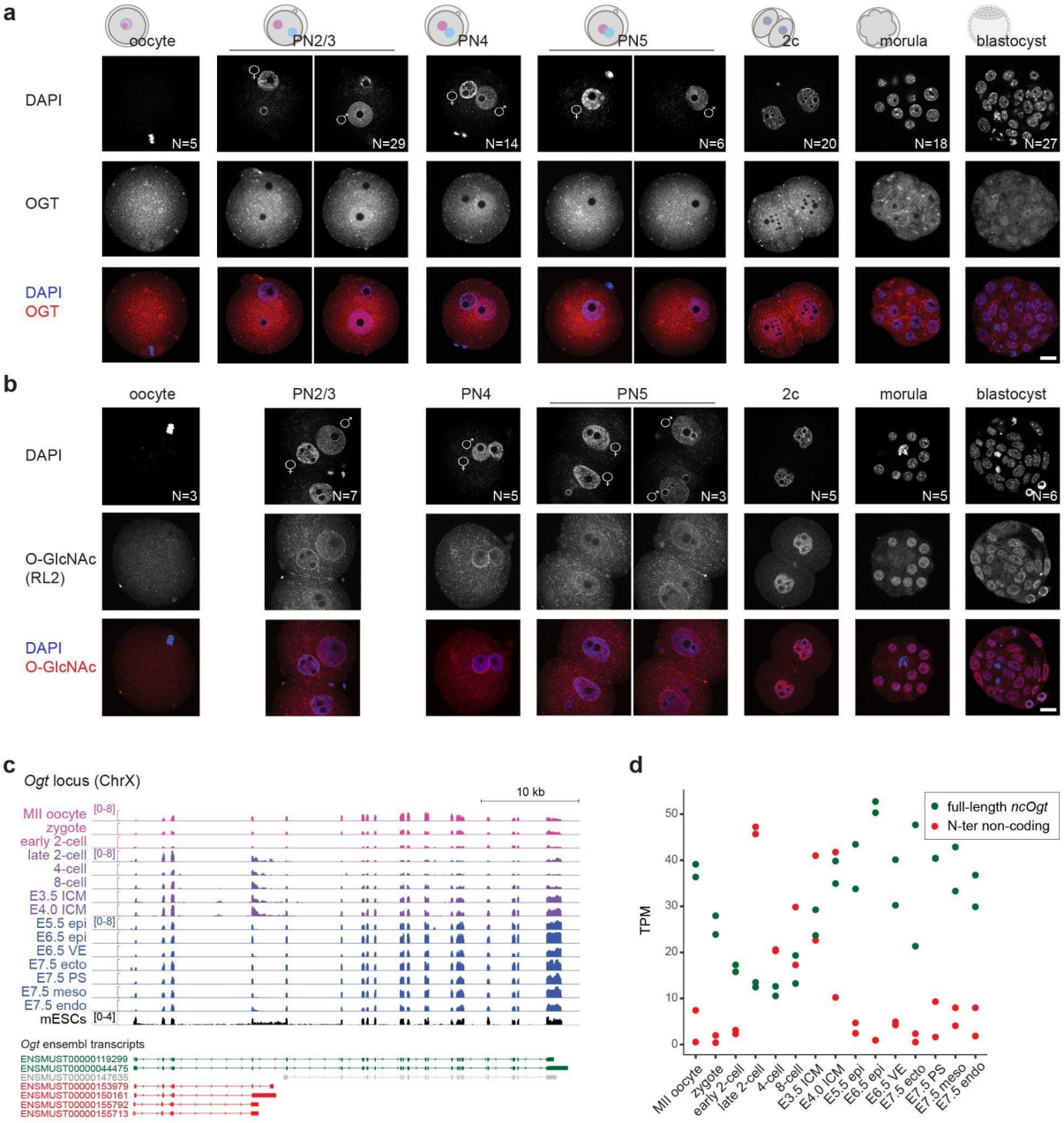


Figure 4: Profiling of the O-GlcNAc modification throughout preimplantation development.

(a) Immunofluorescence staining of the OGT protein (ab177941) in MII oocytes, zygotes at various pronuclear stages, 2-cell embryos (22-26 h post-IVF), morulae (48 h post-IVF) and blastocysts (68 h post-IVF) generated through IVF.

(b) Immunofluorescence staining of the O-GlcNAc modification (RL2 antibody) in the same preimplantation stages as in (a).

(a-b) Embryos are mounted on coverslips and imaged using a Nikon AX scanning confocal. One z plane is shown. Scale bar indicates 20 µm. PN = pronuclear stage. DNA is stained with DAPI. The total number of embryos imaged is indicated for each stage.

(c) Genome browser view of the *Ogt* gene showing the alignment of mRNA-Seq (Smart-Seq) reads from GSE66582 (Wu et al., 2016) (MII oocyte to E4.0 ICM), GSE76505 (Zhang et al., 2017) (E5.5 to E7.5) and E14 mESCs cultured in the lab. Read counts are normalized to Transcript Per Million (TPM). One biological replicate per stage is shown.

(d) Salmon quantification of the *Ogt* transcript isoforms from the same datasets as in (c). Because mapping of reads to a specific long or shorter-N-terminus isoform is not precise enough, TPM of isoforms ENSMUST00000119299 and ENSMUST00000044475 were averaged (full-length *ncOgt*), as well as TPM of ENSMUST00000153979, ENSMUST00000150161, ENSMUST00000155792 and ENSMUST00000155713 (N-ter non-coding). ENSMUST00000147635 is excluded from the plot.

2) Different strategies for the inducible and fast perturbation of O-GlcNAc in the live early embryo

The peri-implantation embryonic lethality caused by maternal inheritance of a loss-of-function *Ogt* gene (O'Donnell et al., 2004a) implies OGT or O-GlcNAc being required for an essential process happening during oocyte maturation or in the embryo before a functional OGT is expressed from the embryonic genome, which happens around the 4-/8-cell stage according to my analysis (Fig.4c-d). Driven by this and by the presence of O-GlcNAc on RNA PolII (Kelly et al., 1993) and many transcription factors (Jackson and Tjian, 1988; Jang et al., 2012; Wulff-Fuentes et al., 2021a), I aimed to uncover the role of the O-GlcNAc modification for embryonic genome activation and in gene expression regulation throughout mouse preimplantation development.

In order to address this important question, I had to overcome two obstacles. The first is the embryonic lethality of maternal *Ogt* loss-of-function (LOF), resulting in the impossibility to maintain even an heterozygous *Ogt*-null mouse line from which to obtain embryos for experiments. The second is that a conditional LOF of *Ogt* in the early embryo would not eliminate the maternal stock of *Ogt* mRNA molecules and OGT proteins. On the other hand, an oocyte-specific disruption of *Ogt* could lead to confounding effects of oocyte maturation defects or impacted fertilization capacity of such oocytes.

To overcome these issues, I implemented two different strategies to achieve an inducible and fast perturbation of OGT and/or O-GlcNAcylation in the early embryo growing *ex-vivo*:

1. A mouse model bearing the endogenous *Ogt* fused to the AID degron, which allows the fast depletion of the OGT protein itself via proteasomal degradation (Results section 2.1). From this mouse line, it should be possible to produce zygotes and achieve in a few hours the depletion of the OGT protein (including the maternal payload) from the early embryo.
2. An enzymatic approach to directly remove the O-GlcNAc modification from the preimplantation embryo by overexpressing a recombinant O-GlcNAcase via mRNA injection of the zygote (Results section 2.2).

Using inducible and fast approaches has the additional advantage that it allows me to study the molecular phenotype very close in time to the start of the perturbation. Hence, I have a higher chance to characterize the primary effects caused by the removal of O-GlcNAc/OGT, before

the cell starts to accumulate secondary effects which I suspect based on the final lethality induced in embryos and cell lines by *Ogt* disruption.

Both of the two systems have advantages and limitations and they can be used to address a different set of questions:

1. By achieving the degradation of the OGT protein, the degron system can be used to address the role of asymmetric presence of OGT in the two zygotic pronuclei (Fig.4a) and any other potential role of OGT protein that is independent of O-GlcNAcylation. A first limitation of this system is that it does not affect the maternal payload of O-GlcNAcylated proteins, which will need a certain time before being diluted by the action of endogenous OGA and by physiological protein turnover. Another limitation is that the degron removes OGT from the whole cell. This could cause cellular phenotypes (such as a cell cycle block) which could impede the study of gene expression changes due to O-GlcNAc perturbation.
2. O-GlcNAc removal using the recombinant OGA has the benefit to remove both embryonic and maternal O-GlcNAc moieties and to do so in a fast manner, with no delay from the moment that such OGA is expressed. OGA expression should be detected a few hours after its mRNA is injected. The second advantage consists in addressing the role of the O-GlcNAc modification *per se*, without disturbing the *Ogt* gene or protein, therefore achieving a separation between catalitically-dependent and catalitically-independent functions of OGT. The third advantage is that the recombinant protein can be targeted to a specific compartment, which in my case will be the nucleus, since I want to focus my investigation on O-GlcNAc-mediated regulation of transcription. The limitation of this approach is that it cannot address the role of OGT protein itself, for example in the zygotic pronuclei showing OGT asymmetric signal.

In the following sections, I will describe each method and all results I obtained with them.

2.1 A mouse model for auxin-inducible degradation of endogenous OGT

As a system for fast depletion of the OGT protein itself, I decided to leverage the auxin-inducible degron (AID) technology, which is based on the transplantation of a plant-specific degradation pathway controlled by a phytohormone, auxin (indole-3-acetic acid; IAA), into mammalian cells (Nishimura et al., 2009). In detail, for this system to work in the cell, on one side the AID degron has to be fused to the protein of interest and, on the other side, the plant TIR1 protein has to be expressed. TIR1 is an F-box protein, which binds to a few mammalian endogenous factors and arranges a functional SCF-type ubiquitin ligase (E3) which specifically targets AID (Gray et al., 1999). AID polyubiquitination and rapid degradation is triggered by the addition of auxin to the culture medium. I opted for the AID degron system because at the time of design of the mouse model it was one of the best degron systems available in terms of size of the degron peptide and in terms of rapidity of protein degradation: a 10-20' half-life had been measured for RAD21 fused to AID after the addition of auxin (Natsume et al., 2016).

I asked the Gene Editing and Embryology Facility (GEEF) at EMBL Rome to create a mouse model bearing the endogenous *Ogt* gene fused with the AID degron at the N-terminus of the longest nucleocytoplasmic RNA isoform (full-length *ncOgt*) ($Ogt^{NterAID\text{-}MYC\text{-}FLAG}$ allele, henceforth Ogt^{AID}) (Fig.5a). In order to limit the size of the added peptide, I chose a minimal version of AID (44 amino acids) (Morawska and Ulrich, 2013; Nora et al., 2017) shown to be functional in mammalian cells (Nora et al., 2017). For this mouse model, only heterozygous Ogt^{AID} females ($Ogt^{AID/WT}$) were viable at birth; they are also healthy and fertile and transmit the mutation at the expected ratio, but mutant males were never found at birth. It is important to keep in mind that males contain only one copy of the X-linked *Ogt* gene, therefore they are hemizygous for any mutation introduced to *Ogt*. By dissecting E7.5 embryos, I confirmed that both female and male embryos express the AID-OGT protein (Fig.5b). One possible explanation for the lethality of hemizygous male embryos $Ogt^{AID/Y}$ could be the lower expression of AID-OGT compared to the wild type protein (Fig.5d,e). This is due to the mere addition of the AID tag, since all other components of the AID system (OsTIR1 and auxin) are still absent. The level of O-GlcNAcylation is also visibly dampened in $Ogt^{AID/Y}$ cells derived from those embryos also in the absence of auxin (Fig.5f). Nonetheless, I found morphologically normal hemizygous $Ogt^{AID/Y}$ males at E12.5, therefore I tested the degron system on male Mouse Embryonic Fibroblasts (MEFs), which are derived from this embryonic stage.

2.1.1 Rapid degradation of endogenous OGT in primary MEFs

The *Gt(ROSA)26*^{OsTIR} mouse line was already available at EMBL Rome and kindly donated to our group by the Lancrin group. This mouse model expresses the *OsTIR1* gene under the promoter of the *Rosa26* gene, which is constitutively and ubiquitously expressed in mice (Soriano, 1999). In order to reconstitute the full AID system, I bred *Ogt*$^{AID/WT}$ females with *OsTIR*-homozygous males of the *Gt(ROSA)26*^{OsTIR} mouse line. From this breeding, I derived male MEFs clones containing both mutations (*OsTIR*,*Ogt*^{AID}) and clones only containing the *OsTIR* transgene as a control. I chose to use male cells because they have only one copy of *Ogt*, either the wild type or that bearing AID. From the same breeding, female progeny is either *Ogt*$^{WT/WT}$ or *Ogt*$^{AID/WT}$, therefore auxin treatment would leave intact the wild type OGT protein molecules which are also present in the cell.
Auxin (500 μM) reproducibly reduced the level of AID-OGT to 20% of the wild type already after 30' minutes of treatment (Fig.5d,e). The strong reduction of OGT endogenous levels is maintained at least until day 4 (Fig.S5a). The degradation of OGT results in a global decrease of O-GlcNAc levels from 24 hours onwards (Fig.5f).

Even though the creation of the *Ogt*^{AID} mouse allele was aimed to perform auxin-inducible OGT depletion in embryos, I leveraged this system to get a first insight into the molecular effect of OGT depletion in the MEFs *in vitro* system.

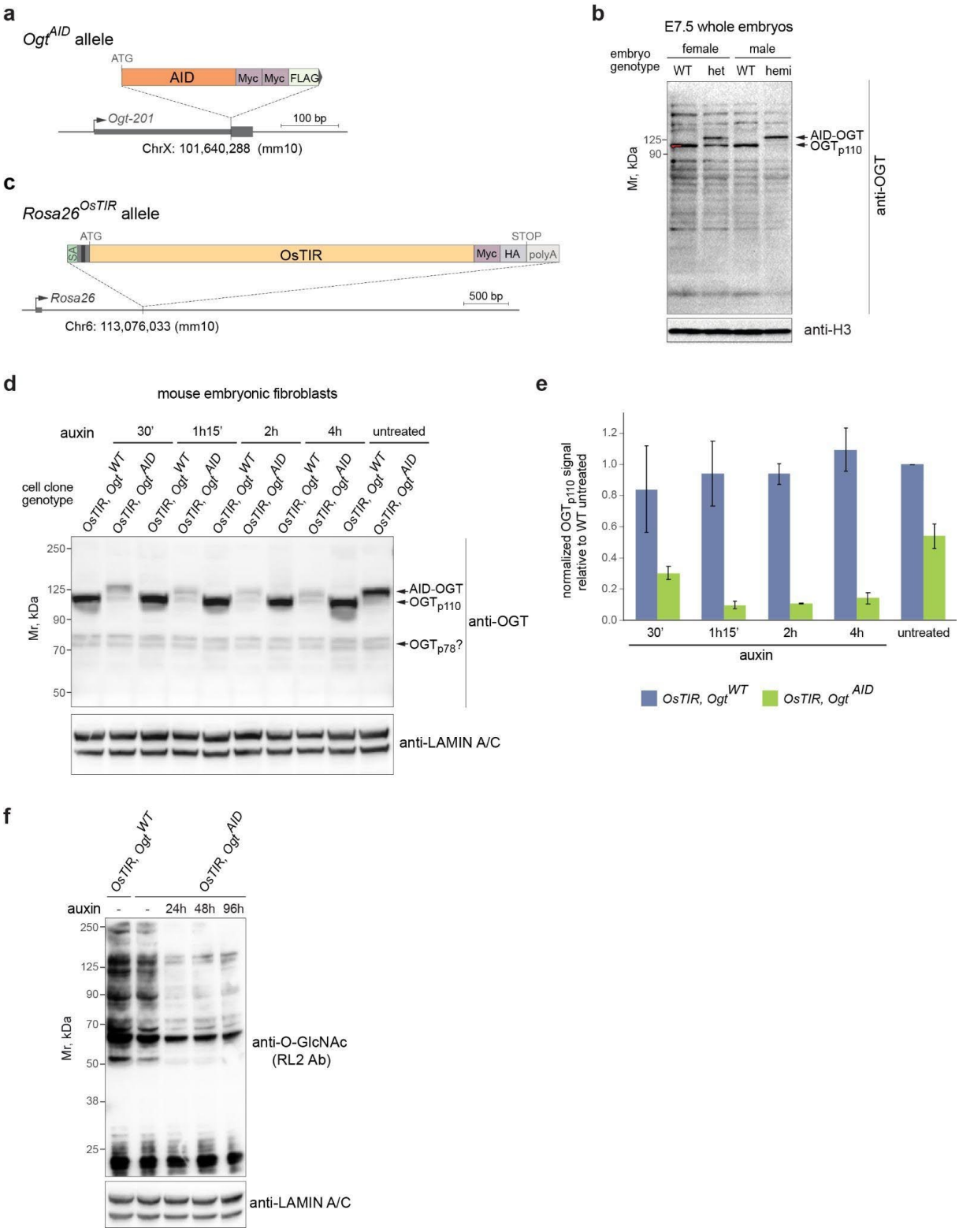


Figure 5: A novel *Ogt*^{AID} mouse allele for fast and inducible OGT depletion proves to be effective in MEFs.

(a) The *Ogt*^{AID} allele: the AID peptide is fused to the N-terminus of the longest *Ogt* isoform (*Ogt-201* or ENSMUST00000044475 or *ncOgt*), together with two Myc tags and one FLAG tag sequence for eventual targeted biochemical pull-down assays. Of note, anti-FLAG and anti-MYC antibodies did not

give a specific signal for AID-OGT when tested using WB or IF in embryos, hence they were not used for any following WB or IF experiment.
(b) Western blot of E7.5 embryos derived from $Ogt^{AID/WT}$ x $Ogt^{WT/Y}$ breedings, showing the expression of AID-OGT in both female heterozygous (het) ($Ogt^{AID/WT}$) and male hemizygous (hemi) ($Ogt^{AID/Y}$) embryos. The molecular mass of ncOGT is 116 kDa, which corresponds to the main band at ~110 kDa. AID-OGT is distinguished from OGT thanks to the size shift due to the addition of AID-2xMyc-FLAG peptide, which is ~8kDA.
(c) The $Rosa26^{OsTIR}$ allele, consisting of the insertion of the *OsTIR1* plant transgene (together with a C-terminal Myc and HA tag) under the promoter of the constitutively and ubiquitously expressed *Rosa26* gene. The transgene expression from the endogenous promoter is possible thanks to its integration at the splicing acceptor (SA) site of *Rosa26*. *Rosa26* expression gets interrupted by the insertion since the resulting transcript is polyadenylated at the polyA site encoded in the insert. Protein translation of the resulting transcripts starts at *OsTIR1* ATG and ends at the stop codon (STOP) encoded at the end of the insert.
The AID degron system needs the contemporary presence of both alleles in (a) and (b) to trigger OGT depletion in the presence of auxin.
(d) Western blot and relative quantification (e) of OGT (ab177941 antibody) in male primary mouse embryonic fibroblasts (MEFs) derived E12.5 embryos from breeding $Ogt^{AID/WT}$ females with *OsTIR*-homozygous males, untreated or treated with the auxin compound for different amounts of time. From the same litter, an *OsTIR*,Ogt^{AID} clone and an *OsTIR*,Ogt^{WT} clone as a control were treated in parallel. A ~90% OGT depletion is achieved after 1 h, but is already visible after 30', and stays stable until 4 h. Barplot heights and error bars show the mean and standard deviation, respectively, of two replicates of the auxin treatment.
(f) Western blot of O-GlcNAc (RL2 antibody) from untreated and auxin-treated *OsTIR*,Ogt^{WT} and *OsTIR*,Ogt^{AID} MEFs, showing the decrease in pan-O-GlcNAc signal in the *OsTIR*,Ogt^{AID} clone after 24 h of treatment and the persistence of the reduction to 4 d (last time point tested).
Note that both in (e) and (f) it emerges that the AID-OGT protein is already ~50% less abundant than wild type OGT independently of auxin (e) and, accordingly, the O-GlcNAc level in untreated *OsTIR*,Ogt^{AID} MEFs is already reduced compared to the *OsTIR*,Ogt^{WT} control (f).

2.1.2 Time-course transcriptomics analysis following OGT depletion in MEFs

In order to probe the effect of OGT acute depletion on gene expression, I performed mRNA-sequencing (mRNA-Seq) of three different clones of *OsTIR*,Ogt^{AID} MEFs treated with auxin for 24, 48 and 96 hours and compared them with the same clones grown in parallel in the absence of auxin. To control for the effect of the auxin drug *per se* ("auxin effect" from now on), I used three littermate *OsTIR*,Ogt^{WT} clones, also treated with auxin for 24, 48, and 96 hours and compared with the untreated condition. Additionally, after observing that the level of the AID-OGT protein is ~50% lower than wild type OGT (Fig.5d,e), I compared gene expression of the two genotypes (*OsTIR*,Ogt^{AID} versus *OsTIR*,Ogt^{WT}) between each other in the absence of auxin ("genotype effect" from now on).

Firstly, auxin effect is small but not negligible (Fig.6e), since it leads to downregulation of some metabolic pathways (such as sterol metabolism) (Fig.6g), which could potentially be confounded with an effect of O-GlcNAc depletion. This reinforces the importance of using a genetic control whereby control cells are treated but lack the AID tag (e.g. *OsTIR*,Ogt^{WT}). I excluded any gene changing at any time point after auxin addition in the control clones from the genes used for the gene ontology (GO) analysis of all the other comparisons.

OGT and O-GlcNAc levels are tightly coupled with OGA levels through both transcriptional and posttranslational mechanisms aiming at maintaining O-GlcNAc homeostasis (Park et al., 2017; Pravata et al., 2019b; Qian et al., 2018; Vaidyanathan et al., 2017b; Willems et al., 2017b; Z. Zhang et al., 2014). Hence, a change in *Ogt* and *Oga* expressions could represent the cellular compensation to decreased O-GlcNAc levels. In *OsTIR*,Ogt^{AID} clones, *Oga* was significantly downregulated at all time points after auxin addition compared to untreated cells, while *Ogt* was always upregulated (Fig.S2b). *Oga* level was already downregulated when comparing the *OsTIR*,Ogt^{AID} genotype with the *OsTIR*,Ogt^{WT} one, independently of auxin. This is in agreement with the western blot result showing lower levels of O-GlcNAc in *OsTIR*,Ogt^{AID} clones before auxin addition (Fig.5f) and confirms that *OsTIR*,Ogt^{AID} is *de facto* an hypomorphic *Ogt* mutant. Curiously, *OsTIR1* is also significantly downregulated after auxin addition (Fig.S2b), suggesting that the cell also uses completely unknown alternative mechanisms in order to maintain a constant O-GlcNAc concentration.

The first non-trivial result of the mRNA-Seq is the transcriptional phenotype overall consisting of low-magnitude changes in expression, with most of the $\log_2$ fold changes (log2FC) below 1 in both directions (Fig.6a-e). This result is somewhat surprisingly considering that *Ogt* is an essential gene. It is possible that the hypomorphic phenotype of untreated *OsTIR*,Ogt^{AID} clones could be partly responsible for the overall mild changes in transcription observed in the same clones after acute depletion. On the other hand, gene expression changes are widespread, with hundreds of significantly deregulated genes and a slightly more pronounced gene expression downregulation (Fig.6a-d,f). As expected, the differentially expressed genes (DEGs; defined as adjusted p-value<0.05, any log2FC) in *OsTIR*,Ogt^{AID} cells at the three time points overlap significantly (Fisher exact test's p-value$<<1e^{-10}$ for all comparisons between time points; Fig.6f).

Gene ontology (GO) over-representation test of the DEGs revealed very few significantly upregulated GO terms, but many commonly downregulated terms between all time points of OGT depletion, with statistical significance or number of DEGs belonging to that gene set increasing from 24 h to 96 h of acute OGT depletion (Fig.6g and S2c). Some of the most significant commonly downregulated terms revolve around actin cytoskeleton function (*actin filament organization*, *contractile fiber*, *ruffle* GOs) and include genes that are part of the actomyosin complex (e.g. *Cnn1*, *Myl9*) as well as transforming growth factor beta 3 (*Tgfb3*), participating in mesenchymal differentiation (Grafe et al., 2017), and its receptor *Tgfbr1*. Of note, *actin filament organization* and *muscle cell differentiation* GO terms are already present when comparing untreated *OsTIR*,Ogt^{AID} versus untreated *OsTIR*,Ogt^{WT} cells, in line with the former representing a hypomorphic OGT condition. Other downregulated genes common to all time points after auxin addition participate in intracellular membrane trafficking (e.g. *endoplasmic reticulum to Golgi vesicle-mediated transport* and *establishment of protein localization*). The presence of these terms could potentially be linked to glycosylation pathways other than O-GlcNAc being perturbed as well, as shown by the *glycoprotein metabolic process* among the most significantly downregulated GO terms in auxin-treated *OsTIR*,Ogt^{AID} at any of the three time points (Fig.6g). The significantly downregulated genes include Ribitol Xylosyltransferase 1 (*Rxytl1*) and ALG14 UDP-N-Acetylglucosaminyltransferase subunit (*Alg14*), the enzyme responsible for N-glycosylation. Both of these enzymes participate in the biosynthesis of branched glycans, which happens in the endoplasmic reticulum (ER) and Golgi apparatus for their successive exposure on the cell membrane (Varki et al., 2022). Indeed, UDP-GlcNAc is the common substrate for both OGT and branched O- and N-glycosylations (Varki et al., 2022). It is possible that the increase in free UDP-GlcNAc in the absence of OGT could re-enter the biosynthesis of branched glycans and thus lead to their misregulation. The cell could then respond by downregulating branched glycans' biosynthetic enzymes.

Besides commonalities among all time points, I noticed that changes observed at the 48 h and 96 h time points are very similar, whereas at 24h there are unique events, such as the upregulation of the proteasome complex and the mitochondrial ribosomal subunits. This could reflect an early transcriptional response to OGT depletion. At later time points this upregulation disappears and the proteasome starts to be downregulated. On the other hand, a downregulated gene set appearing from later time points is the cytosolic ribosome (Fig.S2c), including many ribosomal proteins (e.g. *Rpl7*, *Rps23*, *Rpl15*, *Rpl8*, *Rplp0*, *Rpl10*), translation factors (*Eif3e*, *Ddx3x*, *Eif3j2*, *Eif3a*) and aminoacyl-tRNA ligases (*Tarsl2*, *Lars*, *Qars*, *Aars*, *Eprs*, *Nars*, *Yars*).

All together these genes suggest that protein translation could be negatively affected by the loss of OGT.

The presence of proteins participating in mesenchymal differentiation is widespread across all GO results. Such proteins are mostly studied in the context of the epithelial-to-mesenchymal (EMT) signaling cascade, which is observed both during gastrulation and tissue morphogenesis but also in cancer cells. EMT is a reversible cellular transformation process characterized by changes in cell junctions, with a switch between intercellular contacts mediated by E-cadherin (CDH1) to contacts with the extracellular matrix driven by N-cadherin (CDH2) and massive cytoskeletal rearrangements. The main molecule responsible for driving EMT during development is transforming growth factor beta (TGF-β), which triggers SMAD-mediated signaling cascade. The latter activates transcription factors (such as the *Snail* or *Zeb* class of TFs) which induce the expression of mesenchymal genes such *Cdh2*, and the repression of epithelial genes such as *Cdh1*, also by recruiting chromatin remodelers such as histone deacetylases HDAC1 and HDAC2 (Grafe et al., 2017). SMAD signaling is complemented by other signaling cascades triggered by TGF-β, such as ERK/MAP kinases (ERK/MAPK), RHO GTPases and PI3 kinase/AKT (PI3K/AKT) pathways (Grafe et al., 2017). Noteworthy, the PI3K/AKT pathway is responsible for activation of the mTORC1 pathway, which orchestrates an increase in translation both via posttranslational modification of the ribosome as well as by activating transcription factors that are able to modulate the expression of ribosomal genes (Gentilella et al., 2015).
In my dataset, the significantly downregulated genes involved in the mesenchymal transition include *Tgfb3* and its receptor *Tgfbr1,* Rho GTPases (*Rhoa*, *Cdc42*, *Rac1*), the histone deacetylase complex (*Hdac1*, *Hdac2* and their cofactor Rbbp7 are all significantly downregulated) and actin-binding proteins. However, there is no upregulation of epithelial pathways. This suggests that OGT depletion impacts the mesenchymal gene network but does not cause any reverse mesenchymal-to-epithelial transition. Importantly, the mesenchymal network is the default differentiation one active in MEFs.

In order to get more insights into the pathways affected by OGT depletion, I performed an analysis of enrichment of molecular signatures using the MsigDB database (Fig.6h). At the 96 h time point, this showed that targets of mTORC1 and the transcription factor c-Myc are enriched among downregulated genes. Targets of the tumor suppressor protein p53 are also enriched, but they are both up- and downregulated, making an interpretation more difficult. In addition, the

Cdkn1a gene, the canonical p53 transcriptional target also known as p21, is not differentially expressed at any time point. The inactivation of mTORC1, which could be driven by the reduction in TGF-β signaling or by other mechanisms, could be partly responsible for the downregulation of genes of the translational machinery. A more direct mechanism leading to decreased expression of genes involved in translation could involve the inactivation of c-Myc, a renown oncogene stimulating cell growth (Dang, 2013). c-Myc targets which are downregulated after OGT depletion include proteasomal subunits, ribosomal proteins, translation initiation factors and *Hdac2* (Fig.6h). Notably, c-Myc targets are already significantly enriched after 48 h of OGT depletion, while the enrichment of mTORC1 signaling appears only later.

In summary, I showed that our novel *Ogt*-degron mouse allele is efficient in achieving fast OGT depletion in primary MEFs, with a strong reduction in O-GlcNAc levels already after 24 h. This resulted in the accumulation over time of transcriptional downregulation affecting the default MEFs' mesenchymal network as well as a gene network involved in translation.

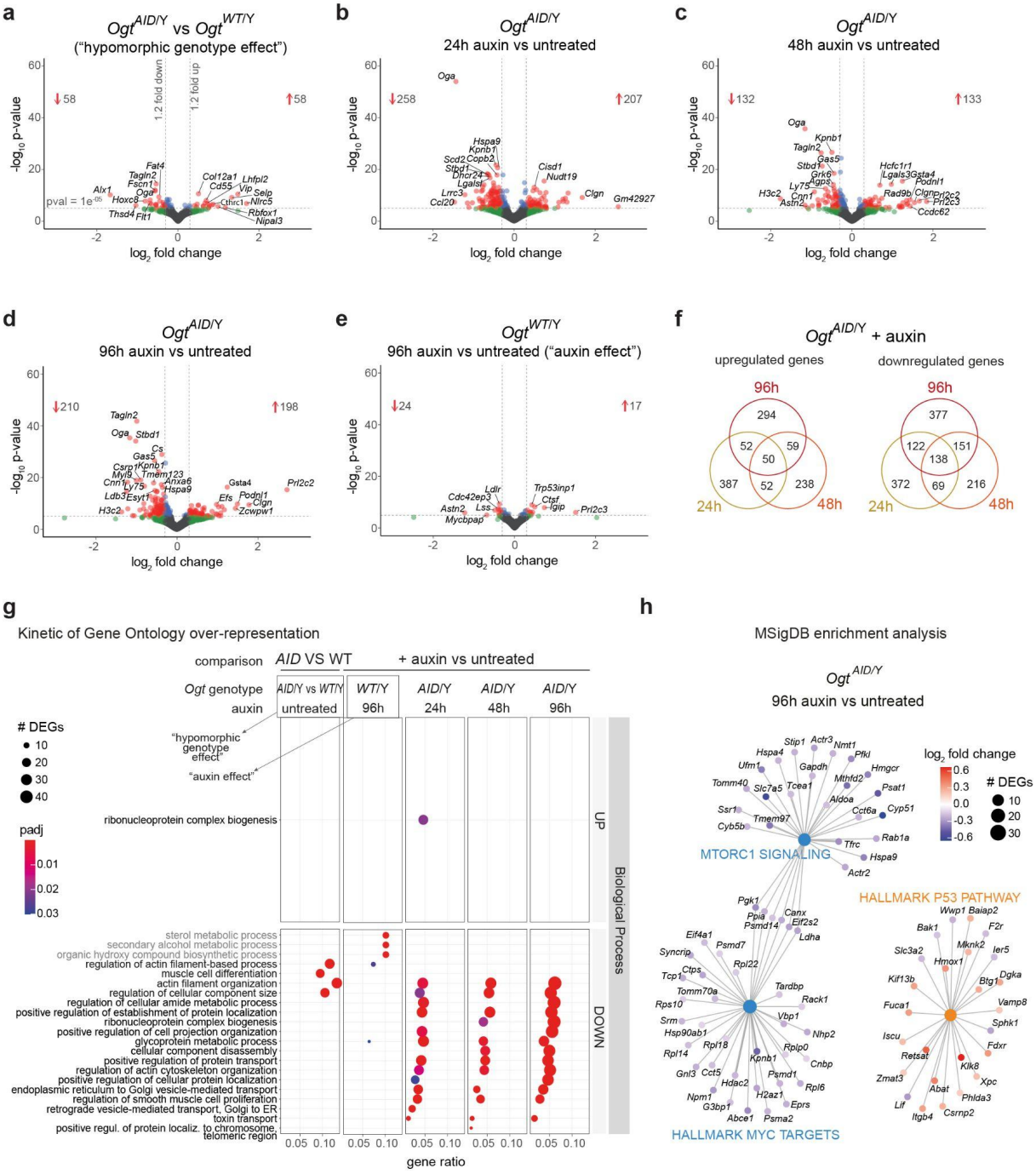


Figure 6: Time-course transcriptional response to OGT depletion in MEFs.

(a-e) Volcano plots from differential expression analysis performed with DESeq2 for the different comparisons: (a) untreated *OsTIR*,Ogt^{AID} versus untreated *OsTIR*,Ogt^{WT} clones ("hypomorphic genotype effect"), (b-d) *OsTIR*,Ogt^{AID} clones after 24 h, 48 h and 96 h of auxin treatment compared to untreated *OsTIR*,Ogt^{AID} clones grown in parallel for the same amount of time and (e) *OsTIR*,Ogt^{WT} control clones treated with auxin for 96 h versus grown untreated for 96 h ("auxin effect") (e). The number of genes with p-value < 10^{-5} and absolute log2FC > 0.3 (i.e. 1.2X increase or decrease in expression) is indicated. Genes that stand out are labeled.

(f) Number of differentially expressed genes (DEGs) (adj. p-value < 0.05, any log2FC) in auxin-treated *OsTIR*,Ogt^{AID} clones versus untreated *OsTIR*,Ogt^{AID} clones at the three time points.

(g) Gene ontology (GO) over-representation analysis of gene expression changes for the same comparisons as in (a-e). The first 25 most enriched Biological Process GO terms are shown, based on p-value across all comparisons. Terms are ordered by gene ratio. Gene ratio = genes belonging to the GO term / total number of up- or downregulated DEGs for that comparison. UP = upregulated DEGs, DOWN = downregulated DEGs. Terms which are enriched due to auxin treatment on control clones are colored in gray.
(h) Analysis of enrichment for Hallmark gene sets of the Molecular Signatures Database (MSigDB) among DEGs after 96 h of OGT depletion.
For (g) and (h), DEGs after OGT depletion which are also differentially expressed in the auxin-treated control cells are excluded from the GO and MDigDb enrichment analyses of auxin-treated Ogt^{AID} clones; DEGs with mean DESeq2-normalized counts < 10 across all samples are excluded from all enrichment analyses.

2.1.3 Seemingly inefficient auxin-induced degradation of endogenous OGT in early embryos

We created the mouse model of *Ogt*-degron with the goal of probing the effect of acute OGT depletion on the developing early embryo. Encouraged by the efficiency of the system *in vitro* in MEFs, I tested whether it was able to achieve auxin-inducible OGT degradation in live embryos grown *ex vivo*.

Firstly, I tested auxin toxicity in embryos from in vitro fertilization (IVF) of wild type FVB oocytes and sperm. I grew the embryos in different concentrations of auxin from the time of fertilization, by adding auxin to the fertilization drop (where oocytes and sperm are placed together), to the blastocyst stage (E4). By reducing auxin to half of the concentration used in MEFs (250 μM), 79% of the zygotes developed to expanded blastocyst, the same percentage as in the untreated condition (78%) (Fig.S2d). Based on this result, I chose the 250 μM auxin concentration for all my next experiments.

Next, I tested whether auxin treatment was able to induce OGT depletion *ex vivo* by immunostaining of the O-GlcNAc modification. Before testing O-GlcNAc depletion in embryos, I used a longer time of auxin treatment than in MEFs because I was expecting a slower rate of OGT degradation with the lower concentration of auxin used, as it was previously shown in yeast (Morawska and Ulrich, 2013). I performed IVF using oocytes from either *OsTIR*,$Ogt^{AID/WT}$ females or control *OsTIR*,$Ogt^{WT/WT}$ siblings with the same wild type sperm and cultured the resulting embryos in auxin from the time of fertilization to the morula stage (E2.5),

when I fixed and stained them with an anti-O-GlcNAc antibody (RL2 clone). The morulae from the *Ogt*$^{AID/WT}$ mothers were a mixture of four genotypes: females *OsTIR*,*Ogt*$^{WT/WT}$ and *OsTIR*,*Ogt*$^{AID/WT}$, males *OsTIR*,*Ogt*$^{WT/Y}$ and *OsTIR*,*Ogt*$^{AID/Y}$. I knew from previous single-embryo genotyping experiments that the rate of formation of expanded blastocyst from zygotes grown in auxin was not significantly different when using *OsTIR*,*Ogt*$^{AID/WT}$ oocytes compared to control oocytes (Fig.S2e), implying that the treatment or, if any, the OGT depletion is not lethal to any of the genotypes. It follows that in the immunofluorescence experiment at the morula stage all four possible embryonic genotypes should be present with the same percentage in both groups, thus ~25%. In case of an efficient OGT depletion, I expected that at least the *OsTIR*,*Ogt*$^{AID/Y}$ males (~25% of embryos), which only contain the AID-tagged form of OGT would show a detectable lower level of O-GlcNAc. Contrary to this prediction, there was no group of outlier embryos in the group of embryos from the *OsTIR*,*Ogt*$^{AID/WT}$ mothers when I quantified the O-GlcNAc signal of the single embryos (Fig.7a-b). These data show that in early embryos the auxin-inducible AID-OGT depletion system is not efficient enough to cause a detectable difference in O-GlcNAc levels by immunostaining.

In spite of the imaging result, I could not exclude a low level of auxin-induced OGT degradation, thus I investigated whether a molecular difference was emerging in auxin-treated *OsTIR*,*Ogt*$^{AID/Y}$ embryos by means of the more sensitive transcriptomics approach. An auxin-dependent effect on gene expression of the hemizygous males would be an indirect indication of some level of OGT depletion.

For this experiment, I bred *Ogt*$^{AID/WT}$ females with *OsTIR*-homozygous males by natural mating, collected the oocytes the day after and grew them for 3 days in normal culture conditions. At E3.5, I divided the morulae from the two types of breedings in two groups and treated one with 250 μM auxin for 24 h, then collected the blastocysts for single embryo mRNA-Seq (Smart-Seq) (Picelli et al., 2014). Like in the imaging experiment, the two groups of blastocysts contained all possible genotypes: females *OsTIR*,*Ogt*$^{WT/WT}$ and *OsTIR*,*Ogt*$^{AID/WT}$, males *OsTIR*,*Ogt*$^{WT/Y}$ and *OsTIR*,*Ogt*$^{AID/Y}$. This allows to sort out the effect of OGT depletion from the effect of the auxin treatment *per se* and from the effect of the hypomorphic *Ogt*^{AID} allele (independently of auxin) using a combination of differential gene expression analyses. Firstly, I assigned the sex and genotype to the embryos *in silico* (see Methods) and performed PCA (Fig.7c-d). Notably, a slight separation between the *Ogt*^{WT} and *Ogt*^{AID} genotypes appears only in males and only in the presence of auxin. Since *Ogt*^{AID} males do not bear any wild type *Ogt* copy, they are the ones expected to be affected by AID-OGT depletion. Hence, the male-

specific separation induced in the PCA by auxin treatment is an indirect indication that the degron system impacts OGT's function upon auxin addition. Accordingly, significant gene expression changes are almost only observed when comparing the two genotypes of auxin-treated males (Fig.7f among 7e-h).

