From Jumping Genes to Domesticated Elements: The Human PGBD Family as a Model for DNA Transposon Evolution

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

Submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Kathrin Radscheit, M.Sc

2023

This work was prepared from May 2018 to May 2023 under the supervision of Dr. Zsuzsanna Izsvák (Max-Delbrück Center for Molecular Medicine, Berlin, Germany). All presented experiments and computational analyses were conducted by myself unless indicated otherwise.

1st Reviewer:

Dr. Zsuzsana Izsvák

Max-Delbrück Center for Molecular Medicine Berlin, Germany

2nd Reviewer:

Prof. Dr. Katja Nowick

Institute for Biology and Zoology,

Freie Universität, Berlin, Germany

date of defense: 13.02.2024

Table of contents

Table of figures

List of tables

Table of abbreviations

Significance Level

ns.	Not significant, p-value > 0.1
.	A tendency, p-value < 0.1
*	Significant, p-value < 0.05
**	Significant, p-value < 0.01
***	Significant, p-value < 0.001

Gene Symbols

ARC	Activity-regulated cytoskeleton-associated protein
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3 related
ATRX	Alpha thalassemia/mental retardation syndrome X-linked
CBP	CREB-binding protein
CENP-B	Centromere protein B
CoREST	REST corepressor 1
CREB	CAMP-response element binding protein
CTCF	CCCTC-binding factor
DDB1	Damage-specific DNA binding protein 1
DNA-PK	DNA-dependent protein kinase, catalytic subunit
DUSP1	Dual specificity protein phosphatase 1
EGR1/2	Early growth factor 1/2
Eno2	Enolase 2
ERCC6	Excision repair cross-complementation group 6
ERα	Estrogen receptor α
FBXL19	F-box and leucine-rich repeat protein 19

FOS	Fos proto-oncogene, AP-1 transcription factor subunit
HPRT1	Hypoxanthine phosphoribosyltransferase 1
IS4	Insertion element 4
JUNB	Jun-B proto-oncogene, AP-1 transcription factor subunit
MAPK	Mitogen-activated protein kinase
MYB	Proto-oncogene myeloblastosis
NEUROG2	Neurogenin 2
NPAS2/4	Neuronal PAS domain protein 2/4
NR4A1	Nuclear receptor subfamily 4 group A member 1
PARP1	Poly [ADP-ribose] polymerase 1
PBLE	piggyBac-like elements, potentially active
PGBD	PiggyBac transposable element derived
PGM	piggyMac
piggyBac	Ti. Ni piggyBac
PIP5K1A/C	Phosphatidylinositol-4-phosphate 5-kinase type 1 α/ γ
RAD18	RAD18 E3 ubiquitin protein ligase
RAG1/2	Recombination activating gene 1/2
Rbfox3	RNA binding fox-1 homolog 3
REST	RE1 silencing transcription factor
SIN3A	SIN3 Transcription Regulator Family Member A
SIRT1	Sirtuin-1
SL1 complex	Selective factor 1 complex
SRF	Serum response factor
STAT3	Signal transducer and activator of transcription 3
Stoml3	Stromatin like 3
TAF1A/B/C	TATA-box binding protein associated factor, RNA polymerase I subunit A/B/C
Tmem181b-ps	Pseudogene transmembrane protein 181b
TOP1(MT)	Topoisomerase I/ Mitochondrial
TOP2A/B	Topoisomerase II α/ β

TOP3A/B	Topoisomerase III α/ β
TP53BP1	Tumor protein p53 binding protein 1
TPB2	Tetrahymena piggyBac-like 2
TRIM28	Tripartite motif containing 28
Tshz2	Teashirt zinc finger homeobox 2
UBTF	Upstream binding transcription factor
XRCC5/6	X-ray repair cross-complementing 5/6
Zfp366	Zinc finger protein 366

Units

bp	Base pairs
cm	Centimeter
Da/kDa	Dalton / kilo
g/mg/µg/ng	Gram/ milli / micro / nano
kb	Kilobases
l/ml/µl	Liters / milli / micro
M/mM/µM	Molar / milli / micro
min	Minutes
rpm	Revolutions per minute
sec/msec	Seconds / milli
U	Units
V	Volt

Others

AP	Affinity purification
ATP	Adenosine triphosphate
AUC	Area under curve
BP	Biological process
CC	Cellular component
CDS	Coding sequence

ChIP	Chromatin immunoprecipitation
CRD	Cystine rich domain
CRISPR/Cas	Clustered regularly interspaced short palindromic repeats/ CRISPR-associated protein
Ct	Cycle threshold
CTR	C-terminal region
DDBD	DNA-binding and dimerization domain
DEG	Differentially expressed genes
DNA	Deoxyribonucleic acid
DPC	DNA-protein crosslinks
DSB	Double-strand break
dsDNA	Double stranded DNA
EDTA	Ethylenediaminetetraacetic acid
e.g.	Latin: exempli gratia, for example
EMSA	Electrophoretic mobility shift assay
ESCs	Embryonic stem cells
et al.	Latin: et alia, and others
FDR	False discovery rate
GFP	Green Fluorescent Protein
GO	Gene ontology
H1	Specific embryonic stem cell line derived from the inner cell mass of a human blastocyst
HA-tag	Human influenza hemagglutinin -tag
HEK293	Human embryonic kidney 293 cells
ICC	Immunocytochemistry
IEG	Immediate early gene
iNeurons	Inducible Neurons
IP	Immunoprecipitation
iPSCs	Induced pluripotent stem cells
kDNA	Kinetoplast DNA

KO	Knockout
lncRNA	Long non-coding RNA
MCF-7	Cell line, Michigan Cancer Foundation-7
MF	Molecular function
MITES	Miniature inverted-repeat transposable elements
mRNA	Messenger RNA
MS	Mass spectrometry
NCBI	National Center for Biotechnology Information
NGS	Next generation sequencing
NHEJ	Non-homologous end-joining
NLS	Nuclear localization signal
NMD	Non-sense mediated decay
NMDA	N-methyl-D-aspartate
NPCs	Neural progenitor cells
NTD	N-terminal domain
PCA	Principal component analysis
PCR	Polymerase chain reaction
PSS	PGBD5 signal sequence
PTC	Premature translation-termination codon
RADAR	Rapid approach to DNA adduct recovery
RLE	Relative log expression
rlogs	Regularized-logarithm transformation
RNA	Ribonucleic acid
RPE-1	Retinal pigment epithelial 1
RPKM	Reads per kilobase of transcript per million mapped reads
rRNA	Ribosomal RNA
RT-PCR	Real-time polymerase chain reaction
SILAC	Stable isotope labeling of amino acids in cell culture
SNP	Single-nucleotide polymorphism
Ti. Ni.	Trichoplusia ni

TIR	Terminal inverted repeat
TMM	Trimmed mean of M-values
TOPcc	Topoisomerase cleavage complex
TPM	Transcript per million
TSS	Transcription start site
UCSC	University of California, Santa Cruz
VP-16	Etoposide
WT	Wild type

Summary

Our lab focuses on studying mobile genetic elements and their contribution to genome evolution. The remnants of mobile genetic elements can be repurposed by the host. For example, they serve as enhancers, alternative poly(A) signals, splice sites, exons, or entirely new genes. This study investigates the domestication of DNA transposons to form entirely new genes.

PiggyBac-derived elements (PGBDs) are transposase-like genes that have been repeatedly domesticated into vertebrate genomes, suggesting that they may have important functions in host biology. The human genome comprises five PGBD genes, PGBD1-5. However, little is known about the molecular mechanisms underlying their domestication or their physiological roles. Therefore, we investigated the evolutionary history and possible molecular functions of the human PGBD family, especially PGBD1 and PGBD5, as an example of host transposase domestication events.

First, we investigated the domestication event of PGBD1 and found that it had captured an N-terminal SCAN and KRAB domain in its ancestral condition. However, while the transposase domain was under strong evolutionary constraints, the *N-terminal domains* (NTDs) were under lesser purifying selection. In the KRAB domain that showed K_a/K_s ratios bigger than 1, amino acid substitution may have been beneficial for PGBD1 function, including 252M, which might have compromised *Tripartite motif containing 28* (TRIM28) binding capacity. Additional losses and decays of the N-terminal domain of PGBD1 highlight the functional dominance of the domesticated transposase domain. Hence, host domains might facilitate the capture of DNA transposases to the host genome, but their function could be submissive to the protein function.

In the second part, we investigated the molecular functions of PGBD5 and found that it is doubtfully an active transposase, and we identified that it interacts with another nuclease, topoisomerase II α (TOP2A). We show that PGBD5 is enhancing TOP2A function *in-vitro* and that it, together with topoisomerase II β (TOP2B), might regulate the transcription of a subset of

immediate early genes (IEGs), namely FOS, NPAS4, NR4A1 and DUSP1 in the brain. Further, we show that PGBD5 interacts with several proteins involved in histone modifications, transcriptional regulation, and DNA repair. Altogether we provide an alternative mechanism by which PGBD5 contributes to DNA *double-strand breaks* (DSBs) and genomic rearrangements. Additionally, we show that PGBD5 adds an extra layer of IEG control that is specific to the brain and thereby contributes to brain-specific signaling.

The thesis contributes to the knowledge of the molecular mechanisms underlying the domestication of transposase-like genes and their physiological roles in the host genome.

Zusammfassung

Die Forschung unseres Labors konzentriert sich auf die Untersuchung mobiler genetischer Elemente und deren Beitrag zur Genomevolution. Rudimente mobiler genetischer Elemente können vom Wirt weiterverwendet werden, um neuen Aufgaben zu dienen. Diesen Prozess nennt man Domestizierung. Sie dienen zum Beispiel als Enhancer, alternative Poly(A)-Signale, Spleißstellen, Exons oder bilden selbstständige neue Gene. In dieser Arbeit wird die Weiterverwendung von DNA-Transposons zur Bildung völlig neuer Gene untersucht.

PiggyBac-derived elements (PGBDs) sind eine Gruppe von Genen die von der *piggyBac* Transposase abstammen. Diese Sequenzen wurden wiederholt in Wirbeltiergenomen domestiziert, was darauf hindeutet, dass sie wichtige Funktionen in der Wirtsbiologie haben. Das menschliche Genom umfasst fünf PGBD-Gene, PGBD1-5. Über die molekularen Mechanismen, die ihrer Domestizierung zugrunde liegen, und ihre physiologischen Funktionen ist jedoch wenig bekannt. Daher untersucht die vorliegende Arbeit die Evolutionsgeschichte und mögliche molekulare Funktionen der menschlichen PGBD-Familie, insbesondere von PGBD1 und PGBD5, als Beispiel für die Domestizierung von Wirtstransposasen.

Im ersten Teil wurde das Domestikationsereignis von PGBD1 untersucht und festgestellt, dass es in seinem Urzustand an eine SCAN- und KRAB-Domäne fusionierte. Während die Transposase-Domäne jedoch starken evolutionären Zwängen unterlag, waren die N-terminalen Domänen einer geringeren reinigenden Selektion ausgesetzt. In der KRAB-Domäne, die ein K_a/K_s -Verhältnis von mehr als 1 aufwies, könnte die Substitution von Aminosäuren für die Funktion von PGBD1 von Vorteil gewesen sein, einschließlich 252M, die die TRIM28-Bindungskapazität beeinträchtigt haben könnte. Zusätzliche Verluste und Veränderungen der N-terminalen Domäne von PGBD1 unterstreichen die funktionelle Dominanz der domestizierten

Transposasedomäne. Wirtsdomänen könnten also die Integration von DNA-Transposasen in das Wirtsgenom erleichtern, aber ihre Funktion könnte für die Proteinfunktion untergeordnet sein.

Im zweiten Teil untersuchten wir die molekularen Funktionen von PGBD5. Zunächst stellten wir fest, dass es sich wahrscheinlich nicht um eine aktive Transposase oder Nuklease handelt. Da es aber mehrfach mit vermehrten doppelsträngigen DNA brüchen assoziiert wurde, untersuchten wir die Protein-Protein Interaktionen von PGBD5 und identifizierten, dass es mit Topoisomerase II α, interagiert. Wir zeigen, dass PGBD5 die Funktion von Topoisomerase II α *in-vitro* verstärkt und dass es zusammen mit Topoisomerase II β die Transkription einer Untergruppe von *Immediate Early Genes* (IEGs), nämlich FOS, NPAS4, NR4A1 und DUSP1, im Gehirn regulieren könnte. Außerdem zeigen wir, dass PGBD5 mit mehreren Proteinen interagiert, die an Histonmodifikationen, der Transkriptionsregulation und der DNA-Reparatur beteiligt sind. Insgesamt liefern wir einen alternativen Mechanismus, durch den PGBD5 zu DNA Doppelstrangbrüchen und Gen-Rearrangements beiträgt. Darüber hinaus zeigen wir, dass PGBD5 zu einer zusätzlichne Kontrollebene zur Regulierung der IEG beiträgt, die spezifisch für das Gehirn ist.

Die Arbeit leistet einen Beitrag zum Wissen über die molekularen Mechanismen, die der Domestizierung von Transposase-Ähnlichen Genen und ihrer physiologischen Rolle im Wirtsgenom zugrunde liegen.

Acknowledgments

I want to thank Dr. Zsuzsanna Izsvak and Prof. Dr. Katja Nowick for reviewing my dissertation. Their support and guidance have been crucial during my doctoral research, and I am genuinely thankful for their presence.

I thank Zsuzusanna Izsvak for warmly welcoming me into her lab and for her trust, interest, and compassion throughout our scientific collaboration at the Max-Delbrück-Center for Molecular Medicine. I would also like to thank Dr. Tamas Rasko for introducing me to laboratory work and teaching me the essential molecular biology methods. Together with Dr. Artilla Szvetnik, they introduced me to the fascinating field of domesticated transposases. Additionally, I would like to thank Dr. Manvendra Singh for keeping me excited about the open questions in the area and mentoring me with bioinformatic guidance.

I would like to acknowledge Prof. Dr. Laurence Hurst, who has inspired me by sharing his perspective on biological questions. His mentorship during the evolutionary analysis of my work has been invaluable.

I would also like to express my appreciation to the members of the Stem-cell Facility, including Dr. Sebastian Diecke, Dr. Silke Frahm-Barske, Dr. Narasimha Swamy Telugu, Sandra Schommer, and Vivian Schulz, for their training and technical support in working with human iPSCs. Furthermore, I am grateful to the Proteomics Facility, particularly Prof. Dr. Matthias Selbach and Christian Sommer, for conducting the mass spectrometry analysis of the PGBD5 interactome and providing insights into proteome analysis. The Advanced Light Microscopy Facility, with Dr. Anje Sporbert and Dr. Anca Margineanu, deserves my thanks for introducing me to confocal microscopy and granting me access to their equipment. I am also grateful to Alex Kentsis for sharing RNA-seq data with me.

I want to express my heartfelt gratitude to all the lab members, past and present, of the Mobile Element Lab, especially Dr. Aleksandra Kondrashkina, Bertrand Tangu Teneng, Chigozie Joseph

Samuel, Franziska Baar, Beatrice Heuser, Sandra Neuendorf, Dr. Zhimin Zhou, Dr. Katarina Stevanovic, Dr. Yuliang Qu, and Felix Lundberg. Each one of you helped me to grow in this time. Our discussions, method-sharing, and mutual support, particularly during the weekend shifts, have been so important to me. We have also shared many memorable moments during lunch breaks, beer hours, hat-making events, BBQs, and bicycle trips to Liepnitzsee. Some of you have become close friends, and I am excited to continue following your journeys.

I am also deeply thankful to my beautiful family and friends outside the lab, especially my son, Leonard Ivanov, and my husband, Andranik Ivanov, who supported me in many ways. They have brought me immense joy and helped me take a break from work, even during the countless weekends I spent in the lab. I also thank my amazing parents, Thomas and Christina Radscheit, who devotedly supported me all my life.

1 Introduction

How do protein domains of host-captured transposases adapt after the domestication event? And how do transposase-derived genes serve the host organism?

DNA Transposons shaped and diversified genomes in multiple ways. One way in which they can change the genome is through domestication events. In this process, genetic material derived from transposons integrates into the host genome and is utilized by the organism to serve host functions. The human PGBD family is a product of such domestication events. The human PGBD gene family comprises five genes, PGBD1-5. The thesis investigates two members of the family: PGBD1 and PGBD5. The first part concentrates on the protein domain changes upon domestication of PGBD1. The second part explores the possible host function of PGBD5 as an example of a domesticated DNA transposon.

1.1 DNA transposons in eukaryotes

DNA transposons, also known as "class II transposable elements," are segments of DNA that can move within a genome from one position to another. They are called DNA transposons because they transpose via DNA intermediates, in contrast to retrotransposons, which transpose via RNA intermediates (Feschotte & Pritham, 2007). The majority of eukaryotic DNA transposons move by encoding a transposase enzyme that recognizes specific sequences on the ends of the transposon, cuts the DNA, and then integrates the transposon into a new location in the genome. This process is called the *cut-and-paste* mechanism. However, two additional transposition mechanisms have been identified. Prokaryotes display further DNA transposition mechanisms, as reviewed by (Hickman & Dyda, 2015).

Helitrons are DNA transposons that transpose via a rolling-circle mechanism (Kapitonov & Jurka, 2001). Mavericks encode several proteins and may replicate using a self-encoded DNA polymerase (Pritham *et al.*, 2007). Because both of these mechanisms rely on single-stranded DNA intermediates, they are thought to transpose by a replicative copy-and-paste mechanism (Feschotte & Pritham, 2007; Kapitonov & Jurka, 2001).

Well-known examples of *cut-and-paste* DNA transposon superfamilies are Tc1/mariner, MuDR/Foldback, hAT, piggyBac, PIF, Merlin, CACTA, P element, Transib and Banshee (Feschotte & Pritham, 2007). Other DNA transposon superfamilies are Helitron and Mevricks (Feschotte & Pritham, 2007). Sources vary about the number of DNA transposon superfamilies ranging from ten to 23 (Feschotte & Pritham, 2007; Kojima, 2018). Because the advances in sequencing technologies, sequence annotation algorithms, and the increasing availability of sequenced genomes led to the discovery of new families or the merge of known superfamilies (Feschotte & Pritham, 2007).

The following paragraph describes the classical *cut-and-paste* transposition mechanism in more detail.

1.1.1 The DDE/D family of "cut-and-paste" transposons

Nearly all characterized DNA *cut-and-paste* transposons belong to the family of DDE/DDD recombinases (Yuan & Wessler, 2011). The motif consists of two conserved aspartic (D) residues and one glutamic (E) or another D. The amino acid triad coordinates divalent metal ions such as Mg^{2+}, which catalyze transposition. They form an RNaseH-like fold that brings the catalytic triad in proximity. One exception are Cryptons, these *cut-and-paste* DNA transposons encode a tyrosine recombinase which probably facilitates recombination of a circular intermediate and the DNA target (Carducci *et al.*, 2020). The following paragraph describes DDE/DDD transposition illustrated in Figure 1.

DNA *cut-and-paste* transposons are flanked by *terminal inverted repeats* (TIRs), which hold the recognition sequence of the transposase enzyme. The transposase enzyme binds to the TIRs at the ends of the transposon and assembles them into a synaptic protein-DNA complex, which is required to perform all catalytic steps (Montaño & Rice, 2011). The synaptic complex formation involves both transposon ends, an oligomer of the transposase, and sometimes accessory proteins and host factors (Montaño & Rice, 2011). The synaptic complex formation and the downstream chemical steps somewhat vary between the transposase families (Montaño & Rice, 2011). However, the mechanisms share some key characteristics (Hickman & Dyda, 2016):

The first step of the *cut-and-paste* transposition mechanism is the cleavage of the DNA at the ends of the transposable element by the synaptic complex, generating single-stranded overhangs. The transposable element is then released from its original location in the genome, leaving behind a DSB.

The transposase enzyme recognizes specific DNA sequences at the target site, where it will insert the transposable element. In order to insert the transposable element, the transposase cuts the DNA at these target sites, leaving behind single-stranded DNA ends. The transposable element then binds to the single-stranded DNA ends, and the transposase enzyme catalyzes the joining of the transposable element to the DNA at the target site.

This joining process often leads to a duplication of the DNA sequence immediately flanking the target site, as the single-stranded ends of the DNA are mainly used as templates to repair the DSB created by the transposase. The duplicated DNA sequence that is generated by this process is called a *target site duplication*.

In brief, the *cut-and-paste* transposition mechanism involves the recognition of DNA sequences at transposon ends and the formation of the synaptic complex that includes both transposon ends. The synaptic complex formation is essential for the transposase enzyme to perform all catalytic steps. It is followed by the cleavage of DNA sequences by the transposase enzyme, excision of the transposable element from its original location, integration into a new location in the genome, and repair of the DSBs generated during the process.

Transposases have the ability to excise and reintegrate any sequence located between two TIRs within a certain proximity (Feschotte & Pritham, 2007). This makes them a powerful molecular tool for transgenesis but also poses certain risks, such as chromosomal rearrangements described later.

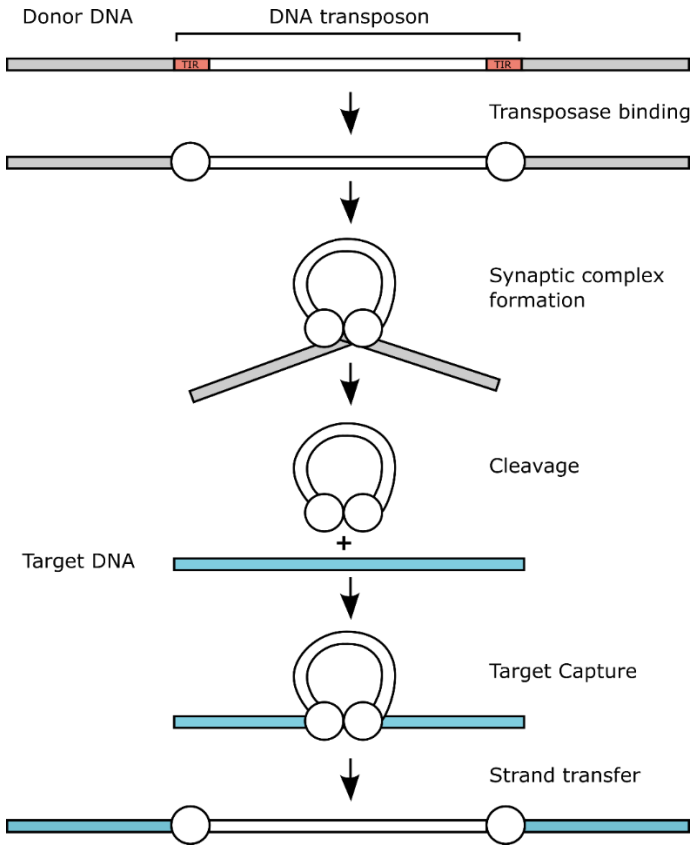


Figure 1: Transposition mechanism of DDE/D DNA *cut-and-paste* transposons. The transposase enzymes (white circles) bind to the transposon DNA ends called TIRs (red). Looping of the transposon DNA brings its ends together, forming a synaptic complex. The transposase then cuts the transposon DNA from the donor DNA (gray). The transposase/DNA complex moves and binds to the target DNA (blue). Through strand transfer, the transposase inserts the transposon DNA into the target DNA, completing the transposition process.

In addition to autonomous DNA transposons, there are also non-autonomous DNA transposons. Non-autonomous DNA *cut-and-paste* transposons are known as *Miniature Inverted-repeat Transposable Elements* (MITEs). These elements are typically 100-600 bp long, contain TIRs

and two flanking target site duplications (Feschotte & Pritham, 2007). MITEs exhibit length homogeneity and are often found in high copy numbers (Feschotte & Pritham, 2007).

Because recognizing terminal sequences is usually the only requirement for DNA transposons to undergo transposition, autonomous DNA transposons can also mobilize internally deleted or rearranged non-autonomous elements. They likely arose from only a few, or even a single, progenitor copies (Feschotte & Pritham, 2007). The deletion of a larger transposon during gap repair or the proximity of two single TIRs flanking host sequences may have been the source of progenitor copies (Feschotte & Pritham, 2007).

It was hypothesized that the propagation of MITEs is facilitated by their ability to evade the host-defense mechanism, as their sequences often lack homology to their associated autonomous transposases (Feschotte & Pritham, 2007).

DNA cut-and-paste transposons are non-replicative; however, they increase their copy number by indirect mechanisms utilizing the host machinery (Feschotte & Pritham, 2007). One way is the transposition during DNA replication from an already replicated site to an unreplicated site. The transposon is replicated twice, resulting in an additional copy (Feschotte & Pritham, 2007; Kunze & Weil, 2007). The second mechanism relies on homologous DNA repair of the DSBs left after excision of the transposon. Is the transposon present on the homologous chromosome, homologous DNA repair might reintroduce the transposon into the donor site (Engels *et al.*, 1990; Feschotte & Pritham, 2007). During S-phase, the homologous sequence can also originate from the sister chromatid (Engels *et al.*, 1990; Feschotte & Pritham, 2007).

1.1.2 How transposons shaped the eukaryotic genome

DNA transposons greatly impacted genome evolution (Feschotte & Pritham, 2007). They shaped and diversified genomes in several ways:

1. Transposons may **disrupt functional genes** by inserting into them or their regulatory regions, potentially causing mutations that alter gene expression or function.

2. They can give rise to new genes or regulatory sequences. Several transposons have been domesticated to serve host functions. The **domestication** event sometimes involve the capture by host protein domains such as the KRAB domain (Cosby *et al.*, 2021). The widespread MITEs, which are non-autonomous, can give rise to miRNAs that regulate cellular pathways (Piriyapongsa *et al.*, 2012).
3. DNA transposons might also induce **epigenetic** changes, especially local heterochromatin formation, due to their repetitive nature and TIR structure.
4. Occasionally, DNA transposons induce **large-scale genomic rearrangements** (Collins & Rubin, 1984; Y. H. M. Gray, 2000). Those include chromosomal inversions, duplications, and deletions of over 100 kb. These rearrangements occur due to DNA *double-strand breaks* (DSBs), which are a part of the transposition mechanism. Two nearby TIRs of separate transposon copies are recognized by the transposase and form a synaptic complex with complete or partial transposition, leading to various outcomes.

1.1.3 Domestication by host transposase capture

DNA transposons have been widely repurposed by hosts for various functions, as evidenced by well-known examples such as *Recombination activating gene 1 and 2* (RAG1 and RAG2) or *Centromere Protein B* (CENP-B). RAG1/2 plays a crucial role in the process of V(D)J recombination, which allows the adaptive immune system to recognize an almost infinite number of antigens and protect against pathogenic microorganisms (Matthews & Oettinger, 2009). Through the combinatorial joining of gene segments, RAG1/2 recombinase creates a diverse array of T-cell receptors and immunoglobulins, enhancing the immune response to microorganisms (Matthews & Oettinger, 2009). CENP-B, on the other hand, is a centromeric protein that binds *sequence-specific* to alpha satellite DNA, making it the only protein of its kind identified to date (Gamba & Fachinetti, 2020). Its role in centromere function is essential for proper chromosome segregation during cell division (Gamba & Fachinetti, 2020). These examples illustrate the remarkable ability of DNA transposons to be co-opted by hosts for crucial biological functions, highlighting the importance of transposons for evolutionary innovation.

Domesticated transposons are characterized by their evolution under functional constraints, the presence of intact orthologs in syntenic regions across different species, evidence of transcription, evidence of biological host function, and often the absence of functional TIRs (Feschotte & Pritham, 2007). In contrast, active transposable elements typically evolve under neutral selection, and their transcription is suppressed by the host (Feschotte & Pritham, 2007).

DNA transposons can be domesticated by fusing the transposase domain with a pre-existing host domain (Cosby *et al.*, 2021). These host domains, including KRAB, SET, and SCAN, are often involved in regulating transcription and organizing chromatin. Among these domains, KRAB is the most frequently observed in the N-terminal region of domesticated transposons, occurring in about one-third of all detectable events (106 independent host-transposase fusions in tetrapods). In most cases, the host-captured transposases retain their DNA binding domain, accounting for approximately 77% of cases (Cosby *et al.*, 2021). Consequently, the fusion of host and transposase domains is hypothesized to generate novel transcriptional regulators (Cosby *et al.*, 2021).

It remains unclear whether the KRAB domain possesses features that promote its capture by transposases (Cosby *et al.*, 2021). KRAB is a ubiquitous domain in tetrapods, with the largest transcription factor family, KRAB-zinc fingers, comprising 487 elements in humans (Cosby *et al.*, 2021). This high occurrence of KRAB domains in host-transposase fusion events could be attributed to their abundance (Cosby *et al.*, 2021). However, even in genomes with low KRAB abundance, such as birds with fewer than ten KRAB-zinc fingers, two independent host-transposase captures to a KRAB domain have been detected (Cosby *et al.*, 2021).

In addition to coding sequences, non-coding sequences that bear a resemblance to DNA transposons are another source of genetic material that can be repurposed for new functions. For instance, MITEs have a palindromic structure that makes them beneficial to becoming miRNA genes (Piriyapongsa & Jordan, 2007). The MADE1 elements are a well-known example of such MITEs, which were recently propagated in primate evolution by a Tc1/Mariner transposase and gave rise to one of the most prominent human miRNA families, *hsa-mir-548* (Liang *et al.*, 2012; Piriyapongsa & Jordan, 2007).

1.2 The *piggyBac* transposon

The *Trichoplusia ni* (T. ni) piggyBac transposon has gained popularity as a molecular tool for transgenesis due to its activity in a variety of cell lines and species, including mammals. One of its key advantages is that its transposition mechanism is specific to TTAA sites and leaves no DNA footprints, as demonstrated by several studies (Fraser *et al.*, 1996; Mitra *et al.*, 2008; Wilson *et al.*, 2007). Additionally, the repair of the broken sites does not require DNA synthesis (Mitra *et al.*, 2008), which is a further advantage of the PiggyBac system.

However, some integration sites can differ from the preferred TTAA sites (about 2%), and host factors are required to repair these sites, resulting in mismatches (Li *et al.*, 2013). Furthermore, the excision site can also be somewhat variable, which ensures that transposons do not get trapped (Li *et al.*, 2013).

In the thesis, the term piggyBac is used for the *Ti. ni* piggyBac transposase, *piggyBac-like elements* (PBLEs) are putative active transposases belonging to the piggyBac superfamily of transposases, and *piggyBac-derived* (PGBD) sequences are domesticated elements.

1.2.1 Chemical steps of piggyBac transposition

The molecular mechanism of transposition by piggyBac is visualized in Figure 2 and was discovered by Mitra *et al.* (2008). This process involves several hydrolysis and transesterification steps. Initially, the transposases bind to the TIRs at the transposon ends and cleaves the 3’ end of each strand through hydrolysis. The resulting free 3’OH group reacts with the complementary 5’ end, creating a TTAA hairpin and catalyzing excision from the donor DNA via transesterification.

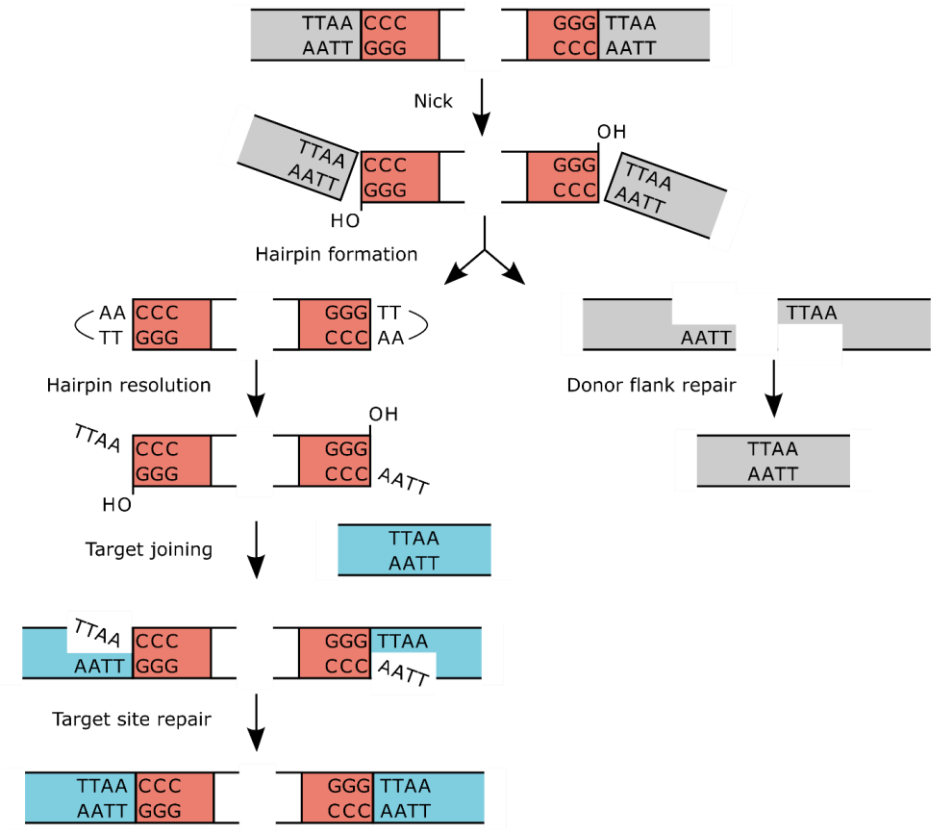


Figure 2: Chemical steps of the piggyBac transposition mechanism. Hydrolysis: The piggyBac transposase performs hydrolysis, releasing the 3'-OH group on the DNA strand for integration. Transesterification and hairpin formation: The 3'-OH attacks flanking DNA, forming a DNA hairpin structure four nucleotides from the transposon end. Hairpin opening and TTAA overhang: piggyBac opens the DNA hairpin, generating a four nt TTAA overhang on each end. Second transesterification and integration: piggyBac catalyzes a second transesterification, joining the ends with TTAA overhangs to the target DNA. Donor flank repair: Complementary DNA strands allow seamless repair at the empty donor site. Adapted from Mitra et al. (2008)

The hairpin structure of the DNA intermediate is resolved rapidly by another hydrolysis step that leads to a hairpin opening and a four-nucleotide TTAA overhang. The resulting free 3'OH groups initiate end joining by attacking the 5' end of the targeted TTAA site in a transesterification step. After the target site ligation, the transposition is completed. The broken DNA ends at the donor site are complementary TTAA overhangs, which are repaired seamlessly.

1.2.2 PiggyBac structure and domain architecture

The overall structure of piggyBac (when incubated with left TIR) is an asymmetric dimer that synapses two parallel TIR ends (Chen *et al.*, 2020). This dimer is theoretically sufficient for the transposition reaction (Chen *et al.*, 2020; Luo *et al.*, 2022). However, piggyBac oligomers are more active due to the following reason: The recognition of the 19 bp palindrome within the TIRs is achieved by a CRD dimer, and the piggyBac monomer contains only a solo CRD (Chen *et al.*, 2020; Luo *et al.*, 2022). Hence, a tetramer was suggested as the most straightforward configuration, as each of the two palindromes is bound by two dimerized CRDs (Chen *et al.*, 2020; Luo *et al.*, 2022).

The piggyBac transposase is a modular protein consisting of several functional subdomains. In the following paragraph, I will review the significance of the protein domains in piggyBac transposition.

1.2.2.1 *N-terminal domain (NTD) (1-116 aa)*

Initially, the role of the *N-terminal domain* (NTD) (1-116 aa) in piggyBac transposition was unclear, despite the elucidation of the piggyBac 3-D structure by cryo-EM (Chen *et al.*, 2020). It was thought to be unnecessary for transposition due to its disordered structure and to mediate protein-protein interactions with host factors (Chen *et al.*, 2020). However, recent studies have explored the role of this region in piggyBac transposition.

Deletion of the first 100 amino acids of the NTD compromised transposition activity, particularly in the excision step, indicating its importance (Wachtl *et al.*, 2022). In a recent study, Luo *et al.* (2022) demonstrated that the NTD plays a regulatory role in piggyBac activity. Specifically, phosphorylation within the N-terminal at three specific sites inhibited transposition by weakening the DNA-protein interaction. Conversely, mutation of the identified phosphorylation

sites or truncation of the first 74 amino acids increased transposition by piggyBac, indicating a possible regulatory layer by which host genomes manipulate piggyBac activity.

Interestingly, the truncation of the first 74 aa did not influence the piggyBac integration profile (Luo *et al.*, 2022). Notably, further truncation of the NTD (missing the first 104 aa) did not affect piggyBac excision but impaired its integration (Luo *et al.*, 2022). This suggests that the NTD is not involved in targeting but is required for integration. Additionally, the study showed that the NTD is involved in piggyBac dimerization even in the absence of transposon DNA, adding to its multifunctionality (Luo *et al.*, 2022). Altogether, the NTD is involved in at least three functions: dimerization, integration, and transposase activity regulation.

1.2.2.2 DNA-binding and dimerization domain (DDBD) (117-263 & 457-535)

The *DNA-binding and dimerization domain* (DDBD) knits the protein together and interacts with TIRs (bp 7-16) (Chen *et al.*, 2020). This domain, along with the catalytic and insertion domain, is also involved in the synapse formation of the TIRs (Chen *et al.*, 2020).

1.2.2.3 *Catalytic domain (264-371 & 433-456) & insertion domain (372-432)*

PiggyBac has a catalytic DDE/D motif, as many other transposases and integrases. Structural alignments and mutational screens have shown that D268, D346, and D447 are the catalytic triad in piggyBac (Keith *et al.*, 2008; Mitra *et al.*, 2008). All three residues are essential for transposition, including nicking, hairpin resolution, and target joining (Mitra *et al.*, 2008).

The catalytic domain in piggyBac has an RNaseH-like fold which is interrupted by the insertion domain as seen in other DDE/D transposases with insertion domains (Chen *et al.*, 2020).

1.2.2.4 C-terminal region including cysteine-rich domain (CRD) (553-594)

Mitra *et al.* claimed in (2008) that the cysteine-rich domain (CRD) is dispensable for piggyBac recombination, and truncated piggyBac (1-558) shows the same activity in *in-vitro* assays as the wild-type piggyBac (1-594). The idea was that CRD might be necessary for chromatin interaction which is absent *in-vitro*. However, they did not provide any supporting experimental data.

Morellet *et al.* in 2018 showed that the CRD is essential for transposition *in-vitro* and *in-vivo*. That a truncated piggyBac (1-558) is not transposing *in-vivo* in mammalian cells is not surprising because the bipartite nuclear localization signal (NLS) of piggyBac is located in the C-terminal (554-KKR-556 and 565-KIRRK-569) region overlapping with the CRD. The missing NLS diminishes the protein transport into the nucleus (Helou, Beauclair, Dardente, Arensburger, *et al.*, 2021). But the study further provided *in-vitro* evidence that showed the inability of truncated piggyBac to induce DSBs at the initial strand cleavage step. Notably, truncated piggyBac was still able to cleave the TTAA hairpin after formation to initiate target joining.

The 3-D cryo-EM structure of full-length piggyBac was elucidated by Chen *et al.* (2020). The study showed that the CRD could only bind the TIRs in the form of a dimer and hypothesized that a piggyBac tetramer would be necessary to perform transposition. Further, they noticed that the CRD was essential to bend DNA by about 40%. This conformational change contributes to TTAA specificity (Chen *et al.*, 2020).

A recent study by Luo *et al.* (2022) confirmed this hypothesis indirectly. The researchers modified piggyBac transposase from a tetrameric complex to a more active dimeric complex. One key was to add another C-terminal domain to piggyBac which would dimerize and allow a single piggyBac to recognize the palindrome structure at the TIRs.

1.2.2.5 *Do some PBLE transpose without CRD?*

It was stated that some distantly related PBLEs transpose without the presence of the CRD, although this transposition resulted in a 10-fold reduced integration efficiency and mostly improper target site duplications or TIRs (in about 81% of integrations) (Helou, Beauclair, Dardente, Arensburger, *et al.*, 2021). That might highlight the importance of the CRD for TTAA-specific seamless excision and integration.

1.2.2.6 Terminal inverted repeats (TIRs)

PiggyBac left, and right TIRs are asymmetric. For *in-vitro* and *in-vivo* transposition, a 35 bp left TIR and a 63 bp right TIR are sufficient (Chen *et al.*, 2020). They share some common sequences and repeats but differ in their overall structural organization (Chen *et al.*, 2020). For piggyBac, this asymmetric arrangement is required for transposition (Chen *et al.*, 2020). This requirement can be overcome by adding a second CRD to the transposase (Luo *et al.*, 2022). Also, modified piggyBacs, such as hyPBase (hyperactive piggyBac), transpose with two left TIRs (Chen *et al.*, 2020). Providing an additional binding site using the left TIR in tandem plus the right TIR increased transposition. Most likely by stabilizing DNA-protein binding (Chen *et al.*, 2020). The CRD recognizes a 19 bp palindrome sequence within the TIRs (Chen *et al.*, 2020; Morellet *et al.*, 2018).

1.3 The PGBD gene family

The human PGBD gene family consists of five genes that have been domesticated in multiple waves. Here I will review the evolutionary origins of PGBD elements and known functions.

The piggyBac transposable element was first discovered in the cabbage looper moth *Trichoplusia ni* (Cary *et al.*, 1989; Fraser *et al.*, 1983), and since then, many PBLEs and PGBD sequences have been identified in various eukaryotic genomes, ranging from primates (Sarkar *et al.*, 2003) to protozoans (Pritham *et al.*, 2005). However, only a subset of these elements are active transposases.

PGBD elements have been divided into nine distinct groups, indicating nine independent domestication events throughout the eukaryotic tree (Bouallègue *et al.*, 2017). Specific elements include *PiggyMac* (PGM) and *Tetrahymena piggyBac-like 2* (TPB2) in ciliates, KOBUTA in amphibians, and NeoPGBD in fishes. PGBD3 and PGBD4 are found exclusively in primates, while PGBD1 and PGBD2 are mammalian-specific. PGBD5 is the oldest element, having integrated into the common ancestor of vertebrates and cephalochordates over 500 million years ago (Bouallègue *et al.*, 2017; Pavelitz *et al.*, 2013). Five of these elements, PGBD1-5, are present in the human genome, but their host functions remain poorly characterized.

1.3.1 Functions of other domesticated PGBD elements

Examining other piggyBac elements can provide insights into their potential functions. For example, PGM is a catalytically active element with intact DDD, and during genomic reprogramming in ciliates (*Paramecium*), it introduces DSBs at TA sites to remove unwanted sequences (Baudry *et al.*, 2009; Dubois *et al.*, 2017). PGM is tightly linked to the classical *non-homologous end-joining* (NHEJ) repair process (PGM-KU complex) to ensure efficient repair of these sites (Kapusta *et al.*, 2011; Marmignon *et al.*, 2014). In addition, five other catalytically inactive PGM-like genes work with PGM to remove sequences by forming a complex and mediating precise positioning (Bischerour *et al.*, 2018).

TPB2 is another active element, and in Tetrahymena (another ciliate species), it functions similarly to PGM in excising unwanted sequences, although those are not flanked by specific sequences but are instead marked by heterochromatin structures (Cheng *et al.*, 2010). TPB2 has been suggested to be targeted to sequences by recognizing certain heterochromatin features and

performing endonuclease activity on DNA (Cheng *et al.*, 2010, 2016). This is different from the targeting of piggyBac, which is unfavorable to heterochromatin (Gogol-Döring *et al.*, 2016).

KOBUTA is another domesticated and likely inactive element with an unknown function in Xenopus (Hikosaka *et al.*, 2007). NeoPGBDs are exclusively found in fish, and their functions have not yet been studied (Bouallègue *et al.*, 2017).

1.3.2 Functions of PGBD1-5

The mammalian PGBD1-5 family is a relatively understudied group of genes; however, a few functions are currently attributed to them. One of the first characterized proteins from this family is PGBD3 and its fusion transcript *Excision repair cross-complementation group 6* (ERCC6)/PGBD3. PGBD3 is integrated into the 5th intron of the ERCC6 gene locus and acts as an alternative 3' terminal exon, with three transcripts expressed from this locus: PGBD3, ERCC6/PGBD3, and ERCC6 (Newman *et al.*, 2008). The PGBD3 sequence contains a 5' splice acceptor site upstream of the coding ORF and a polyadenylation sequence at the 3' end of the ORF, acting as a natural exon trap (Newman *et al.*, 2008).

Research has shown that the fusion transcript ERCC6/PGBD3 plays a role in DNA repair and the induction of genes involved in innate immunity and interferon-like antiviral response (Bailey *et al.*, 2012). It has also been found to bind AP-1 complex proteins at DNA and modulate their target gene expression (L. T. Gray *et al.*, 2012). However, it is unknown how the transposase domain contributes to these functions.

Interestingly, restoring the catalytic motif from DND to DDD did not restore transposon activity with MER85 substrate, suggesting that other factors or domains may be involved in this process (L. T. Gray *et al.*, 2012; Kolacsek *et al.*, 2022).

PGBD4 is a member of the PGBD family with a conserved DDD triad, and it is associated with MER75 TIRs that flank the PGBD4 genomic locus. However, despite its intact transposase motif,

PGBD4 does not transpose with the MER75b substrate, suggesting that its function may not involve transposition (Kolacsek *et al.*, 2022). The role of PGBD4 in the genome remains unclear.

The PGBD3 and PGBD4 are primate-specific and exhibit recent integration events throughout the genome. Multiple pseudogenes, solo TIRs, and MITEs are dispersed within the genome, suggestive of recent transpositional activity (Kolacsek *et al.*, 2022). It has been shown that PGBD3 recognizes a 16 bp imperfect palindrome sequence, rather than the 13 bp TIR sequence, for binding to MER85 (L. T. Gray *et al.*, 2012). *Electrophoretic mobility shift assay* (EMSA) experiments have highlighted the importance of both TIRs and palindromic sequences for PGBD3 binding (L. T. Gray *et al.*, 2012). The human genome harbors approximately 900 MER85 elements (Bailey *et al.*, 2012). Although PGBD3 and ERCC6/PGBD3 recognize their associated TIRs (MER85), the binding does not affect nearby gene expression (L. T. Gray *et al.*, 2012).

PGBD1 is a fusion gene that contains an N-terminal SCAN domain and a C-terminal transposase-derived protein domain (Sarkar *et al.*, 2003) and has been implicated in several neurological disorders. Genome-wide association studies have linked PGBD1 to both schizophrenia and Alzheimer's disease (Belbin *et al.*, 2011; Yue *et al.*, 2011). PGBD1 and PGBD3 are examples of transposase-derived sequences that fused to host domains by host transposase capture described above (Cosby *et al.*, 2021).

PGBD2 is a protein with an unknown function that is highly expressed in syncytiotrophoblasts, specialized cells that play a crucial role in placental development and function (*Human Protein Atlas*, 2023). A transcript PGBD2-SZT2 has been one of four chimeric transcripts shown to be expressed in preeclamptic patients (Shi *et al.*, 2023). A genome-wide association study investigating blood clotting in a healthy Chinese population found that an SNP in PGBD2 correlated with activated partial thromboplastin time (F. Zhang *et al.*, 2022). A missense mutation in PGBD2 and deletion of the complete PGBD2 locus in two patients have been associated with a congenital diaphragmatic hernia (Zhu *et al.*, 2018). That condition is characterized by malformation of the diaphragm and hypoplasia of the lungs, is one of the most common and severe birth defects, and is associated with high morbidity and mortality rates.

PGBD5 has been suggested to be an active recombinase in humans, with potential transposase activity in mammalian cell cultures (Henssen *et al.*, 2015, 2017a). It has been proposed to promote site-specific DNA rearrangements in childhood solid tumors, indicating its potential role in tumorigenesis (Henssen *et al.*, 2017a). PGBD5 literature is further reviewed in an upcoming Chapter (see 2.4).

1.3.3 Phylogenetic relationships of PGBDs

A study by Bouallègue *et al.* (2017) investigated the phylogenetic relationships among the PGBD protein family. The analysis suggested that PGBD1, PGBD2, and PGBD3 are closely related, with PGBD2 being the closest relative of PGBD1. PGBD4 was found to cluster together with KOBUTA, PGM, and TPB2, while PGBD5 was somewhat distinct from the other PGBDs, likely due to its early domestication.

To identify potentially active elements that gave rise to the domesticated PGBD families, the authors performed a phylogenetic analysis of PBLEs and PGBDs. The analysis revealed homologous sequences of PGBD3 in spider *Stegodyphus mimosarum* and aphid *Acyrthosiphum pisum*, which might have originated from a common ancestor PBLE. This common ancestor most likely possessed a 5' splice acceptor site and a 3' polyadenylation signal, as found in PGBD3, and those sequences are the most closely related sequences of PGBD1 and PGBD2 as well. However, the RNA processing features were lost in PGBD1 and PGBD2.

PGBD4 and the PBLE of the bat *Pteropus vampyrus* grouped together, while PGBD5 resided in an outgroup with no homology to any other PBLE. Interestingly, the authors found a fragmented and unlikely domesticated piggyBac-related sequence in hemichordates.

1.4 Exploring the functions of PGBD5: A literature review

1.4.1 The debate about PGBD5's transposase activity

Several studies have investigated the controversial issue of whether PGBD5 retains catalytic activity and functions as an active transposase. Pavelitz *et al.* (2013) suggested that PGBD5 lacks the conserved DDD triad necessary for transposition and is unlikely to have catalytic activity. However, Henssen *et al.* (2015) demonstrated that PGBD5 is an active transposase that catalyzes piggyBac transposition in human cells. Their study utilized a transposition assay based on neomycin-resistant colony formation quantification and showed that GFP-PGBD5 promotes transgenesis by forming neomycin-resistant colonies. They also found that intact TIR and hairpin formation was necessary for PGBD5 transposition, indicating that it is substrate specific. The authors suggested a newly identified DDD triad that catalyzed PGBD5 transposition and showed that mutations in this triad reduce the number of transposon copies in HEK293T cells. They also found that PGBD5-mediated transposition occurs at TTAA sequences (82% of all integrations), which is a unique feature of piggyBac transposition, suggesting that PGBD5 mediates the same transposition mechanism. Finally, the authors concede that PGBD5 facilitated transposition on nacked substrate but may behave differently on chromatinized DNA.

Henssen *et al.* (2016) aimed to explore PGBD5 recombinase activity in human cell lines on chromatinized DNA substrate. While they did not observe transposition, they found recombinase activity of PGBD5. They generated fibroblasts that stably overexpress *Green fluorescent protein* (GFP)-PGBD5 or GFP and found that GFP-PGBD5 overexpression resulted in more complex structural variations, especially inversions, compared to GFP. The authors suggested that PGBD5 itself induces DSBs and that the flanking sequences of breakpoints might reflect PGBD5-specific recognition sequences. They generated a putative *PGBD5 signal sequence* (PSS) motif based on 13 pre-selected sequences in the genomic locus *Hypoxanthine phosphoribosyltransferase 1* (HPRT1), which comprises 43 kb. However, this study is limited to the HPRT1 locus, and further studies are needed to investigate the generalizability of the PSS motif.

The study by Henssen, Reed, *et al.* (2017) demonstrated that PGBD5, a transposable element belonging to the piggyBac-like (PBLE) family, is responsible for oncogenic mutations in human tumors. The authors showed that PGBD5 transposes with PSS sequences as substrate, which are abundant in the human genome but usually chromatinized and not naked as in the constructs used in the study. The study found that PSS motifs are enriched at sites of genomic rearrangements in rhabdoid tumors and that PGBD5 binds to DNA sequences enriched in PSS motifs. The study also demonstrated that PGBD5 is catalytically active at the PSS motif sequence and that its enzymatic activity is required for its transformation activity. Additionally, the study showed that PGBD5-induced tumors display specific intrachromosomal deletions, inversions, duplications, and translocations, with breakpoints enriched in PSS motifs.

In a subsequent study by Helou, Beauclair, Dardente, Arensburger, *et al.* (2021), the authors challenged the prevailing belief that the CRD domain is required for piggyBac transposition (see Chapter 2.2.2). The researchers truncated piggyBac C-terminally and added an NLS to the sequence. The authors found that truncated piggyBac lacking the CRD domain, as well as murine and human PGBD5 that are naturally deficient in that domain, is able to transpose with piggyBac substrate. However, the truncation of the CRD domain resulted in a high number of improper transposition events (81%). The study also found that murine and human PGBD5 showed lower efficiency (10% and 4.6% respectively) in canonical transposition events compared to piggyBac. Integration sites were often found in genic regions, predominantly in neuronal genes, including genes involved in synapse formation, regulation of small GTPase-mediated signal transduction, and nervous system development.

