Svoluji k zapůjčení své dizertační práce ke studijním účelům a prosím, aby byla vedena přesná evidence vypůjčovatelů. Převzaté údaje je vypůjčovatel povinen řádně citovat.

Charles University

Faculty of Science

Molecular and Cellular Biology, Genetics and Virology

Mgr. Martina Benešová

TET1 overexpression, DNA hypomethylation and aberrant expression of human endogenous retrovirus ERVWE1 in germ cell tumors

Zvýšená produkce mRNA dioxygenázy TET1, nízká hladina metylace DNA and změny exprese lidského endogenního retroviru ERVWE1 v nádorech zárodečných buněk

Doctoral thesis

Supervisor: Mgr. Kateřina Trejbalová, PhD.

Laboratory of Viral and Cellular Genetics

Institute of Molecular Genetics of the ASCR, v. v. i.

Prague, 2019

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Declaration:

I declare that I wrote the doctoral thesis myself. All the information sources used in the thesis have been cited. Neither this thesis, nor its substantial part were used for gaining another or the same academic degree.

Prague, 14. 11. 2019

….....

Acknowledgement

I would like to acknowledge Kateřina Trejbalová for the supervision of my laboratory work, all the cunsultations, and helpfulness, Jiří Hejnar for valuable consultations, and the members of the Laboratory of Viral and Cellular Genetics for both consultations and a pleasant working enviroment. Further, I would like to acknowledge Zdenka Vernerová (Department of Pathology, Third Faculty of Medicine, Charles University in Prague) for sample collecting and their microscopic analysis and Rachel Amoroux (MRC London Institute of Medical Sciences, London, UK and Institute of Clinical Sciences, Faculty of Medicine) for LC/MS analysis. Many thanks further belons to my family, especially to my husband and little daughter, for their love, patience and support.

Grant support

The work was supported by grant No. 13-3700S (Czech Science Foundation, RNDr. Jiří Hejnar, CSc.) and grant No. NT14601-3/2013 (Grant Agency of Ministry of Health of the Czech Republic, RNDr. Jiří Hejnar, CSc., doc. MUDr. Zdenka Vernerová, CSc., MUDr. Petr Klézl). The work was further institutionally supported by RVO: 68378050 and the research project PRVOUK – Oncology P27 (Charles University in Prague, doc. MUDr. Zdenka Vernerová, CSc., MUDr. Petr Klézl, MUDr. Arpád Szabó).

Abstract

TGCTs are tumors of male germ cells. They comprise of seminomas and non-seminomas (embryonal carcinoma, yolk sac tumor, choriocarcinoma, and teratoma). GCT types differ in the stage of differentiation, from undifferentiated seminoma to more differentiated non-seminomas. In our studies, we aimed to characterize specific epigenetic features of GCT types that enable transcription derepression of the human endogenous retrovirus *ERVWE1* in these tumors. We detected upregulated mRNA expression of *TET1-3* dioxygenases in GCTs, especially of *TET1* in seminomas. Moreover, seminomas showed low global levels of 5mC and 5hmC. TET1 knock-down in a seminoma-derived cell line resulted in a decreased amount of 5hmC and unchanged 5mC level. These results stress the dynamics of cytosine modifications, which has not been precisely described yet. Further, we observed high level of *ERVWE1* transcript together with efficient RNA splicing in seminomas. Detected *ERVWE1* transcription is independent of the expression of other examined endogenous retroviruses. *ERVWE1* transcription derepression corresponds with the low global level of 5mC detected in seminomas, which involves extensive DNA hypomethylation of the *ERVWE1* promoter. We propose the high TET1 dioxygenase expression as s marker of undifferentiated GCTs. Furthermore, we propound the high *ERVWE1* RNA expression and its efficient splicing as a marker of seminomas and the seminoma component of mixed GCT.

Key words: TET1, 5-hydroxymethylcytosine, 5-methylcytosine, seminoma, germ cell tumor, human endogenous retrovirus, ERVWE1, promoter DNA methylation, RNA transcription, RNA splicing

Abstrakt

Testikulární nádory zárodečných buněk se dělí na seminomy a neseminomy (embryonální karcinom, nádor žloutkového váčku, choriokarcinom a teratom). Tyto nádory se mezi sebou liší ve stupni diferenciace, kdy seminom představuje jejich nejméně diferencovaný typ. Našim cílem bylo charakterizovat specifické epigenetické vlastnosti nádorů mužských zárodečných buněk, které zde umožňují transkripci lidského endogenního retroviru ERVWE1. Náš výzkum odhalil zvýšenou hladinu mRNA *TET1-3* dioxygenáz, zejména pak *TET1* v seminomech. Pro seminomy byla také charakteristická nízká celková hladina 5mC a 5hmC. TET1 knock-down v buněčné linii odvozené od seminomu vedl ke snížení množství 5hmC, zatímco hladina 5mC se nezměnila. Tyto výsledky poukazují na význam málo prozkoumané dynamiky cytosinových modifikací. Nízká celková hladina 5mC v seminomech se odráží i v nízké hladině metylace DNA promotorové oblasti *ERVWE1*. Nízká hladina metylace DNA v promotorové oblasti je jedním z faktorů, které umožňují pozorovanou intenzivní transkripcí *ERVWE1* a účinný sestřih vzniklé RNA. Transkripce dalších zkoumaných lidských endogenních retrovirů neodpovídá produkci pozorované u *ERVWE1*. Z našich výsledků vyplývá, že vysoká hladina dioxygenázy TET1 je charakteristickou vlastností nediferencovaných typů nádorů zárodečných buněk. Navíc, vysoká hladina RNA *ERVWE1* a její účinný sestřih je charakteristický pro seminomy a seminomovou složku smíšených nádorů zárodečných buněk.

Klíčová slova: TET1, 5-hydroxymetylcytosin, 5-metylcytosin, seminom, nádor zárodečných buněk, lidský endogenní retrovirus, ERVWE1, metylace DNA v promotorové oblasti, transkripce RNA, sestřih RNA

Contents

Abbreviations

A	adenine
AA	amino acid
AC	adenylyl cyclase
AID	activation-induced cytidine deaminase
AlkB	alkane 1-monooxygenase
AP-2	adaptor protein-2
AP-2γ	adaptor protein-2γ
APC	adenomatous polyposis coli
APOBEC	apolipoprotein B mRNA editing enzyme catalytic polypeptide-like
ASCT1	alanine/serine/cysteine/threonine transporter 1
ASCT2	alanine/serine/cysteine/threonine transporter 2
ATP	adenosine triphosphate
BCL9L	B-cell Lymphoma 9-like protein
BER	base excision repair
BeWo	human choriocarcinoma-derived cell line
BHQ1	black hole quencher-1
bp	base pairs
BS-seq	bisulfite sequencing
C	cytosine
CaCl2	calcium chloride
CAMKI	calcium/calmodulin-dependent protein kinase 1
cAMP	cyclic adenosine monophosphate
Ca2+	calcium
$CaCl_2$	calcium chloride
CBP	CREB binding protein
CCR	chemokine receptor
cDNA	complementary deoxyribonuclease acid
Cdx2	caudal type homeobox 2
CD8+	cluster of differentiation 8+
CD30	cluster of differentation 30
CHC	choriocarcinoma
CIMP	CpG island methylator phenotype
CKIα	casein kinase Iα
CMS	5-methylene sulfonate
CpG	5´- cytosine – phosphate – guanin - 3´
CREB	cAMP-responsive element binding protein
CREM	cAMP-responsive element modulator
CRIPTO	Cryptic family gene
CT	cytotrophoblast
CXCR4	C-X-C chemokine receptor type 4
CXXC	cytosine-X-X-cytosine
CYT	cytoplasmic tail
C/EBPα	CCAAT Enhancer Binding Protein Alpha
DAZL1	deleted in azoospermia-like-1
dCTP	deoxycytidine triphosphate
ddPCR	droplet digital PCR
dhmCTP	deoxyhydrogenmethylcytidine triphosphate
DMEM	Dulbecco´s modified Eagle medium
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
DNMT1	DNA methyltransferase 1
DNMT2	DNA methyltransferase 2
DNMT3A	DNA methyltransferase 3A
DNMT3B	DNA methyltransferase 3B
DNMT3L	DNA methyltransferase 3-like
DUSP23	dual-specificity phosphatase 23
Dvl	disheveled
EC	embryonal carcinoma