Notwithstanding the overall mild gene expression change after AID-OGT depletion in embryos, the most significantly deregulated genes might be meaningful. Notably, one of the most upregulated genes is *Dppa3*, also known as *Stella*, which contributes to the DNA demethylation of the maternal genome during the first embryonic cleavages (Li et al., 2018). In mESCs, it has a similar role than in embryos in preventing DNMT1 methylation (Greenberg, 2021), it is a marker of naive pluripotency and it was reported to be part of the network that prevents mESCs' differentiation to the mesoderm lineage (Waghray et al., 2015). Other upregulated genes are *Washc3*, which is part of the WASH complex promoting the nucleation of actin cytoskeleton at the surface of endosomes (Derivery et al., 2009; Waghray et al., 2015), *Snapc5* - an activator of transcription of small nuclear RNAs, the small RNAs involved in splicing (Henry et al., 1998) - and the ribosomal protein *Rpl39l*. The most downregulated genes with reported functions are *Acyl-CoA Synthetase Long Chain Family Member 5* (*Acslg*), an isozyme which acts both in lipid synthesis and degradation via β-oxidation, *Msh Homeobox 2* (*Msh2*), an *Hox* gene which regulate transcription in bone development (Jabs et al., 1993) and *Sphingosine Kinase 2* (*Sphk2*), catalyzing the phosphorylation of sphingosine into sphingosine 1-phosphate. Sphingosine 1-phosphate mediates many cellular processes including migration, proliferation and apoptosis (Mendelson et al., 2013).

In conclusion, I showed via indirect transcriptomics evidence that the auxin system applied to preimplantation development induces some reduction in the OGT level in the blastocyst stage but is not efficient enough for a corresponding significant decrease in the O-GlcNAc signal. The probable reduced level of OGT after auxin-induced degradation is likely impacting O-GlcNAc levels on a limited number of targets and partially. Therefore, the resulting gene expression changes can be meaningful when discussed in the context of the transcriptional effect of O-GlcNAc depletion via Btgh injection, presented in section 2.2.

2.1.4 Possible explanations for the inefficiency of the AID-OGT system applied to the preimplantation embryo

The reasons for the inefficiency of O-GlcNAc depletion after auxin treatment in preimplantation embryos could be multiple. The auxin degron system was previously shown to be efficient in oocytes (Camlin and Evans, 2019) and in blastocysts (Gu et al., 2018) with 500 μM auxin to induce a significant reduction in the AID-tagged protein of interest in a few hours. Therefore, the lower concentration of auxin used in my experiment could be one relevant factor. However, the authors kept the blastocysts in auxin for only a few hours and did not test the toxicity of this concentration if used through the whole preimplantation development. According to my test for toxicity, 500 μM of auxin was detrimental to the development to the blastocyst stage. Another variable is the level of expression of *OsTIR*, which was shown to correlate with the level of depletion of the AID-tagged protein in oocytes, where the authors overexpressed *OsTIR* via mRNA injection of the oocytes (Camlin and Evans, 2019). They showed that 2.2 μg/μl of *OsTIR* mRNA led to a 30% depletion of AID-GFP signal, while 0.5 μg/μl only led to 15% depletion (Camlin and Evans, 2019). I could also implement the strategy of *OsTIR* mRNA injection to test if it can achieve better efficiency of AID-OGT depletion. In the blastocyst, the authors integrated *OsTIR* in the *Actb* locus (Gu et al., 2018), which might induce a higher level of expression than *Rosa26* in the early embryos.

Finally, it is possible that the auxin degron system is in general not suitable to deplete O-GlcNAc in the first embryonic stages. In fact, my analysis of transcript isoforms of *Ogt* across preimplantation stages shows that there is very little transcription, if any, of full-length *ncOgt* until the ~8-cell stage, hence ncOGT proteins mostly come from the maternal payload of ncOGT protein and mRNA (Fig.4c-d). Thus I expect auxin treatment to mostly cause degradation of maternally provided AID-OGT protein until the 8-cell stage. With embryonic cleavages, the blastomeres get smaller, therefore the concentration of maternal OGT could still be enough to sustain new O-GlcNAcylation (hence auxin treatment can result in a visible O-GlcNAc reduction). However, this depends on the rate of degradation of maternal OGT proteins and mRNAs, which is unknown in the early embryo. If this rate is fast enough to cause a decreased cellular concentration of functional ncOGT, O-GlcNAc cellular level would not increase during the first cleavages and the O-GlcNAc signal would be mainly contributed by the maternal payload. It follows that at the morula stage it would be hard to see a decrease in O-GlcNAc signal after auxin-induced degradation of OGT from the zygote. To test this hypothesis, I could probe the O-GlcNAc or OGT signal after auxin treatment in the late

blastocyst - when new embryonic ncOGT is already accumulating - or I could start the treatment in the oocyte (although this might impact the fertilization rate). However, even with this experiment a decrease of signal could be difficult to detect if the depletion is not highly efficient, as suggested by the little amount of gene expression change in the blastocysts after auxin treatment of the morulae (Fig.7h). In addition, if this hypothesis is true, it would mean that the auxin system is not suitable to study the role of O-GlcNAc at earlier preimplantation stages. It might still be useful to test the role of OGT protein in the zygotic pronuclei.

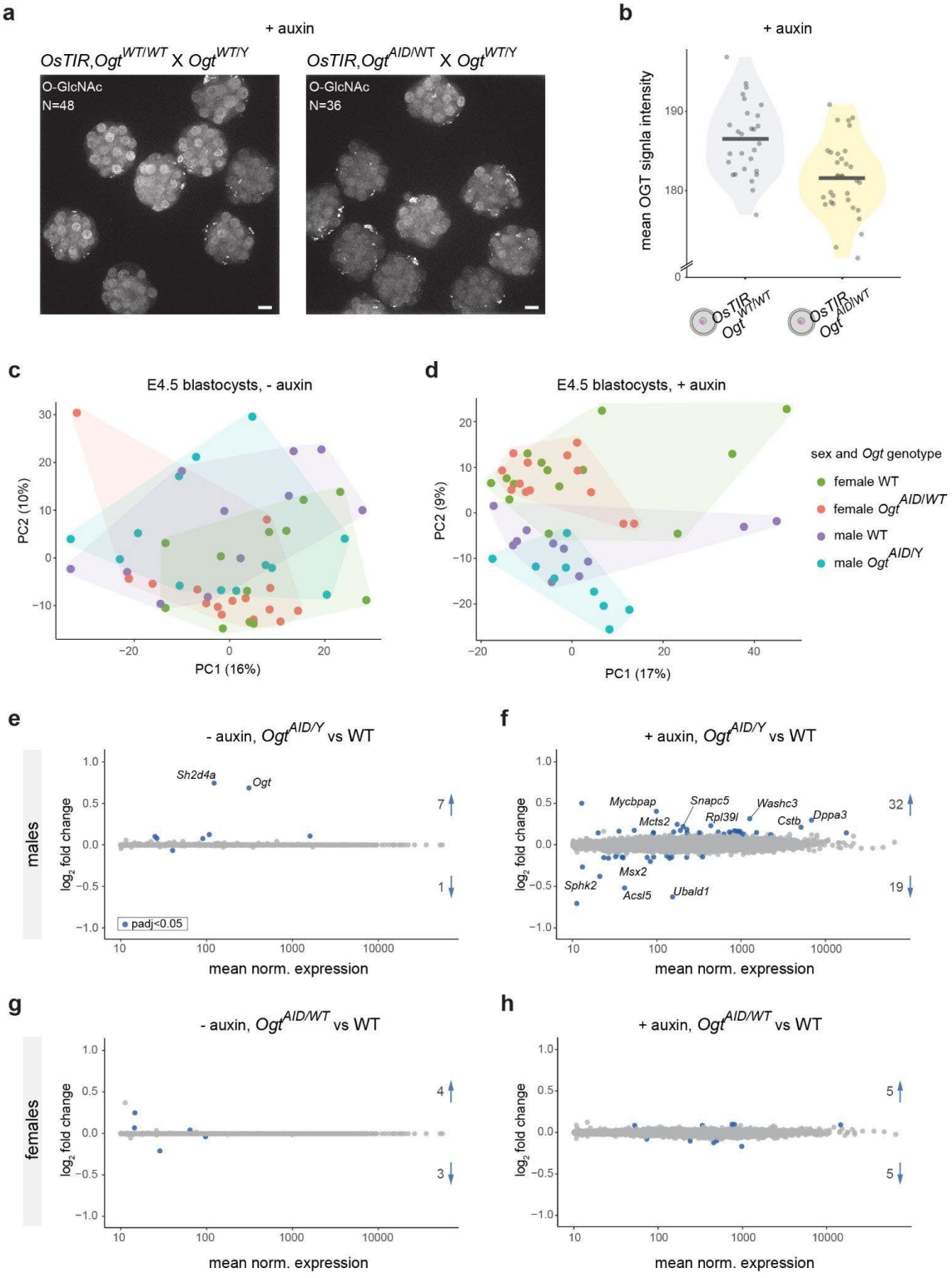


Figure 7: Seemingly inefficient auxin-induced degradation of endogenous OGT in early embryos. (a) Representative images of O-GlcNAc staining (RL2 antibody) in morulae from *OsTIR*,*Ogt*$^{AID/WT}$ and control *OsTIR*,*Ogt*$^{WT/WT}$ oocytes, developed in 250 µM auxin from the time of fertilization. Embryos are mounted in drops and imaged using an X-light V3 Spinning disk confocal. One z plane is shown. Scale bar indicates 20 µm. The total number of imaged embryos is indicated.

(b) Quantification of the O-GlcNAc fluorescence signal from the morulae in (a) as described in Methods. Average signal per group is marked.
(c-d) Principal component analysis (PCA) of E4.5 blastocysts from natural mating of $Ogt^{AID/WT}$ females with *OsTIR*-homozygous males, grown in normal medium or in medium supplemented with 250 μM auxin for the 24 h before collection. Blastocysts were genotyped for sex and *Ogt* genotype based on the RNA-Seq data, as described in Methods. PCA was performed as described in Methods. The variance explained by each principal component is in parentheses.
(e-g) MA-plots of DESeq2 differential gene expression analysis for the Ogt^{AID} versus wild type *Ogt* genotype comparison, separately for males treated and untreated, females treated and untreated. Only genes with mean of DESeq2-normalized counts ≥ 10 are shown. All genes with adj. p-value < 0.05, any Log2FC are colored and their number is indicated. Genes with absolute log2FC ≥ 0.2 are labeled.

2.2 Effective depletion of O-GlcNAc in embryonic nuclei using a recombinant OGA

2.2.1 Experimental design and proof of principle

In order to deplete O-GlcNAc from the nuclei of the live mouse embryo, I exploited the homologous O-GlcNAcase from the human gut symbiont *Bacteroides thetaiotaomicron* (BtGH84), whose structure and catalytic mechanism have been well-characterized *in vitro (Dennis et al., 2006)* and which has been shown to remove O-GlcNAc at targeted mammalian genomic loci when fused to dCas9 (Boulard et al., 2020). The use of this bacterial homologue is suitable for perturbation in mammalian cells for four main reasons. First, it is smaller than the mammalian one (737 residues versus 916 of the mouse and human ones); second, while the active site is remarkably conserved (Dennis et al., 2006), the evolutionary distance of the other domains lowers the chances of interaction with the endogenous interactors of mouse OGA; third, it does not bear the pseudo histone acetyltransferase (HAT) domain of mammalian OGA (Dennis et al., 2006); forth, a single amino acid mutation that makes it catalytically inactive has been found and tested (Boulard et al., 2020; Dennis et al., 2006).

I fused BtGH84 to N-terminal and C-terminal nuclear localization signals (NLS) and to an EGFP reporter and injected the mRNA of this construct (called "Btgh" from now on) to zygotes 2 hours after IVF (~PN3) (Fig.8a). I used two different controls of specificity of the effect of O-GlcNAc removal: the first control is non-injected embryos, the second is the injection of a catalytically inactive construct bearing a single amino acid mutation (D242A) (Dennis et al., 2006) (hence "dBtgh" for "dead Btgh"). The comparison of the phenotypes of the three groups enables to pinpoint the consequence of loss of O-GlcNAc on nuclear proteins, accounting for both the effects due to the injection procedure and potential enzymatic-independent effects of the recombinant BtGH84 protein. I assessed the efficiency and specificity of this method by immunofluorescence detection of O-GlcNAc. After injection of Btgh, the O-GlcNAc signal was dramatically reduced to near undetectable levels at the early 2-cell stage (i.e. before EGA) (Fig.8b-c), while it remained comparable to control (non-injected) levels following dBtgh expression. Because the antibody used for O-GlcNAc staining (clone RL2) could be biased towards specific epitopes such as the NPC, I tested nuclear O-GlcNAc depletion using another anti-O-GlcNAc antibody (clone HGAC85), which shows O-GlcNAc signal at the nuclear membrane and at perinucleolar foci. The HGAC85 signal also disappears after Btgh injection (Fig.S1c). Importantly, O-GlcNAc depletion persists until the late blastocysts stage (E4) - the

last stage I tested by immunostaining - where the absence of the signal is also exclusively restricted to the nucleus (Fig.8d).

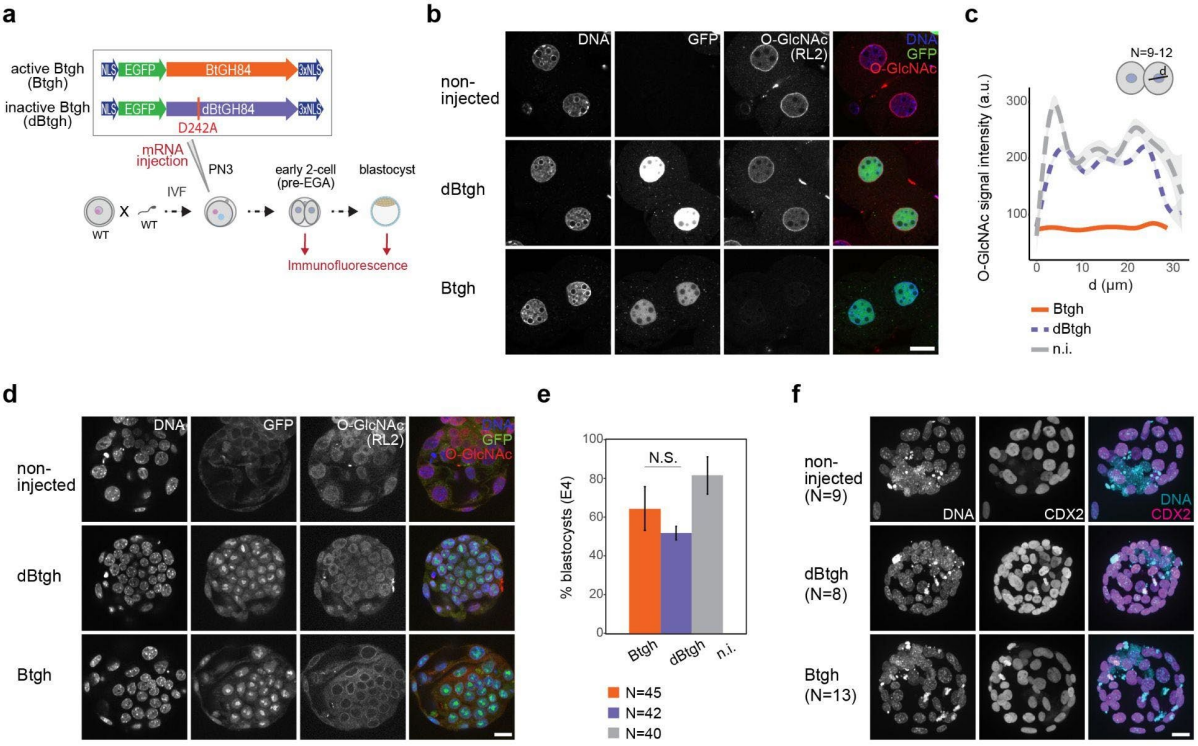


Figure 8: The effective depletion of O-GlcNAc from mouse embryonic nuclei using a recombinant OGA allows normal development to blastocysts.

(a) Experimental design for results in (b-d). Zygotes were generated through IVF and injected 2h post-IVF with the mRNA of the active or inactive form of NLS-EGFP-BtGH84-3xNLS (Btgh and dBtgh, respectively) or left uninjected. The three groups of embryos were cultured *ex-vivo* and fixed both as pre-EGA 2-cell and at the blastocyst stage.

(b) Confocal imaging of pre-EGA 2-cell embryos from zygotes injected with Btgh/dBtgh or non-injected, stained for the O-GlcNAc modification (RL2 antibody).

(c) Quantification of the O-GlcNAc signal across single nuclei of 2-cell embryos from the experiment in (a-b). N=9-12 nuclei per experimental group were quantified, with procedure described in Methods. Quantification was performed from two replicates of the microinjection experiment, which were pooled in one single plot.

(d) Confocal imaging of E4 blastocysts developed from zygotes injected with Btgh/dBtgh or non-injected, stained for the O-GlcNAc modification (RL2 antibody).

(b,d) Embryos are mounted on coverslips and imaged using a Nikon AX scanning confocal. One z plane is shown. Scale bar indicates 20 µm. DNA is stained with DAPI. The GFP signal comes from the Btgh/dBtgh protein.

(e) Percentage of healthy late (E4) blastocysts which developed *ex-vivo* from Btgh/dBtgh injected and non-injected zygotes. Bar heights and error bars indicate the average and standard deviation, respectively, for four replicates of the microinjection experiment. Total number of starting 2-cell embryos is stated in figure legend. The lower developmental rate of both injected groups is expected and due to the injection procedure. N.S = not significant, based on unpaired Student's t-test.

(f) Immunofluorescence staining of OGT and the trophoblast marker CDX2 in blastocysts developed from Btgh/dBtgh injected and non-injected embryos. Embryos are mounted in drops and imaged using an X-light V3 Spinning disk confocal. Max projection of all z planes is shown. Scale bar indicates 20 μm. DNA is stained with DAPI. The total number of imaged blastocysts is indicated.

2.2.2 Normal blastocyst formation but delayed transcriptome of nuclear O-GlcNAc-depleted embryos

I investigated the effect of nuclear O-GlcNAc depletion on the rate of late blastocyst (E4) formation. I found no measurable impact of nuclear O-GlcNAc depletion on *ex vivo* blastocyst developmental rates (Fig.8e). I assessed the specification of the trophectoderm (TE) and inner cell mass (ICM) lineages by staining the TE marker CDX2 (Strumpf et al., 2005). I found CDX2 to be exclusively present in the external layer of cells of the blastocyst (namely the TE), as expected, in all imaged blastocysts (N of expanded blastocysts imaged = Btgh: 13, dBtgh: 8, non-injected: 9) (Fig.8f).
My data show that the development to the blastocyst stage and establishment of the two main cell lineages of the blastocyst can take place with significantly lower levels of nuclear O-GlcNAc.

The presence of O-GlcNAcylation of RNA polymerase II and transcription factors (Jackson and Tjian, 1988; Kelly et al., 1993; Wulff-Fuentes et al., 2021a) supports a role for the O-GlcNAc modification in regulating transcription. However, this possibility remains to be tested *in vivo*. Therefore, I assessed the effect of O-GlcNAc depletion on preimplantation gene expression. To this aim, I collected post-EGA 2-cell embryos, morulae and blastocysts cultured *ex vivo* from non-injected and injected zygotes for single embryo mRNA-sequencing (Smart-Seq) (Picelli et al., 2014) (Fig.9a). Although no developmental delay was evident from embryo morphology (Fig.8f), I asked whether any delay was captured in the transcriptome. By performing PCA of all embryos from the three stages using the 500 genes with the highest variance, I found the transcriptome of O-GlcNAc-depleted morulae to be more distant from the blastocyst stage than the control morulae (Fig.9b), hence pointing to a developmental delay of O-GlcNAc-depleted morulae better captured in the transcriptome. The same observation can be made when performing the PCA using all differentially expressed genes (adj. p-value<0.05 in the Btgh versus dBtgh or Btgh versus non-injected comparison at any stage) (Fig.9b). Notably, in the PCA pseudotime dBtgh-injected morulae are found between Btgh-injected and non-

injected embryos. The intermediate transcriptional delay shown by dBtgh-injected morulae could also be due to the injection procedure, but also by a low level of activity of the dBtgh protein, which cannot be excluded. From the same PCA, the level of transcriptional delay affecting the morulae is not detected anymore at the blastocyst stage (Fig.4b).

I proceeded analyzing the stage-specific transcriptional phenotype of O-GlcNAc-depleted preimplantation embryos.

2.2.3 Nuclear O-GlcNAc depletion does not prevent embryonic genome activation

I specifically investigated the effect of nuclear O-GlcNAc depletion on the process of EGA. To this aim, I clustered non-injected and Btgh/dBtgh-injected 2-cell embryos from our datasets with published mRNA-Seq data of single 2-cell embryos collected 18 h (pre-EGA) and 28 h (post-EGA) after ICSI (Conine et al., 2018), which allows an exact timing of fertilization and thus a more precise timing of EGA. I defined “strictly EGA" genes by intersecting published EGA genes (Li et al., 2018) and genes significantly upregulated (adj. p-value<0.05 and log2FC>1) from 18 h to 28 h in 2-cell embryos; I defined “strictly maternal” genes as the ones significantly downregulated (adj. p-value<0.05 and log2FC<-1) from 18 h to 28 h in 2-cell embryos; from the pool of these two groups of genes, I selected the first 250 showing highest variance in the ICSI dataset. Notably, the 2-cell stage data from my experiments clustered with post-EGA 2-cells, with no additional clustering observed between Btgh-injected embryos and controls (Fig.9d). When I looked at the expression of some EGA markers (*Sp110*, *Zfp352*, *Zscan4c*, *B020004J07Rik*, *Arg2*, *Tcstv3*), only two of them showed a significant upregulation (*Arg2*, *Tcstv3*) after O-GlcNAc depletion (Fig.S3a) and none a downregulation.
This result shows that EGA is globally unaffected by nuclear O-GlcNAc removal.

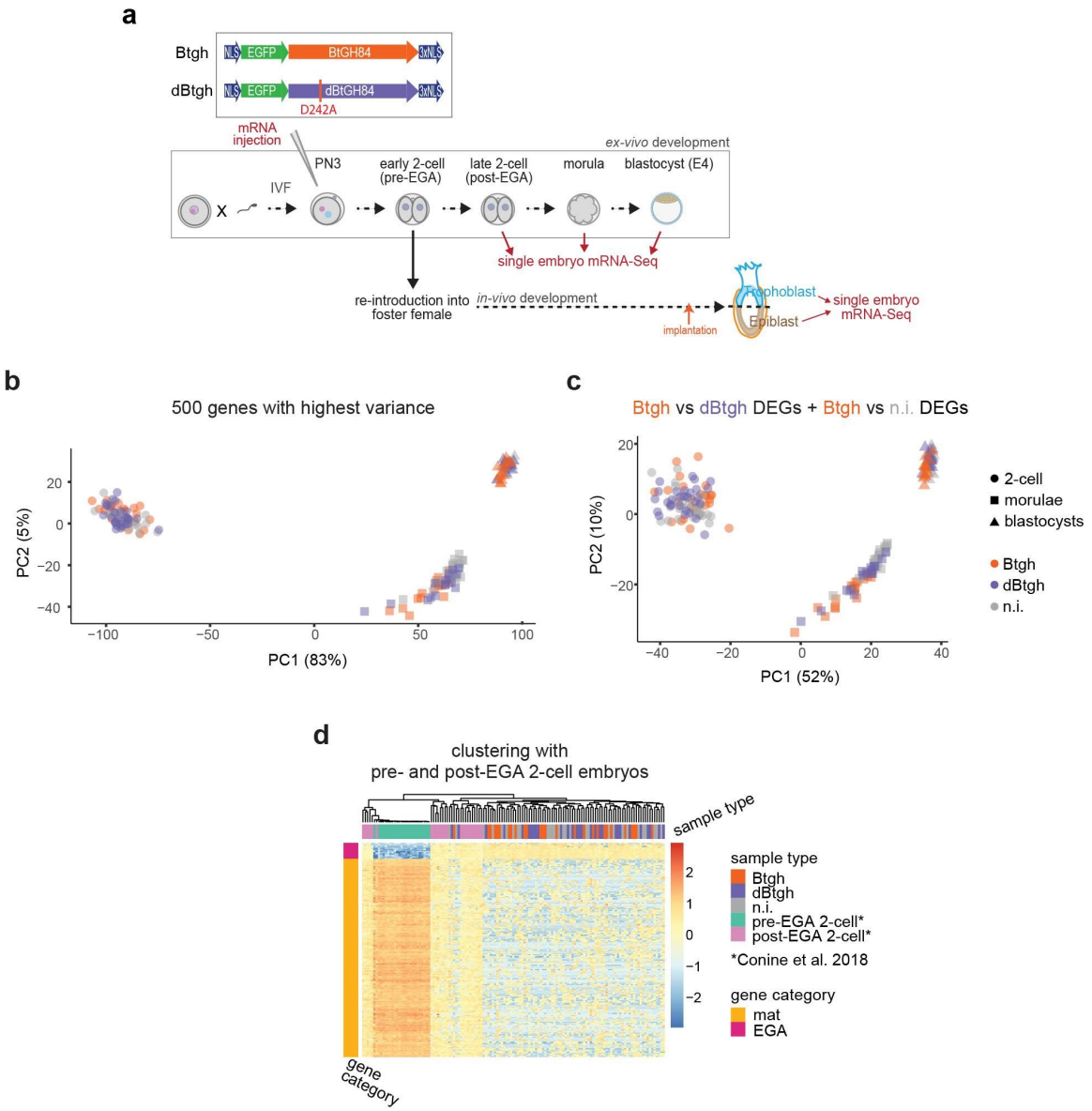


Figure 9: Developmental delay in O-GlcNAc-depleted morulae captured by transcriptomics but normal EGA in O-GlcNAc-depleted 2-cell embryos.

(a) Experimental design for all Smart-Seq experiments performed after Btgh injection. Related to results in Fig.9-13.

(b-c) PCA of all single embryos from the three Smart-Seq datasets of preimplantation stages, using (b) the first 500 genes with the highest variance across all samples or (c) all differentially expressed genes (DEGs) in O-GlcNAc-depleted embryos (adj. p-value<0.05 in the Btgh versus dBtgh or Btgh versus non-injected comparison at any stage). PCA was performed as described in Methods. The variance explained by each principal component is in parentheses.

(d) Heatmap of the three experimental groups of 2-cell embryos from my experiment together with pre-EGA and post-EGA 2-cell embryos generated through ICSI (GSE111864) (Conine et al., 2018). All samples are pooled together in one DESeq2 dataset, normalized using DESeq2 and $\log_2$ transformed. Then, only the first 250 strictly maternal (mat) and strictly ZGA (ZGA) genes with the highest variance in the ICSI dataset are used to build the heatmap. Heatmap values are scaled by rows and samples are clustered using Pearson correlation.

2.2.4 The expression of retrotransposons across preimplantation stages is largely unaffected by O-GlcNAc removal

A peak in the expression of the MERVL class of LTR retrotransposons is a unique feature of the 2-cell transcriptome (Peaston et al., 2004). Both full-lenght MERVL and MERVL 5' LTRs (annotated as MT2_Mm) are transiently associated with ZGA (Franke et al., 2017; Peaston et al., 2004) and were proposed to be essential for this process (Hendrickson et al., 2017; Sakashita et al., 2023). I used a specific bioinformatic pipeline to quantify the expression of RNA transposable elements (retrotransposons; in this section, TEs for simplicity) in our single embryo mRNA-Seq data. Both full-length MERVL transcripts (MERVL-int) and MT2_Mm fragments are equally expressed - if not slightly upregulated in some embryos - between Btgh-injected and control 2-cell embryos (Fig.10a). Additionally, I found no differentially expressed RNA transposons in O-GlcNAc-depleted 2-cell embryos (Fig.10b). O-GlcNAc-depleted morulae show a few significantly deregulated retrotransposons (adj. p-value<0.1) (Fig.10b), but the difference in expression is very low in magnitude ($0.1 < \text{absolute log2FC} < 0.4$) compared to what is observed for many families of RNA transposons across preimplantation development (Fig.10c) (mRNA-Seq data from GSE66582 (Wu et al., 2016)) and to what is expected from sequences present in multiple copies in the genome. The magnitude of TEs' deregulation is comparable to the expression difference observed for single copy genes after O-GlcNAc removal (see below), therefore it could be attributable to changes in chromatin landscape at TEs' genomic loci which are very close to gene transcription. Notably, upregulated TEs in Btgh-injected morulae include MTA_Mm-int, an ERVL-MaLR element typical of the 1-cell embryo (Franke et al., 2017) which should be downregulated between the 4-cell and 8-cell stage (Fig.10c). This could indicate that the deregulation of TEs observed in Btgh-injected morulae is, at least in part, a consequence of the slight developmental delay observed at this stage at the transcriptional level after O-GlcNAc depletion (Fig.9b). The trend of sustained upregulation of MERVL-int and MT2_Mm elements at this stage (Fig.10a) supports this hypothesis, since these TEs undergo gradual repression from the 2-cell to the 8-cell stage and beyond in unperturbed embryos (Fig.10c).

My data show that nuclear O-GlcNAc removal does not affect the dynamic regulation of retrotransposons' expression during preimplantation development.

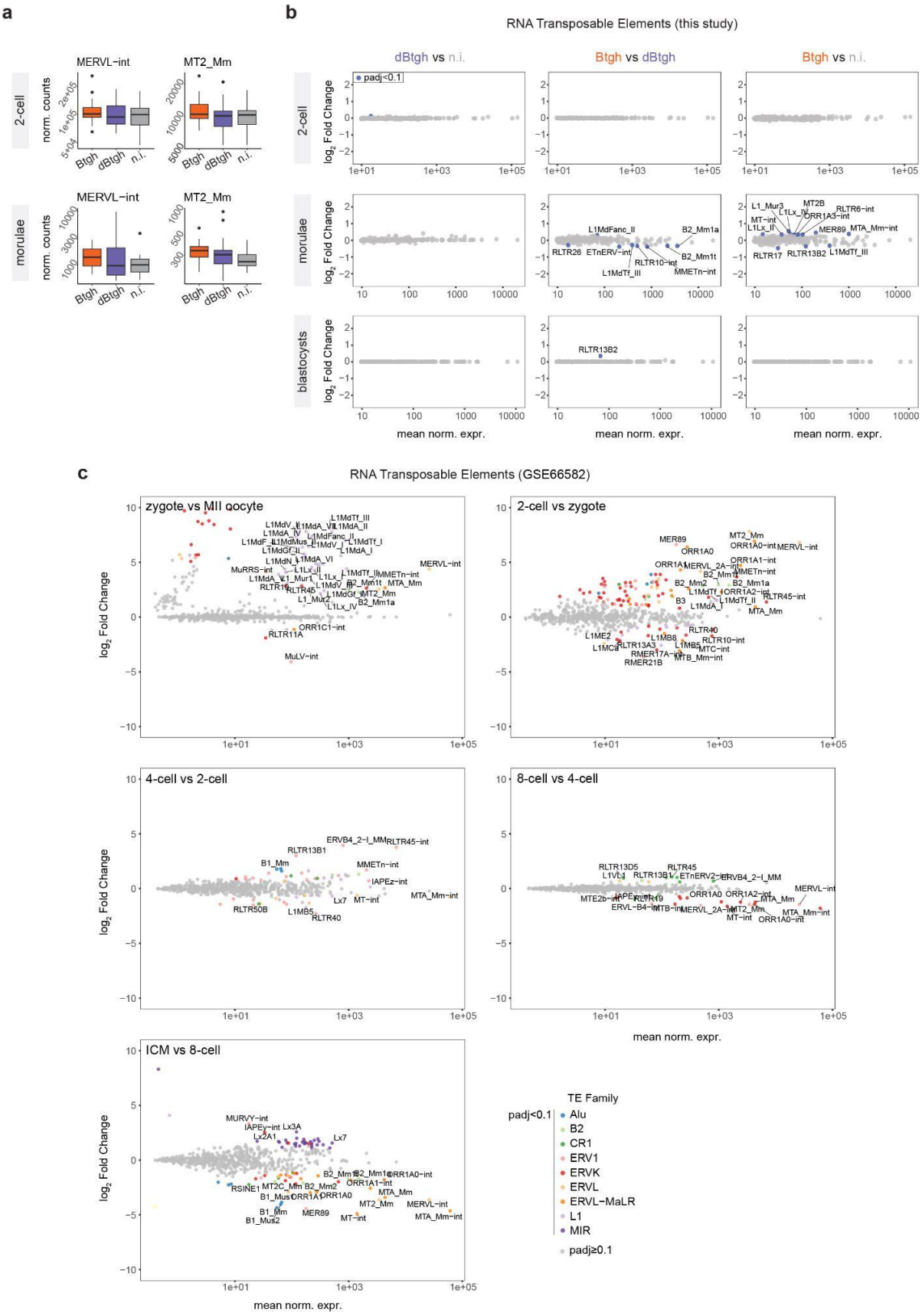

a
MERVL−int
MT2_Mm
2-cell
morulae
norm. counts
Btgh
dBtgh
n.i.
b
RNA Transposable Elements (this study)
dBtgh vs n.i.
Btgh vs dBtgh
Btgh vs n.i.
padj<0.1
log2 Fold Change
mean norm. expr.
L1MdFanc_II
B2_Mm1a
RLTR26
ETnERV−int
B2_Mm1t
RLTR10−int
L1MdTf_III
MMETn−int
L1_Mur3
MT2B
L1Lx_IV
RLTR6−int
ORR1A3−int
MT−int
L1Lx_II
MER89
MTA_Mm−int
RLTR17
RLTR13B2
L1MdTf_III
RLTR13B2
c
RNA Transposable Elements (GSE66582)
zygote vs MII oocyte
2-cell vs zygote
4-cell vs 2-cell
8-cell vs 4-cell
ICM vs 8-cell
log2 Fold Change
mean norm. expr.
MERVL−int
MuLV−int
ORR1C1−int
RLTR11A
MER89
ORR1A0
ORR1A0−int
ORR1A1−int
MERVL_2A−int
MMETn−int
B2_Mm2
B2_Mm1a
ORR1A2−int
RLTR45−int
B3
L1MdA_I
MTA_Mm
RLTR40
L1ME2
L1MB8
RLTR10−int
L1MB5
MTC−int
L1MC4
RLTR13A3
RMER17A−int
MTB_Mm−int
RMER21B
ERVB4_2−I_MM
RLTR45−int
RLTR13B1
B1_Mm
MMETn−int
IAPEz−int
Lx7
MT−int
MTA_Mm−int
RLTR50B
L1MB5
RLTR40
RLTR13D5
L1VL1
RLTR45
RLTR13B1
ETnERV2−int
MTE2b−int
RLTR19
ERVL−B4−int
MTB−int
MERVL_2A−int
MT2_Mm
MT−int
MURVY−int
IAPEy−int
Lx3A
Lx2A1
Lx7
RSINE1
MT2C_Mm
B2_Mm2
B2_Mm1t
ORR1A1−int
MTA_Mm
B1_Mus
ORR1A2
ORR1A0
MT2_Mm
MERVL−int
B1_Mm
B1_Mus2
MER89
MT−int
MTA_Mm−int
TE Family
padj<0.1
Alu
B2
CR1
ERV1
ERVK
ERVL
ERVL−MaLR
L1
MIR
padj≥0.1

Figure 10: The expression of retrotransposons across preimplantation is largely unaffected by O-GlcNAc removal.

(a) DESeq2-normalized counts of two retrotransposons associated with mouse EGA. MERVL-int denotes the full-length MERVL element, MT2_Mm the MERVL 5' LTR. Boxplots hinges correspond to first and third quartiles; median is shown inside; whiskers extend to the largest and smallest values no further than 1.5 * IQR from the hinge (IQR = inter-quartile range, or distance between the first and third quartiles); data beyond the end of the whiskers are plotted individually. y-axes are in $\log_{10}$ scale.

(b) MA-plots of DESeq2 differential gene expression analysis of RNA transposable elements comparing the three experimental groups of embryos in the three preimplantation stages. Only TEs with mean of DESeq2-normalized counts ≥ 10 are shown. All TEs with adj. p-value < 0.1, any Log2FC are colored, the ones with absolute log2FC ≥ 0.2 are labeled.

(c) MA-plots of DESeq2 differential gene expression analysis of RNA transposable elements between successive stages of preimplantation development using published Smart-Seq data GSE66582 (Wu et al., 2016). 2-cell here denotes post-EGA 2-cell embryos. Pre-EGA 2-cell embryos were discarded because one replicate out of two was low quality. Only TEs with mean of DESeq2-normalized counts ≥ 10 are shown. TEs with adj. p-value < 0.1 are colored by family. TEs which stand out are labeled.

2.2.5 Transcriptional phenotype of O-GlcNAc-depleted preimplantation embryos

2.2.5.1 2-cell embryos

The transcriptional phenotype at the 2-cell stage after nuclear O-GlcNAc depletion was overall subtle, with the three conditions only slightly separating along PC1 after principal component analysis (Fig.11a). Interestingly, the group of genes mostly associated with PC1 (but not PC2 and only mildly PC3) (hereafter "PC1-genes") showed significantly higher average expression in Btgh-injected embryos compared to both controls (Fig.11b and Fig.S3b). Since the maternal to embryonic transition is characterized by a progressive degradation of maternal mRNAs alongside the accumulation of newly transcribed embryonic RNAs following EGA, I checked whether the higher expression of PC1-genes could be driven by either maternal or embryonic transcripts, as expected from an alteration of degradation or new transcription, respectively. Unexpectedly, PC1-genes comprised maternal and EGA transcripts, both types showing higher expression after O-GlcNAc depletion (Fig.11b). These genes are also overall not highly expressed genes (note the y-axis scale in Fig.11b) and the higher average expression of each of these genes in O-GlcNAc-depleted embryos is often due to a bigger number of embryos showing a detectable expression (Fig.S3c).

Despite a subtle separation by PCA, at the 2-cell stage the analysis of differential gene expression between the three conditions showed that O-GlcNAc depletion has a measurable, albeit small in magnitude, effect on gene expression (Fig.11c). Gene set enrichment analysis (GSEA) revealed that downregulated genes tend to encode proteins involved in protein translation (Fig.11d). Accordingly, genes encoding for ribosomal proteins (*Rpl12*, *Rpl11*, *Rps5*, *Rps15a*, *Rps26*, *Rps2*, *Rpl10*) as well as for factors of translation initiation and mRNA metabolism (*Eif4b*, *Eif3a*, *Ythdf2*, *Larp4*) are among significantly downregulated ones after O-GlcNAc depletion (Fig.S3d-e). When averaging the expression of all ribosomal proteins, Btgh-injected embryos showed a mild but significant downregulation (Fig.11e).
There are always small embryo-to-embryo variations in the picoliters of mRNA released by the needle during microinjections. Since the injection of the recombinant mRNA might represent an increase in intracellular translation demand, I asked whether the higher expression of ribosomal proteins was associated with a higher number of injected Btgh mRNA molecules than dBtgh ones. This turned out not to be the case (Fig.11e). Thus the global downregulation of genes encoding ribosome-associated proteins is attributed to nuclear deglycosylation.

Translation and mRNA metabolism are the most dynamically changing pathways in the 2-cell embryo (Gao et al., 2017), therefore my GSEA result could point to a developmental delay already starting at this stage. However, the profile of EGA and maternal gene expression does not support this hypothesis: O-GlcNAc-depleted and control 2-cell embryos look indistinguishable in terms of completion of maternal RNA degradation and upregulation of newly transcribed genes, two processes that gradually take place during the 24 h-long 2-cell stage (Fig.9c).

It is noticeable that the two most upregulated genes in O-GlcNAc-depleted 2-cell stage embryos are *Eif4a3l1* and *Eif4a3l2* (Fig.11c), while their paralogue *Eif4a* is downregulated. These three genes are mouse orthologs of human EIF4A3, a core component of the exon junction complex (Chan et al., 2004), whose function is tightly linked with mRNA nuclear export and cytosolic translation and which has been shown to coordinate ribosomal biogenesis (Kanellis et al., 2021). To the best of my knowledge, the differential regulation of these three related proteins in the mouse embryo has never been reported.