Helou, Beauclair, Dardente, Piégu, *et al.* (2021) investigated whether PGBD5 also transposes other PBLEs. The hypothesis was that PGBD5 transposes or integrates closely related PBLEs more efficiently than distantly related elements such as piggyBac. They identified Tcr-pble from *Timema cristinae* as the closest related element to PGBD5 based on homology. The study found that human and murine PGBD5 is able to integrate closely related PBLEs, such as Tcr-pble, with equal efficiency as piggyBac. However, overexpression of PGBD5 was cytotoxic, and most integrations were non-canonical transposition events that might rely solely on PGBD5's potential nuclease activity.

The finding that PGBD5 is catalyzing transposition with piggyBac substrate raised concerns about the safety of piggyBac-mediated gene transfer. The study by Beckermann *et al.* (2021) attempted to replicate the results of previous studies and conducted additional *in-vitro* assays to explore the issue in more detail. They found that PGBD5 was unable to transpose piggyBac transposons in human cells, as evidenced by negative excision assays, colony count, and plasmid rescue analysis. Finally, they show that PGBD5 is unable to bind the TIRs of piggyBac, most likely because of the missing CRD domain. The authors note that the difference in methods to detect transposon integrations might explain some of the differences in results. The researchers of Henssen *et al.* (2015, 2017) and Helou *et al.* (2021) used *Polymerase chain reaction* (PCR)-amplification and *Next generation sequencing* (NGS) of insertion sites. This method is very sensitive; however, it is also subject to potential PCR artifacts or ligation of PCR-amplified products. Beckermann *et al.*(2021) used plasmid rescue instead a method that does not require PCR.

In the discussion, Beckermann and colleagues highlight some of the inconsistencies between the studies of Helou *et al.* and Henssen *et al*. found that PGBD5 precisely excises piggyBac transposons leading to a single PCR band and neatless repaired TTAA sites in the excision assay. Further, their insertion site analysis showed that 65 out of 67 transposon junctions had intact TIRs and TTAA sites. In contrast, Helou *et al.* reported poor fidelity of less than 5% in their NGS analysis of potential PGBD5 insertions. Additionally, Helou *et al.* reported cytotoxic effects of PGBD5 overexpression that are used for colony count normalization but were not sufficiently investigated.

Beckermann *et al.* also criticizes the conclusions made by Helou *et al.*, in particular their suggestion that piggyBac transposition does not require the CRD domain. Helou *et al*. show a 10-fold reduction in integration efficiency with truncated piggyBac and loss of fidelity of TIRs and target site duplications that were only found in 19% of the integrations compared to 96.7% with full-length piggyBac. Additionally, several colony count assays did not show significant effects with truncated piggyBac. Beckermann *et al*. conclude that the CRD domain is necessary for correct piggyBac transposition with high fidelity and efficiency. The authors speculate that

truncated piggyBac might still retain some nuclease activity that led to these results. However, PGBD5s potential nuclease activity still needs to be demonstrated.

The study of Beckermann *et al.* concentrated on the cross-interactions of piggyBac (insect), piggyBat (Bat), and PGBD5 (human). Kolacsek *et al.* (2022) further investigated the potential transposition activity of PGBD5 with piggyBac, MER75, and MER85 substrates. The authors demonstrated that PGBD5 was not able to excise or transpose either of the substrates.

The potential transposase activity of PGBD5 has been a topic of intense discussion within the transposon research community. While some studies have supported this idea, they have all been conducted by the same lab led by Alex Kentsis (Henssen *et al.*, 2015, 2016; Henssen, Koche, *et al.*, 2017) or in collaboration with his lab (Helou, Beauclair, Dardente, Arensburger, *et al.*, 2021; Helou, Beauclair, Dardente, Piégu, *et al.*, 2021). In contrast, several other labs attempting to replicate these results have failed, prompting some scientists to request further *in-vitro* assays to support the claims of PGBD5's transposase activity.

It is worth noting that while piggyBac lacking the CRD domain may retain some non-specific nuclease activity due to the presence of the intact catalytic triad DDD (Beckermann *et al.*, 2021), it is unable to excise DNA but can facilitate hairpin opening (Morellet *et al.*, 2018). However, the same cannot be assumed for PGBD5. This protein lacks the conserved DDD motif, and it has not been demonstrated that the alternative triad is brought into close proximity once the protein is folded.

1.4.2 Other literature investigating PGBD5 functions

In Pavelitz *et al.* (2013), the expression pattern of PGBD5 was investigated, revealing strong expression in the brain and weak expression in the spinal cord, as well as low expression in leukemia cell lines but not in healthy B- and T-cells. The study also showed murine Pgbd5 expression in specific regions of the brain, including the olfactory bulb, hippocampus, and cerebellum, with specificity to granule cells in the olfactory bulb and hippocampus. Furthermore,

PGBD5 was found to be bound by several proteins, including *RE1 silencing transcription factor* (REST), *REST Corepressor 1* (CoREST), *SIN3A transcription regulator family member A* (SIN3), *Tripartite motif containing 28* (TRIM28), *Signal transducer and activator of transcription 3* (STAT3), and *CCCTC-binding factor* (CTCF), within its promoter and gene body. However, it was shown that PGBD5 does not bind to DNA or chromatin, despite localizing to the nucleus.

A study by Zapater *et al.* (2023) demonstrates that depletion or dysfunction of PGBD5 in mouse or human were associated with developmental delays, intellectual disabilities, learning, and motor deficits, as well as recurrent seizures. The authors suggested that those arose from somatic DNA rearrangements introduced by Pgbd5s nuclease activity and that neuronal somatic DNA rearrangements are required for normal brain development. However, it remains to be proven that Pgbd5 acts directly at DNA. The authors conceded that Pgbd5 might interact with co-factors, including other nucleases and chromatin remodeling factors, to promote DNA rearrangements. It also remains to be proven that genomic rearrangements in *Neural progenitor cells* (NPCs) and neurons are physiologically relevant or byproducts.

Another recent study (Simi *et al.*, 2023) has identified Pgbd5 as a crucial mediator of neuronal cell commitment. The study suggested that Pgbd5 plays a key role in modulating the cell cycle exit of NPCs in mice. The study demonstrated that cells expressing Pgbd5 exhibit markers of neuronal cell determination and maintenance. In contrast, depletion of Pgbd5 increased the expression of genes typically involved in the proliferation and renewal of NPCs. Additionally, Pgbd5 was implicated in neuronal migration and cortex development, where its activity is associated with an increase in DNA DSBs. Furthermore, the study highlighted that Pgbd5 has a significant impact on cell metabolism.

Henssen, Reed, *et al.* (2017) investigated the potential of therapeutic targeting of PGBD5 in pediatric solid tumors. The study found that cells expressing PGBD5 require intact NHEJ repair and DNA damage signaling, suggesting that PGBD5 induces DSBs. The authors identified two potent chemicals, AZD6738 and KU60019, which are *Ataxia telangiectasia and Rad3 related* (ATR) and *Ataxia telangiectasia mutated* (ATM)-selective kinase inhibitors that interfered with cell growth and survival in PGBD5-expressing cells. AZD6738, which selectively inhibits ATR

kinase activity, showed the most substantial effect and compromised the growth and survival of PGBD5-expressing cells. Knockdown of endogenous PGBD5 induced resistance to AZD6738 in tumor cells, indicating that ATR inhibition by AZD6738 is specifically lethal for PGBD5-expressing cells. The study also found that AZD6738 treatment impaired the growth of neuroblastoma and medulloblastoma tumors in a mouse model, with DNA damage accumulation and apoptosis predominantly observed in the G1 phase of the cell cycle.

A study by Xie *et al.* (2022) aimed to improve the treatment of patients suffering from papillary thyroid cancer, a common malignant tumor. The researchers found that differences in glucose metabolism-related genes in those patients reflected clinicopathological features of papillary thyroid cancer. They could identify six genes, including PGBD5 which were sufficient to create a model for prognostic outcomes. PGBD5 was the only of the six genes that was enriched in the papillary thyroid cancer high-risk group. Knockdown of PGBD5 in papillary thyroid cancer cells drastically decreased their proliferation.

Another study by Mutalip *et al.* (2014) found that PGBD5 was associated with 17βH-neriifolin (cardiac glycoside) treatment induced cell death in ovarian cancer. These results suggest that PGBD5 might be a valuable target in ovarian cancer therapeutics.

Analysis of genome-wide association studies investigating immune-related genetic enrichment in frontotemporal dementia identified 15 loci, including the PGBD5 gene locus, as potentially critical factors in frontotemporal dementia pathogenesis (Broce *et al.*, 2018). This is the first study linking PGBD5 to the immune system.

Overall, PGBD5 has been shown to be a neural-specific gene involved in neuronal migration and cell cycle exit throughout development. Its expression is associated with increased DSBs during neuronal differentiation and requires intact DNA repair machinery. Ectopic expression of PGBD5 has been associated with several cancers, such as ovarian cancer, neuroblastoma, medulloblastoma, rhabdoid tumors, and Erwin sarcoma. It has been associated with cell proliferation in cancer, glucose metabolism, and the neuronal immune system.

1.5 Topoisomerase II and transcription of immediate early genes

1.5.1 Overview of human topoisomerases

Topoisomerases are enzymes that play a crucial role in resolving topological problems that arise due to the complex organization of DNA within cells. Human DNA is highly compacted yet must remain accessible for various cellular functions such as transcription, replication, and repair. Additionally, each cell contains circular mitochondrial DNA and folded RNAs present in various subcellular compartments. Topoisomerases exist in all three domains of life (Pommier *et al.*, 2022). The human genome comprises six topoisomerases with partially redundant but also specific functions (Pommier *et al.*, 2022).

Type IB topoisomerases, TOP1 and TOP1MT, release supercoiled DNA by binding to dsDNA and cleaving a single strand (Pommier *et al.*, 2022). The broken strand is then rotated around the intact strand and re-ligated without the need for co-factors (Pommier *et al.*, 2022). Type IIA topoisomerases, TOP2A and TOP2B release DNA supercoils, catenanes, and knots by inducing temporary DNA DSBs, facilitating strand passage of the intact strand, and re-ligating the broken DNA ends with the help of *Adenosine triphosphate* (ATP) and divalent metal ions such as Mg2+ (Pommier *et al.*, 2022). Type IA topoisomerases, TOP3A and TOP3B, bind and nick single-stranded nucleic acids. While TOP3A is active at DNA, TOP3B is active at both DNA and RNA. Its activity is independent of ATP but requires divalent metal ions. The action of TOP3 enzymes also depends on host factors (Pommier *et al.*, 2022).

All types of topoisomerases form temporary *DNA-protein crosslinks* (DPCs) prior to nucleic acid nicking, with type IB forming 3’ DPCs and type IA and type IIA forming 5’ DPCs (Pommier *et al.*, 2022). Altogether, topoisomerase types differ in their polarity, substrate specificity, nucleic acid relaxation mechanisms, and cofactor requirements (Pommier *et al.*, 2022).

In some cases, the religation within the *topoisomerase cleavage complex* (TOPcc) fails, resulting in trapped topoisomerases (TOP-DPCs) (Pommier *et al.*, 2022). These stalled TOPccs can be caused by endogenous and environmental factors and require DNA repair machinery for

resolution (Pommier *et al.*, 2022). The intentional trapping of TOP-DPCs by chemicals is widely used in anticancer and antibacterial chemotherapies (Pommier *et al.*, 2022).

Etoposide (VP-16) is an anticancer drug and a chemical used to study Top II function in research. Etoposide binds to Top II-DNA covalent complexes and initiates topoisomerase II poisoning (Buzun *et al.*, 2020). It prevents the religation of the DNA strands, which results in permanent DSBs that lead to cell death (Buzun *et al.*, 2020).

Merbarone is a catalytic Top II inhibitor. It prevents the 5′-phosphotyrosyl bond in the Top IIα-DNA complex and is thereby inhibiting Top II upstream of the DSBs formation (Buzun *et al.*, 2020).

Topoisomerases mediate several essential functions during replication and transcription, and they are vital to genome stability and organization. In the following, I will concentrate on the involvement of topoisomerases type IIA during transcription.

1.5.2 Comparing type IIA topoisomerases: TOP2A and TOP2B

The vertebrate genome contains two paralogues of type IIA topoisomerase, TOP2A and TOP2B, which arose from a whole-genome duplication event during early chordate evolution (McLysaght *et al.*, 2002). TOP2A and TOP2B are structurally similar proteins with identical protein domain architecture (Austin & Marsh, 1998). They mainly differ in their *C-terminal regions* (CTR) (Austin & Marsh, 1998; Madabhushi, 2018). Those regions contribute to their regulation and affect their activity. TOP2B-CTR has decreased the strand passage activity and DNA binding affinity of TOP2B compared to TOP2A-CTR, based on experiments with artificial chimeric constructs (Gilroy & Austin, 2011; Meczes *et al.*, 2008).

In addition to these differences, TOP2A and TOP2B have distinct expression patterns and cell cycle dependencies. TOP2A is expressed mainly during S1, G2, and M phases and is responsible for cell cycle progression, while TOP2B expression is stable throughout the cell cycle and does not influence cell proliferation (Capranico *et al.*, 1992; Woessner *et al.*, 1991). Additionally, the

cellular localization of the proteins differs during the cell cycle (Chaly *et al.*, 1996). TOP2A is tissue-specific and highly expressed in tissues containing proliferating cells, such as bone marrow, intestine, and spleen, whereas TOP2B is ubiquitously expressed (Capranico *et al.*, 1992). Interestingly, there is a transition from TOP2A to TOP2B expression in granule and Purkinje cell progenitors as they exit the cell cycle (K. Tsutsui *et al.*, 2001). Moreover, TOP2A is non-essential in terminally differentiated cells (Pommier *et al.*, 2022).

TOP2B has a special role in nervous system development (Madabhushi, 2018). It is essential for neural development, as shown in Top2b knockout mice (Lyu & Wang, 2003). Further, Top2b activity is crucial for neurite outgrowth of cerebellar and cortical neurons, targeting of retinal ganglion cell axons, and proper wiring of the visual system (Nevin *et al.*, 2011; Zaim & Isik, 2018). De-novo mutations in human TOP2B were associated with intellectual disabilities, hypotonia, progressive microcephaly, and autistic features (Lam *et al.*, 2017).

Notably, TOP2A and TOP2B usually homodimerize to exert their catalytic function; however, in Hela cells, a substantial portion of TOP2 dimers was found to be catalytically active TOP2A-TOP2B heterodimers (Biersack *et al.*, 1996).

1.5.3 Topoisomerases and transcription

In a relaxed state, DNA has a helical coiled conformation. Winding of DNA by, e.g., helicases and ATP translocases open the DNA duplex and generate supercoils. Winding in the same direction leads to positive supercoils, and winding in the opposite direction causes negative supercoils. In the case of transcription, RNA polymerases progress along the DNA, which generates positive supercoils ahead and negative supercoils behind them. This model is also called the *twin supercoiled domain model* (Liu & Wang, 1987). Topoisomerases are required to relax these supercoils. Notably, positive supercoils in front of RNA polymerases hinder transcription elongation, while negative supercoils behind the transcription machinery are associated with increased transcription initiation through the promotion of DNA melting (Hirose & Suzuki, 1988; Matsumoto & Hirose, 2004; Rifka *et al.*, 2015). The negative supercoils behind

RNA polymerase are generally less efficiently relaxed, leading to an accumulation of negative supercoils (Pommier *et al.*, 2022).

Both types of supercoils are effectively resolved by TOP1 and TOP2 topoisomerases (Pommier *et al.*, 2022). TOP1 acts at underwound DNA duplexes, and TOP2 acts at DNA duplex crossovers (Pommier *et al.*, 2022). However, it was observed that catalytically active TOP1 accumulates predominantly along the gene body and is associated with transcriptional elongation by removing positive supercoils (Baranello *et al.*, 2016). TOP2 was more efficient at relaxing supercoils that accumulate within chromatin (Salceda *et al.*, 2006). Additionally, TOP2 is required for the transcription of highly transcribed and very long genes (King *et al.*, 2013; Kouzine *et al.*, 2013). Notably, TOP1 also promotes the transcription of long genes, but when genes exceed a certain length > 200 kb, it is not sufficient to do so (King *et al.*, 2013).

DNA binding sites of TOP2A and TOP2B are enriched at the promoters, enhancers, and gene bodies of actively transcribed genes. Genome-wide characterization of TOP2B binding sites in neurons showed that TOP2B preferentially binds upstream of actively transcribed genes (Madabhushi *et al.*, 2015). TOP2B binding sites in neurons overlap with transcription factors specialized to neuronal-activity-dependent genes, such as *CAMP-response element binding protein* (CREB), *Serum response factor* (SRF), and *CREB-binding protein* (CBP) but also with chromatin structural protein CTCF (Madabhushi *et al.*, 2015). TOP2B activity is especially crucial for the late stages of neuronal differentiation, and the loss of TOP2B only affects a subset of neuronal genes (Madabhushi, 2018).

1.5.4 Type IIA topoisomerases induce the expression of immediate early genes

Inhibition of either TOP2A or TOP2B was found to induce the expression of a subset of *immediate early genes* (IEGs) (Herrero-Ruiz *et al.*, 2021; Madabhushi *et al.*, 2015). IEGs are a group of genes that are rapidly and transiently induced in response to extracellular and intracellular signals (Gallo *et al.*, 2018). They are the first transcriptional response of cells to a stimulus and include well-studied examples such as *Fos proto-oncogene* (FOS), *Activity*

regulated cytoskeleton associated protein (ARC), *Neuronal PAS domain protein 4* (NPAS4) , *Dual specificity protein phosphatase 1* (DUSP1), *Nuclear receptor subfamily 4 group A member 1* (NR4A1), and *Early growth factor 1* (EGR1).

FOS is a commonly used marker for neuronal activity, and its expression is associated with synaptic plasticity, learning, and memory (Tanaka *et al.*, 2014; Weber Boutros *et al.*, 2022). Depletion of FOS in mice hippocampi resulted in increased neuronal excitability, more severe kainic acid-induced seizures, and neuronal cell death (J. Zhang *et al.*, 2002).

ARC is an effector IEG. It modulates synapses to modulate their neurotransmitter response and is associated with long-term memory consolidation as well as spatial memory in the hippocampus of rodents (Tanaka *et al.*, 2014; Weber Boutros *et al.*, 2022).

NPAS4 promotes the balance of excitation and inhibition in neurons (X. Sun & Lin, 2016). Further, it was associated with memory formation (X. Sun & Lin, 2016), and dysfunction led to psychiatric disorders (Rossi *et al.*, 2021).

EGR1 induces changes in the brain methylome upon neuronal activation together with TET1 (Z. Sun *et al.*, 2019). It was associated with motor learning (Brito *et al.*, 2022), long-term neuronal plasticity (Mataga *et al.*, 2001), and brain development (Weber Boutros *et al.*, 2022). *DUSP1* maintains homeostasis of *Mitogen-activated protein kinase* (MAPK) signaling and facilitates several neural protective functions in the brain (Pérez-Sen *et al.*, 2019).

NR4A1 is a transcription factor induced as a response to stress (Jeanneteau *et al.*, 2018). It modulates mitochondrial function and synaptic growth (Jeanneteau *et al.*, 2018).

IEGs exhibit a bursting transcription pattern, e.g., FOS expression peaks around 30-60 min after the stimulus (Herrero-Ruiz *et al.*, 2021). Many of these genes are transcription factors themselves, and they induce the expression of other genes, thereby propagating a gene expression cascade (Weber Boutros *et al.*, 2022). In the brain, this cascade ultimately modulates synaptic plasticity, a process critical for learning and memory (Boutros *et al.*, 2022, Gallo *et al.*, 2018).

1.5.4.1 The supercoil inhibition model

A recent study by Herrero-Ruiz *et al.* (2021) showed that TOP2-induced transcriptional changes are dependent on its ability to release negative DNA supercoils at transcription start sites. TOP2A was inhibited using different chemical compounds such as Merbarone and Etoposide in human-telomerase-immortalized retinal pigment epithelial 1 (RPE-1) cells. The transcriptome analysis revealed mainly an upregulation of genes, indicating that TOP2A had a repressive role. As mentioned before, the upregulated genes showed an enrichment of IEGs. *Chromatin immunoprecipitation* (ChIP) analysis showed that upregulated genes had a higher occupancy of TOP2A than randomly selected genes with similar expression levels under basal conditions (cell cycle arrest G0/G1). Overall, TOP2A peaks correlated with RNA polymerase II and active histone marks such as H3K4me3 and H3K27ac. The study showed that inhibition of TOP2A accumulated negative supercoiled DNA at the transcription start site of genes. The accumulated negative supercoils allowed for promoter-proximal pause release. Notably, the mechanism was independent of DNA DSBs and cellular stress induction.

1.5.4.2 The breakage model

Another proposed mechanism of IEG induction by TOP2 states that TOP2-mediated DSBs are critical to the release of RNA polymerase promoter-proximal pausing by changing the chromatin structure. The pause-release allowed for the rapid transcription of genes.

Madabhushi *et al.* (2015) were surprised when they noticed that Etoposide (a TOP2 inhibitor) treatment in neurons induced the transcription of IEGs. The idea was further inspired by the observation of accumulated DSBs upon neuronal activity in IEG promoter regions. Several studies have shown that neuronal stimulation by various paradigms and chemical compounds induces DSBs in neurons (Konopka & Atkin, 2022). DSBs in neurons have been traditionally associated with destructive processes such as aging and neurodegeneration (Konopka & Atkin, 2022). However, a growing body of evidence suggests that they have physiological roles in brain

plasticity, including memory formation (Konopka & Atkin, 2022) and memory recall (Weber Boutros *et al.*, 2022).

Madabhushi *et al.* (2015) observed that N-methyl-D-aspartate (NMDA) treatment, similar to Etoposide, induced transcription of IEGs. The transcription initiation was accompanied by increased TOP2B-ccs in those promoters with either treatment. That finding indicated that frequently observed DSBs upon neuronal stimulation are caused by Top2 b-induced DNA cleavage. IEG promoters such as Fos and *Neuronal PAS domain protein 2* (Npas2) were bound by Tob2b under basal conditions, and NMDA treatment increased genome-wide Top2b binding drastically. Further, the authors showed that Top2b was critical for the induction of DSBs upon stimulation and that initiation of DSBs with *Clustered regularly interspaced short palindromic repeats/ CRISPR-associated protein* (CRISPR/Cas) in IEG promoters was sufficient to drive their expression. When DSBs break repair was inhibited, the cells showed a continuous expression of Fos.

Those findings were coherent with previous research that reported TOP2B mediated DSBs in promoters of various receptor signaling pathways. Ju *et al.* (2006) were the first to report that DNA cleavage was necessary for the transcription of estrogen receptor α (ERα) target genes in human breast cancer cells MCF-7. They found that upon estradiol treatment, ERα and TOP2B are recruited to the promoter of estrogen-responsive gene pS2. The induction of gene expression was blocked when TOP2B was inhibited. Although *in-vitro* TOP2B DSBs are re-ligated by the enzyme itself, the study showed that classical repair pathways were recruited to repair the lesions, e.g., *X-ray repair cross-complementing 6* (XRCC6*), X-ray repair cross-complementing 5* (XRCC5), *DNA-dependent protein kinase* (DNA-PK), and *Poly ADP-ribose polymerase 1* (PARP1). They found similar stimulus-activated DSBs in target gene promoters of the androgen receptor, retinoic acid receptor, and thyroid hormone receptor.

Those results were confirmed by further studies that showed enriched TOP2B mediated DSBs to diverse physiological stimuli, such as androgens, insulin, glucocorticoids, retinoic acid, and serum (Madabhushi 2018).

The genome-wide binding sites of TOP2B were increased around the binding sites of CTCF and the cohesion complex (Uusküla-Reimand *et al.*, 2016). Further, TOP2B interacts with CTCF and members of the cohesion complex through protein-protein interactions (Uusküla-Reimand *et al.*, 2016). CTCF sites were significantly closer to the *Tanscription start site* (TSS) of genes that governed DSBs upon NMDA treatment (Madabhushi *et al.*, 2015).

Those studies resulted in a DNA "breakage model" (Uusküla-Reimand & Wilson, 2022). The gene promoters are already loaded with RNA polymerases which are, however, paused by promoter-proximal pausing. CTCF/cohesion complex forms chromatin loops that form a topological barrier and hinder transcription by preventing promoter-enhancer interactions. Promoters and enhancers are also prebound by required transcription factors. Once topoisomerase II induces DSBs, an interaction of promoter and enhancer is facilitated, and the gene is consequently transcribed after the pause release (Madabhushi *et al.*, 2015).

The new findings of Herrero-Ruiz *et al.* (2021), however, challenge this model. Treatment with Etoposide which results in stable DSBs, showed a reduced induction of gene expression compared to NMDA and merbarone or serum (Herrero-Ruiz *et al.*, 2021; Madabhushi *et al.*, 2015). Merbarone inhibits TOP2A upstream of DSBs initiation. It was proving that a reduced catalytic activity of TOP2 is sufficient to induce gene expression without the involvement of TOP2ccs. Deletion of TOP2A or TOP2B did not affect the induction of c-FOS. Instead, it increased basal levels and elevated the serum response of c-FOS. That indicates that TOP2 has a rather repressive function for those promoters. Induced DSBs in FOS promoter using CRISPR/Cas hindered expression upon serum activation of c-FOS. However, this model was insufficient to explain all data supporting the DNA breakage model.

Further research is required to determine which model explains the induction of gene expression upon TOP2 inhibition better. Or whether both models complement each other.

1.6 Aim of the study

This thesis investigates the domestication of DNA transposons to form entirely new genes. Therefore, we investigated the evolutionary history and possible molecular functions of the human PGBD family, especially PGBD1 and 5, as an example of host transposase domestication events.

Part I

One theory on how the host utilizes newly acquired transposases was stated by Cosby *et al.* in 2021. The researchers found that domesticated transposases are often fused to host domains. The authors suggested that the transposase domains might be domesticated to serve as DNA binding interfaces, while host domains such as KRAB, SCAN, or SET mediate gene regulatory functions. However, this implies that host domains are under higher evolutionary constraints than transposase domains, which would only undergo evolutionary constraints regarding DNA binding interfaces.

Therefore, we aim to investigate the evolutionary forces acting on the PGBD family, especially the PGBD1 element, and test the theory on the example of this gene family. To do so, we will calculate phylogeny-supported ratios of nonsynonymous and synonymous substitution rates of the entire genes and their subdomains. This analysis will help us identify regions of evolutionary constraint indicative of functionality and shed light on the domesticated features of the PGBD family.

Part II

PGBD elements have been frequently domesticated to the vertebrate genome. Still, it is unclear whether common features, such as DNA or protein binding capacities, are recycled during domestication or whether each event is unique. In a broader scope, we aim to elucidate the

molecular functions of the PGBD family and finally compare each domestication event. In the scope of this thesis, we aim to gain insights into the molecular functions of PGBD5.

PGBD5 has gained significant attention in recent years due to compelling research suggesting it functions as an active transposase in human cells (Henssen *et al.*, 2015). Since then, numerous studies have investigated PGBD5's putative recombinase and nuclease activity (Beckermann *et al.*, 2021; Henssen *et al.*, 2016; Henssen, Koche, *et al.*, 2017; Henssen, Reed, *et al.*, 2017; Jubierre Zapater *et al.*, 2023). While some studies have supported these claims, such as the association of PGBD5 with genomic rearrangements and DNA damage in the form of DNA DSBs, other experiments have refuted these theories by demonstrating that PGBD5 lacks enzymatic activity.

We will investigate the potential for PGBD5 to act as a nuclease. We will evaluate structural information and predicted fold and compare these to its ancient relative, piggyBac. Additionally, we will recreate the protein-protein interactome of PGBD5 to elucidate which factors, such as nucleases, PGBD5 may interact with to initiate DSBs. In parallel, we will analyze the interactome of PGBD5 to gain insights into the cellular pathways in which it is involved. Combining this with analysis of the transcriptome of Pgbd5 knockout mice, we will identify relevant pathways and ultimately better understand the biological function of PGBD5.

2 Materials and methods

This chapter contains a description of all materials and methods used in this study. Table S 2 includes a list of used antibodies and chemicals and Table S 3 lists all machines.

2.1 Domain architecture of the PGBD family

The protein sequences of PGBD1-5 were obtained from the NCBI database and used as input queries in the Phyre 2 server (Kelley *et al.*, 2015) to perform structural alignments. The gene subdomains were manually annotated by examining the structural alignment output and cross-referencing it with the piggyBac annotation provided in the publication of Chen *et al.* (2020). Additionally, the TIRs were annotated using the Repbase track (Jurka, 2000), which is a comprehensive database of repetitive DNA elements, available on the UCSC genome browser (Kent WJ *et al.*, 2002). This approach allowed us to accurately identify and annotate the key structural and functional features of PGBD1-5 genes (Table 1).

Table 1: NCBI reference sequences and domain annotations of the human PGBD family

Gene symbol	Reference protein sequence	NTD	DDBD1	Catalytic 1	Insertion domain	Catalytic 2	DDBD2	CRD
PGBD1	NP_115896.1	1-290	405-541	541-651	652-725	726-750	751-804	-
PGBD2	NP_733843.1	1-126	127-264	265-374	375-449	450-473	474-520	557-586
PGBD3	NP_736609.2	1-597	598-732	733-843	844-917	918-940	941-1015	1026-1055
PGBD4	NP_689808.2	1-89	90-244	245-354	355-417	418-441	442-518	537-578
PGBD5	NP_001245240.1	1-107	108-251	252-367	368-438	439-462	463-516	-

2.1.1 Analysis of evolutionary constraints with PAML

2.1.1.1 K_a/K_s ratio calculation

PGBD *messenger RNA* (mRNA) *coding sequences* (CDS) were retrieved from the NCBI database, and sequences that showed the highest bitscores, when blasted to the human query PGBD sequence, were manually selected. The sequences were aligned using MUSCLE (Edgar, 2004) with default parameters to generate a multiple-sequence alignment of the translated amino acids. This alignment was performed using the UGENE software (Okonechnikov *et al.*, 2012). The taxonomy trees available on NCBI were manually modified to generate unrooted trees and utilized for the subsequent analysis.

PGBD1 and PGBD2 mammalian: (KOALA, (MOUSE, (PONAB, (HUMAN, PANTR), MACMU)), (HORSE, (PIG, PHYMC), CALUR, DESRO));

ERCC6/PGBD3 primates: ((HYLML, NOMLE), (HUMAN, PONAB, (PANTR, PANPA), GORGO), (PILTE, COLAB, RHIRO, CHLSB, MANLE, THEGE, PAPAN(MACNE, MACMU, MACFA), CERAT));

PGBD4 primates: (AOTNA, (HUMAN, PONAB, PANTR, GORGO), (PILTE, COLAB, (RHIRO, RHIBE), MANLE, THEGE, (MACNE, MACMU, MACFA), CERAT), (CEBCI, SAIBB, CALJA));

PGBD5 vertebrates: (((9AMPH, GEOSA), ((CAMFR, PHOSS, PIG), HORSE, ((HALGR), (PANPR, FELCA), CANLF), (AOTNA, (HUMAN, PONAB, (PANTR, PANPA), GORGO), (MACNE, MACMU)))), (SCLFO, ANGAN));

The K_a/K_s ratios were calculated using PAML (version 4.9, Yang, 2007) (M0).

2.1.1.2 Neutral and adaptive evolution

To evaluate the adaptive evolution found in the KRAB region of PGBD1, the primate sequences within the mammalian tree were compared to other mammalian sequences by M1a versus M2a in PAML. In this comparison, primates were used as foreground and all other mammals as background, and the hypothesis was evaluated in a chi2 test (with *degrees of freedom* = 2). Foreground branches were manually marked and tested against the unmarked background branches in PAML. Furthermore, the adaptive evolution of the KRAB(-like) region in primates was assessed by performing a comparison of M1 versus M2, using a chi2 test (with *degrees of freedom* = 1). This test was performed using the primate phylogenic tree of PGBD1.

PGBD1 primates: (PROCO, TARSY, (SAIBB, AOTNA, ((PONAB, (HUMAN, (PATNR, PANPA), GORGO)), ((9PRIM, COLAB, RHIBE), (CHLSB, MANLE, THEGE, (MACNE, MACMU, MACFA), CERAT)))));

2.1.2 Dating horizontal gene transfer

The Ensembl synteny browser (Yates *et al.*, 2020) was unable to allocate a syntenic region between monotremes and humans around the PGBD1 and PGBD2 loci (Figure S 2). To investigate whether the absence of PGBD1 and PGBD2 in these taxa was due to homology search failure or true absence, a recent method was used called AbSENSE (Weisman *et al.*, 2020). To assess the evolutionary distances between nine species pairs (human-rhesus macaque, human-Ma's night monkey, human-goat, human-camel, human-koala, human-platypus, human-American alligator, human-green anole and African clawed frog), orthologs were retrieved from the BUSCO curated vertebrate dataset (Simão *et al.*, 2015). A total of 73 genes were common to all selected species, and isoforms were selected based on their IsoSel (Philippon *et al.*, 2017). Gene sequences were aligned using MUSCLE (default) and concatenated into one alignment. The evolutionary distances were calculated using Protdist (PHYLIP, default) (Felsenstein, 2009), with human as the focal species. Bitscores were calculated using BLATSP (NCBI).

Finally, significance testing was performed according to the method described in Weisman *et al.* (2020). These analyses allowed us to assess the likelihood that the absence of PGBD1 and PGBD2 in monotremes and reptiles was due to homology search failure or true absence.

2.1.3 Phylogenetic tree generation of PGBD1 and PGBD2

A total of approximately 12,000 sequences containing the Pfam domain Transposase *Insertion element 4* (IS4) were retrieved from the Interpro Uniprot database (Finn *et al.*, 2017). These sequences were aligned using the MAFFT program (Katoh *et al.*, 2002) with default settings. From an initial tree calculated using the UPGMA algorithm (Michener & Sokal, 1957), a subtree was manually selected that included the cluster of PGBD1 and PGBD2 plus some closely related sequences. Sequences shorter than 250 bp and identical sequences (CD-HIT 100% identical) were removed. The PGBD1 and PGBD2 sequences (XP_020822236.1 and XP_020822393.1) from koala were manually added to the alignment. The selected transcripts were realigned using MUSCLE with default settings, and a phylogenetic tree was constructed using MrBayes (Huelsenbeck & Ronquist, 2000; Ronquist & Huelsenbeck, 2003) with a mixed rate model, single chain, and an average standard deviation of split frequencies < 0.05. The resulting tree was visualized with iTOL (Letunic & Bork, 2016). Protein domains were annotated using the HMMscan (Finn *et al.*, 2011) program from the Pfam database. For visualization purposes, another tree was constructed using representative PGBD1 and PGBD2 sequences along with invertebrate sequences using MrBayes with the same settings. Protein domains were annotated using HMMscan and CDD (NCBI), and the KRAB-like domains were annotated using Phyre 2.

2.1.4 Conservation of PGBD1 across rodent model organisms

The PGBD1 exon architecture and conservation track was obtained from the UCSC genome browser (hg19). To detect the conservation of the SCAN, KRAB and catalytic domain, multiple

sequence alignment of mammalian PGBD1 sequences were performed using sequences retrieved from the NCBI (mouse: XP_030103153.1, rat: XP_017456282.1). MUSCLE was used to align the sequences to their respective structural templates (SCAN: c3lhrA_, KRAB: d1v65a_).

2.2 PGBD5 interactome

2.2.1 Cell culture and transfection

Human Embryonic Kidney (HEK293) cells were cultured in Dulbecco's modified minimal essential medium (DMEM) supplemented with 10% fetal bovine serum, 1% L-glutamine, and 1% penicillin-streptomycin (Thermo Fisher Scientific). To ensure the quality of the cultures, routine screening for mycoplasma infection was conducted by RT-PCR.
For transfection, the jetPRIME transfection reagent (Polyplus) was used according to the manufacturer's instructions. Specifically, 4 x 10^5 cells were seeded onto 6-well plates or 2 x 10^6 cells were seeded onto 10 cm dishes. The lipid-DNA mixture was prepared and transfected into the cells using 500 ng or 2 µg of DNA, as specified by the protocol.

2.2.2 Stable isotope labelling by amino acids in cell culture (SILAC) followed by affinity purification mass spectrometry (AP-MS)

- Performed and summarized by Dr. Tamas Rasko and Christian Sommer

Two populations of HEK293 cells were cultivated in cell culture for three weeks. One population of cells was fed with growth medium containing normal amino acids ("light cell population"). The second population of cells was cultured in growth medium, containing amino acids labelled with stable heavy isotopes ($^{13}C_6$-$^{15}N_4$ L-arginine; $^{13}C_6$-$^{15}N_2$ L-lysine) ("heavy cell population"). In our experimental approach, untagged PGBD5 and *Human influenza hemagglutinin* (HA)-tagged PGBD5 were overexpressed in both conditions. The overexpressing plasmids encoding

the HA-tagged or untagged PGBD5 were transfected into the two cell populations of HEK293 cells, respectively. Three days post transfection protein purification was performed using EZview^{TM} red colored Anti-HA agarose affinity gel, following the recommendations of the manufacturer. Purified protein mixtures were prepared as follows: (a) as forward experiment: protein mixture of Heavy HA-PGBD5 and protein mixture of Light PGBD5; (b) as label-swap experiment: protein mixture of Light HA-PGBD5 and protein mixture of Heavy PGBD5. The purified protein mixtures including HA-tagged target proteins and their interacting partners were subjected to LS-MS/MS (mass spectrometry): Samples were processed by methanol-chloroform extraction, reduced, alkylated and digested with LysC and trypsin using standard protocols. After offline desalting, peptides were analyzed by LC-MS/MS on a Proxeon EASY-nLC II system, connected to a Q Excative mass spectrometer (Thermo Scientific). Chromatography was performed using a 120 min acetonitrile gradient on a 25cm long inhouse prepared column (ReproSil-Pur 120 C18-AQ, 3 µm (Dr. Maisch GmbH HPLC)). The instrument was operated in the data dependent mode with the following settings for the full scans: resolution 70,000, AGC target value 3E6, maximum injection time 20 msec. The following settings were chosen for the MS2 scans: resolution 17,500, AGC target value 1E6, maximum injection time 60 msec. Raw files were processed with MaxQuant (version 1.4.1.2).

2.2.3 SILAC-AP-MS analysis

2.2.3.1 Calculation of significance

The significance was calculated as described in (Cox & Mann, 2008) with small modifications. Contaminants and proteins without a detected unique peptide were excluded. Subsequently, a logarithmic transformation to the normalized H/L ratios (log2 with pseudo count 1) was applied and a Gaussian distribution of the transformed data was confirmed by the Shapiro-Wilk test. The 15.87th, 50th and 84.15th percentiles were calculated and called r_{-1}, r_0 and r_1, respectively. The z-transformation is defined as:

$z = \frac{r - r_0}{r_1 - r_0}$ for $r > r_0$ and $z = \frac{r_0 - r}{r_0 - r_{-1}}$ for $r < r_0$, where r are the transformed H/L ratios.

To determine the statistical significance of the results, p-values were calculated using the significance A formula. It should be noted that in the original paper, significance A produces two-sided p-values. For this analysis, the directionality of the test was considered and adjusted the distribution to one side accordingly, resulting in a one-sided p-value calculation.

$$Significance\ A = \frac{1}{2}\ erfc\ (\frac{z}{\sqrt{2}})\ , \text{for } z \geq 0.$$

$$Significance\ A = 1 - \left(\frac{1}{2}\ erfc\ (\frac{z}{\sqrt{2}})\right), \text{for } z < 0.$$

The label-swap experiment conducted in this study involved a reversed sidedness. That is why, adjustments for sidedness were made when $z \geq 0$. Additionally, to control for false positives, p-values were adjusted for multiple testing using the Benjamini-Hochberg method. Only proteins were considered that were significant, FDR < 0.05, in both label-swap experiment (intersection). The p-values obtained from both experiments were combined using the Berger method, which was implemented using the scran R package (Lun *et al.*, 2016). The script used for the analysis was witten in Rstudio (R version 3.6.3) and employed the erfc function from the pracma package (Borchers, 2022).

2.2.3.2 Gene set enrichment analysis of PGBD5 interactome

For the gene set enrichment analysis, all gene sets from the Molecular Signature Database v7.2 (Liberzon *et al.*, 2011; Subramanian *et al.*, 2005) were downloaded. Gene sets with at least 5 and less than 500 gene members were selected for enrichment testing. The enrichment analysis was performed using a two sets of genes, foreground and background and utilized the "HG" test from the tmod R package (Zyla *et al.*, 2019). Some unique peptides mapped to multiple genes, and the

true origin of the peptides could not be determined. For the analysis, all genes associated with the mapped peptides were considered.

2.2.4 Validation of the interactome: (Co-)immunoprecipitation

For the *co-immunoprecipitation* (co-IPs) experiments, 7.5 x 10^6 HEK293 cells were lysed in 1 mL lysis buffer containing 50 mM TRIS HCl pH 8.0, 10 mM EDTA, 150 mM NaCl, 5% glycerol, 1% NP-40 supplemented with protease inhibitors (Pierce), Phosphatase inhibitor (Active motif) and Benzonase (Novagen) for 30 mins at 4°C on a head-over rotator. After incubation, the lysates were centrifuged for 20 min at 12,000 rpm at 4°C to remove unbroken cells and cell debris. Supernatants were collected and the protein concentrations were measured in a BCA protein assay (Pierce).

In parallel, 50 µl of DynabeadsTM Protein G or A was washed as recommended by the manufacturer and resuspended in 200 µl 0.02% Tween 20 in PBS. The mixture was incubated with primary antibodies (Table 2) for 10 min at room temperature on a turning wheel. After incubation, the magnetic beads were collected and washed with 200 µl 0.02% Tween 20 in PBS. Subsequently, 400 µl of cell lysates were added to the beads, and the mixture was incubated overnight at 4°C on a turning wheel. The beads were then washed three times with 200 µl wash buffer (5 mM TRIS HCl pH 8.0, 1 mM EDTA, 15 mM NaCl, 0.5% glycerol, 0.1% NP-40).

The proteins were eluted from the beads by adding a mixture of 20 µl 50 mM glycine pH 2.8 for 2 min, and the pH was balanced with an additional 20 µl 1M Tris-HCL pH=8.0. Before the samples were loaded onto the gel, they were denatured in 10 µl of 5 x SDS-loading buffer (10% SDS, 10 mM DTT, 20 % glycerol, 0.2 M Tris-HCl pH 6.8, 0.05% Bromophenol blue) and boiled at 99°C for 5 min.

Table 2: Validation of interactome: Antibodies used in co-IP experiments.

Antibody	**catalogue**	**company**	**Amount** [µg/IP]	**Dynabeads**
UBF rabbit	A301-859A	Bethyl Laboratories	20	Protein A
TAF1C rabbit	A303-698A	Thermofisher	30	Protein A
PGBD5 mouse	MBS355128	MyBioSource	40	Protein G
PGBD5 mouse	NBP2-67048	Novusbio	30	Protein G
IGG rabbit	02-6102	Thermofisher	As specific ABs	Protein A
IGG mouse	ab91353	Abcam	As specific ABs	Protein G

2.2.5 Immunocytochemistry staining and microscopy

Coverslips were seeded with 100,000 cells per well in 12-well cell culture plates. After 48 hours of transfection, cells were fixed with 4% paraformaldehyde (Sigma) and permeabilized with 0.1% Triton X-100 in PBS for 2 min. Primary antibodies were added to the coverslips and incubated overnight at 4°C. The coverslips were washed three times with PBS and then incubated with secondary antibodies for 60 min. Antibodies and dilutions are listed in Table 3. For the repeat experiments the fluorescence dyes of the secondary antibodies were swapped. After an additional washing step, Hoechst (Thermo Fisher Scientific) was used to stain the nuclei according to manufacturer's instructions and the samples were mounted using glycine based antifade mounting media (Vectashield). Images were captured using a Leica TCS SP8 inverted confocal microscope.

Table 3: Validation of interactome: antibodies and dilutions used in ICC experiments.

Antibody	catalogue	company	Dilution
UBF rabbit	A301-859A	Bethyl Laboratories	1/200
TAF1C rabbit	A303-698A	Thermofisher	1/200
FBXL19 rabbit	ab172961	Abcam	1/200
PGBD5 mouse	NBP2-67048	Novusbio	1/500
Alexa Fluor® 647 goat anti-rabbit	A-21244	Life Technologies	1/500
Alexa Fluor® 647 goat anti-mouse	A-21236	Life Technologies	1/500
Alexa Fluor® 488 donkey anti-rabbit	A-21206	Thermo Scientific	1/500
Alexa Fluor® 488 goat anti-mouse	A31620	Invitrogen	1/500

2.2.6 Chromatin immunoprecipitation followed by Western Blot (ChIP-WB)

HEK293 cells were fixed for 10 min by adding formaldehyde (methanol-free, Thermo Fisher Scientific) to the media to a final concentration of 1%. The reactions were quenched for 5 min by the addition of glycine to a final concentration of 125 µM. After two washed with ice-cold PBS, cells were centrifuged for 5 min at 1,000 x g at 4°C. Cells were resuspended in ChIP lysis buffer (59 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8, 1% Triton X-100, 0.1% Sodium Deoxycholate, 0.1% SDS and freshly added Protease inhibitors). After cell lysis and chromatin extraction for 10 min on ice, chromatin was sonicated using a BioRuptor Pico sonicator (Diagenode), followed by centrifugation at 12,000 × *g* for 15 min at 4°C. For each sample and IP, 25 µg chromatin were diluted 1:10 in RIPA buffer were used for further processing. Chromatin was precleared with protein G Dynabeads (Life Technolgies) and blocked with 0.2 mg/ml BSA and 50 µg/ml yeast tRNA (Invitrogen) and incubated with the respective

antibodies for 1 hour at 4°C. Antibody-bound chromatin was purified using blocked protein G Dynabeads overnight at 4°C. ChIP washes, protein elution and Western Blot were performed as for the IPs described above.

ChIP DNA was eluted in ChIP elution buffer (1% SDS, 100 mM $NaHCO_3$) and reversed cross-linked overnight at 65°C with 200 mM NaCl and RNase A (Sigma). The reverse cross-linked samples were treated with 20 µg/ml Proteinase K (Thermo Fisher Scientific) and purified using phenol-chloroform extraction to check for the length of sonicated DNA by agarose gel.

2.3 Transcriptome analysis of Pgbd5-knockout in mice

Prior to analysis, the quality of the sequenced libraries was assessed using FastQC (Andrews, 2010) and MultiQC (Ewels *et al.*, 2016). The genome sequence and annotation (gencode.vM10.annotation.gtf) of the mouse genome mm10 were downloaded from the GENCODE repository (Frankish *et al.*, 2021). The reads were mapped using STAR (--alignIntronMin 20 --alignIntronMax 1000000 --chimSegmentMin 15 --chimJunctionOverhangMin 15 --outFilterMultimapNmax 20, (Dobin & Gingeras, 2015), with tolerance for up to 20 multimappers. The quality of mapping was ensured using MultiQC, and RSeQC (Wang *et al.*, 2012). The gene count matrix was generated using featureCounts (default parameters, Liao *et al.*, 2014). The library was reverse stranded, but the hippocampus samples exhibited a high number of strand-unassigned reads (~13-21%), thus the unstranded option was selected for further analysis. However, both options yielded comparable results.

For the *Transcript per million* (TPM) calculation, read counts were normalized to their median transcript lengths as reported by featureCounts (calculated from gencode.vM10.annotation.gtf) and the cumulative number of uniquely mapped reads per sample. *Reads per kilobase of transcript per million mapped reads* (RPKMs) , *Trimmed mean of M-values* (TMMs), Upper quartiles were calculated with implemented functions of the R NOISeq package (Tarazona *et al.*, 2015).

After optimizing the normalization with *relative log expression* (RLE, (Gandolfo & Speed, 2018) and *principal component analysis* (PCA) plots, differential gene expression analysis was performed using Deseq2 (Love *et al.*, 2017) (normalization: rlogs, p-value adjustment: Benjamini-Hochberg). Genes with a log2 fold change of at least 0.3 and an absolute *false discovery rate* (FDR) smaller than 0.05 for the cerebellum and 0.1 for the hippocampus were considered significant.

The correlation of log2 fold changes between the two brain regions was determined using Pearson correlation and linear regression of R stats package. A combined p-value was calculated using the "berger" method of the R scran package.

Gene set enrichment analysis was performed using R tmod (Zyla *et al.*, 2019). The enrichment was tested with the CERNO-test, and the provided ranked gene list was sorted by FDR. The selected gene sets exhibited an *area under the curve* (AUC) of at least 0.75 and an FDR < 0.01.

2.4 PGBD5-knockout iPSC cell line

The hiPSC cell line BIHi005-A-24 was acquired from the Stem Cell Unit at MDC, Diecke Lab. The cell line was genetically modified to contain a doxycycline-inducible *Neurogenin 2* (NEUROG2) cassette, which allows for rapid differentiation into glutamatergic neurons. This system is commonly referred to as *inducible neurons* (iNeurons). The donor of the cell line was a male aged between 25-29 years old. Additionally, to the wild type cell line, a PGBD5 knockout cell line was generated by the Stem Cell Unit. The deletion of 10 bps within exon 3 resulted in a frameshift and hence knockout of the full-length protein.

The same cell lines were used in the neuronal differentiation protocol in Chapter 2.6.2.

2.4.1 Western Blot

Received cell pallets were lysed with lysis buffer (150 mM NaCl, 50 mM Tris pH 8, 1% Nonidet P40, 5% Glycerol, 10 mM EDTA) containing protease inhibitors (Pierce). Cells were lysed for 30 min on a head-over rotator at 4°C and after centrifugation for 20 min at 12.000 rpm, 4°C, the supernatants were frozen and stored at –80°C. Protein concentrations were measured in a BCA protein assay (Pierce). Samples were diluted with lysis buffer and 5 x sample buffer to a final protein concentration of 1 µg/µl. Protein samples are denatured while boiling for 5 min at 99°C. 20 µg protein was loaded to a gel (1,5 mm, 12% Acrylamide, stain-free, fast-cast gel, Bio-Rad). Gel electrophoresis was run at 90 V in SDS-Running buffer. The gel was activated for 5 min and imaged with Bio-Rad Chemi Doc MP Imaging System. Proteins were transferred to PVDF membrane with Bio-Rad Trans-Blot Turbo Transfer System (1.3 A, 25 V, 10 min). The membrane was blocked with 5% milk in TBS-T for overnight at 4°C, while shaking.

The primary antibody α PGBD5(7-F8-5) (1:5000, #NBP2-67048, Novusbio) or α PGBD5 (1:5000, MBS355128, MyBioSource) incubated for one hour at room temperature. After three washing steps, each 10 min in TBS-T, the secondary antibody Goat anti-Mouse IgG Secondary Antibody, HRP (1:10.000, ThermoFisher Scientific). The membrane was washed as described above and band were visualized with ECL (Amersham).

2.4.2 Cell culture

2.4.2.1 Thawing and maintenance

Cryo-frozen wild type cells were taken from liquid nitrogen tank and thawed in 37°C water bath. The cells were diluted in 5 mL E8 (Gibco) medium in a 15 mL falcon. After centrifugation at 300 rpm for 5 min, the supernatant was removed, and the cell pallet was resuspended in 2 mL E8 medium containing Penicillin-Streptomycin (PenStrep, 100 U/ml, Gibco) and Y-27632 (Rock inhibitor, Peprotech). The cells were transferred to one well of a 6-well plate. The plate was pre-

coated with Matrigel (hESC-Qualified Matrix, ThermoFisher Scientific) in DMEM/F12 medium (Gibco) for 30 min at 37 °C. The cells are cultured in E8 medium containing PenStrep with daily medium change. The cells were split every 3-4 days at 60-70% confluency. The cells are kept in a hypoxia incubator (37°C, 5%CO_2, 5%O_2). The procedure for the knockout clones was similar with following modifications: StemFlex medium (Gibco) containing cloneR (Stemcell Technologies) substituted E8 medium and Rock inhibitor. In the following days StemFlex medium was slowly changed to E8 medium as indicated in Table 4.

2.4.2.2 Splitting

Cells were washed with DPBS (no calcium, no magnesium, Life Technologies) and 1 mL Accutase (Sigma)/well detach the cells from plate. After 5 min incubation at 37°C the reaction was stopped by adding 2 mL of medium (DMEM/F12, GlutaMAX (Gibco)). The cells were transferred to a 15 mL falcon and centrifuged at 300 rpm for 5 min. The supernatant was removed, and cells were resuspended in 1 mL medium (E8 supplemented with PenStrep and Rock inhibitor). The cells were split (1:6) to a 6-well plate containing E8 supplemented with PenStrep and Rock inhibitor. The 6-well plate have been pre-coated with Matrigel.

Table 4: StemFlex/ E8 medium transition scheme.