EDTA	ethylenediaminetetraacetic acid
ENV	envelope glycoprotein
EnvPb	envelope glycoprotein, group Pb
EnvV	envelope glycoprotein, group V
Epac1	exchange protein directly activated by cAMP 1
ERV	endogenous retrovirus
ERVFRDE1	locus coding Env glycoprotein, chromosome 6q24.1, family FRD, member 1
ERVWE1	locus coding Env glycoprotein, chromosome 7q 21 – 22, family W, member 1
ERV3	endogenous retrovirus 3
ES	embryonal stem cell
EVT	extravillous trophoblast
FAM	6-fluorescein amidite
FBS	fetal bovine serum
FP	fusion peptide
FW	forward
FBXL11/KDM2A	F-box and leucine-rich repeat protein 11/lysine-specific demethylase 2A
F-12	F-12 nutrient mixture
G	guanine
Gadd45	Growth Arrest And DNA-Damage-Inducible 45
GAG	group specific antigen
Gata2	GATA binding protein 2
GCMa	glial cell missing a
GCM1	glial cell missing 1
GCNIS	germ cell neoplasia *in situ*
GCT	germ cell tumor
GFP	green fluorescent protein
GLI2	glioma-associated oncogene 2
gPr73	glycoprotein precursor 73
gp24	glycoprotein 24
gp50	glycoprotein 50
GSK-3β	glycogen synthase kinase 3β
G-protein	guanine nucleotide-binding protein
H	hydrogen
hCG	human chorionic gonadotrophin
HDACs	histone deacetylases
HeLa	human cervical cancer-derived cell line
HELLP	hemolysis, elevated liver enzymes, low platelets
HERV	human endogenous retrovirus
HERV-Fb1	human endogenous retrovirus, family Fb1
HERV-FRD	human endogenous retrovirus, family FRD
HERV-K	human endogenous retrovirus, family K
HERV-W	human endogenous retrovirus, family W
hES	human embryonic stem cell
HEX	hexachloro-fluorescein
HOXA9	homeobox A9
hPL	human placental lactogen
H2A.Z	histone H2A family member Z
H3K4me1	histone 3 lysine 4 monomethylation
H3K4me2/3	histone 3 lysine 4 di/trimethylation
H3K9ac	histone 3 lysine 9 acetylation
H3K9me2	histone 3 lysine 9 dimethylation
H3K9me3	histone 3 lysine 9 trimethylation
H3K27me3	histone 3 lysine 27 trimethylation
H3K36 me3	histone 3 lysine 36 trimethylation
IDH1	isocitrate dehydrogenase 1
IDH2	isocitrate dehydrogenase 2
IFN	interferon
Igf2	insulin-like growth factor 2
IHC	immunohistochemistry
IPTG	isopropyl β-D-1-thiogalactopyranoside
ISD/I	immunosuppresive domain
IUGR	intrauterine growth restriction
JAR	human choriocarcinoma-derived cell line

JEG-3	human choriocarcinoma-derived cell line
JMJD3	Jumonji Domain-Containing Protein 3
K	lysine
kbp	kilobase pairs
KIT	tyrosine-protein kinase
Klf4	Kruppel Like Factor 4
$KRuO_4$	potassium perruthenate
LB	lysogeny broth medium
LC/MS	liquid chromatography-mass spectrometry
LINE-1	long interspersed nuclear element – 1
LHCGR	luteinizing hormone/gonadotrophin receptor
LTR	long terminal repeat
LTR9b	long terminal repeat element 9b
LZ	leucine zipper
MaLR	mammalian apparent-LTR retrotransposon
Mdb4	Methyl-CpG Binding Domain 4
MEM-D	minimum essential medium
$MgCl_2$	magnesium chloride
MGMT	O-6-Methylguanine-DNA Methyltransferase
miR-199-3P	microRNA-199-3P
miR-371a-3p	microRNA-371a-3p
MLL	myeloid/lymphoid leukemia
$MnCl_2$	manganese chloride
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger ribonucleic acid
MSFD2	major facilitator superfamily domain containing 2
N	asparagine
N	nitrogen
$NaHCO_3$	sodium hydrogen carbonate
NANOG	Nanog homeobox
NaOH	sodium hydroxide
NEIL	endonuclease VIII-like protein
NOS	not otherwise specified
O	oxygen
OCT3/4	octamer-binding transcription factor 3/4
ORF	open reading frame
oxBS-Seq	oxidative bisulfite sequencing
PBS	phosphate buffered saline
PBMC	peripheral blood lymphocyte
PCR	polymerase chain reaction
PGC	primordial germ cell
PKA	protein kinase A
PLAP	placental-like alkaline phosphatase
POL	polymerase
POLR2A	α-subunit of DNA-directed RNA polymerase II
polyA	polyadenylic acid
POU5F1	POU Domain Class 5, Transcription Factor 1
p63	protein 63
R	arginine
Rap1	Ras-related protein
RASSF1A	Ras Association Domain Family Member 1A
RbCl	rubidium chloride
RNA	ribonucleic acid
RPMI-1640	culture medium, developed at Roswell Park Memorial Institute
RPP30	ribonuclease P protein subunit P30
rRNA	ribosomal ribonucleic acid
RT	reverse transcriptase
RV	reverse
qRT-PCR	quantitative real-time polymerase chain reaction
SALL4	Sal-like protein 4
SCGB3A1	secretoglobin Family 3A Member 1
SDF1	stromal cell-derived factor 1
SE	seminoma

SENP1	sentrin-specific protease 1
siRNA	small interfering RNA
SLC1A4	solute carrier family 1 member 4
SLC1A5	solute carrier family 1 member 5
SMUG 1	Single-Strand-Selective Monofunctional Uracil-DNA Glycosylase 1
SOX2	sex determining region Y box-2
SP	signal peptide
Sp-1	specificity protein 1
SU	surface subunit
Suv39h	histone H3 lysine 9 specific histone methyltransferases
ST	syncytiotrophoblast
Syncytin-Car1	envelope glycoprotein Syncytin in Carninora, member 1
Syncytin-Mab1	envelope glycoprotein Syncytin in Mabuya, member 1
Syncytin-Mar1	envelope glycoprotein Syncytin in Marmotini, member 1
Syncytin-Opo1	envelope glycoprotein Syncytin in Opossum, member 1
Syncytin-Rum1	envelope glycoprotein Syncytin in Ruminantia, member 1
Syncytin-Ten1	envelope glycoprotein Syncytin in Tenrecidae, member 1
S-gal	3,4-Cyclohexenoesculetin β-D-galactopyranoside
T	thymine
T	tumor
T2	type 2 bacteriophage
T4	type 4 bacteriophage
T6	type 6 bacteriophage
TAE	Tris-acetate EDTA
TAB-Seq	TET-assisted bisulfite sequencing
TBP	TATA-box binding protein
TCA	tricarboxylic cycle
TCF	T-cell factor
TE	teratoma
Tfcp2l1	transcription factor CP2 like 1
TGCT	testicular germ cell tumor
TDG	thymine-DNA glycosylase
TCF	T-cell factor
TET	ten-eleven translocation
TET1	ten-eleven translocation 1
TET2	ten-eleven translocation 2
TET3	ten-eleven translocation 3
TGF-β	transforming growth factor β
TM	transmembrane subunit
TNM	tumor-nodes-metastasis
TSC	trophoblast stem cell
TSE	trophoblast specific enhancer
T-Cam2	human seminoma-derived cell line
U	unit
U	uracil
UDP-glucose	uridine diphosphate glucose
URE	upstream regulatory element
V	volt
W	tryptophan
Wnt	Int-1 and Wingless
Wnt10b	Int-1 and Wingless ligand 10b
YST	yolk sac tumor
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5gmC	β-glucosyl-5-hydroxymethylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
βGT	β-glucosyltransferase

Introduction

Syncytin-1, the envelope protein of human endogenous retrovirus ERVWE1, induces cell-cell fusion of placental trophoblast. This endogenous retrovirus have been widely studied in our laboratory, the Laboratory of viral and cellular genetics. It was found that *ERVWE1* expression in placenta is connected with the low level of *ERVWE1* promoter CpG methylation. Furthermore, in the non-placental tissues Syncytin-1 was found to be repressed by epigenetic mechanisms. Nevertheless, *ERVWE1* expression and efficient RNA splicing together with the low level of promoter CpG methylation, were described in testicular germ cell tumors (TGCTs). However, the intriguing observations were done on the small panel of samples and TGCT type-specific expression was not determined. Therefore, we collected large panel of TGCT samples and controls to analyze the phenomenon of *ERVWE1* expression in TGCTs more deeply. Moreover, we wanted to find the specific epigenetic features of TGCTs that allows the expression of the human endogenous retrovirus ERVWE1.

TGCTs are the tumors derivated from male germ cells. Germ cell tumors (GCTs) originate from a gonocyte or primordial germ cell (PGC) with delayed or blocked maturation, from which the germ cell carcinoma *in situ* (GCNIS) is established. It is believed that GCNIS is formed during the fetal development and stays dormant for a certain period. From GCNIS, seminomas or embryonal carcinomas originate. Moreover, it is supposed, that from seminoma, embryonal carcinoma can be established as well. From embryonal carcinoma, yolk sac tumor, choriocarcinoma and teratoma are formed. These four types of tumors are collectively called non-seminomas. Furthermore, there are mixed GCTs that contain two or more of the above-mentioned GCT types. Importantly, GCTs can be distinguished by the stage of differentiation, while GCNIS and seminomas represent undifferentiated types. Furthermore, the stage of differentiation corresponds with the amount of global DNA methylation. Therefore, GCNIS and semiomas display a low level of DNA methylation, embryonal carcinomas a higher level and yolk sac tumors, choriocarcinomas, and teratomas are hypermethylated. The level of DNA methylation in GCT types is associated with differences in gene expression. Due to previously observed expression of human endogenous retrovirus ERVWE1 in GCTs, we focused on analysis of the selected epigenetic features of different GCT types that allow this transcription derepression.