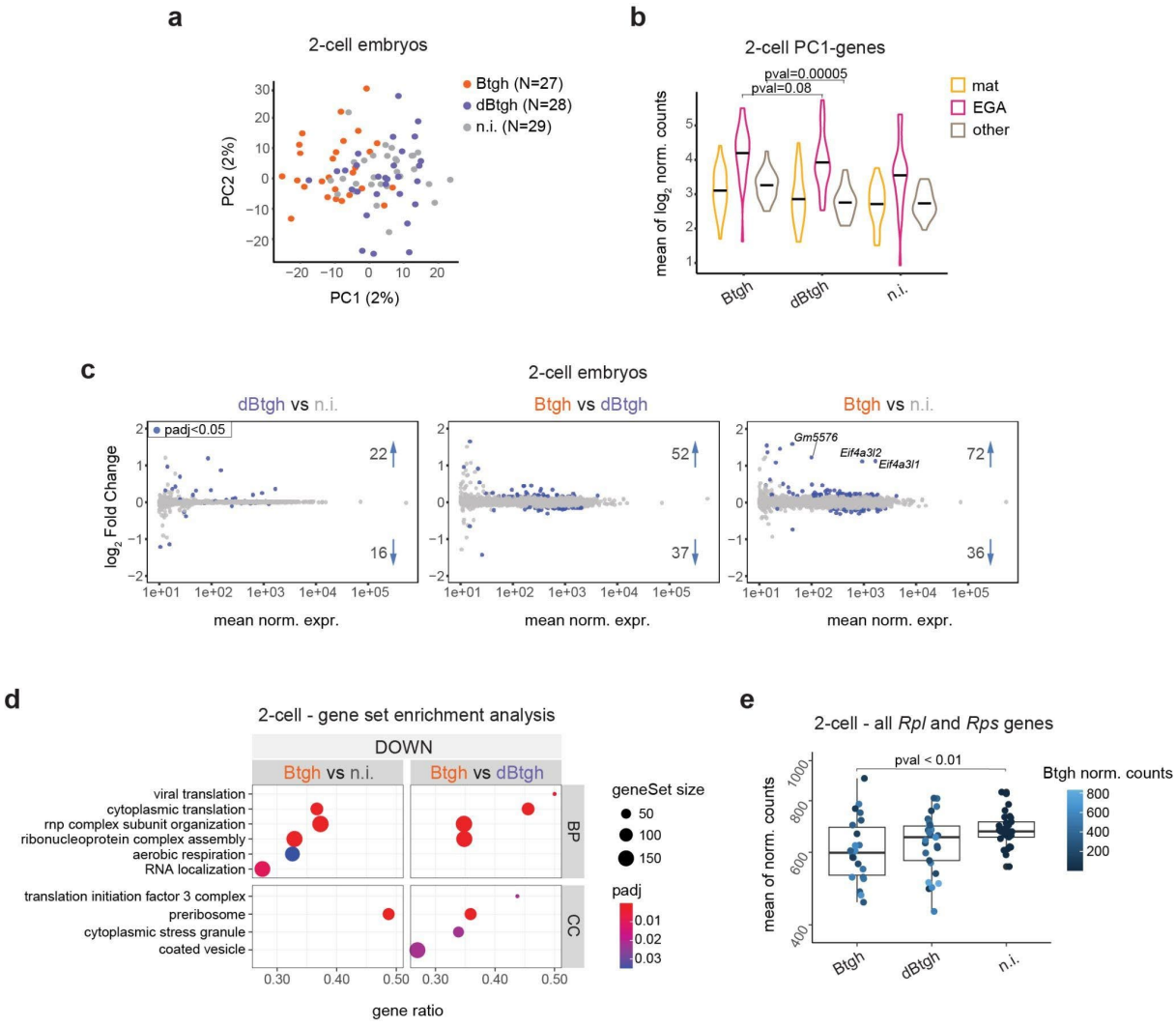


Figure 11: Transcriptional phenotype of O-GlcNAc-depleted 2-cell embryos.

(a) PCA of the 2-cell embryos. Total number of single embryos in the dataset for each experimental group is indicated in legend. PCA was performed as described in Methods. The variance explained by each principal component is in parentheses.

(b) Average of DESeq2-normalized and $\log_2$ transformed expression of the 150 genes mostly associated with PC1 of (a), in the three experimental groups of 2-cell embryos. Genes are defined as "strictly maternal" (mat) or "strictly ZGA" (ZGA) as described in the text, as "other" if not falling in these two categories. Mean for each group of embryos is marked. P-values from unpaired Wilcoxon rank sum exact test are shown if < 0.05.

(c) MA-plots of DESeq2 differential gene expression analysis for the different comparison between experimental groups of 2-cell embryos. Only genes with mean of DESeq2-normalized counts ≥ 10 are shown. All genes with adj. p-value < 0.05, any Log2FC are colored and their number is indicated. Two upregulated genes standing out described in the text (*Eif4a3l1* and *Eif4a3l2*) are labeled, together with the third one with a similar change in expression, *Gm5576*. The latter is a pseudogene described in NCBI as orthologous to human EIF4A3 (https://www.ncbi.nlm.nih.gov/gene/434050), as for *Eif4a3l1* and *Eif4a3l2*. It is difficult to judge whether *Gm5576* upregulation is caused by the transcription of the pseudogene or by mapping errors due to high similarity with the orthologous *Eif4a3l1* and *Eif4a3l2*.

(d) Gene set enrichment analysis (GSEA) of gene expression change in nuclear O-GlcNAc-depleted 2-cell embryos versus controls. The first 10 terms based on Normalized Enrichment Score (NES) are

shown. No upregulated term is found among the first significant ones. Terms are ordered based on gene ratio. Size of dots is proportional to the number of total genes of a gene ontology (GO) term. Gene ratio = fraction of total genes of the GO term which are differentially expressed in the dataset. BP = Biological Process GO, CC = Cellular Component GO.
(e) Single-embryo mean of DESeq2-normalized counts for all ribosomal proteins' genes from the Ribosomal Protein Gene Database (Nakao et al., 2004) for mus musculus. y-axis is in $\log_{10}$ scale. Dots are colored based on DESeq2-normalized counts of the Btgh/dBtgh contruct. P-value (pval) was computed using unpaired Student's t-test.

2.2.5.2 Morulae and blastocysts

At later preimplantation stages, the magnitude of gene expression changes due to O-GlcNAc depletion remains modest, although at the morula stage it temporarily increases, also in terms of number of DEGs (defined as adj. p-value<0.05, any log2FC ; Fig.12a). I showed that the overall transcriptome of O-GlcNAc-depleted embryos, as well as the differentially expressed genes, reveal a developmental delay at the morula stage (Fig.9b). Hence, I expect that a consistent fraction of gene expression differences in Btgh-injected morulae is a reflection of this delay. Moreover, the transcriptional delay could still partly be visible at the blastocyst stage and explain a number of differentially expressed genes in Btgh-injected blastocysts.

I performed GSEA at later preimplantation stages and compared the result for morulae and blastocysts. Notably, this analysis shows that the gene sets affected by O-GlcNAc depletion are highly shared between the morula and blastocyst datasets (Fig.12b and S4a). In particular, upregulated genes are mostly related to mitotic chromosome separation and metaphase/anaphase transition of the cell cycle. Downregulated genes are involved in various essential cellular functions: the mitochondrial respiratory chain complex (*respiratory chain complex IV*, *cytochrome complex*, *ATP synthase complex*, *myelin sheath* GOs), intracellular vesicles, transport through membranes and protein localization to membranes (*ER lumen*, *Golgi-associated vesicle*, *vesicle coat*, *cell surface*, *cell cortex* GOs), function of the actomyosin cytoskeleton (*contractile fiber*, *cell cortex*, *ciliary base*, *apical part of cell*, *dendritic spine* GOs), the proteasome (*proteasome core complex* GO) and DNA compaction (*nucleosome*, *DNA packaging* GOs), the latter including canonical and non-canonical histone variants. Noteworthy, the downregulation of genes participating in aerobic respiration, as well as vesicle coat, already appeared in O-GlcNAc-depleted 2-cell embryos (Fig.11d).

Ribosome subunits show a more complex expression behavior, with a temporary upregulation in the morula followed by a downregulation in the blastocyst (Fig.S4b). Importantly, in agreement with the apparently normal blastocyst lineage specification (Fig.8f), I did not observe a deregulation of the pluripotency or trophoblast network.

I tested whether the misregulation of genes related to metabolic respiration could be indicative of a general defect in the dynamics of metabolic pathways normally observed during preimplantation, consisting of an increase of respiration from the 2-cell to the blastocyst and then a switch to glycolysis as the predominant metabolic driver after implantation (Malkowska et al., 2022). However, I observed the expected dynamics of metabolic genes across stages, with no differences between Btgh-injected embryos and controls (Fig.S4c).

The upregulation of mitotic genes could be an indication of a defect in chromosome segregation and potentially affect the incidence of aneuploidy in the embryo. To address this question, for my IVF experiments I used evolutionary distant mouse strains (female FVB/NCrl x male PWD/Ph) and performed transcriptome analyses on F1 hybrid embryos. The presence of single nucleotide polymorphism (SNPs) between the two strains allowed me to assign reads to either the maternal or paternal allele and compute the ratio of total maternal versus total paternal reads for each embryo. This ratio should be around 0.5 in case of normal ploidy. This analysis revealed that aneuploidy is not more common in the Btgh-injected group as compared to the control groups (Fig.12c).

As discussed above, part of the gene expression change described by the GSEA result could be due to the slight developmental delay of the morulae. For example, *gastrulation* appears as one of the enriched terms among downregulated genes specifically in the GSEA result of Btgh-injected morulae. I inquired about the expected dynamics of expression of the downregulated genes belonging to this term across preimplantation development by analyzing published mRNA-Seq data (Wu et al., 2016). 19 out of 23 of them turned out to be upregulated between the 8-cell and the E3.5 stage, often drastically (Fig.S4d). This implies that the apparent downregulation of these genes in O-GlcNAc-depleted morulae could reflect their level at a slightly earlier point in development and argues in favor of their downregulation being a result of the delay (caused by nuclear O-GlcNAc depletion) and not a primary effect of O-GlcNAc-mediated transcriptional regulation.

2.2.6 O-GlcNAc-mediated transcriptional response across preimplantation

I analyzed the dynamics of differentially expressed genes (DEGs) (adj. p-value < 0.05 in the Btgh versus dBtgh or Btgh versus non-injected comparison) across the three studied preimplantation stages. There is a low number of overlapping DEGs (Fig.S4e) and low correlation of log2FoldChanges between stages (Fig.12d). However, a subset of DEGs is persistently and coherently deregulated from the 2-cell to the morula and a smaller subset even up to the blastocyst (Fig.12d). Most of the constantly upregulated genes (7 in total) are involved in the transition to the mitotic phase of the cell cycle (*Ccnf*, *Bora*, *Cks1b*) or in the mitotic spindle (*Bora*, *Map4*). Furthermore, among the consistently upregulated genes we find the mitochondrial glutathione transporter *Slc25a39 (Wang et al., 2021)*, *B4galt3*, one of the 7 beta-1,4-galactosyltransferases (beta4GalT) responsible for the synthesis of complex-type N-linked oligosaccharides (Shaper et al., 1986) and Casein Kinase 1 Epsilon (*Csnk1e*), a central component of the circadian clock (Partch et al., 2006). The constantly downregulated genes (6 in total) include two ribosomal proteins (*Rpl10* and *Rpl5*), Poly(U) Binding Splicing Factor 60 (*Puf60*), involved in pre-mRNA splicing (Page-McCaw et al., 1999), a subunit of mitochondrial ATP synthase (*Atp5pb*), Clathrin Light Chain A (*Clta*), which coats intracellular vesicles, and CCHC-Type Zinc Finger Nucleic Acid Binding Protein (*Cnbp*), also involved in transcriptional repression by binding the sterol regulatory element (SRE) (Rajavashisth et al., 1989). A more recent study implicated *Cnbp* in the promotion of translation by showing its binding at G-rich elements of mRNAs (Benhalevy et al., 2017).

Interestingly, when I plotted the dynamics of constantly deregulated genes across preimplantation stages in unperturbed embryos from public data (Wu et al., 2016), I noticed that the downregulated ones are transcriptionally upregulated throughout preimplantation (Fig.12e, bottom panel). The dynamics in normal conditions of the coherently downregulated genes in the absence of O-GlcNAc suggests that O-GlcNAc is one of the contributor to their developmental activation across preimplantation stages, or at least at the 2-cell stage, which is missing when the modification is depleted. On the other hand, coherently upregulated genes show a variety of preimplantation expression dynamics in unperturbed embryos (Fig.12e, upper panel), hence it is less likely that in normal conditions they are downregulated by O-GlcNAc. Their identity points towards an acute response to an effect of the absence of O-GlcNAc on the cell cycle.

In order to gain insight into potential mediators of the regulation of coherently changing genes by O-GlcNAc, I performed ChIP Enrichment Analysis (ChEA) (Lachmann et al., 2010) of DEGs in O-GlcNAc-depleted 2-cell embryos, since the coherent response already starts at this stage but 2-cell embryos do not seem to be affected yet by the developmental delay, which could confound the ChEA result (Fig.12f). In addition, I decided to be stringent and exclusively use DEGs from the comparison of Btgh-injected versus dBtgh-injected embryos, in order to eliminate any effect of the injection procedure on gene expression. Noteworthy, I found enriched transcription factors only among the list of downregulated DEGs. The most enriched TFs include *Creb1* and *Crem*, both binding to cAMP-responsive elements on DNA; *Stat3*, which responds to various signals to stimulate cell growth (Hirano et al., 2000); the histone modifiers *Ep300* (Tropberger et al., 2013) and *Kdm5b* (Yamane et al., 2007); the oncogene *c-Myc* (Dang, 2013). Importantly, none of the regulators appearing in the ChEA result are significantly differentially expressed (Fig.S4f), suggesting that their potential differential binding after O-GlcNAc depletion is likely attributed to post-translational regulation. Most of these TFs have been reported to be modified by O-GlcNAc in wild type cells (Wulff-Fuentes et al., 2021a), although this has to be taken with caution considering that thousands of O-GlcNAcylated TFs have been reported (Wulff-Fuentes et al., 2021a).

Overall, my transcriptomics analyses across three key preimplantation stages showed that the effect of nuclear O-GlcNAc removal on preimplantation gene expression is consistently low in magnitude but significantly affects hundreds of genes. Part of this effect can be explained by a slight developmental delay induced by O-GlcNAc depletion from the morula stage onwards. Another part of this gene expression change is shared and almost constant between O-GlcNAc-depleted 2-cell embryos, morulae and blastocysts, hence it is likely to be a direct effect of O-GlcNAc-mediated transcriptional regulation. The shared response affects in opposite ways the mitotic spindle checkpoint on one side, translation, the mitochondrial electron transport chain and membrane-associated functions on the other.

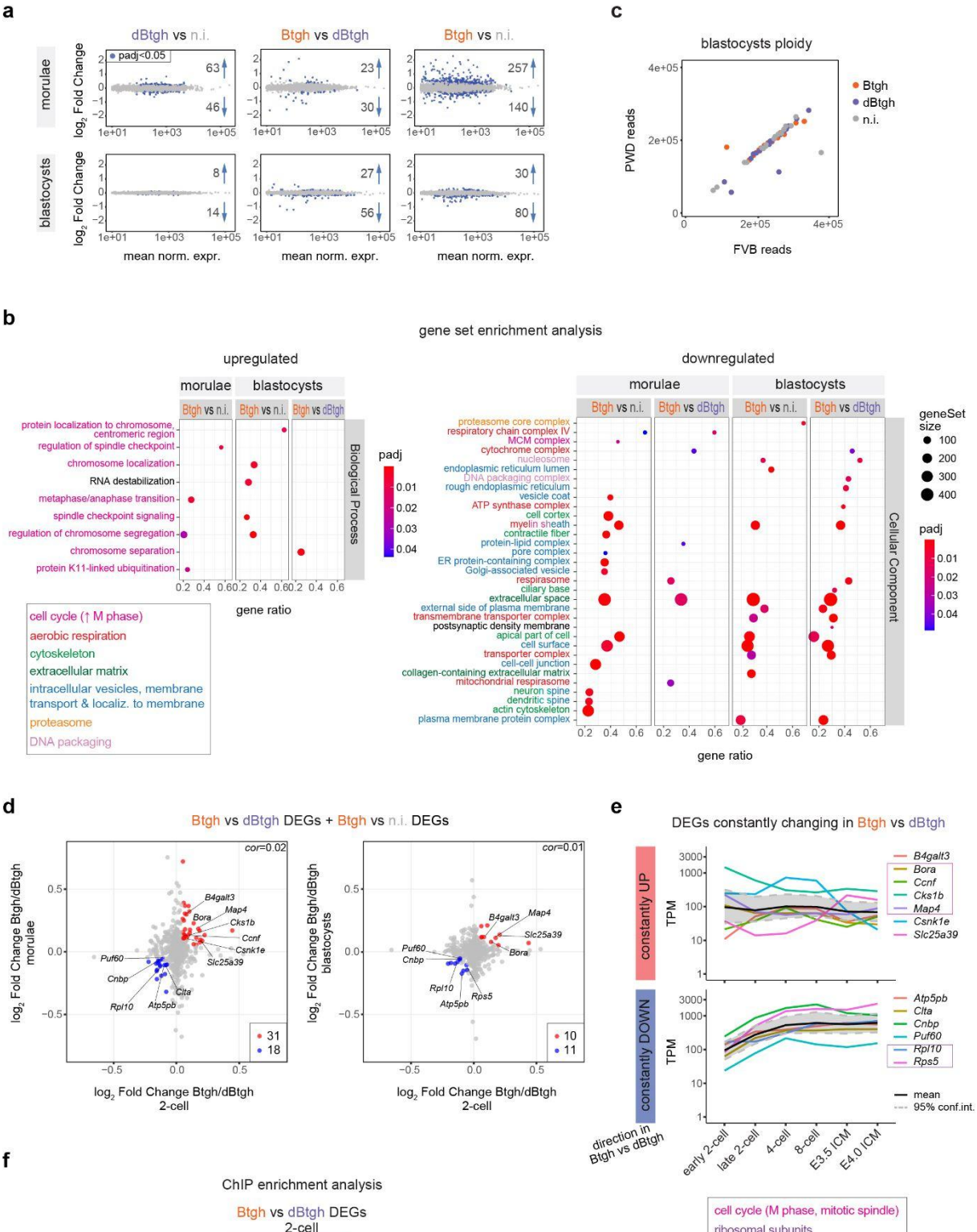

a
dBtgh vs n.i.
Btgh vs dBtgh
Btgh vs n.i.
morulae
blastocysts
padj<0.05
log2 Fold Change
mean norm. expr.
c
blastocysts ploidy
PWD reads
FVB reads
Btgh
dBtgh
n.i.
b
gene set enrichment analysis
upregulated
downregulated
morulae
blastocysts
Btgh vs n.i.
Btgh vs dBtgh
protein localization to chromosome, centromeric region
regulation of spindle checkpoint
chromosome localization
RNA destabilization
metaphase/anaphase transition
spindle checkpoint signaling
regulation of chromosome segregation
chromosome separation
protein K11-linked ubiquitination
Biological Process
gene ratio
padj
cell cycle (↑ M phase)
aerobic respiration
cytoskeleton
extracellular matrix
intracellular vesicles, membrane transport & localiz. to membrane
proteasome
DNA packaging
proteasome core complex
respiratory chain complex IV
MCM complex
cytochrome complex
nucleosome
endoplasmic reticulum lumen
DNA packaging complex
rough endoplasmic reticulum
vesicle coat
ATP synthase complex
cell cortex
myelin sheath
contractile fiber
protein-lipid complex
pore complex
ER protein-containing complex
Golgi-associated vesicle
respirasome
ciliary base
extracellular space
external side of plasma membrane
transmembrane transporter complex
postsynaptic density membrane
apical part of cell
cell surface
transporter complex
cell-cell junction
collagen-containing extracellular matrix
mitochondrial respirasome
neuron spine
dendritic spine
actin cytoskeleton
plasma membrane protein complex
Cellular Component
geneSet size
d
Btgh vs dBtgh DEGs + Btgh vs n.i. DEGs
cor=0.02
cor=0.01
log2 Fold Change Btgh/dBtgh morulae
log2 Fold Change Btgh/dBtgh blastocysts
log2 Fold Change Btgh/dBtgh 2-cell
B4galt3
Bora
Map4
Cks1b
Ccnf
Csnk1e
Slc25a39
Puf60
Cnbp
Clta
Rpl10
Atp5pb
Rps5
e
DEGs constantly changing in Btgh vs dBtgh
constantly UP
constantly DOWN
TPM
direction in Btgh vs dBtgh
early 2-cell
late 2-cell
4-cell
8-cell
E3.5 ICM
E4.0 ICM
mean
95% conf.int.
cell cycle (M phase, mitotic spindle)
ribosomal subunits
f
ChIP enrichment analysis
Btgh vs dBtgh DEGs 2-cell
CREB1 ChIP-Seq hippocampus rat
CREM ChIP-Seq GC1-SPG mouse
STAT3 ChIP-Seq EML mouse bone marrow
P300 Chip−Seq B-cells human
KDM5B ChIP-Seq mESCs
MYC ChIP-ChIP mESCs
MYC ChIP-Seq human breast cancer cell line
SPI1 ChIP-Seq erythroleukemia mouse
E2F1 ChIP-Seq hepatocytes mouse
SPI1 ChIP-Seq NB4 Human
DOWN
pval
combined score

Figure 12: Transcriptional phenotype of O-GlcNAc-depleted morulae and blastocysts and coherent response across stages.

(a) MA-plots of DESeq2 differential gene expression analysis for the different comparison between experimental groups of morulae and blastocysts. Only genes with mean of DESeq2-normalized counts ≥ 10 are shown. All genes with adj. p-value < 0.05, any Log2FC are colored and their number is indicated.

(b) Gene Set Enrichment Analysis (GSEA) of gene expression change in nuclear O-GlcNAc-depleted morulae and blastocysts versus controls. For the upregulated terms, the most 20 significant Biological Process (BP) GOs per stage based on Normalized Enrichment Score (NES) are shown, because more informative than the Cellular Component (CC). For the downregulated terms, the most 20 significant CC GOs per stage based on NES are shown, because more informative than the BP. See Fig.S4a for the complementary upregulated CC and downregulated BP. Terms are ordered based on gene ratio and colored based on manual grouping of cellular functions. Ribosome-related terms are excluded and shown in Fig.S4b. Size of dots is proportional to the number of total genes of a GO term. Gene ratio = fraction of total genes of the GO term which are differentially expressed in the dataset.

(c) Total number of maternal (FVB) and paternal (PWD) reads for the single blastocysts Smart-Seq samples. Outliers represent potential aneuploid embryos. Assignment of reads to the FVB or PWD genome is described in Methods.

(d) Comparison of $\log_2$ fold changes of DEGs in Btgh-injected embryos versus any of the control groups between preimplantation stages. DEGs with log2FC > 0.05 at both stages and log2FC ≥ 0.1 in at least one of the two stages of each plot are coloured; among those, the ones which are coloured in both plots are also labeled. cor = Pearson correlation.

(e) Expression dynamics of genes labeled in (d) (i.e. constantly deregulated at all stages) throughout preimplantation development (mRNA-Seq data from GSE66582 (Wu et al., 2016)). The two biological replicates per stage were averaged. Mean for all genes is drawn, as well as 95% confidence interval for all genes, computed using basic nonparametric bootstrap (R function 'mean.cl.boot'). y-axis shows Transcript Per Million (TPM) and is in $\log_{10}$ scale.

(f) Result of ChIP Enrichment Analysis (ChEA) (Lachmann et al., 2010) for DEGs in Btgh-injected versus dBtgh-injected 2-cell embryos. The most significant terms are shown based on combined score, a ChEA-specific score that takes into account p-value and number of binding sites of a regulator. The ratio of number of DEGs targeted by the TF / total number of input DEGs is indicated on each bar. Regulators with less than 5 genes of overlap or not expressed in 2-cell embryos (mean DESeq2-norm. expr. < 10) were excluded. No upregulated enriched terms were found.

2.2.7 Loss of nuclear O-GlcNAc preimplantation causes developmental delay after implantation

Ogt-null embryos arrest their development around the time of implantation (E4-5) (O'Donnell et al., 2004a), therefore I asked whether nuclear O-GlcNAc could be vital during the implantation process. I injected zygotes with either the Btgh or dBtgh construct, which were transferred at the early 2-cell stage into pseudopregnant females. The surgical procedure was performed by EMBL Rome GEEF staff. The lab manager in the Boulard lab, Dr. Agnieszka Sadowska, dissected the females six days later (seven days after IVF, thus we consider the resulting embryos to be E7), and took microscopy images of all embryos (Fig.9a). I found no significant difference in the percentage of embryos recovered at E7 between the two groups of injections (Btgh: 43%, N=92 total embryos transferred, dBtgh: 48%, N=134 total embryos transferred) (Fig.13a). Staging of the embryos on the basis of appearance of the mesoderm (Downs and Davies, 1993) showed a very similar percentage of embryos to be Early Streak or Mid-Streak in both groups (Table S7). However, I found nuclear O-GlcNAc-depleted embryos to be slightly but significantly smaller in size at E7 (Fig.13b). This difference is mainly driven by 2 out of 5 of the litters of Btgh-injected embryos, thus the phenotype might be partially penetrant.

I then investigated whether the depletion of nuclear O-GlcNAc during preimplantation stages had an effect on postimplantation gene expression. To this end, after dissection embryos were cut in two halves, corresponding to the epiblast and trophectoderm tissues (Fig.13c). I processed them for single-tissue single-embryo mRNA-sequencing. The injected Btgh/dBtgh mRNA is no longer detectable at the E7 stage, while a low count of reads can still be detected in E4 blastocyst data (Fig.S5a). I checked the expression of *Ogt* and *Oga* and observed *Ogt* upregulation and *Oga* downregulation in Btgh-injected embryos at the morula and blastocyst stages, while this phenomenon disappears at E7 (Fig.S5b). As already mentioned earlier, a compensatory change in expression of the O-GlcNAc enzymes has been reported when O-GlcNAc homeostasis is disturbed (Park et al., 2017; Pravata et al., 2019b; Qian et al., 2018; Vaidyanathan et al., 2017b; Willems et al., 2017b; Z. Zhang et al., 2014) and I also observed it in MEFs after OGT depletion (Fig.S2b). Based on *Oga* and *Ogt* expression (Fig.S5b), I infer that active O-GlcNAc removal is no longer ongoing at E7, hence I interpret any gene expression changes at this stage as a result of a perturbation during previous developmental stages.

I observed no difference in gene expression in trophectoderm samples derived from Btgh-injected embryos compared to dBtgh injected ones (Fig.13d). However, there are few DEGs in the epiblast (10 up, 7 down). GSEA uncovered downregulation of mesoderm development and upregulation of genes involved in DNA methylation in the epiblast of Btgh-injected embryos (Fig.13e). The development of the mesoderm characterizes the transition between E6.5 and E7.5 (Kojima et al., 2014; Pijuan-Sala et al., 2019). On the other hand, the onset of mammalian gastrulation coincides with *de novo* DNA methylation, with the enzymes involved in this process peaking in expression at E5.5 and gradually decreasing between E6.5 and E7.5 (Fig.S5c). Taking this into consideration, the transcriptional abnormalities in the epiblasts very likely reflect a delay in postimplantation development resulting from nuclear O-GlcNAc depletion during preimplantation stages. In agreement, I found all significantly upregulated genes in the Btgh epiblasts to decrease in expression in the transition from E5.5 to E7.5 in unperturbed embryos (Zhang et al., 2017), while the opposite was true for the downregulated ones (Fig.S5d).

I used the expression of E7 mesodermal markers (from a single cell RNA-Seq dataset of mouse gastrulation (Pijuan-Sala et al., 2019)) and de-novo DNMTs to investigate the penetrance of the developmental delay suggested by both morphological assessment and differential expression analysis. Importantly, 9 out of 12 of the Btgh epiblasts showed lower expression of E7 mesodermal markers (and in most of the cases also higher expression of de-novo DNMTs) than any of the controls (Fig.13f), supporting a highly penetrant phenotype of developmental delay.

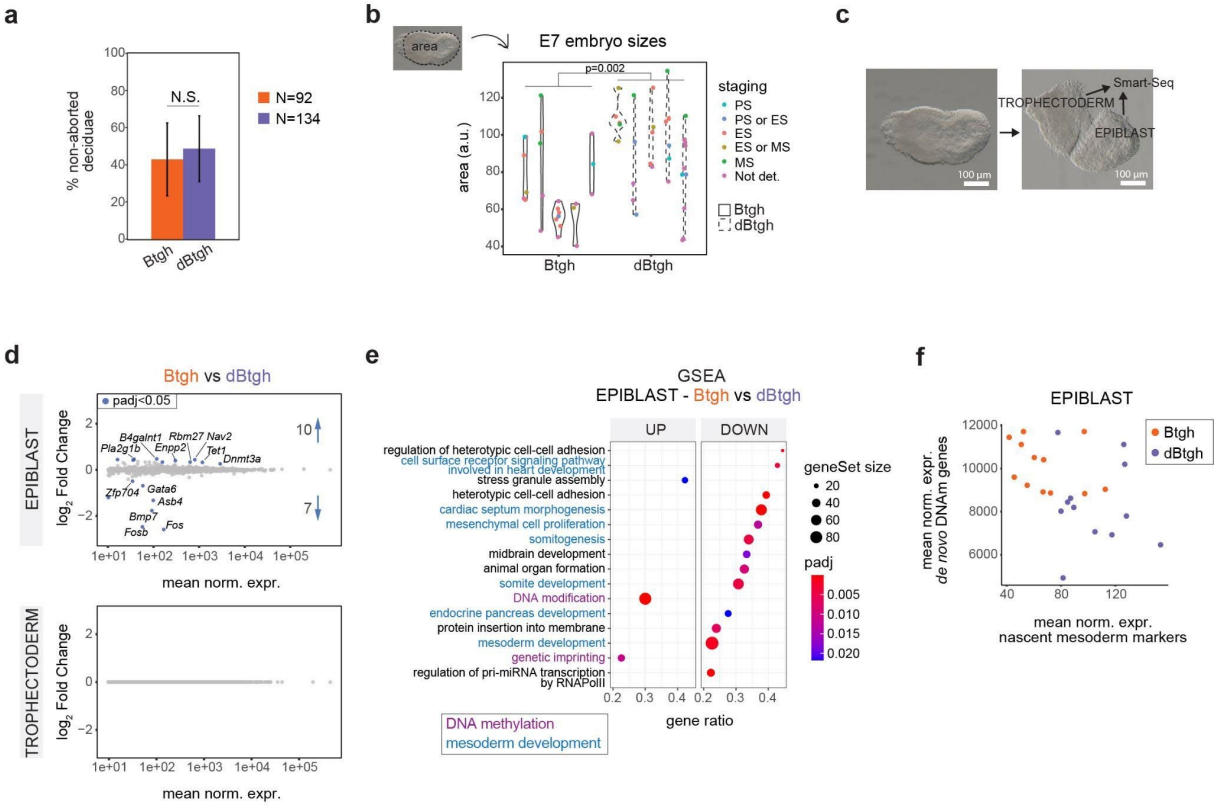


Figure 13: Loss of nuclear O-GlcNAc preimplantation causes developmental delay after implantation

(a) Percentage of Btgh/dBtgh-injected embryos found seven days after IVF (E7). Bar heights and error bars indicate the average and standard deviation, respectively, for five replicates of the microinjection experiment. N = 92 and 134 total injected 2-cell embryos transferred to pseudopregnant females for Btgh and dBtgh, respectively. N.S = not significant, based on unpaired Student's t-test.

(b) Area of Btgh/dBtgh-injected E7 embryos for all dissected litters, measured from the microscopy images taken during dissection. Each violin represents a litter. Embryos are coloured based on manual morphological staging of the image. One litter from dBtgh-injected embryos which was a clear outlier (all MS or later stages) was excluded. P-value was computed with unpaired Wilcoxon rank sum exact test.

(c) Scheme of the embryo cut performed during dissection for collection of the single embryonic halves for Smart-Seq.

(d) MA-plots from DESeq2 differential expression analysis of E7 epiblast or trophectoderm tissues from Btgh-injected embryos versus dBtgh-injected ones. Only genes with mean of DESeq2-normalized counts ≥ 10 are shown. All genes with adj. p-value < 0.05, any Log2FC are colored and their number is indicated.

(e) Gene Set Enrichment Analysis (GSEA) of gene expression change in E7 epiblasts derived from Btgh-injected embryos. The 16 most significant Biological Process (BP) GO terms based on NES are shown. Terms are ordered based on gene ratio and colored based on manual grouping of cellular functions. Size of dots is proportional to the number of total genes of a GO term. Gene ratio = fraction of total genes of the GO term which are differentially expressed in the dataset.

(f) Mean DESeq2-normalized counts of E7 mesoderm markers from (Pijuan-Sala et al., 2019) versus mean DESeq2-normalized counts of enzymes associated with *de novo* DNA methylation (*Dnmt3a*, *Dnmt3b*, *Tet1*), whose expression level decreases from E6.5 to E7.5, in E7 epiblasts

3) A suite of mouse alleles bearing catalytic point mutations in the *Ogt* gene

Owing that the *Ogt*-null mutation is embryonic lethal, the study of an hypomorphic *Ogt* allele is another valuable strategy to investigate O-GlcNAc biology *in vivo*. To this end, I exploited the results of a study which used the high resolution structure of an OGT homologue to test *in vitro* the effect of mutating specific residues predicted by the structure to be important for catalysis (Martinez-Fleites et al., 2008) (Fig.14a). Based on this study, I asked the Gene Editing and Embryology Facility (GEEF) at EMBL Rome to generate four mouse models bearing single amino acid mutations disrupting OGT's catalytic activity to various degrees (Fig.13b). The mutant mouse alleles were made by the GEEF using CRISPR-Cas9 genomic editing of the zygote, in an FVB background because both the bigger big size of the litters and of the single embryos from this strain make it a perfect model to study embryonic development.

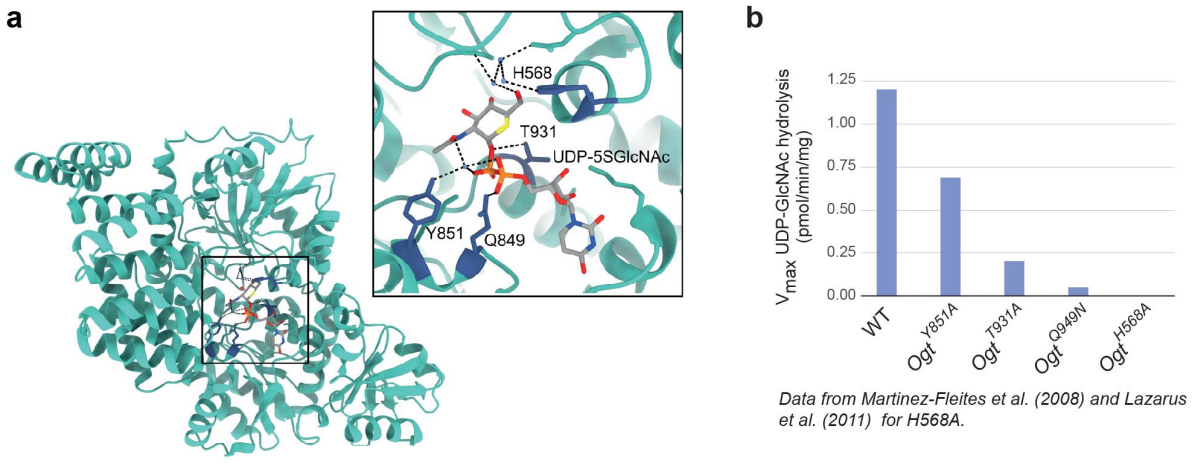
3.1 The viability of the hypomorphic mutants mirrors the impairment of *in vitro* activity

In the process of creation of a novel mouse allele, two initial steps have to be successful: i. the creation of chimeric mutant animals at the F0 generation (henceforth called "founders") after gene editing and reimplantation of the zygotes and ii. the transmission of the mutation from the founder to the F1 generation, also known as germline transmission (GLT) because it implies the presence of the mutation in the founder's germline. GLT is positive if mutant pups are found at weaning (21 days after birth or P21) in the F1 generation. Both the success in the creation of founders and a positive GLT are influenced by the efficiency of gene editing at the targeted locus, however for simple mutations (such as single amino acid mutations) they can give a first indication of the compatibility of the novel mouse allele with life. For example, in the case of our suite of *Ogt* hypomorphic alleles, the combination of number of founders and GLT showed a striking qualitative correlation with the level of disruption of OGT catalytic activity reported by the mutations *in vitro* (Fig.14c, first two columns). In detail, no founders were obtained for the mutation which completely abolishes OGT *in vitro* activity (Y568A), while none of the four founders obtained for the second most disruptive mutation, Q949N, transmitted the mutation to the F1 generation. I investigated whether Q949N mutant embryos could still be found at early stages of development by breeding a founder with two wild type females and genotyping single blastocysts at E3.5 for both sex and presence of the mutation. I found an expected ratio of female and male embryos (49% and 51%, respectively, N = 72), but none of them was mutant.

This result indicated that - at least for this one founder - the mutation was already not present in the germline.

For the T931A mutation (~20% of wild type OGT catalytic activity) GLT was positive, but the mutation was present in the F1 generation at a lower than expected percentage and only in heterozygous females (described in detail in next section). Finally, the Y851A mutation (~70% of wild type OGT catalytic rate) was transmitted to the F1 generation at mendelian ratio. Moreover, both heterozygous females and homozygous males were obtained, which were healthy and fertile. This allowed me and the lab manager Dr. Agnieszka Sadowska to inquire about eventual sub-lethal phenotypes of the mutant embryos, by collecting the number of live animals after breeding $Ogt^{Y851A/Y851A}$ females with $Ogt^{Y851A/Y}$ males. The collection of these data is still ongoing, but preliminary results show a sub-lethality both before birth and between birth and the day of weaning (P21; Fig.14c, third column).

In the following section, I will present all phenotypic data collected for mouse line Ogt^{T931A}, which is the allele with the most disruptive OGT mutation which still transmitted the mutation to the F1 generation.



Data from Martinez-Fleites et al. (2008) and Lazarus et al. (2011) for H568A.

c

	Number of founders	Germline transmission of the mutant allele	Lethal phenotype
Ogt^{Y851A}	4 mutant F0 / 26 total F0	At mendelian ratio 11 mutant F1 / 22 total F1	Prenatal and postnatal sub-mortality. Surviving hemizygous males and homozygous females were healthy and
Ogt^{T931A}	6 mutant F0 / 56 total F0	Low frequency 4 mutant F1 / 59 total F1	No hemyzygous male pup was born. Hemizygous blastocysts were obtained. Heterozygous females were healthy.
Ogt^{Q949N}	4 mutant F0 / 29 total F0	No germline transmission 0 mutant F1 / 26 total F1	Unknown
Ogt^{H568A}	0 mutant F0 / 31 total F0	Unknown	Unknown

Figure 14: A suite of mouse alleles bearing catalytic point mutations in the Ogt gene.
(a) Crystal structure of human OGT isoform 1 (UniProt O15294-1, one of the two ncOGT isoforms) with UDP-5SGlcNAc and a peptide substrate (PDB 4GJJ). For clarity, only one of the chains is shown. The area in the square is shown with higher detail as an insert. Human OGT 1 is identical (except for two single amino acid mutations outside of the catalytic core) to *Mus musculus* OGT isoform 1 (UniProt Q8CGY8-1) hence residues' numbering is the same. The residues we mutated in mice are shown in sticks representation in blue and the UDP-5SGlcNAc in sticks representation coloured by heteroatom. T931 and Q849 establish direct interactions with the donor substrate. Y851 interacts with the donor substrate via hydrogen bonds with an intermediate water molecule. Finally, H568 does not establish direct contacts with the donor substrate but coordinates different water molecules in the crystal structure near the binding site; H568 is considered to be the catalytic base.
I am thankful to Javier Lizarrondo who made the rendering of OGT's structure.
(b) Catalytic rate measured *in vitro* for the four OGT's mutants studied. Data are from (Martinez-Fleites et al., 2008) and for H568A also from (Lazarus et al., 2011).
(c) Number of animal bearing the correct point mutation at the F0 generation (founders), number of mutants in the F1 generation (frequency of germline transmission (GLT)) and eventual lethal phenotype observed from the F1 generation onwards for the four hypomorphic mutants of OGT described in (a) and (b).