Day	**StemFlex : E8 medium**
1	StemFlex only
2	3:1
3	1:1
4	1:3
5	E8 only

2.4.3 Inhibition of topoisomerase II with Etoposide

E8 media was supplemented with either 200 µM Etoposide (VP-16) in DMSO or an equivalent amount of DMSO for the control. The treatment incubated for 30 min or 60 min at 37°C in hypoxia.

2.4.4 RNA extraction

Cells were washed with DPBS and 300 µl Trizol (ThermoFisher Scientific) per 1 well of a 6 well-plate was directly applied to the cells. After detaching the cells with a cell scraper, the cells were transferred to 1.5 mL Eppendorf tube and vortexed for 1 min. Cells were stored in a -80°C freezer.

RNA was extracted with Direct-zol RNA micro-prep kit (Zymo Research) according to manufacturer's instructions and eluted in RNase and DNase free water. The RNA was measured with a Spectrophotometer (DeNovix).

RNA was reverse transcribed to cDNA with High-Capacity RNA to cDNA reagents (Applied Biosystems). One reaction contains 1 µg RNA in total volume of 20 µl. The incubation was performed on a thermal cycler (Alpha Cycler, PCRmax) with steps indicated in Table 5.

Table 5: RNA to cDNA synthesis: Thermal cycler steps.

Step	Temperature	Time
Incubation	37 °C	60 min
Stopping reaction	95 °C	5 min
Hold	4 °C	continuously

2.4.5 Real-time polymerase chain reaction (RT-PCR)

CDNA was diluted 1:100 prior to *real-time polymerase chain reaction* (RT-PCR). The primers were synthesized by BioTez Berlin-Buch GmbH and are listed in Table 6. The reactions were prepared with Power SYBR green PCR Master Mix (Applied Biosystems) according to manufacturer's instructions. In brief, 2 µl diluted cDNA was mixed with 500 nM forward primer, 500 nM reverse primer, RNase and DNase free water (Sigma Aldrich) and 5 µl Power SYBR green PCR Master Mix in a total reaction volume of 10 µl. The RT-PCR was conducted on CFX96 Real Time PCR System (Bio-Rad), using the program outlined in Table 7. The data was analyzed by Maestro Software (Bio-Rad). The expression of all genes was normalized to ACTB and 18s *ribosomal RNA* (rRNA) expression.

Gene name	**Forward Primer (5' to 3')**	**Reverse Primer (5' to 3')**
PGBD5	TACAAGGTCCAGCCCTTCCT	GCACGTGGCAATGAATACAG
TOP2B	ATGATTTGGCTGGTTCGTGT	CCACCCCAGTTTCATCCAAT
FOS	GCCTCTCTTACTACCACTCACC	AGATGGCAGTGACCGTGGGAAT
EGR1	AGCAGCACCTTCAACCCTCAGG	GAGTGGTTTGGCTGGGGTAACT
ACTB	CTGGAACGGTGAAGGTGACA	AAGGGACTTCCTGTAACAATGCA
DUSP1	CAACCACAAGGCAGACATCAGC	GTAAGCAAGGCAGATGGTGGCT
GAPDH	ATGGAAATCCCATCACCATCTT	CGCCCCACTTGATTTTGG
18s rRNA	GATGGTAGTCGCCGTGCC	GCCTGCTGCCTTCCTTGG

Table 6: RT-PCR primer list.

Table 7: RT-PCR cycling scheme with pre-heated lid temperature of 105 °C.

Step	**Temperature**	**Time**	**Cycles**
Initial denaturation	95°C	10 min	1
Denaturation	95°C	15 sec	40
Annealing	60°C	1 min	40
Plate read			40
Melting curve			1

2.4.5.1 Standard and melting curves

Standard curves were built for every primer pair. The cDNA of the samples from one experiment were mixed. This master mix was then diluted in a 1:10 serial dilution and five samples ranging from 20 ng to 0.002 ng were used for standard curves. The mean of technical triplicates was calculated, to reduce the variability. Standard curves were the functions of *Cycle threshold* (Ct) values over cDNA concentrations, used in the reactions. With the slope of that function, the primer pair efficiency was calculated. Efficiencies within a range of 90 – 110% were acceptable. Melting curve analysis of PCR products in the presence of a fluorescent dsDNA binding protein such as SYBR Green, was used for an assessment of the purity.

2.5 Topoisomerase II specific decatenation activity

2.5.1 Protein purification

Cells were transfected using jetPRIME transfection reagent (Polyplus), according to the manufacturer's instruction. Briefly, 2 x 10^6 cell were seeded in a 10 cm dish and one day later, cells were transfected with pcDNA3.1-HA-PGBD5 plasmid, using 2 µg of DNA. 72 hours after transfection, cells were washed with DPBS and collected in 500 µl/dish ice-cold lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 150 mM NaCl, 5% glycerol, 1% NP-40 supplemented with protease inhibitors (Pierce), Phosphatase inhibitor (Active motif) and Benzonase (Novagen)). The cells were sonicated (Bioruptor Pico) for 15 sec and incubated 15 min on ice before centrifugation (12,000 x g, 4°C, 10 min). The supernatant was transferred to a new Eppendorf tube and 500 µl lysate was placed in a purification column. The protein purification was done with the HA-tagged purification kit (MBL) according to manufacturer's instructions. To generate a control, the same amount of non-transfected cells underwent the HA-purification procedure.

Briefly, the lysate together with the HA agarose beads incubated for 1 hour at 4°C on a head-over-rotator. Afterwards, the beads were washed three times in a purification column with wash buffer and were finally eluted in 40 µl HA elution peptide (HA peptide, 2 mg in 1 ml PBS).

2.5.1.1 Protein concentration measurements

The protein concentration was measured by Nanodrop (DS-11 FX+, DeNovix) with E1% 0.12. The molar extinction coefficient according to Beer-Lambert law was calculated with Expasy's ProtPram tool (Gasteiger *et al.*, 2005). The protein concentration ranged from 350 to 450 mg/ml.

2.5.1.2 Purity estimation

The purity measurement of the extraction was conducted through Western Blot analysis. For this purpose, samples were denatured in 5x SDS sample buffer at 99°C for 5 min. A gel containing 1.5 mm, 12% Acrylamide, and stain-free fast-cast gel (Bio-Rad) was loaded with 2.5 µg of protein, gel electrophoresis was run at 90 V in SDS-Running buffer. The stain-free gel was activated for 5 min and imaged with Chemi Doc MP Imaging System (Bio-Rad).

The images were analyzed with ImageJ to determine the purity of the extraction. In this regard, the image background was subtracted, and the intensity of the whole lane, an empty lane of the same size, and the HA-PGBD5 band were measured. The purity of the extraction was estimated by the following formula:

$$purity = \frac{\sum Intensity\ whole\ lane - \sum Intensity\ empty\ lane}{\sum Intensity\ HA\_PGBD5\ lane}$$

2.5.1.3 TOP2A co-IP

Furthermore, proteins were transferred to PVDF membrane with Trans-Blot Turbo Transfer System (1.3 A, 25 V, 10 min, Bio-Rad). The membrane was blocked with 5% milk in TBS-T for overnight at 4°C, while shaking. The primary antibody α TOP2A (1:6667, Abcam) incubated for one hour at room temperature. After three washing steps, each 10 min in TBS-T, the secondary antibody Goat anti-Rabbit IgG Secondary Antibody, HRP (1:6.667, Thermo Fisher Scientific) was given to the membrane. The membrane was washed as described above and bands were visualized with ECL (Amersham).

2.5.2 DNA decatenation assay

The reagents for the decatenation assay were mixed in PCR tubes, acquired from Profoldin, and adjusted to final concentrations of 10 mM Tris-HCl (pH 8), 50 mM NaCl, 0.1 mM EDTA, 50 mM KCl, 5 mM MgCl2, 15 µg/ml BSA, 3 µg/ml concatenated DNA, 0.2 mM ATP, and 10 U/ml human topoisomerase II α, and 85 µg/µl purified HA-Pgbd5 or control extract in a reaction volume of 25 µl. The control extracts used were either the HA-purified extract of untransfected HEK293 cells (described in 2.5.1) or HA elution peptide (HA peptide, 2 mg in 1 ml PBS). Optionally, 200 µM Etoposide (Merck) in 50% DMSO or for control 50% DMSO was added to the reaction.

The reactions incubated for 2.5 hours at 37°C. After incubation, 2 µl of 10% SDS was added and incubated another 1 minute at 37°C. The reactions were stopped by adding 5 µl of 0.5 M EDTA. Subsequently, 2 µl of Proteinase K (5 mg/ml, Longlife) were added to dissociate the proteins from the DNA. This mixture was incubated for an additional two hours at 37°C. The samples were mixed with 6 x loading dye (TriTrack, Thermo Scientific). A 0.8% agarose gel was prepared and loaded with 15 µl of the sample, a 1 kb plus DNA ladder, and 1.5 µg of linearized *kinetoplast DNA* (kDNA). The gel was run at 90 V for 2.5 hours and then imaged using a Chemi Doc MP Imaging System (Bio-Rad).

2.5.2.1 Linearized kDNA

Most rings of the concatenated DNA from Crithidia fasciculata have a size of 2,3 kb. The DNA sequence was retrieved from NCBI to identify a unique restriction site. Restriction enzyme digestion was perfromed with Xho1 (New England Biolabs) according to manufacturer's recommendations. Subsequently, the product was electrophoresed on an agarose gel at 90 V and the DNA band was excised with a scalpel and purified using the Zymoclean Gel DNA Recovery Kit. The kDNA was then amplified via PCR using KAPA Hifi Hotstart Ready mix (Roche) and the following primers: reverse primer GCAATCAATGTGTACCACGC and forward primer GTGCGATGTTGTGTTGATAG. The PCR was conducted with the following settings:

Table 8: PCR steps of linearizec kDNA

Step	**Temperature**	**Time**	**Cycles**
Initial denaturation	95°C	3 min	1
Denaturation	98°C	20 sec	30
Annealing	58°C	20 sec	30
Extension	72°C	3 min	30
Final extension	72°C	6 min	1

The DNA concentration was measured by Nanodrop (DS-11 FX+, DeNovix) and diluted with sterile water to a final concentration of 100 ng/µl.

2.5.3 Decatenation assay: Quantification of relaxed kDNA

The quantifications of the decatenation assays were performed with ImageJ (version 1.53k). In a first step, the background was subtracted from the raw images (*tiff* file format) with a rolling ball

radius of 50 pixels and background noise of the images were reduced by the *despeckle* option. Afterwards, rectangular ROIs were selected, and the band intensities were measured and normalized as follows:

$$RawIntDen = Mean\ Intensities * Area$$

$$IntDen = RawIntDen * pixel\ width\ \frac{inch}{1000} * pixel\ height\ \frac{inch}{1000}$$

With *RawIntDen* … Raw Intensity density and
NormIntDen … Normalized intensity density.

Afterwards the corresponding background was subtracted (Table 9).

Table 9: Control lanes for decatenation assay.

Foreground Sample	**Background Sample**	**Condition**
Pgbd5 + TopIIa	Pgbd5	Pgbd5
TopIIa	-	Control of Pgbd5
Pgbd5 + TopIIa + VP-16	Pgbd5 + VP-16	Pgbd5 + VP-16
TopIIa + VP-16	VP-16	Control of Pgbd5 + VP-16

The fold inductions of the samples over the control samples were calculated. The plots were generated with the ggplot2 package (Wickham, 2016) in RStudio. T-tests were performed with in-built R-functions in RStudio with *var.equal = FALSE.*

2.6 Expression patterns of PGBD5

Datasets were retrieved from different databases such as the Human Protein Atlas (Uhlén *et al.*, 2015), EMBO Single Cell Atlas (2022), BrainSpan (*BrainSpan*, 2022), and plots were generated in RStudio (RStudio version 1.0.143, R version 3.6) employing the ggplot2 package.

GTEx data was obtained from the GTEx portal (*GTEx Consortium*, 2022) and is based on Analysis V8 of the RNA-seq data. Unlike the Human Protein Atlas, the GTEx dataset includes all individual measurements, making it possible to examine tissue expression variation. The means, medians, standard deviations, and linear regression models were calculated using R base functions.

The age of donors in the BrainSpan data was initially provided in weeks, months, and years. To ensure consistency, the age was converted to days. Gene expression levels were quantified as -log2 RPKMs (reads per kilobase of transcript per million mapped reads). To examine the relationship between gene expression and age, a linear regression model was constructed using -log2 RPKMs as the dependent variable and the corresponding days as the independent variable. To assess the correlation between gene expression and age, Pearson correlation coefficients were calculated using -log2 RPKM values and log2-transformed age in days as inputs.

2.6.1 Single-cell RNA-seq analysis of a murine visual cortex

2.6.1.1 Preprocessing and differential gene expression analysis

Matrix files containing preprocessed counts of 1767 cells have been downloaded from EMBO Single Cell Atlas (E-GEOD: 71585, Tasic *et al.*, 2016). Seurat (v4.1.0) (Satija *et al.*, 2015) within RStudio was used for data processing. All cells with a minimum of 3000 expressed genes were kept. Counts were normalized with the *centered log ratio* (CLR) method. Clustering was performed using the *FindClusters* function with 0.5 resolution. The first 20 dimensions were

used of the principal component analysis for the construction of the *shared-nearest neighbor* (SNN) graph to generate data visualization using *Uniform manifold approximation and projection* (UMAP). All neuronal clusters were identified by neuronal marker gene expression *Enolase 2* (Eno2) and *RNA binding fox-1 homolog 3* (Rbfox3). Five non-neuronal cluster were removed from the analysis. Manual inspection of the histogram of Pgbd5 expression was used to determine the point of increasing Pgbd5 expression which was about 0.5. The cells were split into two groups, cells expressing high levels of Pgbd5 (expression >= 0.5) and cells expressing low levels of Pgbd5 (expression < 0.5). Afterwards, differential expression analysis was performed using the *FindMarkers* function. Genes with an absolute log2 fold change of at least 0.3 and an FDR of 0.05 were considered as differentially expressed.

2.6.1.2 Gene set enrichment analysis of Pgbd5 co-expressed genes in mouse visual cortex

For the gene set enrichment analysis, all mouse gene sets from the Molecular Signature Database version 2022.1 (Liberzon *et al.*, 2011; Subramanian *et al.*, 2005) and KEGG pathways (Kanehisa & Sato, 2020) were downloaded. Gene sets with at least 5 and less than 500 gene members were selected for enrichment testing. The enrichment analysis was performed using a ranked gene list approach and employed the "CERNO" algorithm from the tmod R package (Zyla *et al.*, 2019).

2.6.2 Neuronal differentiation

For the neuronal differentiation protocol, the cell line BIHi005-24 was used. Characteristics of the cell line are described in Chapter 2.4.

2.6.2.1 Thawing and maintenance

Cryo-frozen wild-type cells were taken from a liquid nitrogen tank and thawed in a 37°C water bath. The cells were diluted in 5 ml Essential 8 (E8, Gibco) medium in a 15 ml falcon. After centrifugation at 300 rpm for 5 min, the supernatant was removed, and the cell pellet was resuspended in 2 ml E8 medium containing Penicillin-Streptomycin (PenStrep, 100 U/ml, Gibco) and Y-27632 (Rock inhibitor, 10µM, Peprotech). The cells were transferred to one well of a 6-well plate. The plate was pre-coated with Matrigel (hESC-Qualified Matrix, ThermoFisher Scientific) in DMEM/F12 medium (Gibco) for 30 min at 37 °C. The cells were cultured in E8 medium (basal medium + supplements) containing PenStrep with daily medium change. The cells were split every 3-4 days at 60-70% confluency. The cells were kept in a hypoxia incubator (37°C, 5%CO_2, 5%O_2).

2.6.2.2 Splitting with accutase

Cells were washed with DPBS (no calcium, no magnesium, Life Technologies), and detached from the plate using 1 ml/well Accutase (Sigma). After 5 min incubation at 37°C, the reaction was stopped by adding 2 ml of medium (DMEM/F12, GlutaMAX (Gibco)). The cells were transferred to a 15 ml falcon and centrifuged at 300 rpm for 5 min. After aspiration of the supernatant, the cells were resuspended in 1 ml medium (E8 supplemented with PenStrep and Rock inhibitor). The cells were split (1:6) into a 6-well plate containing E8 supplemented with PenStrep and Rock inhibitor. The 6-well plate was pre-coated with Matrigel.

2.6.2.3 Splitting with Versene

Cells were washed in DPBS, and 1 ml/well Versene (1:5000, Gibco) detached the cells from the plate. After 5 min incubation at 37°C, the reaction was stopped by adding 2 ml of medium (DMEM/F12, GlutaMAX (Gibco)). The cells were detached by gentle pipetting (2x). The cells

were split (1:6) to a 6-well plate containing E8 supplemented with PenStrep and Rock inhibitor. The 6-well plates were pre-coated with Matrigel.

2.6.2.4 Neuronal differentiation protocol

One day before the induction, the cells were split with Accutase and plated at a density of 4.5 * 10^4/cm² onto a Matrigel-coated plate (either 6 or 12-well) in E8 medium containing Rock inhibitor.

The following day (day 0), the medium was replaced by F12-N2 medium (DMEM-F12 (+ L-Glutamine + 15mM HEPES, Gibco), N2 supplement (1x, Gibco), NEAA (1x, Gibco), hBDNF (10ng/ml, PeproTech), hNT-3 (10ng/ml, R&D Systems), laminin (0.2µg/ml, Sigma Aldrich) and PenStrep (100 U/ml, Gibco)) containing 2µg/ml freshly added doxycycline (Sigma Aldrich). After the media change, the cells were kept in a normoxia incubator (37°C, 5% CO_2).

On day 1, the media was changed with an F12-N2 medium containing freshly added doxycycline.

On day 2 the media was replaced by NB-B27 medium (Neurobasal medium, B-27 supplement (Gibco), GlutaMAX (Gibco), hBDNF (10ng/ml), hNT-3 (10ng/ml), laminin (0.2µg/ml) and PenStrep (100 U/ml, Gibco)) supplemented with 2µg/ml freshly added doxycycline.

On day 3, the media was exchanged with NB-B27 supplemented with 2µg/ml of freshly added doxycycline.

From day 4 to 15, 50% of the media was changed with NB-B27 supplemented with 2µg/ml of freshly added doxycycline every other day. Thawed doxycycline was kept for not more than one week at 4°C. The prepared medium was stored for a maximum of two weeks at 4°C.

2.6.2.5 Co-cultures with astrocytes

Mouse astrocytes were obtained from MDC's stem cell core unit (Dieckes lab). They were prepared in-house in the Kettenmann lab.

Cryo-frozen mouse astrocytes were taken from a liquid nitrogen tank and thawed in a 37°C water bath. The cells were diluted in 5 ml Dulbecco's Modified Eagle Medium (DMEM 1x + GlutaMAX +4.5 g/L D-Glucose + Pyruvate, Gibco) supplemented with fetal bovine serum (FBS, 10%, Life Technologies), NEAA (1x, Gibco) and PenStrep (100 U/ml, Gibco) in a 15 ml falcon. After centrifugation at 300 rpm for 5 min, the supernatant was removed, and the cell pellet was resuspended in 2 ml medium. The cells were transferred to a 10cm dish. The dish was pre-coated with Attachment factor (Sigma Aldrich). The cells were cultured with daily medium change in a normoxia incubator (37°C, 5% CO_2). After 5-7 days, the astrocytes were split with Accutase and NB-B27 medium into the neuronal cultures (Day 2 of differentiation) at a density of 4.5 * $10^4/cm^2$.

From day 4, the NB-B27 medium was supplemented with Cytosine -D-arabinofuranoside (Ara-c, 5 µM, Sigma Aldrich) to stop the further expansion of astrocytes and thereby make sure that they do not overgrow the neuronal culture.

Furthermore, from day 14 on neuron-astrocyte co-cultures received Peitz-medium (NB-B27 + GDNF (10 ng/µl, Peprotech) + ascorbic acid (200 µM, Sigma Aldrich) + cAMP (500 µM, Merck Millipore)). Those are essential supplements for the development of firing potentials in neurons.

2.6.2.6 Neuronal stimulation with NMDA

Neurons were stimulated with 50 µM N-Methyl-D-aspartic acid (NMDA; Sigma Aldrich) in serum-free Neurobasal medium supplemented with B27 for 10 min. Prior to stimulation, the old medium was removed and saved in a 50 ml Falcon tube. The stimulation medium (Neurobasal/B27 supplemented with 50 µM NMDA and without GlutaMAX) or control medium (Neurobasal/B27 without GlutaMAX) was added to the cells (500 µl for each 6-well, 350 µl for each 12-well) and incubated for 10 min in a humidified incubator at 37°C under normoxic conditions. After stimulation, the medium was aspirated and the saved, centrifuged medium was added back to the wells.

2.6.2.7 Cell collection

Cells were harvested either in lysis buffer (50 mM TRIS HCl pH 8.0, 10 mM EDTA, 150 mM NaCl, 5% glycerol, 1% NP-40 supplemented with protease inhibitors, Phosphatase inhibitor and Benzonase) for protein extraction or in Trizol for RNA extraction. Prior to cell collection the cells were washed with DPBS.

2.6.3 MRNA synthesis and transfection

2.6.3.1 MRNA synthesis and transfection

PGBD5 and HA-PGBD5 mRNAs were synthetized from plasmid with the HiScribe T7 ARCA mRNA Kit including tailing (New England Biolabs). The template plasmids were pcDNA3.1-PGBD5 or pcDNA3.1-HA-PGBD5. Both contained a T7 promoter. The mRNA synthesis was conducted as recommended by the manufacturer. In brief, anti-reverse cap analog (ARCA)/NTP mix, template DNA, T7 RNA polymerase mix, and nuclease-free water were mixed and incubated for 120 min at 37°C. After removal of DNA by DNase I treatment at 37°C for 15 min, a poly(A) tailing reaction was set up. The *in-vitro* transcription (IVT) reaction, 10x Poly(A) Polymerase reaction buffer, Poly(A) Polymerase and nuclease-free water were mixed and incubated for 60 min at 37°C in a total reaction volume of 20 µl. Finally, mRNA was purified by Phenol:Chloroform extraction and ethanol precipitation (as described in the next Chapter 2.6.3.2).

The mRNA was transfected using Lipofectamine MessengerMAX transfection reagent (Invitrogen) according to manufacturer's instructions. The transfected mRNAs were either HA-PGBD5, PGBD5 or for the control GFP mRNA (BIOZOL). MessengerMAX reagent was diluted 1:25 in Opti-MEM (Gibco). After 10 min of incubation, the mRNA was diluted in Opti-MEM. MessengerMAX and mRNA dilutions were mixed 1:1 and incubated for another 5 min at room

temperature. Finally, 50 µl reaction mixture was added to each 12-well with a total of 1 µg mRNA.

2.6.3.2 Phenol:Chloroform extraction and ethanol precipitation of RNA

To adjust the reaction volume to 200 µl, 180 µl of nuclease-free water were added, followed by the addition of 20 µl of 3 M sodium acetate, pH 5.5 (Life Technologies). The mixture was thoroughly mixed, and then subjected to one extraction with 1:1 phenol:chloroform mixture, followed by two extractions with chloroform. The aqueous solution was then transferred to a new Eppendorf tube. To precipitate the RNA, the sample was incubated with 2 volumes of ethanol at -80°C for 2 hours. The pellet was collected by centrifugation at 4°C and 15,000 rpm, and the supernatants were removed. The pellet was rinsed with 500 µl of ice-cold 70% ethanol and centrifuged for 10 min at 4°C. The ethanol was removed and the pellet was air-dried. Finally, the RNA was resuspended in 50 µl of 0.1 mM EDTA and stored at -80°C until further use.

2.6.4 Neuronal differentiation: Time course of PGBD5 expression

2.6.4.1 Cell lysis and Western Blots

HiPSCS were differentiated into glutamatergic neurons with or without astrocytes (see *Neuronal differentiation).* The cells were collected during the time course until day 15. Cell lysis, denaturation, Western Blot, and visualization were performed as mentioned before (Chapter 2.4.1) with the following modifications: Cells of one 12-well were lysed in 350µl lysis buffer containing 50 mM TRIS HCl pH 8.0, 10 mM EDTA, 150 mM NaCl, 5% glycerol, 1% NP-40 supplemented with protease inhibitors (Pierce), Phosphatase inhibitor (Active motif) and Benzonase (Novagen). Protein concentrations were measured in a BCA protein assay (Pierce). However, the BCAs failed because the final solution turned yellow instead of the typical violet. The BCAs with the same lysis buffer worked in HEK293, and the pH did not differ from the

standards. Without concentration adjustments, 5 x SDS sample buffer was added to the samples, and proteins were denatured at 99°C for 5 min. The gel (1.5 mm, 12% Acrylamide, stain-free, fast-cast gel, Bio-Rad) was loaded with 20 µl protein lysate. After gel-electrophoresis and transfer, the membrane was cut in pieces that were incubated with different antibodies. The top of the membrane was blocked in 5% milk in TBS-T, and the bottom (10-25 kDA area) was blocked in 2.5% BSA for 1h at room temperature while shaking.

The primary antibodies as listed in Table 10, were incubated for 60 min at room temperature. The membrane was washed three times for 10 min each in TBS-T. In the next step, the secondary antibody was diluted in solution (Table 10) and added to the membrane. The membrane was washed as described above, and the bands were visualized using ECL (Amersham) detection method.

Table 10: List of antibodies used in neuronal differentiation experiments.

1st AB	Dilution	Company	2nd AB	Dilution	Company	Solution
α TOP2B (ab72334)	1:5000	Abcam	HRP Anti-rabbit IgG	1:6667	ThermoFisher Scientific	5% milk in TBS-T
α PGBD5 (7-F8-5)	1:5000	Novusbio	Goat anti-Mouse IgG	1:10.000	ThermoFisher Scientific	5% milk in TBS-T
α yH2AX (ab81299)	1:2500	Abcam	HRP Anti-rabbit IgG	1:10.000	ThermoFisher Scientific	2.5% BSA

2.6.4.2 Quantifications of protein amounts from Western Blots

The Western Blot band intensities were normalized and analyzed with ImageJ (version 1.53k). The preprocessing steps are similar to the ones described in 2.5.3. Briefly, the background was substracted from the raw images (*tiff* file format) with rolling ball radius of 50 pixels. Afterwards,

rectangular ROIs were selected, and the band intensities were measured. The *Raw Intensity Density* and *Intensity Density* were calculated by the formulas described in 2.5.3.
Afterwards, the band intensities were normalized to the protein load or Actin.

$$NormIntDen_{\ protein\ x} = IntDen_{protein\ x} * \frac{IntDen_{protein\ load\ 1st}}{IntDen_{protein\ load\ x}}$$

With *IntDen* … Intensity Density and
NormIntDen … normalized intensity density

For some plots the intensities are given in percentages. Here I normalized as follows:

$$Int\ in\ \%_{protein\ x} = NormIntDen_{protein\ x} - \frac{NormIntDen_{protein\ min}}{NormIntDen_{protein\ max} - NormIntDen_{protein\ min}}$$

With *Int* … Intensity and
NormIntDen … normalized intensity density

The data was visualized plots employing ggplot2 in RStudio.

2.6.5 RNA extraction & RT-PCR

Cells were washed with DPBS and 500 µl Trizol (ThermoFisher Scientific) per 1 well of a 6 well-plate or 300 µl Trizol per 1 well of a 12 well-plate were directly applied to the cells. All the other steps of the extraction were as described in Chapter 2.4.3.
The RT-PCR was performed as described in Chapter 2.4.5. The expression of all genes was normalized to ACTB expression.

3 Results

3.1 Protein domain characteristics of the piggyBac-derived gene family

In the first part of the results section, we present an overview of the PGBD family protein domain architecture compared to piggyBac. The 3-D structure of piggyBac has been recently determined by the Fred Dyda lab (Chen *et al.*, 2020) and is now accessible in various databases. Utilizing this structure and the subdomain categorization that Chen *et al.* (2020) established, we evaluated evolutionary constraints acting on these domains in the piggyBac-derived human gene family, PGBD1-5.

3.1.1 Protein domain annotation

We used two algorithms for annotating protein domains, HMMER (Finn *et al.*, 2011) and Phyre 2 (Kelley *et al.*, 2015). HMMER is a widely used and effective homology search algorithm that utilizes *hidden Markov models* (HMM) to generate sequence profiles. It searches the sequence of interest against a database of HMM profiles. The method allows the identification of similarities of even distantly related sequences. Phyre 2 is a tool that predicts and analyses protein structures. It performs a comprehensive analysis by comparing a protein sequence of interest with a vast sequences database. This process involves constructing evolutionary and statistical profiles for the sequence, which is scanned against a database of profiles representing known structures. In contrast to HMMER, Phyre 2 generates alignments based on structural homology searching rather than sequence homology.

We screened the human PGBD1-5 genes against the structural database using the Phyre 2 server and successfully aligned them to the piggyBac transposase structural template (c6x68D). The

analysis resulted in high-confidence scores for all five proteins. Hence, the piggyBac-like fold is the most likely 3-D-structure of all members of the human PGBD family. Figure 3 schematically illustrates these alignments and highlights the conserved protein subdomains that were identified through this analysis.

Alignment scores were: PGBD1: confidence 100%, coverage: 47%, identities 20 %, PGBD2: confidence 100%, coverage: 74%, identities 20 %, PGBD3: confidence 100%, coverage: 74%, identities 20 %, PGBD4: confidence 100%, coverage: 79%, identities 24 %, PGBD5: confidence 100%, coverage: 74%, identities 18 %.

Interestingly, all PGBD proteins have a distinct NTD that is not part of the transposase IS4 domain. The transposase IS4 domain stretches from the beginning of the first DDBD to the end of the CRD. However, PGBD1 and PGBD5 showed truncated IS4 domains that lack CRDs. Additionally, we found that the catalytic motif DDD is mutated in all proteins except PGBD4.

In our analysis, PGBD3 and PGBD4 resembled the piggyBac transposase most closely regarding protein domain and TIR integrity. In fact, their DNA substrates are still flanking the gene locus, with MER85 flanking PGBD3 and MER75 flanking PGBD4. On the other hand, PGBD5 was the most aberrant in terms of TIR integrity, conservation of catalytic triad, and CRD domain integrity. Furthermore, we observed an Alu insertion within the 3’ TIR of PGBD4, separating it into two parts, which resulted in a shortened TIR remaining at the transposon end. PGBD1 had additional N-terminal SCAN and KRAB-like domains, commonly found in the large class of zinc finger transcription factors. Although not illustrated here, PGBD3 had integrated within the ERCC6 gene locus.

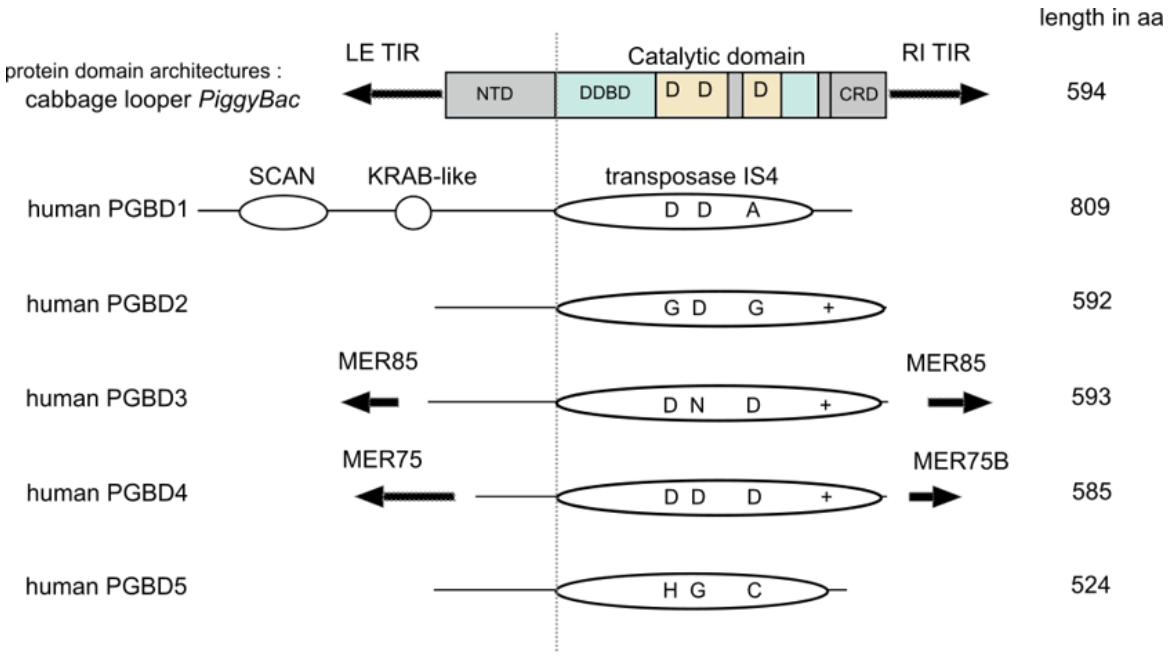


Figure 3: Schematic comparison of protein domain architecture between the human PGBD family and piggyBac. The top section shows the domain architecture of piggyBac. TIRs (Terminal inverted repeats) are indicated by arrows. The protein subdomains include NTD (N-terminal domain), DDBDs (DNA-binding and dimerization domains) shown in light blue, catalytic domains shown in yellow, an insertion domain that separates the catalytic domain into two parts, and a CRD (cysteine-rich domain). The aligned human PGBD1-5 sequences are presented below. The transposase IS4 domains were aligned with piggyBac. The conservation and substitution of the catalytic triad DDD are indicated within the catalytic regions of the proteins. The symbol '+' indicates the structural conservation of the CRD domain. The protein length in amino acids is illustrated on the right end.

The high-confidence structural alignments resulting from a screen against a massive database of 3-D structures identified the piggyBac structural template as the best hit against human PGBD5, demonstrating that PGBD5 has a piggyBac-like fold. However, inspecting the domain architecture more closely, we find that PGBD5 is unlikely a catalytically active transposase, as assumed in several studies (Helou, Beauclair, Dardente, Arensburger, *et al.*, 2021; Helou, Beauclair, Dardente, Piégu, *et al.*, 2021; Henssen *et al.*, 2015, 2016; Henssen, Koche, *et al.*, 2017; Henssen, Reed, *et al.*, 2017; Jubierre Zapater *et al.*, 2023). We used structural alignments in contrast to previous studies that generated sequence-based alignments (Bouallègue *et al.*, 2017; Helou, Beauclair, Dardente, Piégu, *et al.*, 2021; Henssen *et al.*, 2015). Structural alignments are

more reliable for distantly related proteins with low levels of sequence identity (Carpentier *et al.*, 2019). Our alignment showed that all three residues of PGBD5's catalytic domain mutated from DDD to HGC. After examining the alternative motif proposed by (Henssen *et al.*, 2015), we found that the first residue of this alternative catalytic triad, D168 (D249 in PiggyBac), is located within the DDBD and not in the catalytic domain of the protein. Because human PGBD5 lacks the CRD domain and an intact DDD catalytic motif, it is improbable to be an active transposase.

3.1.2 Evolutionary constraints on subdomains

K_a/K_s ratios estimate evolutionary constraints acting on protein-coding genes. They are calculated as the ratio of nonsynonymous substitutions per nonsynonymous site (K_a) over synonymous substitutions in synonymous sites (K_s) in a given time (Hurst, 2002). A K_a/K_s ratio of 1 indicates neutral selection. Values greater than 1 indicate positive selection meaning that mutations in the sequence had benefited the host. In contrast, values below 1 indicate purifying selection meaning that mutations in that gene have rather negative consequences for the host. Functional genes show overall purifying selection.

To further investigate the functional relevance of the PGBD genes, we examined the evolutionary forces acting on their entire sequences and subdomains. For this purpose, we used PAML (Yang, 2007) and analyzed 10-20 relevant mRNA sequences (CDS) per gene. We used PGBD3/ERCC6 fusion transcripts with N-terminal ERCC6 host domains.

The analysis revealed that all five genes are subject to purifying selection, indicating that they are functional and essential for the host. However, the two genes harboring additional N-terminal domains, PGBD1 and PGBD3, showed relatively weaker purifying selection, especially in their newly acquired N-terminal regions (PGBD4 NTD ~0.37 and PGBD3 NTD ~0.43). This finding suggests that the proteins underwent modest adaptation after fusion with the host domains. The adaptation primarily occurred through modifications of the N-terminal host domains rather than the transposase domains. All in all, we find evidence that the integrity of the transposase domain is vital for protein function.

Furthermore, we observed that all genes showed strong purifying selection on their transposase-derived, with the strongest selection in PGBD5 (Figure 4). This finding highlights the critical role of this domain in the function of all PGBDs.

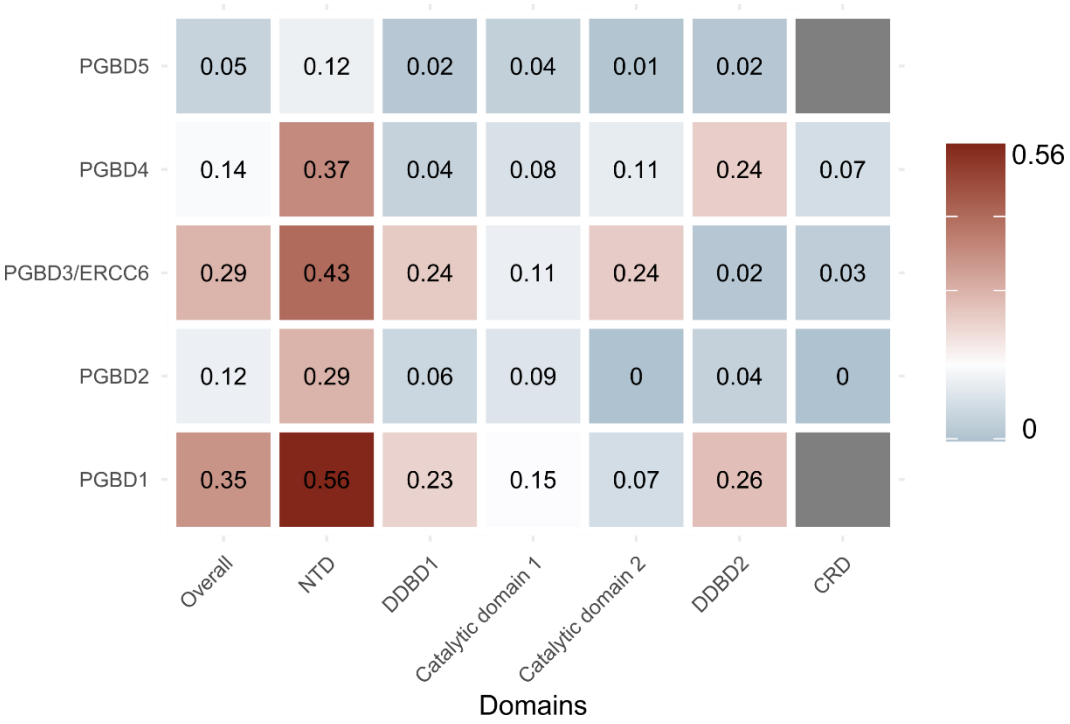


Figure 4: Evolutionary forces acting on the human PGBD gene family and their subdomains. The color scale represents K_a/K_s ratios, with corresponding labels indicating different levels of selection pressure. Blue signifies robust purifying selection, while red means a relatively weaker one. X-axis: Protein subdomains: Overall, NTD (N-terminal domain), DDBD1/2 (DNA-binding and dimerization domains 1 & 2), catalytic domains 1 & 2, and CRD (cysteine-rich domain). Y-axis: PGBD1-5 gene family.

Altogether, the study analyzed the protein domain and subdomain architectures compared to the insect piggyBac and compared them along their own phylogenetic trees.

PGBD3 and PGBD4, the evolutionary youngest genes, resemble the piggyBac protein domain architecture most closely (Figure 4). They not only have high-confidence structural alignments to all necessary protein domains for transposition, but their genes are also flanked by specific TIRs (MER85 and MER75, respectively). Plus, PGBD4 is the only gene with a conserved

catalytic DDD triad. In contrast, PGBD5 lacks a CRD, TIRs, and the catalytic DDD motif (Figure 4) and is most likely not transposing. That is in line with the transposition assay results of Beckermann *et al.* (2021) but contrary to those of Henssen *et al.* (2015) and Henssen, Reed, *et al.* (2017).

Exploration of evolutionary forces using K_a/K_s ratios on the PGBDs and their subdomains revealed that they are overall under purifying selection, indicating their functionality. However, those forces were somewhat weaker in the N-terminal regions, especially in PGBD1 and PGBD3, suggesting that they might have acquired more functional modifications over evolution than their transposase-derived domains.

3.2 Evolutionary roots of PGBD1

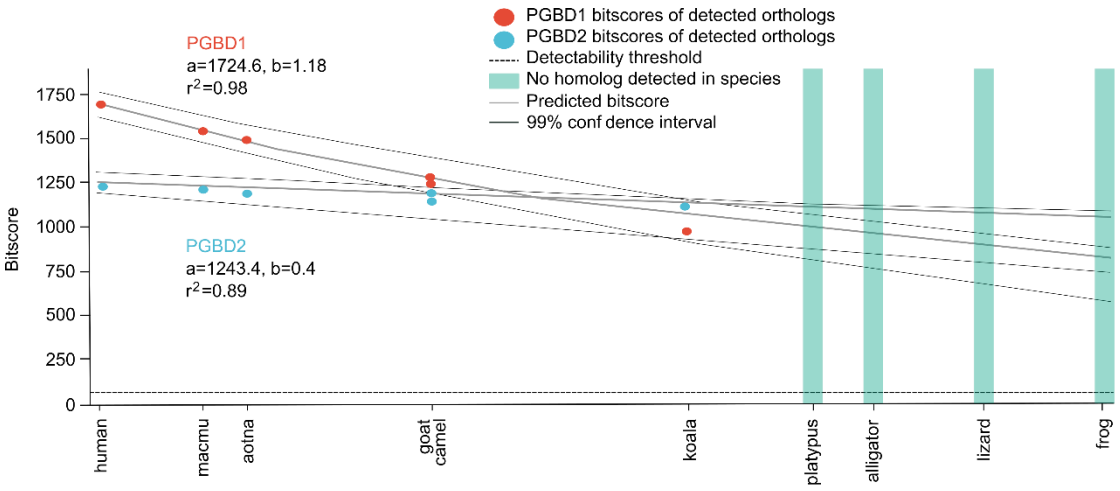
One of our Lab's main focuses is the study of gene functions of human transposase-derived genes, particularly the human PGBD family. We extensively studied PGBD1, which co-evolved with the mammalian-specific paraspeckle structure and regulates this complex by suppressing NEAT1 lncRNA in NPCs (Raskó *et al.*, 2022). Within the scope of this publication, I dissected the evolutionary origin of PGBD1 and its protein domain architecture. A summary of Raskó *et al.* (2022) findings can be found in the discussion Chapter 4.1.

3.2.1 PGBD1 and PGBD2 are mammalian specific genes

Previous studies have reported that PGBD1 is specific to mammals based on homology searching (Bouallègue *et al.*, 2017). However, inferring taxonomic presence or absence through homology searching may have limitations. To overcome this limitation, Weisman *et al.* (2020) developed a method called abSENSE, to determine whether the lack of detectable homology is likely due to the failure of homology searching. The tool requires alignment scores of the protein of interest in at least two species and pairwise evolutionary distances between selected species of interest

as input. AbSENSE estimates alignment scores for all species chosen with unknown homologs by a linear regression model and uses confidence intervals to assess the probability of homology detection failure.

Therefore, we used curated common orthologues from the BUSCO database (Simão *et al.*, 2015) to calculate the evolutionary distances between species (described in Chapter 2.1.2). These distances served as a basis for determining the evolutionary substitution rate of the protein of interest, which was derived from the alignment scores of its known homologs.



Species	Homolog bitscore	
	PGBD1	PGBD2
Human	1693.0	1241.0
macmu	1608.0	1233.0
aotna	1543.0	1207.0
goat	1264.0	1199.0
camel	1349.0	1099.0
koala	1059.0	1081.0
platypus	No homolog detected	
alligator	No homolog detected	
lizard	No homolog detected	
frog	No homolog detected	

Species missing a detected homolog	Probability of homolog detection failure (E=0.001)		Predicted bitscore		99 percent conf dence interval	
	PGBD1	PGBD2	PGBD1	PGBD2	PGBD1	PGBD2
platypus	0.0	0.0	953.23	1018.54	805.26 - 1088.61	908.03 - 1127.61
alligator	0.0	0.0	919.99	1006.45	770.28 -1057.21	854.08 - 1119.63
lizard	0.0	0.0	843.68	977.55	689.82 - 984.95	814.25 - 1100.68
frog	0.0	0.0	767.91	947.08	611.97 - 911.53	908.03 - 1080.26

Figure 5: Homology detection failure analysis for PGBD1 and PGBD2. The X-axis displays selected species along with their corresponding relative evolutionary distances. The colored points on the graph represent the BLASTP scores, indicating the similarity between the protein of interest and its ortholog in the respective species. For each protein, the figure shows the linear regression model's best-fit values of a and b. R^2 values are Pearson correlations that reflect the accuracy of the fit.

Applying this method, we found that PGBD1 and PGBD2 are specific to non-monotreme mammals, consistent with the findings of Bouallègue *et al.* (2017) (PGBD1: probability of homolog detection failure = 0 (E = 0.001), 99% confidence interval, a = 1724.6, b = 1.18, r2 = 0.98; PGBD2: probability of homolog detection failure = 0 (E = 0.001), 99% confidence interval, a = 1251.7, b = 0.32, r2 = 0.78) (Figure 5). Altogether the results provide further evidence that PGBD1 and 2 are mammalian-specific genes.

3.2.2 Gains and losses of PGBD1's N-terminal domains

To explore the conjugation between SCAN, KRAB-like, and transposase-derived domains observed in PGBD1, we constructed a phylogenetic tree incorporating all transposase IS4 domain-containing sequences (~12k). We manually selected a subtree from this tree consisting of PGBD1, PGBD2, and some closely related sequences (Figure 6 and Figure S 1).

The phylogenetic tree revealed that PGBD1 and PGBD2 are mammal-specific, consistent with previous findings by Bouallègue *et al.* (2017). Furthermore, we observed high similarity between these two sequences (on average ~63% sequence similarity according to pairwise distance matrix calculated in UGENE), with their similarity extending beyond the boundaries of the transposase IS4 domain (Figure 9). The close relationship between PGBD1 and PGBD2, combined with their apparent simultaneous integration into the mammalian genome, suggests gene duplication or parallel integration.

However, during the domestication, PGBD1 acquired a SCAN domain that is absent from all PGBD2 orthologs (Figure 6), and there is no evidence of homology across the relevant region in any PGBD2 sequence (Figure 9).

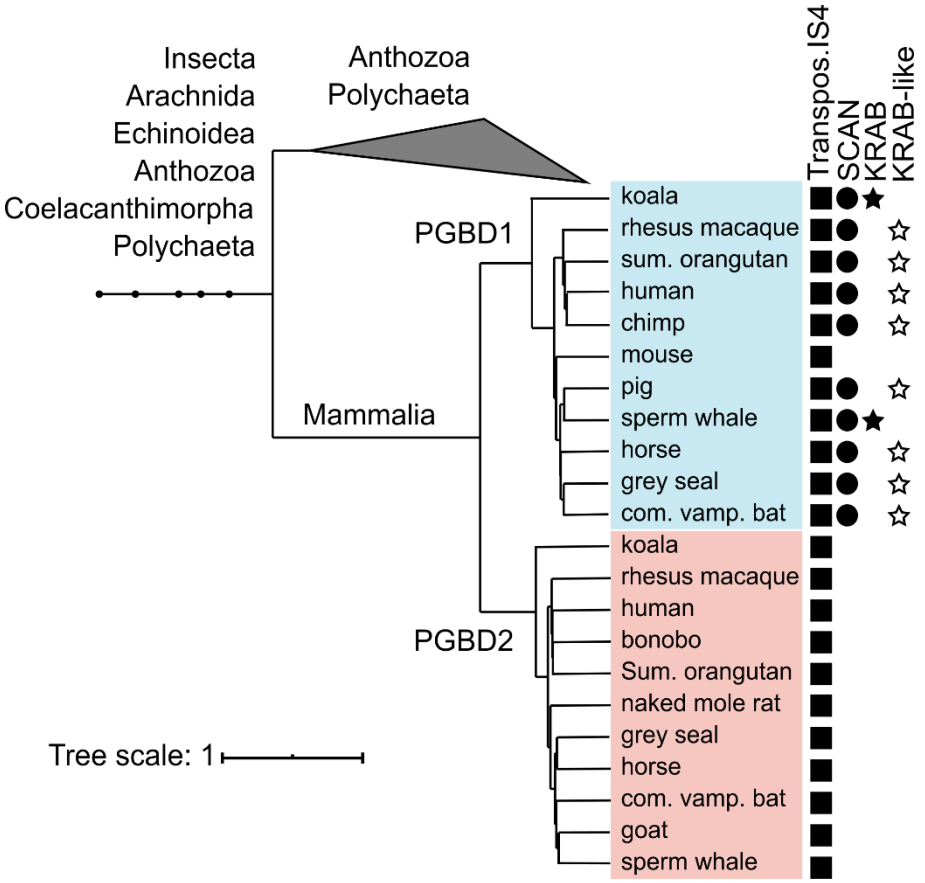


Figure 6: Phylogenetic tree of PGBD1 and PGBD2 showing the presence of transposase-derived, SCAN, and KRAB Domains. Human PGBD1 and PGBD2, along with closely related sequences containing the transposase IS4, were aligned using the MUSCLE alignment algorithm. The tree was constructed using *MrBayes*. Protein domains were annotated using *hmmerscan*. The KRAB-like domain was annotated using Phyre 2.

Acquisition of the SCAN domain likely occurred once in the common ancestor of eutherians and marsupials shortly after the duplication (or parallel integration) that gave rise to PGBD1. This is supported by the fact that PGBD1 is found in a genomic domain rich in SCAN domain proteins, suggesting gene fusion occurred after integration into this site.

In-silico analyses indicated that there had been multiple independent losses of the SCAN domain in different lineages, such as rodents, cats, grey lemurs, and some marsupials (Figure 6 & Figure S 1). To validate one such loss, we examined murine Pgbd1 in detail. Although we observed a relatively high homology between the transposase domains of the murine and human PGBD1 (87%), we could not observe homology to the SCAN domain by HMMER or Phyre 2 search, nor

could we observe homology to the SCAN domain within multiple sequence alignment when employing the annotated sequences of mouse and rat (Figure 7). The loss of the SCAN domain in rat Pgbd1 was also validated experimentally (Raskó *et al.*, 2022).

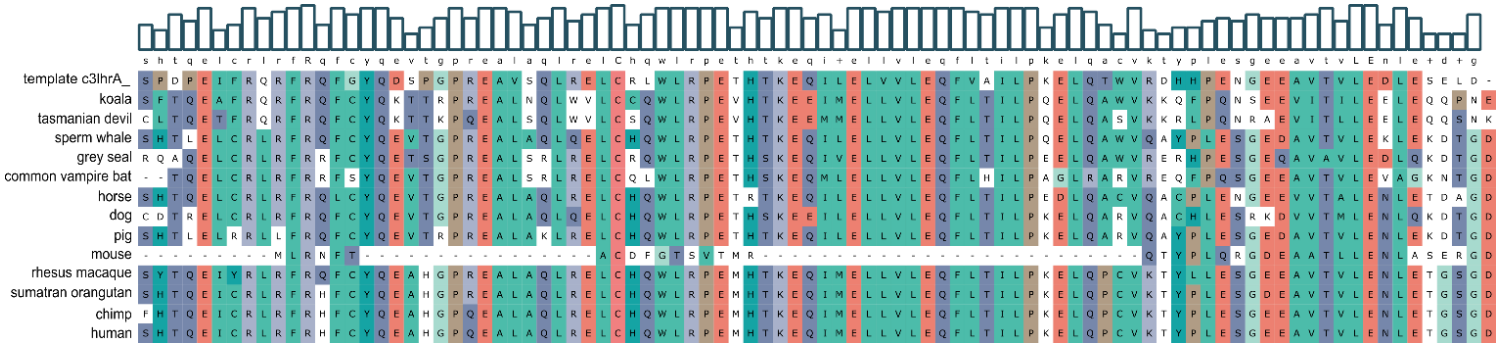


Figure 7: Alignment of selected PGBD1 sequences to the SCAN domain template. The first row of the alignment shows the best hit structural template (c3lhrA_) identified by Phyre 2, representing the SCAN domain of the ZNF24 gene.