As GCNIS and seminomas display the lowest levels of DNA methylation, expression from the genes that are silenced under physiological conditions can occur. This refers to

endogenous retroviruses. Endogenous retroviruses are remnants of infectious retroviruses that became the integral parts of the host genome. Most of the human endogenous retroviruses have accumulated inactivating mutations and STOP-codons during the evolution, and therefore they are not replication competent and do not express functional viral proteins. Nevertheless, there are exceptions. Some of the viral proteins are translated and even gained a function in the host organism. This is the case of human endogenous retrovirus ERVWE1. Under physiological conditions, the ERVWE1 envelope protein Syncytin-1 is expressed in human placenta, where it mediates fusion of cytotrophoblasts into the syncytiotrophoblast layer. Therefore, Syncytin-1 is necessary for the proper formation of the human placenta. The expression from the *ERVWE1* locus is known to be regulated by the presence of a specific transcription factor, by the level of promoter DNA methylation and by RNA splicing. This very strict control of Syncytin-1 production is needed due to its fusogenic properties.

In this work, we examined a wide panel of human GCT samples of different types for selected epigenetic characteristics and for the expression pattern of the ERVWE1 endogenous retrovirus.

Aims of the Thesis

PGCs are embryonic precursors of gametes. DNA modifications are crucial for the maintenance of the characteristic features of PGCs and for their further differentiation (Hajkova et al. 2002; Seisenberger et al. 2012). GCTs can arise from PGCs with delayed or blocked maturation, therefore changes in DNA modifications occur during the tumor transformation (Skakkebék et al. 1998; Oosterhuis and Looijenga 2005; Rajpert-De Meyts 2006; Kristensen et al. 2013). Due to previously observed ERVWE1 derepression in GCTs (Trejbalová et al. 2011), we focused on analysing selected epigenetic features that can lead to *ERVWE1* transcription in these tumors. Our aims were following:

- to determine the TET1, TET2 and TET3 expression pattern in different types of GCTs
- to define *DNMT* mRNA expression in GCTs
- to quantify the global levels of 5mC and 5hmC in GCTs
- to characterize the effect of TET protein down-regulation on the 5mC and 5hmC levels in a model cell line

The very low global level of 5mC observed in seminomas (Smiraglia et al. 2002; Netto et al. 2008) could favour expression from epigenetically-silenced genes – such as ERVWE1/Syncytin-1 and ERVFRDE1/Syncytin-2. The expression from these genes of retroviral origin could be employed for the GCT type determination. Therefore, we focused on the characterization of ERVWE1 and ERVFRDE1 expression in GCTs. Our aims were following:

- to determine the *ERVWE1* and *ERVFRDE1* mRNA expression pattern in different types of GCTs
- to examine the known regulators of *ERVWE1/ERVFRDE1* transcription in GCTs
- to determine the specificity of *ERVWE1/ERVFRDE1* expression in GCTs in comparison with other endogenous retroviruses

Literature Review

The first part of the literature review is focused on DNA methylation as one of the principal epigenetic mark influencing gene expression. DNA methylation establishment and maintenance as well as the ways of DNA demethylation are described. Special attention is paid to the alteration of the DNA methylation pattern in various types of tumors.

1 DNA methylation

Methylation is one of the most important DNA modifications in vertebrates. This epigenetic mark is formed mainly by addition of the methyl group to the fifth position of cytosine in the context of CpG dinucleotide. In this way 5-methylcytosine (5mC) is formed (Fig. 1). The donor of the methyl group is predominantly S-adenosylmethionine.

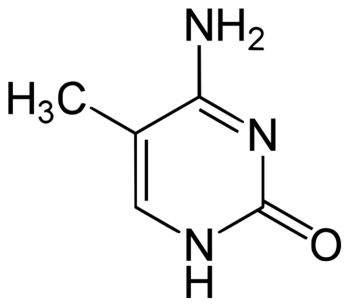

Fig. 1: Skeletal formula of 5mC. *5mC is a DNA pyrimidine base. C = carbon, H = hydrogen, N = nitrogen, O = oxygen. The figure was taken from https://en.wikipedia.org/wiki/5-Methylcytosine#/media/File:5-Methylcytosine.svg.*

DNA methylation is established and maintained enzymatically by DNA methyltransferases (DNMTs) – DNMT1, DNMT2, DNMT3A, and DNMT3B. DNMT1 is important for maintaining the methylation pattern during the replication process (Bestor et al. 1988), and therefore is called a maintenance methyltransferase. DNMT3A and DNMT3B, the *de novo* methyltransferases, are involved in establishing the methylation pattern in the developing embryo (Okano et al. 1999). DNMT2 was classified among DNMTs due to its sequential similarity with other DNMTs (Yoder and Bestor 1998); however, later it was found that its substrate is RNA (Goll et al. 2006). DNMT3L belongs to the DNMT group; nevertheless, the protein lacks the catalytic domain and hence does not act as a methyltransferase (Hata et al. 2002). However, DNMT3L is important for direct stimulation of DNMT3A and DNMT3B proteins (Suetake et al. 2004). Classificatin of DNMTs into the

maintenance and *de novo* categories is not really strict, because some redundancy between the methyltransferase functions was observed.

The presence of 5mC is particularly important at promoters, where it is linked with both short-time and long-time transcription repression (Gama-Sosa et al. 1983). Moreover, DNA methylation can be presented at gene bodies, and DNA methylation out of the CpG dinucleotides is possible (R. Lister et al. 2009); however their effect on transcription is hardly predictable (Malone et al. 2001). DNA methylation-induced transcription repression is mediated by recruiting methyl-binding proteins, whose presence prevents binding of transcription factors to DNA and blocks gene expression (Hendrich and Bird 1998). Due to the effect of DNA methylation on transcription, DNA methylation has an irreplaceable function both during the fetal development and the lifetime of the organism.

1.1 DNA methylation during the development, differentation and tumorogenesis

The potential of DNA methylation to regulate transcription from a specific gene or from the whole chromosome has a serious impact on functioning of the whole organism. During development of a human organism and in the course of its lifetime, changes in the level of DNA methylation physiologically occur. One of the examples is X chromosome inactivation. In eutherian mammals, one X chromosome of the female sex is inactivated by DNA methylation to compensate for the dosage of two X (Riggs 1975). Changes in the level of DNA methylation are also observed during the process called imprinting, when one of the two inherited parental alleles is silenced by joint activities of DNA methylation, histone modifications and non-coding RNAs, reviewed in (Elhamamsy 2017). Further, during the embryonic development, the level of DNA methylation influences transcription from specific genes that regulate cellular and tissue differentiation. In the differentiation process, the cells/tissues lose their developmental potential and gain functional specialization.

Moreover, DNA methylation is important in transcription repression of endogenous retroviruses and transposons (Gama-Sosa et al. 1983). The potential expression of retroelements is dangerous, because their efficient replication followed by provirus integration can cause mutations in the genome or can lead to changes in the gene expression (Chapter 3 in the Literature review). Therefore, expression of repetitive elements must be strictly regulated.

In the mouse fetal development, there are two phases of the genome DNA demethylation. A similar scenario is supposed to occur in humans as well. The first wave is passive DNA demethylation which occurs during the PGC migration to the genital ridge.

Even though the majority of the DNA is demethylated, there are specific sites in the genome where DNA methylation is actively maintained, such as retroelements (Hajkova et al. 2002). In the process of passive DNA methylation, the 5mC pattern is re-established during the DNA replication. The second wave, active DNA demethylation which is performed enzymatically (Chapter 1.3 in the Literature review), takes place in PGCs as soon as they reach the gonad (Seisenberger et al. 2012). During this wave almost all the genome is demethylated. Despite the extensive DNA demethylation event, this does not lead to expression of elements such as retrotransposons (Seisenberger et al. 2012). The findings highlight the importance of the transcriptional silencing of retroelements.

Pathological changes in the level of DNA methylation can have a serious impact on the functioning of both cells and the whole tissue. Altered DNA methylation can modify expression of tumor suppressor genes or oncogenes and therefore disrupt cell homeostasis, resulting in potential tumor formation. In most of the tumors, there is a lower level of DNA methylation in comparison with the adjacent healthy tissue (Goelz et al. 1985), like in ovarian carcinoma (Cheng et al. 1997), prostate cancer (Bedford and van Helden 1987), hepatocellular carcinoma (C. H. Lin et al. 2001), seminomas, and embryonal carcinomas (Netto et al. 2008). However, there are some tumors with opposite characteristics (Frigola et al. 2006), like part of colorectal tumors, reviewed in (Toyota and Issa 1999). The phenomenon, which is characterized by a very high level of DNA CpG methylation at some genes, is called CpG island methylator phenotype (CIMP). In the case of colorectal cancer, the level of CIMP can be used for tumor classification (Yagi et al. 2010) and furthermore corresponds to the disease prognosis and reaction to chemotherapy (Exner et al. 2015; Van Rijnsoever et al. 2003). As the tumor formation is associated with changes in the level of DNA methylation of specific genes or of the whole genome, the presence of 5mC in tumors is broadly studied.

1.2 5-Methylcytosine detection

The level of DNA methylation can be measured by bisulfite sequencing (Frommer et al. 1992). The method allows 5mC detection in a specific locus, or can be combined with the Next Generation Sequencing for the whole genome analysis. Furthermore, the 5mC presence can be examined by methyl-specific restriction enzymes, liquid chromatography-mass spectrometry (LC/MS), and by real time methylation-specific PCR.