3.2 Maternal effect of the hypomorphic mutant *Ogt*^{T931A}

The T931A substitution lowers *in vitro* OGT's activity to ~20% of the wild type. For this mouse line, 6 founders were obtained, of which two females transmitted the mutation to the F1 generation, implying that their germline is mutated. However, for these two heterozygous founders, the percentage of mutant pups at weaning was lower than expected for females and no hemizygous male was recovered (Fig.15a). This means that the mutation is sub-lethal either before birth or in the first days of life. Since males inheriting the mutation have only the mutant copy of *Ogt* (they are hemizygous for the mutation), they are more affected by *Ogt* disruption. At the blastocyst stage (E4) I found both hemizygous mutant males and heterozygous mutant females, but their percentage was already lower than expected at this stage (8% and 10% for hemizygous males and heterozygous females, respectively, when 25% is the expected) (Fig.15a), therefore cellular defects already accumulate in the oocyte before fertilization or during preimplantation development.

I analyzed the transcriptome of the F1 male blastocysts in order to investigate: i. the molecular phenotype of mutant *Ogt*$^{T931A/Y}$ compared to *Ogt*$^{WT/Y}$ and ii. the presence of differences between *Ogt*$^{WT/Y}$ embryos deriving from the F0 heterozygous *Ogt*$^{T931A/WT}$ mothers and *Ogt*$^{WT/Y}$ embryos from wild type mothers. A difference observed in the phenotype of genetically identical embryos (all wild type) which is solely due to the different genotype of the mother (wild type or heterozygous *Ogt*$^{T931A/WT}$) will be called a maternal effect phenotype. Three main hypotheses could explain a maternal effect phenotype: i. the maternal inheritance of an abnormal payload of glycosylated proteins due to presence pre-meiosis of the hypomorphic OGT protein, ii. a putative impaired communication with the somatic cells surrounding the oocyte or iii. a different gestational environment. All of these would in turn affect embryonic development independently of the genotype of the progeny.

For the transcriptomics experiment, I performed IVF of oocytes from two F0 *Ogt*$^{T931A/WT}$ females and two sibling wild type females using the same wild type sperm and collected single E4 blastocysts for mRNA-Seq (Picelli et al., 2014). Blastocysts from the two different pairs of mothers looked morphologically indistinguishable (Fig.S6a), therefore all healthy-looking ones were collected from both groups. Surprisingly, wild type blastocysts clearly separate in the PCA space based on the genotype of their mothers, in other words there is a clear maternal effect (Fig.15b). I can exclude batch effects thanks to the experimental design, specifically the choice

of sibling pairs of females and the use of the same sperm for all the four of them. Even more surprisingly, the only two mutant male blastocysts found among all embryos (deriving from the heterozygous mothers) cluster with the wild type embryos from wild type mothers and positioned between the latter and their sibling wild type embryos on the PCA (Fig.15c). I looked at the expression of the two enzymes responsible for the O-GlcNAc cycle, namely *Ogt* and *Oga*, whose levels are known to change in opposite ways as a cellular response to changes in O-GlcNAc levels (Park et al., 2017; Pravata et al., 2019b; Qian et al., 2018; Vaidyanathan et al., 2017b; Willems et al., 2017b; Z. Zhang et al., 2014). I found a downregulation of *Ogt* expression and an upregulation of *Oga* expression in the wild type blastocysts from heterozygous $Ogt^{T931A/WT}$ mothers (Fig.15d). This is an indication of an increased level of O-GlcNAc in the previous embryonic stages. In good agreement with the PCA clustering, the level of *Ogt* and *Oga* in the two mutant blastocysts is more similar to that of the embryos from wild type females (Fig.15d), suggesting that their O-GlcNAc level in the stage immediately before the stage of collection is more similar to the level found with wild type mothers.

I inquired about the features of gene expression change observed in wild type male blastocysts due to the maternal effect. Firstly, the magnitude of changes (Fig.15e) is overall similar to the one observed in Btgh-injected embryos (Fig.11c and 12a), another model of O-GlcNAc perturbation used in this study. Secondly, gene ontology terms related to ATP synthesis via oxidative phosphorylation and ribosome biogenesis are over-represented among upregulated genes (Fig.15f), while no significant terms are found for downregulated genes. This result is noticeable because the same pathways were enriched among downregulated genes after O-GlcNAc depletion via Btgh injection. Together with the expression of *Ogt* and *Oga,* this argues that the maternal effect observed in the blastocyst is a consequence of higher O-GlcNAc levels during earlier embryonic stages.

I speculate a possible explanation for the increased level of O-GlcNAc in wild type embryos from heterozygous mothers, which assumes the female diploid germline being the initial player of the maternal effect (Fig.15g). The presence of hypomorphic OGT protein molecules in the diploid female germ cells could lead to increased transcription of the *Ogt* gene from both the wild type and the hypomorphic allele, hence increased concentration of the OGT protein (including the wild type copy of OGT) in the protein payload of the mature haploid MII oocyte. This might imply a slightly increased O-GlcNAcylation in the mature oocyte and the early embryo produced by $Ogt^{T931A/WT}$ female. Importantly, elevated O-GlcNAc levels in the early

embryo can be compensated only from the 4/8-cell stage onwards, when the production of catalytically functional OGT from the embryonic genome seems to start (Fig.4c-d). By collecting E4 blastocysts, the transcriptomic snapshot captures the effect on gene expression of the O-GlcNAc level in the embryonic stages right before. At the morula stage, the O-GlcNAc level in wild type embryos from heterozygous mothers is supposedly still higher than in the embryos from wild type mothers (Fig.15g). In fact, there is a lag between an updated diminished level of OGT (due to less transcription of *Ogt* from the 4/8-cell stage) and an effectively diminished O-GlcNAc level, because O-GlcNAc residues inherited from previous developmental stages needs to be diluted or removed by OGA.

This speculation also explains the partial "rescue" of the maternal effect in the hemizygous embryos: as for the wild type embryos from the same heterozygous mothers, the hemizygous have higher O-GlcNAc level than the embryos from wild type mothers during the first embryonic cleavages, hence are predicted to reduce *Ogt*'s transcription from the 4/8-cell stage. However, $Ogt^{T931A/Y}$ hemizygous express an OGT protein that has ~80% reduced activity, leading to a faster rescue of O-GlcNAc levels to the ones observed for the embryos with wild type mothers (Fig.15g).

I analyzed the expression in the hemizygous $Ogt^{T931A/Y}$ blastocysts of some significantly upregulated genes due to maternal effect, focusing on genes belonging to the mitochondrial respiratory chain and ribosomal biogenesis pathways (Fig.15h). Notably, in most of the cases the mutant blastocysts have an intermediate expression of those genes between wild type (WT) blastocysts coming from WT mothers and WT blastocysts from heterozygous mothers. The model proposed above can explain this phenomenon, for the reason that follows. Fig.15h shows examples of genes that are upregulated in blastocysts (WT) produced by $Ogt^{T931A/WT}$ females (e.g. *Atp5pb*, *Rpl10*). Fig.12e shows that the same genes increase their expression across preimplantation stages in the wild type, starting from the 2-cell stage, therefore their final mRNA level in late blastocyst can in principle be influenced by changes in their gene expression regulation at various past stages. According to the model in Fig.15g, before the rescue of O-GlcNAc levels (i.e. before the morula stage), hemizygous embryos experience the same higher level of O-GlcNAc than their WT siblings, while from around the morula stage their O-GlcNAc level reaches similar levels to the WT blastocysts from WT mothers (Fig.15g). If O-GlcNAc modulates the expression of these genes at both earlier and later preimplantation stages, it follows that in hemizygous embryos these O-GlcNAc regulated genes will ultimately show an intermediate expression between the other two groups of embryos. It is likely that the impact

of the immediately previous stage on the mRNA level observed in the blastocyst is on average stronger, as shown for example by *Rpl10*, for which gene the hemizygous blastocysts completely resemble the WT blastocysts from WT mother. In agreement, at the morula or early blastocyst stage hemizygous embryos are predicted to have the same level of O-GlcNAc than WT embryos from WT mother.

There is one exception worth noticing: *H19*, the most upregulated genes due to the maternal effect (Fig.15e), whose expression in $Ogt^{T931A/Y}$ blastocysts resembles that of its siblings from heterozygous mothers, differently from most of the other genes. *H19* is an imprinted gene which starts to be expressed from the blastocyst stage onwards only from the paternal allele (Bartolomei et al., 1991). According to my model, the O-GlcNAc level in hemizygous blastocysts is more similar to WT blastocysts from WT mother than to their siblings (Fig.15g), thus the model does not seem to stand in front of this result. *H19* is paternally imprinted, meaning that it is monoallelically expressed from the maternal allele due to DNA methylation on the paternal allele (Bell and Felsenfeld, 2000), thus a defect in *de novo* methylation (occurring in the prospermatogonia) can be excluded. It is conceivable however that the silencing of the paternally inherited *H19* allele could be altered in the embryo with altered O-GlcNAcylation. In support of this idea there is the identification of *Ogt* in a CRISPR screen for regulators of imprinted genes (Butz et al., 2022). Without knowing the allelic identity of the RNA-Seq reads I cannot attribute the increase of *H19* expression to loss of imprinting. I had the allelic information in F1 hybrid blastocysts and E7 embryos developed from Btgh-injected zygotes, thus in these two datasets I computed the ratio of maternal over paternal reads (henceforth allelic ratio) for known imprinted genes with enough reads coverage on SNPs. Unfortunately, H19 did not have enough coverage, but nuclear O-GlcNAc depletion at preimplantation stages did not cause any significant change in the allelic ratio of any of the other imprinted genes analyzed (Fig.S6b). This argues against a role of O-GlcNAc in imprinted expression in the embryo. It remains possible that OGT (which is perturbed in the embryos from $Ogt^{T931A/WT}$ but not in the Btgh system) has a role in the maintenance of imprinting which is independent of O-GlcNAc or that Btgh activity is not enough to compete O-GlcNAcylation at the imprinted loci.

Finally, *H19* being a negative regulator of embryonic growth (Leighton et al., 1995), its expression could be stimulated mono-allelically from the blastocyst stage, in line with O-GlcNAc modulation of growth starting from the process of translation (see Discussion). These hypotheses remain to be tested.

In conclusions, I showed that the T92A hypomorphic mutant leads to a maternal effect, consisting in the upregulation of a small set of genes related to protein translation and mitochondrial respiration in the embryos deriving from heterozygous mothers, independently of the embryonic genotype. In addition, I provided a possible mechanism explaining the maternal effect, which is the inheritance by the early embryos of a higher OGT and O-GlcNAc level from the oocyte. In the hemizygous embryos, the impaired catalytic activity of the newly transcribed embryonic OGT leads to a faster rescue of O-GlcNAc level to the ones observed in embryos from WT mothers.

Finally, although the O-GlcNAc level of hemizygous blastocysts is almost comparable to wild type levels, the absence of live mutant males at birth suggests that this level cannot be maintained much further. From the blastocyst onwards, the only present OGT protein carries the point mutation that strongly reduces its activity. Most likely, in these embryos a compensatory increase in *Ogt* transcription is not sufficient to maintain O-GlcNAc homeostasis along the next phases of development.

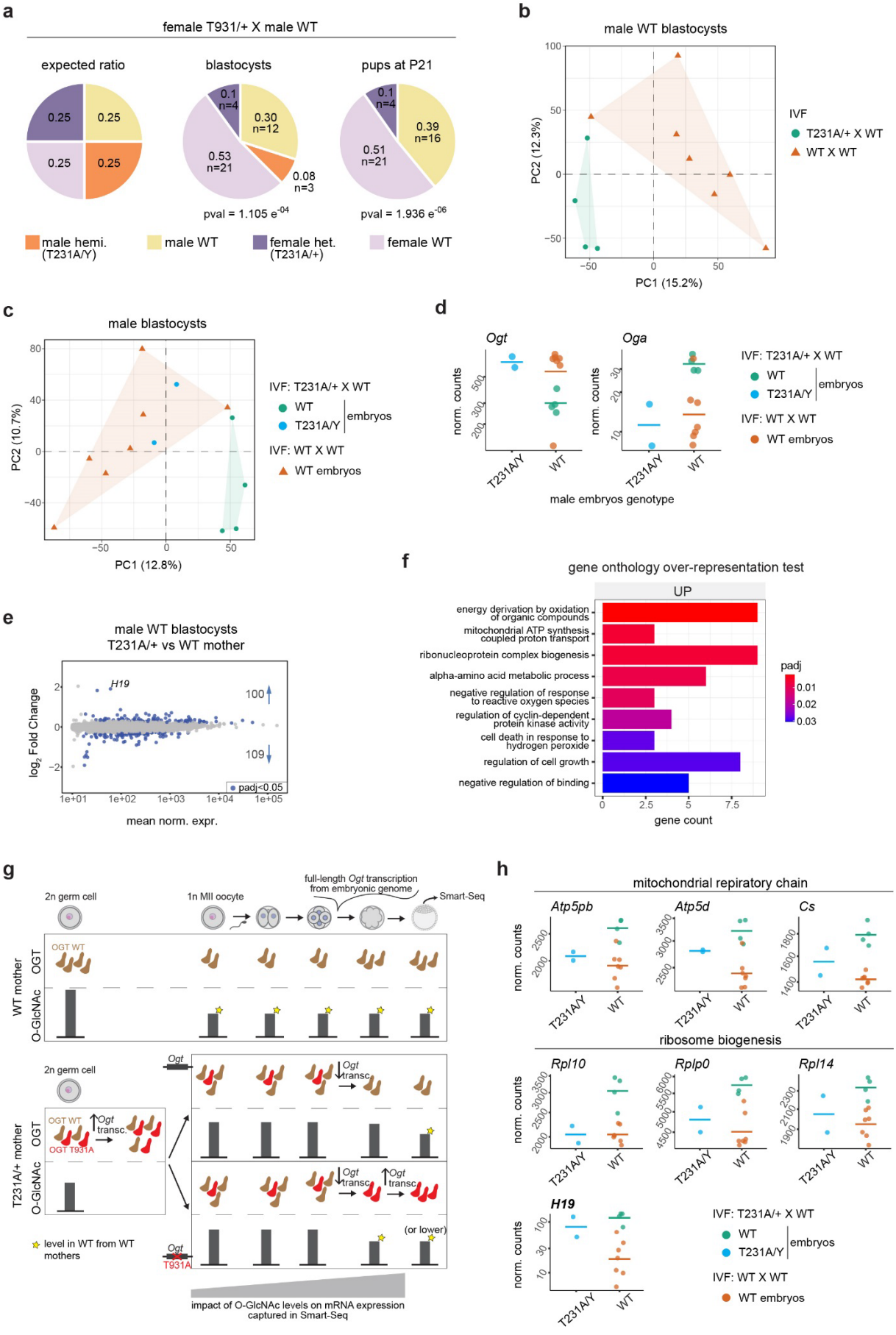

a
female T931/+ X male WT
expected ratio
blastocysts
pups at P21
0.25
0.25
0.25
0.25
0.1 n=4
0.30 n=12
0.53 n=21
0.08 n=3
0.1 n=4
0.39 n=16
0.51 n=21
pval = 1.105 e-04
pval = 1.936 e-06
male hemi. (T231A/Y)
male WT
female het. (T231A/+)
female WT
b
male WT blastocysts
PC2 (12.3%)
PC1 (15.2%)
IVF
T231A/+ X WT
WT X WT
c
male blastocysts
PC2 (10.7%)
PC1 (12.8%)
IVF: T231A/+ X WT
WT
T231A/Y
embryos
IVF: WT X WT
WT embryos
d
Ogt
Oga
norm. counts
T231A/Y
WT
male embryos genotype
IVF: T231A/+ X WT
WT
T231A/Y
embryos
IVF: WT X WT
WT embryos
e
male WT blastocysts
T231A/+ vs WT mother
H19
100
109
log2 Fold Change
padj<0.05
mean norm. expr.
f
gene onthology over-representation test
UP
energy derivation by oxidation of organic compounds
mitochondrial ATP synthesis coupled proton transport
ribonucleoprotein complex biogenesis
alpha-amino acid metabolic process
negative regulation of response to reactive oxygen species
regulation of cyclin-dependent protein kinase activity
cell death in response to hydrogen peroxide
regulation of cell growth
negative regulation of binding
gene count
padj
0.01
0.02
0.03
g
2n germ cell
1n MII oocyte
full-length Ogt transcription from embryonic genome
Smart-Seq
WT mother
OGT
O-GlcNAc
OGT WT
T231A/+ mother
OGT T931A
Ogt transc.
level in WT from WT mothers
(or lower)
impact of O-GlcNAc levels on mRNA expression captured in Smart-Seq
h
mitochondrial repiratory chain
Atp5pb
Atp5d
Cs
ribosome biogenesis
Rpl10
Rplp0
Rpl14
H19
norm. counts
T231A/Y
WT
IVF: T231A/+ X WT
WT
T231A/Y
embryos
IVF: WT X WT
WT embryos

Figure 15: Maternal effect of the hypomorphic mutant *Ogt*^{T931A}.
(a) Distribution of genotypes as expected in the newborns by mendelian ratio (left) and as observed in E4 blastocysts (middle) and pups at weaning (right) from *Ogt*$^{T931A/WT}$ females x *Ogt*$^{WT/WT}$ males. P-value was computed using the chi-squared test for independence.
(b) PCA of male E4 blastocysts from IVF of oocytes from a pair of *Ogt*$^{T931A/WT}$ females and a pair of sibling *Ogt*$^{T931A/WT}$ females with the same wild type sperm, after excluding the two hemizygous *Ogt*$^{T931A/Y}$ blastocysts from the dataset. Blastocysts had been genotyped for sex and *Ogt* genotype using their cDNA, as described in Methods. PCA was performed as described in Methods. The variance explained by each principal component is in parentheses.
(c) Same as in (b), but including the two hemizygous blastocysts in the analysis.
(d) DESeq2-normalized counts of the two enzymes of the O-GlcNAc cycle. Mean of counts for the three groups of embryos is shown. y-axis is in $\log_{10}$ scale.
(e) MA-plots from DESeq2 differential expression analysis of WT blastocysts from *Ogt*$^{T931A/WT}$ mothers versus WT blastocysts from WT mothers, hence gene expression changes are a maternal effect. Only genes with mean of DESeq2-normalized counts ≥ 10 are shown. All genes with adj. p-value < 0.05, any Log2FC are colored and their number is indicated.
(f) Gene ontology (GO) over-representation analysis of DEGs in (e). All enriched Biological Process GO terms found for the upregulated DEGs are shown, based on adj. p-value. No terms were found for downregulated genes. Gene count = number of DEGs belonging to the term.
(g) Speculative model to explain the maternal effect and the higher similarity of homozygous blastocysts to the WT blastocysts from WT mothers than to the WT siblings from heterozygous mother. The model is described in the text.
(h) DESeq2-normalized counts of some relevant significantly upregulated genes in WT blastocysts due to maternal effect (colored in (e)). *H19* is bold because it is the only one for which the expression level in the hemizygous blastocysts is comparable to that observed in their WT siblings. A possible explanation is described in the text. Mean of counts for the three groups of embryos is shown. y-axis is in $\log_{10}$ scale.

Discussion

O-GlcNAcylation is a very intriguing and still obscure post-translational modification. It is present on thousands of intracellular proteins, however it is catalyzed by one single enzyme, OGT, which is remarkably conserved across the animal kingdom and has a conserved role in development in evolutionary distant organisms, such as flies and mammals. In mammals, OGT has evolved as essential for embryonic development. More specifically, the maternal copy of the gene is necessary for the embryo to develop after the blastocyst stage, which strongly argues for a role either in the maturation of the oocyte or in the first embryonic stages. In spite of this, the role of OGT and O-GlcNAc in the developing mammalian embryo has remained untested. In fact, answering this question in mammals is technically challenging, due to the embryonic lethality caused by the lack of a functional *Ogt*. Moreover, OGT's pleiotropy makes the understanding of the roles of OGT and O-GlcNAc during early development a very complex problem. O-GlcNAc has been described to regulate many cytosolic pathways and cellular machineries and the nuclear compartment is even more enriched of O-GlcNAcylated proteins (Wulff-Fuentes et al., 2021a). Even though many reported O-GlcNAcylated residues have no validation or lack a proper investigation of the functional significance for the protein target, the number of OGT targets remains considerable. This picture implies that OGT activity can act on many different cellular activities at the same time. If we consider that the donor substrate used by OGT, UDP-GlcNAc, is the product of a metabolic pathway that channels 2-3% of the glucose of the cell, we can then conceive this modification as suitable to mediate the adaptation of the cell status to changes in the environment or in the intracellular demands, potentially also through regulating gene expression.

In this study, I start characterizing the role of O-GlcNAc in the early mammalian embryo for the first time. I provide the first spatiotemporal profiling of OGT and O-GlcNAc throughout preimplantation development, then probe their function in the live embryo thanks to the development of novel functional strategies.

In order to understand the spatiotemporal dynamics of OGT and O-GlcNAc from the zygote to implantation, I used both immunofluorescence (IF) and transcriptional data. From these two techniques, three main concepts emerge. First of all, in the zygote there seems to be a separation of function between OGT and its catalytic product O-GlcNAc, based on the asymmetric

presence of OGT in the two pronuclei and the parallel symmetric pronuclear O-GlcNAcylation (Fig.4a). I am aware that this claim is only based on the immunofluorescence result and that this could be invalidated in case of a poorly specific signal of one of the antibodies used. I proved the specificity of the O-GlcNAc IF signal in preimplantation embryos after O-GlcNAc removal through recombinant OGA (Fig.8b), while I could not validate the OGT IF signal in preimplantation embryos using the strategies I developed. However, the monoclonal ab177941 antibody used to detect OGT gives almost a unique band on western blot and this band is specific based on the shift induced by the AID tag (Fig.5b) and by auxin-induced disappearance (Fig.5d). Furthermore, OGT pronuclear asymmetry had already been observed using another OGT antibody (Nagaraj et al., 2017). To my knowledge, there are not many other examples of proteins which are asymmetric between zygotic pronuclei that could explain a non-specific signal of the anti-OGT antibody. Instead, an asymmetric pronuclear signal is shown by some chromatin modifications. One of those is 5-hydroxymethylcytosine (5hmC), which is catalyzed by TET enzymes and is enriched at the paternal pronucleus as OGT. Notably, TETs are one of the most abundant OGT interactors when performing biochemical pull-downs of OGT (Vella et al., 2013a). The possibility of OGT function in the paternal pronucleus being related to the process of paternal DNA demethylation is intriguing, also because such a process is only partially understood and not all players have been identified yet (Amouroux et al., 2016). Unfortunately, I could not test this possibility using my strategies. The OGT-degron mouse model was the most adequate system to this aim, but it showed to be inefficient in the early embryo. At the end, I will propose some experiments to tackle this question.

The second concept emerging from the spatiotemporal profiling of O-GlcNAc was its abundance in the nucleus from the 2-cell until the blastocyst stage. Notably, this was again not the case for OGT, whose signal is more homogeneous in the cell. The analysis of the transcript isoforms throughout preimplantation stages provide a possible explanation, which is the third important concept emerging from this first part of the study: there is little transcription of the full-length isoform of *Ogt* (encoding the nuclear-cytoplasmic catalytic ncOGT) at EGA and for the next couple of embryonic cleavages, after which *ncOgt* transcription raises again. Moreover, at EGA, a shorter transcript is produced which contains an intron (intron 4) previously described to trigger either nonsense-mediated decay or nuclear retention of *Ogt* (Park et al., 2017; Tan et al., 2020). Blastomeres undergoing the major wave of ZGA modulate the composition of the spliceosome and employ a temporary unique splicing regulation, leading to the highest levels of exon skipping reported for any cell or tissue type, often resulting in ORF interruption (Wyatt et al., 2022). Hence, the mechanism of production of the short *Ogt* transcript could be the result

of an EGA-specific alternative splicing. In any case, this shorter isoform does not seem to bear the catalytic domain. This suggests that the concentration of ncOGT does not increase during embryonic cleavages. I hypothesize that this temporary phenomenon could be necessary to maintain O-GlcNAc homeostasis in blastomeres, which become smaller at each division since this is not accompanied by embryonic growth.

In order to test the role of O-GlcNAc across preimplantation key stages, I developed two different strategies which allow a separation of function between OGT and O-GlcNAc. The first approach is a novel *Ogt*-degron mouse allele, for the acute and fast depletion of endogenous OGT in the preimplantation live embryo. The second one is the injection of a recombinant bacterial isoform of OGA (called Btgh) in the zygote, for the fast removal of the O-GlcNAc modification throughout preimplantation. This second strategy allows even a finer separation of functions: the recombinant OGA can be targeted to a specific cellular compartment by fusing its coding sequence with different localization signaling peptides. In my case, I fused it with nuclear localization signals in order to specifically investigate the function of O-GlcNAc in the nucleus and in transcriptional regulation. For a pleiotropic modification like O-GlcNAcylation, the possibility to exclude the cytosolic compartment from the perturbation is a great advantage, because it increases the probability to probe a molecular phenotype that is due to a change in transcriptional regulation. In this regard, I have to mention that, besides being present on factors localized in the nucleoplasm and on chromatin, O-GlcNAc is highly abundant on the disordered FG-Nucleoporins (FG-Nups) subunits of the Nuclear Pore Complex, thus perturbing nuclear O-GlcNAc might also affect the functionality of the Nuclear Pore Complex. Hence, differences in gene expression observed after nuclear O-GlcNAc depletion could derive from the modulation of the function or structure of a transcriptional regulator but also potentially from the differential nuclear import/export of proteins or RNA molecules. In addition, specific Nups were shown to be involved in transcriptional regulation themselves (Toda et al., 2017).
In spite of the nuclear targeting, a possible question that could be raised to my approaches is that they remain non-specific in terms of the O-GlcNAcylated target or pathway that is perturbed. In my view this is a very necessary step in the investigation of O-GlcNAc function in the early mammalian embryo, because the role of O-GlcNAc in this system is completely unknown. An untargeted approach allows the unbiased measurement of the main molecular phenotype emerging in the absence of O-GlcNAc in the embryo (including the O-GlcNAc maternal payload when using the Btgh system). The resulting data are rich and complex, but

they can reveal what is the most prominent functional significance of this modification in the early embryo and inform the choice of the most relevant follow-up experiments.

As far as the OGT-degron strategy is concerned, it was inefficient to cause a decrease in O-GlcNAc levels after OGT depletion from the zygote to the morula stage. An insufficient level of expression of OsTIR or the lower auxin concentration used in the embryos are two possible explanations. Moreover, in light of the inheritance of the maternal payload of O-GlcNAc, a degron system to deplete OGT could be in general inappropriate to tackle the role of O-GlcNAc in the early embryo. However, it remains a powerful approach to study the role of OGT (independently of O-GlcNAc) in the zygote, since it could in principle halve OGT's concentration in ~1h (Camlin and Evans, 2019; Gu et al., 2018). This makes it worth the creation of a more efficient *Ogt*-degron mouse model. In fact, nowadays, there are more effective versions of the AID system available (Yesbolatova et al., 2020).

On the other hand, I showed that Btgh injection was extremely efficient to deplete nuclear O-GlcNAc as early as in the pre-EGA 2-cell embryos and until the blastocyst stage, in an exquisitely nuclear-specific manner. This empowered me to study the impact of significantly lower levels of O-GlcNAc on two processes for which scientific literature points to a potential involvement of O-GlcNAc. The first one is EGA, the first transcriptional event of the new individual. I speculated a possible role of O-GlcNAc in EGA mostly based on the mapping of O-GlcNAc on the CTD of RNA PolII. PolII undergoes a series of phosphorylation events that are responsible for the transition through the various phases of the transcriptional cycle. Since O-GlcNAc competes with phosphorylation at some residues, it could participate in the regulation of the transcriptional cycle. In addition, in the last decade, a model emerged associating CTD phosphorylations with liquid-liquid phase separation and the formation of liquid-like droplets facilitating transcription (Harlen and Churchman, 2017). The competition of O-GlcNac with phosphorylation has been implicated in the aggregation properties of low-complexity domains (Lv et al., 2022; O'Donnell et al., 2004a), thus it was tempting to speculate the participation of O-GlcNAc in regulating transcriptional phase separation. Besides a possible role on PolII, the O-GlcNAcylation of transcription factors involved in EGA could be necessary for the onset of embryonic transcription, a phenomenon which is only partially understood in mammals. Using the Btgh system, I found that a significant depletion of nuclear O-GlcNAc did not impede the occurrence of EGA at the right time. While Btgh injection reduces O-GlcNAc to undetectable levels in the nucleus, I cannot formally exclude that residual levels of O-

GlcNAc on PolII CTD or specific factors are enough to sustain the process. Nonetheless, I can conclude that O-GlcNAc homeostasis is not necessary for EGA.

The second early developmental process for which scientific literature suggested the participation of O-GlcNAc is the establishment of the pluripotency network. In mESCs, the O-GlcNAcylation of OCT4 has been shown to be necessary for the activation of some of its pluripotency target genes (Jang et al., 2012). Surprisingly, there was no detectable effect of nuclear O-GlcNAC depletion on the specification of the blastocyst. I can provide possible explanations for this apparent discrepancy. Firstly, pluripotency is a very transient state *in vivo*, from which the embryo needs to rapidly transit away at the start of gastrulation. Instead, the *in vitro* culture of embryonic stem cells is by definition characterized by the capacity to indefinitely maintain pluripotency without differentiating. O-GlcNAc might be dispensable for the transitional *in vivo* pluripotency. Secondly, the evidence provided for the functional relevance of O-GlcNAcylated OCT4 is based on the substitution of the O-GlcNAcylation residue, which can have unexpected effects on the function of this pioneer factor. Moreover, the substitution of serine or threonine to alanine blocks both O-GlcNAcylation and phosphorylation, which confounds the results.

After specifically investigating the role of O-GlcNAc in EGA and pluripotency, I dissected for the first time the transcriptional response to O-GlcNAc depletion *in vivo.* In spite of an apparently normal morphology and a normal rate of embryonic cleavage, O-GlcNAc-depleted morulae showed a slight developmental delay, only detectable in the transcriptome (Fig.9b-c). This implies that some process which starts to be perturbed at a previous stage is having a developmental impact. The transcriptome of O-GlcNAc-depleted 2-cell embryos showed a downregulation of genes related to protein synthesis, aerobic respiration and vesicle coating; the decrease in one or more of these pathways could be responsible for slowing development at the next stages. The delay of O-GlcNAc-depleted morulae also means that gene expression changes from the morula onwards will be a combination of an ongoing effect of O-GlcNAc depletion on gene expression and of a transcriptome resembling a slightly earlier developmental stage. To start disentangling the two effects, I looked at the transcriptional response to O-GlcNAc depletion found at later stages which was already appearing in 2-cell embryos. A few genes enriched in specific cellular pathways were coherently deregulated starting from the 2-cell stage: on one side, genes participating in the mitotic phase of the cell cycle were upregulated, together with an enzyme involved in branched glycan biosynthesis (*B4galt3*) and with glutathione transporter *Slc25a39*; on the other, two ribosomal proteins, one subunit of

mitochondrial ATP synthase and clathrin, which coats intracellular vesicles, were downregulated. The constant change in expression of these genes across the three stages suggests that if the related pathways are observed in O-GlcNAc-depleted morulae and blastocysts this is likely not due to the delayed transcriptome but to a transcriptional effect of O-GlcNAc depletion across preimplantation development. A second hint to discover the cellular pathways regulated by O-GlcNAc in early embryos came from the OGT-degron system. In fact, in spite of the unsuccessful application in the embryo, the OGT-degron was fast and highly efficient to deplete OGT in MEFs, providing new molecular insights into the effect of OGT depletion. This information turned out to be precious when comparing it to the transcriptomics results in preimplantation embryos after nuclear O-GlcNAc depletion via Btgh injection.

I will now describe some of the common features of the transcriptome of OGT-depleted MEFs and the transcriptome of O-GlcNAc-depleted preimplantation embryos, which will lead to my final interpretation.

The first observation is that gene expression changes were overall modest in magnitude, albeit affecting hundreds of genes, whispering a coordinated effect of O-GlcNAc depletion on many cellular functions. Secondly, in both systems, although the number of upregulated and downregulated differentially expressed genes was very similar, only downregulated genes were participating in common cellular functions. Instead, upregulated genes in MEFs were almost not enriched for any gene ontology term, while in Btgh-injected embryos by only one specific pathway, namely the transition to the mitotic phase of the cell cycle. The genes which are constantly upregulated in O-GlcNAc-depleted embryos are not repressed in unperturbed embryos at the same developmental stages (Fig.12e), hence it is unlikely that they are repressed by O-GlcNAc in normal conditions. I interpret the upregulation of cell cycle genes as an acute response to O-GlcNAc depletion, which could for example be caused by the impairment of the cell cycle, a hypothesis which can have various non-exclusive explanations. OGT catalyzes the maturation of cell cycle factor HCF-1 through O-GlcNAcylation followed by proteolysis. The resulting HCF-1 subunits regulate the synthesis of S phase genes and the progression through the M phase, respectively (Julien and Herr, 2003; Tyagi et al., 2007b). OGT is unperturbed in the Btgh system and the maturation of HCF-1 should be unaffected, however it is unknown whether the O-GlcNAc moieties on the mature HCF-1 subunits are relevant for the subunits' function in the cell cycle. A second explanation for cell cycle impairment could be an impact of the lack of O-GlcNAc on cytoskeletal functions. In fact, although Btgh is efficiently targeted

to the nucleus at interphase, the recombinant protein spreads to the whole cell when the nuclear envelope breaks during mitosis. The staining of OGT in the oocyte showed the enrichment of OGT on the mitotic spindle (Fig.4a), but the same is not observed for O-GlcNAc (Fig.4b). At the time of writing this dissertation, it was reported that germ cells of aging male mice show a change in O-GlcNAc homeostasis and that restoring the latter can partially rescue the increase of meiosis defects with age (Qian et al., 2023). Conserved O-GlcNAc sites have been found on both β-tubulin and actin, where they have been predicted to be important for polymerization of the respective monomers and to be involved in the neuronal phenotype of OGT-XLID patients (Mitchell et al., 2022). A third explanation for cell cycle impairment could be the transcriptional downregulation of cytoskeletal functions by O-GlcNAc depletion itself, as shown by the gene sets analysis in both MEFs and Btgh-injected embryos (Fig.6g and 12b). In MEFs, this could be linked to the reduced activity of the mesenchymal transcriptional network following OGT depletion, but this cannot be the case for Btgh-injected embryos. A fourth explanation is provided by a very recent study showing that the O-GlcNAcylation of a specific residue of histone H4 contributes to the activation of replication origins and that removing O-GlcNAc reduces the number of active origins during S phase (Zou et al., 2023). Finally, if the cell cycle defect involves chromosome separation at mitosis, the cell manages to overcome the problem because I found that O-GlcNAc-depleted embryos are not aneuploid.

As mentioned above, cell cycle genes are not upregulated in OGT-depleted MEFs. The difference with the Btgh system could be due to the slower rate of cellular division of MEFs (length of one MEFs' cell cycle is >24 h) compared to the rate of embryonic cleavages, happening around every 12 h (Ciemerych and Sicinski, 2005). Therefore, I cannot exclude that there would be mitotic genes upregulated also in MEFs with a longer axin treatment.

While the upregulation response might be an acute response to O-GlcNAc depletion, the downregulation response after O-GlcNAc depletion in MEFs and the three stages of preimplantation embryos give insights into cellular functions whose default transcriptional levels in normal conditions are regulated by O-GlcNAc. Two processes, probably interrelated, affected in MEFs and embryos are intracellular vesicle trafficking and protein localization to membranes and the extracellular space. For the MEFs dataset, I provided one possible explanation for this effect, which involves the increase in free UDP-GlcNAc after OGT depletion. Such an increase could compromise the equilibrium of donors' concentration for the biosynthesis of branched protein glycosylations, which are abundant on membranes and face the extracellular space. The biosynthesis of branched glycans resides in the ER-Golgi network

(Varki et al., 2022). A defect in glycans' biosynthesis could be sensed by the cell and trigger an adaptation response involving genes of this network to avoid glycoprotein misfolding. OGT catalytic activity is unperturbed (although the gene is upregulated but only at the blastocyst stage, Fig.S5b) in the Btgh system, hence I would not expect the concentration of free UDP-GlcNAc to change significantly. However, the upregulation of branched glycans' glycosyltransferase *B4galt3* in O-GlcNAc-depleted embryos could point to a similar mechanism also induced by Btgh activity.

Remarkably, a significant number of components of two essential cellular machineries are downregulated in MEFs and embryos after O-GlcNAc depletion: the ribosome and the proteasome. Other proteins involved in protein translation, such as enzymes of mRNA metabolism and translation initiation factors, are also transcriptionally dampened. The processes of protein translation and degradation are coupled and related to the cellular rate of protein synthesis. Noteworthy, translation is the first downregulated pathway appearing in O-GlcNAC-depleted embryos, already at the 2-cell stage (Fig.11d). This is also one of the most transcriptionally dynamic processes in the 2-cell embryo (Gao et al., 2017). Because of this, the downregulation of translation-related genes could be a mere consequence of a developmental delay already starting at the 2-cell stage, but this possibility is not supported by the seemingly normal maternal-to-embryonic transition shown by O-GlcNAc-depleted 2-cell embryos (Fig.9c and Results section 2.2.5.1). Moreover, the same result in MEFs argues in favor of the transcriptional regulation of these genes by O-GlcNAc. As per all genes deregulated by O-GlcNAc depletion, the magnitude of change of the mRNA level of ribosome-related genes is small (10-20% decrease). Yet, it could be relevant for such abundant proteins whose stoichiometry can in principle control the cellular rate of translation (as discussed again later).

The effect of O-GlcNAc depletion on genes associated with translation might be even more relevant if combined with the other cellular pathway that is recurrently and constantly downregulated in Btgh-injected embryos, namely the mitochondrial respiratory chain. The mammalian preimplantation embryo mostly relies on mitochondrial respiration for ATP production and OxPhos genes are gradually upregulated from the 2-cell stage to the blastocyst (Malkowska et al., 2022). This could be one of the reasons why this pathway is observed in embryos but not in MEFs after O-GlcNAc depletion.

It was shown that a higher rate of mitochondrial respiration can increase the level of protein translation (Diaz-Cuadros et al., 2023). Apparently, this was not mediated by ATP production but by the increase in the NAD^+/NADH ratio (Diaz-Cuadros et al., 2023). Solely from my

transcriptomics readout, I am not able to predict the consequence of the reduced expression of a group of OxPhos genes on mitochondrial metabolic flux and even less on the level of specific products (such as NAD^+) versus others. What I wish to point out is the possibility that gene expression changes affecting the respirasome and translation-related genes could together contribute to the same O-GlcNAc-mediated final effect on protein synthesis.

In summary, after comparing the transcriptional effect of removing O-GlcNAc in MEFs and in embryos at different stages, I can conclude that the decrease in translation-related and OxPhos-related genes is most likely not a secondary effect of the slight developmental delay observed in the O-GlcNAc-depleted morula's transcriptome. Instead, these results uncover a potential regulation of these two pathways by O-GlcNAc in the early embryo. Such a regulation involves only specific subunits and the effect is of relatively low magnitude, therefore it does not resemble a transcriptional activation, instead it has the features of a modulation. The transcriptional modulation of OxOhos and translation by O-GlcNAc has several implications: it could be an O-GlcNAc-mediated means of adapting embryonic metabolism to the changes in the maternal environment, for example in the case of stress during pregnancy or different nutrient levels (Howerton et al., 2013).