In addition to the SCAN domain, we identified a KRAB domain in PGBD1. Automated protein domain search algorithms could only detect the KRAB domain in a few species, including marsupials (e.g., koalas) (Figure 6). This supports the hypothesis that KRAB inclusion is, like SCAN inclusion, the ancestral condition, but with numerous loss or decay events occurring in different lineages. When we expanded the KRAB domain annotation with Phyre 2 to structural rather than sequence similarity, we identified KRAB-like domains in many more species (Figure 6 & Figure 8), e.g., in human PGBD1 (region aa 211-267, KRAB structural template: d1v65a_, confidence score: 99.3%, identities: 16%).

While we find evidence that PGBD1 is somewhat structurally variable, we have extended the analysis of the subdomains of PGBD1. As mentioned earlier, our analysis of the PGBD1 gene indicated that it is mainly subject to purifying selection, suggesting it is an important functional gene. The overall K_a/K_s ratio of PGBD1 is <= 0.35 (Figure 9).

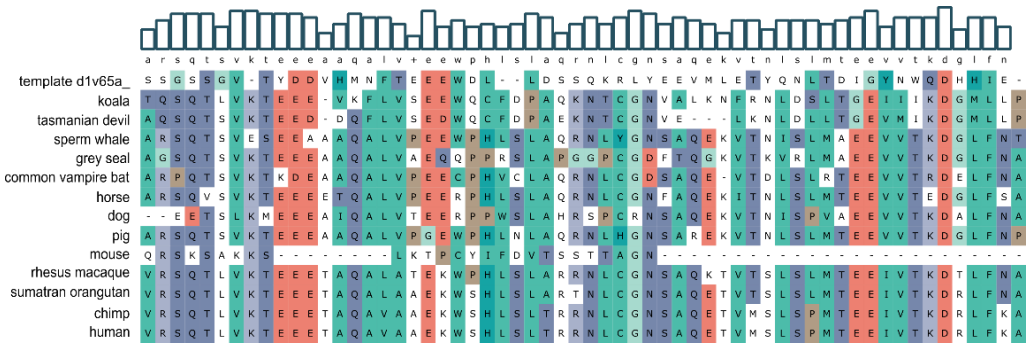


Figure 8: Alignment of selected PGBD1 sequences to the KRAB domain template. The first row of the alignment shows the best hit structural template (d1v65a_) identified by Phyre 2, representing the KRAB domain of the ZSCAN8 gene. The alignment includes sequences from diverse mammalian species, such as koalas and gray seals, where the KRAB domain was reported alongside other species.

However, we observed some variation in the KRAB domain of PGBD1. Our analysis of branch-specific ratios revealed that some organisms, such as horses, grey seals, vampire bats, and primates, had K_a/K_s ratios > 1 in this region. We investigated whether neutral or adaptive forces better account for the evolution of the KRAB-like domain. Therefore, we tested both hypotheses, revealing that adaptive evolution provides a better explanation ($chi^2 = 7.47$, *degrees of freedom* = 2, p-value = 2.38e-02). We repeated the analysis in sequences derived exclusively from primates since the problem of saturated synonymous sites could arise at large evolutionary distances. Again, we found that adaptive evolution explained the processes in this region significantly better ($chi^2 = 7.43$, *degrees of freedom* = 1, p-value = 6.42e-03). We identified two positively selected sites in this region, 227V and 252M, which may represent beneficial mutations. Of particular interest is position 227, which is a well-conserved residue in other KRAB domains and is important for binding the transposable element suppressor TRIM28 (Peng *et al.*, 2009; Tycko *et al.*, 2020). The functional significance of 252M is less clear, as it is not well-conserved in other KRAB domains. These results suggest that PGBD1 is a functional gene subject to purifying selection overall. However, the KRAB-like domain is subject to positive selection in some lineages for reasons that require further investigation.

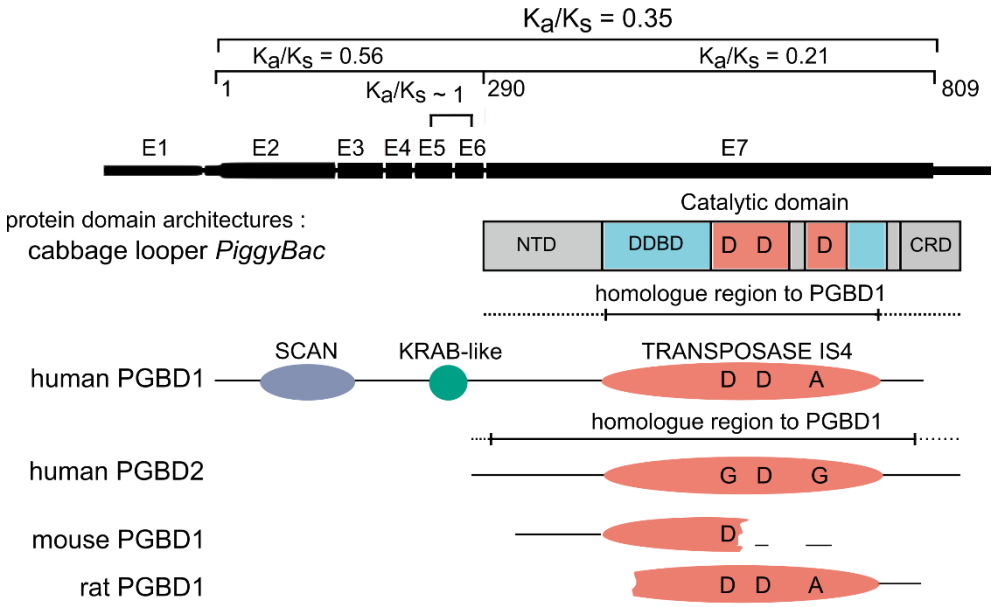


Figure 9: Summary of the protein domain analysis of PGBD1. The figure illustrates the domain structure of PGBD1 in comparison to PiggyBac, human PGBD2, rat PGBD1, and mouse PGBD1. The transposase-derived domain (IS4) consists of dimerization and DNA binding domains (DDBD) as well as the catalytic domains found in PiggyBac. Other domain abbreviations include NTD (N-terminal domain) and CRD (C-terminal cysteine-rich domain). The exons are represented as E1-7. Within the transposase-derived domains, the "D"s represent the catalytic triad DDD (D268, D346, D447), where D447 is replaced by (A) in PGBD1. PGBD1 and PGBD2 exhibit high similarity, with an average pairwise similarity score of approximately 63% in the aligned region, which extends beyond the annotated transposase IS4 domain (spanning 1324 bp). The PGBD1 sequences in rodent animal models are truncated, resulting in degenerated copies. The K_a/K_s values for the entire PGBD1 and various subdomains are shown. The KRAB domain exhibits a K_a/K_s value of approximately 1.

Altogether, the analysis provided further evidence that PGBD1 and PGBD2 are mammal specific based on phylogenetic relationships of the transposase IS4 protein family. PGBD1 had fused to a SCAN-KRAB domain in the ancestral condition, but later the KRAB domain diverged from the consensus KRAB, indicated by a positive selection of amino acid changes in this region. That might have compromised the TRIM28 binding capacity of that region.

3.3 Molecular functions of PGBD5

The first part of the thesis focused on the evolutionary history of the PGBD family, while the second part explores the molecular functions of domesticated DNA transposases, using PGBD5 as a case study.

3.3.1 Interactome of PGBD5

To investigate the molecular function of PGBD5, we reconstructed the PGBD5 interactome using *affinity purification* (AP) followed by *mass spectrometry* (MS). This method allowed the identification of whole protein complexes rather than binary protein-protein interactions and was applicable in mammalian cells. Specifically, we cultured HEK293 cells in either heavy or regular isotope-containing medium to achieve *Stable isotope labeling of amino acids in cell culture* (SILAC), a technique used to reduce false positives. The use of HEK293 cells is advantageous over other cell lines because they efficiently incorporate those isotopes.

We conducted two APs with HA-tagged PGBD5 as bait, differing from each other by the isotope-culture medium used for background and signal, referred to as forward and label-swap experiments. Statistical significance was determined by calculating p-values following the method described by Cox & Mann (2008) and adjusting them for multiple testing using the Benjamin-Hochberg method. We considered only proteins meeting the significance criteria, FDR < 0.05, in both APs as interactors. This approach provided a conservative statistical analysis, focusing solely on high-confidence interactors, as illustrated in Figure 10. However, a slightly less stringent threshold of a p-value < 0.05 was used for the gene set enrichment analysis, as shown in Figure 11.

The PGBD5 interactome revealed that PGBD5 might be involved in the epigenetic regulation of gene transcription, DNA-templated transcription, and DNA repair. Among the interactors, we identified several protein complexes that are essential transcriptional regulators, including the

histone methyltransferase complex-associated proteins (e.g., SIRT1, PELP1) and the *selective factor 1* (SL1) complex (e.g., UBTF, TAF1A, TAF1B, TAF1C).

Additionally, we identified *Phosphatidylinositol phosphate kinases* (PIP5K1A and PIP5K1C), which play essential roles in various cellular processes, including synaptic plasticity in the brain. Studies have shown that PIP5K1A and PIP5K1C regulate receptor-mediated calcium signaling (Vasudevan *et al.*, 2009), synaptic vesicle docking and fusion (Honigmann *et al.*, 2013), and synaptic plasticity (Hofbrucker-MacKenzie *et al.*, 2023; Unoki *et al.*, 2012), processes that are vital for proper synaptic function and brain activity.

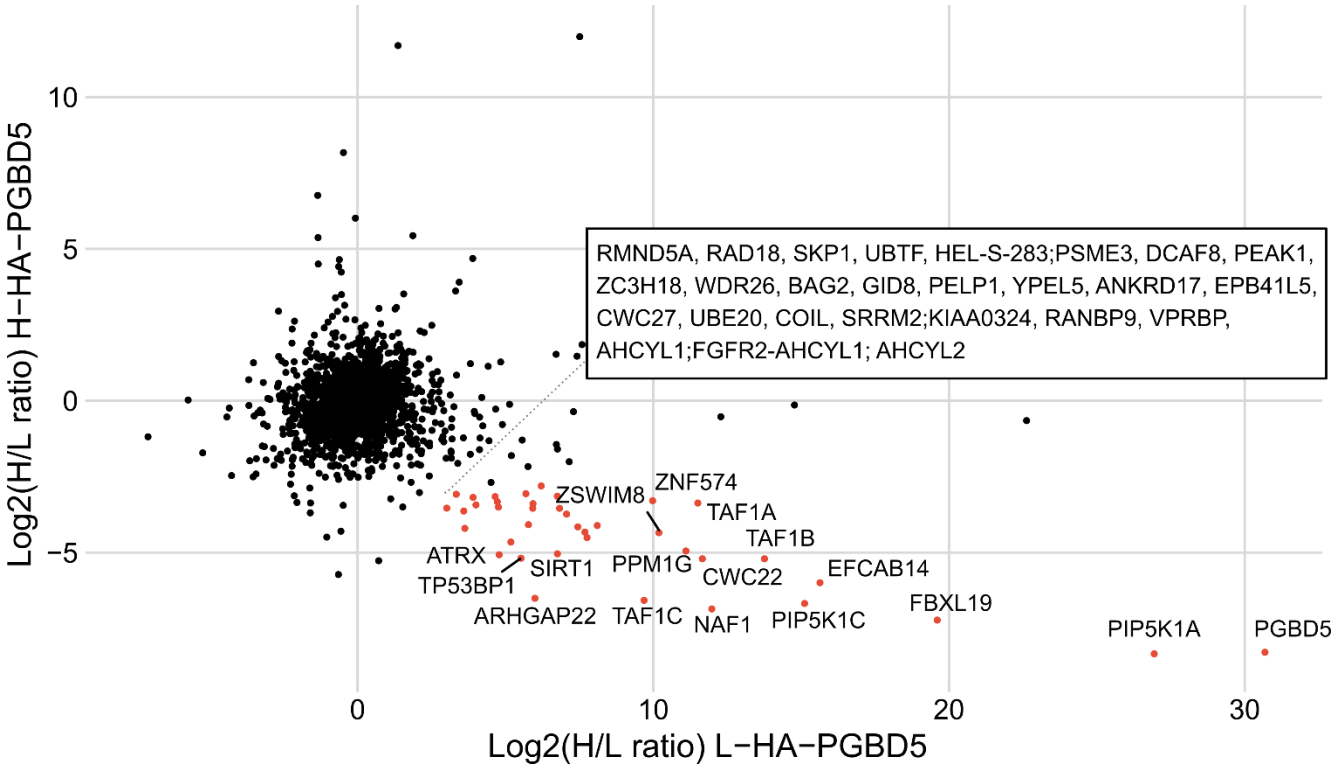


Figure 10: Interactome of PGBD5 revealed by SILAC-based AP-MS analysis. The protein interactors of PGBD5 were identified using two affinity purifications (APs), namely the forward and label-swap experiments. Significantly enriched proteins with an FDR < 0.5 are denoted by red dots and corresponding gene labels. In cases where the true origins of the detected peptides were ambiguous, genes are grouped and separated by a semicolon (;) for clarity.

We identified enriched pathways and protein complexes by performing gene set enrichment analysis using the significant genes as a signal and all detected genes with at least one unique peptide as a background. The results emphasized the role of PGBD5 in regulating gene transcription indicated by enriched gene sets such as *Epigenetic regulation of gene expression* and *RNA polymerase I transcription termination* (Figure 11).

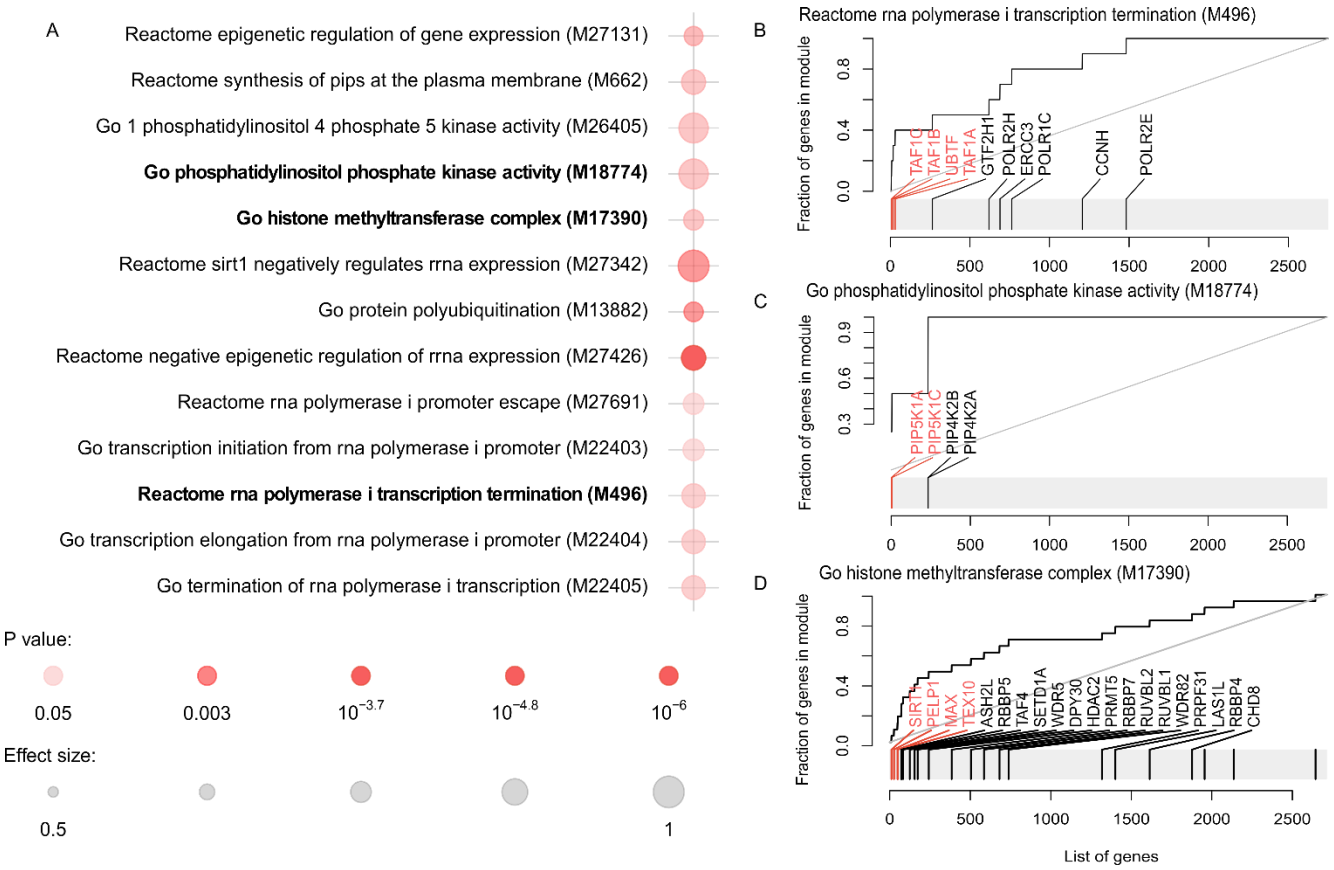


Figure 11: Gene set enrichment analysis of the PGBD5 interactome. A) Enriched gene sets were evaluated using tmod. The red color intensity reflects the significance level (P-value), while the dot size indicates the effect size, as denoted by the fold enrichment of the respective pathway. B-D) Evidence plots of selected gene sets: B) Reactome: RNA polymerase 1 transcription termination. C) GO: Phosphatidylinositol phosphate kinase activity. D) GO: Histone methyltransferase complex. In the evidence plots, the X-axis represents the sorted genes, while the Y-axis represents the fraction of the genes. Genes from the PGBD5 interactome showing significant enrichment (P-value < 0.05) are highlighted in red.

Significant interactors within the PGBD5 network analysis included several genes associated with DNA repair processes. These genes comprised *Tumor protein p53 binding protein 1* (TP53BP1), *Sirtuin-1* (SIRT1), *Alpha thalassemia/mental retardation syndrome X-linked* (ATRX), *RAD18 E3 ubiquitin-protein ligase* (RAD18), and *Damage-specific DNA binding protein 1* (DDB1) (Figure 12).

TP53BP1 is known to promote the end-joining of distal DNA ends during DSB repair (Panier & Boulton, 2014) and serves as a marker for active DNA repair through the NHEJ pathway (Weber Boutros *et al.*, 2022). *RAD18*, an E3 ubiquitin ligase, facilitates post-replication repair and is associated with homology-directed repair (Nambiar *et al.*, 2019). Notably, RAD18 suppresses the localization of TP53BP1 to DSBs, indicating a competitive relationship between these two repair pathways (Nambiar *et al.*, 2019). *SIRT1*, through its deacetylation activity, participates in various cellular processes, including NHEJ DNA repair by deacetylating XRCC6 (Jeong *et al.*, 2007). *ATRX*, a chromatin remodeling protein, plays a crucial role in heterochromatin formation and is essential for extended DSB DNA repair during homologous recombination (Juhász *et al.*, 2018; Teng *et al.*, 2021). Additionally, *DDB1* is a component of the E3 ubiquitin ligase complex CRL4, facilitating nucleotide excision repair (Lans *et al.*, 2019). Moreover, DDB1 has been implicated in other DNA repair pathways, such as NHEJ (Feng *et al.*, 2021) and homologous recombination repair (Moss *et al.*, 2010).

These findings highlight the involvement of PGBD5 with various DNA repair mechanisms, emphasizing a connection of PGBD5 to DNA DSBs and their repair.

While most of these complexes were identified by gene ontology enrichment analysis of the MSigDB database, we manually added a custom set *DNA repair* to our analysis because of its known association with PGBD5 (Figure 12).

Taken together, our PGBD5 interactome analysis reveals a regulatory role of PGBD5 in gene transcription. Moreover, the interaction of PGBD5 with DNA repair proteins further reinforces its association with DNA damage processes.

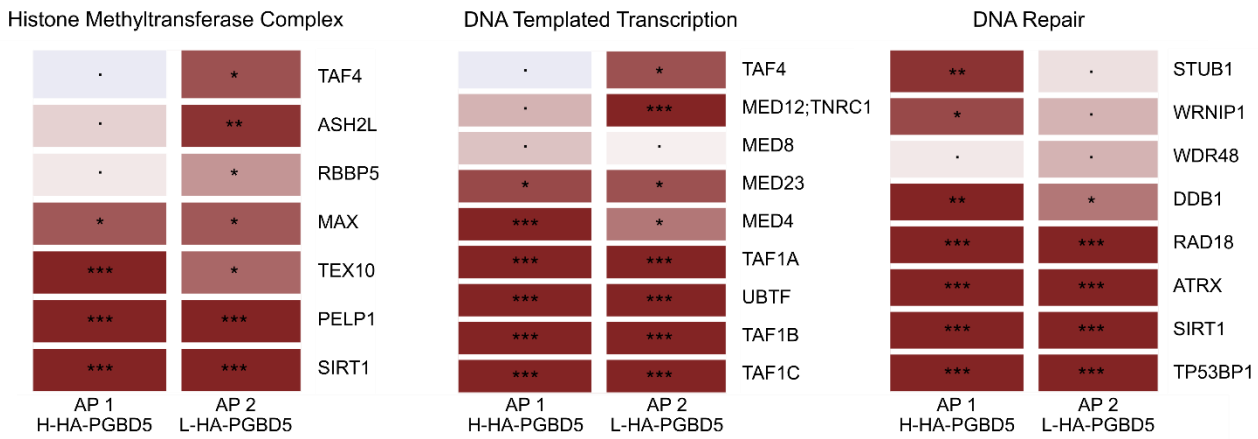


Figure 12: Interactors and significance levels of three selected gene sets. Left) Histone methyltransferase complex. Middle) DNA templated transcription. Right) DNA repair. The significance levels are indicated by color, and asterisks denote the corresponding level of significance. Asterisks denote p-values (see Table of abbreviations)

3.3.1.1 PGBD5 Interacts with TOP2A

In the search for interaction partners of PGBD5 that might facilitate DNA DSBs, we identified *topoisomerase II α* (TOP2A) as a top candidate. Although TOP2A was insignificant in the SILAC-based AP-MS experiment, it was close to the significance threshold and significant in one of the two APs (Figure 13).

We verified the interactome using an HA-tagged overexpression construct in combination with an HA antibody and an endogenous PGBD5 antibody (Figure 14). We successfully validated an interaction of PGBD5 with *Upstream binding transcription factor* (UBTF), *TATA-box binding protein associated factor* (TAF1C), and TOP2A. However, all interactions showed only a small enrichment, indicating that these proteins may interact transiently or indirectly.

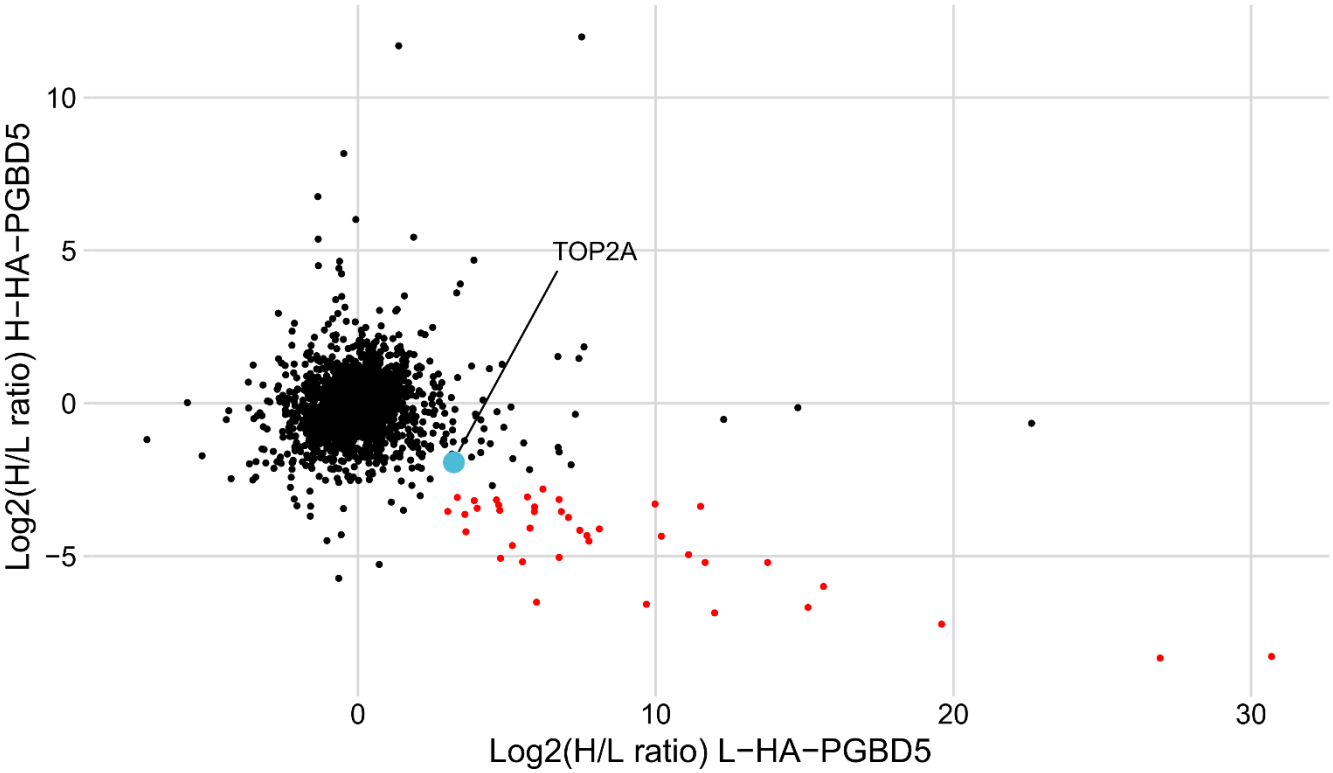


Figure 13: Potential interaction of PGBD5 with TOP2A revealed in SILAC-based AP-MS experiments. The protein interactors of PGBD5 were identified through two APs: the forward and label-swap experiments. Significantly enriched proteins with an FDR < 0.5 are represented by red dots. The figure highlights TOP2A, which exhibited significance solely in the forward AP and was not considered a high-confidence interactor.

Interestingly, a comparison of PGBD5 and *F-box and leucine-rich repeat protein 19* (FBXL19) interactomes revealed several shared interaction partners, including SKP1, PSME3, MED23, MED8, and MED12. The FBXL19 interactome was previously characterized by Dimitrova *et al.* (2018) in mouse *embryonic stem* (ES) cells, where it was found to play a role in priming genes for transcription during lineage commitment (Dimitrova *et al.*, 2018). We observed co-localization of FBXL19 with HA-PGBD5 but not with endogenous PGBD5 (Figure 14). However, it is important to note that we encountered several inconsistencies when using the endogenous PGBD5 antibodies, including low specificity that affected all experiments involving endogenous PGBD5 antibodies (as discussed in Chapter 4.2). Therefore, we cannot exclude the possibility of an interaction between FBXL19 and endogenous PGBD5 due to these limitations in antibody performance.

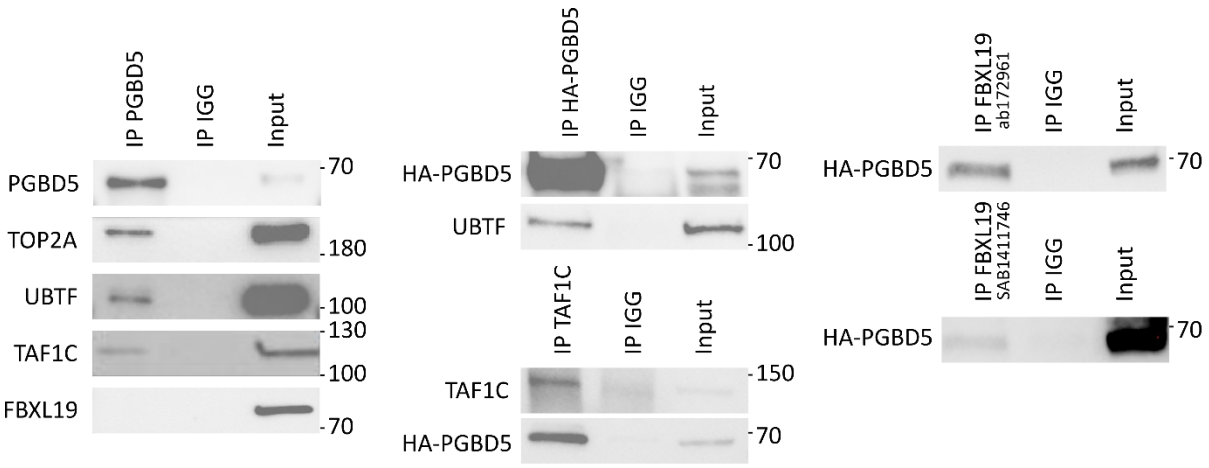


Figure 14: Validation of several interaction partner by co-IPs. Co-IP experiments were conducted in HEK293 cells, followed by Western Blot analysis using specific antibodies as indicated, along with an IgG control. The purified protein samples were probed with different antibodies as indicated to detect the presence of interactions. In the case of HA-tagged PGBD5, the protein was previously overexpressed from a transfected plasmid and subsequently pulled down.

3.3.1.2 PGBD5 localizes to the nucleoplasm and interacts with TAF1C on DNA

We speculated that the cellular localization of PGBD5 would reveal insights into its specific interaction with transcriptional regulators. Several PGBD5 interactors, such as *TATA-box binding protein associated factors* TAF1C, TAF1B, TAF1A, and UBTF, were found to localize to cell nucleoli (Kresoja-Rakic & Santoro, 2019). To investigate the cellular distribution of PGBD5, we conducted immunocytochemistry (ICC) staining (Figure 15). The staining revealed a uniform distribution of PGBD5 protein throughout the nucleus, in contrast to UBTF and FBXL19, which mainly localized to nucleoli. TAF1C showed an unexpected expression pattern. It did not localize to the nucleoli but was uniformly distributed throughout the nucleus. The limited overlap between PGBD5 and its interaction partners is consistent with the findings of the co-immunoprecipitation experiments. The experiment clarified that the interaction with

transcription-related proteins is not based on their common co-localization to nucleoli. We aimed to confirm the role of PGBD5 in transcription by investigating whether it interacts with transcriptional regulators at DNA.

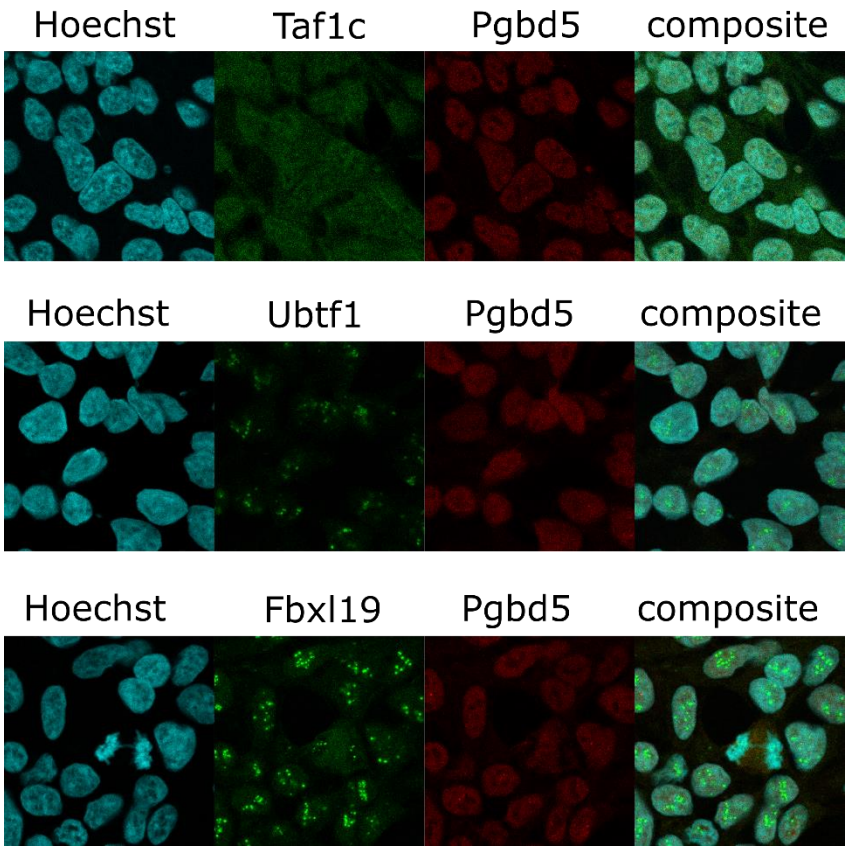


Figure 15: PGBD5 exhibits uniform nuclear localization and does not localize to nucleoli. Immunofluorescence staining was performed to visualize PGBD5 and its interactors. The nuclei were stained with Hoechst for reference. PGBD5 protein distribution throughout the nucleus was found to be uniform, while its interactors displayed distinct localization patterns.

We combined ChIP experiments with Western Blot (ChIP-WB) to achieve this. Our findings revealed a direct interaction between PGBD5 and TAF1C on DNA (Figure 16). In contrast, this assay observed no detectable interaction between PGBD5 and UBTF, FBXL19, or TOP2A. However, the reliability of our conclusions is limited due to the previously discussed issue with

the PGBD5 antibody, as outlined in Chapter 4.2. Consequently, further experiments utilizing an improved PGBD5 antibody or an HA-tagged construct are necessary to obtain conclusive results.

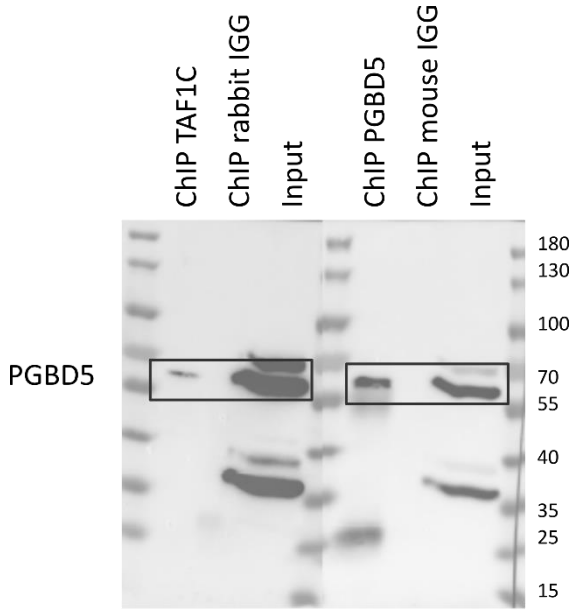


Figure 16: PGBD5 and TAF1C interact at DNA identified by ChIP-WB. ChIPs were performed, followed by Western Blot analysis using specific antibodies against endogenous TAF1C and PGBD5, along with a corresponding IgG control. The purified DNA samples were probed with a PGBD5 antibody to detect the presence of PGBD5-TAF1C complexes.

3.3.1.3 Summary

The PGBD5 interactome was comprehensively characterized using SILAC-based AP-MS. Among PGBD5 interactors, we identified proteins that are epigenetic regulators of gene expression, regulators of transcription, and DNA repair proteins. These interactions strongly suggest the role of PGBD5 in regulating gene transcription.

In particular, we aimed to identify PGBD5 interactors capable of inducing DNA DSBs, a phenomenon previously associated with PGBD5 (Henssen, Koche, *et al.*, 2017; Henssen, Reed,

et al., 2017; Jubierre Zapater *et al.*, 2023; Simi *et al.*, 2023). Interestingly, we discovered an interaction between PGBD5 and TOP2A, a known protein associated with DNA DSB induction. Co-immunoprecipitation experiments validated these interactions, confirming the presence of TAF1C, UBTF, and TOP2A as PGBD5 interactors.

Altogether the results suggest an involvement of PGBD5 in cellular processes such as (epigenetic) regulation of gene transcription and DNA repair. Further, we identified an interaction with TOP2A, thus implicating it as a potential candidate involved in the induction of DNA DSBs.

3.3.2 Transcriptome analysis of Pgbd5 knockout in mice

We further investigated PGBD5's molecular function by analyzing RNA-sequencing data of Pgbd5 knockout and wild-type mice in two brain regions, the cerebellum and the hippocampus. The raw data used in our analysis was generously provided by our collaborators at the Alex Kentsis lab. The genetically modified mice were generated by deleting the fourth exon of Pgbd5. We refer to Jubierre Zapater *et al.* (2023) for details about establishing germline knockout mice. Our analysis aimed to uncover the pathways impacted by Pgbd5 depletion, thereby enhancing our understanding of the underlying molecular mechanisms associated with Pgbd5.

3.3.2.1 Differential gene expression analysis

We evaluated the effectiveness of various normalization methods through the analysis *of Relative Log Expression* (RLE) plots (Figure S 3 & Figure S 5) and *Principal Component Analyses* (PCA) (Figure S 4 & Figure S 6). Our findings indicate that, except for TPM normalization, all other methods successfully normalized variation in RLE plots and effectively distinguished between knockout and wild-type mice in PCAs.

Utilizing Deseq2 (Love *et al.*, 2017) for differential gene expression analysis, which employed *regularized-logarithm transformation* (rlogs) for normalization, we observed a more pronounced impact of Pgbd5 knockout on gene expression in the cerebellum compared to the hippocampus. Specifically, we identified a significantly greater number of differentially expressed genes in the cerebellum (782 genes, FDR < 0.05 and absolute log2 fold change > 0.3) in comparison to the hippocampus (8 genes, FDR < 0.1 and absolute log2 fold change > 0.3). (Figure 17).

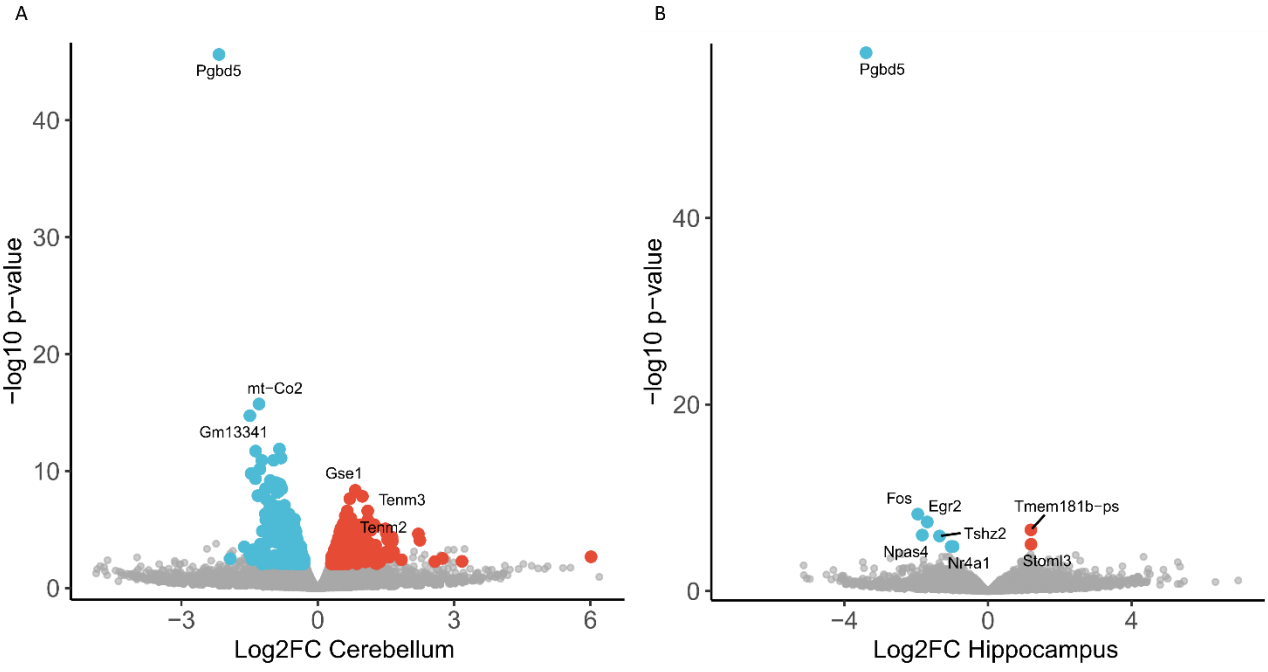


Figure 17: Transcriptomic analysis of Pgbd5 knockout mice in two brain regions, cerebellum and hippocampus. A) Volcano plot of 782 DEGs in the cerebellum. B) Volcano plot of 8 DEGs in the hippocampus. Red dots indicate significantly upregulated, and blue dots indicate significantly downregulated genes.

We found that within the hippocampus, the majority of the differentially expressed genes were downregulated IEGs such as Fos, *Early growth factor 2* (Egr2), Npas4, Nr4a1, and Dusp1, as well as DNA transcription factors including Fos, Nr4a1, Egr2, Npas4, and *Teashirt zinc finger homeobox 2* (Tshz2). Conversely, the upregulated genes in the hippocampus were the *pseudogene transmembrane protein 181b* (Tmem181b-ps) and *Stromatin Like 3* (Stoml3). However, it is worth noting that Stoml3 did not remain in the list of deregulated genes after

applying log foldchange shrinkage, suggesting that its upregulation may not be statistically significant. On the other hand, Dusp1 only exhibited significance after applying log foldchange shrinkage.

Among the 50 most deregulated candidate genes in the hippocampus, we found four more transcription factors *Proto-oncogene myeloblastosis* (Myb), *Zinc finger protein 366* (Zfp366), and *Proto-oncogene jun-B* (Junb), of which Junb is another IEG (Figure 18).

Among the top 50 deregulated genes in the cerebellum (Figure 18) we found several genes involved in cellular respiration (Cox6c, Ndufs4, Cox7b, Atp5l, Atp5k, Uqcrh, mt-Co2, mt-Atp6, mt-Nd3, mt-Cytb, and mt-Nd4l), two ribosomal genes (Rps24 and Rpl22l1), the JAK-STAT signaling pathway proteins (Il22, Iltifb), and transmembrane proteins (Tenm3 and Tenm2). The observed changes in gene expression suggest that the Pgbd5 deficiency in the cerebellum had a significant impact on cellular metabolism, translation, and signal transduction.

We conducted gene set enrichment analysis on all deregulated genes within the cerebellum, utilizing a ranked list approach implemented in the R tmod package (Zyla *et al.*, 2019). This analysis was performed against five databases from MsigDB: KEGG, GO *Biological Process* (BP), GO *Molecular Function* (MF), GO *Cellular Component* (CC), and Reactome pathways. Among the pathways that exhibited significant deregulation, we once again observed processes associated with cellular respiration, including *Aerobic Electron Transport Chain, NADH Dehydrogenase Complex Assembly*, and *Proton Motive Force Driven ATP synthesis*. Furthermore, we examined the KEGG database and identified significant enrichment of the Ribosome gene set. The third category of deregulated pathways was associated with neuronal cell signaling, specifically *Retrograde Trans Synaptic Signaling*, among others detailed in Table S 1. Figure 19 shows the enriched pathways of KEGG and GO BP and the share of either significantly up (red) or down (blue) regulated genes in that pathway. Table S 1 includes the results of all databases (with AUC > 0.75 and FDR < 0.05).

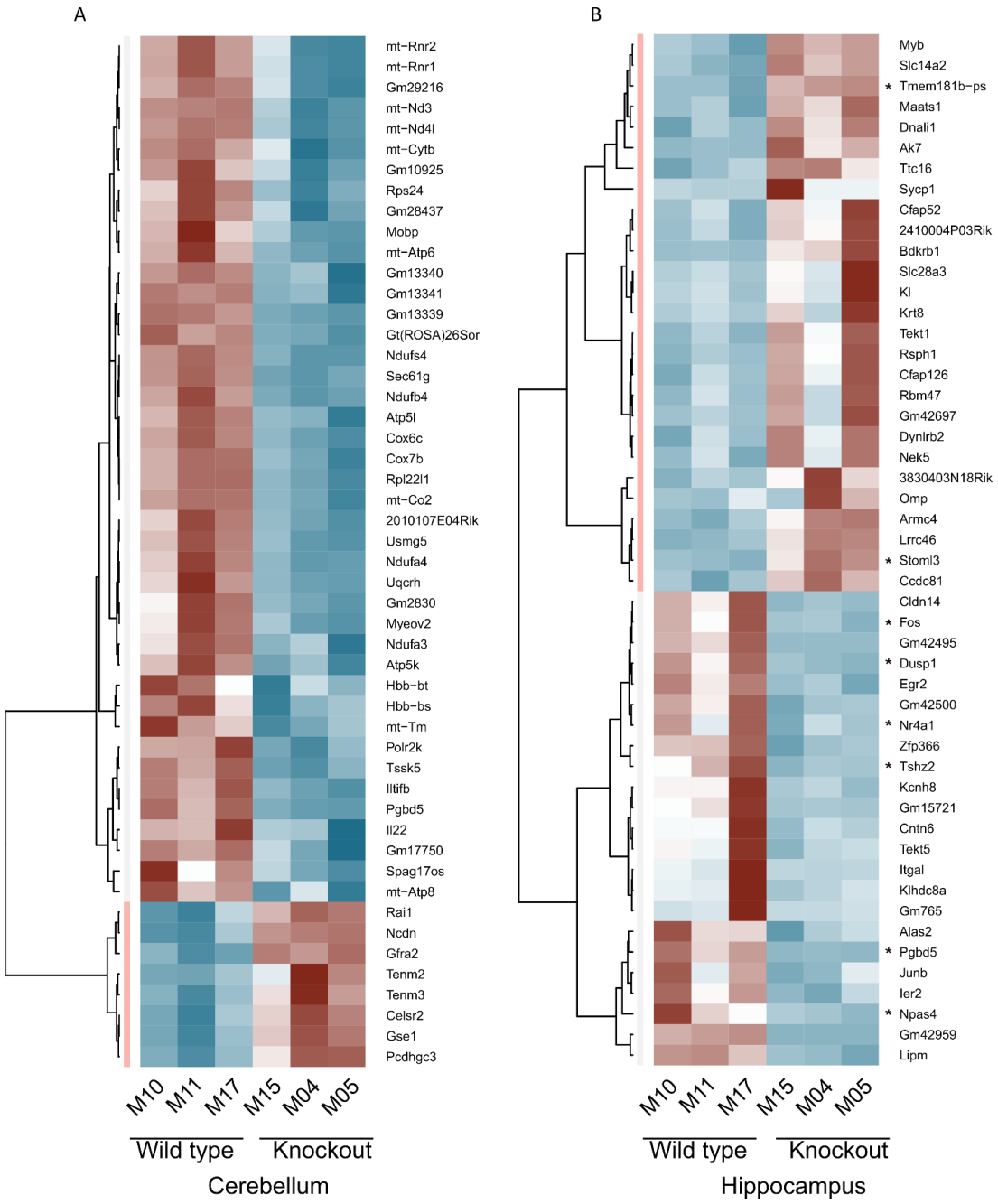


Figure 18: DEGs of Pgbd5 knockout mice in cerebellum and hippocampus. A) Heatmap of the top 50 DEGs in the cerebellum. All 50 genes show significant differential expression with a FDR < 0.05 and a log2FC > 0.3. B) Heatmap of the top 50 DEGs in the hippocampus. In contrast to the cerebellum, only a few genes exhibit significant deregulation with an FDR < 0.1 and log2FC > 0.3. Asterisks denote these significantly deregulated genes in the hippocampus. Color Scale: Red indicates upregulation, and blue indicates downregulation. Expression in rlogs was z-scaled.

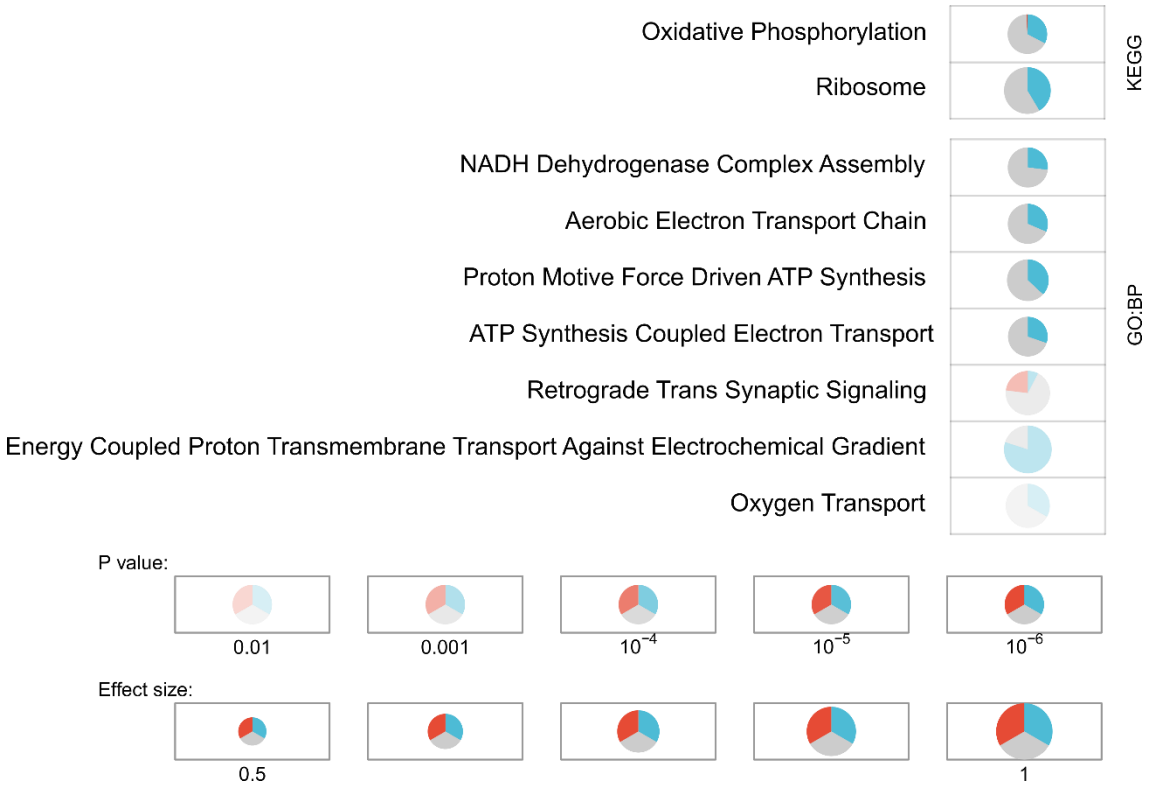


Figure 19: Gene set enrichment analysis. DEGs were sorted by FDR and tested against KEGG and GO Biological Process databases. The area under curve (AUC) represents the effect size of enrichment, indicating the strength of association between DEGs and the gene sets in the databases. The pie chart illustrates the distribution of DEGs. Blue represents downregulated genes, red indicates upregulated genes, and grey represents genes without significant differential expression.

In Figure 20 selected gene sets and their evidence plots are illustrated. The evidence plots demonstrate that many mitochondrial and ribosomal genes were significantly downregulated in the Pgbd5 knockout mice in the cerebellum.

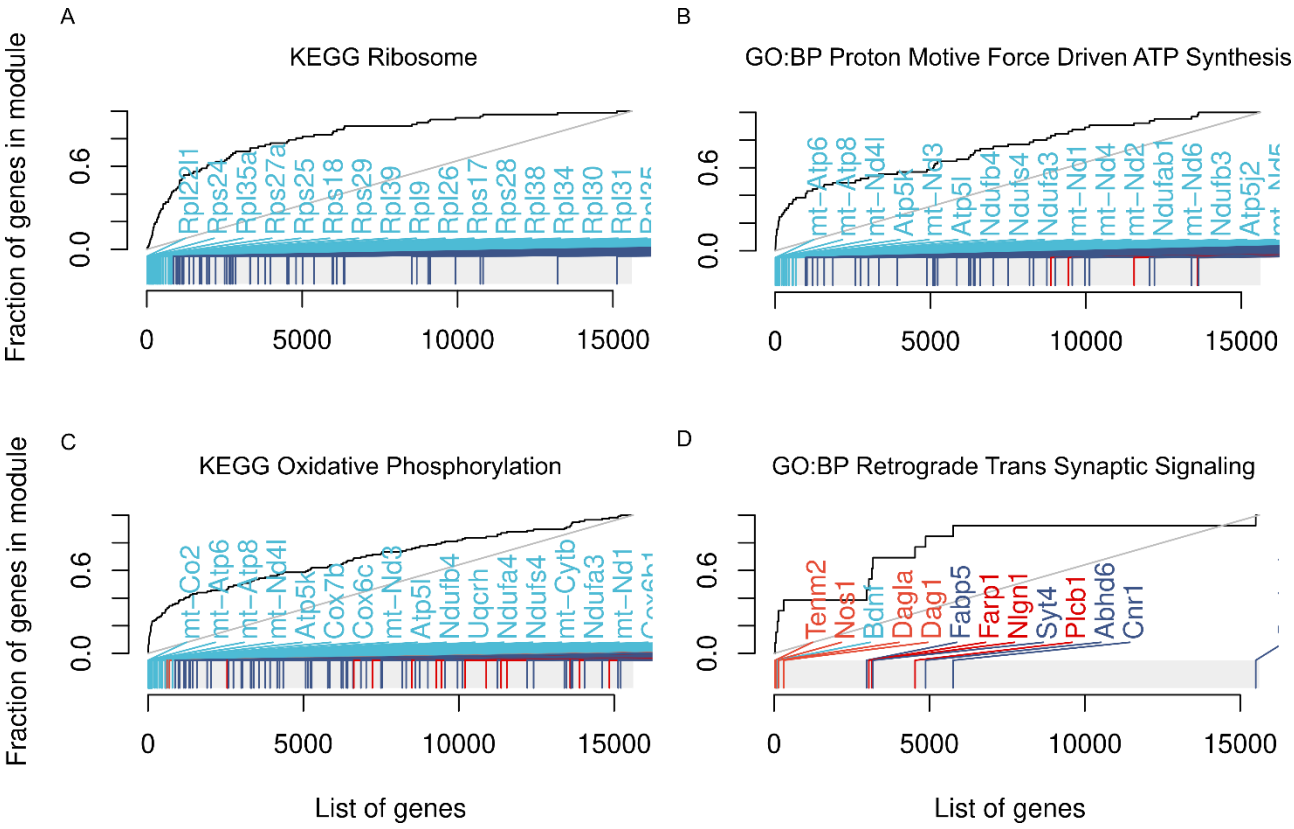


Figure 20: Evidence plots of selected gene sets. A) KEGG Ribosome. B) GO:BP Proton motive force driven ATP synthesis. C) KEGG Oxidative phosphorylation. D) GO:BP Retrograde trans synaptic signaling. In the evidence plots, the X-axis represents the sorted genes, while the Y-axis represents the fraction of the genes sorted by FDR. Blue represents downregulated genes, red indicates upregulated genes, light color represents significant differential expression and dark color indicates not significantly deregulated genes.