In the process of bisulfite sequencing, genomic DNA is first denatured. During the denaturation, the two complementary strains are separated. Thereafter, sodium bisulfite is

added. Cytosine and 5mC react with sodium bisulfite in different ways – cytosine is deaminated into uracil, while 5mC remains unchanged. This is caused by the fact that the conversion of 5mC to uracil is very slow (R. Y. H. Wang et al. 1980). Bisulfite-treated DNA is then amplified by PCR reaction using primers specific for the modified sequence, sequenced, and compared with the non-treated DNA. The method of bisulfite sequencing is summarized in Fig. 2.

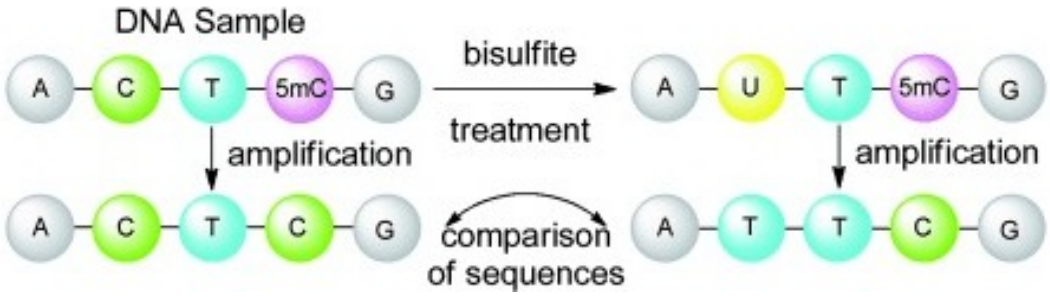

Fig. 2: Overview of bisulfite sequencing. *The analyzed DNA sample is treated with sodium bisulfite – C is converted into U and 5mC remains unchanged. Bisulfite-treated DNA is amplified in PCR reaction, sequenced, and compared with non-treated DNA. C in non-treated DNA is converted to T in bisulfite-treated DNA, 5mC in non-treated DNA remains as C in bisulfite-treated DNA. A = adenine, C = cytosine, G = guanine, T = thymine, U = uracil, 5mC = 5-methylcytosine. The figure was taken from (Schüler and Miller 2012).*

The total level of 5mC in a specific locus or in the whole genome can be measured by LC/MS technique. The DNA of interest is digested to single nucleotides, separated by liquid chromatography and analyzed by mass spectrometry. The total amount of 5mC or otherwise modified C is counted.

The presence of 5mC can be further analyzed by the real-time methylation-specific PCR (Yoshioka et al. 2018). In this method, two PCR reactions with bisulfite treated DNA are performed, one with primers specific for a fully methylated sequence and the second with primers specific for a fully unmethylated sequence. The amount of both products is compared. The method allows detection of the 5mC presence only in the sequence of the used primers, not in the whole amplified locus.

1.3 Active DNA demethylation

1.3.1 5-Hydroxymethylcytosine

The process of active DNA demethylation is performed enzymatically (Chapter 1.3.2), while the 5-hydroxymethylcytosine (5hmC) intermediate is established. The skeletal formula of 5hmC is depicted in Fig. 3. Even though 5hmC is mainly detected in DNA, it is present in

RNA as well, but to a lesser extent (Yoshioka et al. 2018).

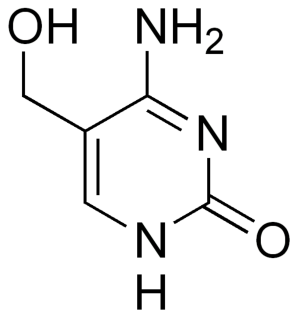

Fig. 3: Skeletal formula of 5hmC. *5hmC is a DNA pyrimidine base. H = hydrogen, N = nitrogen, O = oxygen. The figure was taken from https://en.wikipedia.org/wiki/5-Hydroxymethylcytosine#/File:Hydroxymethylcytosine.png.*

Successively, the 5hmC presence was detected in the DNA of bacteriophages T2, T4 and T6 (Wyatt and Cohen 1953), in the brains of the rat, mouse and frog (Penn et al. 1972), specifically in the Purkinje and granule cell nuclei (Kriaucionis and Heintz 2009). However, its significance remained unclear until the discovery of the activity of TET family enzymes (Ten-eleven translocation) that oxidize 5mC to 5hmC (Tahiliani et al. 2009; Ito et al. 2010), Chapter 1.3.2.1 in Literature review.

5HmC can be present at both promoter and gene bodies and can influence both transcription activation and transcription repression (Wu, D'Alessio, Ito, Wang, et al. 2011). Furthermore, the presence of 5hmC at genes is able to influence the chromatin structure, and therefore the gene expression in this location e.g., via binding of specific 5hmC-binding proteins (Tahiliani et al. 2009). Even though the exact outcome of the 5hmC presence on promoters and gene bodies is still unknown, it can point to the ongoing process of active DNA demethylation.

1.3.1.1 5-Hydroxymethylcytosine detection

It is crucial to distinguish between the presence of 5mC and 5hmC, due to their possible different effects on the gene expression. However, bisulfite sequencing (Chapter 1.2 in the Literature review) is unable to discriminate between 5mC and 5hmC. Both 5mC and 5hmC are converted very slowly in the presence of sodium bisulfite. During the bisulfite treatment, 5hmC is transformed to 5-methylenesulfonate (CMS), whose deamination occurs even more slowly than in the case of 5mC, reviewed in (Schüler and Miller 2012). Therefore, many analyzed DNA methylation patterns can possibly contain 5hmC instead of 5mC inside (Jin et al. 2010). The presence of 5hmC can be detected by TET-assisted oxidative bisulfite

sequencing, oxidative bisulfite sequencing with $KRuO_4$, hairpin oxidative bisulfite sequencing, and by the LC/MS technique (Chapter 1.2 in the Literature review).

TET-assisted oxidative bisulfite sequencing (TAB-Seq) is based on the principle that the analyzed DNA is oxidized by TET enzymes (M. Yu et al. 2012) (Fig. 4). First, the DNA is incubated with β-glucosyltransferase, which protects 5hmC from further oxidation by creating β-glucosyl-5-hydroxymethylcytosine. In a further step, the mouse recombinant TET1 protein is added. The TET1 protein oxidizes 5mC to cytosine and β-glucosyl-5-hydroxymethylcytosine remains unchanged. Next, bisulfite sequencing is employed. The analyzed sequence is then amplified in PCR reaction. TAB-Seq must be performed together with bisulfite sequencing to distinguish between 5mC and 5hmC.

Another used method is oxidative bisulfite sequencing (oxBS-Seq) with $KRuO_4$ (Booth et al. 2012) (Fig. 4). $KRuO_4$ specifically oxidizes 5hmC to 5-formylcytosine (5fC). In the next step, the analyzed sample is treated with sodium bisulfite, which transforms both 5fC and cytosine to uracils. 5mC remaines unchanged. The analyzed sequence is amplified by PCR reaction. OxBS-Seq must be performed together with bisulfite sequencing to compare the sequences and distinguish between 5mC and 5hmC.

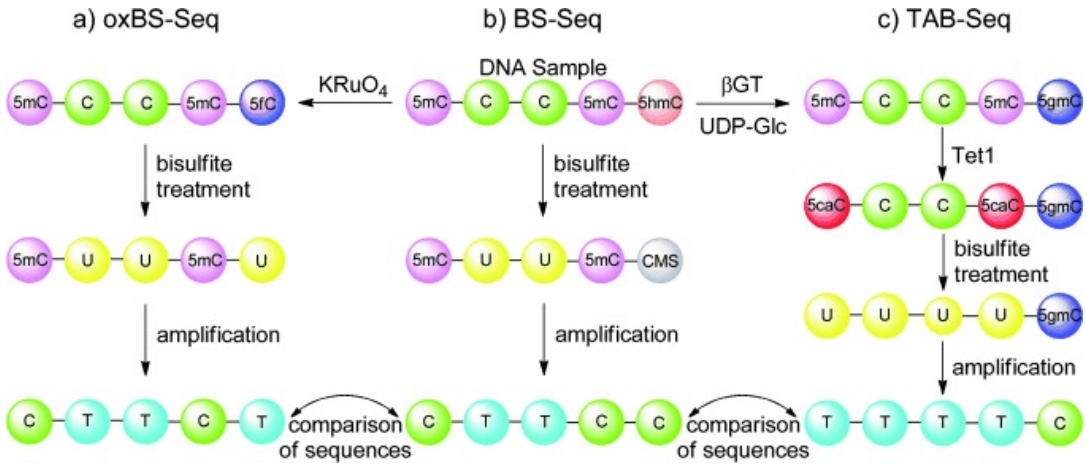

Fig. 4: Schematic representation of oxidative bisulfite sequencing – oxBS-Seq and TAB-Seq. *(a) In oxBS-Seq, the analyzed DNA is treated with $KRuO_4$, which transforms 5hmC to 5fC, while 5mC remains unchanged. In the next step, the analyzed DNA is treated with sodium bisulfite, which converts both 5fC and C to U. 5mC still remains unchanged. The last step is amplification of the analyzed sequence by PCR. All present U are detected as T, 5hmC as C. The obtained sequence is then compared with the bisulfite-treated sequence (b). (c) In TAB-Seq, the analyzed DNA sample is treated with β-glucosyltransferase (βGT), which attaches uridin diphosphate glucose (UDP-glucose) to 5hmC, while β-glucosyl-5-hydroxymethylcytosine (5gmC) is formed. In the next step, recombinant mouse TET1 protein is added. TET1 transforms 5mC to 5-carboxycytosine (5caC), but 5gmC remains unchanged. After that, bisulfite treatment occurs. Both 5caC and C are transformed to U, and 5gmC still remains unchanged. The last step is amplification of the analyzed sequence by PCR. All present U are detected as T, 5gmC (previous 5hmC) as C. The obtained sequence is then compared with bisulfite-treated sequence (b). The figure was taken from (Schüler and Miller 2012).*

The last recently used method is the hairpin oxidative bisulfite sequencing (Giehr et al. 2018) (Fig. 5). The methods described above are strand-specific, while hairpin oxidative bisulfite sequencing allows analysis of DNA methylation and hydroxymethylation on both DNA strands. In general, DNA methylation is symmetrically localized on both strands, but in some cases the strains can have different patterns. In the first step of the harpin oxidative bisulfite sequencing, the analyzed DNA is digested in the close proximity of the CpG of interest. Then a hairpin linker is attached. In next step, OxBS-Seq is performed (described above). The result of the hairpin oxidative bisulfite sequencing is compared with bisulfite sequencing and non-treated DNA, and the exact presence of 5hmC and 5mC is ascertained.