The regulation of the translational machinery by the O-GlcNAcylation of translation initiation factors eIF4A and eIF4G has previously been reported (Li et al., 2019; Shu et al., 2021), for example in response to nutrient stress (Shu et al., 2021). In my study, I observe a novel transcriptional modulation of the translational machinery by O-GlcNAc. My first speculation is that this could lead to a modulation of the rate of protein synthesis. Moreover, I am proposing a possible scenario where O-GlcNAc orchestrates a dampening of transcription not only of translation-related genes but also other genes involved in the production of the energy potential needed for the protein biosynthetic process.

I wish to underline that the hypothesis of O-GlcNAc modulation of protein synthesis needs to be functionally tested. It needs to be measured whether the transcriptional modulation of ribosome-related genes (maybe together with OxPhos-related genes) observed in this study is associated with an actual reduced rate of protein synthesis. The measurement of translation elongation rate in O-GlcNAc-depleted preimplantation embryos is one of the experiments I propose as a follow-up of this study. Moreover, it would be interesting to explain if and how the transcriptional modulation of a subset of ribosomal proteins (RPs) - together with other

translation factors - could lead to an overall decrease of translation (or a shift towards specific types of mRNAs). It is possible that the specific RPs regulated by O-GlcNAc are rate limiting for the biogenesis or the function of ribosomes. For example, they might be the ones whose upregulation is specifically required at certain developmental times. Along this line, an analysis of the expression dynamics of RPs across preimplantation stages is ongoing. Another aspect to consider is that, as shown in mice and yeasts (Slavov et al., 2015), RPs that compose ribosomes associated with a few mRNA molecules are different to the ones composing polysomes. Polysomes have been associated with a higher translational rate and, in good agreement, fitness of yeast strains correlates with the level of expression of polysome-associated RPs (Slavov et al., 2015). Curiously, RPs which are constantly downregulated in O-GlcNAc-depleted preimplantation embryos (*Rpl10*, *Rps5*) are found in the polysome fraction in mESCs (Slavov et al., 2015), hence they could be particularly rate limiting for the overall cellular level of protein synthesis. The early embryo might have a different RPs' stoichiometry than mESCs, but a similar principle being behind the choice of the specific subunits transcriptionally regulated by O-GlcNAc is plausible.

If O-GlcNAc-mediated transcriptional modulation of translation-associated genes is effectively leading to an increased rate of protein synthesis, my second speculation is that O-GlcNAc can ultimately modulate embryonic growth. This idea is supported by a study showing in humans and mice that the rate of protein synthesis correlates with the different rate of growth of the two mammalian species (Diaz-Cuadros et al., 2023). Of note, developmental delay is a common feature of individuals with mutations in human OGT (Pravata et al., 2020). Other two results, if taken together, seem to describe an O-GlcNAc-mediated mechanism to control growth also in the adult organism: *Ogt* deletion in the paraventricular neurons of the hypothalamus impairs satiety and leads to overgrowth (Lagerlöf et al., 2016), while deleting *Ogt* in another type of hypothalamic neurons increases fat browning and protects against obesity (Ruan et al., 2014).

Notably, the hypothesis of O-GlcNac-mediated modulation of embryonic growth could explain the phenotype of E7 embryos grown *in vivo* from O-GlcNAc-depleted zygotes: a higher proportion of O-GlcNAc-depleted embryos has a smaller size postimplantation (Fig.13b). The effect of O-GlcNAc on cell growth is unlikely to be observed in O-GlcNAc-depleted preimplantation embryos because cell division until the blastocyst stage is not accompanied by embryonic growth, which starts at the mid-to-late blastocyst transition. Importantly, the same E7 embryos also show a retarded gastrulation (Fig.13f). Since the onset of gastrulation has been

found to be coupled with the number of cells (Kojima et al., 2014), I propose a possible explanation for retarded gastrulation which involves a reduced cell number due to intercellular competition. It has been reported that differentiation of pluripotent cells is followed by the start of an intercellular competition which results in the apoptotic elimination of cells with lower fitness (Clavería et al., 2013). Lower fitness can be represented, for example, by the activation of p53 (Bowling et al., 2018) or by reduced c-Myc's activity (Clavería et al., 2013). In the case of p53, the induction of apoptosis has been shown to be mediated by the reduction of mTORC1 signaling (Bowling et al., 2018). O-GlcNAc depletion is likely not homogenous in all cells after Btgh starts being diluted. In O-GlcNAc-depleted cells in the peri-implantation embryo, the lower rate of protein synthesis and maybe metabolic fluxes could be sensed as reduced fitness and lead to the elimination of cells with lower O-GlcNAc. This effect could be mediated by c-Myc, mTORC1 or p53 signaling or by a different pathway. c-Myc has been reported to be O-GlcNAcylated (T.-Y. Chou et al., 1995; T. Y. Chou et al., 1995) and it is one of the TFs whose binding sites are enriched among downregulated genes in Btgh-injected 2-cell embryos (Fig.12f). In addition, c-Myc's targets, together with targets of mTORC1 signaling, are enriched among downregulated genes in OGT-depleted MEFs. Another option leading to the elimination of O-GlcNAc-depleted cells not involving the effect on metabolism could involve the impairment of cell cycle by reduced O-GlcNAc, which I discussed earlier. A disruption of the mitotic spindle or a DNA break can trigger the activation of transcription factor p53, which can block the cell cycle via the transcriptional induction of cell cycle inhibitors such as p21 (Chen, 2016). In O-GlcNAc-depleted embryos, the expression of *Trp53* (encoding for p53) and *Cdkn1a* (encoding for p21) is on average slightly higher (but not significantly; adj. P-value > 0.05) only at the 2-cell stage. It is possible that at later stages the level of these genes is already heterogeneous among cells, hence difficult to detect without a single-cell resolution. However, my data do not clearly support p53 activation, at least at preimplantation stages.

There are other hypotheses for retarded gastrulation which do not involve intracellular competition. The first hypothesis is a reduced activity of Myc in the blastocyst, which was shown to induce a dormant condition resembling diapause (Scognamiglio et al., 2016). Diapause is a reversible embryonic state of metabolic, translational and proliferative quiescence, which does not affect pluripotency (Murphy, 2020). Genes affected by *Myc* KD are different from the ones found deregulated in our study and Btgh-injected blastocysts did not show the typical oblong morphology of diapause, therefore a reduced activity of c-Myc is not enough to explain the transcriptional phenotype of O-GlcNAc-depleted embryos. However, it

could contribute to the gastrulation delay. In addition, the example of Myc-induced diapause demonstrates the principle that a reduction in the biosynthetic potential can pause development without affecting pluripotency, hence the same could be possible after O-GlcNAc depletion, where indeed pluripotency is unaffected.

Another hypothesis to explain the retarded gastrulation is that the epithelial-to-mesenchymal transition (EMT), which is necessary for the ingression of cells forming the primitive streak, is delayed by O-GlcNAc depletion. In support of this idea, mesenchymal differentiation is downregulated in OGT-depleted MEFs.

Finally, O-GlcNAc reduction could impair the establishment of DNA methylation (DNAme) in the differentiating mesoderm. From a few studies, DNAme of regulatory elements is emerging as playing a role during lineage differentiation (Cheng et al., 2022; Sardina et al., 2018). Loss of function of the three TET proteins from implantation led to a massively perturbed gastrulation which was associated with DNA hypermethylation at some genomic loci, including at putative enhancers (Cheng et al., 2022). The hypothesis of O-GlcNAc regulation of DNAme has not been tested in this study and, to the best of my knowledge, in literature it would be exclusively and indirectly supported by OGT interaction with TET proteins (Vella et al., 2013a), by O-GlcNAc-mediated repression of LTRs transposable elements (Boulard et al., 2020) and by the recently reported O-GlcNAc modification of DNMT1 (Shin et al., 2022). In my study, monoallelic expression of imprinted genes, which depends on DNAme, is maintained in O-GlcNAc-depleted embryos (Fig.S6b), as well as the repression of transposable elements (Fig.10b). On the other side, the imprinted non-coding RNA *H19* is the most upregulated gene in $Ogt^{T931A/Y}$ blastocysts; *Dppa3*, which contributes to the DNA demethylation of the maternal genome during embryonic cleavages (Li et al., 2018), is one of the most upregulated genes in auxin-treated Ogt^{AID} blastocysts. Differently from the injection of Btgh, in $Ogt^{T931A/Y}$ embryos and in the OGT-degron system the perturbation directly affects the OGT protein. It is possible that the removal of nuclear O-GlcNAc by Btgh is not sufficient to impede OGT's function in DNAme dynamics, for example if this lies in the OGT-TETs complexes. Within the OGT-TET complex, O-GlcNAcylation might be constantly ongoing and spatially concentrated such that the activity of Btgh might be inconsequential. Alternatively, a catalytically-independent role for OGT in DNAme dynamics could also solve this discrepancy. In any case, based on the data in this study and in literature, supposing that OGT may have a role in DNAme, it is not possible to predict whether this is in DNAme maintenance or removal.

Finally, in this study, I designed and characterized mouse models bearing hypomorphic variants of the *Ogt* gene. If viable, these models would have allowed the study of the molecular phenotype of mice and embryos with significantly reduced OGT activity. One of these mutants, bearing the T931A mutation and showing embryonic sub-lethality, gave the most noticeable result. The study of this mouse model showed a clear maternal effect phenotype, namely the transcriptomes of the embryos with the same (wild type) genotype was different depending on the genotype of the mother (either wild type or $Ogt^{T931A/WT}$). Remarkably, the maternal effect consisted in a low-magnitude but spread transcriptional response characterized by the upregulation of genes of mitochondrial respiration and associated to translation (Fig.15f), including the same genes which are part of the shared transcriptional response in Btgh-injected embryos (Fig.15h). The remarkable overlap of the molecular pathways in the two systems of O-GlcNAc perturbation further substantiates the regulation of these processes by O-GlcNAc during *in vivo* early mammalian development. Moreover, the different direction of change of the same pathways in the Btgh system allowed me to predict an elevated level of O-GlcNAc in the embryos from $Ogt^{T931A/WT}$ mothers. Based on this, I used the well established prior knowledge about the feedback mechanism between OGT and O-GlcNAc in O-GlcNAc homeostasis to propose a model explaining the maternal effect. This model involves the inheritance from the oocyte of a higher level of OGT proteins and eventually of O-GlcNAc payload (Fig.15g).
These data divulge the importance of the inheritance of a post-translational modification from the oocyte to the embryo. Moreover, in the case of O-GlcNAc, the identity of the affected molecular pathways suggests that such an inheritance has implications for embryonic growth.

By providing the first molecular characterization of a developing mouse embryo with a perturbed O-GlcNAc homeostasis, this study shows that the most salient effect is a delayed preimplantation and postimplantation development, associated with the modulation of the pathways of intracellular vesicle trafficking, mitochondrial respiration and translation, together with the activation of cell cycle genes.

How does this reconcile with the phenotype of *Ogt*-null fly embryos, where the most visible outcome was the disruption of Polycomb function? In my postimplantation experiment, the activity of Btgh is likely not ongoing anymore and the phenotype observed is the consequence of removing O-GlcNAc in the preimplantation (and maybe peri-implantation) embryo on postimplantation development. An effect of O-GlcNAc depletion on Polycomb - which is

relevant for lineage differentiation - might only be visible if such depletion is maintained until postimplantation stages. However, *Ezh2*-null embryos were reported to arrest their development at E5.5 (O'Carroll et al., 2001), hence O-GlcNAc seems to be dispensable at least for the function of mammalian PRC2 core's proteins.
Furthermore, my work does not pretend to describe all functions of O-GlcNAc in the mammalian early embryo, but the most phenotypically relevant one. O-GlcNAc is highly pleiotropic, within the cell as well as for the whole embryo. The molecular mechanism affected by O-GlcNAc perturbation which is more phenotypically relevant might vary depending on the organism and the specific phase of development.

The following questions are raised by the data I presented but could not be answered and await further investigations. Firstly, the significance of the asymmetric distribution of OGT in the two pronuclei. To tackle this question, I propose the creation of a better OGT-degron system or a system to induce the export of OGT from the pronucleus. This can be combined with the perturbation in the zygote of the O-GlcNAc modification, instead symmetrically present in the parental pronuclei. To this aim, the experimental design of Btgh injections could be modified in order to induce O-GlcNAc depletion already in early pronuclear stages. Given the overlapping pattern of 5hmC asymmetry and the interaction of TET with OGT, it would be informative to profile 5hmC after O-GlcNAc removal in the PN1 zygote.
Secondly, what are the molecular events connecting O-GlcNAc with the transcriptional regulation of translation-related pathways? This question is very difficult to answer due to the thousands of different proteins modified by O-GlcNAc. In my view, it is difficult to imagine a mechanism where one modification, which can impact protein structure or function in different ways depending on the target, can have a general coherent response on the cell biosynthetic potential as the one speculated from the data (and also often implicitly speculated in literature). I find it more conceivable that the most prominent cellular output of O-GlcNAc is given by the direct regulation of a few factors or processes that are key players in the control of cell metabolism. Cancer cells are a good example where the change in the activity of a few oncogenes (for example via a genetic mutation) is able to drive a cascade that completely shifts the cell's proliferative and metabolic status. The O-GlcNAcylation of a master regulator of cell growth such as c-Myc or a component of mTORC1 signaling is conceivable and it was reported for c-Myc (T.-Y. Chou et al., 1995; T. Y. Chou et al., 1995). PUF60, one of the genes which are constantly downregulated in O-GlcNAc-depleted embryos (with log2FC ~ 0.15), besides having a role in splicing, has been reported in complex with the transcriptional machinery and

to repress activator-induced (but not basal) transcription at *Myc* promoter (Liu et al., 2000). The latter finding is only based on one study, however it is interesting that mutations in human PUF60 are associated with a syndrome including intellectual disability, short stature and other body anomalies (Low et al., 2017), reminiscent of OGT-XLID syndrome. This could be a hint for this gene being one of the mediators of O-GlcNAc function. Another very intriguing possibility is the regulation by O-GlcNAc of ribosomal RNA (rRNA) biogenesis, at the step of their transcription or post-trancriptional processing. In fact, rRNA synthesis and the transcription of ribosomal proteins are highly coordinated processes (Albert et al., 2016). rRNA genes are transcribed by RNA PolI, which does not seem to bear O-GlcNAc moieties (Jackson and Tjian, 1988), but recently RNA PolII has been found at rRNA loci (Abraham et al., 2020). Nucleolar PolII generates antisense transcripts which are necessary to maintain nucleolar condensates (Abraham et al., 2020). The idea of O-GlcNAc and rRNA synthesis comes from the peculiar signal shown by one anti-O-GlcNAc antibody at foci surrounding nucleolar-like-bodies (NLBs) in zygotes and 2-cell embryos (Fig.S1b-c). Only at these stages, rDNA clusters localize at NLBs' periphery. The immunofluorescence staining of O-GlcNAc combined with rDNA FISH could help to define whether O-GlcNAc colocalizes with rDNA in the zygote and 2-cell embryos. Furthermore, experiments are ongoing in the lab to test the effect of O-GlcNAc removal on rRNA transcription and processing in mESCs.

Being the first study investigating the O-GlcNAc biology of early mammalian embryos, I did not presume to exhaustively answer this complex biological question. With the results obtained, I propose a mechanism of intergenerational regulation of embryonic growth by O-GlcNAc through transcriptional modulation of genes involved in translation. I hope that my results will now open new avenues of investigation to better understand the biology of this fascinating modification in the physiologically relevant system that is the live embryo.

References

Abe, K., Yamamoto, R., Franke, V., Cao, M., Suzuki, Y., Suzuki, M.G., Vlahovicek, K., Svoboda, P., Schultz, R.M., Aoki, F., 2015. The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3′ processing. Embo J 34, 1523–1537. https://doi.org/10.15252/embj.201490648

Abraham, K.J., Khosraviani, N., Chan, J.N.Y., Gorthi, A., Samman, A., Zhao, D.Y., Wang, M., Bokros, M., Vidya, E., Ostrowski, L.A., Oshidari, R., Pietrobon, V., Patel, P.S., Algouneh, A., Singhania, R., Liu, Y., Yerlici, V.T., Carvalho, D.D.D., Ohh, M., Dickson, B.C., Hakem, R., Greenblatt, J.F., Lee, S., Bishop, A.J.R., Mekhail, K., 2020. Nucleolar RNA polymerase II drives ribosome biogenesis. Nature 585, 298–302. https://doi.org/10.1038/s41586-020-2497-0

Adenot, P.G., Mercier, Y., Renard, J.-P., Thompson, E.M., 1997. Differential H4 acetylation of paternal and maternal chromatin precedes DNA replication and differential transcriptional activity in pronuclei of 1-cell mouse embryos. Development 124, 4615–4625. https://doi.org/10.1242/dev.124.22.4615

Afgan, E., Baker, D., Batut, B., Beek, M.V.D., Bouvier, D., Ech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research 46, W537–W544. https://doi.org/10.1093/nar/gky379

Ahmad, K., Henikoff, S., 2002. The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. Mol Cell 9, 1191–1200. https://doi.org/10.1016/s1097-2765(02)00542-7

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., Bazett-Jones, D.P., 2010. Global Chromatin Architecture Reflects Pluripotency and Lineage Commitment in the Early Mouse Embryo. Plos One 5, e10531. https://doi.org/10.1371/journal.pone.0010531

Aiken, C.E.M., Swoboda, P.P.L., Skepper, J.N., Johnson, M.H., 2004. The direct measurement of embryogenic volume and nucleo-cytoplasmic ratio during mouse pre-implantation development. Reproduction 128, 527–535. https://doi.org/10.1530/rep.1.00281

Albert, B., Knight, B., Merwin, J., Martin, V., Ottoz, D., Gloor, Y., Bruzzone, M.J., Rudner, A., Shore, D., 2016. A Molecular Titration System Coordinates Ribosomal Protein Gene Transcription with Ribosomal RNA Synthesis. Mol Cell 64, 720–733. https://doi.org/10.1016/j.molcel.2016.10.003

Alexander, K.A., Wang, X., Shibata, M., Clark, A.G., García-García, M.J., 2015. TRIM28 Controls Genomic Imprinting through Distinct Mechanisms during and after Early Genome-wide Reprogramming. Cell Reports 13, 1194–1205. https://doi.org/10.1016/j.celrep.2015.09.078

Allan, R.K., Ratajczak, T., 2011. Versatile TPR domains accommodate different modes of target protein recognition and function. Cell Stress Chaperones 16, 353–367. https://doi.org/10.1007/s12192-010-0248-0
Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W.S., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., Encheva, V., Kudo, N.R., Koseki, H., Sasaki, H., Hajkova, P., 2016. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nature Cell Biology 18, 225–233. https://doi.org/10.1038/ncb3296
Andergassen, D., Smith, Z.D., Kretzmer, H., Rinn, J.L., Meissner, A., 2021. Diverse epigenetic mechanisms maintain parental imprints within the embryonic and extraembryonic lineages. Developmental Cell 56, 2995-3005.e4. https://doi.org/10.1016/j.devcel.2021.10.010
Aoki, F., Worrad, D.M., Schultz, R.M., 1997. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. Developmental Biology 181, 296–307. https://doi.org/10.1006/dbio.1996.8466
Aramburu, I.V., Lemke, E.A., 2017. Floppy but not sloppy: Interaction mechanism of FG-nucleoporins and nuclear transport receptors. Semin Cell Dev Biol 68, 34–41. https://doi.org/10.1016/j.semcdb.2017.06.026
Asami, M., Lam, B.Y.H., Ma, M.K., Rainbow, K., Braun, S., VerMilyea, M.D., Yeo, G.S.H., Perry, A.C.F., 2022. Human embryonic genome activation initiates at the one-cell stage. Cell Stem Cell 29, 209-216.e4. https://doi.org/10.1016/j.stem.2021.11.012
Balana, A.T., Levine, P.M., Craven, T.W., Mukherjee, S., Pedowitz, N.J., Moon, S.P., Takahashi, T.T., Becker, C.F.W., Baker, D., Pratt, M.R., 2021. O-GlcNAc modification of small heat shock proteins enhances their anti-amyloid chaperone activity. Nature Chemistry 1–10. https://doi.org/10.1038/s41557-021-00648-8
Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., Bourc'his, D., 2016. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354, 909–912. https://doi.org/10.1126/science.aah5143
Bartolomei, M.S., Zemel, S., Tilghman, S.M., 1991. Parental imprinting of the mouse H19 gene. Nature 351, 153–155. https://doi.org/10.1038/351153a0
Bell, A.C., Felsenfeld, G., 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405, 482–485. https://doi.org/10.1038/35013100
Benhalevy, D., Gupta, S.K., Danan, C.H., Ghosal, S., Sun, H.-W., Kazemier, H.G., Paeschke, K., Hafner, M., Juranek, S.A., 2017. The Human CCHC-type Zinc Finger Nucleic Acid-Binding Protein Binds G-Rich Elements in Target mRNA Coding Sequences and Promotes Translation. Cell Reports 18, 2979–2990. https://doi.org/10.1016/j.celrep.2017.02.080
Berletch, J.B., Yang, F., Xu, J., Carrel, L., Disteche, C.M., 2011. Genes that escape from X inactivation. Hum Genet 130, 237–245. https://doi.org/10.1007/s00439-011-1011-z
Bogue, M.A., Ball, R.L., Philip, V.M., Walton, D.O., Dunn, M.H., Kolishovski, G., Lamoureux, A., Gerring, M., Liang, H., Emerson, J., Stearns, T., He, H., Mukherjee, G.,

Bluis, J., Desai, S., Sundberg, B., Kadakkuzha, B., Kunde-Ramamoorthy, G., Chesler, E.J., 2022. Mouse Phenome Database: towards a more FAIR-compliant and TRUST-worthy data repository and tool suite for phenotypes and genotypes. Nucleic Acids Res 51, D1067–D1074. https://doi.org/10.1093/nar/gkac1007
Borsos, M., Torres-Padilla, M.-E., 2016. Building up the nucleus: nuclear organization in the establishment of totipotency and pluripotency during mammalian development. Gene Dev 30, 611–621. https://doi.org/10.1101/gad.273805.115
Bošković, A., Bender, A., Gall, L., Ziegler-Birling, C., Beaujean, N., Torres-Padilla, M.E., 2012. Analysis of active chromatin modifications in early mammalian embryos reveals uncoupling of H2A.Z acetylation and H3K36 trimethylation from embryonic genome activation. http://dx.doi.org/10.4161/epi.20584 7, 747–757. https://doi.org/10.4161/epi.20584
Bošković, A., Eid, A., Pontabry, J., Ishiuchi, T., Spiegelhalter, C., Ram, E.V.S.R., Meshorer, E., Torres-Padilla, M.-E., 2014. Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. Gene Dev 28, 1042–1047. https://doi.org/10.1101/gad.238881.114
Boulard, M., Rucli, S., Edwards, J.R., Bestor, T.H., 2020. Methylation-directed glycosylation of chromatin factors represses retrotransposon promoters. Proceedings of the National Academy of Sciences of the United States of America 117, 14292–14298. https://doi.org/10.1073/pnas.1912074117
Boutz, P.L., Bhutkar, A., Sharp, P.A., 2015. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. Gene Dev 29, 63–80. https://doi.org/10.1101/gad.247361.114
Bowling, S., Gregorio, A.D., Sancho, M., Pozzi, S., Aarts, M., Signore, M., Schneider, M.D., Martinez-Barbera, J.P., Gil, J., Rodríguez, T.A., 2018. P53 and mTOR signalling determine fitness selection through cell competition during early mouse embryonic development. Nat Commun 9, 1763. https://doi.org/10.1038/s41467-018-04167-y
Bulut-Karslioglu, A., Macrae, T.A., Oses-Prieto, J.A., Covarrubias, S., Percharde, M., Ku, G., Diaz, A., McManus, M.T., Burlingame, A.L., Ramalho-Santos, M., 2018. The Transcriptionally Permissive Chromatin State of Embryonic Stem Cells Is Acutely Tuned to Translational Output. Cell Stem Cell 22, 369-383.e8. https://doi.org/10.1016/j.stem.2018.02.004
Burt, R.A., Alghusen, I.M., Ephrame, S.J., Villar, M.T., Artigues, A., Slawson, C., 2022. Mapping the O-GlcNAc Modified Proteome: Applications for Health and Disease. Frontiers Mol Biosci 9, 920727. https://doi.org/10.3389/fmolb.2022.920727
Butz, S., Schmolka, N., Karemaker, I.D., Villaseñor, R., Schwarz, I., Domcke, S., Uijttewaal, E.C.H., Jude, J., Lienert, F., Krebs, A.R., Wagenaar, N.P. de, Bao, X., Zuber, J., Elling, U., Schübeler, D., Baubec, T., 2022. DNA sequence and chromatin modifiers cooperate to confer epigenetic bistability at imprinting control regions. Nat Genet 1–9. https://doi.org/10.1038/s41588-022-01210-z
Camlin, N.J., Evans, J.P., 2019. Auxin-inducible protein degradation as a novel approach for

protein depletion and reverse genetic discoveries in mammalian oocytes. Biology of Reproduction 101, 704–718. https://doi.org/10.1093/biolre/ioz113
Camus, A., Perea-Gomez, A., Moreau, A., Collignon, J., 2006. Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. Dev Biol 295, 743–755. https://doi.org/10.1016/j.ydbio.2006.03.047
Capotosti, undefined F., Guernier, S., Lammers, F., Waridel, P., Cai, Y., Jin, J., Conaway, J.W., Conaway, R.C., Herr, W., 2011. O-GlcNAc Transferase Catalyzes Site-Specific Proteolysis of HCF-1. Cell 144, 376–388. https://doi.org/10.1016/j.cell.2010.12.030
Carey, B.W., Finley, L.W.S., Cross, J.R., Allis, C.D., Thompson, C.B., 2015. Intracellular α-ketoglutarate maintains the pluripotency of embryonic stem cells. Nature 518, 413–416. https://doi.org/10.1038/nature13981
Castel, S.E., Levy-Moonshine, A., Mohammadi, P., Banks, E., Lappalainen, T., 2015. Tools and best practices for data processing in allelic expression analysis. Genome Biology 16. https://doi.org/10.1186/s13059-015-0762-6
Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., Smith, A., 2003. Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. Cell 113, 643–655. https://doi.org/10.1016/s0092-8674(03)00392-1
Chan, C.C., DOSTIE, J., DIEM, M.D., FENG, W., MANN, M., RAPPSILBER, J., DREYFUSS, G., 2004. eIF4A3 is a novel component of the exon junction complex. Rna 10, 200–209. https://doi.org/10.1261/rna.5230104
Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., Ma'ayan, A., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. Bmc Bioinformatics 14, 128. https://doi.org/10.1186/1471-2105-14-128
Chen, J., 2016. The Cell-Cycle Arrest and Apoptotic Functions of p53 in Tumor Initiation and Progression. Csh Perspect Med 6, a026104. https://doi.org/10.1101/cshperspect.a026104
Chen, Q., Chen, Y., Bian, C., Fujiki, R., Yu, X., 2013. TET2 promotes histone O-GlcNAcylation during gene transcription. Nature 493, 561–564. https://doi.org/10.1038/nature11742
Cheng, S., Mittnenzweig, M., Mayshar, Y., Lifshitz, A., Dunjić, M., Rais, Y., Ben-Yair, R., Gehrs, S., Chomsky, E., Mukamel, Z., Rubinstein, H., Schlereth, K., Reines, N., Orenbuch, A.H., Tanay, A., Stelzer, Y., 2022. The intrinsic and extrinsic effects of TET proteins during gastrulation. Cell 185, 3169-3185.e20. https://doi.org/10.1016/j.cell.2022.06.049
Chou, T.Y., Dang, C.V., Hart, G.W., 1995. Glycosylation of the c-Myc transactivation domain. Proc National Acad Sci 92, 4417–4421. https://doi.org/10.1073/pnas.92.10.4417
Chou, T.-Y., Hart, G.W., Dang, C.V., 1995. c-Myc Is Glycosylated at Threonine 58, a Known Phosphorylation Site and a Mutational Hot Spot in Lymphomas *. J Biol Chem 270, 18961–18965. https://doi.org/10.1074/jbc.270.32.18961
Chu, C.-S., Lo, P.-W., Yeh, Y.-H., Hsu, P.-H., Peng, S.-H., Teng, Y.-C., Kang, M.-L., Wong, C.-H., Juan, L.-J., 2014. O-GlcNAcylation regulates EZH2 protein stability and function.

Proceedings of the National Academy of Sciences of the United States of America 111, 1355–60. https://doi.org/10.1073/pnas.1323226111
Chuong, E.B., Rumi, M.A.K., Soares, M.J., Baker, J.C., 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nat Genet 45, 325–329. https://doi.org/10.1038/ng.2553
Ciemerych, M.A., Sicinski, P., 2005. Cell cycle in mouse development. Oncogene 24, 2877–2898. https://doi.org/10.1038/sj.onc.1208608
Clavería, C., Giovinazzo, G., Sierra, R., Torres, M., 2013. Myc-driven endogenous cell competition in the early mammalian embryo. Nature 500, 39–44. https://doi.org/10.1038/nature12389
Clift, D., Schuh, M., 2013. Restarting life: fertilization and the transition from meiosis to mitosis. Nat Rev Mol Cell Bio 14, 549–562. https://doi.org/10.1038/nrm3643
Conine, C.C., Sun, F., Song, L., Rivera-Pérez, J.A., Rando, O.J., 2018. Small RNAs Gained during Epididymal Transit of Sperm Are Essential for Embryonic Development in Mice. Developmental Cell 46, 470-480.e3. https://doi.org/10.1016/j.devcel.2018.06.024
Crosetto, N., Bienko, M., 2020. Radial Organization in the Mammalian Nucleus. Frontiers Genetics 11, 33. https://doi.org/10.3389/fgene.2020.00033
Czajewski, I., Aalten, D.M.F. van, 2023. The role of O-GlcNAcylation in development. Development 150. https://doi.org/10.1242/dev.201370
Dahlet, T., Lleida, A.A., Adhami, H.A., Dumas, M., Bender, A., Ngondo, R.P., Tanguy, M., Vallet, J., Auclair, G., Bardet, A.F., Weber, M., 2020. Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. Nature Communications 11, 3153. https://doi.org/10.1038/s41467-020-16919-w
Dang, C.V., 2013. MYC, Metabolism, Cell Growth, and Tumorigenesis. Csh Perspect Med 3, a014217. https://doi.org/10.1101/cshperspect.a014217
Daou, S., Mashtalir, N., Hammond-Martel, I., Pak, H., Yu, H., Sui, G., Vogel, J.L., Kristie, T.M., Affar, E.B., 2011. Crosstalk between O-GlcNAcylation and proteolytic cleavage regulates the host cell factor-1 maturation pathway. Proc National Acad Sci 108, 2747–2752. https://doi.org/10.1073/pnas.1013822108
Dennis, R.J., Taylor, E.J., MacAuley, M.S., Stubbs, K.A., Turkenburg, J.P., Hart, S.J., Black, G.N., Vocadlo, D.J., Davies, G.J., 2006. Structure and mechanism of a bacterial β-glucosaminidase having O-GlcNAcase activity. Nature Structural and Molecular Biology 13, 365–371. https://doi.org/10.1038/nsmb1079
Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Méndez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., Shih, A.H., Levine, R.L., Bernard, O., Mercher, T., Solary, E., Urh, M., Daniels, D.L., Fuks, F., 2013. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. The EMBO Journal 28, 3185–3195. https://doi.org/10.1038/emboj.2009.258
Derivery, E., Sousa, C., Gautier, J.J., Lombard, B., Loew, D., Gautreau, A., 2009. The Arp2/3

Activator WASH Controls the Fission of Endosomes through a Large Multiprotein Complex. Dev Cell 17, 712–723. https://doi.org/10.1016/j.devcel.2009.09.010
Diaz-Cuadros, M., Miettinen, T.P., Skinner, O.S., Sheedy, D., Díaz-García, C.M., Gapon, S., Hubaud, A., Yellen, G., Manalis, S.R., Oldham, W.M., Pourquié, O., 2023. Metabolic regulation of species-specific developmental rates. Nature 613, 550–557. https://doi.org/10.1038/s41586-022-05574-4
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England) 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635
Downs, K.M., Davies, T., 1993. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. Development 118, 1255–1266. https://doi.org/10.1242/dev.118.4.1255
Dumollard, R., Carroll, J., Duchen, M.R., Campbell, K., Swann, K., 2009. Mitochondrial function and redox state in mammalian embryos. Semin Cell Dev Biol 20, 346–353. https://doi.org/10.1016/j.semcdb.2008.12.013
Eckersley-Maslin, M.A., Alda-Catalinas, C., Reik, W., 2018. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. Nature Reviews Molecular Cell Biology 2018 19:7 19, 436–450. https://doi.org/10.1038/s41580-018-0008-z
Edwards, J.R., Yarychkivska, O., Boulard, M., Bestor, T.H., 2017. DNA methylation and DNA methyltransferases. Epigenetics & Chromatin 10, 23. https://doi.org/10.1186/s13072-017-0130-8
Eid, A., Rodriguez-Terrones, D., Burton, A., Torres-Padilla, M.E., 2016. SUV4-20 activity in the preimplantation mouse embryo controls timely replication. Genes & Development 30, 2513–2526. https://doi.org/10.1101/gad.288969.116
Erickson, J.R., Pereira, L., Wang, L., Han, G., Ferguson, A., Dao, K., Copeland, R.J., Despa, F., Hart, G.W., Ripplinger, C.M., Bers, D.M., 2013. Diabetic hyperglycaemia activates CaMKII and arrhythmias by O-linked glycosylation. Nature 2013 502:7471 502, 372–376. https://doi.org/10.1038/nature12537
Evans, M.J., Kaufman, M.H., 1981. Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154–156. https://doi.org/10.1038/292154a0
Evertts, A.G., Zee, B.M., DiMaggio, P.A., Gonzales-Cope, M., Coller, H.A., Garcia, B.A., 2013. Quantitative Dynamics of the Link between Cellular Metabolism and Histone Acetylation. J Biol Chem 288, 12142–12151. https://doi.org/10.1074/jbc.m112.428318
Faust, C., Lawson, K.A., Schork, N.J., Thiel, B., Magnuson, T., 1998. The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. Development 125, 4495–4506. https://doi.org/10.1242/dev.125.22.4495
Ferguson-Smith, A.C., 2011. Genomic imprinting: The emergence of an epigenetic paradigm. Nature Reviews Genetics 12, 565–575. https://doi.org/10.1038/nrg3032
Ferrer, C.M., Lynch, T.P., Sodi, V.L., Falcone, J.N., Schwab, L.P., Peacock, D.L., Vocadlo,

D.J., Seagroves, T.N., Reginato, M.J., 2014. O-GlcNAcylation Regulates Cancer Metabolism and Survival Stress Signaling via Regulation of the HIF-1 Pathway. Molecular Cell 54, 820–831. https://doi.org/10.1016/j.molcel.2014.04.026
Fong, J.J., Nguyen, B.L., Bridger, R., Medrano, E.E., Wells, L., Pan, S., Sifers, R.N., 2012. β-N-Acetylglucosamine (O-GlcNAc) Is a Novel Regulator of Mitosis-specific Phosphorylations on Histone H3*. J Biol Chem 287, 12195–12203. https://doi.org/10.1074/jbc.m111.315804
Franke, V., Ganesh, S., Karlic, R., Malik, R., Pasulka, J., Horvat, F., Kuzman, M., Fulka, H., Cernohorska, M., Urbanova, J., Svobodova, E., Ma, J., Suzuki, Y., Aoki, F., Schultz, R.M., Vlahovicek, K., Svoboda, P., 2017. Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. Genome Res 27, 1384–1394. https://doi.org/10.1101/gr.216150.116
Fujiki, R., Hashiba, W., Sekine, H., Yokoyama, A., Chikanishi, T., Ito, S., Imai, Y., Kim, J., He, H.H., Igarashi, K., Kanno, J., Ohtake, F., Kitagawa, H., Roeder, R.G., Brown, M., Kato, S., 2011. GlcNAcylation of histone H2B facilitates its monoubiquitination. Nature 480, 557–560. https://doi.org/10.1038/nature10656
Gagnon, J., Daou, S., Zamorano, N., Iannantuono, N.V., Hammond-Martel, I., Mashtalir, N., Bonneil, E., Wurtele, H., Thibault, P., Affar, E.B., 2015. Undetectable histone O-GlcNAcylation in mammalian cells. http://dx.doi.org/10.1080/15592294.2015.1060387 10, 677–691. https://doi.org/10.1080/15592294.2015.1060387
Gambetta, M.C., Müller, J., 2015. A critical perspective of the diverse roles of O-GlcNAc transferase in chromatin. Chromosoma 124, 429–442. https://doi.org/10.1007/s00412-015-0513-1
Gambetta, M.C., Müller, J., 2014a. O-GlcNAcylation Prevents Aggregation of the Polycomb Group Repressor Polyhomeotic. Developmental Cell 31, 629–639. https://doi.org/10.1016/j.devcel.2014.10.020
Gambetta, M.C., Müller, J., 2014b. O-GlcNAcylation Prevents Aggregation of the Polycomb Group Repressor Polyhomeotic. Developmental Cell 31, 629–639. https://doi.org/10.1016/j.devcel.2014.10.020
Gambetta, M.C., Oktaba, K., Müller, J., 2009a. Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. Science 325, 93 96. https://doi.org/10.1126/science.1169727
Gambetta, M.C., Oktaba, K., Müller, J., 2009b. Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. Science 325, 93 96. https://doi.org/10.1126/science.1169727
Gao, Y., Liu, X., Tang, B., Li, C., Kou, Z., Li, L., Liu, W., Wu, Y., Kou, X., Li, J., Zhao, Y., Yin, J., Wang, H., Chen, S., Liao, L., Gao, S., 2017. Protein Expression Landscape of Mouse Embryos during Pre-implantation Development. Cell Reports 21, 3957–3969. https://doi.org/10.1016/j.celrep.2017.11.111
Gao, Y., Wells, L., Comer, F.I., Parker, G.J., Hart, G.W., 2001. Dynamic O-glycosylation of nuclear and cytosolic proteins: Cloning and characterization of a neutral, cytosolic β-N-acetylglucosaminidase from human brain. Journal of Biological Chemistry 276, 9838–9845.

https://doi.org/10.1074/jbc.m010420200
Garcia-Fernàndez, J., 2005. The genesis and evolution of homeobox gene clusters. Nat Rev Genet 6, 881–892. https://doi.org/10.1038/nrg1723
Gentilella, A., Kozma, S.C., Thomas, G., 2015. A liaison between mTOR signaling, ribosome biogenesis and cancer. Biochimica Et Biophysica Acta Bba - Gene Regul Mech 1849, 812–820. https://doi.org/10.1016/j.bbagrm.2015.02.005
Grafe, I., Alexander, S., Peterson, J.R., Snider, T.N., Levi, B., Lee, B., Mishina, Y., 2017. TGF-β Family Signaling in Mesenchymal Differentiation. Csh Perspect Biol 10, a022202. https://doi.org/10.1101/cshperspect.a022202
Gray, W.M., Pozo, J.C. del, Walker, L., Hobbie, L., Risseeuw, E., Banks, T., Crosby, W.L., Yang, M., Ma, H., Estelle, M., 1999. Identification of an SCF ubiquitin–ligase complex required for auxin response in Arabidopsis thaliana. Gene Dev 13, 1678–1691. https://doi.org/10.1101/gad.13.13.1678
Greally, J.M., 2018. A user's guide to the ambiguous word "epigenetics." Nat Rev Mol Cell Bio 19, 207–208. https://doi.org/10.1038/nrm.2017.135
Greenberg, M.V.C., 2021. Get Out and Stay Out: New Insights Into DNA Methylation Reprogramming in Mammals. Frontiers in Cell and Developmental Biology 8, 629068. https://doi.org/10.3389/fcell.2020.629068
Greenberg, M.V.C., Bourc'his, D., 2019. The diverse roles of DNA methylation in mammalian development and disease. Nature Reviews Molecular Cell Biology 1–18. https://doi.org/10.1038/s41580-019-0159-6
Groff, A.F., Resetkova, N., DiDomenico, F., Sakkas, D., Penzias, A., Rinn, J.L., Eggan, K., 2019. RNA-seq as a tool for evaluating human embryo competence. Genome Research 29, 1705. https://doi.org/10.1101/gr.252981.119
Grosswendt, S., Kretzmer, H., Smith, Z.D., Kumar, A.S., Hetzel, S., Wittler, L., Klages, S., Timmermann, B., Mukherji, S., Meissner, A., 2020. Epigenetic regulator function through mouse gastrulation. Nature 584, 102–108. https://doi.org/10.1038/s41586-020-2552-x
Gu, B., Posfai, E., Rossant, J., 2018. Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. Nature Biotechnology 2018 36:7 36, 632–637. https://doi.org/10.1038/nbt.4166
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., Iqbal, K., Shi, Y.G., Deng, Z., Szabó, P.E., Pfeifer, G.P., Li, J., Xu, G.L., 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 2011 477:7366 477, 606–610. https://doi.org/10.1038/nature10443
Guan, M., Bogani, D., Marschall, S., Raspa, M., Takeo, T., Nakagata, N., Fray, M., 2014. In vitro fertilization in mice using the MBCD-GSH protocol. Current protocols in mouse biology 4, 67–83. https://doi.org/10.1002/9780470942390.mo140059
Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., Surani, M.A., 2013. Synergistic Mechanisms of DNA Demethylation during Transition to Ground-State