Further, we explored the potential similarities in gene expression changes between the cerebellum and hippocampus of Pgbd5 deficient mice (

Figure 21). To do so, we examined the correlation between log2fold changes in the cerebella and hippocampi after excluding low-count genes. The p-values of the deregulated genes in both datasets were combined, and a comprehensive p-value was calculated for each gene, represented in

Figure 21.

Overall, the hippocampus and cerebellum correlation was close to zero (-0.08, Pearson method). However, a handful of genes were differentially expressed in both tissues following Pgbd5 knockout. In addition to Pgbd5, several IEGs (Fos, Npas4, Nr4a1, and Dusp1), previously identified in the hippocampus, were significantly downregulated in both

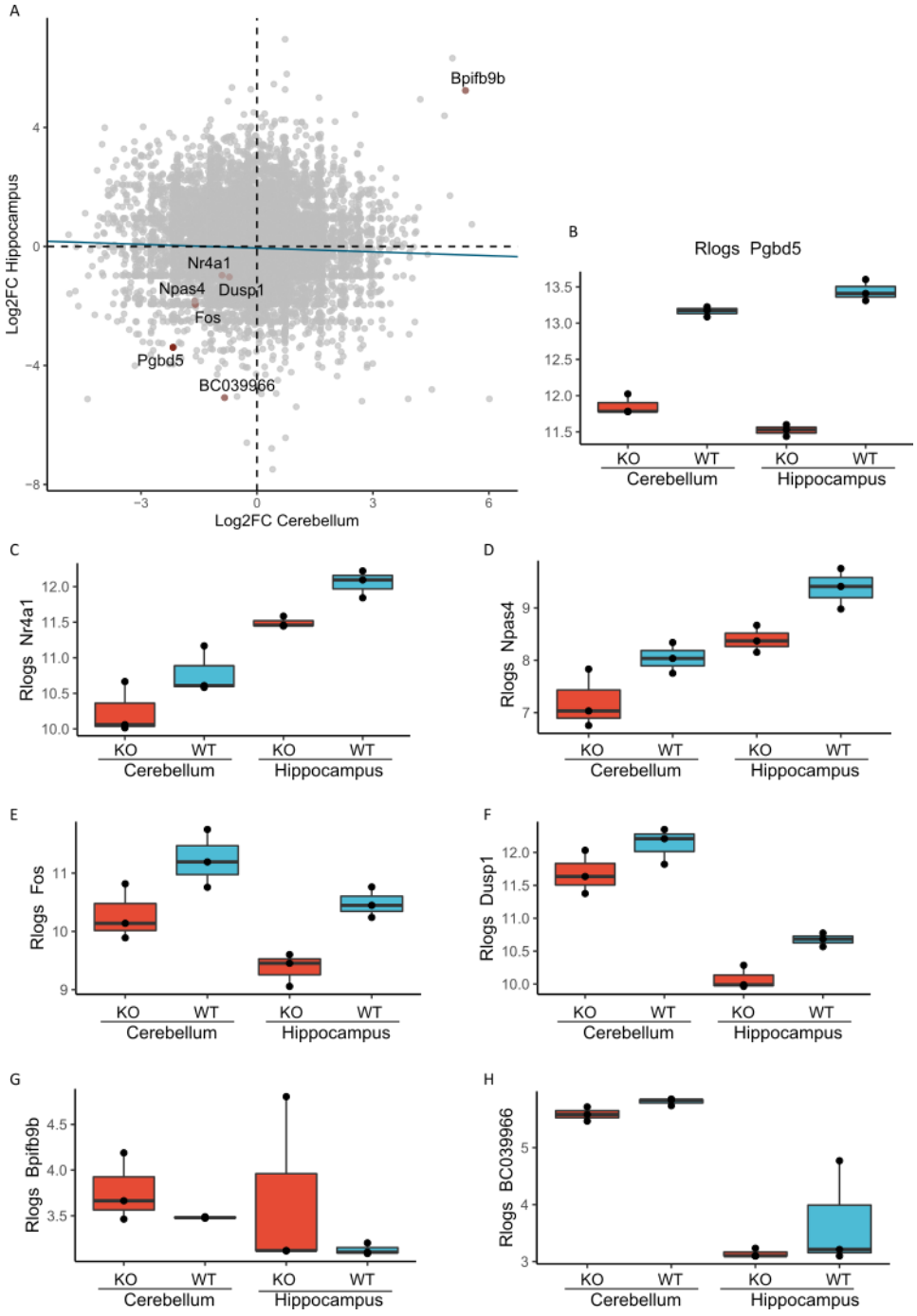


Figure 21: Common deregulated genes in cerebellum and hippocampus. A) The plot shows the correlation between the log2 fold changes of the common DEGs in both brain regions. The color indicates the significance thresholds, with red representing significant DEGs and grey showing non-significant DEGs. B - H) Expression profile of common DEGs. B) Pgbd5. C) Nr4a1. D) Npas4. E) Fos. F) Dusp1. G) Bpifb9b. H) BC039966.

tissues. Additionally, we found that the antisense lncRNA BC039966 was downregulated, while *BPI fold containing family B, member 9B* (Bpifb9b), was the only common upregulated gene. Boxplot diagrams in

Figure 21 show the differential expression pattern of each gene. While Pgbd5, Nr4a1, Npas4, Fos, and Dusp1 show convincing differential expression in both tissues, BC039966 and Bpifb9b appear significant, possibly due to outlier samples in the hippocampus.

3.3.2.2 Comparison of transcriptome and interactome

The following paragraph will discuss the transcriptomic changes considering the PGBD5 interactome identified in our previous experiment.

In the interactome analysis, the identified proteins were involved in transcriptional and epigenetic gene regulation (e.g., SIRT1, PELP1, TEX10, UBTF, TAF1A, TAF1B, TAF1C, MED4, MED23), DNA repair proteins (e.g., TP53BP1, SIRT1, ATRX. RAD18, DDB1) and phosphatidylinositol phosphate kinases (PIP5K1A and PIP5K1C).

Furthermore, we identified TOP2A, a type II DNA topoisomerase involved in DNA replication and transcription. Although the interaction did not pass the significance threshold in the SILAC-based AP-MS analysis, we successfully confirmed it through co-immunoprecipitation experiments. The commonly deregulated genes Fos, Npas4, and Nr4a1 are regulated by TOP2B and TOP2A (Herrero-Ruiz *et al.*, 2021; Madabhushi *et al.*, 2015; Stott *et al.*, 2021). While TOP2A is active in proliferating cells, TOP2B is ubiquitously present in all cell types, including post-mitotic cell types such as neurons. Therefore, Pgbd5 may regulate these genes with TOP2B, which is consistent with our observation of the deregulation of these genes in both the cerebellum and hippocampus of Pgbd5-deficient mice. The gene regulation of TOP2B is associated with the occurrence of DNA DSBs, which is also consistent with identifying DNA repair proteins that interact with PGBD5 in our interactome analysis.

The observed transcriptomic changes in ribosomal translation may be related to the interaction between PGBD5 and the RNA polymerase 1 pre-initiation complex (SL1 complex), which is composed of TAF1A, TAF1B, TAF1C, and UBTF. This complex is critical for the transcription of rRNA, and rRNAs are essential for the assembly of ribosomes. Therefore, the deregulation of ribosomal translation observed in Pgbd5-deficient mice may be attributed to the disruption of the SL1 complex and impaired rRNA transcription.

3.3.2.3 Summary

Altogether Pgbd5 knockout in mice showed different effects. While there were only a handful of deregulated genes in Pgbd5 deficient hippocampi, the Pgbd5 deficiency in cerebella caused the deregulation of mainly three essential pathways. Cellular respiration, ribosomal translation, and neuronal cell signaling were affected. However, IEGs such as Fos, Npas4, Nr4a1, and Dups1 were significantly downregulated in both tissues.

Overall, the comparison of the transcriptomic changes with the PGBD5 interactome suggests that PGBD5 is involved in the regulation of important biological processes. The interaction of PGBD5 with proteins involved in transcriptional and epigenetic gene regulation and DNA repair proteins suggests that PGBD5 may play a role in regulating gene expression and DNA damage response. The identification of TOP2A in the interactome (and the potential interaction with paralogue TOP2B), together with the downregulation of the IEGs Fos, Npas4, Nr4a1, and Dusp1, suggests that PGBD5 may be involved in the regulation of genes that are important for neuronal activity and plasticity. The interaction of PGBD5 with the RNA polymerase 1 pre-initiation complex suggests that PGBD5 may also regulate rRNA transcription and ribosome assembly.

However, it is important to note that the identified interactions may not all be direct or functional, and further validation is needed to confirm the functional relevance of these interactions. Additionally, it is possible that the observed changes in gene expression are indirect effects of PGBD5 deficiency, and further studies are needed to elucidate the exact mechanisms by which PGBD5 regulates these processes.

3.3.3 Knockout of PGBD5 in human iPSCs

So far, our results revealed that depletion of Pgbd5 in mice significantly impacted the expression of IEGs in the hippocampus and cerebellum. However, the homozygous knockout of Pgbd5 in mice might have led to developmental defects that deregulated the IEGs, independent of Pgbd5's direct actions. To address this issue, we aimed to confirm the differential expression of IEGs upon PGBD5 depletion in a non-cancerous human cell line.

3.3.3.1 Validation of PGBD5 knockout

We obtained six PGBD5 knockout iPSC clones generated by the MDC stem cell facility. The knockout was achieved by the genomic deletions of 10 bp using CRISPR/Cas within the 3rd exon resulting in a frameshift validated by Sanger sequencing. We conducted Western Blot and RT-PCR analysis to confirm the knockout at the protein level. However, Western Blot analysis with the endogenous antibody NBP2-67048 did not show a knockout phenotype, indicating that this antibody may not be specific to PGBD5 in the used cell line (Figure 22). Similarly, Western Blot

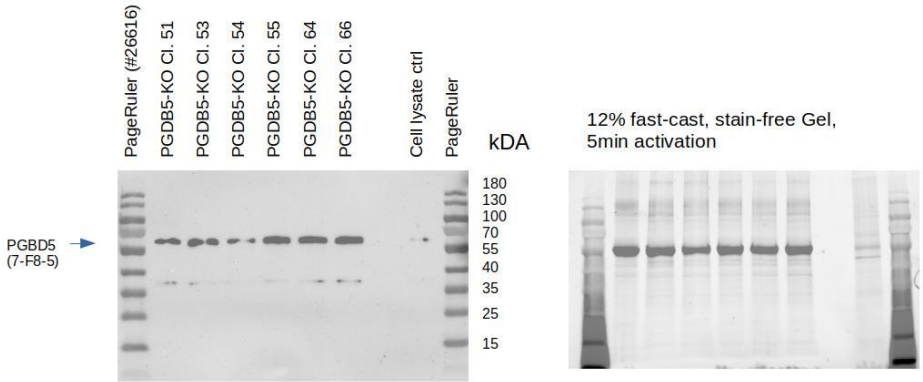


Figure 22: Western Blot analysis of PGBD5 knockout-iPSC clones probing with anti-PGBD5 antibody: NBP2-67048. Left) Western Blot bands obtained from the knockout clones and the wild-type lysate, probed with the anti-PGBD5 antibody NBP2-67048. Right) The protein load was assessed from a stain-free gel.

analysis with MBS355128 resulted in a weaker band at the size of full-length PGBD5 compared to the wild type. Still, a band was detected in all knockout clones, suggesting that this antibody is also not specific to PGBD5 (Figure 23).

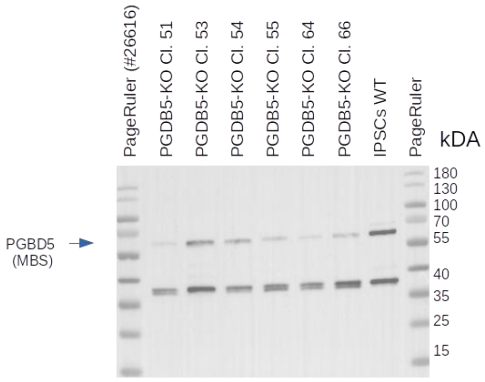


Figure 23: Western Blot analysis of PGBD5 knockout-iPSC clones using anti-PGBD5 antibody: MBS355128. Western Blot bands were obtained from the knockout clones, and the wild-type lysate was probed with the anti-PGBD5 antibody MBS355128.

Although all clones had a validated genomic knockout introducing a frameshift, PGBD5 antibodies failed to confirm a knockout in any of the clones, raising concerns about the specificity

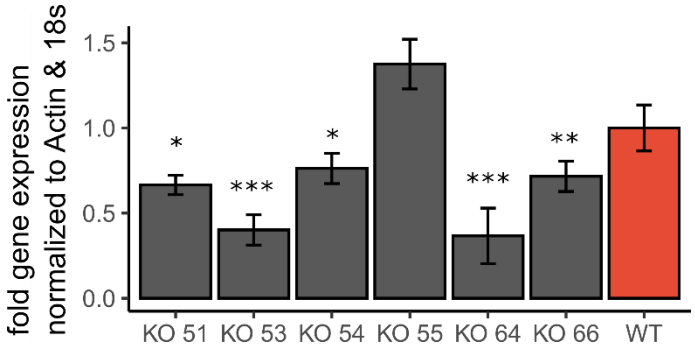


Figure 24: MRNA expression of PGBD5 in PGBD5 knockout-iPSCs measured by RT-PCR. Error bars indicate the standard deviation. Pairwise t-tests were used to confirm downregulation and significance levels are indicated by asterisks.

of the antibodies. To further confirm the depletion, we explored other assays. One approach was to investigate whether the frameshift deletion introduced *premature translation-termination codons* (PTCs), which are recognized and degraded by a gene expression regulation mechanism called *non-sense mediated mRNA decay* (NMD). This mechanism is critical in preventing the production of potentially harmful C-terminally truncated proteins (Kim *et al.*, 2022). The 10 bp deletion in exon 3 of PGBD5 would result in a PTC at either amino acid 346 (reference: human PGBD5 CDS protein NM_001258311.2).

All but one clone showed reduced mRNA levels (Figure 24) in RT-PCR compared to the wild-type cell line (two-sided t-test: clone 51: P-value = 0.037, clone 53: P-value < 0.001, clone 54: P-value = 0.015, clone 64: P-value < 0.001, clone 66: P-value = 0.006). Solely clone 55 had increased expression.

3.3.3.2 Topoisomerase II regulates IEGs in iPSCs

First, we evaluated whether topoisomerase II regulates the expression of IEGs in our iPSC cell line (Figure 25). We inhibited topoisomerase II using Etoposide (VP-16) at a concentration of

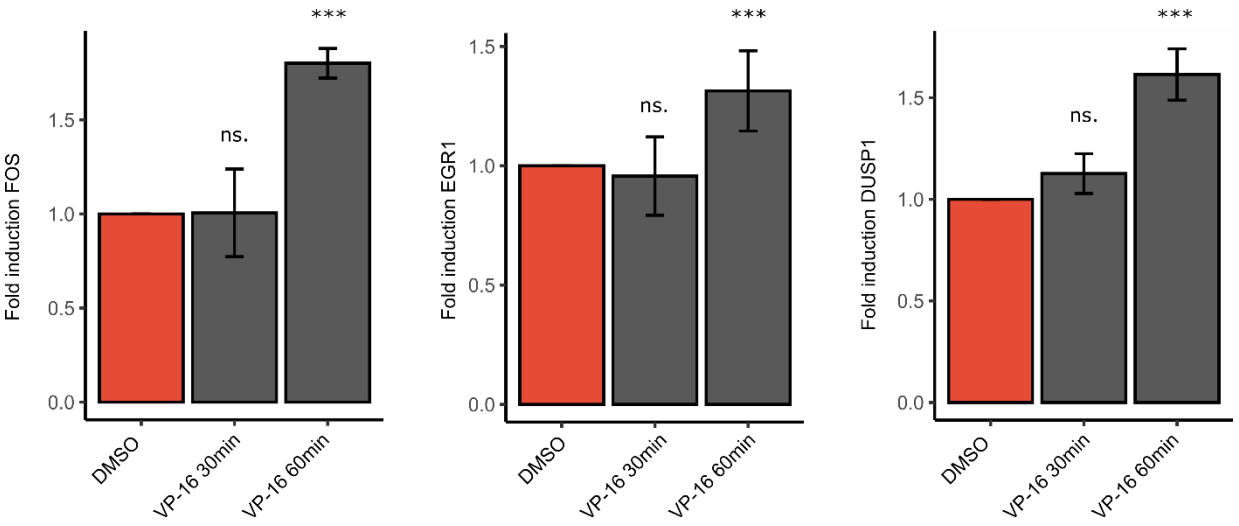


Figure 25: Inhibition of topoisomerase II increases expression of IEGs. Left) FOS, Middle) EGR1 and Right) DUSP1. Error bars indicate the standard deviation. Pairwise t-tests were used to confirm differential expression and significance levels are indicated by asterisks.

200µM for either 30 or 60 minutes. We then measured the expression levels of FOS, DUSP1, and EGR1 using RT-PCR. We chose these three IEGs as they show baseline expression in ES cells, unlike NR4A1 and NPAS4, which are not expressed in iPSCs.

After subjecting the cells to 60 minutes of Etoposide treatment, all tested IEGs displayed a noticeable upregulation compared to the control group treated with DMSO (two-sided t-test: FOS: P-value < 0.001, EGR1: P-value < 0.001, DUSP1: P-value < 0.001). This implies that topoisomerase II plays a regulatory role in controlling the expression of these IEGs in our iPSC cell line.

3.3.3.3 FOS & EGR1 expression in PGBD5 knockout clones

Next, we examined the impact of PGBD5 knockout on the expression of IEGs, specifically FOS and EGR1 (Figure 26). Our analysis revealed that the basal expression of FOS was consistently upregulated in nearly all tested clones (clone 53: P-value = 0.042, clone 54: P-value = 0.020, clone 64: P-value = 0.064, clone 66: P-value = 0.046), as determined by a two-sided t-test. Although clone 64 did not show statistically significant upregulation, there was a noticeable trend. On the other hand, the depletion of PGBD5 did not have a significant effect on the expression of EGR1 (clone 53: P-value = 0.98, clone 54: P-value = 0.29, clone 64: P-value = 0.30, clone 66: P-value = 0.08). However, clone 64 exhibited a tendency towards altered expression.

In a previous experiment, we observed that FOS displayed a greater upregulation of approximately 1.7-fold upon topoisomerase inhibition treatment, while EGR1 exhibited a more modest increase of 1.25-fold. Based on previous findings, we formulated the hypothesis that PGBD5 may modulate the activity of topoisomerase II. This aligns with our current results, as we observed a more pronounced effect on the expression of FOS, whereas we detected no changes in EGR1 expression. Our findings revealed that topoisomerase II also plays a crucial role in regulating the induction of IEGs in human iPSCs. Moreover, we observed that the expression of the IEG FOS was affected by the depletion of PGBD5, suggesting that PGBD5

influences the expression of specific IEGs. Interestingly, unlike the results obtained from mice brain samples, the depletion of PGBD5 led to an upregulation of FOS in our experiment. This discrepancy suggests that PGBD5 may exert diverse effects on IEG expression depending on the context, e.g., unstimulated vs. stimulated environment.

These results imply that PGBD5 might exert its regulatory influence on a subset of IEGs by enhancing the function of topoisomerase II. However, further experiments should be conducted

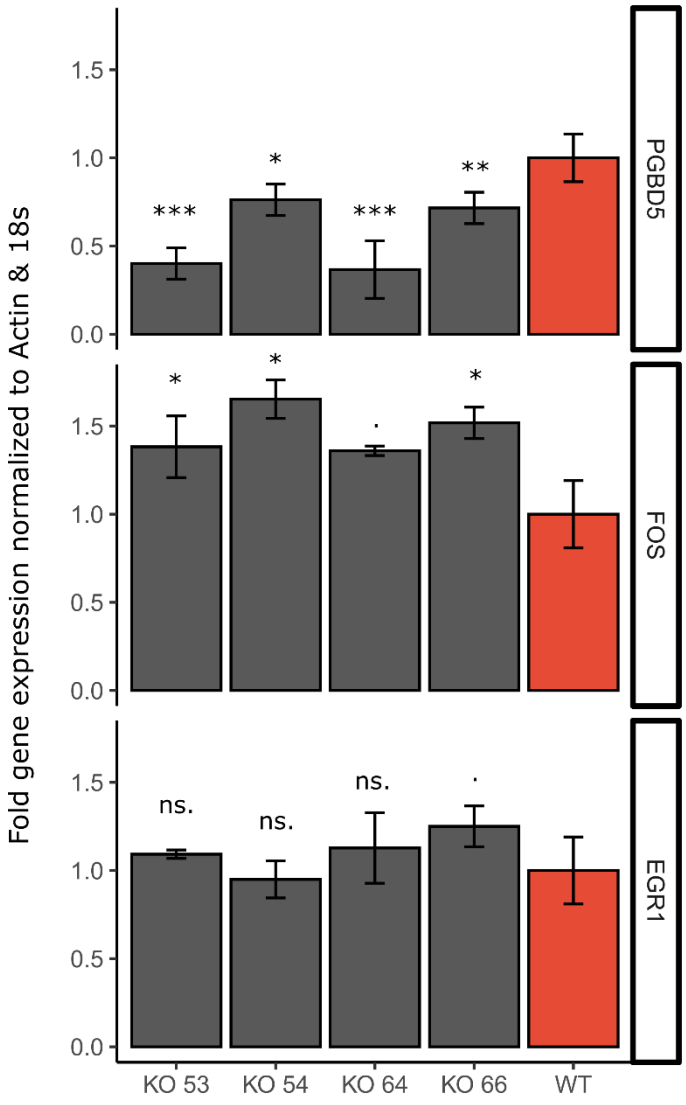


Figure 26: MRNA expression of PGBD5, FOS and EGR1 in PGBD5 knockout-iPSCs measured by RT-PCR. Error bars indicate the standard deviation. Pairwise t-tests were used to confirm differential expression and significance levels are indicated by asterisks.

to validate these claims and ensure statistical robustness. Repetition of the experiment with increased sample size and inclusion of additional IEGs would provide more robust evidence and a more comprehensive understanding of the relationship between PGBD5, topoisomerase II, and IEG expression in iPSCs.

3.4 PGBD5 enhances topoisomerase II activity in-vitro

In our search for a protein capable of inducing DNA DSBs, we conducted SILAC-based AP-MS experiments to generate a PGBD5 interactome. While not statistically significant, our findings indicated TOP2A as the most promising candidate. To confirm their interaction, we successfully performed co-IP. Additionally, we observed that a subset of IEGs regulated by topoisomerase II showed consistent deregulation between two brain regions in Pgbd5 deficient mice. Moreover, FOS exhibited differential expression upon PGBD5 depletion in human iPSCs. In our subsequent investigation, we aimed to determine whether Pgbd5 directly affects the enzymatic activity of topoisomerase II α through a topoisomerase II-specific *in-vitro* decatenation assay.

Mitochondrial *kinetoplast DNA* (kDNA) is comprised of interlocked DNA circles, which can only be decatenated by topoisomerase type II proteins. We focused on the relaxed DNA circles of Crithidia fasciculata, where the most prevalent size is 2.3 kb. In our decatenation assay, we examined the impact of PGBD5 on the catalytic activity of topoisomerase II by measuring the relaxation of DNA following the reaction.

We purified HA-tagged Pgbd5 protein from HEK293 cells. We used only eluates with at least 70% purity for the decatenation assay. The purity was quantified by measuring the protein load in a stain-free agarose gel (Figure 27, left panel). The purification process successfully weakened the interaction with topoisomerase II α, possibly through the sonication process, and no measurable amount of topoisomerase II α was detected in the HA-Pgbd5 eluates (Figure 27, right panel).

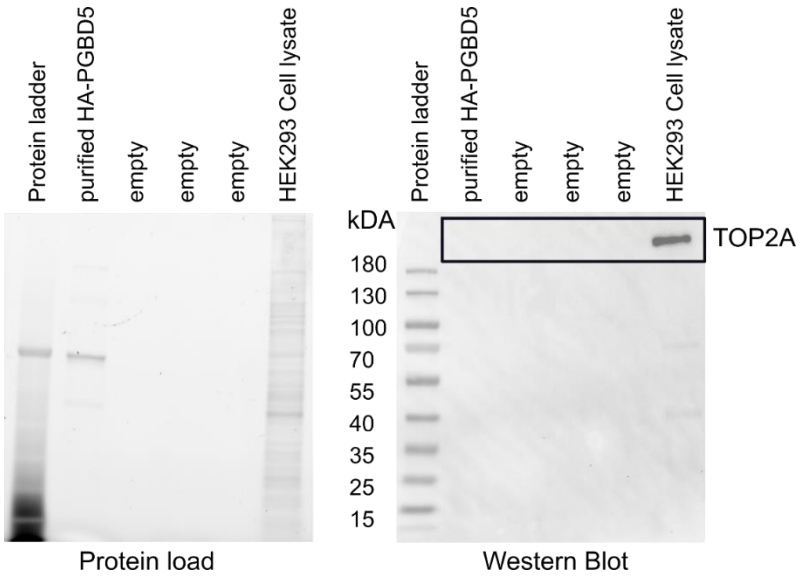


Figure 27: Protein purification of HA-tagged Pgbd5. Left) Protein load. Right) Negative co-IP of purified HA-Pgbd5 probed with topoisomerase II α antibody.

We performed the decatenation assay with two control samples: HA peptide eluate in *Phosphate-buffered saline* (PBS) and purified cellular extract with HA beads in untransfected HEK293 cells. We added the control lysate to all samples except those with Pgbd5 lysate. While PBS exhibited no background activity (Figure 28), the untransfected cellular extract control displayed a high background. To quantify the results, the background was subtracted by comparing the reaction with TOP2A to the same reaction without TOP2A. The addition of *Ethylenediaminetetraacetic acid* (EDTA) at the beginning of the reaction greatly reduced the background activity. A minor amount was still detected when ATP, necessary for decatenation by TOP2A, was omitted. The presence of ATP and small amounts of TOP2A in the cellular extracts could contribute to this background activity. Additionally, we used Etoposide, a well-known inhibitor of topoisomerase II, in this assay. Etoposide traps the enzyme on DNA, preventing the religation of broken DNA strands and inhibiting the relocation of topoisomerase II to other regions from relaxing supercoiled DNA.

In this experiment, topoisomerase activity was successfully inhibited by approximately 50% using Etoposide. The purpose of this was to establish an additional reference point for quantifying

the effects of PGBD5. Prior to this experiment, we determined the required concentration of Etoposide through a titration series.

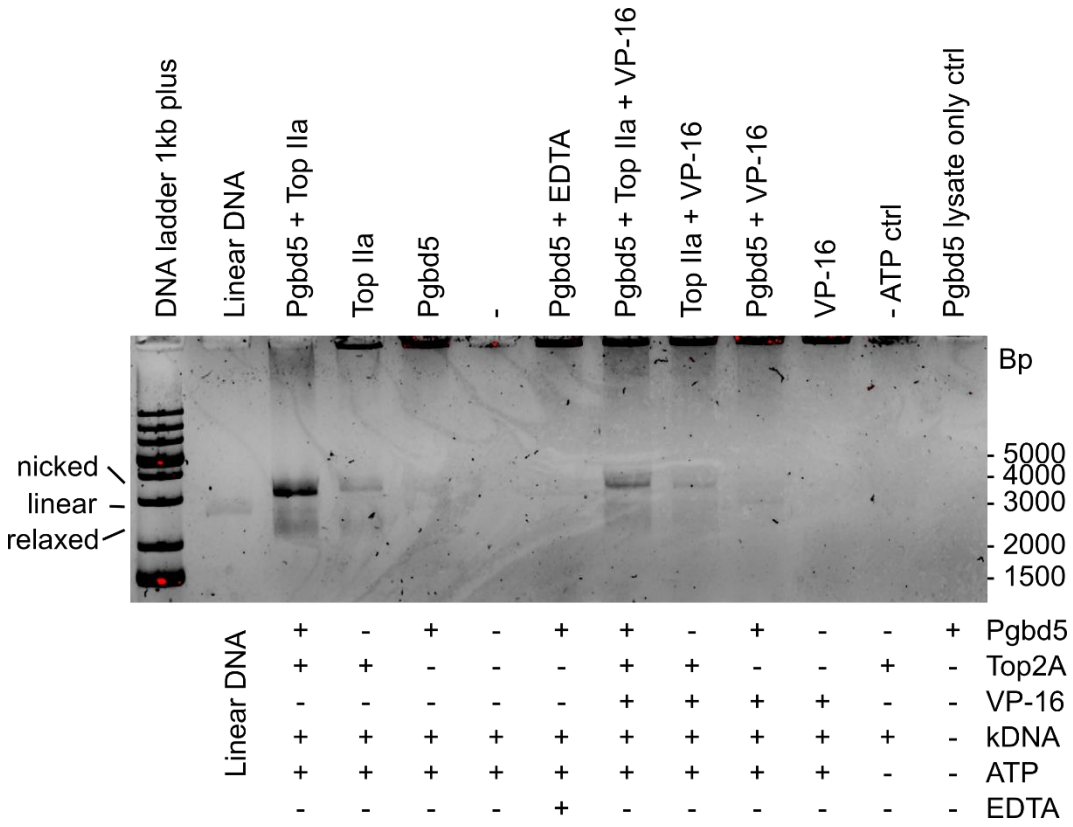


Figure 28: Effect of Pgbd5 on topoisomerase II specific decatenation assay. KDNA was linearized using the unique restriction site XholI (lane 1). KDNA was incubated with Top IIα and purified HA-Pgbd5 extract either in the presence (lane 7) or absence (lane 2) of Etoposide (VP-16). Control lanes were established that contained no Top IIα (4 & 9) or no HA-Pgbd5 (lanes 3 & 8) or lacking both (lanes 5 & 10). Two reactions provided negative controls: one sample was supplemented with EDTA at the beginning of the reaction (lane 6), and another sample lacked ATP (lane 11), a required co-factor of topoisomerase IIα. The last lane contained purified HA-Pgbd5 extract and controlled for DNA contamination in the sample.

The data shows that HA-tagged Pgbd5 increased the amount of relaxed DNA (two-sided t-test, *degrees of freedom* = 3, p-value = 0.0069) by approximately 2-fold. The increase was about 1.5-fold when in the presence of Etoposide (two-sided t-test, *degrees of freedom* = 3, p-value = 0.0042) (Figure 28 & Figure 29).

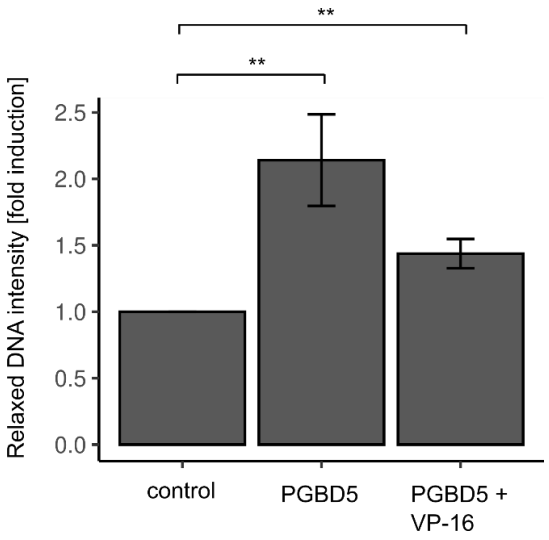


Figure 29: Quantification of decatenation assay. The relaxed DNA band intensities were quantified in ImageJ. The band intensities were normalized to their corresponding controls. Their means were compared with a t-test, and asterisks denote the significance level.

Altogether, the results indicate that PGBD5 enhances topoisomerase II α decatenation activity *in-vitro* and potentially enhances topoisomerase activity *in-vivo.* That experiment further strengthens the evidence that PGBD5 regulates a subset of IEGs with TOP2.

However, it is important to note that the protein purity of PGBD5 was only > 70%. Thus, instead of a direct effect, PGBD5 might pull down factors that are promoting the increase in TOP2A activity.

3.4.1 Expression patterns of PGBD5

This paragraph focuses on investigating the expression patterns of PGBD5, aiming to understand the expression pattern of PGBD5 in tissues and cell types. Our main objectives were to identify the factors that drive PGBD5 expression, confirm its suggested brain specificity as indicated by Pavelitz *et al.* (2013), and ultimately determine the physiological relevance of PGBD5. Additionally, we aimed to assess the suitability of *in-vitro* neuronal differentiation systems for

studying the molecular function of PGBD5 in a physiologically relevant context. Specifically, we sought to verify whether PGBD5 enhances TOP2B function and regulates IEGs in neurons.

3.4.1.1 PGBD5 is expressed in the brain, particularly in neurons

To gain insight into the possible function of PGBD5, we determined the expression profile of its RNA in different cell types and tissues. For this purpose, we utilized several publicly available databases and datasets. First, we examined human tissue data in the Human Protein Atlas database (2023). The results showed that PGBD5 expression was highest in several brain regions, particularly in the hippocampal formation and cerebral cortex. Additionally, high levels of expression were observed in the pancreas and cervix, as shown in Figure 30.

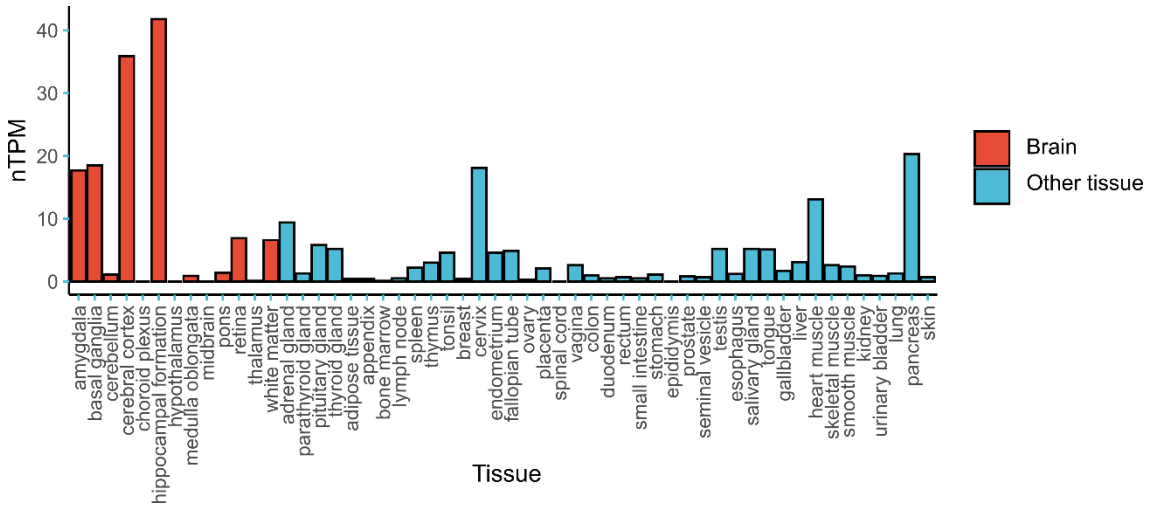


Figure 30: Tissue-specific expression of PGBD5. PGBD5 is highly expressed in brain tissues, cervix, and pancreas. The X-axis represents different tissues, and the Y-axis displays expression levels in normalized TPM (nTPM). The calculation method for nTPM can be found in the Human Protein Atlas (2023)

On a single-cell level, we found exceptionally high expression of PGBD5 in excitatory and inhibitory neurons, according to data from the Human Protein Atlas, Single-Cell RNA (2023), and EMBO Single-Cell Atlas (2022) (Figure 31).

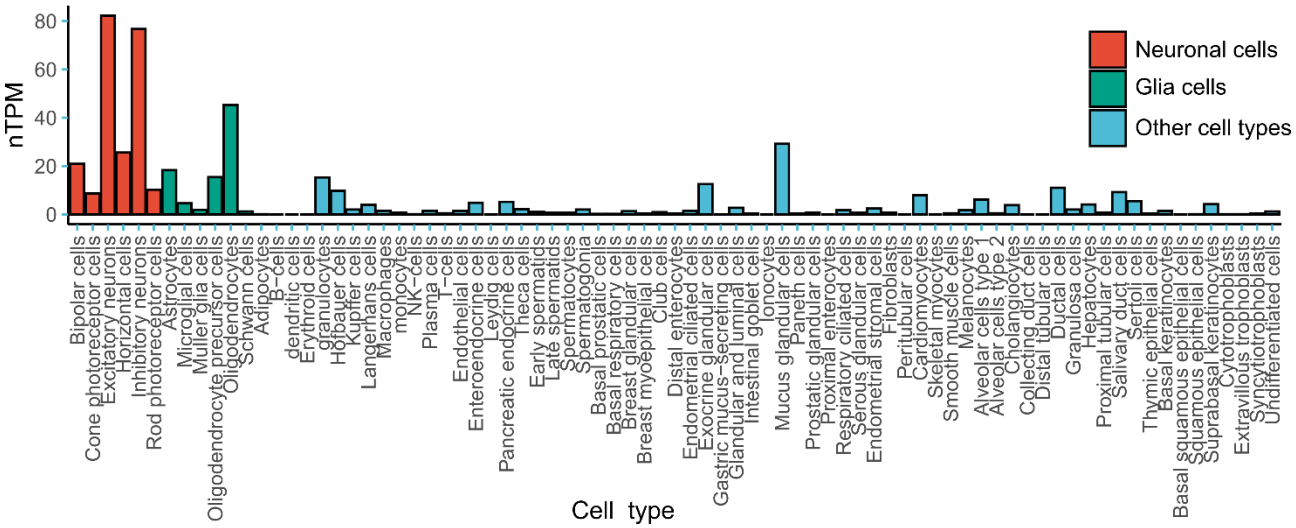


Figure 31: Cell type-specific expression of PGBD5. PGBD5 is highly expressed in excitatory and inhibitory neurons as well as in oligodendrocytes. The X-axis represents different cell types, while the Y-axis displays expression levels in normalized TPMs (nTPM). The calculation method for nTPM can be found in the Human Protein Atlas (2023)

Despite this, a significant variation was observed across datasets, even within the same brain regions (Figure 32A). A linear regression analysis revealed that the variation was positively correlated with gene expression in the hippocampus (t-test, p-value < 2e-16, adjusted $R^2 = 0.94$, *degrees of freedom* = 56198) (Figure 32B). We assessed the magnitude of PGBD5's variation by comparing PGBD5's standard deviation to that of genes with a similar mean expression (+/- 5 RPKM). The comparison showed that PGBD5 was in the top 13% regarding variation (302 out of 2503 genes had higher variation) (Figure 32C). Various external and internal factors may cause variations in expression levels, such as circadian rhythm, age, season, food intake, hormonal fluctuations, or other neural stimuli.

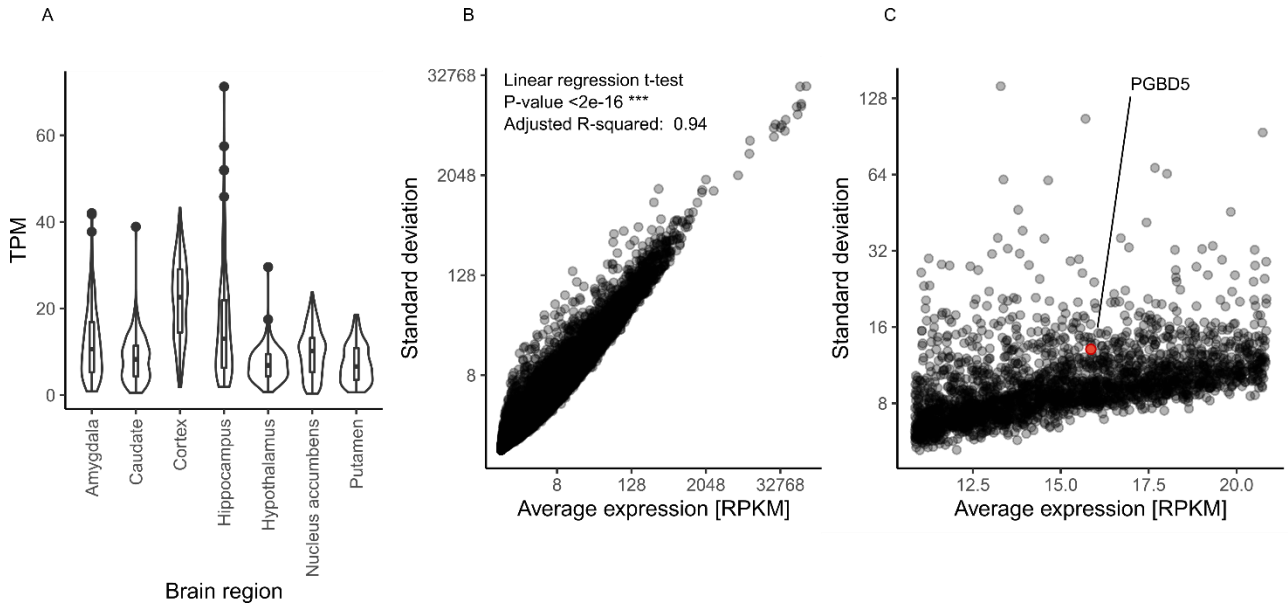


Figure 32: PGBD5 expression variation in brain tissues. A) PGBD5 exhibits high expression variation in selected brain tissues of the GTEx dataset. B) The standard deviation of PGBD5 expression positively correlates with gene expression in the hippocampus data from the GTEx dataset. C) Within the PGBD5 expression window of +/- 5 RPKM, it ranks among the top 13% of variably expressed genes

This variance in expression resulted in hippocampi with low PGBD5 expression compared with the expression of all other genes (among 25.6% highest expressed genes, 13456/52566 gene features showed higher expression) or relatively high PGBD5 expression (among 3.2% highest expressed genes, 1682/52566 gene features showed higher expression) illustrated in Figure 33.

In conclusion, PGBD5 is highly expressed in excitatory and inhibitory neurons in the brain. Its expression exhibits a high degree of variation between donors, ranking among the top 13% of genes in its expression range in terms of variation.

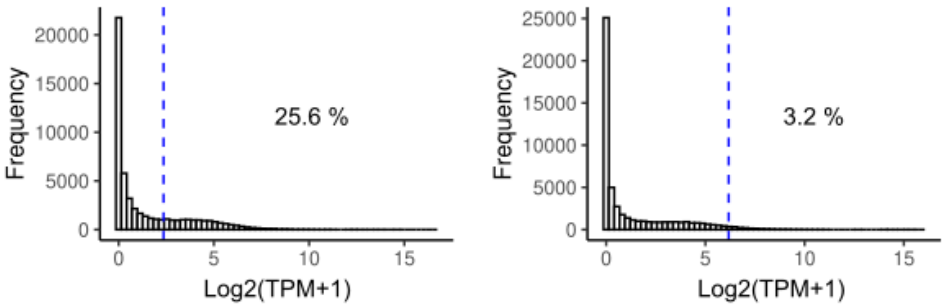


Figure 33: PGBD5 expression variation of human hippocampi. Left panel) PGBD5 is among the 25.6% highest expressed genes. Right panel) PGBD5 is among the 3.2 % highest expressed genes. The blue dotted lines indicate the expression level of PGBD5.

3.4.1.2 Investigating the source of PGBD5 expression variation: Age as a contributor

PGBD5 expression is highly variable, and age is one factor that can contribute to differences between donors. The BrainSpan (2022) database provides gene expression data for individuals throughout the human lifespan. Although PGBD5 expression is similar between tissues, the donor age partly explains expression variations in most tissues (Figure 34). The Table 11 displays the correlation and results of linear regression-based t-tests between donor age in days and PGBD5 expression in -log2 RPKM for each tissue.

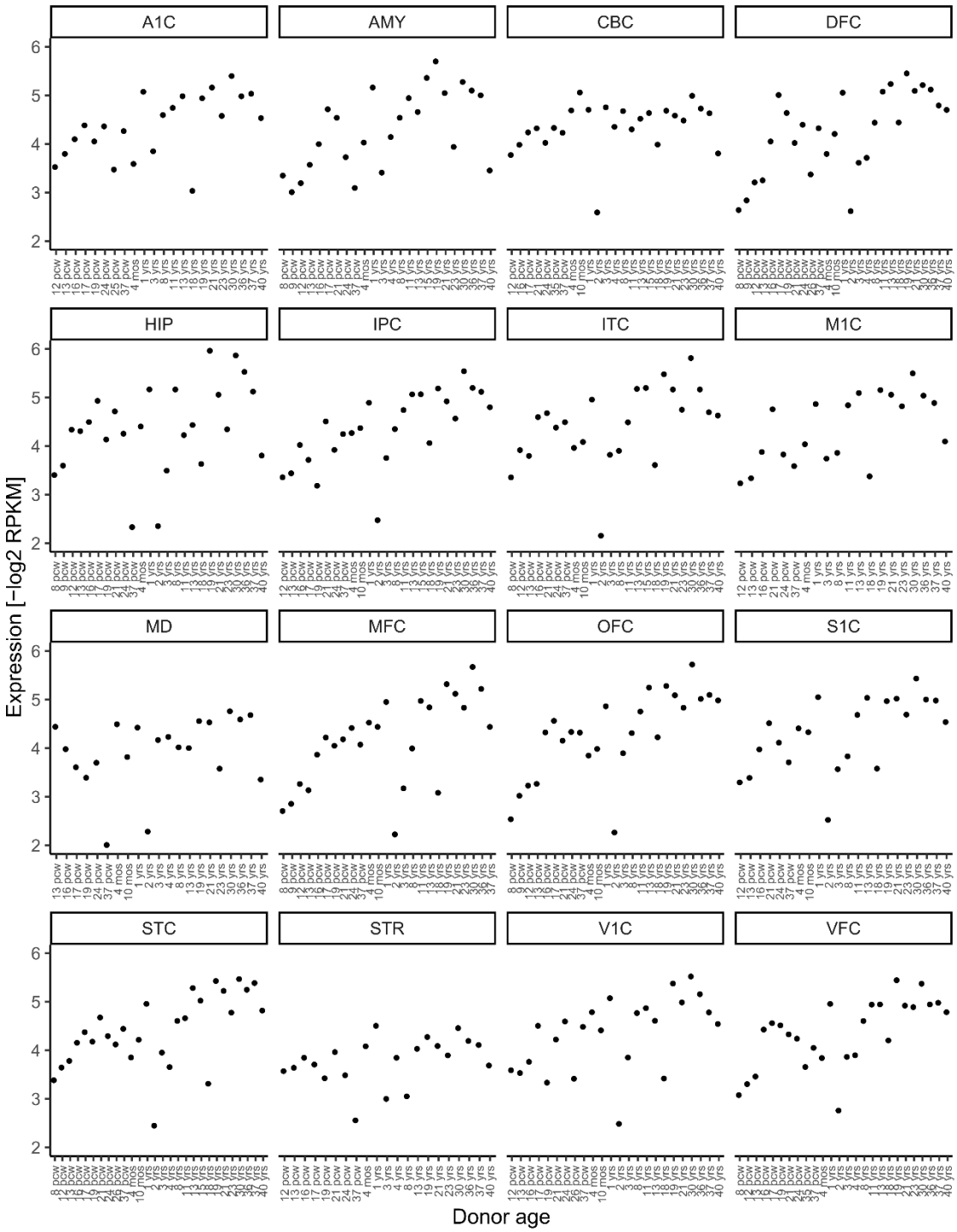


Figure 34: Expression of PGBD5 over age in various brain tissues. The tissue abbreviations are listed in the table below. The other abbreviations are pcw for post-conception weeks, mos for months, and yrs for years.

Structure name	Structure abbreviation	Correlation (Age, RPKM)	T-value (RPKM ~ Age)	P-value	Significant	Adjusted R^2	DF
primary auditory cortex	A1C	0.57	2.87	0.0095	**	0.26	20
amygdaloid complex	AMY	0.61	2.38	0.0264	*	0.17	22
cerebellar cortex	CBC	0.28	1.02	0.317	n.s.	0.00	23
dorsolateral prefrontal cortex	DFC	0.64	3.47	0.0019	**	0.30	25
hippocampus	HIP	0.37	2.17	0.0405	*	0.13	23
posteroventral parietal cortex	IPC	0.69	4.30	0.000269	***	0.42	23
inferolateral temporal cortex	ITC	0.47	2.78	0.0108	*	0.23	22
primary motor cortex	M1C	0.61	2.69	0.0151	*	0.25	18
mediodorsal nucleus of thalamus	MD	0.34	1.54	0.139	n.s.	0.06	19
anterior cingulate cortex	MFC	0.57	3.18	0.00422	**	0.27	23
orbital frontal cortex	OFC	0.7	4.18	0.000359	***	0.41	23
primary somatosensory cortex	S1C	0.55	3.10	0.00561	**	0.29	20
posterior superior temporal cortex	STC	0.55	3.87	0.000663	***	0.34	26
striatum	STR	0.38	1.99	0.0613	.	0.13	19
primary visual cortex	V1C	0.51	2.64	0.0149	*	0.21	22
ventrolateral prefrontal cortex	VFC	0.64	3.94	0.000611	***	0.37	24

Table 11: PGBD5 expression increases with age. Correlation and linear regression statistics of PGBD5 expression over donor's age.

In conclusion, PGBD5 expression increases with age. This discovery is intriguing concerning DNA damage, as PGBD5 expression sensitizes cells to DNA damage.

3.4.1.3 High expression of PGBD5 in cells with recent neuronal activity

We found that age is a factor that contributes to differences in PGBD5 expression between donors. Still, there was also significant variation in PGBD5 expression within the same cell types of a single donor's tissue. To better understand these differences, we investigated them using single-cell data. We divided the cells into two groups, those with high and low expression of PGBD5, and identified genes and pathways that distinguish the two groups.

The analysis demonstrated high expression of PGBD5 in neurons and neuroglia but low expression in other neuronal subtypes (Figure 35). PGBD5 was highly expressed in both GABAergic and glutamatergic neurons.