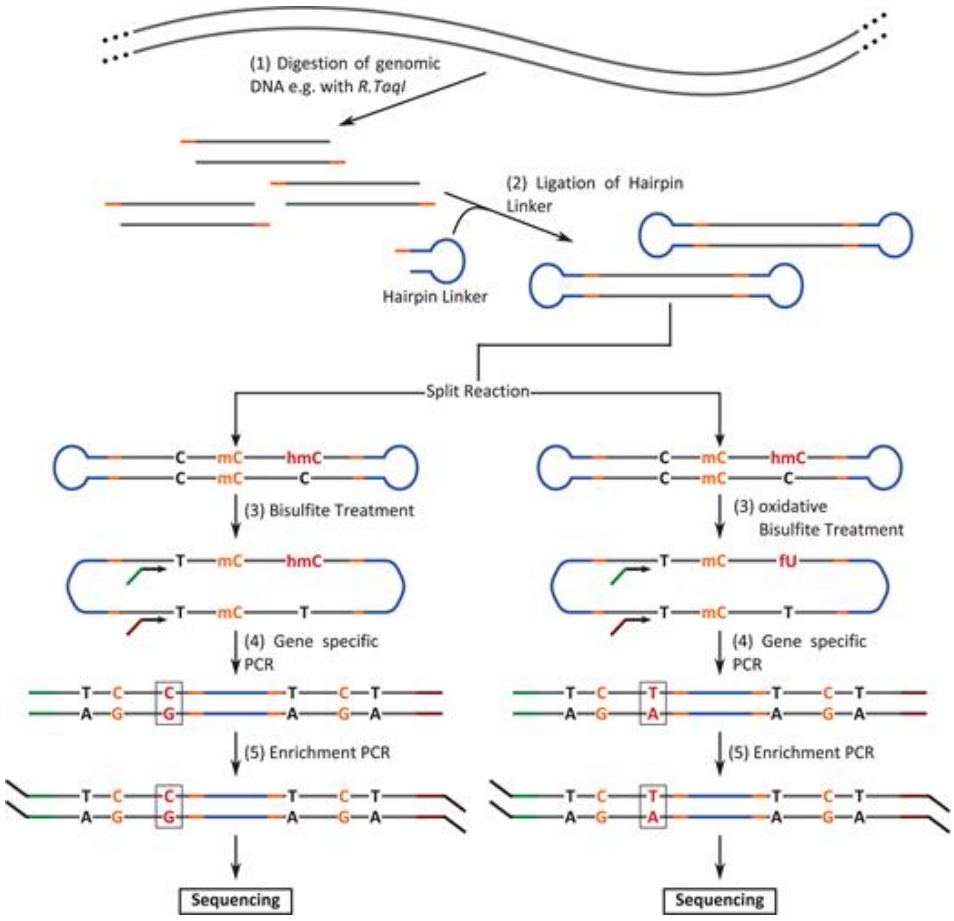

Fig. 5: Schematic representation of harpin oxidative bisulfite sequencing. *(1) DNA is digested with a restriction endonuclease – for example TaqI. (2) The hairpin linker is ligated to the digested DNA. (3) In two parallel reactions, bisulfite and oxidative bisulfite treatment is performed. (4) Gene-specific PCR is done and (5) the locus is then enriched by PCR, sequenced, and compared to each other, and to the non-treated sequence. The figure was taken from (Giehr et al. 2018).*

1.3.2 The process of active DNA demethylation

Active DNA demethylation is mediated enzymatically by the family of TET enzymes (Tahiliani et al. 2009; Ito et al. 2010), Chapter 1.3.2.1 in the Literature review. Nevertheless, there are other possible enzymatic mechanisms of active DNA demethylation.

The family of AID/APOBEC enzymes can de-aminate 5mC to T creating a T/G mismatch. The mismatch can be excised by Mdb4 glycosylase, while the Gadd45 protein promotes interaction between deaminase/glycosylase pairs (Morgan et al. 2004; Rai et al. 2008). The family of AID/APOBEC enzymes can also de-aminate 5hmC to 5-hydroxymethyluracil, which can be converted by TDG/SMUG1 glycosylases and the base excision repair (BER) mechanism to C (Cortellino et al. 2011; J. U. Guo et al. 2011). Moreover, the family of AlkB enzymes is able to oxidize 5mC to 5fC and further to 5-carboxycytosine (5caC) *in vitro* (Bian et al. 2019). 5caC can be converted to C by the TDG/BER mechanism like in TET-mediated DNA demethylation (Chapter 1.3.2.1 in the Literature review). Furthermore, DNMTs were found to be able to demethylate 5mC (C. C. Chen et al. 2013). Nevertheless, the required conditions, e.g. oxidative environment and elevated level of calcium (Ca^{2+}), are not natural (C. C. Chen et al. 2013) and this mechanism is probably not functional *in vivo*.

It is supposed that more proteins will be found to be able to mediate active DNA demethylation in the future. However, it is crucial to distinguish between the mechanisms that are functional *in vivo* and under very specific non-natural conditions *in vitro*.

1.3.2.1 TET enzymes

The family of TET enzymes plays a crucial role in the process of active DNA demethylation. There are three known members of the TET family – TET1, TET2 and TET3. The name Ten-eleven translocation came from the fact that *TET1* was described for the first time as a fusion partner of the myeloid/lymphoid leukemia (MLL) gene in patients with acute myeloid leukemia. The *TET1* gene in this *TET1/MLL* fusion originated from chromosome 10 and was translocated to the *MLL* gene on chromosome 11 (Ono et al. 2002; Lorsback et al. 2003). The process of TET-mediated DNA demethylation is supposed to be the most frequent mechanism of active DNA demethylation, and is therefore broadly studied.

TET enzymes initiate active DNA demethylation. Specifically, TET proteins convert 5mC to 5hmC (Tahiliani et al. 2009; Ito et al. 2010). In the enzymatic change, 5hmC is oxidized to 5fC, which is then decarboxylated to 5caC (Ito et al. 2011). All the steps are

performed by TET proteins. Further, 5caC can be excised by TDG glycosylase and the BER mechanism creates cytosine (He et al. 2011). TDG can be released from the apurinic site by the family of NEIL proteins (Schomacher and Niehrs 2017). Both 5fC and 5caC are substrates for TDG; therefore, they can be converted to C by the TDG-BER mechanism as well (Maiti and Drohat 2011). The process of TET-TDG mediated active DNA demethylation is summarized in Fig. 6. Collectively, TET enzymes initiate active DNA demethylation by converting 5mC to 5caC which is then replaced with C by the cellular DNA repair mechanisms.