Pluripotency. Stem Cell Rep 1, 518–531. https://doi.org/10.1016/j.stemcr.2013.11.010
Hahne, H., Gholami, A.M., Kuster, B., 2012. Discovery of O-GlcNAc-modified Proteins in Published Large-scale Proteome Data*. Mol Cell Proteomics 11, 843–850. https://doi.org/10.1074/mcp.m112.019463
Haltiwanger, R.S., Holt, G.D., Hart, G.W., 1990. Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetylglucosaminyltransferase. J Biol Chem 265, 2563–2568. https://doi.org/10.1016/s0021-9258(19)39838-2
Hanover, J.A., Yu, S., Lubas, W.B., Shin, S.H., Ragano-Caracciola, M., Kochran, J., 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. https://doi.org/10.1016/s0003-9861(02)00578-7
Hao, Y., Fan, X., Shi, Y., Zhang, C., Sun, D. en, Qin, K., Qin, W., Zhou, W., Chen, X., 2019. Next-generation unnatural monosaccharides reveal that ESRRB O-GlcNAcylation regulates pluripotency of mouse embryonic stem cells. Nature Communications 10, 1–13. https://doi.org/10.1038/s41467-019-11942-y
Harlen, K.M., Churchman, L.S., 2017. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat Rev Mol Cell Bio 18, 263–273. https://doi.org/10.1038/nrm.2017.10
Hartweck, L.M., Scott, C.L., Olszewski, N.E., 2002. Two O -Linked N -Acetylglucosamine Transferase Genes of Arabidopsis thaliana L. Heynh. Have Overlapping Functions Necessary for Gamete and Seed Development. Genetics 161, 1279–1291. https://doi.org/10.1093/genetics/161.3.1279
Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.-W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., Nix, D.A., Peterson, C.M., Tapscott, S.J., Carrell, D.T., Cairns, B.R., 2017. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat Genet 49, 925–934. https://doi.org/10.1038/ng.3844
Hennig, B.P., Velten, L., Racke, I., Tu, C.S., Thoms, M., Rybin, V., Besir, H., Remans, K., Steinmetz, L.M., 2018. Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. G3: Genes, Genomes, Genetics 8, 79–89. https://doi.org/10.1534/g3.117.300257/-/dc1
Henry, R.W., Mittal, V., Ma, B., Kobayashi, R., Hernandez, N., 1998. SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III. Gene Dev 12, 2664–2672. https://doi.org/10.1101/gad.12.17.2664
Hirano, T., Ishihara, K., Hibi, M., 2000. Roles of STAT3 in mediating the cell growth, differentiation and survival signals relayed through the IL-6 family of cytokine receptors. Oncogene 19, 2548–2556. https://doi.org/10.1038/sj.onc.1203551
Holt, G.D., Hart, G.W., 1986. The subcellular distribution of terminal N-acetylglucosamine

moieties. Localization of a novel protein-saccharide linkage, O-linked GlcNAc., Journal of Biological Chemistry.
Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., Ren, B., 2013. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat Genet 45, 1198–1206. https://doi.org/10.1038/ng.2746
Howerton, C.L., Bale, T.L., 2014. Targeted placental deletion of OGT recapitulates the prenatal stress phenotype including hypothalamic mitochondrial dysfunction. Proc National Acad Sci 111, 9639 9644. https://doi.org/10.1073/pnas.1401203111
Howerton, C.L., Morgan, C.P., Fischer, D.B., Bale, T.L., 2013. O-GlcNAc transferase (OGT) as a placental biomarker of maternal stress and reprogramming of CNS gene transcription in development. Proc National Acad Sci 110, 5169 5174.
https://doi.org/10.1073/pnas.1300065110
Iaco, A.D., Planet, E., Coluccio, A., Verp, S., Duc, J., Trono, D., 2017. DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat Genet 49, 941–945. https://doi.org/10.1038/ng.3858
Iaco, A.D., Verp, S., Offner, S., Grun, D., Trono, D., 2019. DUX is a non-essential synchronizer of zygotic genome activation. Development 147, dev177725.
https://doi.org/10.1242/dev.177725
Ingham, P.W., 1984. A gene that regulates the bithorax complex differentially in larval and adult cells of Drosophila. Cell 37, 815–823. https://doi.org/10.1016/0092-8674(84)90416-1
Inoue, A., Zhang, Y., 2011. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science 334, 194. https://doi.org/10.1126/science.1212483
Iqbal, K., Jin, S.-G., Pfeifer, G.P., Szabó, P.E., 2011. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc National Acad Sci 108, 3642–3647. https://doi.org/10.1073/pnas.1014033108
Ishiuchi, T., Abe, S., Inoue, K., Yeung, W.K.A., Miki, Y., Ogura, A., Sasaki, H., 2021. Reprogramming of the histone H3.3 landscape in the early mouse embryo. Nature Structural and Molecular Biology 28, 38–49. https://doi.org/10.1038/s41594-020-00521-1
Ito, R., Katsura, S., Shimada, H., Tsuchiya, H., Hada, M., Okumura, T., Sugawara, A., Yokoyama, A., 2013. TET3-OGT interaction increases the stability and the presence of OGT in chromatin. Genes to Cells 19, 52–65. https://doi.org/10.1111/gtc.12107
Jabs, E.W., Müller, U., Li, X., Ma, L., Luo, W., Haworth, I.S., Klisak, I., Sparkes, R., Warman, M.L., Mulliken, J.B., Snead, M.L., Maxson, R., 1993. A mutation in the homeodomain of the human MSX2 gene in a family affected with autosomal dominant craniosynostosis. Cell 75, 443–450. https://doi.org/10.1016/0092-8674(93)90379-5
Jackson, S.P., Tjian, R., 1988. O-glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation. Cell 55, 125–133.
https://doi.org/10.1016/0092-8674(88)90015-3
Jain, K., Verma, P.J., Liu, and J., 2014. Isolation and Handling of Mouse Embryonic

Fibroblasts, in: York, S.S.M.N. (Ed.), .
Janetzko, J., Walker, S., 2014. The making of a sweet modification: Structure and function of O-GlcNAc transferase. Journal of Biological Chemistry 289, 34424–34432. https://doi.org/10.1074/jbc.r114.604405
Jang, H., Kim, T.W., Yoon, S., Choi, S.Y., Kang, T.W., Kim, S.Y., Kwon, Y.W., Cho, E.J., Youn, H.D., 2012. O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network. Cell Stem Cell 11, 62–74. https://doi.org/10.1016/j.stem.2012.03.001
Jínek, M., Rehwinkel, J., Lazarus, B.D., Izaurralde, E., Hanover, J.A., Conti, E., 2004. The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin α. Nature Structural and Molecular Biology 11, 1001–1007. https://doi.org/10.1038/nsmb833
Joiner, C.M., Li, H., Jiang, J., Walker, S., 2019. Structural characterization of the O-GlcNAc cycling enzymes: insights into substrate recognition and catalytic mechanisms. Current Opinion in Structural Biology 56, 97–106. https://doi.org/10.1016/j.sbi.2018.12.003
Jukam, D., Shariati, S.A.M., Skotheim, J.M., 2017. Zygotic Genome Activation in Vertebrates. Developmental Cell 42, 316–332. https://doi.org/10.1016/j.devcel.2017.07.026
Julien, E., Herr, W., 2003. Proteolytic processing is necessary to separate and ensure proper cell growth and cytokinesis functions of HCF-1. EMBO Journal 22, 2360–2369. https://doi.org/10.1093/emboj/cdg242
Jürgens, G., 1985. A group of genes controlling the spatial expression of the bithorax complex in Drosophila. Nature 316, 153–155. https://doi.org/10.1038/316153a0
Kanellis, D.C., Espinoza, J.A., Zisi, A., Sakkas, E., Bartkova, J., Katsori, A.-M., Boström, J., Dyrskjøt, L., Broholm, H., Altun, M., Elsässer, S.J., Lindström, M.S., Bartek, J., 2021. The exon-junction complex helicase eIF4A3 controls cell fate via coordinated regulation of ribosome biogenesis and translational output. Sci Adv 7, eabf7561. https://doi.org/10.1126/sciadv.abf7561
Keembiyehetty, C., Love, D.C., Harwood, K.R., Gavrilova, O., Comly, M.E., Hanover, J.A., 2015. Conditional knock-out reveals a requirement for O-Linked N-Acetylglucosaminase (O-GlcNAcase) in metabolic homeostasis. Journal of Biological Chemistry 290, 7097–7113. https://doi.org/10.1074/jbc.m114.617779
Keembiyehetty, C.N., Krzeslak, A., Love, D.C., Hanover, J.A., 2011. A lipid-droplet-targeted O-GlcNAcase isoform is a key regulator of the proteasome. J Cell Sci 124, 2851–2860. https://doi.org/10.1242/jcs.083287
Kelly, W.G., Dahmus, M.E., 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., Hart, G.W., 1989. Glycosylation of chromosomal proteins: Localization of O-linked N-acetylglucosamine in Drosophila chromatin. Cell 57, 243–251.

https://doi.org/10.1016/0092-8674(89)90962-8
Khidekel, † Nelly, Arndt, † Sabine, Lamarre-Vincent, † Nathan, Lippert, † Alexander, Poulin-Kerstien, † Katherine G., Ramakrishnan, ‡, § Boopathy, Qasba, ‡ and Pradman K., Hsieh-Wilson*, † Linda C., 2003. A Chemoenzymatic Approach toward the Rapid and Sensitive Detection of O-GlcNAc Posttranslational Modifications. https://doi.org/10.1021/ja038545r
Kim, G., Cao, L., Reece, E.A., Zhao, Z., 2017. Impact of protein O-GlcNAcylation on neural tube malformation in diabetic embryopathy. Scientific Reports 7, 1–8. https://doi.org/10.1038/s41598-017-11655-6
Kingsley, D.M., Kozarsky, K.F., Hobble, L., Krieger, M., 1986. Reversible defects in O-linked glycosylation and LDL receptor expression in a UDP-GalUDP-GalNAc 4-epimerase deficient mutant. Cell 44, 749–759. https://doi.org/10.1016/0092-8674(86)90841-x
Kojima, Y., Tam, O.H., Tam, P.P.L., 2014. Timing of developmental events in the early mouse embryo. Semin Cell Dev Biol 34, 65–75. https://doi.org/10.1016/j.semcdb.2014.06.010
Kreppel, L.K., Blomberg, M.A., Hart, G.W., 1997. Dynamic glycosylation of nuclear and cytosolic proteins: Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. Journal of Biological Chemistry 272, 9308–9315. https://doi.org/10.1074/jbc.272.14.9308
Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., Ma'ayan, A., 2010. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics 26, 2438–2444. https://doi.org/10.1093/bioinformatics/btq466
Lagerlöf, O., Slocomb, J.E., Hong, I., Aponte, Y., Blackshaw, S., Hart, G.W., Huganir, R.L., 2016. The nutrient sensor OGT in PVN neurons regulates feeding. Science (New York, N.Y.) 351, 1293–6. https://doi.org/10.1126/science.aad5494
Lairson, L.L., Henrissat, B., Davies, G.J., Withers, S.G., 2008. Glycosyltransferases: Structures, Functions, and Mechanisms. Annu Rev Biochem 77, 521–555. https://doi.org/10.1146/annurev.biochem.76.061005.092322
Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., Reik, W., 2003. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis 35, 88–93. https://doi.org/10.1002/gene.10168
Latorre-Muro, P., O'Malley, K.E., Bennett, C.F., Perry, E.A., Balsa, E., Tavares, C.D.J., Jedrychowski, M., Gygi, S.P., Puigserver, P., 2021. A cold-stress-inducible PERK/OGT axis controls TOM70-assisted mitochondrial protein import and cristae formation. Cell Metabolism. https://doi.org/10.1016/j.cmet.2021.01.013
Lazarus, B.D., Love, D.C., Hanover, J.A., 2006. Recombinant O-GlcNAc transferase isoforms: Identification of O-GlcNAcase, yes tyrosine kinase, and tau as isoform-specific substrates. Glycobiology 16, 415–421. https://doi.org/10.1093/glycob/cwj078
Lazarus, M.B., Jiang, J., Kapuria, V., Bhuiyan, T., Janetzko, J., Zandberg, W.F., Vocadlo, D.J., Herr, W., Walker, S., 2013. HCF-1 is cleaved in the active site of O-GlcNAc transferase. Science (New York, N.Y.) 342, 1235–9. https://doi.org/10.1126/science.1243990

Lazarus, M.B., Nam, Y., Jiang, J., Sliz, P., Walker, S., 2011. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. Nature 469, 564–567. https://doi.org/10.1038/nature09638
Lee, D.R., Lee, J.E., Yoon, H.S., Roh, S.I., Kim, M.K., 2001. Compaction in preimplantation mouse embryos is regulated by a cytoplasmic regulatory factor that alters between 1- and 2-cell stages in a concentration-dependent manner. J. Exp. Zool. 290, 61–71. https://doi.org/10.1002/jez.1036
Lee, M.T., Bonneau, A.R., Takacs, C.M., Bazzini, A.A., DiVito, K.R., Fleming, E.S., Giraldez, A.J., 2013. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature 503, 360–364. https://doi.org/10.1038/nature12632
Leighton, P.A., Ingram, R.S., Eggenschwiler, J., Efstratiadis, A., Tilghman, S.M., 1995. Disruption of imprinting caused by deletion of the H19 gene region in mice. Nature 375, 34–39. https://doi.org/10.1038/375034a0
Lewis, B.A., Burlingame, A.L., 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. https://doi.org/10.1074/jbc.m115.684365
Lewis, B.A., Levens, D., 2021. Establishment of a Human RNA Pol II Pausing System and the Identification of O-GlcNAc Cycling Regulating Pol II Pausing and Elongation. Biorxiv 2020.01.23.917237. https://doi.org/10.1101/2020.01.23.917237
Lewis, E.B., 1978. A gene complex controlling segmentation in Drosophila. Nature 276, 565–570. https://doi.org/10.1038/276565a0
Li, B., Kohler, J.J., 2014. Glycosylation of the Nuclear Pore. Traffic 15, 347–361. https://doi.org/10.1111/tra.12150
Li, B., Li, H., Lu, L., Jiang, J., 2017. Structures of human O-GlcNAcase and its complexes reveal a new substrate recognition mode. Nat Struct Mol Biol 24, 362–369. https://doi.org/10.1038/nsmb.3390
Li, E., Beard, C., Jaenisch, R., 1993. Role for DNA methylation in genomic imprinting. Nature 366, 362–365. https://doi.org/10.1038/366362a0
Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., Ferguson-Smith, A.C., 2008. A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. Dev Cell 15, 547 557. https://doi.org/10.1016/j.devcel.2008.08.014
Li, Xuexia, Zhu, Q., Shi, X., Cheng, Y., Li, Xueliu, Xu, H., Duan, X., Hsieh-Wilson, L.C., Chu, J., Pelletier, J., Ni, M., Zheng, Z., Li, S., Yi, W., 2019. O-GlcNAcylation of core components of the translation initiation machinery regulates protein synthesis. https://doi.org/10.1073/pnas.1813026116
Li, Y., Zhang, Z., Chen, J., Liu, W., Lai, W., Liu, B., Li, X., Liu, L., Xu, S., Dong, Q., Wang, M., Duan, X., Tan, J., Zheng, Y., Zhang, P., Fan, G., Wong, J., Xu, G.L., Wang, Z., Wang, H., Gao, S., Zhu, B., 2018. Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. Nature 564, 136–140. https://doi.org/10.1038/s41586-018-

0751-5
Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M.M., Kirov, N., Rushlow, C., 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature 456, 400–403. https://doi.org/10.1038/nature07388
Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656
Lin, C.-J., Koh, F.M., Wong, P., Conti, M., Ramalho-Santos, M., 2014. Hira-Mediated H3.3 Incorporation Is Required for DNA Replication and Ribosomal RNA Transcription in the Mouse Zygote. Dev Cell 30, 268–279. https://doi.org/10.1016/j.devcel.2014.06.022
Liu, B., Xu, Q., Wang, Q., Feng, S., Lai, F., Wang, P., Zheng, F., Xiang, Y., Wu, J., Nie, J., Qiu, C., Xia, W., Li, L., Yu, G., Lin, Z., Xu, K., Xiong, Z., Kong, F., Liu, L., Huang, C., Yu, Y., Na, J., Xie, W., 2020. The landscape of RNA Pol II binding reveals a stepwise transition during ZGA. Nature 587, 139–144. https://doi.org/10.1038/s41586-020-2847-y
Liu, C., Du, Y., 2019. Microinjection: Methods and Protocols.
Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G.W., Gong, C.X., 2004. O-GlcNAcylation regulates phosphorylation of tau: A mechanism involved in Alzheimer's disease. Proceedings of the National Academy of Sciences of the United States of America 101, 10804–10809. https://doi.org/10.1073/pnas.0400348101/suppl_file/00348fig6.pdf
Liu, J., He, L., Collins, I., Ge, H., Libutti, D., Li, J., Egly, J.-M., Levens, D., 2000. The FBP Interacting Repressor Targets TFIIH to Inhibit Activated Transcription. Mol Cell 5, 331–341. https://doi.org/10.1016/s1097-2765(00)80428-1
Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., Wong, K.-Y., Sung, K.W., Lee, C.W.H., Zhao, X.-D., Chiu, K.-P., Lipovich, L., Kuznetsov, V.A., Robson, P., Stanton, L.W., Wei, C.-L., Ruan, Y., Lim, B., Ng, H.-H., 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38, 431–440. https://doi.org/10.1038/ng1760
Love, Dona C, Ghosh, S., Mondoux, M.A., Fukushige, T., Wang, P., Wilson, M.A., Iser, W.B., Wolkow, C.A., Krause, M.W., Hanover, J.A., 2010. Dynamic O-GlcNAc cycling at promoters of Caenorhabditis elegans genes regulating longevity, stress, and immunity. Proceedings of the National Academy of Sciences of the United States of America 107, 7413–8. https://doi.org/10.1073/pnas.0911857107
Love, D.C., Kochan, J., Cathey, R.L., Shin, S.-H., Hanover, J.A., 2003. Mitochondrial and nucleocytoplasmic targeting of O-linked GlcNAc transferase. Journal of Cell Science 116, 647–654. https://doi.org/10.1242/jcs.00246
Love, Dona C., Krause, M.W., Hanover, J.A., 2010. O-GlcNAc cycling: Emerging roles in development and epigenetics. Semin Cell Dev Biol 21, 646–654. https://doi.org/10.1016/j.semcdb.2010.05.001
Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion

for RNA-seq data with DESeq2. Genome Biology 15, 550. https://doi.org/10.1186/s13059-014-0550-8
Low, K.J., Ansari, M., Jamra, R.A., Clarke, A., Chehadeh, S.E., FitzPatrick, D.R., Greenslade, M., Henderson, A., Hurst, J., Keller, K., Kuentz, P., Prescott, T., Roessler, F., Selmer, K.K., Schneider, M.C., Stewart, F., Tatton-Brown, K., Thevenon, J., Vigeland, M.D., Vogt, J., Willems, M., Zonana, J., Study, D.D.D., Smithson, S.F., 2017. PUF60 variants cause a syndrome of ID, short stature, microcephaly, coloboma, craniofacial, cardiac, renal and spinal features. Eur J Hum Genet 25, 552–559. https://doi.org/10.1038/ejhg.2017.27
Lubas, W.A., Frank, D.W., Krause, M., Hanover, J.A., 1997a. O-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. The Journal of biological chemistry 272, 9316–24. https://doi.org/10.1074/jbc.272.14.9316
Lubas, W.A., Frank, D.W., Krause, M., Hanover, J.A., 1997b. O-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. The Journal of biological chemistry 272, 9316–24. https://doi.org/10.1074/jbc.272.14.9316
Lv, P., Du, Y., He, C., Peng, L., Zhou, X., Wan, Y., Zeng, M., Zhou, W., Zou, P., Li, C., Zhang, M., Dong, S., Chen, X., 2022. O-GlcNAcylation modulates liquid–liquid phase separation of SynGAP/PSD-95. Nature Chemistry 2022 1–10. https://doi.org/10.1038/s41557-022-00946-9
Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., Pfaff, S.L., 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 2012 487:7405 487, 57–63. https://doi.org/10.1038/nature11244
Malkowska, A., Penfold, C., Bergmann, S., Boroviak, T.E., 2022. A hexa-species transcriptome atlas of mammalian embryogenesis delineates metabolic regulation across three different implantation modes. Nature Communications 2022 13:1 13, 1–12. https://doi.org/10.1038/s41467-022-30194-x
Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S., 2002. The Protein Kinase Complement of the Human Genome. Science 298, 1912–1934. https://doi.org/10.1126/science.1075762
Marotta, N.P., Lin, Y.H., Lewis, Y.E., Ambroso, M.R., Zaro, B.W., Roth, M.T., Arnold, D.B., Langen, R., Pratt, M.R., 2015. O-GlcNAc modification blocks the aggregation and toxicity of the protein α-synuclein associated with Parkinson's disease. Nature Chemistry 7, 913–920. https://doi.org/10.1038/nchem.2361
Marshall, S., Bacote, V., Traxinger, R.R., 1991. Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role of hexosamine biosynthesis in the induction of insulin resistance. J Biol Chem 266, 4706–4712. https://doi.org/10.1016/s0021-9258(19)67706-9
Marshall, S., Nadeau, O., Yamasaki, K., 2004. Dynamic Actions of Glucose and Glucosamine on Hexosamine Biosynthesis in Isolated Adipocytes DIFFERENTIAL EFFECTS ON

GLUCOSAMINE 6-PHOSPHATE, UDP-N-ACETYLGLUCOSAMINE, AND ATP LEVELS*. J Biol Chem 279, 35313–35319. https://doi.org/10.1074/jbc.m404133200
Martinez-Fleites, C., Macauley, M.S., He, Y., Shen, D.L., Vocadlo, D.J., Davies, G.J., 2008. Structure of an O-GlcNAc transferase homolog provides insight into intracellular glycosylation. Nature Structural and Molecular Biology 15, 764–765. https://doi.org/10.1038/nsmb.1443
McClain, D.A., Lubas, W.A., Cooksey, R.C., Hazel, M., Parker, G.J., Love, D.C., Hanover, J.A., 2002. Altered glycan-dependent signaling induces insulin resistance and hyperleptinemia. Proc National Acad Sci 99, 10695–10699. https://doi.org/10.1073/pnas.152346899
Mendelson, K., Evans, T., Hla, T., 2013. Sphingosine 1-phosphate signalling. Development 141, 5–9. https://doi.org/10.1242/dev.094805
Mitchell, C.W., Czajewski, I., Aalten, D.M.F. van, 2022. Bioinformatic prediction of putative conveyers of O-GlcNAc Transferase intellectual disability. The Journal of biological chemistry 0, 102276. https://doi.org/10.1016/j.jbc.2022.102276/attachment/52797df3-70a6-41d6-935c-0ab0953cac88/mmc4.xlsx
Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., Yamanaka, S., 2003. The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. Cell 113, 631–642. https://doi.org/10.1016/s0092-8674(03)00393-3
Miyazawa, H., Aulehla, A., 2018. Revisiting the role of metabolism during development. Development (Cambridge) 145. https://doi.org/10.1242/dev.131110
Miyazawa, H., Yamaguchi, Y., Sugiura, Y., Honda, K., Kondo, K., Matsuda, F., Yamamoto, T., Suematsu, M., Miura, M., 2017. Rewiring of embryonic glucose metabolism via suppression of PFK-1 and aldolase during mouse chorioallantoic branching. Development (Cambridge) 144, 63–73. https://doi.org/10.1242/dev.138545
Molden, R.C., Goya, J., Khan, Z., Garcia, B.A., 2014. Stable Isotope Labeling of Phosphoproteins for Large-scale Phosphorylation Rate Determination*. Mol Cell Proteomics 13, 1106–1118. https://doi.org/10.1074/mcp.o113.036145
Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., Köster, J., Reich, M., 2021. Sustainable data analysis with Snakemake. F1000Research 2021 10:33 10, 33. https://doi.org/10.12688/f1000research.29032.1
Monsigny, M., Roche, A., Sene, C., Maget-Dana, R., Delmotte, F., 1980. Sugar-Lectin Interactions: How Does Wheat-Germ Agglutinin Bind Sialoglycoconjugates? Eur J Biochem 104, 147–153. https://doi.org/10.1111/j.1432-1033.1980.tb04410.x
Morawska, M., Ulrich, H.D., 2013. An expanded tool kit for the auxin-inducible degron system in budding yeast. Yeast 30, 341–351. https://doi.org/10.1002/yea.2967
Muha, V., Authier, F., Szoke-Kovacs, Z., Johnson, S., Gallagher, J., McNeilly, A.,

McCrimmon, R.J., Teboul, L., Aalten, D.M.F. van, 2021. Loss of O-GlcNAcase catalytic activity leads to defects in mouse embryogenesis. J Biol Chem 296, 100439. https://doi.org/10.1016/j.jbc.2021.100439
Murphy, B.D., 2020. Under Arrest: The Embryo in Diapause. Developmental Cell 52, 139–140. https://doi.org/10.1016/j.devcel.2020.01.002
Myers, S.A., Panning, B., 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 of the United States of America 108, 9490–9495. https://doi.org/10.1073/pnas.1019289108
Nagaraj, R., Sharpley, M.S., Chi, F., Braas, D., Zhou, Y., Kim, R., Clark, A.T., Banerjee, U., 2017. Nuclear Localization of Mitochondrial TCA Cycle Enzymes as a Critical Step in Mammalian Zygotic Genome Activation. Cell 168, 210-223.e11. https://doi.org/10.1016/j.cell.2016.12.026
Nakamura, T., Liu, Y.-J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., Nakano, T., 2012. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature 486, 415–419. https://doi.org/10.1038/nature11093
Nakao, A., Yoshihama, M., Kenmochi, N., 2004. RPG: the Ribosomal Protein Gene database. Nucleic Acids Res 32, D168–D170. https://doi.org/10.1093/nar/gkh004
Natsume, T., Kiyomitsu, T., Saga, Y., Kanemaki, M.T., 2016. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. Cell Reports 15, 210–218. https://doi.org/10.1016/j.celrep.2016.03.001
Nelson, D.L., Cox, M., 2021. Lehninger Principles of Biochemistry: International Edition.
Newport, J., Kirschner, M., 1982. A major developmental transition in early xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell 30, 675–686. https://doi.org/10.1016/0092-8674(82)90272-0
Nguyen, D.K., Disteche, C.M., 2006. Dosage compensation of the active X chromosome in mammals. Nat Genet 38, 47–53. https://doi.org/10.1038/ng1705
Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., Smith, A., 1998. Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. Cell 95, 379–391. https://doi.org/10.1016/s0092-8674(00)81769-9
Nie, H., Ju, H., Fan, J., Shi, X., Cheng, Y., Cang, X., Zheng, Z., Duan, X., Yi, W., 2020. O-GlcNAcylation of PGK1 coordinates glycolysis and TCA cycle to promote tumor growth. Nature Communications 11, 1–14. https://doi.org/10.1038/s41467-019-13601-8
Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., Kanemaki, M., 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nature Methods 6, 917–922. https://doi.org/10.1038/nmeth.1401
Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N.,

Dekker, J., Mirny, L.A., Bruneau, B.G., 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell 169, 930-944.e22. https://doi.org/10.1016/j.cell.2017.05.004
Nugent, B.M., O'Donnell, C.M., Epperson, C.N., Bale, T.L., 2018. Placental H3K27me3 establishes female resilience to prenatal insults. Nat Commun 9, 2555. https://doi.org/10.1038/s41467-018-04992-1
O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., Jenuwein, T., 2001. The Polycomb-Group GeneEzh2 Is Required for Early Mouse Development. Mol Cell Biol 21, 4330–4336. https://doi.org/10.1128/mcb.21.13.4330-4336.2001
O'Connor, L., Gilmour, J., Bonifer, C., 2016. The Role of the Ubiquitously Expressed Transcription Factor Sp1 in Tissue-specific Transcriptional Regulation and in Disease. Yale J Biology Medicine 89, 513–525.
O'Donnell, N., Zachara, N.E., Hart, G.W., Marth, J.D., 2004a. Ogt-Dependent X-Chromosome-Linked Protein Glycosylation Is a Requisite Modification in Somatic Cell Function and Embryo Viability. Mol Cell Biol 24, 1680 1690. https://doi.org/10.1128/mcb.24.4.1680-1690.2004
O'Donnell, N., Zachara, N.E., Hart, G.W., Marth, J.D., 2004b. Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. Molecular and cellular biology 24, 1680–90. https://doi.org/10.1128/mcb.24.4.1680-1690.2004
Ogawa, M., Okajima, T., 2019. Structure and function of extracellular O-GlcNAc. Curr Opin Struc Biol 56, 72–77. https://doi.org/10.1016/j.sbi.2018.12.002
Ohn, T., Kedersha, N., Hickman, T., Tisdale, S., 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. https://doi.org/10.1038/ncb1783
Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., Heard, E., 2004. Epigenetic Dynamics of Imprinted X Inactivation during Early Mouse Development. Science 303, 644–649. https://doi.org/10.1126/science.1092727
Page-McCaw, P.S., Amonlirdviman, K., Sharp, P.A., 1999. PUF60: A novel U2AF65-related splicing activity. Rna 5, 1548–1560. https://doi.org/10.1017/s1355838299991938
Pantaleon, M., Tan, H.Y., Kafer, G.R., Kaye, P.L., 2010. Toxic Effects of Hyperglycemia Are Mediated by the Hexosamine Signaling Pathway and O-Linked Glycosylation in Early Mouse Embryos1. Biol Reprod 82, 751–758. https://doi.org/10.1095/biolreprod.109.076661
Park, S.K., Zhou, X., Pendleton, K.E., Hunter, O.V., Kohler, J.J., O'Donnell, K.A., Conrad, N.K., 2017. A Conserved Splicing Silencer Dynamically Regulates O-GlcNAc Transferase Intron Retention and O-GlcNAc Homeostasis. Cell Reports 20, 1088–1099. https://doi.org/10.1016/j.celrep.2017.07.017
Partch, C.L., Shields, K.F., Thompson, C.L., Selby, C.P., Sancar, A., 2006. Posttranslational regulation of the mammalian circadian clock by cryptochrome and protein phosphatase 5.

Proc National Acad Sci 103, 10467–10472. https://doi.org/10.1073/pnas.0604138103
Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nature Publishing Group 14, 417–419. https://doi.org/10.1038/nmeth.4197
Peaston, A.E., Evsikov, A.V., Graber, J.H., Vries, W.N. de, Holbrook, A.E., Solter, D., Knowles, B.B., 2004. Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos. Developmental Cell 7, 597–606. https://doi.org/10.1016/j.devcel.2004.09.004
Perreault, S.D., Barbee, R.R., Slott, V.L., 1988. Importance of glutathione in the acquisition and maintenance of sperm nuclear decondensing activity in maturing hamster oocytes. Dev Biol 125, 181–186. https://doi.org/10.1016/0012-1606(88)90070-x
Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., Sandberg, R., 2014. Full-length RNA-seq from single cells using Smart-seq2. Nature Protocols 9, 171–181. https://doi.org/10.1038/nprot.2014.006
Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., Reik, W., Srinivas, S., Simons, B.D., Nichols, J., Marioni, J.C., Göttgens, B., 2019. A single-cell molecular map of mouse gastrulation and early organogenesis. Nature 566, 490–495. https://doi.org/10.1038/s41586-019-0933-9
Pravata, V.M., Muha, V., 1+, G., Ferenbach, A.T., Kakade, P., Vandadi, V., Wilmes, A.C., Borodkin, V.S., Joss, S., Stavridis, M.P., Aalten, D.M.F.V., 2019a. Catalytic deficiency of O-GlcNAc transferase leads to X-linked intellectual disability 116, 14961–14970. https://doi.org/10.1073/pnas.1900065116
Pravata, V.M., Muha, V., 1+, G., Ferenbach, A.T., Kakade, P., Vandadi, V., Wilmes, A.C., Borodkin, V.S., Joss, S., Stavridis, M.P., Aalten, D.M.F.V., 2019b. Catalytic deficiency of O-GlcNAc transferase leads to X-linked intellectual disability 116, 14961–14970. https://doi.org/10.1073/pnas.1900065116
Pravata, V.M., Omelková, M., Stavridis, M.P., Desbiens, C.M., Stephen, H.M., Lefeber, D.J., Gecz, J., Gundogdu, M., Õunap, K., Joss, S., Schwartz, C.E., Wells, L., Aalten, D.M.F. van, 2020. An intellectual disability syndrome with single-nucleotide variants in O-GlcNAc transferase. European Journal of Human Genetics 2020 28:6 28, 706–714. https://doi.org/10.1038/s41431-020-0589-9
Prioleau, M.-N., Huet, J., Sentenac, A., Méchali, M., 1994. Competition between chromatin and transcription complex assembly regulates gene expression during early development. Cell 77, 439–449. https://doi.org/10.1016/0092-8674(94)90158-9
Qian, K., Wang, S., Fu, M., Zhou, J., Singh, J.P., Li, M.D., Yang, Y., Zhang, K., Wu, J., Nie, Y., Ruan, H.B., Yang, X., 2018. Transcriptional regulation of O-GlcNAc homeostasis is disrupted in pancreatic cancer. Journal of Biological Chemistry 293, 13989–14000. https://doi.org/10.1074/jbc.ra118.004709

Qian, Z., Li, C., Zhao, S., Zhang, H., Ma, R., Ge, X., Jing, J., Chen, L., Ma, J., Yang, Y., Zheng, L., Zhang, K., He, Z., Xue, M., Lin, Y., Jueraitetibaike, K., Feng, Y., Cao, C., Tang, T., Sun, S., Teng, H., Zhao, W., Yao, B., 2023. Age-related elevation of O-GlcNAc causes meiotic arrest in male mice. Cell Death Discov 9, 163. https://doi.org/10.1038/s41420-023-01433-x

Quadros, R.M., Miura, H., Harms, D.W., Akatsuka, H., Sato, T., Aida, T., Redder, R., Richardson, G.P., Inagaki, Y., Sakai, D., Buckley, S.M., Seshacharyulu, P., Batra, S.K., Behlke, M.A., Zeiner, S.A., Jacobi, A.M., Izu, Y., Thoreson, W.B., Urness, L.D., Mansour, S.L., Ohtsuka, M., Gurumurthy, C.B., 2017. Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. Genome Biol 18, 92. https://doi.org/10.1186/s13059-017-1220-4

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., Trono, D., 2011. In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. Molecular Cell 44, 361–372. https://doi.org/10.1016/j.molcel.2011.08.032

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Radermacher, P.T., Myachina, F., Bosshardt, F., Pandey, R., Mariappa, D., Müller, H.-A.J., Lehner, C.F., 2014. O-GlcNAc reports ambient temperature and confers heat resistance on ectotherm development. Proc National Acad Sci 111, 5592–5597. https://doi.org/10.1073/pnas.1322396111

Rajavashisth, T.B., Taylor, A.K., Andalibi, A., Svenson, K.L., Lusis, A.J., 1989. Identification of a Zinc Finger Protein That Binds to the Sterol Regulatory Element. Science 245, 640–643. https://doi.org/10.1126/science.2562787

Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W., Lewis, B.A., 2012. Evidence of the Involvement of O-GlcNAc-modified Human RNA Polymerase II CTD in Transcription in Vitro and in Vivo *. J Biol Chem 287, 23549–23561. https://doi.org/10.1074/jbc.m111.330910

Risso, D., Ngai, J., Speed, T.P., Dudoit, S., 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol 32, 896–902. https://doi.org/10.1038/nbt.2931

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D.B., Trono, D., 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature 463, 237–240. https://doi.org/10.1038/nature08674

Ruan, H.-B., Dietrich, M.O., Liu, Z.-W., Zimmer, M.R., Li, M.-D., Singh, J.P., Zhang, K., Yin, R., Wu, J., Horvath, T.L., Yang, X., 2014. O-GlcNAc Transferase Enables AgRP Neurons to Suppress Browning of White Fat. Cell 159, 306–317.

https://doi.org/10.1016/j.cell.2014.09.010
Ruan, H.B., Han, X., Li, M.D., Singh, J.P., Qian, K., Azarhoush, S., Zhao, L., Bennett, A.M., Samuel, V.T., Wu, J., Yates, J.R., Yang, X., 2012. O-GlcNAc Transferase/Host Cell Factor C1 Complex Regulates Gluconeogenesis by Modulating PGC-1α Stability. Cell Metabolism 16, 226–237. https://doi.org/10.1016/j.cmet.2012.07.006
Sacco, F., Perfetto, L., Castagnoli, L., Cesareni, G., 2012. The human phosphatase interactome: An intricate family portrait. Febs Lett 586, 2732–2739. https://doi.org/10.1016/j.febslet.2012.05.008
Sakabe, K., Wang, Z., Hart, G.W., 2010. β-N-acetylglucosamine (O-GlcNAc) is part of the histone code. Proc National Acad Sci 107, 19915–19920. https://doi.org/10.1073/pnas.1009023107
Sakaidani, Y., Nomura, T., Matsuura, A., Ito, M., Suzuki, E., Murakami, K., Nadano, D., Matsuda, T., Furukawa, K., Okajima, T., 2011. O-Linked-N-acetylglucosamine on extracellular protein domains mediates epithelial cell–matrix interactions. Nat Commun 2, 583. https://doi.org/10.1038/ncomms1591
Sakashita, A., Kitano, T., Ishizu, H., Guo, Y., Masuda, H., Ariura, M., Murano, K., Siomi, H., 2023. Transcription of MERVL retrotransposons is required for preimplantation embryo development. Nat Genet 55, 484–495. https://doi.org/10.1038/s41588-023-01324-y
Santenard, A., Ziegler-Birling, C., Koch, M., Tora, L., Bannister, A.J., Torres-Padilla, M.-E., 2010. Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. Nat Cell Biol 12, 853–862. https://doi.org/10.1038/ncb2089
Santos, F., Peters, A.H., Otte, A.P., Reik, W., Dean, W., 2005. Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. Dev Biol 280, 225–236. https://doi.org/10.1016/j.ydbio.2005.01.025
Sardina, J.L., Collombet, S., Tian, T.V., Gómez, A., Stefano, B.D., Berenguer, C., Brumbaugh, J., Stadhouders, R., Segura-Morales, C., Gut, M., Gut, I.G., Heath, S., Aranda, S., Croce, L.D., Hochedlinger, K., Thieffry, D., Graf, T., 2018. Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. Cell Stem Cell 23, 727-741.e9. https://doi.org/10.1016/j.stem.2018.08.016
Schalock, R.L., Luckasson, R., Tassé, and M.J., 2021. Intellectual Disability: Definition, Diagnosis, Classification, and Systems of Supports, 12th Edition.
Scheffler, K., Uraji, J., Jentoft, I., Cavazza, T., Mönnich, E., Mogessie, B., Schuh, M., 2021. Two mechanisms drive pronuclear migration in mouse zygotes. Nat Commun 12, 841. https://doi.org/10.1038/s41467-021-21020-x
Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9, 671–675. https://doi.org/10.1038/nmeth.2089
Schrode, N., Saiz, N., Talia, S.D., Hadjantonakis, A.-K., 2014. GATA6 Levels Modulate Primitive Endoderm Cell Fate Choice and Timing in the Mouse Blastocyst. Dev Cell 29, 454–467. https://doi.org/10.1016/j.devcel.2014.04.011