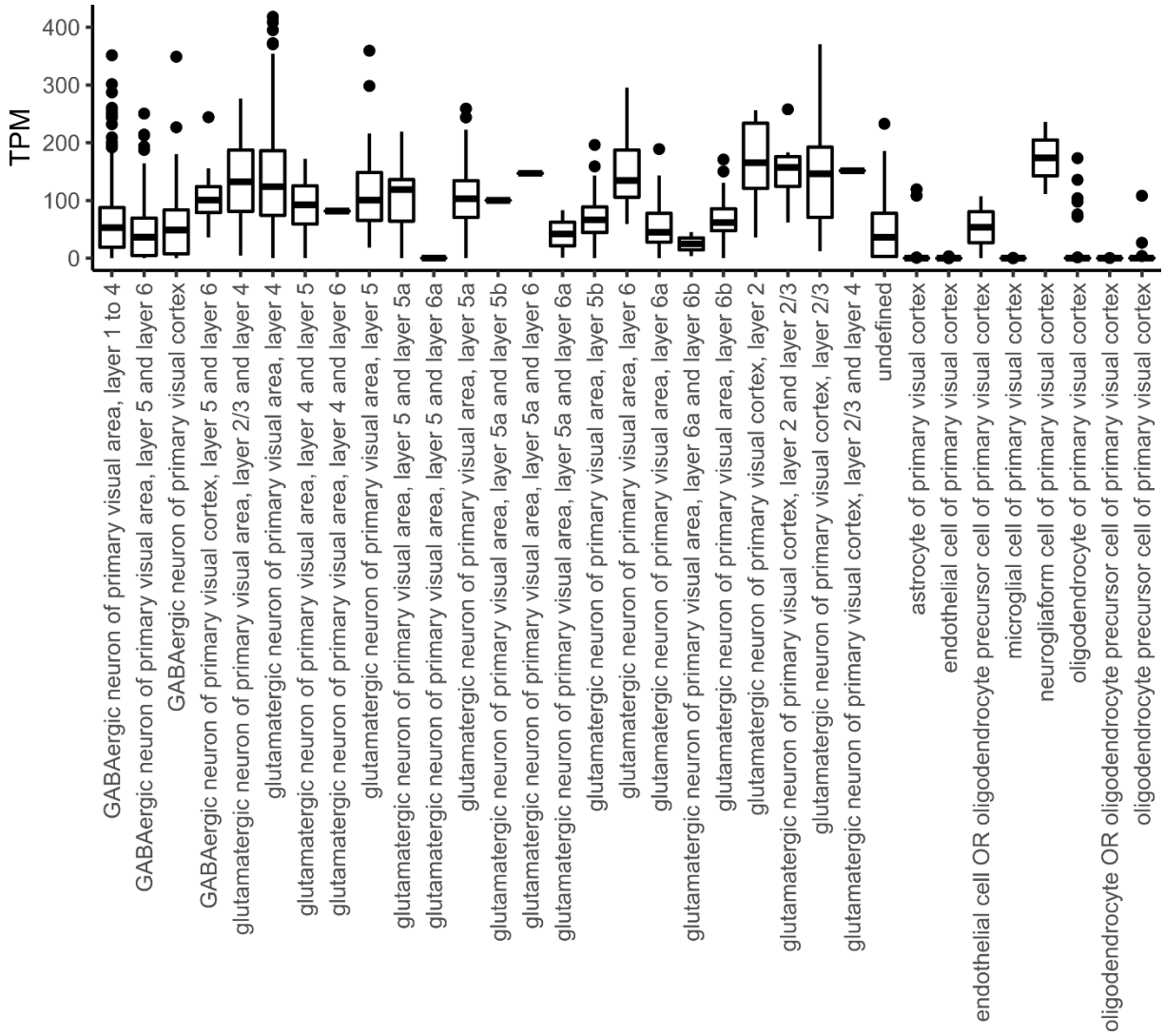


Figure 35: Pgbd5 cell type-specific expression in the visual cortex of mice. Pgbd5 is highly expressed in excitatory and inhibitory neurons.

We sought to understand the conditions under which PGBD5 transcription is induced and analyzed single-cell data from the mouse visual cortex. We divided the cells into two groups (Figure 36A), those with high PGBD5 expression (1312 cells) and those with low PGBD5 expression (301 cells), and performed a differential gene expression analysis. We evaluated a cell-type bias by comparing the distribution of the neuronal marker gene Eno2 in the two groups, and both showed a similar expression pattern of Eno2, as indicated in Figure 36B.

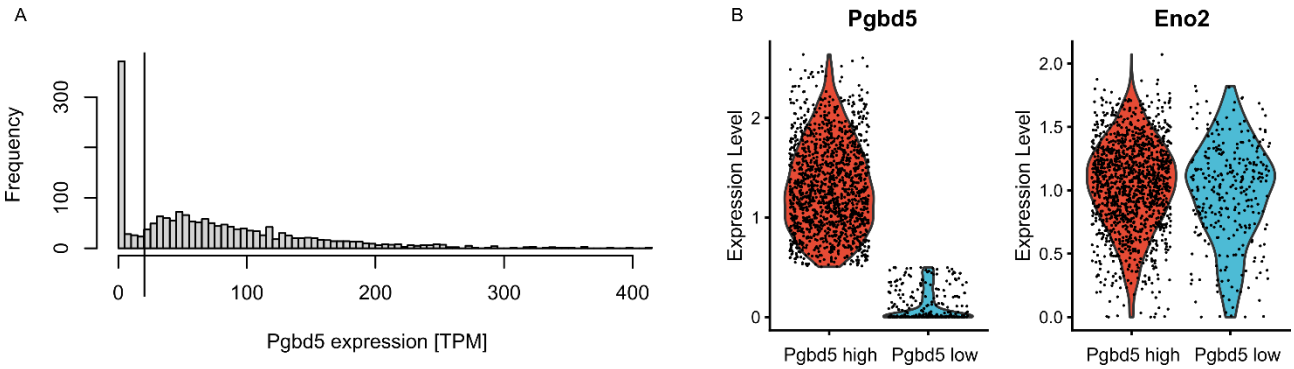


Figure 36: Pgbd5 expression cut-off in the visual cortex single-cell data. A) Histogram of Pgbd5 expression. The line indicates the cut-off of Pgbd5 high vs. low expressing cells. B) Distribution of neuronal marker Eno2 expression in Pgbd5 high vs. low-expressing cells.

Among the 332 co-expressed genes (average log2FC > 0.3 and FDR < 0.05), we identified Arc and Egr1. These two IEGs are expressed in response to visual and motor-related stimulation in the visual cortex (Mahringer *et al.*, 2022). The enrichment of Egr1 and Arc in the high PGBD5 group suggests the recent activity of these neurons. In addition, gene ontologies of the differentially expressed genes showed enrichment in the regulation of synapse organization, synaptic signaling, and long-term potentiation, among others (Figure 37).

In conclusion, the results show that PGBD5 is predominantly expressed in glutamatergic and GABAergic neurons as well as in neuroglia and oligodendrocytes. The co-expression of PGBD5 with IEGs Egr1 and Arc suggests that recent neuronal activity may elevate PGBD5 transcription. Many co-expressed genes are involved in synaptic plasticity and regulation, emphasizing that these neurons may have been recently active.

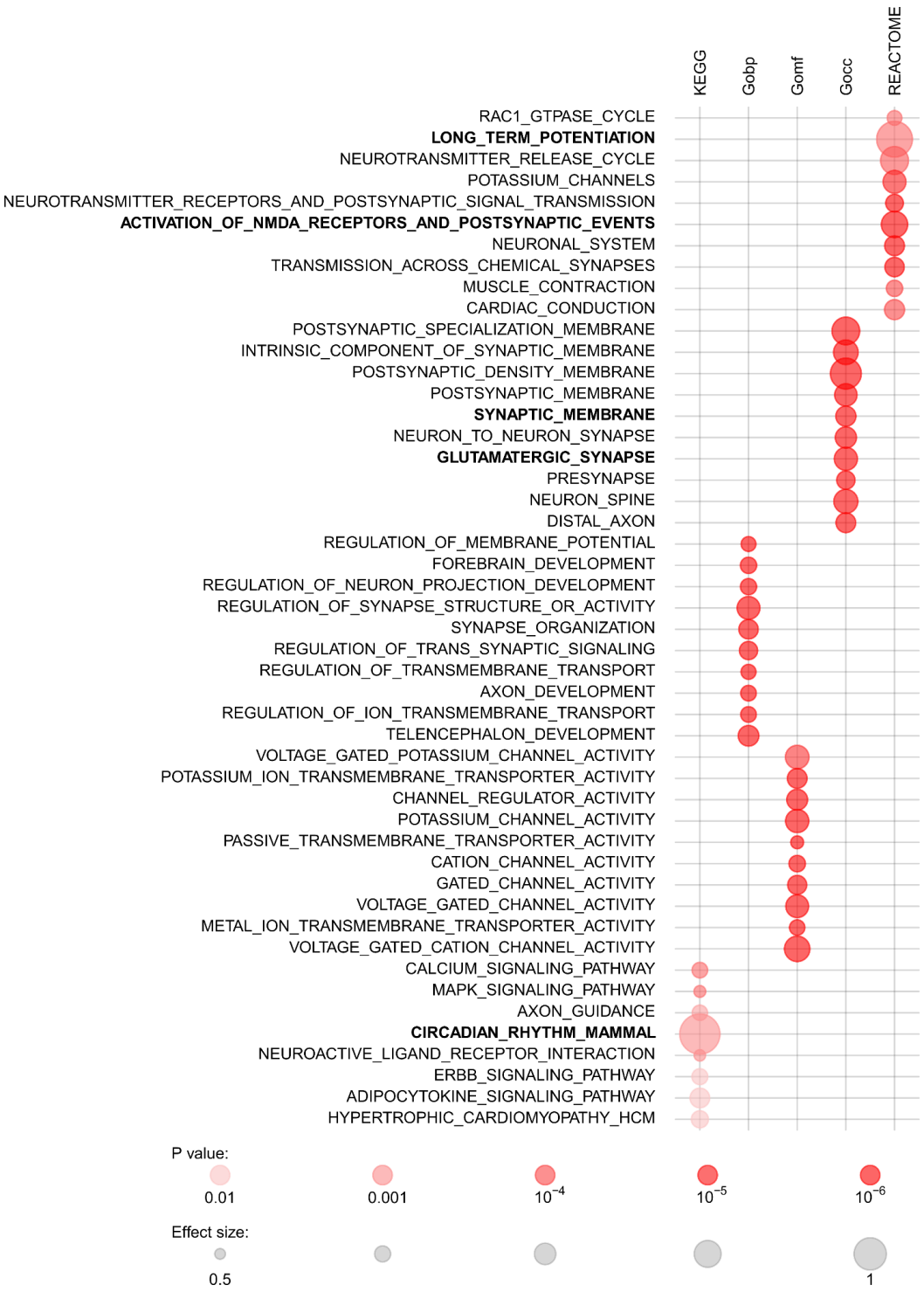


Figure 37: Enriched gene sets of Pgbd5-high vs. low differential expressed genes. Enriched gene sets were evaluated using tmod. The significance level (P-value) is represented by the intensity of the red color, while the dot size indicates the effect size, as indicated by the fold enrichment of the respective pathway. Only gene sets with an effect size > 0.75, and a FDR < 0.01 were included. The figure displays the top 10 terms from each database.

3.4.2 Neuronal differentiation in the dish, a model to study PGBD5?

After looking into the expression pattern, we wanted to evaluate whether human neuronal differentiation methods are suitable for studying PGBD5. Our lab is dedicated to reducing animal models' use in research. Hence, we utilized human neuronal differentiation methods to study PGBD5. The MDC in-house stem-cell unit kindly helped us by providing established protocols, assistance, and materials for neuronal differentiation from human iPSCs.

The goal was to examine how the presence of PGBD5 affects TOP2B-driven transcriptional changes, specifically in the expression of IEGs and DNA DSBs. We initially planned to conduct ChIP-seq experiments in neurons, including TOP2B, PGBD5, and yH2AX antibodies. Therefore, the model needed to express a fair amount of those three proteins. Additionally, we had to show that the model is sufficiently replicating the results of the study by Madabhushi *et al.* (2015), where NMDA treatment induced TOP2B-mediated transcriptional changes of IEGs.

3.4.2.1 PGBD5 expression in published datasets of human neuronal differentiation models

We examined the expression of PGBD5 in various published datasets of human neuronal differentiation models. In the first dataset, we explored the expression of PGBD5 in H1 *human embryonic stem cells* (hESCs) that underwent differentiation into various ventrally derived cell types over a 125-day protocol (Close *et al.*, 2017). The results showed that PGBD5 was expressed in mature neurons but had lower expression in progenitor and precursor cells (Figure 38). However, the overall expression was lower than in the *in-vivo* mouse visual cortex data.

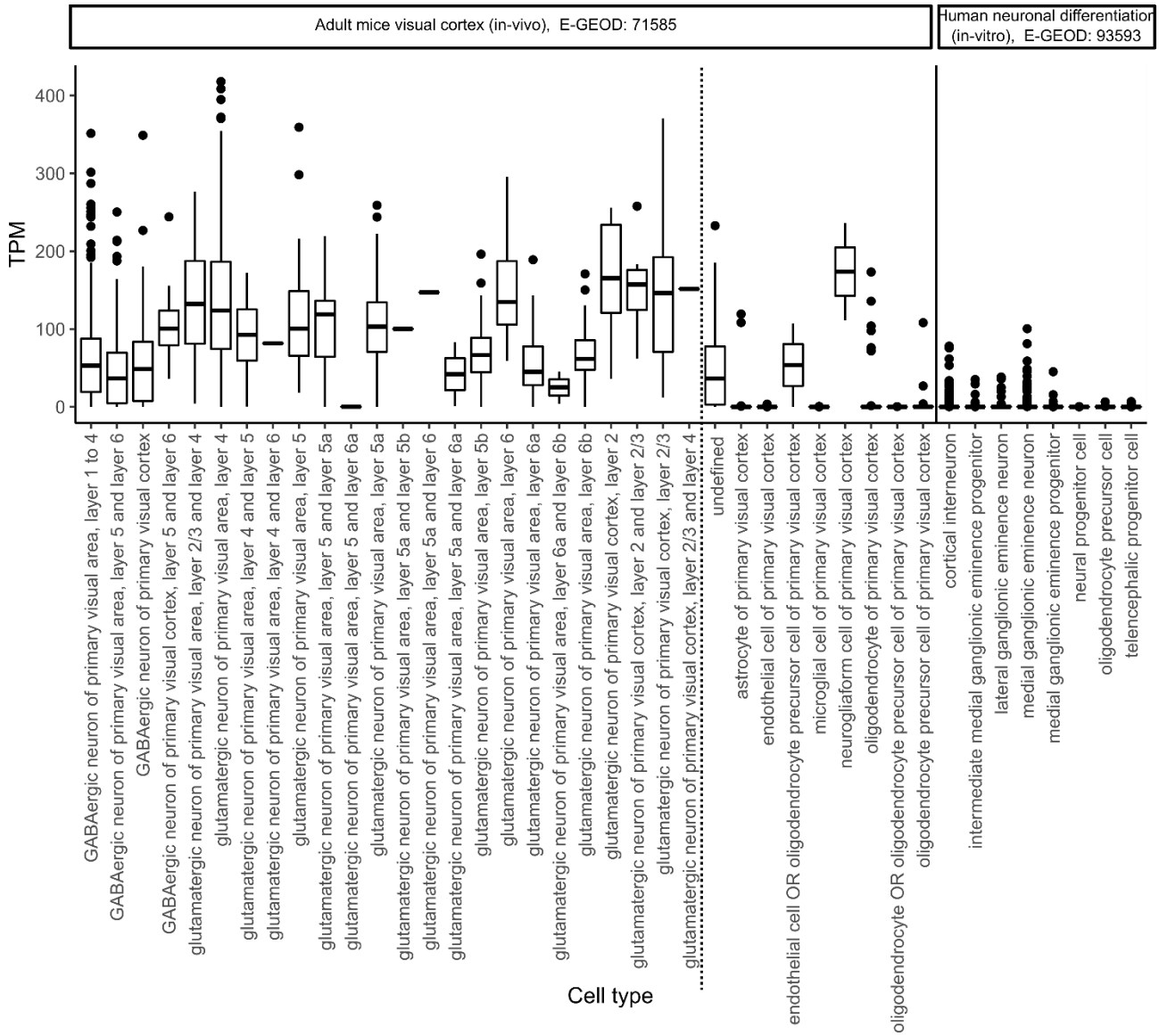


Figure 38: PGBD5 expression in human neuronal cell differentiation. Left) Expression in the mouse visual cortex: PGBD5 expression in visual cortex neurons of adult mice shows higher levels compared to *in-vitro* differentiation. Right) Human neuronal differentiation from H1 ES cells: PGBD5 expression during *in-vitro* differentiation of human neuronal cells exhibits overall lower expression levels.

Further, we investigated the expression of PGBD5 in multipotent cells to explore the potential of PGBD5 to impact neuronal differentiation (Figure 39). To do this, we analyzed single-cell data during human embryogenesis, which showed that PGBD5 is expressed in the oocyte with decreasing expression levels until the late blastocyst stage, where no expression was detected. In

the hESC stage, the expression varied greatly between individual cells (ranging from 0 to 20 RPKM).

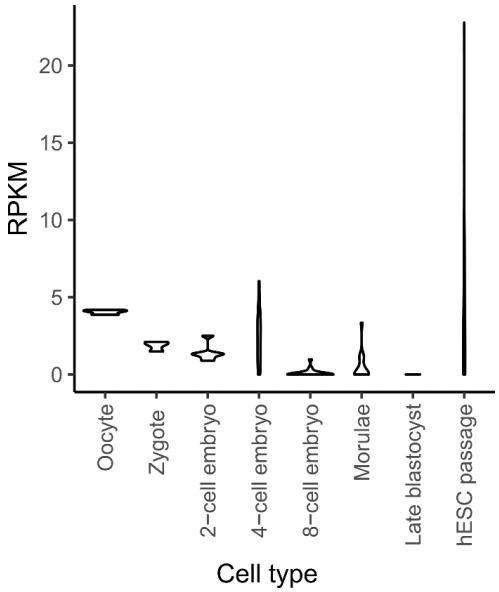


Figure 39: PGBD5 expression during human embryogenesis. PGBD5 expression shows a decline throughout embryogenesis until the late blastocyst stage. It then reaches its peak expression level at the passage of hESCs.

Finally, we investigated PGBD5 expression in a time-course study of hiPSCs differentiated into neurons, including five developmental stages (self-renewal, early neuronal differentiation, NPCs, assembled rosettes, and differentiated neurons) (Figure 40). The hiPSCs were derived from five donors and 13 subclonal lines (Burke *et al.*, 2020). The time course showed that PGBD5 was expressed in self-renewing cells, but its expression decreased during the first days of differentiation and was lowest at days 9 and 15 (NPC stage). However, its expression increased again at day 21 (rosette stage) and remained stable until the end of the protocol (day 77). Although the overall expression of PGBD5 was low, with the highest expression being approximately 1.6 RPKM, it suggests that gene editing of PGBD5 in these cells (e.g., through a knockout) could potentially impact neuronal differentiation.

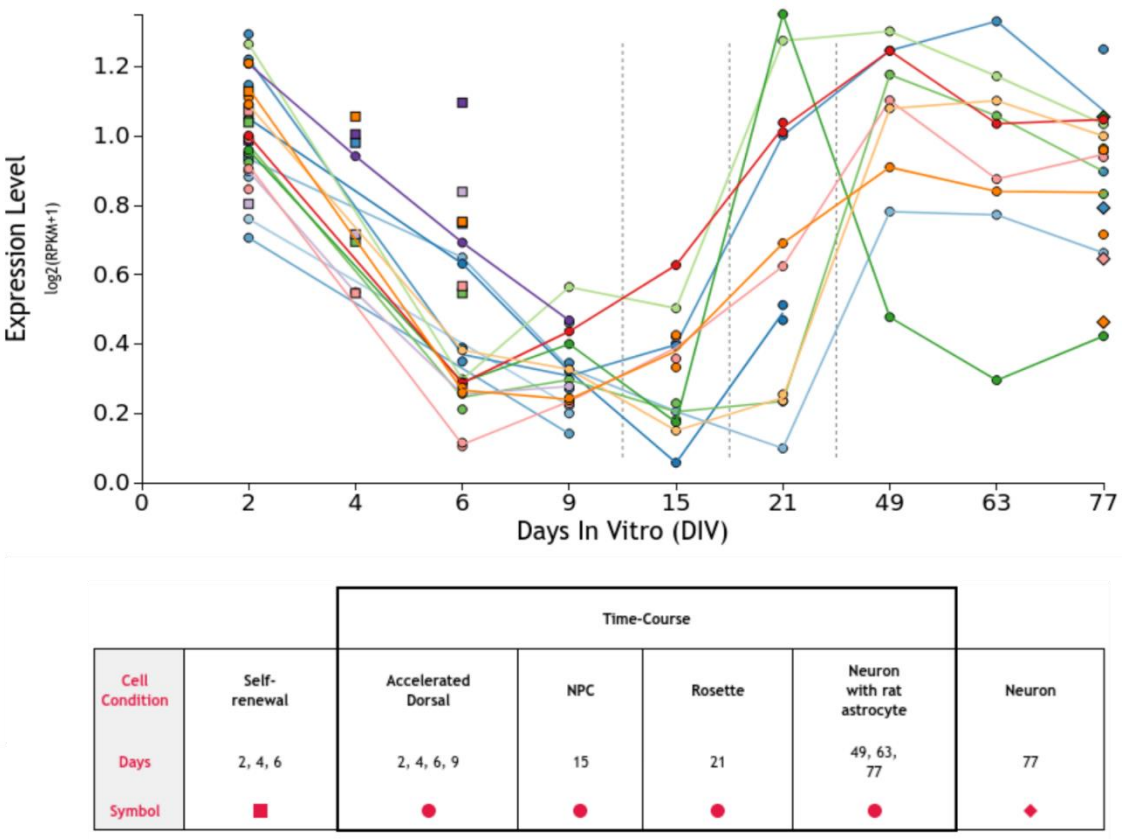


Figure 40: PGBD5 expression during differentiation of human iPSCs to neurons. hiPSC transcriptomics data of corticogenesis from 5 iPSC donors and 13 subclonal lines across nine time points, encompassing five broad conditions: self-renewal, early neuronal differentiation, neural progenitor cells (NPCs), assembled rosettes, and differentiated neuronal cells. The colors in the figure represent the expression levels of PGBD5 in the 13 subclonal lines. Figure was generated by web application (Burke *et al.*, 2020)

In conclusion, PGBD5 is expressed in mature neurons and self-renewing cells (hiPSCs and hESCs) but only vanishingly low in NPCs. The overall poor expression observed in *in-vitro* assays compared to *in-vivo* suggests that environmental factors in the cellular environment, such as neuronal stimulation, may affect PGBD5 transcription.

3.4.2.2 Protocol for hiPSC neuronal differentiation and co-culture with astrocytes

The following paragraph describes the neuronal differentiation protocol used in this study. As mentioned, we received an established protocol, training, and genetically modified hiPSCs for neuronal differentiation from the MDC stem cell unit (Diecke lab). The unit also evaluated marker genes' expression to determine the neurons' cell state and firing potential. The neuronal differentiation protocol was carried out in several steps, as outlined below (Figure 41). Astrocytes support the vitality of neuronal cultures and contribute to their physiological function.

On day -1, the hiPSCs were seeded and incubated for 24 hours before the differentiation protocol began. On day 0, the hiPSC media was replaced with F12-N2 media containing doxycycline, which induced the expression of NEUROG2 and caused the cells to differentiate into a neuronal cell type. On day 1, the F12-N2 media was exchanged with a fresh batch, and the cells were now NPC-like. On day 2, the F12-N2 media was replaced with NB-B27 media and mouse astrocytes were optionally added to create a co-culture. On day 3, another media change to NB-B27 was performed. Starting from day 4, half of the NB-B27 media was changed every other day. When co-cultured with astrocytes, Ara-c was added to the media to prevent the astrocytes from overgrowing the culture. By day 14, the neurons exhibited marker genes associated with mature neurons, and the culturing media was changed to Peitz medium. When co-cultured with astrocytes, the neurons showed firing potential at day 22, and this potential continued to increase until day 50 of the differentiation process. Fresh doxycycline was added to the media throughout the entire protocol.

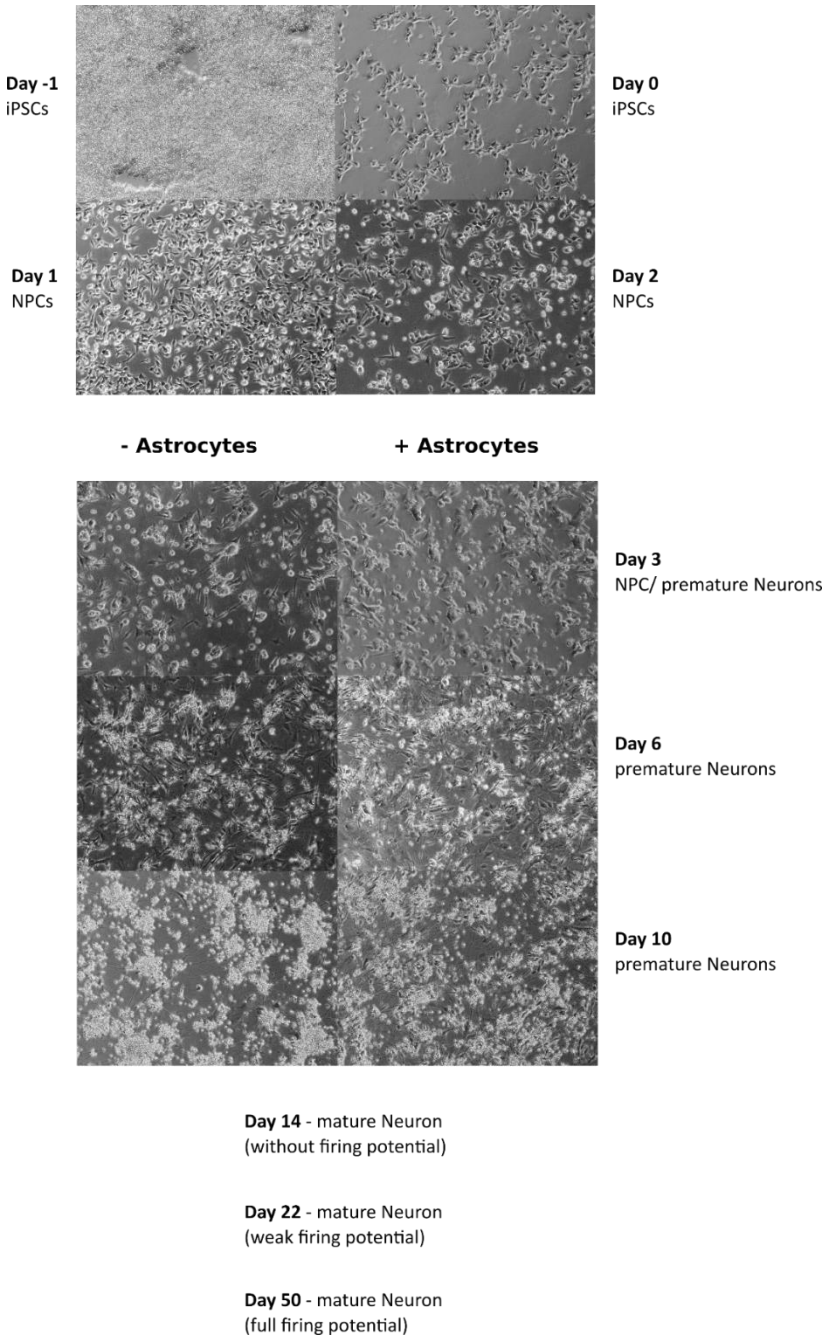


Figure 41: Morphological changes during neuronal differentiation protocol from hiPSCs to glutamatergic iNeurons. Day -1: Initial seeding of cells with typical hiPSC morphology. Day 0: NEUROG2 expression induction by doxycycline. Day 1: Transition to an NPC-like appearance. Day 2: Optional addition of mouse astrocytes to create a co-culture. Day 3: Continued increase in cell number. Fresh doxycycline addition throughout the protocol. Day 14: Expression of marker genes indicating mature neuron state. Day 22 to day 50: Neurons co-cultured with astrocytes demonstrate firing potential with increasing activity.

3.4.2.3 PGBD5 expression during neuronal differentiation and NMDA treatment

The following paragraph describes a series of experiments that were performed to determine the expression patterns of two proteins, PGBD5, and TOP2B, during the differentiation of human induced pluripotent stem cells (hiPSCs) into glutamatergic neurons and the effect of NMDA treatment on differentiated neurons. The results of the experiments are discussed in the context of previous findings from public datasets and the implications for the role of PGBD5 and TOP2B in the regulation of neuronal activity and DNA damage.

We investigated the expression of PGBD5 during neuronal differentiation and under NMDA treatment. We differentiated human induced pluripotent stem cells (hiPSCs) into glutamatergic neurons with or without the addition of astrocytes and determined the PGBD5 expression pattern. Cells were collected and analyzed at various time points until day 15.

Our results showed that PGBD5 protein was expressed only on day 5 of differentiation (Figure 42). As expected, we detected no expression in the NPC state; however, the protein levels were overall shallow. There was no difference in expression with the addition of astrocytes.

Based on PGBD5 expression patterns, precisely the high expression in mature neurons from *in-vivo* experiments and the lack of expression in *in-vitro* neuronal cultures, we hypothesized that PGBD5 expression might increase in response to neuronal activity. We tested the hypothesis by conducting another time-course experiment where we stimulated neurons with NMDA. The cells were subjected to NMDA treatment for 10 min on either one, two, or four consecutive days before collection. In parallel, we measured yH2AX levels as a marker for DNA DSBs, and they additionally served as an indicator of successful neuronal stimulation.

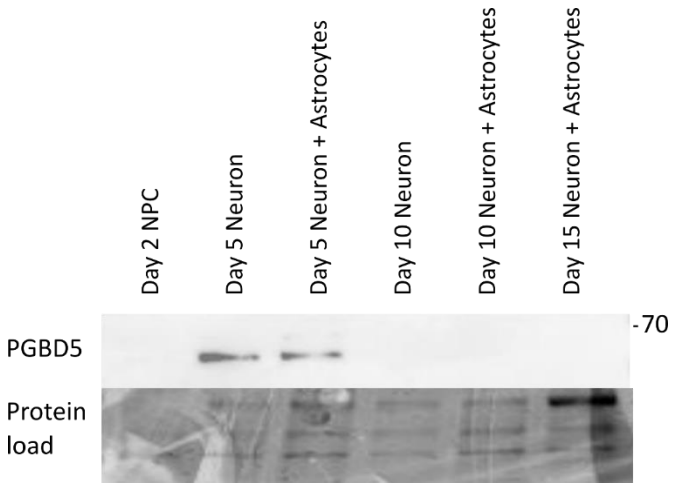


Figure 42: PGBD5 protein expression during neuronal differentiation protocol in mono and co-cultures. Western Blot analysis was performed using an anti-PGBD5 antibody to probe different cell stages. Protein load assessment was conducted using a stain-free gel.

Contrary to the previous experiment, PGBD5 expression peaked on day 5 and decreased in subsequent days (Figure 43). However, we detected a fair amount of PGBD5 during the NPC state (day 2). This discrepancy with PGBD5 RNA expression may be due to its relatively long half-life of approximately 25 hours. Expression was also reintroduced to some extent on day 10.

In addition, TOP2B expression was also measured during the time course (Figure 43). TOP2B expression increased until peaking at day 7 before declining.

In conclusion, our results suggest that PGBD5 expression decreases after day 5 of differentiation and that astrocyte co-cultures show similar expression patterns as neuronal monocultures. TOP2B is lowly expressed throughout the differentiation.

Despite the visual increase in PGBD5 expression in stimulated neurons, the effect was insignificant due to lacking statistical power. We were also concerned about introducing a high background to the control samples by supplementing the media with GlutaMAX.

After each stimulation, we replaced the old media with new media in all wells, including the control. New media contained GlutaMAX, an L-glutamine-based supplement (Thermo Fischer Scientific, January 2023). It can induce excitotoxicity in neuronal cell lines and neurons (Kritis *et al.*, 2015). Glutamine is a precursor for the neurotransmitter glutamate, and when added to the

culture medium, GlutaMAX degrades into L-glutamine (and L-alanine) and, consequently, glutamate (Thermo Fischer Scientific, January 2023). Glutamate stimulates glutamate receptors, including NMDA receptors, and may have similar effects as NMDA stimulation.

We repeated the experiment with modifications. We changed the media to its original after NMDA stimulation, and the control and stimulation media did not contain GlutaMAX.

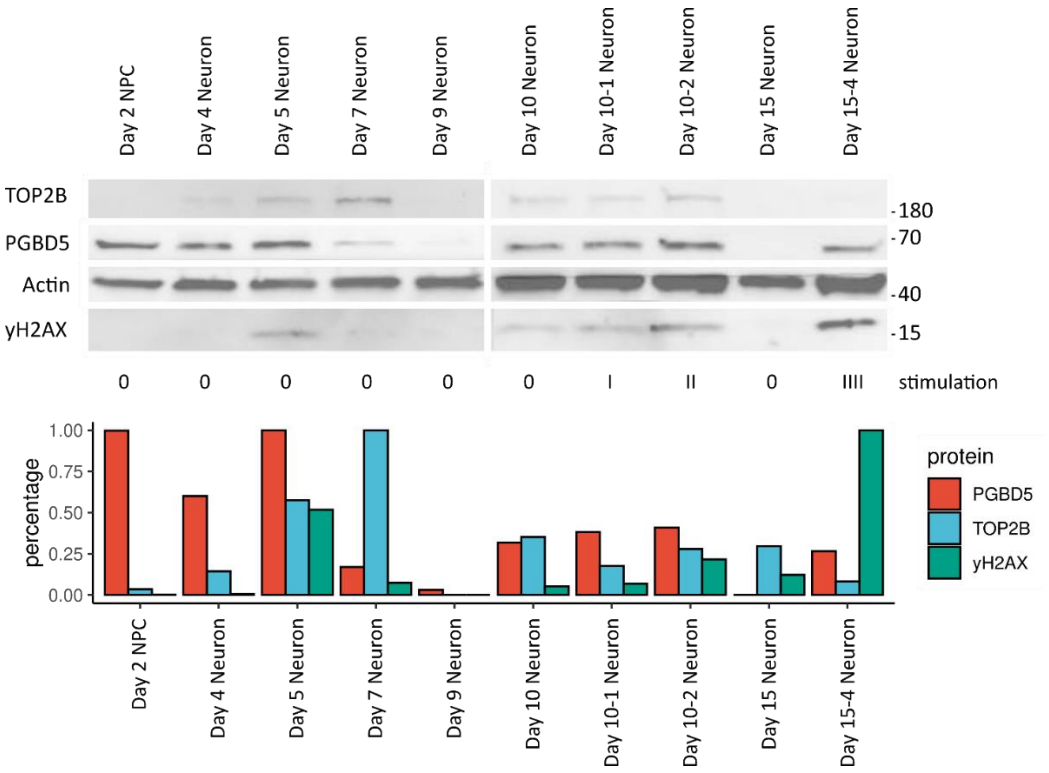


Figure 43: PGBD5 protein expression during neuronal differentiation protocol in mono culture with and without stimulation. The stimulation was performed one, two, or four consecutive days before cell collection. Top panel) Western Blot analysis was performed, and the protein expression was probed using the indicated antibodies. Lower panel) Quantification of the Western Blot results using Actin for load normalization.

The aim was to determine whether neuronal stimulation increases the expression of PGBD5 and to examine the time frame for measuring DNA DSBs indicated by yH2AX expression after stimulation. This was important because PGBD5 was relatively under-expressed in *in-vitro* models of neuronal cell cultures, and stimulation could have helped us to increase the protein

amounts. It also enabled us to estimate the timing of cell collection to study the increased DNA damage.

We differentiated hiPSCs into 15-day-old neurons. On day 14, some neurons underwent NMDA treatment for 10 min (pre-treated condition), and on day 15, all neurons were subjected to the same treatment and collected at 25, 45, or 65 min after recovery. The expression levels of PGBD5, yH2AX, and TOP2B were measured using Western Blot analysis.

The results, summarized in Figure 44, indicate that pre-treated neurons expressed higher levels of PGBD5 than neurons without stimulation on the previous day. This suggests that NMDA treatment increased PGBD5 expression within 24 hours. Furthermore, yH2AX levels increased 65 min after the treatment compared to 25 min after the treatment, making 65 min after the treatment the preferential time point for the yH2AX in ChIP experiments. The expression of TOP2B was confirmed, but it did not show an obvious pattern.

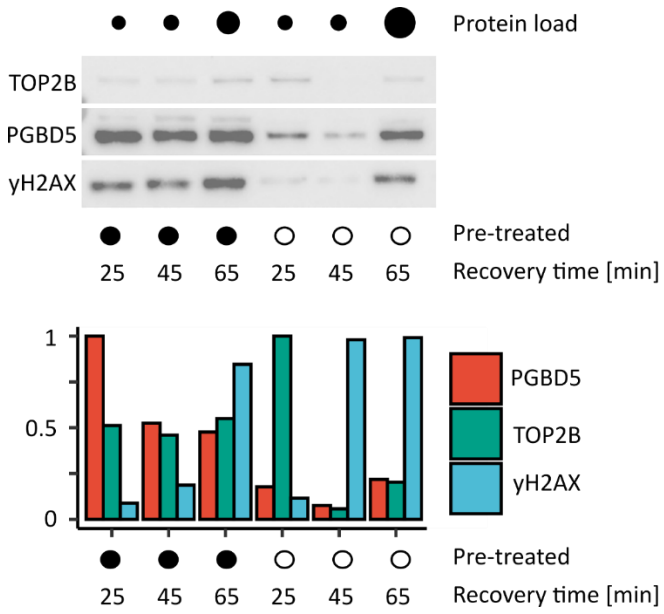


Figure 44: NMDA treatment triggers PGBD5 protein expression in glutamatergic neurons assessed by Western Blot. Glutamatergic Neurons were harvested on day 15. Pre-treated indicates that cells were treated with NMDA one day prior to collection. The recovery time indicates the duration between NMDA treatment and cell collection. Top panel) Western Blot analysis was performed, and the protein expression was probed using the indicated antibodies. Lower panel) Quantification of the Western Blot results, with protein load from a stain-free gel used for normalization.

In addition, we performed RT-PCRs to understand those changes at the mRNA level (Figure 45 & Figure 46). If neuronal stimulation increases PGBD5 protein expression, an increase in mRNA level was expected, although on a shorter time scale. Furthermore, we aimed to test whether the stimulation of neurons with NMDA treatment triggers the transcription of IEGs, an indicator of neuronal stimulation. Finally, PGBD5 mRNA was overexpressed in these cells to examine its influence on EGR1, FOS, and TOP2B expressions.

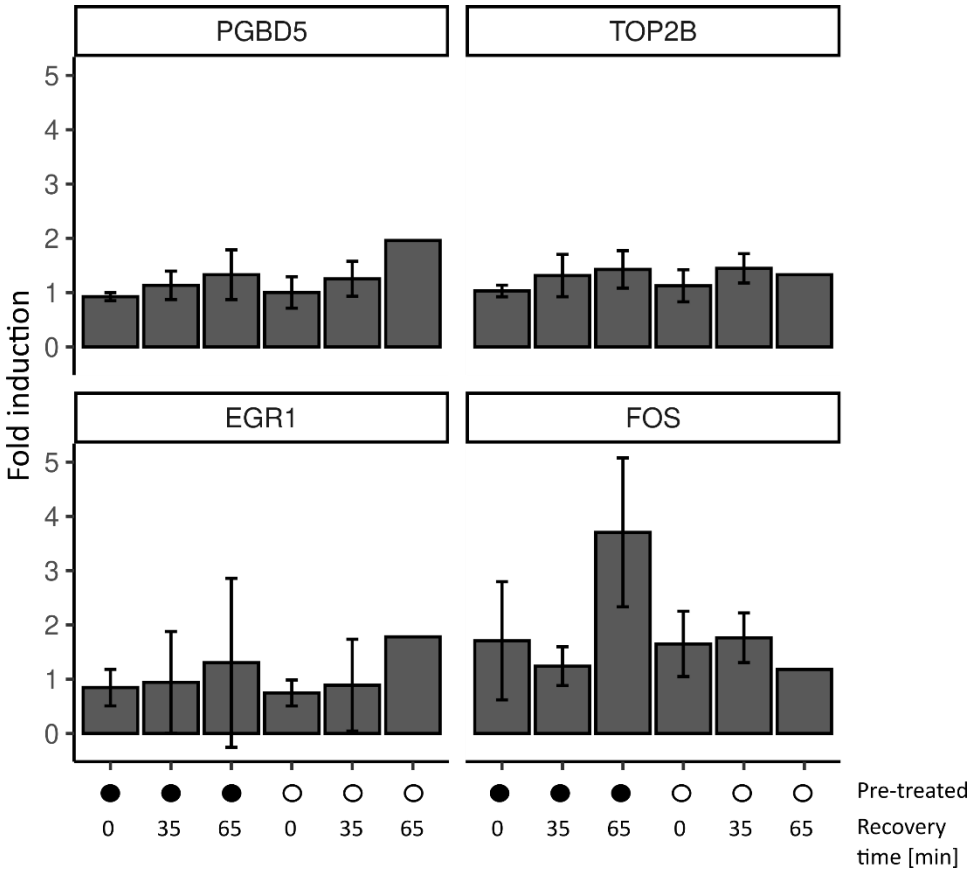


Figure 45: Effect of NMDA treatment on PGBD5 expression in glutamatergic neurons assessed by RT-PCR. Glutamatergic neurons were harvested on day 15. Pre-treated indicates that cells were treated with NMDA one day prior to collection. The recovery time indicates the duration between NMDA treatment and cell collection. MRNA levels of PGBD5, TOP2B, EGR1, and FOS were measured and normalized to Actin.

As shown in Figure 45, we evaluated the relative expressions of PGBD5, TOP2B, EGR1, and FOS in response to stimulation (35 or 65 min after recovery). Importantly, no typical transcriptional burst of IEGs EGR1 and FOS was observed after stimulation, indicating that the stimulation was unsuccessful. Also, we did not find a change in the expression of PGBD5, possibly because of the failed stimulation.

Finally, we transfected PGBD5 mRNA into neurons. Neurons are post-mitotic and difficult to transfect with conventional transfection systems based on plasmids or protein transfection. A well-established alternative is mRNA transfection. Figure 46 shows that the mRNA transfection resulted in a fold induction of approximately 300. We ensured good transfection efficiencies by fluorescence microscopy of GFP mRNA-transfected cells. However, Western Blot analysis showed that the mRNAs did not translate into protein (Figure S 7). The issue requires further investigation.

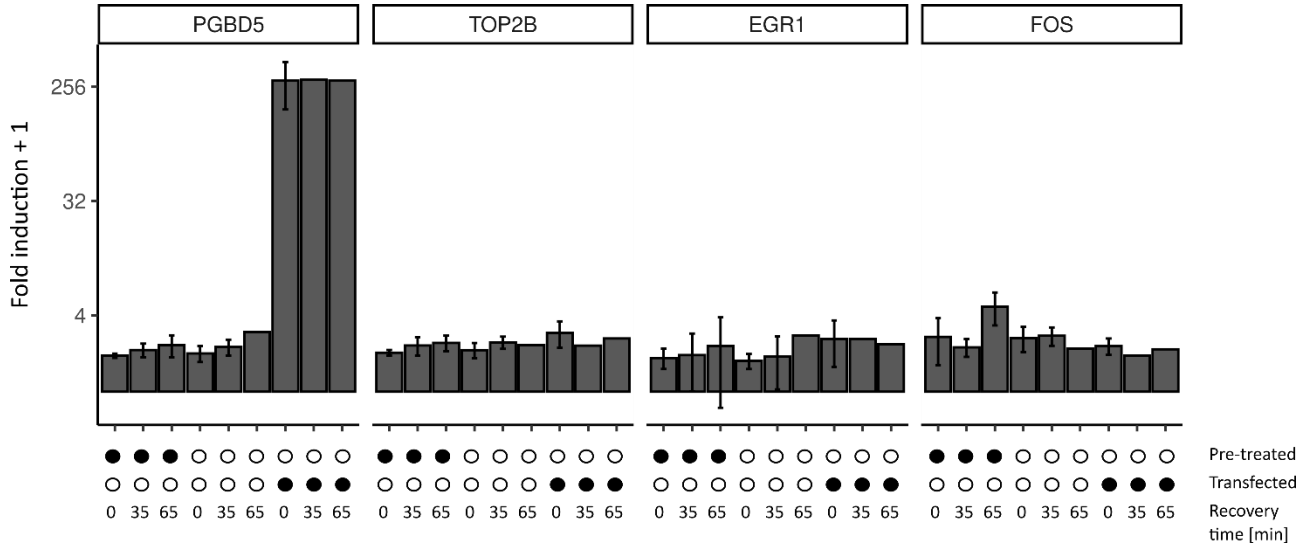


Figure 46: Effect of NMDA treatment on PGBD5 expression in glutamatergic neurons assessed by RT-PCR. Glutamatergic neurons were harvested on day 15. Pre-treated indicates that cells were treated with NMDA one day prior to collection. The recovery time indicates the duration between NMDA treatment and cell collection. MRNA levels of PGBD5, TOP2B, EGR1, and FOS were measured and normalized to Actin, including additional samples of PGBD5 mRNA overexpression.

The results indicate that NMDA treatment of 15-day-old glutamatergic iNeurons led to increased expression of PGBD5 at the protein level. Furthermore, levels of yH2AX showed an increase 65 minutes after NMDA treatment, making this a suitable time point for harvesting cells to examine their DNA damage using yH2AX ChIP experiments. The induction of DNA DSBs, an indicator of neuronal activity, was observed with NMDA treatment. However, this treatment did not result in the upregulation of EGR1 and FOS IEGs at the RNA level, nor did it increase PGBD5 levels. It is important to interpret the Western Blot results cautiously, as the endogenous antibody may lack specificity (4.2). Attempts to pull down PGBD5 or TOP2B in IPs with the appropriate antibodies were unsuccessful, possibly because of the low expression of the proteins. Overall, the model was unsuitable for studying the interaction of PGBD5 and TOP2B and needed further improvement.

4 Discussion

4.1 Part I: PGBD1 not only contains a SCAN but also a KRAB domain, which may have been functionally modified after the domestication event

The study investigated the evolutionary history and functional characteristics of PGBD1 and PGBD2, two genes exclusively found in non-monotreme mammals. The phylogenetic analysis of more than 12,000 sequences of the transposase IS4 family showed that PGBD1 and PGBD2 are mammalian-specific genes, and investigation of missing homology suggests that this is not due to homology detection failure.

The study further annotated the protein domains of PGBD1 and PGBD2 throughout the evolutionary tree, revealing that the N-terminal SCAN and KRAB domains were present in the ancestral condition but underwent losses or decay throughout evolution. The analysis also showed that PGBD1 and PGBD2 are under purifying selection, with the transposase domain under stronger purifying selection than their N-terminal domains. The KRAB domain of PGBD1 was shown to be under positive selection, including one amino acid that might have compromised its TRIM28 binding capacity.

PGBD1 is an exciting example of a host-transposase fusion event (Cosby *et al.*, 2021). In fact, Cosby *et al.* found that the KRAB domain is frequently found in domesticated fusion events compared to other host protein domains. Our study builds on this observation and suggests that in the ancestral condition, the transposase domain of PGBD1 is fused with a SCAN-KRAB domain. Our results suggest that it likely played an important role in acquiring the transposase domain and its expression from a host promoter. However, the KRAB domain of PGBD1 is aberrant in most species and has diverged over evolution. Hence, it appears to have played a submissive role in the protein function. We even identified positive selection in the region, indicating that changes in the KRAB region were beneficial for PGBD1 gene function.

The study's findings are an integral part of the article by Raskó *et al.* (2022), which shows that PGBD1 is a transcriptional activator and repressor of multiple neuronal genes. As a repressor, it suppresses NEAT1 in NPCs. NEAT1 is a long non-coding RNA and the essential core of the mammalian-specific paraspeckle complex. Depletion of PGBD1 in NPCs induced paraspeckle formation and neuronal differentiation. The study suggested that PGBD1 co-evolved with the paraspeckle complex to control the new structure. Further, the study showed that PGBD1 is catalytically inactive and interacts with other SCAN domain-harboring proteins. The DNA binding analysis identified specific SCAN binding motifs but also unknown DNA binding motifs. The results implied that the SCAN domain is functional in PGBD1.

The KRAB domain is commonly found in zinc finger proteins. KRAB-zinc fingers resemble the largest family of transcriptional regulators in humans (Ecco *et al.*, 2017). The KRAB domain recruits TRIM28, and together they repress sequences derived from transposable elements (Ecco *et al.*, 2017). However, KRAB zinc fingers with an additional N-terminal SCAN domain mainly do not bind TRIM28 (Ecco *et al.*, 2017). They are poorly characterized, but a few examples showed that they bind preferentially to promoters and have repressive and activating effects on gene expression (Ecco *et al.*, 2017; Fedotova *et al.*, 2017).

The fact that SCAN-KRAB-zinc fingers mostly do not bind TRIM28 and still exert repressive functions indicates that this group of transcription factors performs repression independently of the TRIM28 pathway. In the case of PGBD1, it might not require the KRAB domain at all. However, the mechanism by which this group of transcriptional regulators exerts its function is still unknown and requires further investigation.

The study’s limitations include that the absence of PGBD1 and PGBD2 in monotremes could have an alternative explanation. The missing homology between human PGBD1 and PGBD2 and monotreme genomes might originate from incomplete genome coverage of sequenced monotreme genome assemblies. Additionally, the KRAB domain diverged but may still play a crucial role in PGBD1 gene regulatory functions.

Our research is focused on gaining a deeper understanding of the functions of domesticated transposases, and to this end, we aim to investigate the specific roles played by the transposase

domain of PGBD1 in DNA binding, as suggested by Cosby *et al.* (2021), as well as determine if the DNA binding function of PGBD1 is solely reliant on its N-terminal host domains. In order to achieve this, we plan to conduct experiments to investigate the DNA binding patterns of individual PGBD1 protein domains.

Additionally, we are interested in exploring the similarities and differences between PGBD1 and its homolog, PGBD2. Although the two genes share sequence homology that extends the borders of their transposase domains, PGBD2 has an intact CRD domain, while PGBD1 has lost this domain but gained a SCAN and a KRAB-like domain. Our investigation will include a comparison of the DNA binding patterns of PGBD2 and PGBD1 as we seek to elucidate the differences between these two genes, which likely arose from a gene duplication event.

4.2 Part II: PGBD5 potentially enhances topoisomerase II activity and regulates IEGs in the brain

In the second part of the project, we aimed to attribute molecular functions to the domesticated DNA transposase PGBD5 and to find insights into whether PGBD5 preserved catalytic activity from its transposase ancestor piggyBac. We found that PGBD5 has a similar folding as piggyBac, as demonstrated in structural alignments with high confidence scores. The 3-D structure is a convenient parameter for classifying transposons by their transposition mechanism (Hickman & Dyda, 2015). However, PGBD5's structural comparison to piggyBac elucidated several features that indicate that it is no longer an active transposase. These include a mutated DDE/D triad, a hallmark of DNA *cut-and-paste* transposition, a lack of a CRD domain that facilitates DNA-specific binding, and missing TIRs resembling DNA transposons' DNA substrate. For the DDE/D triad crucial for piggyBac transposition, including nicking of the DNA (Mitra *et al.*, 2008), Henssen *et al.* (2015) suggested an alternative DDD triad. However, we find that the triad's first aspartic acid residue is located within the DDBD and not in the catalytic part of the protein. Nevertheless, the proximity of the three aspartic residues in the folded state is vital for its mechanism (Nesmelova & Hackett, 2010).

Albeit we find that the transposase and nicking activity of PGBD5 is doubtful, several papers report DNA breaks and genomic rearrangements that are associated with PGBD5 expression (Henssen, Koche, *et al.*, 2017; Henssen, Reed, *et al.*, 2017; Jubierre Zapater *et al.*, 2023; Simi *et al.*, 2023). Hence, we aimed to identify co-factors with which PGBD5 might facilitate those events. We elucidated PGBD5's protein-protein interaction partners using SILAC-based AP-MS technology. We found that it interacts with multiple genes classified as histone methyltransferases, regulators of transcription, and DNA repair proteins, suggesting that PGBD5 itself exerts gene regulatory functions. Chromatin accessibility factors and DNA repair factors contribute to genomic rearrangements and DNA breaks. Thus, the interaction with almost any of these factors reflects a possible way by which PGBD5 facilitates these events. However, we also identified a protein that directly induces DSBs. We identified an interaction of TOP2A with PGBD5. TOP2A is an enzyme that initiates temporary DSBs and re-ligates the broken DNA ends. While *in-vitro,* the religation is independent of DNA repair factors, *in-vivo* studies have shown that the DNA repair pathways are often engaged in the repair (Riccio *et al.*, 2020). TOP2 activity solves topological problems arising from the DNA structure and organization, but the initiated DNA DSBs also pose a risk, as they are associated with cancerous genomic rearrangements (Canela *et al.*, 2017). Hence this study elucidated several co-factors with which PGBD5 might facilitate DNA DSBs and genomic rearrangements.