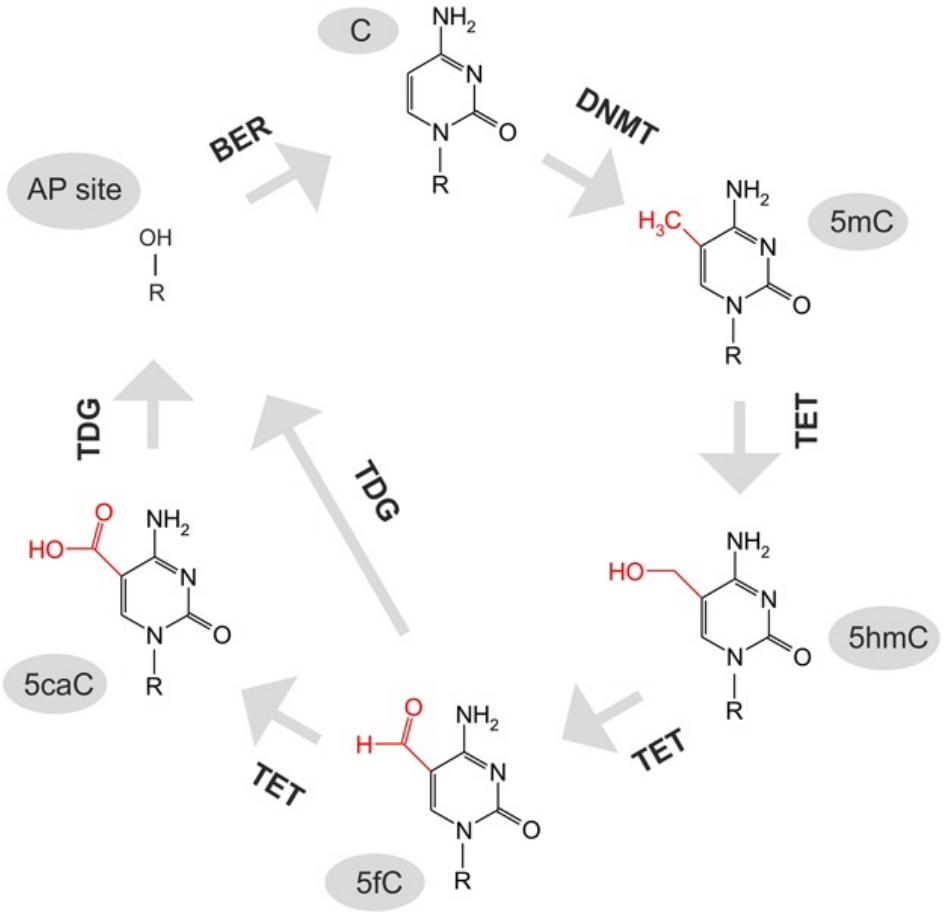

Fig. 6: Schematic representation of TET-TDG-mediated active DNA demethylation. *5mC is converted by the family of TET enzymes into 5hmC and further to 5fC and 5caC. Both 5fC and 5caC can be recognized by TDG, which creates an apurinic site. The apurinic site is converted by the BER repair mechanism to C. Schematic description of the compounds is given. H = hydrogen, N = nitrogen, O = oxygen, R = represents the rest of the compound. The figure was taken from (Schomacher and Niehrs 2017).*

The level of 5mC and the intermediates of the active DNA demethylation process differs. In embryonic stem (ES) cells, the amount of 5hmC is 10 to 100 lower than the amount of 5mC (Tahiliani et al. 2009; Ito et al. 2011) and the level of 5fC and 5caC in the genome is 100 to 1,000 times lower than the amount of 5hmC (Ito et al. 2011). These data point to the

fact, that active DNA demethylation is a rapid process. However, the exact kinetics of the active DNA demethylation process is not precisely understood yet.

The DNA demethylation activity of TET proteins is mediated by their oxygenase domains. For the oxidizing function, two essential cofactors are needed – iron and α-ketoglutarate (Tahiliani et al. 2009). α-Ketoglutarate is produced from isocitrate in the tricarboxylic cycle (TCA) by isocitrate dehydrogenase 1 and 2 (IDH1, IDH2). The presence of specific mutations in *IDH1/IDH2* genes leads to production of 2-hydroxyglutarate instead of α-ketoglutarate (Yan et al. 2009). 2-Hydroxyglutarate serves as a competitive inhibitor to α-ketoglutarate, and its binding to TET proteins blocks the enzymatic activity (Dang et al. 2009; W. Xu et al. 2011) (Fig. 7). Therefore, the enzymatic activity of TET enzymes can be negatively regulated by the presence of specific mutations in *IDH1/IDH2* genes.

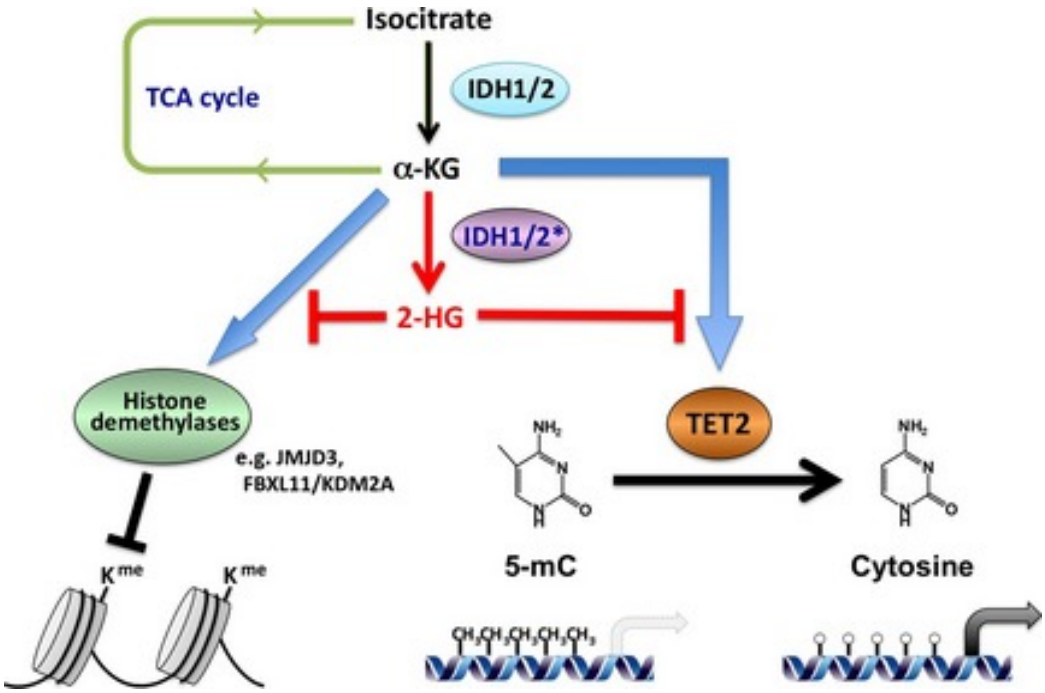

Fig. 7: Schematic representation of the inhibition of TET enzymes activity by mutations in the IDH1/IDH2 genes. *In the TCA cycle, isocitrate is converted to α-ketoglutarate, which serves as a cofactor of TET enzymes in the process of active DNA demethylation, where 5mC is oxidized to 5hmC and further transformed to cytosine. Furthermore, α-ketoglutarate participates in the process of histone demethylation by histone demethylases such as JMJD3 and FBXL11/KDM2A. In the presence of specific mutations in IDH1/IDH2 genes, 2-hydroxyglutarate instead of α-ketoglutarate is produced. 2-hydroxyglutarate serves as a competitive inhibitor to α-ketoglutarate, and blocks the activities of TET enzymes and disrupts the function of some of the histone demethylases. The figure was taken from (Nakajima and Kunimoto 2014).*

The binding of TET1 and TET3 proteins to DNA is performed by the CXXC domain, which preferentially recognize methylated DNA (Zhang et al. 2010; Y. Xu et al. 2012). The

CXXC domain was lost during the evolution in TET2; on the other hand, TET2 interacts with IDAX CXXC protein, which directly binds to unmethylated DNA (Ko et al. 2013). Moreover, TET2 is able to interact with specific transcription factors such as C/EBPα, Klf4 and Tfcp2l1, which recruit the enzyme to the specific locus in the genome (Sardina et al. 2018). It is supposed that TET2 finds the DNA that should be demethylated via the transcription factors. More research should be done in this field to elucidate this hypothesis.

1.3.2.2 Activities of TET proteins and their expression characteristics

The function of the TET enzymes is to perform active DNA demethylation (Tahiliani et al. 2009; Ito et al. 2010; He et al. 2011), especially at promoters, enhancers and other distal regulatory elements (Lu et al. 2014). Even though all three TET enzymes share the same function in general, their expression and activity can differ in various tissues and cell types.

Most of the data known about the TET proteins were obtained from the mouse models. Just a minority of the data are from human tissue samples. It is supposed that the majority of the findings stand both for mice and humans. However, there can be some differences between these two species, and therefore more research in the future is needed.

TET1

The TET1 enzyme has several crucial functions. It is necessary for maintaining the proper level of DNA methylation at imprinted regions (SanMiguel et al. 2018). Next, TET1 is highly expressed in mouse ES cells, where it participates in the lineages specifications (Tahiliani et al. 2009; Dawlaty et al. 2011) and mesenchymal cell lineage determination (Cakouros et al. 2019). Furthermore, TET1 is involved in inner cell mass formation during embryonal development (Ito et al. 2010).

To elucidate the function of the TET1 enzyme in the fetal development, *TET1* knock-out in the mouse model was performed. TET1 depletion in the embryo led to a decreased level of 5hmC in the ES cells. However, this 5hmC reduction did not affect the ES cell pluripotency or viability and development of the embryo (Dawlaty et al. 2011). These findings challenged the previous hypothesis that TET1 is essential for maintaining the pluripotent state of mouse ES cells (Ito et al. 2010; Wu, D'Alessio, Ito, Xia, et al. 2011). Maternal *TET1* knock-out led to smaller ovary size and to a lower amount of produced oocytes (Yamaguchi et al. 2012). This reduction of germ cell number was caused by meiotic defects; therefore, TET1 is one of the meiotic regulators in PGCs (Yamaguchi et al. 2012). Paternal *TET1* knock-out led to

variability in body size, and about half of the paternal knock-out mice died within three days after the birth (Yamaguchi et al. 2013). Moreover, paternal *TET1* knock-out showed hypermethylation of differently methylated regions in the sperm (Yamaguchi et al. 2013). Collectively, TET1 depletion showed that TET1 is not responsible for the active DNA demethylation wave, which physiologically occurs in PGCs during the development (Hajkova et al. 2002; Yamaguchi et al. 2012). Nevertheless, TET1 is a very important regulator in PGCs, ES cells, and during the mouse fetal development.