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., Cavalli, G., 2007. Genome Regulation by Polycomb and Trithorax Proteins. Cell 128, 735–745. https://doi.org/10.1016/j.cell.2007.02.009
Schwartz, Y.B., Pirrotta, V., 2007. Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet 8, 9–22. https://doi.org/10.1038/nrg1981
Scognamiglio, R., Cabezas-Wallscheid, N., Thier, M.C., Altamura, S., Reyes, A., Prendergast, Á.M., Baumgärtner, D., Carnevalli, L.S., Atzberger, A., Haas, S., von Paleske, L., Boroviak, T., Wörsdörfer, P., Essers, M.A.G., Kloz, U., Eisenman, R.N., Edenhofer, F., Bertone, P., Huber, W., van der Hoeven, F., Smith, A., Trumpp, A., 2016. Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. Cell 164, 668–680. https://doi.org/10.1016/j.cell.2015.12.033
Sekine, O., Love, D.C., Rubenstein, D.S., Hanover, J.A., 2010. Blocking O-Linked GlcNAc Cycling in Drosophila Insulin-producing Cells Perturbs Glucose-Insulin Homeostasis*. J Biol Chem 285, 38684–38691. https://doi.org/10.1074/jbc.m110.155192
Senft, A.D., Macfarlan, T.S., 2021. Transposable elements shape the evolution of mammalian development. Nature Reviews Genetics 2021 22:11 22, 691–711. https://doi.org/10.1038/s41576-021-00385-1
Shafi, R., Iyer, S.P., Ellies, L.G., O'Donnell, N., Marek, K.W., Chui, D., Hart, G.W., 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. Proc National Acad Sci 97, 5735 5739. https://doi.org/10.1073/pnas.100471497
Shaper, N.L., Shaper, J.H., Meuth, J.L., Fox, J.L., Chang, H., Kirsch, I.R., Hollis, G.F., 1986. Bovine galactosyltransferase: identification of a clone by direct immunological screening of a cDNA expression library. Proc National Acad Sci 83, 1573–1577. https://doi.org/10.1073/pnas.83.6.1573
Shen, W., Gong, B., Xing, C., Zhang, L., Sun, J., Chen, Y., Yang, C., Yan, L., Chen, L., Yao, L., Li, G., Deng, H., Wu, X., Meng, A., 2022. Comprehensive maturity of nuclear pore complexes regulates zygotic genome activation. Cell 185, 4954-4970.e20. https://doi.org/10.1016/j.cell.2022.11.011
Sheng, G., Arias, A.M., Sutherland, A., 2021. The primitive streak and cellular principles of building an amniote body through gastrulation. Science 374, abg1727. https://doi.org/10.1126/science.abg1727
Shi, F.-T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M.A., Wan, M., Songyang, Z., 2013. Ten-Eleven Translocation 1 (Tet1) Is Regulated by O-Linked N-Acetylglucosamine Transferase (Ogt) for Target Gene Repression in Mouse Embryonic Stem Cells*. J Biol Chem 288, 20776–20784. https://doi.org/10.1074/jbc.m113.460386
Shin, E.M., Huynh, V.T., Neja, S.A., Liu, C.Y., Raju, A., Tan, K., Tan, N.S., Gunaratne, J., Bi, X., Iyer, L.M., Aravind, L., Tergaonkar, V., 2021. GREB1: An evolutionarily conserved protein with a glycosyltransferase domain links ERα glycosylation and stability to cancer. Sci

Adv 7, eabe2470. https://doi.org/10.1126/sciadv.abe2470
Shin, H., Leung, A., Costello, K.R., Senapati, P., Kato, H., Schones, D.E., 2022. Glucose-regulated O-GlcNAcylation of DNMT1 inhibits DNA methyltransferase activity and maintenance of genomic methylation. Biorxiv 2022.05.11.491514. https://doi.org/10.1101/2022.05.11.491514
Shu, X.E., Mao, Y., Jia, L., Qian, S.-B., 2021. Dynamic eIF3a O-GlcNAcylation controls translation reinitiation during nutrient stress. Nature Chemical Biology 2021 1–8. https://doi.org/10.1038/s41589-021-00913-4
Sinclair, D.A.R., Syrzycka, M., Macauley, M.S., Rastgardani, T., Komljenovic, I., Vocadlo, D.J., Brock, H.W., Honda, B.M., 2009. Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc). Proceedings of the National Academy of Sciences of the United States of America 106, 13427–13432. https://doi.org/10.1073/pnas.0904638106
Slavov, N., Semrau, S., Airoldi, E., Budnik, B., van Oudenaarden, A., 2015. Differential Stoichiometry among Core Ribosomal Proteins. Cell Reports 13, 865–873. https://doi.org/10.1016/j.celrep.2015.09.056
Smallwood, S.A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., Kelsey, G., 2011. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nature Genetics 43, 811–814. https://doi.org/10.1038/ng.864
Smith, Z.D., Shi, J., Gu, H., Donaghey, J., Clement, K., Cacchiarelli, D., Gnirke, A., Michor, F., Meissner, A., 2017. Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. Nature 2017 549:7673 549, 543–547. https://doi.org/10.1038/nature23891
Snow, C.M., Senior, A., Gerace, L., 1987. Monoclonal antibodies identify a group of nuclear pore complex glycoproteins. J Cell Biology 104, 1143–1156. https://doi.org/10.1083/jcb.104.5.1143
Sohn, K.-C., Do, S.-I., 2005. Transcriptional regulation and O-GlcNAcylation activity of zebrafish OGT during embryogenesis. Biochem Bioph Res Co 337, 256–263. https://doi.org/10.1016/j.bbrc.2005.09.049
Soneson, C., Love, M.I., Robinson, M.D., 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research 2016 4:1521 4, 1521. https://doi.org/10.12688/f1000research.7563.2
Soriano, P., 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. Nature Genetics 21, 70–71. https://doi.org/10.1038/5007
Soshnikova, N., Duboule, D., 2009. Epigenetic Temporal Control of Mouse Hox Genes in Vivo. Science 324, 1320–1323. https://doi.org/10.1126/science.1171468
Stephens, M., 2017. False discovery rates: a new deal. Biostatistics 18, 275–294. https://doi.org/10.1093/biostatistics/kxw041

Strumpf, D., Mao, C.-A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., Rossant, J., 2005. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development 132, 2093–2102. https://doi.org/10.1242/dev.01801
Sun, M.-A., Wolf, G., Wang, Y., Senft, A.D., Ralls, S., Jin, J., Dunn-Fletcher, C.E., Muglia, L.J., Macfarlan, T.S., 2021. Endogenous retroviruses drive lineage-specific regulatory evolution across primate and rodent placentae. Mol Biol Evol 38, msab223-. https://doi.org/10.1093/molbev/msab223
Sun, Y., Nien, C.-Y., Chen, K., Liu, H.-Y., Johnston, J., Zeitlinger, J., Rushlow, C., 2015. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. Genome Res 25, 1703–1714. https://doi.org/10.1101/gr.192542.115
Svoboda, P., 2018. Mammalian zygotic genome activation. Seminars in Cell & Developmental Biology 84, 118–126. https://doi.org/10.1016/j.semcdb.2017.12.006
Swamy, M., Pathak, S., Grzes, K.M., Damerow, S., Sinclair, L.V., Aalten, D.M.F.V., Cantrell, D.A., 2016. Glucose and glutamine fuel protein O-GlcNAcylation to control T cell self-renewal and malignancy. Nature Immunology 17, 712–720. https://doi.org/10.1038/ni.3439
Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., Hunt, D.F., 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc National Acad Sci 101, 9528–9533. https://doi.org/10.1073/pnas.0402700101
Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., Rao, A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935. https://doi.org/10.1126/science.1170116
Tam, P.P.L., Loebel, D.A.F., 2007. Gene function in mouse embryogenesis: get set for gastrulation. Nat Rev Genet 8, 368–381. https://doi.org/10.1038/nrg2084
Tan, Z.-W., Fei, G., Paulo, J.A., Bellaousov, S., Martin, S.E.S., Duveau, D.Y., Thomas, C.J., Gygi, S.P., Boutz, P.L., Walker, S., 2020. O-GlcNAc regulates gene expression by controlling detained intron splicing. Nucleic Acids Research. https://doi.org/10.1093/nar/gkaa263
Teissandier, A., Servant, N., Barillot, E., Bourc'His, D., 2019. Tools and best practices for retrotransposon analysis using high-throughput sequencing data. Mobile DNA 10, 52. https://doi.org/10.1186/s13100-019-0192-1
Toda, T., Hsu, J.Y., Linker, S.B., Hu, L., Schafer, S.T., Mertens, J., Jacinto, F.V., Hetzer, M.W., Gage, F.H., 2017. Nup153 Interacts with Sox2 to Enable Bimodal Gene Regulation and Maintenance of Neural Progenitor Cells. Cell Stem Cell 21, 618-634.e7. https://doi.org/10.1016/j.stem.2017.08.012
Torres, C.R., Hart, G.W., 1984. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. J Biol Chem 259, 3308–3317. https://doi.org/10.1016/s0021-9258(17)43295-9

Tropberger, P., Pott, S., Keller, C., Kamieniarz-Gdula, K., Caron, M., Richter, F., Li, G., Mittler, G., Liu, E.T., Bühler, M., Margueron, R., Schneider, R., 2013. Regulation of Transcription through Acetylation of H3K122 on the Lateral Surface of the Histone Octamer. Cell 152, 859–872. https://doi.org/10.1016/j.cell.2013.01.032
Tyagi, S., Chabes, A.L., Wysocka, J., Herr, W., 2007a. E2F Activation of S Phase Promoters via Association with HCF-1 and the MLL Family of Histone H3K4 Methyltransferases. Molecular Cell 27, 107–119. https://doi.org/10.1016/j.molcel.2007.05.030
Tyagi, S., Chabes, A.L., Wysocka, J., Herr, W., 2007b. E2F Activation of S Phase Promoters via Association with HCF-1 and the MLL Family of Histone H3K4 Methyltransferases. Molecular Cell 27, 107–119. https://doi.org/10.1016/j.molcel.2007.05.030
Usui, N., Ogura, A., Kimura, Y., Yanagimachi, R., 1997. Sperm nuclear envelope: breakdown of intrinsic envelope and de novo formation in hamster oocytes or eggs. Zygote 5, 35–46. https://doi.org/10.1017/s0967199400003543
Vaidyanathan, K., Niranjan, T., Selvan, N., Teo, C.F., May, M., Patel, S., Weatherly, B., Skinner, C., Opitz, J., Carey, J., Viskochil, D., Gecz, J., Shaw, M., Peng, Y., Alexov, E., Wang, T., Schwartz, C., Wells, L., 2017a. Identification and characterization of a missense mutation in the O-linked β-N-acetylglucosamine (O-GlcNAc) transferase gene that segregates with X-linked intellectual disability. Journal of Biological Chemistry 292, 8948–8963. https://doi.org/10.1074/jbc.m116.771030
Vaidyanathan, K., Niranjan, T., Selvan, N., Teo, C.F., May, M., Patel, S., Weatherly, B., Skinner, C., Opitz, J., Carey, J., Viskochil, D., Gecz, J., Shaw, M., Peng, Y., Alexov, E., Wang, T., Schwartz, C., Wells, L., 2017b. Identification and characterization of a missense mutation in the O-linked β-N-acetylglucosamine (O-GlcNAc) transferase gene that segregates with X-linked intellectual disability. Journal of Biological Chemistry 292, 8948–8963. https://doi.org/10.1074/jbc.m116.771030
Valvezan, A.J., Manning, B.D., 2019. Molecular logic of mTORC1 signalling as a metabolic rheostat. Nat Metabolism 1, 321–333. https://doi.org/10.1038/s42255-019-0038-7
Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N.H., Prestegard, J.H., Schnaar, R.L., Seeberger., and P.H., 2022. Essentials of Glycobiology, 4th edition.
Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., Pasini, D., 2013a. Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. Molecular Cell 49, 645–656. https://doi.org/10.1016/j.molcel.2012.12.019
Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., Pasini, D., 2013b. Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. Molecular Cell 49, 645–656. https://doi.org/10.1016/j.molcel.2012.12.019
Waghray, A., Saiz, N., Jayaprakash, A.D., Freire, A.G., Papatsenko, D., Pereira, C.-F., Lee,

D.-F., Brosh, R., Chang, B., Darr, H., Gingold, J., Kelley, K., Schaniel, C., Hadjantonakis, A.-K., Lemischka, I.R., 2015. Tbx3 Controls Dppa3 Levels and Exit from Pluripotency toward Mesoderm. Stem Cell Rep 5, 97–110. https://doi.org/10.1016/j.stemcr.2015.05.009
Walsh, C.P., Chaillet, J.R., Bestor, T.H., 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet 20, 116–117. https://doi.org/10.1038/2413
Wang, X., Yuan, Z.-F., Fan, J., Karch, K.R., Ball, L.E., Denu, J.M., Garcia, B.A., 2016. A Novel Quantitative Mass Spectrometry Platform for Determining Protein O-GlcNAcylation Dynamics. Molecular & Cellular Proteomics 15, 2462–2475. https://doi.org/10.1074/mcp.o115.049627
Wang, Y., He, Y., Su, C., Zentella, R., Sun, T., Wang, L., 2020. Nuclear Localized O-Fucosyltransferase SPY Facilitates PRR5 Proteolysis to Fine-Tune the Pace of Arabidopsis Circadian Clock. Mol Plant 13, 446–458. https://doi.org/10.1016/j.molp.2019.12.013
Wang, Y., Yen, F.S., Zhu, X.G., Timson, R.C., Weber, R., Xing, C., Liu, Y., Allwein, B., Luo, H., Yeh, H.-W., Heissel, S., Unlu, G., Gamazon, E.R., Kharas, M.G., Hite, R., Birsoy, K., 2021. SLC25A39 is necessary for mitochondrial glutathione import in mammalian cells. Nature 599, 136–140. https://doi.org/10.1038/s41586-021-04025-w
Weinberger, L., Ayyash, M., Novershtern, N., Hanna, J.H., 2016. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat Rev Mol Cell Bio 17, 155–169. https://doi.org/10.1038/nrm.2015.28
White, M.D., Angiolini, J.F., Alvarez, Y.D., Kaur, G., Zhao, Z.W., Mocskos, E., Bruno, L., Bissiere, S., Levi, V., Plachta, N., 2016. Long-Lived Binding of Sox2 to DNA Predicts Cell Fate in the Four-Cell Mouse Embryo. Cell 165, 75–87. https://doi.org/10.1016/j.cell.2016.02.032
White, M.D., Plachta, N., 2019. Specification of the First Mammalian Cell Lineages In Vivo and In Vitro. Csh Perspect Biol 12, a035634. https://doi.org/10.1101/cshperspect.a035634
Wicklow, E., Blij, S., Frum, T., Hirate, Y., Lang, R.A., Sasaki, H., Ralston, A., 2014. HIPPO Pathway Members Restrict SOX2 to the Inner Cell Mass Where It Promotes ICM Fates in the Mouse Blastocyst. Plos Genet 10, e1004618. https://doi.org/10.1371/journal.pgen.1004618
Willems, A.P., Gundogdu, M., Kempers, M.J.E., Giltay, J.C., Pfundt, R., Elferink, M., Loza, B.F., Fuijkschot, J., Ferenbach, A.T., Gassen, K.L.I.V., Aalten, D.M.F.V., Lefeber, D.J., 2017a. Mutations in N-acetylglucosamine (O-GlcNAc) transferase in patients with X-linked intellectual disability. Journal of Biological Chemistry 292, 12621–12631. https://doi.org/10.1074/jbc.m117.790097
Willems, A.P., Gundogdu, M., Kempers, M.J.E., Giltay, J.C., Pfundt, R., Elferink, M., Loza, B.F., Fuijkschot, J., Ferenbach, A.T., Gassen, K.L.I.V., Aalten, D.M.F.V., Lefeber, D.J., 2017b. Mutations in N-acetylglucosamine (O-GlcNAc) transferase in patients with X-linked intellectual disability. Journal of Biological Chemistry 292, 12621–12631. https://doi.org/10.1074/jbc.m117.790097
Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J.,

Helin, K., 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 473, 343–348. https://doi.org/10.1038/nature10066
Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., Walter, J., 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nature Communications 2, 1–8. https://doi.org/10.1038/ncomms1240
Wu, G., Han, D., Gong, Y., Sebastiano, V., Gentile, L., Singhal, N., Adachi, K., Fischedick, G., Ortmeier, C., Sinn, M., Radstaak, M., Tomilin, A., Schöler, H.R., 2013. Establishment of totipotency does not depend on Oct4A. Nat Cell Biol 15, 1089–1097. https://doi.org/10.1038/ncb2816
Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., Li, W., Li, Y., Ma, J., Peng, X., Zheng, H., Ming, J., Zhang, W., Zhang, J., Tian, G., Xu, F., Chang, Z., Na, J., Yang, X., Xie, W., 2016. The landscape of accessible chromatin in mammalian preimplantation embryos. Nature 534, 652–657. https://doi.org/10.1038/nature18606
Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., Yu, G., 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation 2, 100141. https://doi.org/10.1016/j.xinn.2021.100141
Wulff-Fuentes, E., Berendt, R.R., Massman, L., Danner, L., Malard, F., Vora, J., Kahsay, R., Stichelen, S.O.-V., 2021a. The human O-GlcNAcome database and meta-analysis. Sci Data 8, 25. https://doi.org/10.1038/s41597-021-00810-4
Wulff-Fuentes, E., Berendt, R.R., Massman, L., Danner, L., Malard, F., Vora, J., Kahsay, R., Stichelen, S.O.-V., 2021b. The human O-GlcNAcome database and meta-analysis. Sci Data 8, 25. https://doi.org/10.1038/s41597-021-00810-4
Wyatt, C.D.R., Pernaute, B., Gohr, A., Miret-Cuesta, M., Goyeneche, L., Rovira, Q., Salzer, M.C., Boke, E., Bogdanovic, O., Bonnal, S., Irimia, M., 2022. A developmentally programmed splicing failure contributes to DNA damage response attenuation during mammalian zygotic genome activation. Science Advances 8, 4935. https://doi.org/10.1126/sciadv.abn4935/suppl_file/sciadv.abn4935_supplementary_file_s1.zip
Yagi, M., Yamanaka, S., Yamada, Y., 2017. Epigenetic foundations of pluripotent stem cells that recapitulate in vivo pluripotency. Lab Invest 97, 1133–1141. https://doi.org/10.1038/labinvest.2017.87
Yamane, K., Tateishi, K., Klose, R.J., Fang, J., Fabrizio, L.A., Erdjument-Bromage, H., Taylor-Papadimitriou, J., Tempst, P., Zhang, Y., 2007. PLU-1 Is an H3K4 Demethylase Involved in Transcriptional Repression and Breast Cancer Cell Proliferation. Mol Cell 25, 801–812. https://doi.org/10.1016/j.molcel.2007.03.001
Yang, X., Ongusaha, P.P., Miles, P.D., Havstad, J.C., Zhang, F., So, W.V., Kudlow, J.E., Michell, R.H., Olefsky, J.M., Field, S.J., Evans, R.M., 2008. Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. Nature 451, 964–969.

https://doi.org/10.1038/nature06668
Yang, Y.R., Song, M., Lee, H., Jeon, Y., Choi, E.J., Jang, H.J., Moon, H.Y., Byun, H.Y., Kim, E.K., Kim, D.H., Lee, M.N., Koh, A., Ghim, J., Choi, J.H., Lee-Kwon, W., Kim, K.T., Ryu, S.H., Suh, P.G., 2012. O-GlcNAcase is essential for embryonic development and maintenance of genomic stability. Aging Cell 11, 439–448. https://doi.org/10.1111/j.1474-9726.2012.00801.x
Yesbolatova, A., Saito, Y., Kitamoto, N., Makino-Itou, H., Ajima, R., Nakano, R., Nakaoka, H., Fukui, K., Gamo, K., Tominari, Y., Takeuchi, H., Saga, Y., Hayashi, K. ichiro, Kanemaki, M.T., 2020. The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. Nature Communications 11, 1–13. https://doi.org/10.1038/s41467-020-19532-z
Yi, W., Clark, P.M., Mason, D.E., Keenan, M.C., Hill, C., Goddard, W.A., Peters, E.C., Driggers, E.M., Hsieh-Wilson, L.C., 2012. Phosphofructokinase 1 glycosylation regulates cell growth and metabolism. Science 337, 975–980. https://doi.org/10.1126/science.1222278
Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., Smith, A., 2008. The ground state of embryonic stem cell self-renewal. Nature 453, 519–523. https://doi.org/10.1038/nature06968
Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., Wang, S., 2010. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics 26, 976–978. https://doi.org/10.1093/bioinformatics/btq064
Zentella, R., Sui, N., Barnhill, B., Hsieh, W.-P., Hu, J., Shabanowitz, J., Boyce, M., Olszewski, N.E., Zhou, P., Hunt, D.F., Sun, T., 2017. The Arabidopsis O-fucosyltransferase SPINDLY activates nuclear growth repressor DELLA. Nat Chem Biol 13, 479–485. https://doi.org/10.1038/nchembio.2320
Zhang, F., Su, K., Yang, X., Bowe, D.B., Paterson, A.J., Kudlow, J.E., 2003. O-GlcNAc Modification Is an Endogenous Inhibitor of the Proteasome. Cell 115, 715–725. https://doi.org/10.1016/s0092-8674(03)00974-7
Zhang, Q., Liu, X., Gao, W., Li, P., Hou, J., Li, J., Wong, J., 2014. Differential Regulation of the Ten-Eleven Translocation (TET) Family of Dioxygenases by O-Linked β-N-Acetylglucosamine Transferase (OGT). Journal of Biological Chemistry 289, 5986–5996. https://doi.org/10.1074/jbc.m113.524140
Zhang, S., Roche, K., Nasheuer, H.P., Lowndes, N.F., 2011. Modification of histones by sugar β-N-acetylglucosamine (GlcNAc) occurs on multiple residues, including histone H3 serine 10, and is cell cycle-regulated. Journal of Biological Chemistry 286, 37483–37495. https://doi.org/10.1074/jbc.m111.284885
Zhang, Y., Xiang, Y., Yin, Q., Du, Z., Peng, X., Wang, Q., Fidalgo, M., Xia, W., Li, Y., Zhao, Z., Zhang, W., Ma, J., Xu, F., Wang, J., Li, L., Xie, W., 2017. Dynamic epigenomic landscapes during early lineage specification in mouse embryos. Nature Genetics 2017 50:1 50, 96–105. https://doi.org/10.1038/s41588-017-0003-x

Zhang, Z., Tan, E.P., VandenHull, N.J., Peterson, K.R., Slawson, C., 2014. O-GlcNAcase expression is sensitive to changes in O-GlcNAc homeostasis. Frontiers in Endocrinology 5, 1–8. https://doi.org/10.3389/fendo.2014.00206
Zheng, H., Huang, B., Zhang, B., Xiang, Y., Du, Z., Xu, Q., Li, Y., Wang, Q., Ma, J., Peng, X., Xu, F., Xie, W., 2016. Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. Mol Cell 63, 1066–1079. https://doi.org/10.1016/j.molcel.2016.08.032
Zheng, Y., Thomas, P.M., Kelleher, N.L., 2013. Measurement of acetylation turnover at distinct lysines in human histones identifies long-lived acetylation sites. Nat Commun 4, 2203. https://doi.org/10.1038/ncomms3203
Zhou, W., Choi, M., Margineantu, D., Margaretha, L., Hesson, J., Cavanaugh, C., Blau, C.A., Horwitz, M.S., Hockenbery, D., Ware, C., Ruohola-Baker, H., 2012. HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. Embo J 31, 2103–2116. https://doi.org/10.1038/emboj.2012.71
Zhu, Y., Liu, T.-W., Cecioni, S., Eskandari, R., Zandberg, W.F., Vocadlo, D.J., 2015. O-GlcNAc occurs cotranslationally to stabilize nascent polypeptide chains. Nat Chem Biol 11, 319–325. https://doi.org/10.1038/nchembio.1774
Zoncu, R., Efeyan, A., Sabatini, D.M., 2011. mTOR: from growth signal integration to cancer, diabetes and ageing. Nat Rev Mol Cell Bio 12, 21–35. https://doi.org/10.1038/nrm3025
Zou, Y., Pei, J., Long, H., Lan, L., Dong, K., Wang, T., Li, M., Zhao, Z., Zhu, L., Zhang, G., Jin, X., Wang, Y., Wen, Z., Wei, M., Feng, Y., 2023. H4S47 O-GlcNAcylation regulates the activation of mammalian replication origins. Nat Struct Mol Biol 1–12. https://doi.org/10.1038/s41594-023-00998-6

Appendices

1) Supplementary Tables

Table S1: List of generated mouse alleles.

Allele name		Mutation type	Targeted exon	Genomic coordinates insertion or substitution (GRCm39/mm39)
Short name	Standardized nomenclature			
Ogt^{Y851A}	FVB/NCrl-Ogt$^{em1(Y851A)}$Emr	substitution Y851A	exon 19 (*Ogt-201*)	substitution: GTACTGTAACTTTAATCAGTTATATAAAATTGACCCATCT -> CTATTGCAATTTCAACCAACTGGCCAAGATCGATCCTAGC, ChrX:100719847->100719886 forward strand
Ogt^{T931A}	FVB/NCrl-Ogt$^{em2(T931A)}$Emr	substitution T931A	exon 20 (*Ogt-201*)	substitution: CACC -> TGCT, ChrX:100722515->100722518 forward strand
Ogt^{Q949N}	FVB/NCrl-Ogt$^{em3(Q949N)}$Emr	substitution Q949N	exon 19 (*Ogt-201*)	substitution: GTACTGTAACTTTAATCAGTTATATAAAATTGACCCATCT -> CTATTGCAATTTCAACAACCTGTACAAGATCGATCCTAGC, ChrX:100719847->100719886 forward strand
Ogt^{H568A}	FVB/NCrl-Ogt$^{em4(H568A)}$Emr	substitution H568A	exon 13 (*Ogt-201*)	substitution: GAATCAC -> GAATGCT, ChrX:100713458->100713464 forward strand
Ogt$^{NterAID-MYC-FLAG}$	FVB/NCrl-Ogt$^{em6(AID)}$Emr	insertion [N-ter AID-MYC-FLAG]	exon 1 (*Ogt-201*)	insertion: ChrX:100683892 forward strand CCTAAAGACCCTGCAAAACCTCCCGCAAAGGCACAGGTGGTTGGATGGCCCCCAGTGAGATCCTATCGAAAGAACGTAATGGTTTCTTGTCAAAAGAGTAGTGGTGGACCTGAGGCCGCAGCCTTCGTGAAGGAACAAAAACTTATCAGCGAGGAAGACCTCGAGCAGAAGCTGATCAGCGAGGAAGACCTGGATTATAAAGACGACGATGATAAA
Rosa26^{OsTIR}	FVB;B6J;129-Gt(ROSA)26Sor$^{tm1(OsTIR)}$Emr	Insertion [OsTIR-myc-HA]	*Rosa26* locus	INSERTION CHR6: 113,076,033 (MM10): ATCTGTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGATGTCATACTTATCCTGTCCCTTTTTTTTCCACAGCTCGCGGTTGAGGACAAACTCTTCGCGGTCTTTGTGCACTTAAGATAACTTCGTATAGCATACATTATACGAAGTTATCCAGTGGGGATCGACGGTATCGATAAGCTTCCACCATGACATACTTTCCTGAAGAGGTCGTCGAACACATTTTTAGCTTCCTGCCTGCACAGAGAGATAGAAACACAGTGAGCCTGGTCTGCAAAGTGTGGTACGAGATCGAACGCCTGAGCCGGAGAGGAGTGTTCGTCGGCAACT

				GCTATGCTGTGAGAGCAGGCAGGGTCGCCGCT AGGTTTCCAAATGTGCGCGCACTGACCGTCAA GGGGAAACCCCACTTCGCCGACTTTAACCTGGT GCCCCCTGATTGGGGAGGATACGCCGGCCCTT GGATCGAGGCAGCCGCTCGCGGCTGTCATGGA CTGGAGGAACTGCGCATGAAGCGAATGGTGGT CTCTGACGAAAGTCTGGAGCTGCTGGCTCGGA GCTTCCCTAGGTTTCGCGCACTGGTGCTGATTT CTTGCGAAGGCTTCAGCACCGATGGACTGGCA GCCGTGGCCTCCCACTGTAAGCTGCTGCGGGA GCTGGACCTCCAGGAGAATGAAGTGGAGGATA GAGGCCCCAGATGGCTGTCTTGCTTCCCAGACT CATGTACCAGCCTGGTGTCCCTGAACTTTGCCT GCATCAAAGGCGAAGTGAATGCTGGGTCCCTG GAGCGGCTGGTCTCAAGAAGCCCCAACCTGAG GTCTCTGCGGCTGAACCGGAGCGTGAGCGTGG ACACTCTGGCTAAGATTCTGCTGAGAACCCCTA ACCTGGAGGATCTGGGAACCGGCAATCTGACA GACGATTTCCAGACAGAATCCTACTTTAAACTG ACTTCTGCCCTGGAGAAGTGTAAAATGCTGAG GAGTCTGTCAGGATTCTGGGATGCTTCACCCGT GTGCCTGAGCTTTATCTACCCTCTGTGTGCACA GCTGACAGGCCTGAACCTGAGCTATGCACCAA CCCTGGACGCCAGTGATCTGACAAAGATGATC TCACGCTGCGTGAAACTCCAGCGACTGTGGGT GCTGGACTGTATTTCCGATAAGGGGCTCCAGGT GGTCGCCAGCTCCTGCAAGGACCTCCAGGAGC TGAGAGTGTTCCCATCTGATTTTTACGTGGCCG GATATAGTGCTGTCACTGAGGAAGGCCTGGTG GCAGTCTCACTGGGATGCCCAAAGCTGAACAG CCTGCTGTATTTCTGTCATCAGATGACTAATGC TGCACTGGTGACCGTCGCCAAGAACTGCCCTA ATTTCACCCGATTTCGGCTGTGTATTCTGGAAC CAGGCAAACCCGACGTGGTCACATCCCAGCCA CTGGATGAAGGGTTTGGAGCTATCGTGAGAGA GTGCAAGGGACTCCAGAGGCTGAGCATTTCCG GCCTGCTGACAGACAAAGTGTTCATGTACATC GGCAAGTATGCTAAGCAGCTGGAGATGCTGAG CATTGCATTTGCCGGAGACTCCGATAAGGGCA TGATGCACGTGATGAACGGGTGTAAGAATCTG CGAAAACTGGAAATCCGGGACAGCCCTTTCGG GGATGCCGCTCTGCTGGGAAACTTTGCCAGAT ACGAGACAATGAGGAGCCTGTGGATGTCTAGT TGCAATGTGACTCTGAAGGGCTGTCAGGTCCTG GCTAGTAAAATGCCTATGCTGAACGTGGAAGT CATTAATGAGCGGGACGGGTCTAACGAAATGG AGGAAAATCATGGCGACCTGCCAAAGGTGGAG AAACTGTATGTGTATCGGACCACCGCAGGGGC AAGAGATGATGCTCCCAACTTTGTGAAGATTCT GGAGGAGCAGAAGCTGATCTCAGAGGAGGACC TGTACCCATACGATGTTCCAGATTACGCTGTCG ACTAGCATATGTACGAAGTTATAAGCTGGAAG

				TTCCTATTCTCTAGAAAGTATAGGAACTTCAAG CTTAGGTGGCACTTTTCGGGCTACCGGGTAGGG GAGGCGCTTTTCCCAAGGCAGTCTGGAGCATG CGCTTTAGCAGCCCCGCTGGGCACTTGGCGCTA CACAAGTGGCCTCTGGCCTCGCACACATTCCAC ATCCACCGGTAGGCGCCAACCGGCTCCGTTCTT TGGTGGCCCCTTCGCGCCACCTTCTACTCCTCC CCTAGTCAGGAAGTTCCCCCCCGCCCCGCAGCT CGCGTCGTGCAGGACGTGACAAATGGAAGTAG CACGTCTCACTAGTCTCGTGCAGATGGACAGC ACCGCTGAGCAATGGAAGCGGGTAGGCCTTTG GGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGC TTTCTGGGCTCAGAGGCTGGGAAGGGGTGGGT CCGGGGGCGGGCTCAGGGGCGGGCTCAGGGGC GGGGCGGGCGCCCGAAGGTCCTCCGGAGGCCC GGCATTCTGCACGCTTCAAAAGCGCACGTCTGC CGCGCTGTTCTCCTCTTCCTCATCTCCGGGCCTT TCGACCTGCAGCAGCACGTGTTGACAATTAATC ATCGGCATAGTATATCGGCATAGTATAATACG ACAAGGTGAGGAACTAAACCATGGGATCGGCC ATTGAACAAGATGGATTGCACGCAGGTTCTCC GGCCGCTTGGGTGGAGAGGCTATTCGGCTATG ACTGGGCACAACAGACAATCGGCTGCTCTGAT GCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCG CCCGGTTCTTTTTGTCAAGACCGACCTGTCCGG TGCCCTGAATGAACTGCAGGACGAGGCAGCGC GGCTATCGTGGCTGGCCACGACGGGCGTTCCTT GCGCAGCTGTGCTCGACGTTGTCACTGAAGCG GGAAGGGACTGGCTGCTATTGGGCGAAGTGCC GGGGCAGGATCTCCTGTCATCTCACCTTGCTCC TGCCGAGAAAGTATCCATCATGGCTGATGCAA TGCGGCGGCTGCATACGCTTGATCCGGCTACCT GCCCATTCGACCACCAAGCGAAACATCGCATC GAGCGAGCACGTACTCGGATGGAAGCCGGTCT TGTCGATCAGGATGATCTGGACGAAGAGCATC AGGGGCTCGCGCCAGCCGAACTGTTCGCCAGG CTCAAGGCGAGCATGCCCGACGGCGAGGATCT CGTCGTGACCCATGGCGATGCCTGCTTGCCGAA TATCATGGTGGAAAATGGCCGCTTTTCTGGATT CATCGACTGTGGCCGGCTGGGTGTGGCGGACC GCTATCAGGACATAGCGTTGGCTACCCGTGAT ATTGCTGAAGAGCTTGGCGGCGAATGGGCTGA CCGCTTCCTCGTGCTTTACGGTATCGCCGCTCC CGATTCGCAGCGCATCGCCTTCTATCGCCTTCT TGACGAGTTCTTCTGAGCGGGACTCTGGGGTTC GAAATGACCGACCAAGCGACGCCCAACCTGCC ATCACGAGATTTCGATTCCACCGCCGCCTTCTA TGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGA CGCCGGCTGGATGATCCTCCAGCGCGGGGATC TCATGCTGGAGTTCTTCGCCCACCCTAGGGGGA GGCTAACTGAAACACGGAAGGAGACAATACCG GAAGGAACCCGCGCTATGACGGCAATAAAAAG

				ACAGAATAAAACGCACGGTGTTGGGTCGTTTG TTCATAAACGCGGGGTTCGGTCCCAGGGCTGG CACTCTGTCGATACCCCACCGAGACCCCATTGG GGCCAATACGCCCGCGTTTCTTCCTTTTCCCCA CCCCACCCCCCAAGTTCGGGTGAAGGCCCAGG GCTCGCAGCCAACGTCGGGGCGGCAGGCCCTG CCAGGATCCGAAGTTCCTATTCTCTAGAAAGTA TAGGAACTTCCTCGAGTTTAAACTAAGCTGATC AGCCTCGACTGTGCCTTCTAGTTGCCAGCCATC TGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACC CTGGAAGGTGCCACTCCCACTGTCCTTTCCTAA TAAAATGAGGAAATTGCATCGCATTGTCTGAG TAGGTGTCATTCTATTCTGGGGGGTGGGGTGGG GCAGGACAGCAAGGGGGAGGATTGGGAAGAC AATAGCAGGCATGCTGGGGATGCGGTGGGCTT ATGGCTTCTGAGGCGGAAAGAACCAGCTGGGG CTCGAT

Table S2. Primers used for PCR genotyping of mouse lines.

Gene target	**Primer direction**	**Sequence (5'-3')**
Ogt$^{NterAID\text{-}MYC\text{-}FLAG}$	forward	AGTAGTGGCGGCAGTAGAAG
	reverse	TAATGGGGATGGTCAGAGGG
OsTIR insert (to genotype for the presence of the insert at least on one allele)	forward	AGAGATAGAAACACAGTGAGCC
	reverse	TCGCAAGAAATCAGCACCAG
OsTIR flanking sequence (to genotype for homozygosity, if used together with primers in row above)	forward	AGTCGCTCTGAGTTGTTATCAG
	reverse	AGGTTAGCCTTTAAGCCTGC
Ogt^{Q949N}	forward	CTTGACTCAAAACCAGGGCC
Ogt^{Q949N}	reverse	ATGGGGAAGGGAGATTCAGC
Ogt^{Y851A}	forward	CTTGACTCAAAACCAGGGCC
	reverse	ATGGGGAAGGGAGATTCAGC
Ogt^{WT} allele and *Ogt*^{T931A} allele at the region containing T931*	reverse	CATGTGGTCAGGTTTGTTGC
	forward	GCGTTTTCCAGCAGTAGGA
only *Ogt*^{T931A} allele*	reverse	GAACATCCATCCCTGTAGCA
	forward	GCGTTTTCCAGCAGTAGGA

* both pair of primers were used for genotyping of the *Ogt*^{T931A} mouse line: a positive signal with the first pair is necessary to interpret a negative signal with the second pair as a true wild type.

Table S3. Primers used for PCR genotyping of MEFs.

Gene target	**Primer direction**	**Sequence (5'-3')**
Ogt$^{NterAID\text{-}MYC\text{-}FLAG}$	forward	AGTAGTGGCGGCAGTAGAAG
	reverse	TAATGGGGATGGTCAGAGGG
OsTIR	forward	AGAGATAGAAACACAGTGAGCC
	reverse	TCGCAAGAAATCAGCACCAG
Sly-Xlr (McFarlane et al., 2013)	forward	GATGATTTGAGTGGAAATGTGAGGTA
	reverse	CTTATGTTTATAGGCATGCACCATGTA

Table S4. Primers and gBlock used for cloning the pRN3P-NLS-EGFP-Btgh/dBtgh-3NLS plasmids.

Sequence type	**Target**	**Primer direction**	**Sequence (5'-3')**
Primer	Btgh-containing plasmid (Boulard et al., 2020)	forward	GGAAGCATCAGGCGGGCTGCAGAAT AGTAAAGGAGAAGAACTTTTCACTG GAGTTGTCCCA
		reverse	GAGAAACATTCTGTCCGCCTTTGTAT AGTTCATCCATGCCATGTGTAATCC
Primer	NLS-mgfp5 gBlock	forward	CAAAGGCGGACAGAATGTTTCTCTCC AACCTCCAC
		reverse	AGTGGTAACCAGATCCGCTCACACTT TCCGTTTCTTCTTAGGATCG
NLS-mgfp5 gBlock			GATCTGAATTCCTGCAGCCCGCCACC ATGGCCTCTCCTAAGAAAAAGAGGA AAGTGGAAGCATCAGGCGGGCTGCA GAATAGTAAAGGAGAAGAACTTTTC ACTGGAGTTGTCCCAATTCTTGTTGA ATTAGATGGTGATGTTAATGGGCACA AATTTTCTGTCAGTGGAGAGGGTGAA GGTGATGCAACATACGGAAAACTTA CCCTTAAATTTATTTGCACTACTGGA AAACTACCTGTTCCATGGCCAACACT TGTCACTACTTTCACTTATGGTGTTC AATGCTTTTCAAGATACCCAGATCAT ATGAAGCGGCACGACTTCTTCAAGA GCGCCATGCCTGAGGGATACGTGCA GGAGAGGACCATCTTCTTCAATGACG ACGGGAACTACAAGACACGTGCTGA AGTCAAGTTTGAGGGAGACACCCTC GTCAACAGGATCGAGCTTAAGGGAA TCGATTTCAAGGAGGACGGAAACAT CCTCGGCCACAAGTTGGAATACAACT ACAACTCCCACAACGTATACATCATG GCCGACAAGCAAAAGAACGGCATCA AAGCCAACTTCAAGACCCGCCACAA CATCGAAGACGGCGGCGTGCAACTC GCTGATCATTATCAACAAAATACTCC AATTGGCGATGGCCCTGTCCTTTTAC CAGACAACCATTACCTGTCCACACAA TCTGCCCTTTCGAAAGATCCCAACGA AAAGAGAGACCACATGGTCCTTCTTG AGTTTGTAACAGCTGCTGGGATTACA CATGGCATGGATGAACTATACAAAG GCGGACAGAATGTTTCT

Table S5. Primers used for PCR genotyping of the *Ogt*^{T931A} blastocysts' and E7 Btgh/dBtgh-injected embryos' cDNA.