Topoisomerases II serves many host functions, and they were also found to exert gene regulatory functions (Herrero-Ruiz *et al.*, 2021; Madabhushi, 2018), such as the regulation of a handful of genes, predominantly IEGs. Our transcriptome analysis of PGBD5 knockout mice identified a decreased expression of four IEGs, namely Fos, Npas4, Nr4a1, and Dusp1, in two different brain regions, the hippocampus and the cerebellum. All of these genes are controlled by topoisomerases II. These findings encouraged us to investigate the interaction of TOP2A and PGBD5 further. We found that PGBD5 potentially enhances TOP2A functions, as indicated by an *in-vitro* decatenation assay. To study the mechanism *in-vivo,* we examined the use of neuronal differentiation systems. We differentiated human iPSCs to neurons and found that this system is not applicable, mainly due to the low expression of TOP2B and PGBD5 in the differentiated neuronal cells.

4.2.1 Limitations of the study

However, several study limitations must be addressed to prove our hypothesis. We did show that the neural-specific gene PGBD5 interacts with TOP2A and increases TOP2A enzymatic activity *in-vitro*. However, *in-vivo* validation for the enhancing effect of PGBD5 in topoisomerase II activity is missing. Additionally, neurons exclusively express TOP2B, a paralogue of TOP2A. The activity and interactions with TOP2B still need to be addressed in co-IP and decatenation assays.

The protein purification of HA-PGBD5 was achieved using HA-agarose beads in HEK293 cells. The purified lysates were used for the decatenation assays. The purification included sonication steps that have weakened its interaction with other proteins, such as TOP2A. However, the purified protein lysate (>70% purity) still contained small amounts of other proteins that could facilitate decatenation. Although we used appropriate controls, a more elegant solution would be to use a protein purification method that contains a size selection step.

The increased purity of PGBD5 would also help us to evaluate whether PGBD5 stabilizes the DNA DSBs, by quantifying the linear product in the catenation assay. We observed increased linear DNA compared to nicked DNA in samples containing PGBD5. However, more repetitions and better quantification methods will be needed to conclude.

Along with those mentioned above, we also observed issues with the PGBD5 antibodies. During Western Blot analysis of iPSCs with PGBD5 knockout, two of the commercially available endogenous PGBD5 antibodies (namely NBP2-67048 and MBS355128) were found to be unspecific in the used iPSCs (Figure 22 & Figure 23). We found in all samples a visible band, although genomic PCR confirmed a knockout by a frameshift in the third exon. Therefore, we investigated the antibody specificity in other cell lines. Overexpression and downregulation of PGBD5 did not change the band intensity in Western Blot analysis (Figure S 8). Using endogenous antibodies, we attempted to pull down overexpressed HA-tagged PGBD5 in HEK293 cells. The pull-down resulted in either extremely low or non-detectable quantities.

Altogether these results showed that the PGBD5 antibodies NBP2-67048 and MBS355128 are unsuitable for PGBD5 detection. After testing all commercially available PGBD5 antibodies (NBP2-67048, MBS355128, HPA065010, orb13159, and ABIN1854961), we only found one antibody, HPA065010, that successfully pulls down HA-tagged PGBD5 (Figure S 10). Although the co-IP experiments were already validated with HA-tagged PGBD5 and anti-HA antibody, some experiments, including ChIP-WB and co-IPs, should be repeated with the new antibody. However, the antibody is only applicable in IP and ICC assays. Hence Western Blot to validate PGBD5 protein expression will not be possible with the new antibody.

4.2.2 Discrepancies found in previous studies

We also identified further issues in the studies Henssen *et al.* (2015) and Henssen, Reed, *et al.* (2017). The shRNA plasmids used in those studies did not result in a knockdown of PGBD5 in HEK293 cells assessed by RT-PCR. Instead, we observed elevated PGBD5 levels after 48h compared to the control targeting GFP (shGFP). When we investigated the shRNAs we found that shPGBD5-3 (5'-CCAGATTTATGTCCACCTGAA-3') is not targeting the PGBD5 locus but an intergenic region on chromosome 13 according to BLAT search in UCSC genome browser. The second shRNA, shPGBD5-1 (5'-CCTCGTCCTCACTCAGTTATT-3'), targets the 3'UTR that is only present in one of the annotated PGBD5 isoforms.

We also found that the RT-PCR primers (forward: 5'-GCTTATTCTTCAGCGCATCC-3'; reverse: 5'-CAGCCTCTGGGTCAGACAAT-3') of these studies measuring PGBD5 mRNA levels targets the reverse complement sequence and will not amplify PGBD5 mRNA. Their PGBD5 plasmid was based on an old NCBI entry and is missing the first ~ 100 amino acids of PGBD5; instead, there is an N-terminal cloning artifact in the sequence. Moreover, the endogenous antibody used in the study is not working anymore, which was confirmed by the lab of Alex Kentsis. Therefore, we are puzzled about how they could achieve consistency and soundness in their findings.

4.2.3 Upcoming experiments

To address the limitations of this study, we plan to conduct co-IP experiments of TOP2B and PGBD5 in mouse brain samples. Specifically, we have already dissected hippocampi and visual cortices from wild-type mice (strain: C57BL/6J, male, 17-18 weeks old) and obtained protein lysates for the upcoming IPs. Through this experiment, we hope to gain insight into the interaction between TOP2B and PGBD5.

Moreover, we aim to validate the enhancing effect of PGBD5 on topoisomerase II *in-vivo*. To achieve this, we plan to quantify topoisomerase poisons in cells with varying levels of PGBD5 by using a popular assay called the *Rapid approach to DNA adduct recovery* (RADAR) assay (Kiianitsa & Maizels, 2013). This assay relies on the purification of genomic DNA from cells, followed by measurements of covalently bound TOP2 protein. By applying this assay, we can determine whether PGBD5 increases TOP2 activity *in-vivo*.

4.2.4 Significance of the study

Until now, the catalytic activity of PGBD5 has been heavily debated. While some studies found that PGBD5 exhibits transposition on naked DNA substrate and recombination activity on chromatinized DNA (Helou, Beauclair, Dardente, Arensburger, *et al.*, 2021; Helou, Beauclair, Dardente, Piégu, *et al.*, 2021; Henssen *et al.*, 2015, 2016; Henssen, Koche, *et al.*, 2017) other studies could not replicate those results and found that it is unable to bind DNA, bind the naked DNA substrate or excise the substrate (Beckermann *et al.*, 2021; Kolacsek *et al.*, 2022). This study is the first to provide an alternative mechanism by which PGBD5 might confer to DNA damage and genomic rearrangements that have been reported (Henssen *et al.*, 2016; Henssen, Koche, *et al.*, 2017; Henssen, Reed, *et al.*, 2017; Jubierre Zapater *et al.*, 2023; Simi *et al.*, 2023). Our study suggests that PGBD5 increases the activity of topoisomerase II, a protein known to promote genomic rearrangements and DSBs. We also show that PGBD5 binds to several DNA repair proteins that might be involved in the DNA repair of topoisomerase II-induced DSBs.

The interaction between PGBD5 and topoisomerase II provides a comprehensive explanation for various phenomena associated with PGBD5 expression. Firstly, PGBD5 has been implicated in cancer cell proliferation (Xie *et al.*, 2022), which can be attributed to its enhancing effect on TOP2A. TOP2A has consistently been shown to promote cell proliferation (Heck & Earnshaw, 1986; Hsieh *et al.*, 2015; Lynch *et al.*, 1997; Miyata *et al.*, 2006).

Moreover, PGBD5 has been found to play a role in the migration of cortical neurons during development and in the cell cycle exit of NPCs transitioning into newborn neurons. Interestingly, both of these processes are also positively regulated by TOP2B (K. M. Tsutsui *et al.*, 2006).

Furthermore, PGBD5 has been found to be associated with DSBs in neocortical NPCs during the process of differentiation. Interestingly, this coincides with the transition from TOP2A to TOP2B expression, during which neuronal promoters become occupied by TOP2B (Tiwari *et al.*, 2012). The increased expression of TOP2B during this period, along with its enhanced occupancy at promoters and the facilitating effect of PGBD5 on TOP2B, potentially explains the observed increase in DSBs. Notably, these promoters belong among other pathways to genes that promote the generation of neurons, and knockout studies of TOP2B have confirmed their increasing transcription (Tiwari *et al.*, 2012). Conversely, downregulated genes included those involved in cell differentiation (Tiwari *et al.*, 2012). These findings closely parallel the observed effects of PGBD5. (Simi *et al.*, 2023) have shown that cells expressing PGBD5 exhibited markers indicative of neuronal cell determination, and depletion of PGBD5 led to increased expression of genes associated with the proliferation and renewal of NPCs (Simi *et al.*, 2023).

Additionally, we discovered that PGBD5 controls IEGs such as FOS, NR4A1, DUSP1, and NPAS4 in conjunction with TOP2, thus adding a layer of IEG regulation. These IEGs play a fundamental role in brain functioning, and their dysregulation has been implicated in various diseases including schizophrenia, anxiety disorder and major depressive disorder (Gammie, 2022; Jaehne *et al.*, 2015; Nehme *et al.*, 2022). Thus, our study elucidated PGBD5 as a potential target for therapeutic interventions aimed at these diseases.

5 References

Andrews, S. (2010). FastQC. *Babraham Bioinformatics*.

Austin, C. A., & Marsh, K. L. (1998). Eukaryotic DNA topoisomerase IIβ. *Bioessays*, *20*(3), 215–226.

Bailey, A. D., Gray, L. T., Pavelitz, T., Newman, J. C., Horibata, K., Tanaka, K., & Weiner, A. M. (2012). The conserved Cockayne syndrome B-piggyBac fusion protein (CSB-PGBD3) affects DNA repair and induces both interferon-like and innate antiviral responses in CSB-null cells. *DNA Repair*, *11*(5), 488–501.

Baranello, L., Wojtowicz, D., Cui, K., Devaiah, B. N., Chung, H.-J., Chan-Salis, K. Y., Guha, R., Wilson, K., Zhang, X., & Zhang, H. (2016). RNA polymerase II regulates topoisomerase 1 activity to favor efficient transcription. *Cell*, *165*(2), 357–371.

Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., & Bétermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. *Genes and Development*, *23*(21). https://doi.org/10.1101/gad.547309

Beckermann, T. M., Luo, W., Wilson, C. M., Veach, R. A., & Wilson, M. H. (2021). Cognate restriction of transposition by piggyBac-like proteins. *Nucleic Acids Research*, *49*(14). https://doi.org/10.1093/nar/gkab578

Belbin, O., Carrasquillo, M. M., Crump, M., Culley, O. J., Hunter, T. A., Ma, L., Bisceglio, G., Zou, F., Allen, M., & Dickson, D. W. (2011). Investigation of 15 of the top candidate genes for late-onset Alzheimer's disease. *Human Genetics*, *129*, 273–282.

Biersack, H., Jensen, S., Gromova, I., Nielsen, I. S., Westergaard, O., & Andersen, A. H. (1996). Active heterodimers are formed from human DNA topoisomerase II alpha and II beta isoforms. *Proceedings of the National Academy of Sciences*, *93*(16), 8288–8293.

Bischerour, J., Bhullar, S., Denby Wilkes, C., Régnier, V., Mathy, N., Dubois, E., Singh, A., Swart, E., Arnaiz, O., & Sperling, L. (2018). Six domesticated PiggyBac transposases together carry out programmed DNA elimination in Paramecium. *Elife*, *7*, e37927.

Borchers, H. W. (2022). *pracma: Practical Numerical Math Functions*. https://CRAN.R-project.org/package=pracma

Bouallègue, M., Rouault, J. D., Hua-Van, A., Makni, M., & Capy, P. (2017). Molecular Evolution of piggyBac Superfamily: From Selfishness to Domestication. *Genome Biology and Evolution*, *9*(2). https://doi.org/10.1093/gbe/evw292

Boutros, S. W., Krenik, D., Holden, S., Unni, V. K., & Raber, J. (2022). Common cancer treatments targeting DNA double strand breaks affect long-term memory and relate to immediate early gene expression in a sex-dependent manner. *Oncotarget*, *13*(1). https://doi.org/10.18632/ONCOTARGET.28180

BrainSpan. (2022, October 24). Http://Brainspan.Org.

Brito, V., Montalban, E., Sancho-Balsells, A., Pupak, A., Flotta, F., Masana, M., Ginés, S., Alberch, J., Martin, C., Girault, J. A., & Giralt, A. (2022). Hippocampal Egr1-Dependent Neuronal Ensembles Negatively Regulate Motor Learning. *Journal of Neuroscience*, *42*(27). https://doi.org/10.1523/JNEUROSCI.2258-21.2022

Broce, I., Karch, C. M., Wen, N., Fan, C. C., Wang, Y., Hong Tan, C., Kouri, N., Ross, O. A., Höglinger, G. U., & Muller, U. (2018). Immune-related genetic enrichment in frontotemporal dementia: An analysis of genome-wide association studies. *PLoS Medicine*, *15*(1), e1002487.

Buzun, K., Bielawska, A., Bielawski, K., & Gornowicz, A. (2020). DNA topoisomerases as molecular targets for anticancer drugs. In *Journal of Enzyme Inhibition and Medicinal Chemistry* (Vol. 35, Issue 1). https://doi.org/10.1080/14756366.2020.1821676

Canela, A., Maman, Y., Jung, S., Wong, N., Callen, E., Day, A., Kieffer-Kwon, K. R., Pekowska, A., Zhang, H., Rao, S. S. P., Huang, S. C., Mckinnon, P. J., Aplan, P. D., Pommier, Y., Aiden, E. L., Casellas, R., & Nussenzweig, A. (2017). Genome Organization Drives Chromosome Fragility. *Cell*, *170*(3). https://doi.org/10.1016/j.cell.2017.06.034

Capranico, G., Tinelli, S., Austin, C. A., Fisher, M. L., & Zunino, F. (1992). Different patterns of gene expression of topoisomerase II isoforms in differentiated tissues during murine development. *BBA - Gene Structure and Expression*, *1132*(1). https://doi.org/10.1016/0167-4781(92)90050-A

Carducci, F., Barucca, M., Canapa, A., Carotti, E., & Biscotti, M. A. (2020). Mobile elements in ray-finned fish genomes. *Life*, *10*(10), 221.

Carpentier, M., Chomilier, J., & Valencia, A. (2019). Protein multiple alignments: Sequence-based versus structure-based programs. *Bioinformatics*, *35*(20). https://doi.org/10.1093/bioinformatics/btz236

Cary, L. C., Goebel, M., Corsaro, B. G., Wang, H. G., Rosen, E., & Fraser, M. J. (1989). Transposon mutagenesis of baculoviruses: Analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology*, *172*(1). https://doi.org/10.1016/0042-6822(89)90117-7

Chaly, N., Chen, X., Dentry, J., & Brown, D. L. (1996). Organization of DNA topoisomerase II isotypes during the cell cycle of human lymphocytes and HeLa cells. *Chromosome Research*, *4*(6). https://doi.org/10.1007/BF02265053

Chen, Q., Luo, W., Veach, R. A., Hickman, A. B., Wilson, M. H., & Dyda, F. (2020). Structural basis of seamless excision and specific targeting by piggyBac transposase. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-17128-1

Cheng, C. Y., Vogt, A., Mochizuki, K., & Yao, M. C. (2010). A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila. *Molecular Biology of the Cell*, *21*(10). https://doi.org/10.1091/mbc.E09-12-1079

Cheng, C. Y., Young, J. M., Lin, C. Y. G., Chao, J. L., Malik, H. S., & Yao, M. C. (2016). The piggyBac transposon-derived genes TPB1 and TPB6 mediate essential transposon-like excision during the developmental rearrangement of key genes in Tetrahymena thermophila. *Genes and Development*, *30*(24). https://doi.org/10.1101/gad.290460.116

Close, J. L., Yao, Z., Levi, B. P., Miller, J. A., Bakken, T. E., Menon, V., Ting, J. T., Wall, A., Krostag, A. R., Thomsen, E. R., Nelson, A. M., Mich, J. K., Hodge, R. D., Shehata, S. I., Glass, I. A., Bort, S., Shapovalova, N. V., Ngo, N. K., Grimley, J. S., … Lein, E. (2017). Single-Cell Profiling of an In Vitro Model of Human Interneuron Development Reveals Temporal Dynamics of Cell Type Production and Maturation. *Neuron*, *93*(5). https://doi.org/10.1016/j.neuron.2017.02.014

Collins, M., & Rubin, G. M. (1984). Structure of chromosomal rearrangements induced by the FB transposable element in Drosophila. *Nature*, *308*(5957). https://doi.org/10.1038/308323a0

Cosby, R. L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E. J., & Feschotte, C. (2021). Recurrent evolution of vertebrate transcription factors by transposase capture. *Science*, *371*(6531). https://doi.org/10.1126/science.abc6405

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*(12). https://doi.org/10.1038/nbt.1511

Dimitrova, E., Kondo, T., Feldmann, A., Nakayama, M., Koseki, Y., Konietzny, R., Kessler, B. M., Koseki, H., & Klose, R. J. (2018). FBXl19 recruits CDK-Mediator to CpG islands of developmental genes priming them for activation during lineage commitment. *ELife*, *7*. https://doi.org/10.7554/eLife.37084

Dobin, A., & Gingeras, T. R. (2015). Mapping RNA-seq Reads with STAR. *Current Protocols in Bioinformatics*, *51*(1). https://doi.org/10.1002/0471250953.bi1114s51

Dubois, E., Mathy, N., Régnier, V., Bischerour, J., Baudry, C., Trouslard, R., & Bétermier, M. (2017). Multimerization properties of PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements. *Nucleic Acids Research*, *45*(6). https://doi.org/10.1093/nar/gkw1359

Ecco, G., Imbeault, M., & Trono, D. (2017). KRAB zinc finger proteins. *Development*, *144*(15), 2719–2729.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5). https://doi.org/10.1093/nar/gkh340

EMBO Single-Cell Atlas. (2022, August 29). Https://Www.Ebi.Ac.Uk/Teichmann-Srv/Esca/.

Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B., & Sved, J. (1990). High-frequency P element loss in Drosophila is homolog dependent. *Cell*, *62*(3). https://doi.org/10.1016/0092-8674(90)90016-8

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19). https://doi.org/10.1093/bioinformatics/btw354

Fedotova, A. A., Bonchuk, A. N., Mogila, V. A., & Georgiev, P. G. (2017). C2H2 zinc finger proteins: The largest but poorly explored family of higher eukaryotic transcription factors. In *Acta Naturae* (Vol. 9, Issue 2). https://doi.org/10.32607/20758251-2017-9-2-47-58

Felsenstein, J. (2009). PHYLIP - Phylogeny Inference Package, Version 3.69. (Seattle, WA: University of Washington). *The American Naturalist*, *171*(6).

Feng, M., Wang, Y., Bi, L., Zhang, P., Wang, H., Zhao, Z., Mao, J.-H., & Wei, G. (2021). CRL4ADTL degrades DNA-PKcs to modulate NHEJ repair and induce genomic instability and subsequent malignant transformation. *Oncogene*, *40*(11), 2096–2111. https://doi.org/10.1038/s41388-021-01690-z

Feschotte, C., & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. In *Annual Review of Genetics* (Vol. 41). https://doi.org/10.1146/annurev.genet.40.110405.090448

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., Chang, H. Y., Dosztanyi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G. L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., … Mitchell, A. L. (2017). InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Research*, *45*(D1). https://doi.org/10.1093/nar/gkw1107

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, *39*(SUPPL. 2). https://doi.org/10.1093/nar/gkr367

Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J. E., Mudge, J. M., Sisu, C., Wright, J. C., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Boix, C., Sala, S. C., Cunningham, F., Domenico, T. Di, Donaldson, S., Fiddes, I. T., Girón, C. G., … Flicek, P. (2021). GENCODE 2021. *Nucleic Acids Research*, *49*(D1). https://doi.org/10.1093/nar/gkaa1087

Fraser, M. J., Ciszczon, T., Elick, T., & Bauser, C. (1996). Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Molecular Biology*, *5*(2). https://doi.org/10.1111/j.1365-2583.1996.tb00048.x

Fraser, M. J., Smith, G. E., & Summers, M. D. (1983). Acquisition of Host Cell DNA Sequences by Baculoviruses: Relationship Between Host DNA Insertions and FP Mutants of Autographa californica and Galleria mellonella Nuclear Polyhedrosis Viruses . *Journal of Virology*, *47*(2). https://doi.org/10.1128/jvi.47.2.287-300.1983

Gallo, F. T., Katche, C., Morici, J. F., Medina, J. H., & Weisstaub, N. V. (2018). Immediate Early Genes, Memory and Psychiatric Disorders: Focus on c-Fos, Egr1 and Arc. *Frontiers in Behavioral Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnbeh.2018.00079

Gamba, R., & Fachinetti, D. (2020). From evolution to function: Two sides of the same CENP-B coin? *Experimental Cell Research*, *390*(2), 111959.

Gammie, S. C. (2022). Evaluation of animal model congruence to human depression based on large-scale gene expression patterns of the CNS. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-021-04020-1

Gandolfo, L. C., & Speed, T. P. (2018). RLE plots: Visualizing unwanted variation in high dimensional data. *PLoS ONE*, *13*(2). https://doi.org/10.1371/journal.pone.0191629

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). *Protein identification and analysis tools on the ExPASy server*. Springer.

Gilroy, K. L., & Austin, C. A. (2011). The impact of the C-Terminal domain on the interaction of human DNA topoisomerase II α and β with DNA. *PLoS ONE*, *6*(2). https://doi.org/10.1371/journal.pone.0014693

Gogol-Doring, A., Ammar, I., Gupta, S., Bunse, M., Miskey, C., Chen, W., Uckert, W., Schulz, T. F., Izsvak, Z., & Ivics, Z. (2016). Genome-wide profiling reveals remarkable parallels between insertion site selection properties of the MLV retrovirus and the piggyBac transposon in primary human CD4+ T cells. *Molecular Therapy*, *24*(3). https://doi.org/10.1038/mt.2016.11

Gray, L. T., Fong, K. K., Pavelitz, T., & Weiner, A. M. (2012). *Tethering of the conserved piggyBac transposase fusion protein CSB-PGBD3 to chromosomal AP-1 proteins regulates expression of nearby genes in humans*.

Gray, Y. H. M. (2000). It takes two transposons to tango: Transposable-element-mediated chromosomal rearrangements. In *Trends in Genetics* (Vol. 16, Issue 10). https://doi.org/10.1016/S0168-9525(00)02104-1

GTEx Consortium. (2022, September 10). Https://Www.Gtexportal.Org/Home/Datasets.

Heck, M. M. S., & Earnshaw, W. C. (1986). Topoisomerase II: A specific marker for cell proliferation. *Journal of Cell Biology*, *103*(6). https://doi.org/10.1083/jcb.103.6.2569

Helou, L., Beauclair, L., Dardente, H., Arensburger, P., Buisine, N., Jaszczyszyn, Y., Guillou, F., Lecomte, T., Kentsis, A., & Bigot, Y. (2021). The C-terminal domain of piggyBac transposase is not required for DNA transposition. *Journal of Molecular Biology*, *433*(7), 166805.

Helou, L., Beauclair, L., Dardente, H., Piégu, B., Tsakou-Ngouafo, L., Lecomte, T., Kentsis, A., Pontarotti, P., & Bigot, Y. (2021). The piggyBac-derived protein 5 (PGBD5) transposes both the closely and the distantly related piggyBac-like elements Tcr-pble and Ifp2: PGBD5 can mobilize with Tcr-pble and Ifp2. *Journal of Molecular Biology*, *433*(7). https://doi.org/10.1016/j.jmb.2021.166839

Henssen, A. G., Henaff, E., Jiang, E., Eisenberg, A. R., Carson, J. R., Villasante, C. M., Ray, M., Still, E., Burns, M., Gandara, J., Feschotte, C., Mason, C. E., & Kentsis, A. (2015). Genomic

DNA transposition induced by human PGBD5. *ELife*, *4*(September). https://doi.org/10.7554/eLife.10565

Henssen, A. G., Jiang, E., Zhuang, J., Pinello, L., Socci, N. D., Koche, R., Gonen, M., Villasante, C. M., Armstrong, S. A., Bauer, D. E., Weng, Z., & Kentsis, A. (2016). Forward genetic screen of human transposase genomic rearrangements. *BMC Genomics*, *17*(1). https://doi.org/10.1186/s12864-016-2877-x

Henssen, A. G., Koche, R., Zhuang, J., Jiang, E., Reed, C., Eisenberg, A., Still, E., Macarthur, I. C., Rodríguez-Fos, E., Gonzalez, S., Puiggròs, M., Blackford, A. N., Mason, C. E., De Stanchina, E., Gönen, M., Emde, A. K., Shah, M., Arora, K., Reeves, C., … Kentsis, A. (2017). PGBD5 promotes site-specific oncogenic mutations in human tumors. *Nature Genetics*, *49*(7). https://doi.org/10.1038/ng.3866

Henssen, A. G., Reed, C., Jiang, E., Garcia, H. D., Von Stebut, J., MacArthur, I. C., Hundsdoerfer, P., Kim, J. H., De Stanchina, E., Kuwahara, Y., Hosoi, H., Ganem, N. J., Dela Cruz, F., Kung, A. L., Schulte, J. H., Petrini, J. H., & Kentsis, A. (2017). Therapeutic targeting of PGBD5-induced DNA repair dependency in pediatric solid tumors. *Science Translational Medicine*, *9*(414). https://doi.org/10.1126/scitranslmed.aam9078

Herrero-Ruiz, A., Martínez-García, P. M., Terrón-Bautista, J., Millán-Zambrano, G., Lieberman, J. A., Jimeno-González, S., & Cortés-Ledesma, F. (2021). Topoisomerase IIα represses transcription by enforcing promoter-proximal pausing. *Cell Reports*, *35*(2). https://doi.org/10.1016/j.celrep.2021.108977

Hickman, A. B., & Dyda, F. (2015). Mechanisms of DNA Transposition. *Microbiology Spectrum*, *3*(2). https://doi.org/10.1128/microbiolspec.mdna3-0034-2014

Hickman, A. B., & Dyda, F. (2016). DNA Transposition at Work. In *Chemical Reviews* (Vol. 116, Issue 20). https://doi.org/10.1021/acs.chemrev.6b00003

Hikosaka, A., Kobayashi, T., Saito, Y., & Kawahara, A. (2007). Evolution of the Xenopus piggyBac transposon family TxpB: Domesticated and untamed strategies of transposon subfamilies. *Molecular Biology and Evolution*, *24*(12). https://doi.org/10.1093/molbev/msm191

Hirose, S., & Suzuki, Y. (1988). In vitro transcription of eukaryotic genes is affected differently by the degree of DNA supercoiling. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(3). https://doi.org/10.1073/pnas.85.3.718

Hofbrucker-MacKenzie, S. A., Seemann, E., Westermann, M., Qualmann, B., & Kessels, M. M. (2023). Long-term depression in neurons involves temporal and ultra-structural dynamics

of phosphatidylinositol-4,5-bisphosphate relying on PIP5K, PTEN and PLC. *Communications Biology*, *6*(1), 366. https://doi.org/10.1038/s42003-023-04726-0

Honigmann, A., van den Bogaart, G., Iraheta, E., Risselada, H. J., Milovanovic, D., Mueller, V., Müllar, S., Diederichsen, U., Fasshauer, D., Grubmüller, H., Hell, S. W., Eggeling, C., Kühnel, K., & Jahn, R. (2013). Phosphatidylinositol 4,5-bisphosphate clusters act as molecular beacons for vesicle recruitment. *Nature Structural & Molecular Biology*, *20*(6), 679–686. https://doi.org/10.1038/nsmb.2570

Hsieh, M. H., Tsai, C. H., Lin, C. C., Li, T. K., Hung, T. W., Chang, L. Te, Hsin, L. W., & Teng, S. C. (2015). Topoisomerase II inhibition suppresses the proliferation of telomerase-negative cancers. *Cellular and Molecular Life Sciences*, *72*(9). https://doi.org/10.1007/s00018-014-1783-0

Huelsenbeck, J. P., & Ronquist, F. (2000). MrBayes: Bayesian inferences of phylogeny. *Bioinformatics*, *17*.

Human Protein Atlas. (2023, January 23). Human Protein Atlas.

Hurst, L. D. (2002). The Ka/Ks ratio: Diagnosing the form of sequence evolution. In *Trends in Genetics* (Vol. 18, Issue 9). https://doi.org/10.1016/S0168-9525(02)02722-1

Jaehne, E. J., Klarić, T. S., Koblar, S. A., Baune, B. T., & Lewis, M. D. (2015). Effects of Npas4 deficiency on anxiety, depression-like, cognition and sociability behaviour. *Behavioural Brain Research*, *281*. https://doi.org/10.1016/j.bbr.2014.12.044

Jeanneteau, F., Barrére, C., Vos, M., De Vries, C. J. M., Rouillard, C., Levesque, D., Dromard, Y., Moisan, M. P., Duric, V., Franklin, T. C., Duman, R. S., Lewis, D. A., Ginsberg, S. D., & Arango-Lievano, M. (2018). The stress-induced transcription factor NR4a1 adjusts mitochondrial function and synapse number in prefrontal cortex. *Journal of Neuroscience*, *38*(6). https://doi.org/10.1523/JNEUROSCI.2793-17.2017

Jeong, J., Juhn, K., Lee, H., Kim, S. H., Min, B. H., Lee, K. M., Cho, M. H., Park, G. H., & Lee, K. H. (2007). SIRT1 promotes DNA repair activity and deacetylation of Ku70. *Experimental and Molecular Medicine*, *39*(1). https://doi.org/10.1038/emm.2007.2

Ju, B.-G., Lunyak, V. V, Perissi, V., Garcia-Bassets, I., Rose, D. W., Glass, C. K., & Rosenfeld, M. G. (2006). A topoisomerase IIß-mediated dsDNA break required for regulated transcription. *Science*, *312*(5781), 1798–1802.

Jubierre Zapater, L., Rodriguez Fos, E., Planas-Felix, M., Lewis, S., Cameron, D., Demarest, P., Nabila, A., Zhao, J., Bergin, P., & Reed, C. (2023). A transposase-derived gene required for human brain development. *BioRxiv*, 2023–2024.

Juhász, S., Elbakry, A., Mathes, A., & Löbrich, M. (2018). ATRX Promotes DNA Repair Synthesis and Sister Chromatid Exchange during Homologous Recombination. *Molecular Cell*, *71*(1). https://doi.org/10.1016/j.molcel.2018.05.014

Jurka, J. (2000). Repbase Update: A database and an electronic journal of repetitive elements. In *Trends in Genetics* (Vol. 16, Issue 9). https://doi.org/10.1016/S0168-9525(00)02093-X

Kanehisa, M., & Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Science*, *29*(1). https://doi.org/10.1002/pro.3711

Kapitonov, V. V, & Jurka, J. (2001). Rolling-circle transposons in eukaryotes. *Proceedings of the National Academy of Sciences*, *98*(15), 8714–8719.

Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J. D., Malinsky, S., & Bétermier, M. (2011). Highly precise and developmentally programmed genome assembly in paramecium requires ligase IV-dependent end joining. *PLoS Genetics*, *7*(4). https://doi.org/10.1371/journal.pgen.1002049

Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14). https://doi.org/10.1093/nar/gkf436

Keith, J. H., Schaeper, C. A., Fraser, T. S., & Fraser, M. J. (2008). Mutational analysis of highly conserved aspartate residues essential to the catalytic core of the piggyBac transposase. *BMC Molecular Biology*, *9*. https://doi.org/10.1186/1471-2199-9-73

Kelley, L. a, Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). Europe PMC Funders Group The Phyre2 web portal for protein modelling , prediction and analysis. *Nature Protocols*, *10*(6).

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, & Haussler D. (2002). *UCSC Browser*. The Human Genome Browser at UCSC.

Kiianitsa, K., & Maizels, N. (2013). A rapid and sensitive assay for DNA-protein covalent complexes in living cells. *Nucleic Acids Research*, *41*(9). https://doi.org/10.1093/nar/gkt171

Kim, Y. J., Nomakuchi, T., Papaleonidopoulou, F., Yang, L., Zhang, Q., & Krainer, A. R. (2022). Gene-specific nonsense-mediated mRNA decay targeting for cystic fibrosis therapy. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-30668-y

King, I. F., Yandava, C. N., Mabb, A. M., Hsiao, J. S., Huang, H. S., Pearson, B. L., Calabrese, J. M., Starmer, J., Parker, J. S., Magnuson, T., Chamberlain, S. J., Philpot, B. D., & Zylka,

M. J. (2013). Topoisomerases facilitate transcription of long genes linked to autism. *Nature*, *501*(7465). https://doi.org/10.1038/nature12504

Kojima, K. K. (2018). Human transposable elements in Repbase: genomic footprints from fish to humans. *Mobile DNA*, *9*(1), 2.

Kolacsek, O., Wachtl, G., Fóthi, Á., Schamberger, A., Sándor, S., Pergel, E., Varga, N., Raskó, T., Izsvák, Z., & Apáti, Á. (2022). Functional indications for transposase domestications–Characterization of the human piggyBac transposase derived (PGBD) activities. *Gene*, *834*, 146609.

Konopka, A., & Atkin, J. D. (2022). The Role of DNA Damage in Neural Plasticity in Physiology and Neurodegeneration. In *Frontiers in Cellular Neuroscience* (Vol. 16). https://doi.org/10.3389/fncel.2022.836885

Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T. M., & Levens, D. (2013). Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nature Structural and Molecular Biology*, *20*(3). https://doi.org/10.1038/nsmb.2517

Kresoja-Rakic, J., & Santoro, R. (2019). Nucleolus and rRNA Gene Chromatin in Early Embryo Development. In *Trends in Genetics* (Vol. 35, Issue 11). https://doi.org/10.1016/j.tig.2019.06.005

Kritis, A. A., Stamoula, E. G., Paniskaki, K. A., & Vavilis, T. D. (2015). Researching glutamate – induced cytotoxicity in different cell lines: A comparative/collective analysis/study. *Frontiers in Cellular Neuroscience*, *9*. https://doi.org/10.3389/fncel.2015.00091

Kunze, R., & Weil, C. F. (2007). The hAT and CACTA superfamilies of plant transposons. *Mobile DNA Ii*, 565–610.

Lam, C. wan, Yeung, W. lan, & Law, C. yiu. (2017). Global developmental delay and intellectual disability associated with a de novo TOP2B mutation. *Clinica Chimica Acta*, *469*. https://doi.org/10.1016/j.cca.2017.03.022

Lans, H., Hoeijmakers, J. H. J., Vermeulen, W., & Marteijn, J. A. (2019). The DNA damage response to transcription stress. In *Nature Reviews Molecular Cell Biology* (Vol. 20, Issue 12). https://doi.org/10.1038/s41580-019-0169-4

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, *44*(W1), W242–W245. https://doi.org/10.1093/nar/gkw290

Li, M. A., Pettitt, S. J., Eckert, S., Ning, Z., Rice, S., Cadiñanos, J., Yusa, K., Conte, N., & Bradley, A. (2013). The piggyBac transposon displays local and distant reintegration preferences and can cause mutations at noncanonical integration sites. *Molecular and Cellular Biology*, *33*(7), 1317–1330.

Liang, T., Guo, L., & Liu, C. (2012). Genome-wide analysis of mir-548 gene family reveals evolutionary and functional implications. *Journal of Biomedicine and Biotechnology*, *2012*.

Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7). https://doi.org/10.1093/bioinformatics/btt656

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., & Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, *27*(12). https://doi.org/10.1093/bioinformatics/btr260

Liu, L. F., & Wang, J. C. (1987). Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(20). https://doi.org/10.1073/pnas.84.20.7024

Love, M. I., Huber, W., & Anders, S. (2017). Analyzing RNA-seq data with DESeq2. *Bioconductor*.

Lun, A. T. L., McCarthy, D. J., & Marioni, J. C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research*, *5*. https://doi.org/10.12688/f1000research.9501.2

Luo, W., Hickman, A. B., Genzor, P., Ghirlando, R., Furman, C. M., Menshikh, A., Haase, A., Dyda, F., & Wilson, M. H. (2022). Transposase N-terminal phosphorylation and asymmetric transposon ends inhibit piggyBac transposition in mammalian cells. *Nucleic Acids Research*, *50*(22). https://doi.org/10.1093/nar/gkac1191

Lynch, B. J., Guinee, D. G., & Holden, J. A. (1997). Human DNA topoisomerase II-alpha: A new marker of cell proliferation in invasive breast cancer. *Human Pathology*, *28*(10). https://doi.org/10.1016/S0046-8177(97)90256-2

Lyu, Y. L., & Wang, J. C. (2003). Aberrant lamination in the cerebral cortex of mouse embryos lacking DNA topoisomerase IIβ. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(12). https://doi.org/10.1073/pnas.1232376100

Madabhushi, R. (2018). The roles of DNA topoisomerase IIβ in transcription. *International Journal of Molecular Sciences*, *19*(7), 1917.

Madabhushi, R., Gao, F., Pfenning, A. R., Pan, L., Yamakawa, S., Seo, J., Rueda, R., Phan, T. X., Yamakawa, H., & Pao, P.-C. (2015). Activity-induced DNA breaks govern the expression of neuronal early-response genes. *Cell*, *161*(7), 1592–1605.

Mahringer, D., Zmarz, P., Okuno, H., Bito, H., & Keller, G. B. (2022). Functional correlates of immediate early gene expression in mouse visual cortex. *Peer Community Journal*, *2*. https://doi.org/10.24072/pcjournal.156

Marmignon, A., Bischerour, J., Silve, A., Fojcik, C., Dubois, E., Arnaiz, O., Kapusta, A., Malinsky, S., & Bétermier, M. (2014). Ku-Mediated Coupling of DNA Cleavage and Repair during Programmed Genome Rearrangements in the Ciliate Paramecium tetraurelia. *PLoS Genetics*, *10*(8). https://doi.org/10.1371/journal.pgen.1004552

Mataga, N., Fujishima, S., Condie, B. G., & Hensch, T. K. (2001). Experience-dependent plasticity of mouse visual cortex in the absence of the neuronal activity-dependent Markeregr1/zif268. *Journal of Neuroscience*, *21*(24), 9724–9732.

Matsumoto, K., & Hirose, S. (2004). Visualization of unconstrained negative supercoils of DNA on polytene chromosomes of Drosophila. *Journal of Cell Science*, *117*(17). https://doi.org/10.1242/jcs.01225

Matthews, A. G. W., & Oettinger, M. A. (2009). Regulation of RAG transposition. *V (D) J Recombination*, 16–31.

McLysaght, A., Hokamp, K., & Wolfe, K. H. (2002). Extensive genomic duplication during early chordate evolution. *Nature Genetics*, *31*(2). https://doi.org/10.1038/ng884

Meczes, E. L., Gilroy, K. L., West, K. L., & Austin, C. A. (2008). The impact of the human DNA topoisomerase II C-terminal domain on activity. *PloS One*, *3*(3), e1754.

Michener, C. D., & Sokal, R. R. (1957). A quantitative approach to a problem in classification. *Evolution*, *11*(2), 130–162.

Mitra, R., Fain-Thornton, J., & Craig, N. L. (2008). piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO Journal*, *27*(7). https://doi.org/10.1038/emboj.2008.41

Miyata, S., Wang, L. Y., Yoshida, C., & Kitanaka, S. (2006). Inhibition of cellular proliferation by diterpenes, topoisomerase II inhibitor. *Bioorganic and Medicinal Chemistry*, *14*(6). https://doi.org/10.1016/j.bmc.2005.10.059

Montaño, S. P., & Rice, P. A. (2011). Moving DNA around: DNA transposition and retroviral integration. *Current Opinion in Structural Biology*, *21*(3), 370–378.

Morellet, N., Li, X., Wieninger, S. A., Taylor, J. L., Bischerour, J., Moriau, S., Lescop, E., Bardiaux, B., Mathy, N., Assrir, N., Bétermier, M., Nilges, M., Hickman, A. B., Dyda, F., Craig, N. L., & Guittet, E. (2018). Sequence-specific DNA binding activity of the cross-brace zinc finger motif of the piggyBac transposase. *Nucleic Acids Research*, *46*(5). https://doi.org/10.1093/nar/gky044

Moss, J., Tinline-Purvis, H., Walker, C. A., Folkes, L. K., Stratford, M. R., Hayles, J., Hoe, K.-L., Kim, D.-U., Park, H.-O., & Kearsey, S. E. (2010). Break-induced ATR and Ddb1–Cul4Cdt2 ubiquitin ligase-dependent nucleotide synthesis promotes homologous recombination repair in fission yeast. *Genes & Development*, *24*(23), 2705–2716.

Mutalip, S. S. M., Yunos, N. M., Abdul-Rahman, P. S. A., Jauri, M. H., Osman, A., & Adenan, M. I. (2014). Mechanisms of action of 17βH-neriifolin on its anticancer effect in SKOV-3 ovarian cancer cell line. *Anticancer Research*, *34*(8).

Nambiar, T. S., Billon, P., Diedenhofen, G., Hayward, S. B., Taglialatela, A., Cai, K., Huang, J. W., Leuzzi, G., Cuella-Martin, R., Palacios, A., Gupta, A., Egli, D., & Ciccia, A. (2019). Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11105-z

Nehme, R., Pietiläinen, O., Artomov, M., Tegtmeyer, M., Valakh, V., Lehtonen, L., Bell, C., Singh, T., Trehan, A., Sherwood, J., Manning, D., Peirent, E., Malik, R., Guss, E. J., Hawes, D., Beccard, A., Bara, A. M., Hazelbaker, D. Z., Zuccaro, E., … Eggan, K. (2022). The 22q11.2 region regulates presynaptic gene-products linked to schizophrenia. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-31436-8

Nesmelova, I. V., & Hackett, P. B. (2010). DDE transposases: Structural similarity and diversity. In *Advanced Drug Delivery Reviews* (Vol. 62, Issue 12). https://doi.org/10.1016/j.addr.2010.06.006

Nevin, L. M., Xiao, T., Staub, W., & Baier, H. (2011). Topoisomerase IIβ is required for lamina-specific targeting of retinal ganglion cell axons and dendrites. *Development*, *138*(12), 2457–2465.

Newman, J. C., Bailey, A. D., Fan, H.-Y., Pavelitz, T., & Weiner, A. M. (2008). An abundant evolutionarily conserved CSB-PiggyBac fusion protein expressed in Cockayne syndrome. *PLoS Genetics*, *4*(3), e1000031.

Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., German Grehov, O. G., Kandrov, D., Rasputin, K., Syabro, M., & Tleukenov, T. (2012). Unipro UGENE: A unified bioinformatics toolkit. In *Bioinformatics* (Vol. 28, Issue 8). https://doi.org/10.1093/bioinformatics/bts091

Panier, S., & Boulton, S. J. (2014). Double-strand break repair: 53BP1 comes into focus. In *Nature Reviews Molecular Cell Biology* (Vol. 15, Issue 1). https://doi.org/10.1038/nrm3719

Pavelitz, T., Gray, L. T., Padilla, S. L., Bailey, A. D., & Weiner, A. M. (2013). PGBD5: A neural-specific intron-containing piggyBac transposase domesticated over 500 million years ago and conserved from cephalochordates to humans. *Mobile DNA*, *4*(1). https://doi.org/10.1186/1759-8753-4-23

Peng, H., Ivanov, A. V., Oh, H. J., Lau, Y. F. C., & Rauscher, F. J. (2009). Epigenetic gene silencing by the SRY protein is mediated by a KRAB-O protein that recruits the KAP1 co-repressor machinery. *Journal of Biological Chemistry*, *284*(51). https://doi.org/10.1074/jbc.M109.032086

Pérez-Sen, R., Queipo, M. J., Gil-Redondo, J. C., Ortega, F., Gómez-Villafuertes, R., Miras-Portugal, M. T., & Delicado, E. G. (2019). Dual-specificity phosphatase regulation in neurons and glial cells. *International Journal of Molecular Sciences*, *20*(8), 1999.

Philippon, H., Souvane, A., Brochier-Armanet, C., & Perrière, G. (2017). IsoSel: Protein isoform selector for phylogenetic reconstructions. *PLoS ONE*, *12*(3). https://doi.org/10.1371/journal.pone.0174250

Piriyapongsa, J., Bootchai, C., Ngamphiw, C., & Tongsima, S. (2012). MicroPIR: An integrated database of microRNA target sites within human promoter sequences. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0033888

Piriyapongsa, J., & Jordan, I. K. (2007). A family of human microRNA genes from miniature inverted-repeat transposable elements. *PloS One*, *2*(2), e203.

Pommier, Y., Nussenzweig, A., Takeda, S., & Austin, C. (2022). Human topoisomerases and their roles in genome stability and organization. In *Nature Reviews Molecular Cell Biology* (Vol. 23, Issue 6). https://doi.org/10.1038/s41580-022-00452-3

Pritham, E. J., Feschotte, C., & Wessler, S. R. (2005). Unexpected diversity and differential success of DNA transposons in four species of Etamoeba protozoans. *Molecular Biology and Evolution*, *22*(9). https://doi.org/10.1093/molbev/msi169

Pritham, E. J., Putliwala, T., & Feschotte, C. (2007). Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene*, *390*(1–2). https://doi.org/10.1016/j.gene.2006.08.008

Raskó, T., Pande, A., Radscheit, K., Zink, A., Singh, M., Sommer, C., Wachtl, G., Kolacsek, O., Inak, G., Szvetnik, A., Petrakis, S., Bunse, M., Bansal, V., Selbach, M., Orbán, T. I., Prigione, A., Hurst, L. D., & Izsvák, Z. (2022). A Novel Gene Controls a New Structure:

PiggyBac Transposable Element-Derived 1, Unique to Mammals, Controls Mammal-Specific Neuronal Paraspeckles. *Molecular Biology and Evolution*, *39*(10). https://doi.org/10.1093/molbev/msac175

Riccio, A. A., Schellenberg, M. J., & Williams, R. S. (2020). Molecular mechanisms of topoisomerase 2 DNA–protein crosslink resolution. In *Cellular and Molecular Life Sciences* (Vol. 77, Issue 1). https://doi.org/10.1007/s00018-019-03367-z

Rifka, V., Jaco, V. D. T., & Cees, D. (2015). Counterintuitive DNA sequence dependence in supercoiling-induced DNA melting. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0141576

Ronquist, F. R., & Huelsenbeck, J. P. (2003). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, *19*(12).

Rossi, J. J., Rosenfeld, J. A., Chan, K. M., Streff, H., Nankivell, V., Peet, D. J., Whitelaw, M. L., & Bersten, D. C. (2021). Molecular characterisation of rare loss-of-function NPAS3 and NPAS4 variants identified in individuals with neurodevelopmental disorders. *Scientific Reports*, *11*(1), 6602. https://doi.org/10.1038/s41598-021-86041-4

Salceda, J., Fernández, X., & Roca, J. (2006). Topoisomerase II, not topoisomerase I, is the proficient relaxase of nucleosomal DNA. *EMBO Journal*, *25*(11). https://doi.org/10.1038/sj.emboj.7601142

Sarkar, A., Sim, C., Hong, Y. S., Hogan, J. R., Fraser, M. J., Robertson, H. M., & Collins, F. H. (2003). Molecular evolutionary analysis of the widespread piggyBac transposon family and related" domesticated" sequences. *Molecular Genetics and Genomics*, *270*, 173–180.

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*, *33*(5), 495–502.

Shi, X., Singh, S., Wu, P., & Li, H. (2023). O42: The landscape of chimeric RNAs in healthy and preeclampsia patient placenta and blood. *Genetics in Medicine Open*, *1*(1).

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19). https://doi.org/10.1093/bioinformatics/btv351

Simi, A., Ansaloni, F., Damiani, D., Codino, A., Mangoni, D., Lau, P., Vozzi, D., Pandolfini, L., Sanges, R., & Gustincich, S. (2023). The Pgbd5 DNA transposase is required for mouse cerebral cortex development through DNA double-strand breaks formation. *BioRxiv*, 2023.05.09.539730. https://doi.org/10.1101/2023.05.09.539730

Stott, R. T., Kritsky, O., & Tsai, L. H. (2021). Profiling DNA break sites and transcriptional changes in response to contextual fear learning. *PLoS ONE*, *16*(7 July). https://doi.org/10.1371/journal.pone.0249691

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43). https://doi.org/10.1073/pnas.0506580102

Sun, X., & Lin, Y. (2016). Npas4: Linking Neuronal Activity to Memory. *Trends in Neurosciences*, *39*(4), 264–275. https://doi.org/https://doi.org/10.1016/j.tins.2016.02.003

Sun, Z., Xu, X., He, J., Murray, A., Sun, M., Wei, X., Wang, X., McCoig, E., Xie, E., Jiang, X., Li, L., Zhu, J., Chen, J., Morozov, A., Pickrell, A. M., Theus, M. H., & Xie, H. (2019). EGR1 recruits TET1 to shape the brain methylome during development and upon neuronal activity. *Nature Communications*, *10*(1), 3892. https://doi.org/10.1038/s41467-019-11905-3

Tanaka, K. Z., Pevzner, A., Hamidi, A. B., Nakazawa, Y., Graham, J., & Wiltgen, B. J. (2014). Cortical Representations Are Reinstated by the Hippocampus during Memory Retrieval. *Neuron*, *84*(2). https://doi.org/10.1016/j.neuron.2014.09.037

Tarazona, S., Furió-Tarí, P., Turrà, D., Di Pietro, A., Nueda, M. J., Ferrer, A., & Conesa, A. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Research*, *43*(21). https://doi.org/10.1093/nar/gkv711

Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., Levi, B., Gray, L. T., Sorensen, S. A., Dolbeare, T., Bertagnolli, D., Goldy, J., Shapovalova, N., Parry, S., Lee, C., Smith, K., Bernard, A., Madisen, L., Sunkin, S. M., … Zeng, H. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience*, *19*(2). https://doi.org/10.1038/nn.4216

Teng, Y. C., Sundaresan, A., O'Hara, R., Gant, V. U., Li, M., Martire, S., Warshaw, J. N., Basu, A., & Banaszynski, L. A. (2021). ATRX promotes heterochromatin formation to protect cells from G-quadruplex DNA-mediated stress. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-24206-5

Tiwari, V. K., Burger, L., Nikoletopoulou, V., Deogracias, R., Thakurela, S., Wirbelauer, C., Kaut, J., Terranova, R., Hoerner, L., Mielke, C., Boege, F., Murr, R., Peters, A. H. F. M., Barde, Y. A., & Schübeler, D. (2012). Target genes of Topoisomerase IIβ regulate neuronal survival and are defined by their chromatin state. *Proceedings of the National Academy of*

Sciences of the United States of America, *109*(16). https://doi.org/10.1073/pnas.1119798109

Tsutsui, K. M., Sano, K., Hosoya, O., & Tsutsui, K. (2006). Expression dynamics and functional implications of DNA topoisomerase IIβ in the brain. In *Anatomical Science International* (Vol. 81, Issue 3). https://doi.org/10.1111/j.1447-073X.2006.00146.x

Tsutsui, K., Tsutsui, K., Hosoya, O., Sano, K., & Tokunaga, A. (2001). Immunohistochemical analyses of DNA topoisomerase II isoforms in developing rat cerebellum. *Journal of Comparative Neurology*, *431*(2). https://doi.org/10.1002/1096-9861(20010305)431:2<228::AID-CNE1067>3.0.CO;2-M

Tycko, J., DelRosso, N., Hess, G. T., Aradhana, Banerjee, A., Mukund, A., Van, M. V., Ego, B. K., Yao, D., Spees, K., Suzuki, P., Marinov, G. K., Kundaje, A., Bassik, M. C., & Bintu, L. (2020). High-Throughput Discovery and Characterization of Human Transcriptional Effectors. *Cell*, *183*(7). https://doi.org/10.1016/j.cell.2020.11.024

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., & Asplund, A. (2015). Tissue-based map of the human proteome. *Science*, *347*(6220), 1260419.

Unoki, T., Matsuda, S., Kakegawa, W., Van, N. T. B., Kohda, K., Suzuki, A., Funakoshi, Y., Hasegawa, H., Yuzaki, M., & Kanaho, Y. (2012). NMDA Receptor-Mediated PIP5K Activation to Produce PI(4,5)P 2 Is Essential for AMPA Receptor Endocytosis during LTD. *Neuron*, *73*(1). https://doi.org/10.1016/j.neuron.2011.09.034

Uusküla-Reimand, L., Hou, H., Samavarchi-Tehrani, P., Rudan, M. V., Liang, M., Medina-Rivera, A., Mohammed, H., Schmidt, D., Schwalie, P., Young, E. J., Reimand, J., Hadjur, S., Gingras, A. C., & Wilson, M. D. (2016). Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biology*, *17*(1). https://doi.org/10.1186/s13059-016-1043-8

Uusküla-Reimand, L., & Wilson, M. D. (2022). Untangling the roles of TOP2A and TOP2B in transcription and cancer. In *Science Advances* (Vol. 8, Issue 44). https://doi.org/10.1126/sciadv.add4920

Vasudevan, L., Jeromin, A., Volpicelli-Daley, L., De Camilli, P., Holowka, D., & Baird, B. (2009). The β- and γ-isoforms of type I PIP5K regulate distinct stages of Ca2+ signaling in mast cells. *Journal of Cell Science*, *122*(14), 2567–2574. https://doi.org/10.1242/jcs.048124

Wachtl, G., Schád, É., Huszár, K., Palazzo, A., Ivics, Z., Tantos, Á., & Orbán, T. I. (2022). Functional characterization of the N-Terminal disordered region of the piggyBac transposase. *International Journal of Molecular Sciences*, *23*(18), 10317.