An altered level of DNA methylation can lead to tumor formation. A reduced level of TET1 expression was observed in colorectal cancers (Kudo et al. 2012), renal carcinoma (Fan et al. 2015), breast cancer (Yang et al. 2015), prostate cancer (Spans et al. 2016), endometrial carcinomas (Ciesielski et al. 2017), and hepatocellular carcinoma (P. Wang et al. 2019). Upregulated TET1 expression was shown in germ cell tumors (Korkola et al. 2005; 2006; 2009). The data imply involvement of altered TET1 expression in the overall level of DNA methylation in the above-mentioned malignancies, but the change of the TET1 expression itself is not sufficient for oncogenic transformation of the cells (Kudo et al. 2012). Therefore, other regulators are involved.

TET2

Similarly to TET1, TET2 is produced in ES cells (Ito et al. 2010; Koh et al. 2011). Furthermore, the activity of TET2 is important for the stability and immune homeostasis of regulatory T-cells (Yue et al. 2019), and TET2 together with TET1 is involved in the mesenchymal cell lineage determination (Cakouros et al. 2019). Next, TET2 is expressed both in human and in mouse myeloid cells. The importance of TET2 in these cells is manifested by the fact that mutations in the *TET2* gene are associated with myeloplastic syndromes (Langemeijer et al. 2009). These data are in line with the results from *TET2* knock-out. *TET2* knock-outs displayed disrupted hematopoiesis and other problems that are characteristic of chronic myelomonocytic leukemia (hepatomegaly, splenomegaly, higher level of white blood cells, and increased bone marrow cellularity) (Li et al. 2011). Moreover, some of the $TET2^{+/-}$ mice displayed lethal myeloid malignancies as well (Li et al. 2011). The data point to the important role of TET2 dioxygenase in myeloid cells.

Altered TET2 expression was observed in various human cancers. Down-regulated TET2 expression was shown in T-cell and B-cell lymphomas (Quivoron et al. 2011), melanoma (Lian et al. 2012), esophageal squamous cell carcinoma (Murata et al. 2015),

glioma (García et al. 2018), and hepatocellular carcinoma (P. Wang et al. 2019). Upregulated TET2 expression was shown medulloblastoma and ependymoma cells (Ramsawhook et al. 2017). Similarly to TET1, the changes in the TET2 expression contribute to tumorigenesis, but other factors are needed for the tumor formation.

TET3

TET3, together with TET2 is involved in maintaining the stability and homeostasis of regulatory T-cells (Yue et al. 2019). Moreover, the TET3 protein, together with TET1 and TET2 is expressed in mouse ES cells (Ito et al. 2010; Koh et al. 2011; Lu et al. 2014). TET3 enzymatically contributes to the total level of 5hmC is ES cells; however, its effect is very slight (about 2 % of all 5hmC) (Lu et al. 2014). This is in agreement with the results of Dawlaty and colleagues, who found out that TET3 is partly able to rescue the phenotype of *TET1/TET2* double knock-out mice (Dawlaty et al. 2014). Further, TET3 is produced in the zygote, where it is essential for the paternal genome DNA demethylation. Demethylation of the paternal DNA is performed by TET3 of maternal origin (Gu et al. 2011). TET3 oxidizes maternal DNA in the zygote as well, but to a lesser extent (Tsukada et al. 2015). *TET3* knock-out mice died after the birth, and therefore the presence of TET3 is essential for embryonal development (Gu et al. 2011). However, the death of new-born mice was caused by the haploinsufficiency of the *TET3* (Inoue et al. 2015). Collectively, TET3 dioxygenase is essential during the embryonal development.

Altered TET3 production was observed in various human cancers. Upregulated in grade 3 renal carcinomas (D. Chen et al. 2017). Down-regulated TET3 expression was detected in glioma (Carella et al. 2019) and head and neck squamous cell carcinoma (Misawa et al. 2018). The low level of TET3 is associated with poorer prognosis in both these malignancies (Misawa et al. 2018; Carella et al. 2019). The data imply that TET3 plays a role in the progress of gliomas and head and neck squamous carcinomas and participate in tumorigenesis of other above-mentioned malignancies.

Cooperation of TET enzymes

There are cellular processes, in which all three members of the TET family cooperate. Beside the collective expression of TET proteins in ES cells (Ito et al. 2010; Koh et al. 2011; Lu et al. 2014), TET enzymes and components of the BER mechanism are active in PGCs (Hajkova et al. 2010; Vincent et al. 2013; Hill et al. 2018), where they are involved in the

active DNA demethylation process. Further, they are important during PGC development. It is supposed that TET1 is the most important TET enzyme during the DNA demethylation process in PGCs, but not the leading one during the initialization process, where two other TETs are probably essential (Hill et al. 2018). Therefore, cooperation of all TET enzymes is needed for proper development of PGCs.

Further insight into the TET enzymes cooperation was provided by *TET1/TET2/TET3*-mouse triple knock-out. This triple knock-out revealed that TET proteins are necessary for maintaining the DNA hypomethylation of bivalent promoters in ES cells (Verma et al. 2018). Bivalent promoters are important for the proper differentiation of the cell during the early stage of embryonal development (Verma et al. 2018). Therefore, the joint activity of the whole family of TET enzymes is needed during the embryonal development.

Collectively, TET enzymes catalyze an essencial cellular process, the active DNA demethylation. Therefore, the disruption of their activity can have serious impacts on the whole organism both during the embryonal development and during the lifetime, when it can lead to formations of various tumors. The altered activity of TET dioxygenases can lead to changes in the level of promoter DNA methylation at various genes, which further affects their transcription. Collectively, even though the sole disrupted activity of TET dioxygenases is not sufficient for the tumor formation, they widely contribute to the tumorigenesis.

The second part of the literature review deals with the germ cell tumors, the tumors that were analyzed in this thesis. The main focus is given to the germ cell tumor formation and differentiation.

2 Testicular germ cell tumors (TGCTs)

Testicular tumors are malignancies of male germ cells, which count for 1 - 2 % of all male neoplasmas, reviewed in (Park et al. 2018). Testicular cancers group together the germ cell tumors (GCTs), the sex-cord gonadal stromal tumors and the secondary testicular tumors (Fig. 8). GCTs counts for the 95% of all testicular tumors, reviewed in (Smiraglia et al. 2002); they divide into seminomas and non-seminomas. Non-seminomas comprise embryonal carcinomas, yolk sac tumors, choriocarcinomas, teratomas and mixed GCTs. Mixed GCTs contain two or more of the above-mentioned components. The classification of testicular tumors is depicted in Fig. 8. From the epidemiological point of view, GCTs can be divided

into prepubertal and postpubertal, while the postpubertal are more common, reviewed in (Netto et al. 2008).

Testicular cancer
- sex-cord gonadal stromal tumors
- secondary testicular tumors
- germ cell tumors (GCTs)
  - seminomas
  - non-seminomas
    - embryonal carcinomas
    - yolk sac tumors
    - teratomas
    - choriocarcinomas
    - mixed germ cell tumors

Fig. 8: Schematic classification of testicular cancers.

GCTs can be treated by cisplatin chemotherapy or by radiation. The treatment selection depends on the GCT type. Specifically, seminomas are sensitive to both radiation and chemotherapy. However, non-seminomas are partly able to repair the DNA damage induced by the radiation. Further, non-seminomas show reduced reaction to the cis-platin treatment, while teratomas do not react to cisplatin (Einhorn 1997; Oosterhuis and Looijenga 2005; Wermann et al. 2010). Therefore, identification of the GCT type is crucial for the adjustment of the proper treatment.

2.1 The formation of GCTs

GCTs originate from PGCs or gonocytes, the precursors of germ cells (Skakkebék et al. 1998; Rajpert-De Meyts 2006). During the fetal development in males, PGCs migrate to the gonad. When they reach the gonad, they are called gonocytes. After the migration, the supporting cells of PGCs differentiate into Sertoli cells. The process is reviewed in (Skakkebék et al. 1998). From the gonocyte, a pre-spermatogonium is formed. The transformation from the gonocyte into spermatogonium is followed by a rapid loss of embryonic characteristics and acquisition of features of a male germ cell such as the expression of male-specific factors. The process is reviewed in (Rajpert-De Meyts 2006). Furthermore, for the gonocyte to spermatogonium maturation, male-specific hormones and factors produced by Leydig and peritubular cells are needed (Kristensen et al. 2013). When the gonocyte differentiation is blocked, germ cell neoplasia *in situ* (GCNIS) is formed (Skakkebék et al. 1998). Moreover, a reduced function of Sertoli cells can contribute to the

blocking of gonocyte differentiation (Skakkebék et al. 1998). The formed GCNIS is dormant for some time. Then, in pre- or postpubertal age, some changes happen, mainly in endocrine secretion of the male-specific hormones, and the dormant GCNIS starts the transformation to GCT (Kristensen et al. 2013). In some cases, the presence of TGCT is a part of the so-called dysgenesis syndrome, which includes cryptochordism and testicular atrophy. Cryptochordism is an inability of the testes to descent during the development. This syndrome is a result of both endogenous and exogenous factors that affect the patient, reviewed in (Oosterhuis and Looijenga 2005). Collectively, the precursor of GCTs – GCNIS – is formed during the perinatal period from a PGC or premeiotic gonocyte that has delayed or blocked maturation (Skakkebék et al. 1998; Oosterhuis and Looijenga 2005; Rajpert-De Meyts 2006; Kristensen et al. 2013).