Gene target	**Primer direction**	**Sequence (5'-3')**
Xist (cDNA)	forward	TCTATCTTGTGGGTCCTGGAG
	reverse	CTCCTCTAAATCCAGGCAATCC
Ddx3y (cDNA)	forward	TGGAGGAGGAAATACAGAGAGC
	reverse	GGAGGACAATTATTTCCAGTTGC
Eif2s3y (cDNA)	forward	TGGCTGTGAAGTTGATGACC
	reverse	CCTTCTGTACGTACACCTAGG
Ogt^{WT} allele and *Ogt*^{T931A} allele*	reverse	CATGTGGTCAGGTTTGTTGC
	forward	GCGTTTTCCAGCAGTAGGA
only *Ogt*^{T931A} allele*	reverse	GAACATCCATCCCTGTAGCA
	forward	GCGTTTTCCAGCAGTAGGA

* both pair of primers were used for genotyping of the *Ogt*^{T931A} blastocysts: a positive signal with the first pair is necessary to interpret a negative signal with the second pair as a true wild type.

Table S6. Details on the generation and filtering steps of the single embryo Smart-Seq datasets.

All rows except the last one are related to the zygote microinjection experiments.

Stage	# pooled embryos in sequencing run	Average # of reads per embryo	# pooled embryos in sequencing run, per group		# embryos discarded for < 10^6 reads	# embryos discarded for high rDNA/mtgene counts	# embryos discarded for low Btgh expression	# embryos discarded for tissue contamination	# embryos discarded for aneuploidy	# embryos for DE analysis
2-cell	96	5.3* 10^6	Btgh	32	3	1	1	Not applicable	Not assessed	27
			dBtgh	32	2	0	1	Not applicable	Not assessed	29
			n.i.	32	3	0	0	Not applicable	Not assessed	29
morula	62	6.3* 10^6	Btgh	22	1	0	3	Not applicable	Not assessed	18
			dBtgh	18	1	0	1	Not applicable	Not assessed	16
			n.i.	22	0	0	0	Not applicable	Not assessed	22
blastocyst	62	6.1* 10^6	Btgh	19	0	2	0	Not applicable	3	14
			dBtgh	20	0	1	0	Not applicable	2	17
			n.i.	23	4	0	0	Not appli	2	17

									cable		
E7	25 EPI + 26 TROPHO	10.5* 10^6	Btgh	EPI	13	1	0	0	0	0	12
				TROPHO	13	1	0	0	2	0	10
			dBtgh	EPI	12	0	0	0	0	0	12
				TROPHO	13	0	0	0	0	0	13
Ogt^{T931A}	17	4.5* 10^6	Homozygous male from heterozygous mother		2	0	0	Not applicable	Not applicable	Not assessed	2
			WT male from heterozygous mother		8	0	4	Not applicable	Not applicable	Not assessed	4
			WT male from WT mother		7	0	0	Not applicable	Not applicable	Not assessed	7

Table S7. Staging of E7 embryos based on widefield microscopy images.

	PS	PS/ES	PS or ES	ES	advanced ES	ES or MS	MS	not determined	Tot	Pre-ES %	Post-ES %
Btgh	3	0	0	6	0	2	2	4	17	23.1	76.9
dBtgh	4	2	2	9	2	1	4	8	32	25.0	75.0

PS: Primitive Streak, ES = Early Streak, PS/ES: between PS and ES, advanced ES = advance Early Streak, MS = Mid Streak.

2) Supplementary Figures

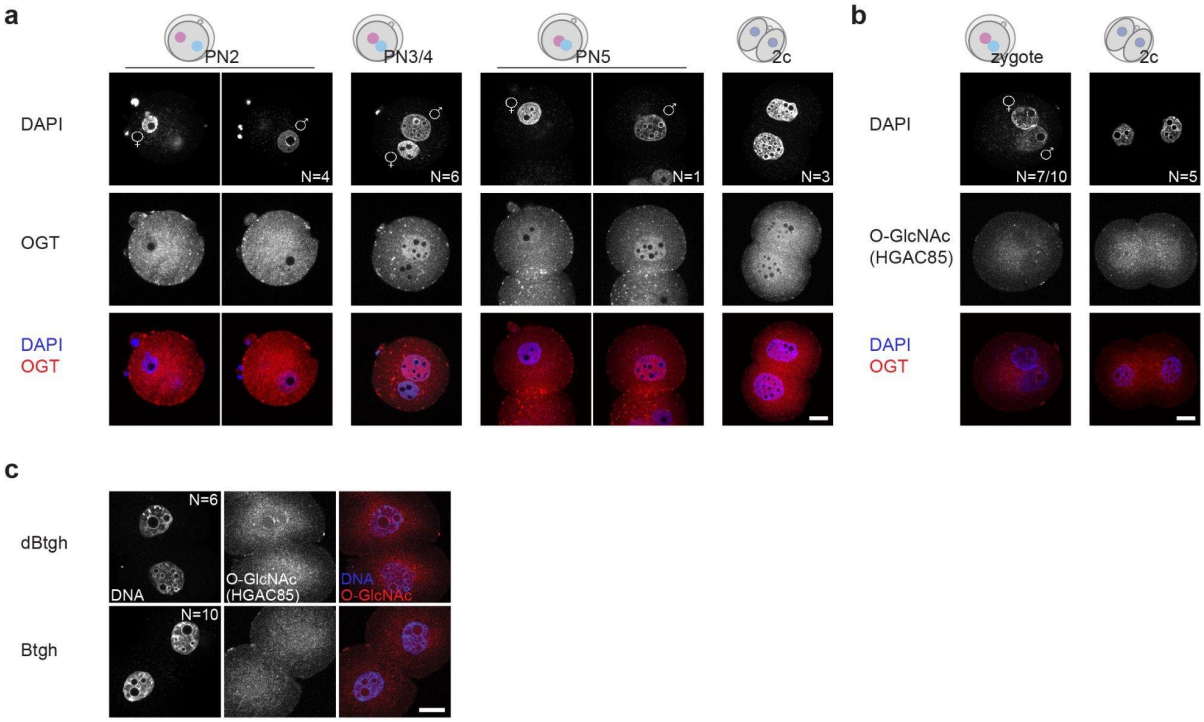


Figure S1, related to Figure 4: Profiling of the O-GlcNAc modification throughout preimplantation development.

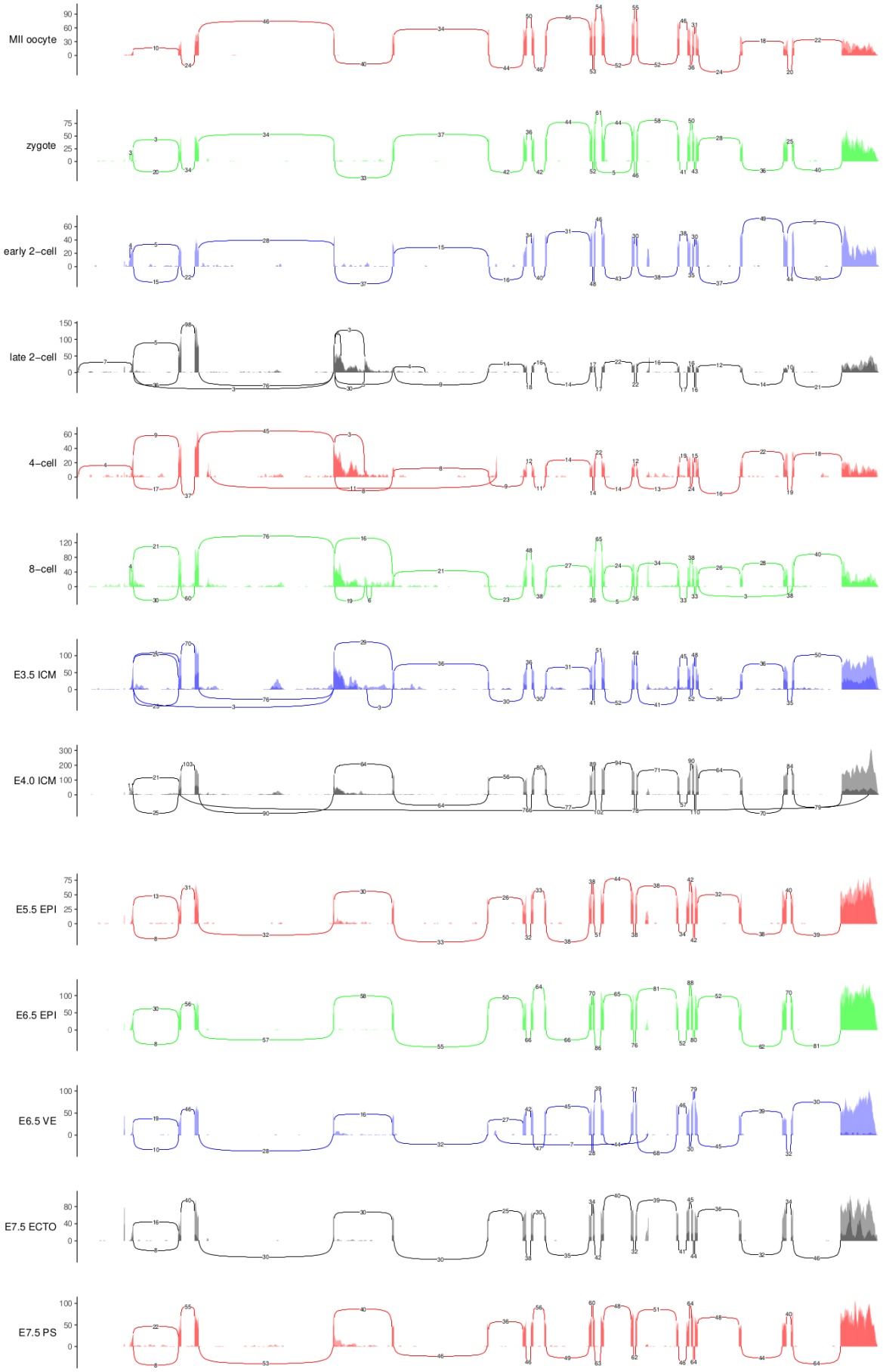
(a) Immunofluorescence staining of the OGT protein (ab177941) at different zygotic pronuclear stages and in 2-cell embryos collected after natural mating throughout the day of plug and 40 h post-hCG, respectively.

(a) Immunofluorescence staining of the O-GlcNAc modification using a different antibody than in Fig.1 (HGAC85), showing signal at perinucleolar foci. In the zygotes, perinuclear foci were present in 7 out of the 10 imaged embryos.

(c) Related to Fig.4 and Fig.8: Confocal imaging of pre-EGA 2-cell embryos from zygotes injected with Btgh/dBtgh or non-injected, stained for the O-GlcNAc modification using a different antibody than in Fig.2 (HGAC85). Both effective O-GlcNAc removal and specificity of the O-GlcNAc signal at perinucleolar foci are confirmed by signal disappearance after Btgh injection, observed in all of the 10 imaged Btgh-injected embryos.

(a-c) Embryos are mounted on coverslips and imaged using a Nikon AX scanning confocal. One z plane is shown. Scale bar indicates 20 µm. PN = pronuclear stage. DNA is stained with DAPI.

(d) (next page)


MII oocyte
zygote
early 2−cell
late 2−cell
4−cell
8−cell
E3.5 ICM
E4.0 ICM
E5.5 EPI
E6.5 EPI
E6.5 VE
E7.5 ECTO
E7.5 PS

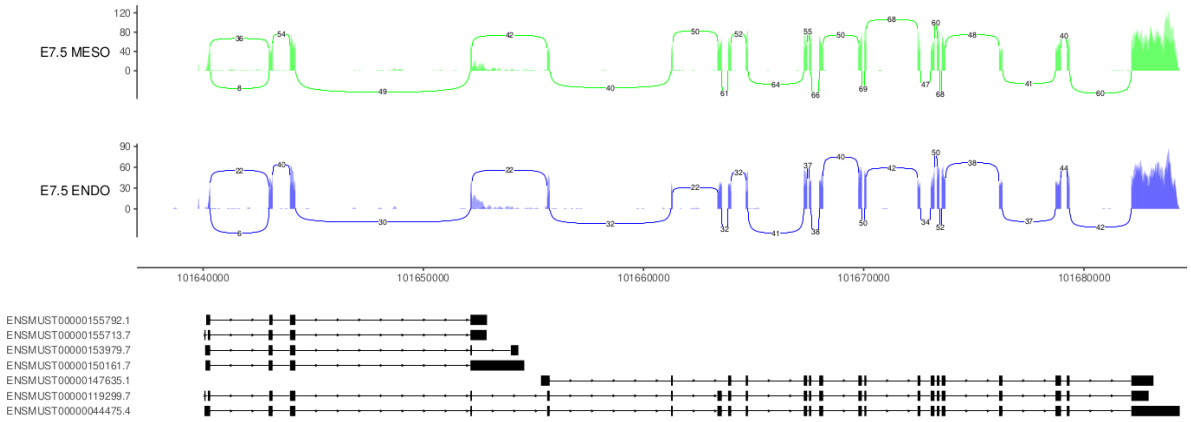


Figure S1

(d) Sashimi plot of the same datasets as in Fig.1c-d. Numbers indicate the raw number of reads spanning each junction. For each stage, biological replicates are overlapped, except for the late 2-cell for which one of the two replicates was discarded because of some high contaminant peaks.

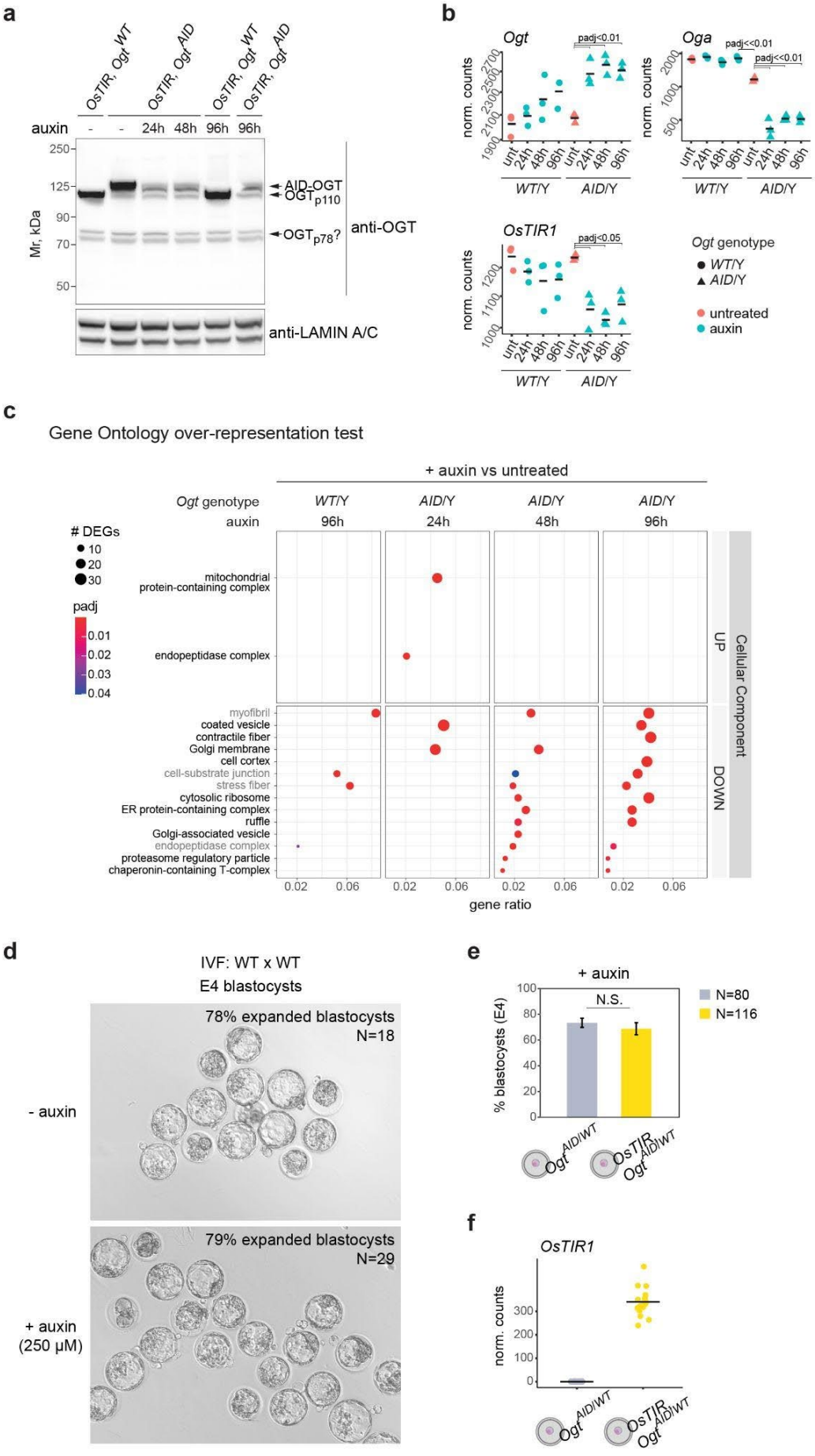

a
OsTIR, Ogt WT
OsTIR, Ogt AID
OsTIR, Ogt WT
OsTIR, Ogt AID
auxin
- - 24h 48h 96h 96h
250
125
90
70
50
Mr, kDa
AID-OGT
OGTp110
OGTp78?
anti-OGT
anti-LAMIN A/C
b
Ogt
padj<0.01
Oga
padj<<0.01
padj<<0.01
OsTIR1
padj<0.05
norm. counts
unt 24h 48h 96h unt 24h 48h 96h
WT/Y AID/Y
Ogt genotype
WT/Y
AID/Y
untreated
auxin
c
Gene Ontology over-representation test
+ auxin vs untreated
Ogt genotype WT/Y AID/Y AID/Y AID/Y
auxin 96h 24h 48h 96h
DEGs
10
20
30
padj
0.01
0.02
0.03
0.04
mitochondrial protein-containing complex
endopeptidase complex
myofibril
coated vesicle
contractile fiber
Golgi membrane
cell cortex
cell-substrate junction
stress fiber
cytosolic ribosome
ER protein-containing complex
ruffle
Golgi-associated vesicle
endopeptidase complex
proteasome regulatory particle
chaperonin-containing T-complex
UP
DOWN
Cellular Component
0.02 0.06 0.02 0.06 0.02 0.06 0.02 0.06
gene ratio
d
IVF: WT x WT
E4 blastocysts
78% expanded blastocysts
N=18
- auxin
79% expanded blastocysts
N=29
+ auxin
(250 µM)
e
+ auxin
N.S.
N=80
N=116
% blastocysts (E4)
Ogt AID/WT
OsTIR Ogt AID/WT
f
OsTIR1
norm. counts
Ogt AID/WT
OsTIR Ogt AID/WT

Figure S2, related to Figure 5-7: A novel *Ogt*^{AID} mouse allele for fast and inducible OGT depletion proves to be effective in MEFs but inefficient in preimplantation embryos.

(a) Western blot of OGT (ab177941 antibody) in male primary *OsTIR,Ogt*^{AID} MEFs and *OsTIR,Ogt*^{WT} controls, untreated or treated auxin, showing that OGT depletion is stable until day 4, the last time point tested.

(b) DESeq2-normalized counts of *Oga*, *Ogt* and *OsTIR*, showing compensatory changes in the expression of each of them after OGT depletion. padj = adj. p-value computed using DESeq2 Wald test and corrected for multiple testing using the Benjamini and Hochberg method. Mean of counts for the three replicates at each condition is shown. y-axis is in $\log_{10}$ scale.

(c) First 25 most enriched (based on p-value across all comparisons) Cellular Component GO term from the same GO over-representation analysis shown in Fig.6g. Terms are ordered by gene ratio. Gene ratio = genes belonging to the GO term / total number of up- or downregulated DEGs for that comparison. UP = upregulated DEGs, DOWN = downregulated DEGs. There are no Cellular Component GO terms enriched due to auxin effect on control clones.

(d) Test for auxin toxicity on wild type embryos. Embryos were generated through IVF and cultured *ex vivo* in the presence of auxin from the moment of fertilization to E4. Representative widefield microscopy images are shown for both conditions, but total number of starting zygotes and percentage of E4 expanded blastocysts are indicated.

(e) Percentage of E4 blastocysts obtained from IVF of *OsTIR,Ogt*^{AID} or control *Ogt*^{AID} oocytes not bearing the *OsTIR* gene with wild type sperm. Embryos of both groups were treated with auxin from the time of fertilization to E4. Barplot heights and error bars indicate the mean and standard deviation, respectively, of three replicate experiments of IVF followed by auxin treatment. Total number of starting 2-cell embryos for the two groups is stated in the legend.

(f) DESeq2-normalized gene counts for OsTIR from the same two groups of embryos in (e). y-axis is in $\log_{10}$ scale. The mean for each group is drawn.

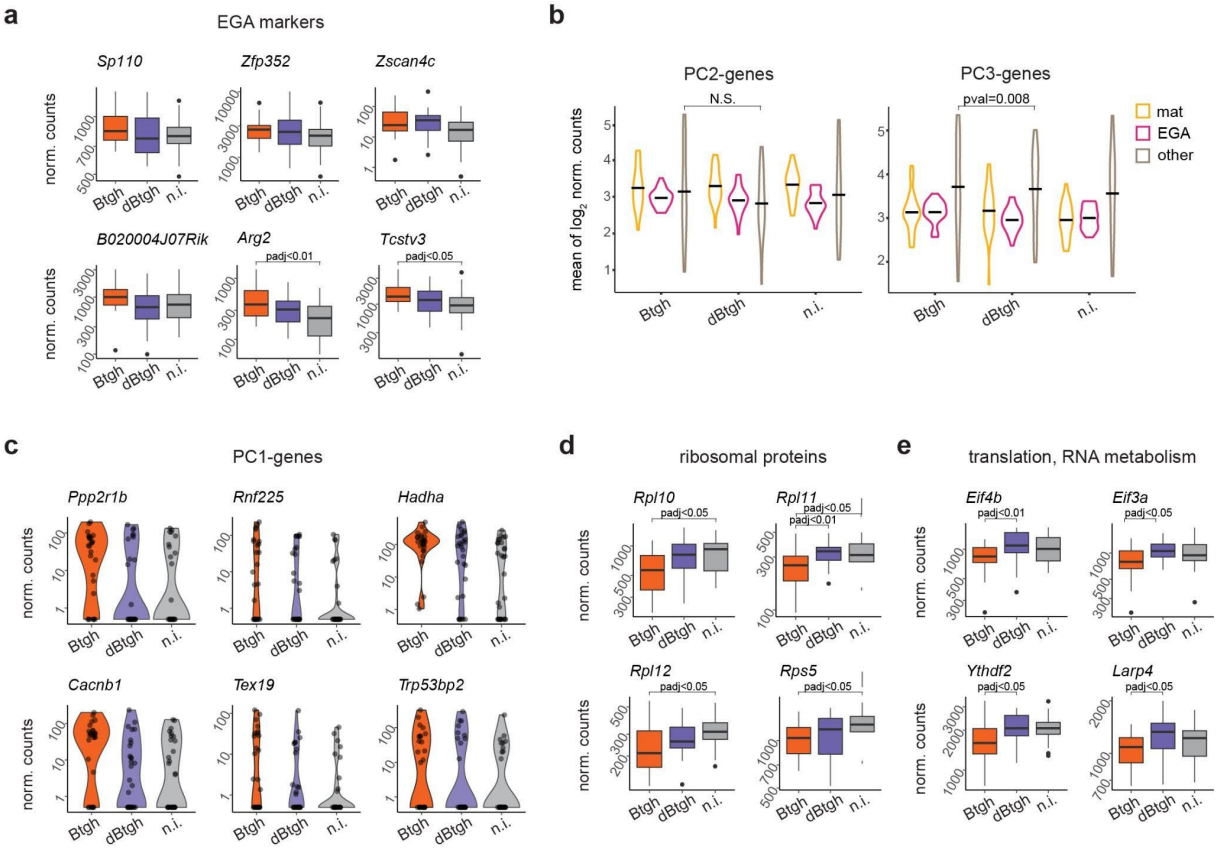


Figure S3, related to Figure 11. Transcriptional phenotype of O-GlcNAc-depleted 2-cell embryos.
(a) (Related to Fig.9) DESeq2-normalized counts of six ZGA-associated genes in the three experimental groups of 2-cell embryos.
(b) Average of DESeq2-normalized and $\log_2$ transformed expression of the 150 genes mostly associated with PC2 and PC3 of Fig.11a, in the three experimental groups of 2-cell embryos. Genes are defined as "strictly maternal" (mat) or "strictly ZGA" (ZGA) as described in the text, as "other" if not falling in these two categories. Mean for each group of embryos is marked. P-values from unpaired Wilcoxon rank sum exact test are shown if < 0.05. N.S. = not significant.
(c) DESeq2-normalized counts of the six genes mostly associated with PC1 of Fig.11a, showing that their expression is undetectable in many embryos giving two bellies to most of the violins.
(d-e) DESeq2-normalized counts of (d) four ribosomal proteins and (e) genes linked to RNA metabolism which are significantly downregulated in 2-cell embryos after nuclear O-GlcNAc depletion.
(a, d-e) padj = adj. p-value computed using DESeq2 Wald test and corrected for multiple testing using the Benjamini and Hochberg method. Boxplots hinges correspond to first and third quartiles; median is shown inside; whiskers extend to the largest and smallest values no further than 1.5 * IQR from the hinge (IQR = inter-quartile range, or distance between the first and third quartiles); data beyond the end of the whiskers are plotted individually.
(a, c-e) y-axes are in $\log_{10}$ scale.

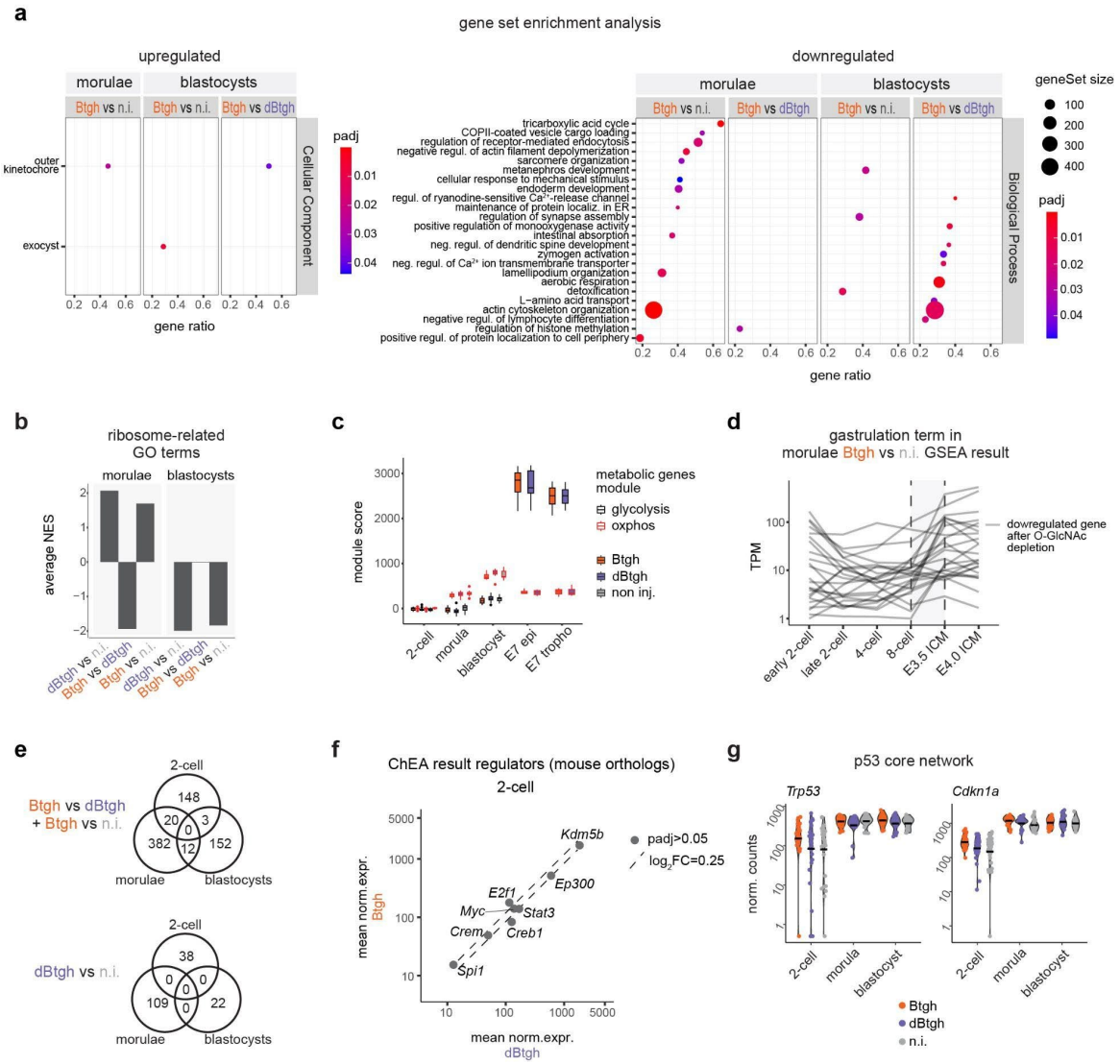


Figure S4, related to Figure 12: Transcriptional phenotype of O-GlcNAc-depleted morulae and blastocysts and coherent response across stages.

(a) Cellular Component (CC) upregulated terms and Biological Process (BP) downregulated terms from GSEA of gene expression change in nuclear O-GlcNAc-depleted morulae and blastocysts versus controls. Complementary to Fig.5c. Terms are ordered based on gene ratio. Ribosome-related terms are excluded and shown in (b). Size of dots is proportional to the number of total genes of a GO term. Gene Ratio = fraction of total genes of the GO term which are differentially expressed in the dataset.

(b) Average Normalized Enrichment Score (NES) of ribosome-related GO terms appearing in GSEA result of morulae and blastocysts.

(c) Dynamics of genes involved in the glycolytic and oxidative phosphorylation (OxPhos) process at all embryonic stages analyzed via mRNA-Seq in this study. The score of each metabolic module is computed as in (Malkowska et al., 2022) (see Methods).

(d) Expression dynamics throughout preimplantation development (mRNA-Seq data from GSE66582 (Wu et al., 2016)) of downregulated genes in O-GlcNAc-depleted morulae which belong to the enriched *gastrulation* GO term. The two biological replicates per stage were averaged. y-axis shows Transcript Per Million (TPM) and is in $\log_{10}$ scale.

(e) Overlap among significantly differentially expressed genes (DEGs) (adj.pval<0.05, any log2FC) with mean DESeq2-normalized counts ≥ 10 in Btgh-injected embryos versus any of the control groups (left) and in dBtgh-injected versus non-injected embryos (right) at the different preimplantation stages.
(f) Gene expression change in Btgh vs dBtgh-injected embryos of the mouse orthologs of potential regulators appearing in ChEA results in Fig.12f. None of them has adj. p-value < 0.05. Axes are in $\log_{10}$ scale.
(g) DESeq-2 normalized gene counts for the genes encoding transcription factor p53 (*Trp53*) and its genomic target p21 (*Cdkn1a*) in the three conditions at the three developmental stages. y-axes are in $\log_{10}$ scale.

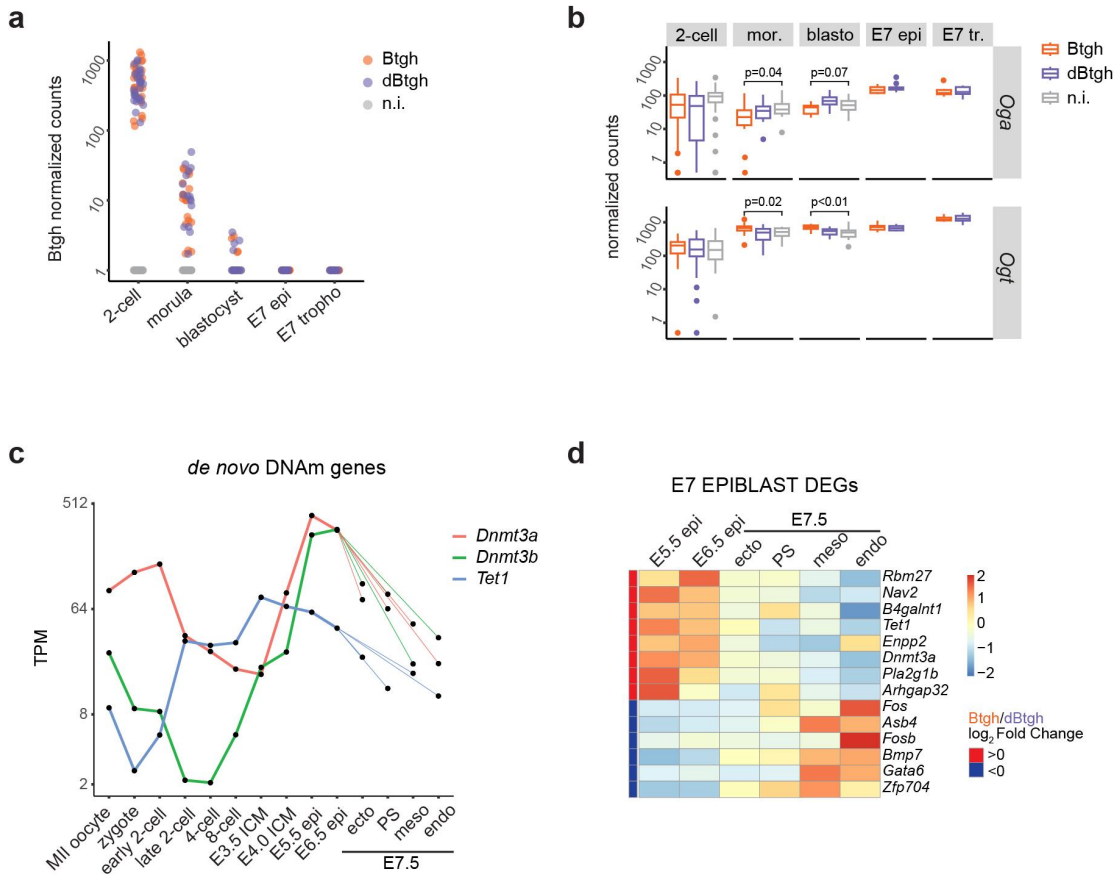


Figure S5, related to Figure 13: Loss of nuclear O-GlcNAc preimplantation causes postimplantation developmental delay.

(a) DESeq2-normalized counts of the Btgh/dBtgh constructs at all embryonic stages analyzed in this study. y-axis is in $\log_{10}$ scale.

(b) DESeq2-normalized counts of *Oga* and *Ogt* at all embryonic stages in this study, showing compensatory downregulation of *Oga* and upregulation of *Ogt* in morulae and blastocysts following nuclear O-GlcNAc depletion. p = adj. p-value computed using DESeq2 Wald test and corrected for multiple testing using the Benjamini and Hochberg method. Boxplots hinges correspond to first and third quartiles; median is shown inside; whiskers extend to the largest and smallest values no further than 1.5 * IQR from the hinge; data beyond the end of the whiskers are plotted individually. y-axis is in $\log_{10}$ scale.

(c) Expression of *de novo* DNA methyltransferases *Dnmt3a* and *Dnmt3b* and active DNA demethylase *Tet1* throughout mouse embryonic development (mRNA-Seq data from GSE66582 (Wu et al., 2016) and GSE76505 (Zhang et al., 2017)). The two biological replicates per stage were averaged. y-axis shows Transcript Per Million (TPM) and is in $\log_2$ scale.

(d) Heatmap showing the expression of all DEGs in Btgh-injected epiblasts (excluding three pseudogenes) (labeled in Fig.13d) at the transition between E5.5 and E7.5 (mRNA-Seq data from GSE76505 (Zhang et al., 2017)). TPM values were $\log_2$-transformed and scaled by rows.

(c-d) epi = epiblast, ecto = ectoderm, PS = primitive streak, meso = mesoderm, endo = endoderm.

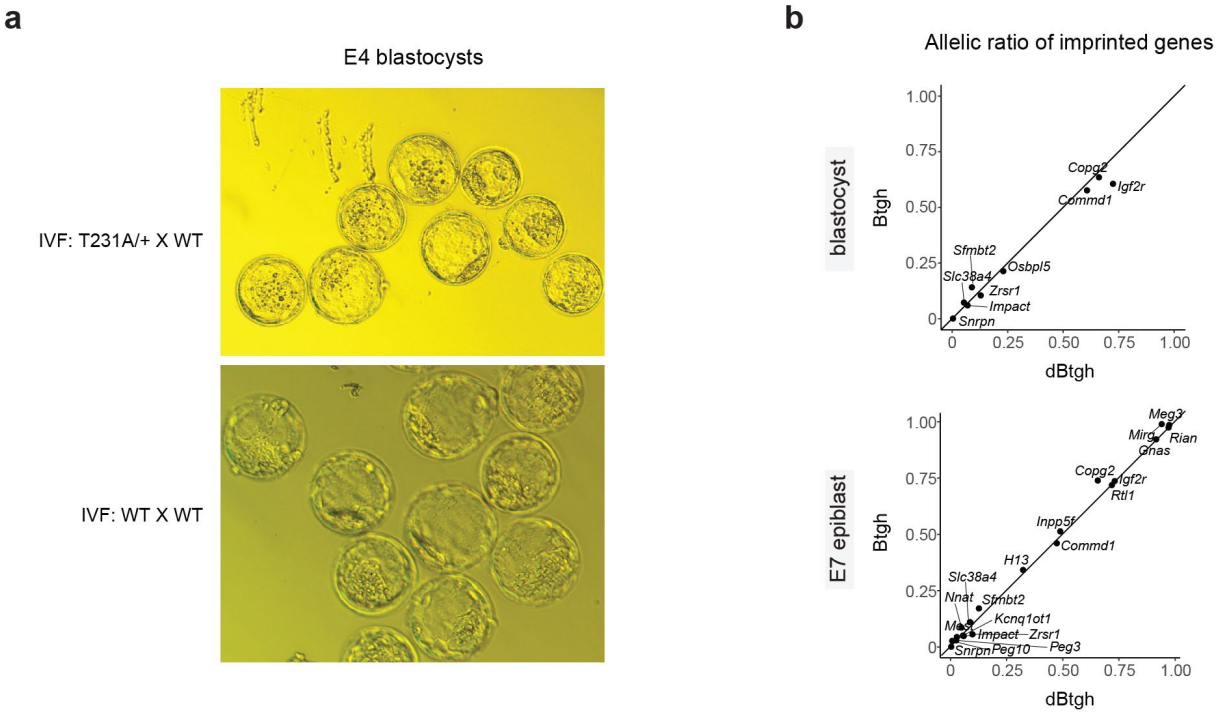


Figure S6, related to Figure 15: Maternal effect of the hypomorphic mutant *Ogt*^{T931A}.
(a) Representative widefield microscopy images taken during collection for Smart-Seq of the blastocysts from the IVF of oocytes from *Ogt*$^{T931A/WT}$ or WT sibling mothers with the same wild type sperm.
(b) Ratio of maternal over paternal reads (allelic ratio) in blastocysts and E7 embryos developed from zygotes injected with Btgh or dBtgh, which shows the absence of changes in the allelic expression of known imprinted genes after O-GlcNAc depletion. Only imprinted genes with enough coverage on SNPs are included in each plot (see Methods). As expected, almost all imprinted genes (except *Inpp5f* and *Commd1* in epiblasts) show a skewed allelic ratio towards 0 or 1, indicating exclusive or preferential expression from the paternal or maternal allele, respectively. Differential allelic expression after O-GlcNAc depletion has been tested using DESeq2 (see Methods) and significant changes were not found for any gene with enough reads coverage (including imprinted genes).