Wang, L., Wang, S., & Li, W. (2012). RSeQC: Quality control of RNA-seq experiments. *Bioinformatics*, *28*(16). https://doi.org/10.1093/bioinformatics/bts356

Weber Boutros, S., Unni, V. K., & Raber, J. (2022). An Adaptive Role for DNA Double-Strand Breaks in Hippocampus-Dependent Learning and Memory. In *International Journal of Molecular Sciences* (Vol. 23, Issue 15). https://doi.org/10.3390/ijms23158352

Weisman, C. M., Murray, A. W., & Eddy, S. R. (2020). Many, but not all, lineage-specific genes can be explained by homology detection failure. *PLoS Biology*, *18*(11). https://doi.org/10.1371/journal.pbio.3000862

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wilson, M. H., Coates, C. J., & George, A. L. (2007). PiggyBac transposon-mediated gene transfer in human cells. *Molecular Therapy*, *15*(1), 139–145.

Woessner, R. D., Mattern, M. R., Mirabelli, C. K., Johnson, R. K., & Drake, F. H. (1991). Proliferation- and cell cycle-dependent differences in expression of the 170 kilodalton and 180 kilodalton forms of topoisomerase II in NIH-3T3 cells. *Cell Growth & Differentiation : The Molecular Biology Journal of the American Association for Cancer Research*, *2*(4).

Xie, W., Zeng, Y., Hu, L., Hao, J., Chen, Y., Yun, X., Lin, Q., & Li, H. (2022). Based on different immune responses under the glucose metabolizing type of papillary thyroid cancer and the response to anti-PD-1 therapy. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/fimmu.2022.991656

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8). https://doi.org/10.1093/molbev/msm088

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., … Flicek, P. (2020). Ensembl 2020. *Nucleic Acids Research*, *48*(D1). https://doi.org/10.1093/nar/gkz966

Yuan, Y.-W., & Wessler, S. R. (2011). The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proceedings of the National Academy of Sciences*, *108*(19), 7884–7889.

Yue, W. H., Wang, H. F., Sun, L. D., Tang, F. L., Liu, Z. H., Zhang, H. X., Li, W. Q., Zhang, Y. L., Zhang, Y., Ma, C. C., Du, B., Wang, L. F., Ren, Y. Q., Yang, Y. F., Hu, X. F., Wang, Y., Deng, W., Tan, L. W., Tan, Y. L., … Zhang, D. (2011). Genome-wide association study

identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nature Genetics*, *43*(12). https://doi.org/10.1038/ng.979

Zaim, M., & Isik, S. (2018). DNA topoisomerase IIβ stimulates neurite outgrowth in neural differentiated human mesenchymal stem cells through regulation of Rho-GTPases (RhoA/Rock2 pathway) and Nurr1 expression. *Stem Cell Research & Therapy*, *9*(1). https://doi.org/10.1186/s13287-018-0859-4

Zhang, F., Mu, G., Liu, Z., Xie, Q., Zhang, H., Zhou, S., Wang, Z., Hu, K., Wang, Z., Zhao, X., Cui, Y., & Xiang, Q. (2022). Genetic Polymorphisms Associated with Prothrombin Time and Activated Partial Thromboplastin Time in Chinese Healthy Population. *Genes*, *13*(10). https://doi.org/10.3390/genes13101867

Zhang, J., Zhang, D., McQuade, J. S., Behbehani, M., Tsien, J. Z., & Xu, M. (2002). c-fos regulates neuronal excitability and survival. *Nature Genetics*, *30*(4), 416–420. https://doi.org/10.1038/ng859

Zhu, Q., High, F. A., Zhang, C., Cerveira, E., Russell, M. K., Longoni, M., Joy, M. P., Ryan, M., Mil-homens, A., Bellfy, L., Coletti, C. M., Bhayani, P., Hila, R., Wilson, J. M., Donahoe, P. K., & Lee, C. (2018). Systematic analysis of copy number variation associated with congenital diaphragmatic hernia. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(20). https://doi.org/10.1073/pnas.1714885115

Zyla, J., Marczyk, M., Kaufmann, S. H. E., Polanska, J., & Weiner 3rd, J. (2019). Gene set enrichment for reproducible science: comparison of CERNO and eight other algorithms. *Bioinformatics*, *35*(24), 5146 – 5154. https://doi.org/10.1093/bioinformatics/btz447

Appendix

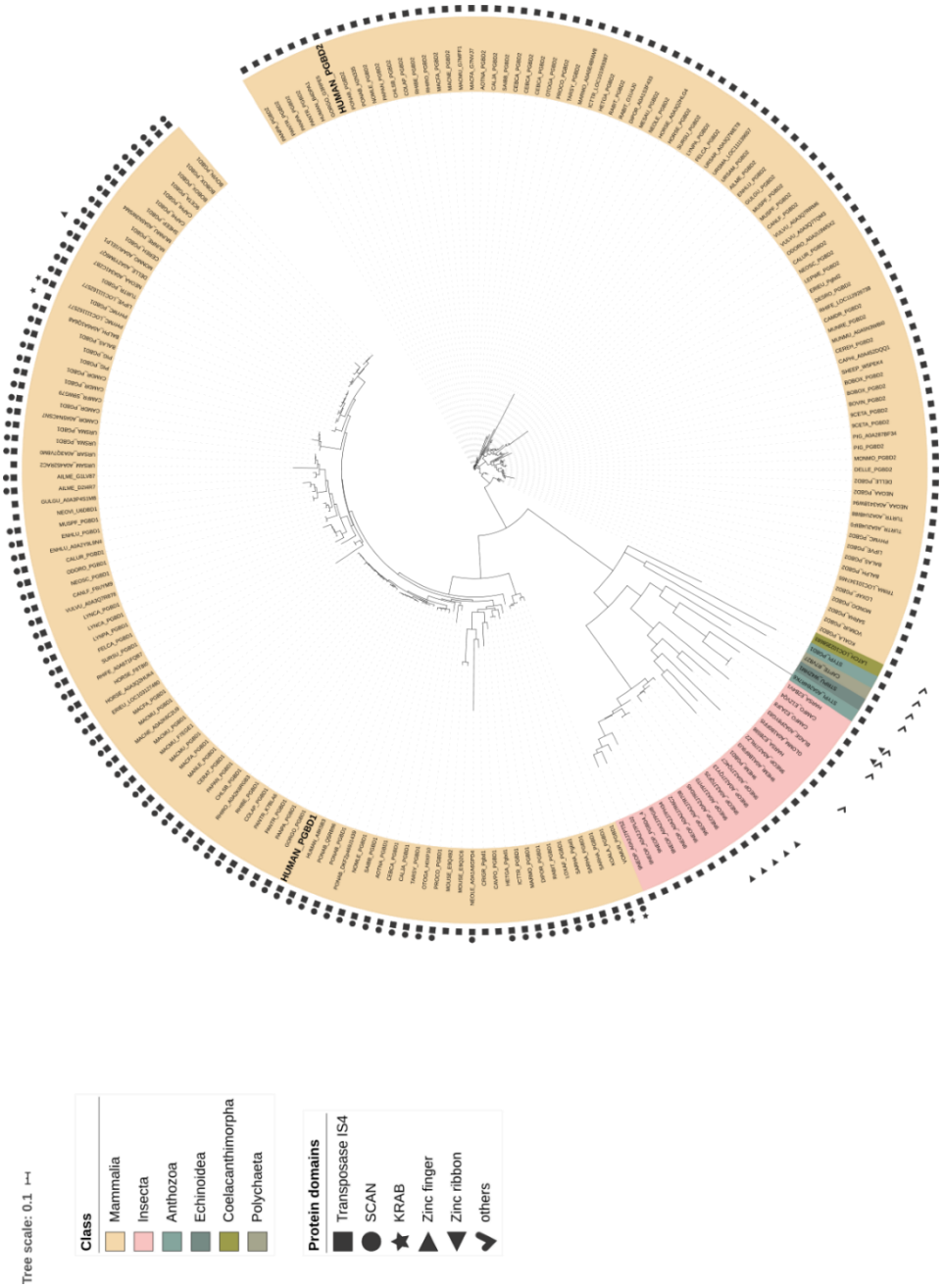


Figure S 1: Phylogenetic tree of PGBD1 and PGBD2 showing the presence of protein domains. Human PGBD1 and PGBD2, along with closely related sequences containing the transposase IS4, were aligned using the MUSCLE alignment algorithm. The tree was constructed using *MrBayes*. Protein domains were annotated using *hmmerscan*.

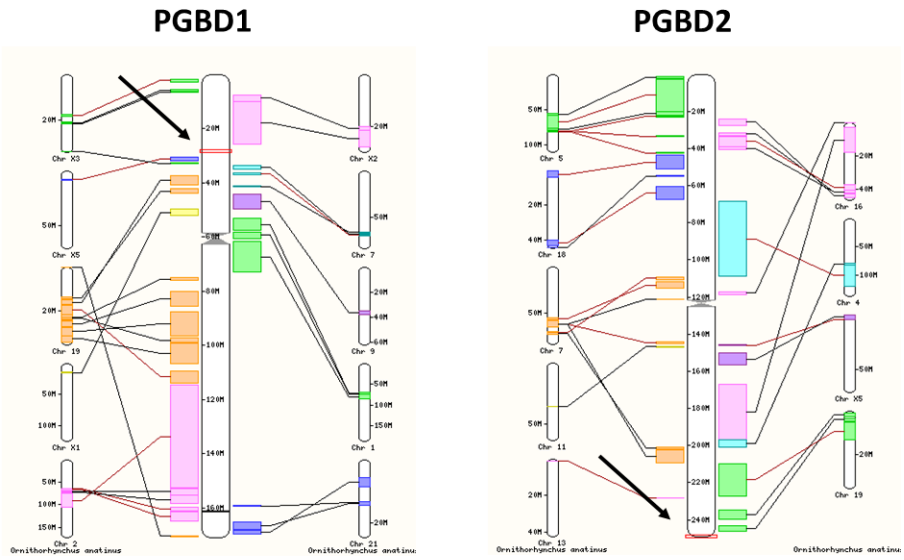


Figure S 2: Syntenic regions for human PGBD1 & PGBD2 are missing in monotremes. Black arrows and red boxes indicate the genomic locations of PGBD1 and PGBD2 within the relevant human chromosomes. The drawings were generated in the Ensembl synteny browser (2020).

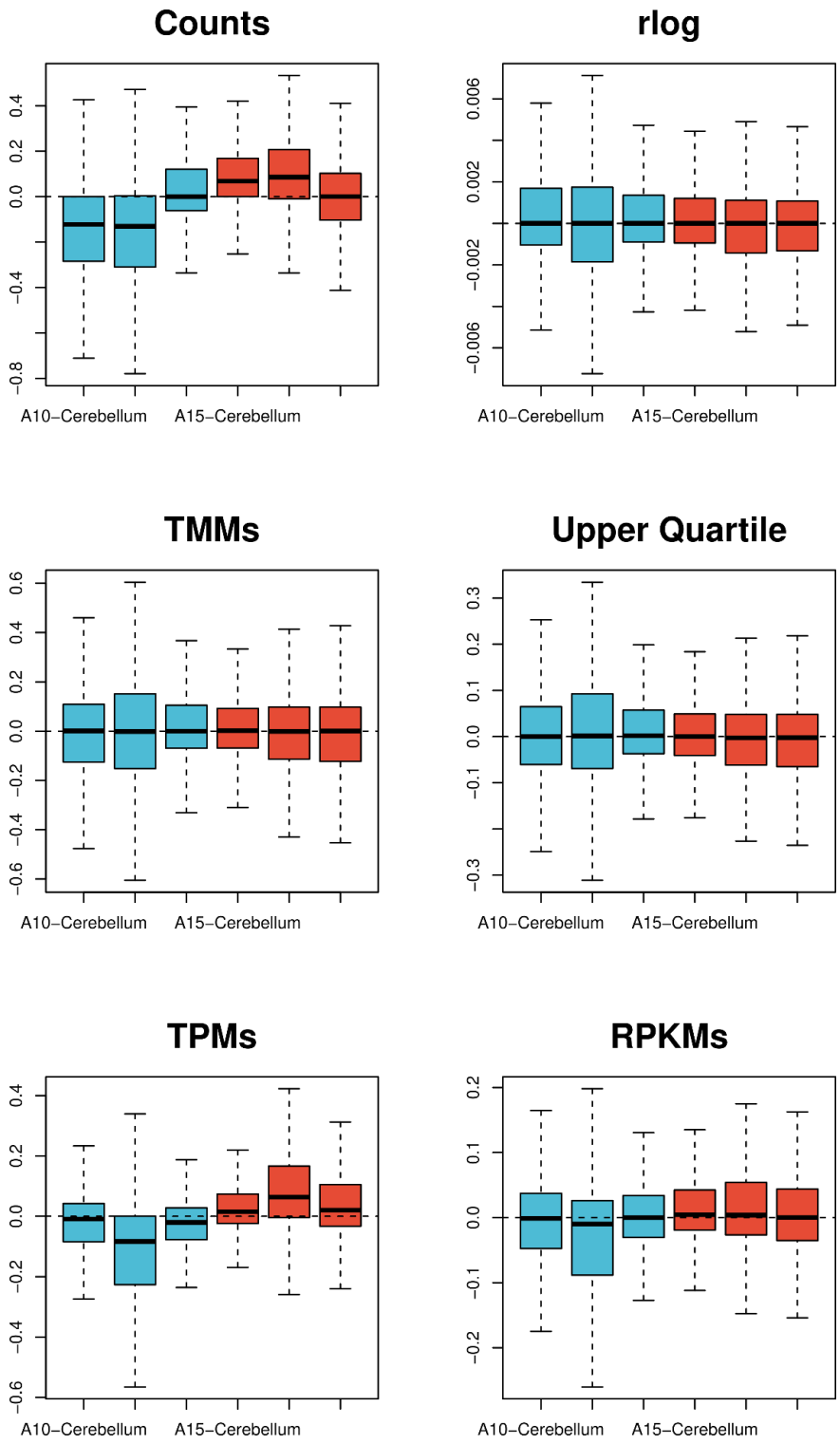


Figure S 3: Investigating normalization methods for RNA-seq data of cerebella in RLE plots. Blue boxes indicate wild type samples and red boxes indicate knockout samples.

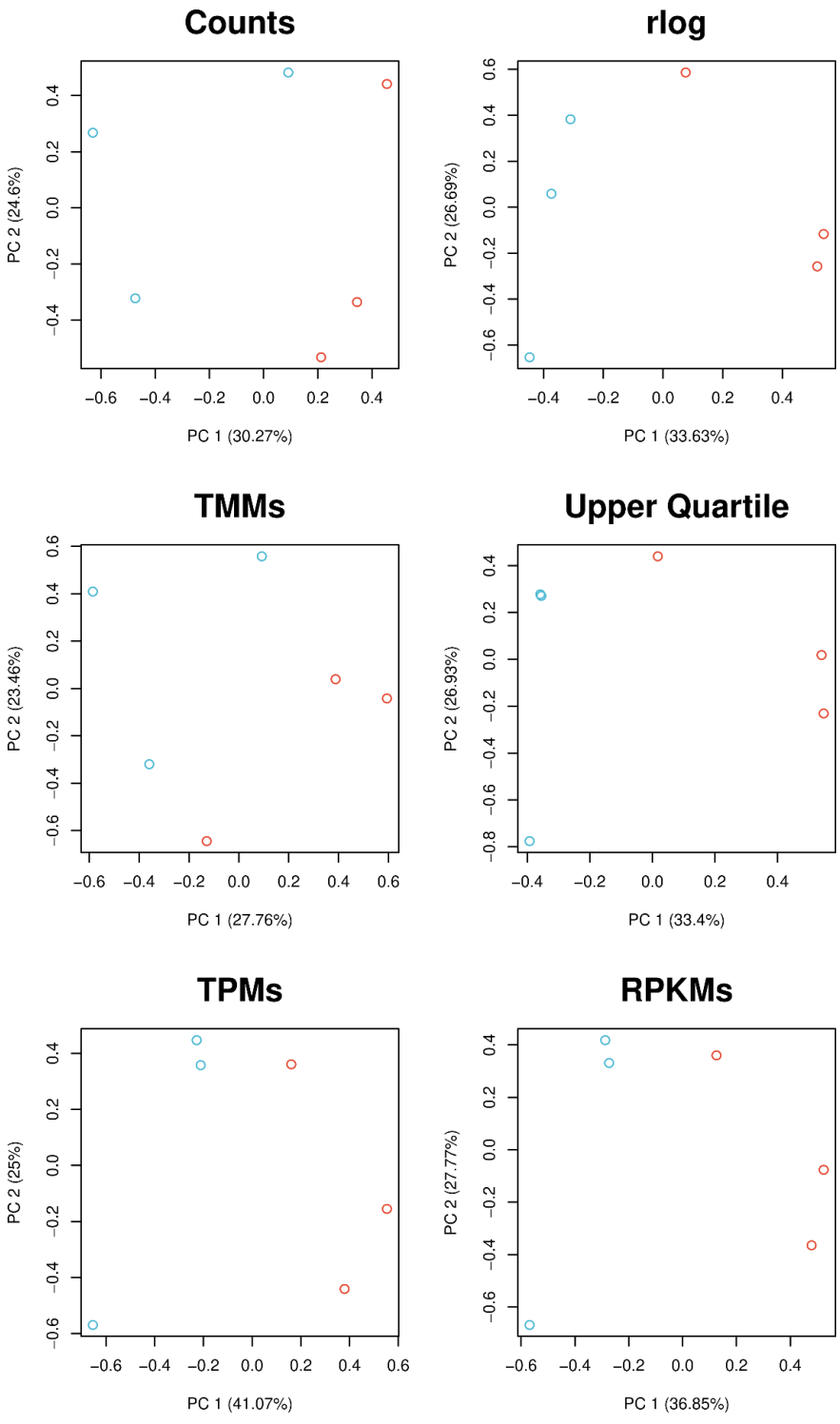


Figure S 4: Investigating normalization methods for RNA-seq data of cerebella in PCA plots. Blue circles indicate wild type samples and red circles indicate knockout samples.

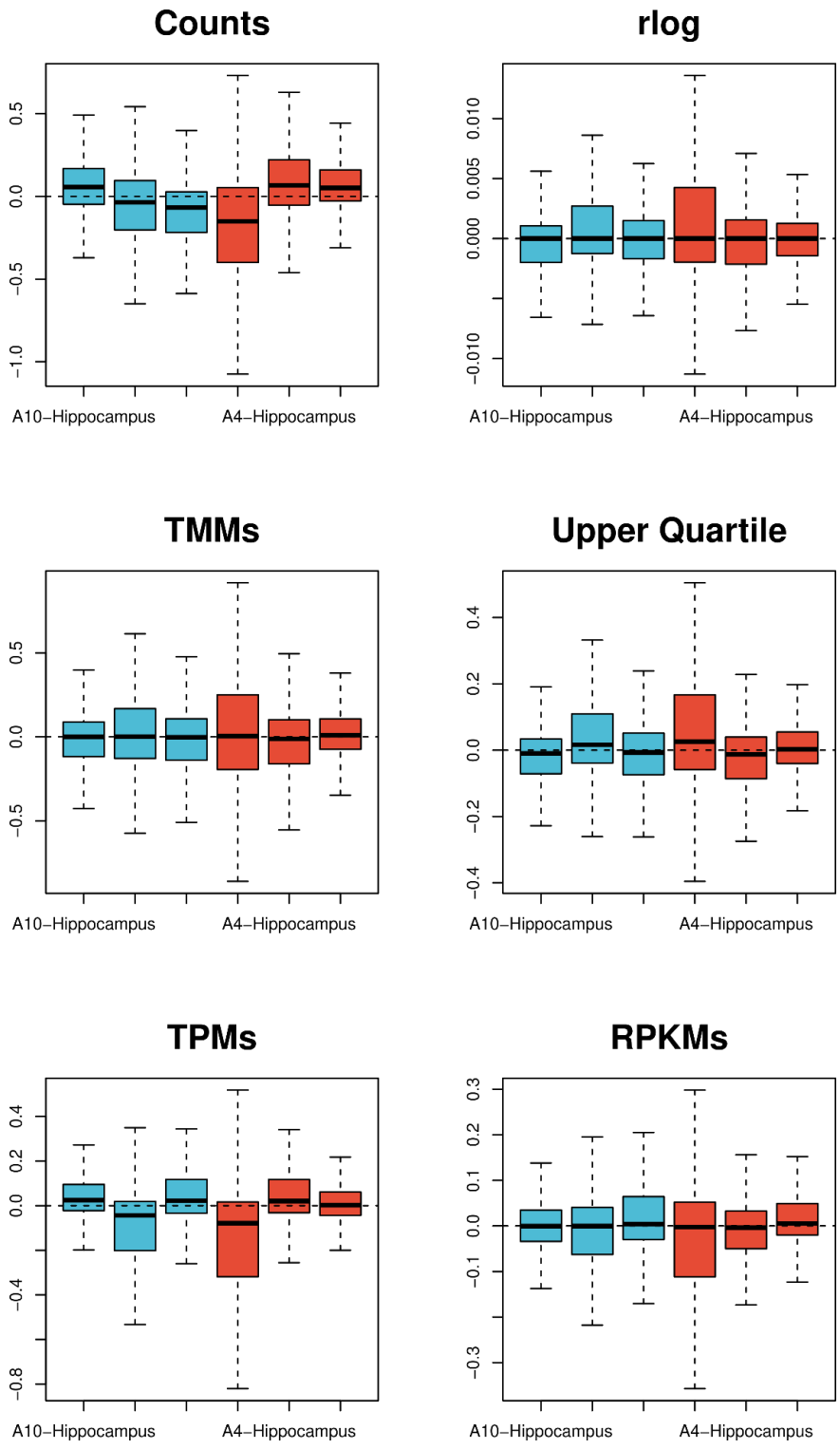


Figure S 5: Investigating normalization methods for RNA-seq data of hippocampi in RLE plots. Blue boxes indicate wild type samples and red boxes indicate knockout samples.

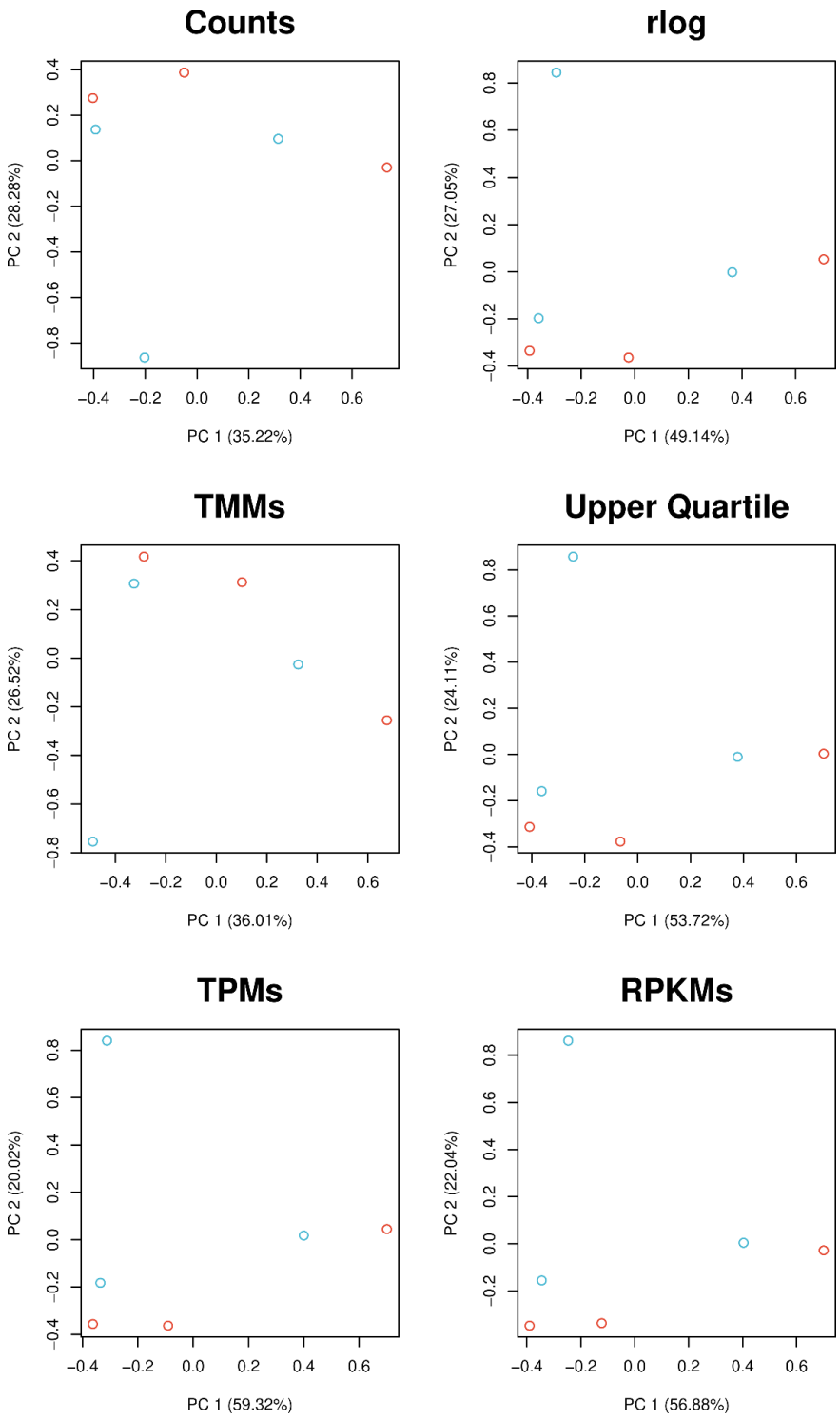


Figure S 6: Investigating normalization methods for RNA-seq data of hippocampi in PCA plots. Blue circles indicate wild type samples and red circles indicate knockout samples.

Table S 1: Gene set enrichment analysis of differentially expressed genes in cerebella of Pgbd5 knockout mice. DEGs were sorted by FDR and tested against KEGG and GO Biological Process databases. The Area Under Curve (AUC) represents the effect size of enrichment, indicating the strength of association between DEGs and the gene sets in the databases. N1 indicates the detected number of DEG belonging to the pathway.

ID	Title	N1	AUC	adj.P.Val
MM5801	Gobp proton motive force driven atp synthesis	65	0.7836	1.57E-17
MM7513	Gobp atp synthesis coupled electron transport	65	0.7627	1.98E-16
MM5979	Gobp aerobic electron transport chain	54	0.7721	8.69E-15
MM4606	Gobp mitochondrial electron transport nadh to ubiquinone	29	0.7691	4.16E-08
MM5802	Gobp energy coupled proton transmembrane transport against electrochemical gradient	5	0.8248	4.31E-06
MM4609	Gobp mitochondrial electron transport cytochrome c to oxygen	9	0.8323	3.06E-04
MM10377	Gobp retrograde trans synaptic signaling	13	0.8058	5.30E-04
MM4608	Gobp mitochondrial electron transport ubiquinol to cytochrome c	12	0.7682	7.68E-04
MM8799	Gobp negative regulation of neurotransmitter transport	14	0.7560	1.64E-03
MM4188	Gobp protein localization to paranode region of axon	5	0.9510	2.24E-03
MM10391	Gobp anterograde dendritic transport of neurotransmitter receptor complex	5	0.8517	2.89E-03
MM5729	Gobp oxygen transport	5	0.8592	2.94E-03
MM11488	Gobp negative regulation of synaptic vesicle exocytosis	5	0.8329	3.66E-03

MM6223	Gobp growth hormone secretion	14	0.7997	8.76E-03
MM10419	Gobp vesicle fusion to plasma membrane	21	0.7528	9.43E-03
MM10441	Gobp protein localization to axon	11	0.7504	1.03E-02
MM11179	Gobp regulation of neuromuscular junction development	7	0.7666	1.20E-02
MM10418	Gobp trans synaptic signaling by bdnf	4	0.8830	1.23E-02
MM5834	Gobp synaptic vesicle maturation	12	0.7709	1.33E-02
MM10387	Gobp anterograde axonal transport of mitochondrion	6	0.8043	1.47E-02
MM11341	Gobp regulation of postsynaptic density organization	19	0.7577	1.81E-02
MM10535	Gobp amyloid beta clearance by transcytosis	7	0.7551	1.85E-02
MM10038	Gobp positive regulation of microtubule nucleation	6	0.7849	1.85E-02
MM10978	Gobp regulation of calcium ion dependent exocytosis of neurotransmitter	5	0.7847	1.92E-02
MM11538	Gobp regulation of camp dependent protein kinase activity	13	0.7871	1.92E-02
MM10409	Gobp regulation of modification of postsynaptic structure	13	0.7890	2.58E-02
MM10815	Gobp regulation of dna damage response signal transduction by p53 class mediator resulting in transcription of p21 class mediator	4	0.9188	2.69E-02
MM8741	Gobp positive regulation of cyclic nucleotide phosphodiesterase activity	4	0.9479	2.75E-02

MM6356	Gobp negative regulation of microtubule polymerization	15	0.7652	2.86E-02
MM10826	Gobp positive regulation of intrinsic apoptotic signaling pathway in response to dna damage	10	0.7594	3.11E-02
MM10428	Gobp trans synaptic signaling by lipid	9	0.8242	3.69E-02
MM7134	Gobp stress granule disassembly	5	0.7578	3.71E-02
MM9216	Gobp spongiotrophoblast differentiation	2	0.9493	4.28E-02
MM6314	Gobp paranodal junction assembly	7	0.8290	4.50E-02
MM10328	Gobp neurotransmitter loading into synaptic vesicle	6	0.7985	4.56E-02
MM11360	Gobp retrograde neuronal dense core vesicle transport	6	0.7841	4.56E-02
MM10771	Gobp regulation of atpase coupled calcium transmembrane transporter activity	8	0.7704	4.66E-02
MM8128	Gobp platelet activating factor metabolic process	5	0.8302	4.75E-02
MM12439	Gocc respirasome	83	0.7730	4.49E-24
MM11986	Gocc cytosolic ribosome	100	0.7865	1.76E-20
MM12064	Gocc nadh dehydrogenase complex	45	0.7752	4.29E-13
MM11985	Gocc cytosolic large ribosomal subunit	52	0.8242	1.11E-12
MM12429	Gocc cytochrome complex	32	0.7606	4.98E-12
MM11823	Gocc mitochondrial respiratory chain complex iv	18	0.8602	1.87E-11
MM12387	Gocc respiratory chain complex iv	21	0.8156	5.65E-11
MM11987	Gocc cytosolic small ribosomal subunit	46	0.7764	1.28E-09

MM12384	Gocc proton transporting atp synthase complex	16	0.8615	1.41E-07
MM12308	Gocc polysomal ribosome	33	0.7847	1.37E-06
MM12622	Gocc integral component of presynaptic active zone membrane	23	0.7655	1.52E-06
MM12201	Gocc proton transporting two sector atpase complex proton transporting domain	17	0.8181	2.69E-06
MM12386	Gocc proton transporting atp synthase complex coupling factor f o	10	0.9101	4.29E-06
MM11822	Gocc mitochondrial respiratory chain complex iii	12	0.7690	6.99E-05
MM12349	Gocc calyx of held	26	0.7609	1.32E-04
MM12555	Gocc atpase dependent transmembrane transport complex	8	0.8396	1.59E-04
MM11883	Gocc sodium potassium exchanging atpase complex	6	0.8048	7.48E-04
MM11975	Gocc proteasome core complex alpha subunit complex	7	0.9476	1.24E-03
MM12139	Gocc haptoglobin hemoglobin complex	3	0.8264	2.02E-03
MM11930	Gocc spectrin associated cytoskeleton	5	0.9440	2.49E-03
MM11860	Gocc proteasome core complex	16	0.7850	4.05E-03
MM11940	Gocc prefoldin complex	7	0.8186	6.00E-03
MM12714	Gocc microvesicle	2	0.9263	6.00E-03
MM12351	Gocc dendritic branch	6	0.8807	7.27E-03
MM12420	Gocc mitotic spindle astral microtubule	6	0.8524	1.11E-02
MM12319	Gocc varicosity	8	0.8281	1.15E-02

MM11859	Gocc chaperonin containing t complex	8	0.8319	1.71E-02
MM11941	Gocc eukaryotic translation initiation factor 4f complex	9	0.7539	1.79E-02
MM12159	Gocc dense core granule membrane	9	0.7742	2.17E-02
MM12385	Gocc proton transporting atp synthase complex catalytic core f 1	6	0.7787	2.24E-02
MM12482	Gocc ptw pp1 phosphatase complex	7	0.7789	2.32E-02
MM12618	Gocc anchored component of postsynaptic membrane	10	0.7738	2.32E-02
MM11838	Gocc signal peptidase complex	7	0.7987	2.60E-02
MM11746	Gocc voltage gated sodium channel complex	12	0.7825	2.77E-02
MM12077	Gocc platelet dense tubular network	4	0.9308	3.08E-02
MM12596	Gocc extrinsic component of synaptic vesicle membrane	8	0.8339	3.16E-02
MM11904	Gocc calcineurin complex	6	0.8488	3.16E-02
MM12457	Gocc translation preinitiation complex	5	0.8608	4.54E-02
MM11856	Gocc actomyosin contractile ring	4	0.8887	4.68E-02
MM11909	Gocc spectrin	5	0.7546	4.73E-02
MM13373	Gomf oxidoreduction driven active transmembrane transporter activity	46	0.7652	2.50E-17
MM14106	Gomf nadh dehydrogenase quinone activity	22	0.7686	6.93E-09
MM12847	Gomf nadh dehydrogenase activity	24	0.7595	1.09E-08
MM13469	Gomf oxidoreductase activity acting on a heme group of donors	16	0.8481	6.73E-08
MM13346	Gomf proton channel activity	14	0.8839	4.59E-07

MM14056	Gomf proton transporting atp synthase activity rotational mechanism	14	0.8839	4.59E-07
MM14134	Gomf microtubule plus end binding	19	0.7789	2.73E-04
MM13781	Gomf d1 dopamine receptor binding	11	0.8058	1.03E-03
MM13979	Gomf neurotrophin binding	10	0.8962	1.26E-03
MM13775	Gomf hemoglobin alpha binding	2	0.9982	1.52E-03
MM13713	Gomf spectrin binding	25	0.7754	1.63E-03
MM13266	Gomf p type potassium transmembrane transporter activity	7	0.7697	3.63E-03
MM13104	Gomf oxygen carrier activity	4	0.8428	4.17E-03
MM13774	Gomf haptoglobin binding	3	0.8264	5.06E-03
MM14191	Gomf ubiquitin protein transferase inhibitor activity	9	0.8595	5.63E-03
MM13168	Gomf ubiquinol cytochrome c reductase activity	6	0.7885	7.73E-03
MM14099	Gomf nerve growth factor binding	6	0.8951	1.26E-02
MM13716	Gomf triplet codon amino acid adaptor activity	14	0.8167	1.27E-02
MM14126	Gomf rage receptor binding	4	0.9266	1.78E-02
MM13265	Gomf p type proton exporting transporter activity	4	0.9484	1.88E-02
MM13096	Gomf inorganic phosphate transmembrane transporter activity	5	0.8491	3.04E-02
MM14426	Gomf thioredoxin dependent peroxiredoxin activity	5	0.8983	4.32E-02
MM13839	Gomf apolipoprotein receptor binding	3	0.8610	4.50E-02
MM14236	Gomf large ribosomal subunit rrna binding	8	0.8146	4.81E-02

MM14299	Gomf voltage gated calcium channel activity involved in cardiac muscle cell action potential	5	0.8943	4.96E-02
MM15415	Reactome eukaryotic translation initiation	106	0.7968	1.58E-22
MM15417	Reactome formation of a pool of free 40s subunits	89	0.8387	2.32E-22
MM14698	Reactome srp dependent cotranslational protein targeting to membrane	81	0.8537	2.49E-21
MM15705	Reactome nonsense mediated decay nmd independent of the exon junction complex ejc	83	0.8279	1.11E-20
MM15640	Reactome nonsense mediated decay nmd	103	0.7798	9.47E-20
MM15416	Reactome activation of the mrna upon binding of the cap binding complex and eifs and subsequent binding to 43s	56	0.8255	1.58E-13
MM14626	Reactome formation of atp by chemiosmotic coupling	15	0.8827	2.44E-07
MM15112	Reactome interaction between l1 and ankyrins	8	0.9123	7.83E-04
MM15475	Reactome advanced glycosylation endproduct receptor signaling	4	0.9048	9.00E-04
MM15052	Reactome reduction of cytosolic ca levels	13	0.8489	9.67E-04
MM15142	Reactome ionotropic activity of kainate receptors	12	0.7858	6.77E-03
MM14950	Reactome hsf1 activation	8	0.8417	1.05E-02

MM15167	Reactome pink1 prkn mediated mitophagy	21	0.7518	1.41E-02
MM15004	Reactome association of tric cct with target proteins during biosynthesis	9	0.8032	2.95E-02
MM14588	Reactome synthesis of pips at the er membrane	5	0.8239	3.01E-02

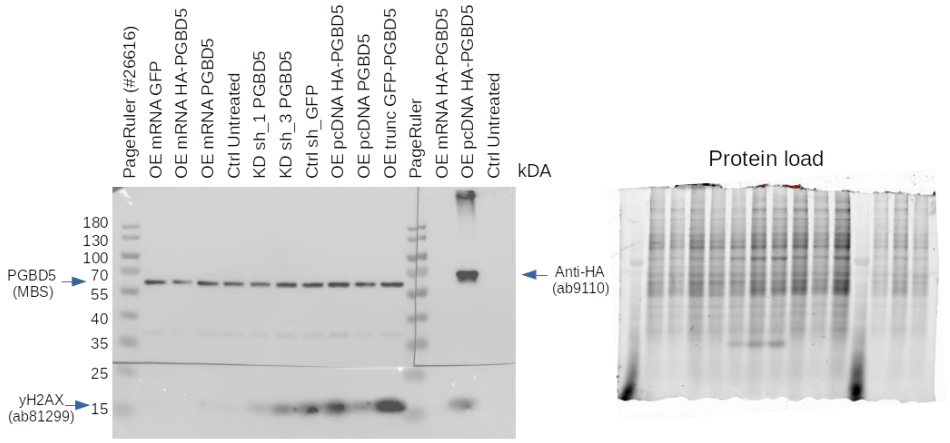


Figure S 7: Western Blot analysis of samples transfected with PGBD5 mRNA. Protein levels were examined using endogenous PGBD5 antibody (left panel) and anti-HA antibody (right panel). Despite varying levels of PGBD5, no enrichment of PGBD5 protein was observed. The protein load was assessed using a stain-free gel. As the antibody was unspecific, the samples probed with anti-HA antibody are of greater interest.

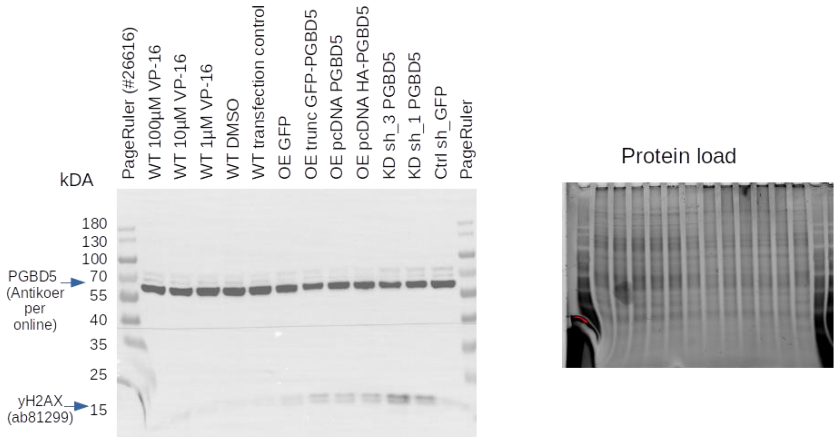


Figure S 8: Western Blot analysis of samples with different levels of PGBD5 expression. Protein levels were assessed using an endogenous PGBD5 antibody. Despite the variations in PGBD5 levels, no significant differences in PGBD5 protein expression were detected. Protein loading was evaluated using a stain-free gel. Similar results were obtained with other endogenous antibodies.

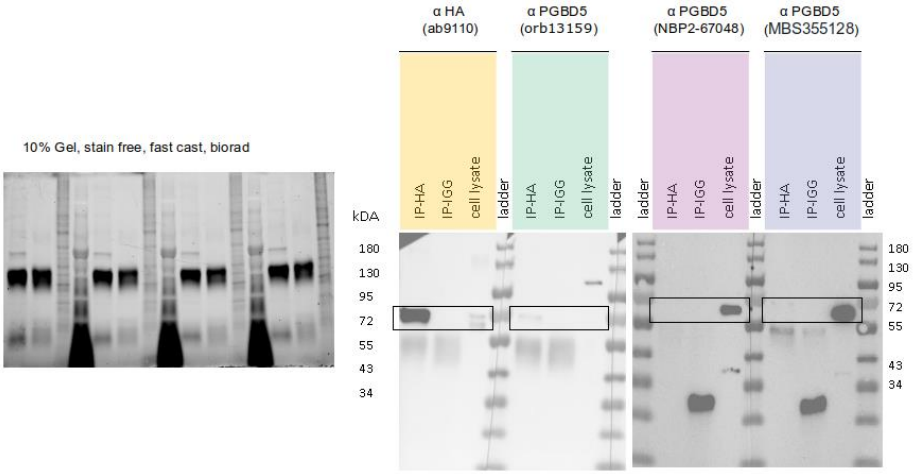


Figure S 9: Endogenous PGBD5 antibodies fail to detect HA-tagged PGBD5 in co-IP. HA-tagged PGBD5 was overexpressed in HEK293 from a transfected plasmid and subsequently pulled down. Co-IP experiments followed by Western Blot analysis were conducted using specific antibodies as indicated, along with an IgG control. The purified protein samples were probed with different antibodies as indicated.

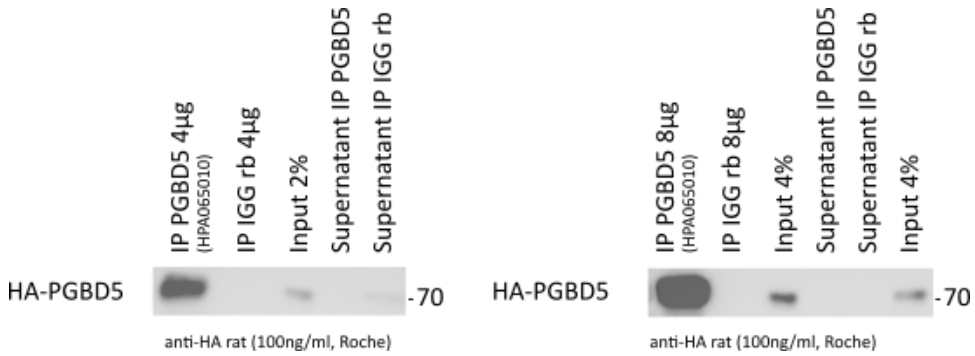


Figure S 10: Endogenous PGBD5 antibody HPA065010 successfully detected HA-tagged PGBD5 in co-IP. HA-tagged PGBD5 was overexpressed in HEK293 from a transfected plasmid and subsequently pulled down. Co-IP experiments followed by Western Blot analysis were conducted using specific antibodies as indicated, along with an IgG control. The purified protein samples were probed with different antibodies as indicated.

Table S 2: List of antibodies and chemicals used in the study

Item	**company**	**cataloge number**
Primary Antibodies		
Anti-PGBD5 (7-F8-5) (mouse)	Novusbio	#NBP2-67048
Anti-Actin	Dianova	#DLN-07276
Anti-HA (rabbit)	Abcam	#ab9110
Anti-HA (rat)	Sigma Aldrich	#11867423001
Anti-PGBD5 (mouse)	Mybiosource	#MBS355128
Anti-PGBD5 (mouse)	Antikoerper-online.de	#ABIN1854961
Anti-PGBD5 (rabbit)	AtlasAntibodies	#HPA065010
Anti-PGBD5 (rabbit)	Biorbyt Ltd	#orb13159
Anti-TAF1C (rabbit)	Thermo Fisher Scientific	#A303-698A
Anti-TOP2A (rabbit)	Abcam	#ab52934
Anti-TOP2B (rabbit)	Abcam	#ab72334
Anti-TOP2B (rabbit)	Abcam	#ab264158
Anti-UBF (F-9) (mouse)	Santa Cruz	#sc-13125
Anti-UBF Polyclonal Antibody (rabbit)	Bethyl Laboratories	#A301-859A

Anti-FBXL19 (rabbit)	abcam	#ab172961

Secondary Antibodies

Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP	Thermo Fisher Scientific	#31432
Goat anti-Rat IgG (H+L) Secondary Antibody, HRP	Thermo Fisher Scientific	#31470
Goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP	Thermo Fisher Scientific	#31462
Alexa Fluor® 647 goat anti-rabbit	Life Technologies	A-21244
Alexa Fluor® 647 goat anti-mouse	Life Technologies	A-21236
Alexa Fluor® 488 donkey anti-rabbit	Thermo Scientific	A-21206
Alexa Fluor® 488 goat anti-mouse	Invitrogen	A31620

Conjugated Antibodies

NucBlue™ Live ReadyProbes™ Reagent (Hoechst 33342)	Thermo Fisher Scientific	#R37605

Chemicals

Agarose	Sigma Aldrich	#A9414-5G
ascorbic acid	Sigma Aldrich	#A8960-5G
Attachment factor solution	Sigma Aldrich	#123-100
B-27 supplement (50x)	Gibco	#17504-044
BCA Protein assay	Pierce	#23225
Benzonase	Novagen	#70746-3
Bromphenol blue	Sigma Aldrich	#B5525-5G
BSA (Fatty acid-free)	Proliant	#68700
BSA (MACS Stock Solution)	Miltenyi Biotec	#130-091-376
cAMP	Merck Millipore	#28745
CellTiter-Glo® Luminescent Cell Viability Assay	Promega	#G7570

cloneR	Stemcell Technologies	#05889
Cytosine -D-arabinofuranoside (AraC)	Sigma Aldrich	#C1768
Direct-zol RNA micro-prep kit	Zymo Research	#R2062
Direct-zol RNA mini-prep kit	Zymo Research	#R2050
Dithiothreitol (DTT)	Invitrogen	#D1532
DMEM/F12 (+L-Glutamine + 15mM HEPES)	Gibco	#11330-032
DMEM/F12, GlutaMax	Gibco	#31331028
Doxycycline hyclate	Sigma Aldrich	#D9891
DPBS, no calcium, no magnesium-500 mL	Life Technolgies	#14190094
Dulbecco's modified Eagle's medium (DMEM 1x + GlutaMAX +4.5 g/L D-Glucose + Pyruvate)	Gibco	#31966-021
ECL Prime Western Blotting Detection Reagent (Amersham)	GE Healthcare	#RPN2232
Essential 8 Media Kit	Gibco	#A1517001
Ethanol	Carl Roth	#P076.2
Ethylenediaminetetraacetic acid (EDTA)	SERVA	#11280.02
Etoposide (VP-16)	Merck	#E1383-25MG
Fast-cast Acrylamide Kit 10%, stain-free	Bio-Rad	#161-0183
Fast-cast Acrylamide Kit 12%, stain-free	Bio-Rad	#161-0185
Fetal Bovine Serum(FBS)	Life Technolgies	#10500064
Formaldehyde (16%, methanol free)	ThermoFisher Scientific	#28906
GDNF	Peprotech	#450-10-10
GFP mRNA	BIOZOL	#OZB-MRNA11-20
GlutaMAX	Gibco	#35050-038
Glycerol	Carl Roth	#3783.2
Glycine	Carl Roth	#3908.3
hBDNF	Peprotech	#450-02-10

HEPES	Sigma Aldrich	#H4034-500G
High-Capacity RNA to cDNA reagents	Applied Biosystems	#4387406
HiScribe T7 ARCA mRNA Kit with tailing	New England Biolabs	#E2060S
hNT-3	R&D Systems	#267-N3-025
Human Topoisomerase 2 DNA Decatenation Assay Kit	ProFoldin	#HDC100KE
jetPRIME® transfection reagent	Polyplus	#n0114-15
laminin	Sigma Aldrich	#L2020
LB-Medium (Lennox)	Carl Roth	#X964.2
Lipofectamine™ MessengerMAX™ Transfection Reagent	Invitrogen	#LMRNA001
Matrigel hESC-Qualified Matrix, 5ml	ThermoFisher Scientific	#11573560
Midori Green	Biozym	#617004
Milchpulver Blotting grade, pulv., fettarm	Carl Roth	#T145.3
Mini-PROTEAN TGX Gels	Bio-Rad	#4561094
Mounting Media (Antifade)	Vectashield	#H-1000
N2 supplement (100x)	Gibco	#17502-048
NaCL Natriumchlorid min. 99,5 %, p.a., ACS, ISO	Carl Roth	# 3957.2
NEAA (100x)	Gibco	#11140-035
Neurobasal medium	Gibco	#2103-049
N-Methyl-D-aspartic acid (NMDA)	Sigma Aldrich	#M3262-100MG
Nonidet-P40 (NP40)	USBIOLOGIE	#C8051909
Opti-MEM™ I Serumreduziertes Medium	Gibco	#10149832
PageRuler Plus Prestained Protein Ladder, 10 to 250 kDa-2 x 250 µL	ThermoFisher Scientific	#11832124
Penicillin-Streptomycin (10.000 U/ml)	Gibco	#11548876 / #15140122
Phenol/Chloroform/Isoamylalkohol	Carl Roth	#A156.1
Phosphatase inhibitors	Active Motif	#37492
Plasmid EasyPure NucleoSpin	MACHEREY-NAGEL	#740727.250

Power SYBR green PCR Master Mix	Applied Biosystems	#4367659
Protease inhibitors	Pierce	#A32955
Protein A Dynabeads	Life Technolgies	#10001D
Protein G Dynabeads	Life Technolgies	#10004D
Proteinase K	Thermo Fisher Scientific	#EO0491
RNase A	Thermo Fisher Scientific	#EN0531
Sodium Acetate (3 M), pH 5.5, RNase-free	Life Technolgies	#AM9740
sodium dodecyl sulfate (SDS)	SERVA	#20783.02
StemFlex Medium	Gibco	#A3349401
Trans-Blot® TurboTM Midi-size Transfer Stacks 8.5cm x 13.5 cm, 40 stacks	Bio-Rad	#1704373
Transfer Buffer Trans-Blot Turbo	Bio-Rad	#10026938
Trichlormethan/ChloroformROTISOLV®	Carl Roth	#7331.2
TRIS-hydrochlorid (TRIS-HCl)	Carl Roth	#9090.2
Tritron X-100	Sigma Aldrich	#T8787-100ML
Trizma base	Sigma Aldrich	#T1503-5KG
Trizol	ThermoFisher Scientific	#15596-018
Trypan Blue solution	Sigma Aldrich	#T8154-100ML
TWEEN 20 Detergent	Merk	#655204-100ml
Versene	Gibco	#15040-033
Water (RNase and DNase free)	Sigma Aldrich	#W3513-100ml
Xtra Midi EF NucleoBond	MACHEREY-NAGEL	#740420.50
Y-27632 (Rock inhibitor)	Peprotech	#1293823
yeast tRNA	Invitrogen	#AM7119

Table S 3: List of machines used in the study

Machine	Company	Notes	Software

Chemi Doc MP Imaging System	Bio-Rad	Western Blot, Agarose Gels	
Trans-Blot Turbo Transfer System	Bio-Rad	Western Blot	
Tecan reader	Thermo Fisher Scientific	BCA	
Proxeon EASY-nLC II system	Thermo Fisher Scientific	LC/MS system	
Q Excative mass spectrometer	Thermo Fisher Scientific	Mass Spectrometer	
Leica TCS SP8	Leica	Inverted Confocal Microscope	LAS X
CFX96 Real Time System, C1000 Touch Thermal Cycler	Bio-Rad	qPCR machine	Maestro
DS-11 FX+ Spectrophotometer/Fluorometer	DeNovix	Nanodrop	
PCRmax Alpha Cycler	PCRmax	PCR machine	
BioRuptor Pico sonicator	Diagenode	Shearing & Protein Lysis	

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbständig und ohne unerlaubte fremde Hilfe angefertigt, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Quellen und Hilfsmitteln wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Berlin, den 3. Mai 2023