The presence of chromosomal aberrations and changes in chromosomal ploidity can have a serious impact on the tumor formation and behaviour. Various chromosomal changes were described in GCTs, but the most common is the duplication of the short arm of chromosome 12. This duplication is associated with the invasiveness of GCT (Rosenberg et al. 2000). Furthermore, the genomes can be hypotriploid in non-seminomas and hypertriploid in seminomas, reviewed in (Prabhakar et al. 2017). The above-described chromosomal changes affect gene expression; especially changes in the expression of tumor suppressor genes and oncogenes can have a serious impact on the tumor formation.

2.2 GCT types characteristic features

Directly from GCNIS, a seminoma or embryonal carcinoma is established, while other non-seminomas arise from the embryonal carcinoma by the differentiation process (Oosterhuis and Looijenga 2005). Further on, it is believed that embryonal carcinoma can also be established from a seminoma (Nettersheim et al. 2015). The differentiation of GCTs is a multistep process during which expression from various genes is changing (Nettersheim et al. 2015; Cheung et al. 2016).

Proteins known to be expressed in GCTs are AP-2γ (activator protein-2γ) in GCNIS and seminoma and slightly in embryonal carcinoma (Pauls et al. 2005), DAZL1 (deleted in azoospermia-like 1) mainly in seminoma (Lifschitz-Mercer et al. 2002), PLAP (placental-like alkaline phosphatase) especially in GCNIS and seminoma (Hamilton-Dutoit et al. 1990), CD30 in embryonal carcinoma, reviewed in (Leroy et al. 2002), glypican 3 in yolk sac tumor and choriocarcinoma (Zynger et al. 2006), and p63 in choriocarcinoma and teratoma

(Wegman et al. 2019). The KIT (tyrosine-protein-kinase) protein is mainly expressed in seminoma, but to a lesser extent in non-seminoma as well (Nikolaou et al. 2007), and CCR and CXCR4 especially in seminoma (Y. H. Chen et al. 2019). The factor that is produced in all GCTs is miR-371a-3p, and therefore it is considered as a GCT marker (Dieckmann et al. 2017). GCNIS, seminoma and embryonal carcinoma express factors that are characteristic for the stem cell pluripotency state, such as OCT3/4 (POU5F1), NANOG and SALL4. Teratoma, yolk sac tumor and choriocarcinoma are more differentiated, and therefore do not express the pluripotency proteins (Looijenga et al. 2003; Almstrup et al. 2004; Kehler et al. 2004; Rajpert-De Meyts et al. 2004; Hoei-Hansen et al. 2005; Cao et al. 2009; Gillis et al. 2011; Miettinen et al. 2014). OCT3/4 and NANOG in GCNIS and seminoma allow reprogramming to embryonal carcinoma and further to more differentiated other non-seminomas (Rajpert-De Meyts 2006). Expression of the above-mentioned proteins is used for clinical distinguishing between the GCT types; however, the protein production is not fully specific just for one GCT type. For example, PLAP and KIT are supposed to be seminoma markers, but they are also expressed during the seminoma to embryonal carcinoma transition (Rajpert-De Meyts 2006; Shen et al. 2018). This is a problem for the treatment, because seminomas and non-seminomas are treated in different ways (Chapter 2 in the Literature review). Therefore, specific markers for distinguishing GCT types are needed.

The GCT type differentiation is accompanied by changes in the level of DNA methylation. GCNIS and seminoma share almost an undetectable amount of 5mC and a low level of 5hmC, while non-seminomas contain a higher level of 5mC (Smiraglia et al. 2002; Netto et al. 2008; Nettersheim et al. 2013; Kristensen et al. 2014). It is supposed, that the DNA methylation detected in seminoma came from the infiltrated lymphocytes (Shen et al. 2018). There is a hypothesis that a low level of 5mC in seminoma leads to expression of HERVs, for example ERVWE1 and ERVFRDE1 (Gimenez et al. 2010; Trejbalová et al. 2011). The presence of HERVs induces type I interferon (INF) signaling, and CD8+ T-lymphocytes infiltrate seminoma (Haffner et al. 2018). More research has to be done to prove this hypothesis. Collectively, the differentiation process of GCTs is associated with changes in the level of DNA methylation, specifically from hypomethylated GCNIS and seminoma to more methylated embryonal carcinoma and hypermethylated choriocarcinoma, yolk sac tumor, and teratoma.

As GCT types differ in the level of DNA methylation, the level of TET dioxygenases was analyzed there. There are studies, that showed low levels of both *TET1* and *TET2* in GCTs (Nettersheim et al. 2013; Kristensen et al. 2014). However, expression of *TET* genes,

particularly *TET1*, was also found to be upregulated in GCTs in comparison with the healthy testes (Korkola et al. 2005; 2006; 2009). Altogether, the level of DNA methylation in GCTs is probably influenced by the TET dioxygenases.

GCT types also differ in the presence of histone modifications. In GCNIS, H3K4me1, H3K9ac and histone variant H2A.Z, which are connected with the active chromatin structure, are present globally. Moreover, there are low levels of histone repressive marks such as H3K9me2 and H3K27me3 (Almstrup et al. 2010). The epigenetic characteristic of GCNIS is similar to the gonocyte, which supports the idea that GCNIS arises from a gonocyte (Almstrup et al. 2010). Similarly to GCNIS, seminoma is characterized by a high level of H3K4me1 and H2A.Z, but, on the other hand, there are high levels of H3K9me2 and H3K27me3 in comparison with GCNIS (Smiraglia et al. 2002; Netto et al. 2008). Embryonal carcinoma displays similar types and levels of histone modifications but higher levels of 5mC in comparison with GCNIS (Smiraglia et al. 2002; Netto et al. 2008). The observed histone modifications in seminoma and embryonal carcinoma correspond with their less-differentiated states.

Gene expression in GCTs is regulated by the interplay of DNA methylation, histone modifications and regulatory proteins. Collectively, the level of DNA methylation and the expression of various genes differ between GCT types and furthermore, the level of DNA methylation is in accordance with the state of differentiation of the tumor. However, the factors for unambiguous discrimination between the GCT types are not known. Therefore, we aimed to analyze the epigenetic features of different GCT types to find the factors distinguishing the GCT types. The low level of DNA methylation in some GCT types allows expression from the genes that are silenced under physiological conditions, like endogenous retroviruses. Therefore, our second objective was to specify the expression patterns of endogenous retroviruses, namely *ERVWE1* and *ERVFRDE1*, in GCTs.

The third part of the literature review is focused primarily on Syncytin-1, an envelope glycoprotein of retroviral origin, the fusogenic activity of which is essential for human placenta formation. Attention is concentrated especially on the regulation of Syncytin-1 expression.

3 Human endogenous retroviruses (HERVs)

Approximately 8 % of the human genome is represented by sequences of retroviral origin (Lander et al. 2001). Endogenous retroviruses originate either from individual infections by exogenous retroviruses or from reinfection by retroviruses that are present in the genome and are able to replicate. As soon as the retrovirus infects the germ cell and integrates into the genome, it can be vertically transmitted into the progeny in Mendelian fashion and becomes an endogenous retrovirus. This process is called endogenization (Fig. 9). At present, the ongoing process of endogenization can be studied in koala *(Phascolarctos cin*ereus) and mule deer (*Odocoileus hemionus*) (Hanger et al. 2000; Elleder et al. 2012).

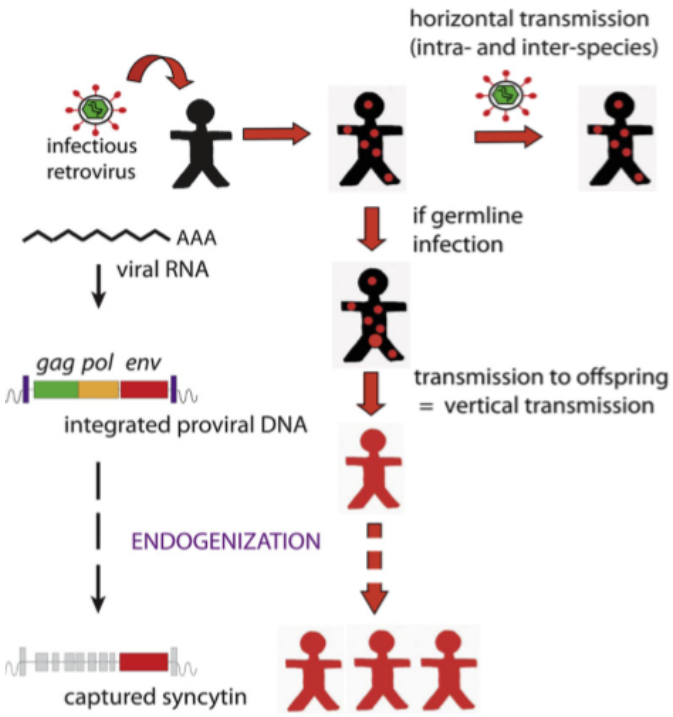

Fig. 9: Schematic representation of the endogenization process. *A human individual is infected by exogenous retrovirus. After the infection, the provirus is integrated into the human genome. The retroviral provirus contains three genes – gag, pol and env. If the proviral DNA is integrated in to the germ cell, the virus can be vertically transmitted to the offspring in Mendelian fashion. In the next generations, the provirus (for example syncytin) can be fixed and inherited within the genome of all cells. The whole process is called endogenization. The figure was taken from (Dupressoir et al. 2012).*

Historically, the classification of human endogenous retroviruses is based on the sequence and on the type of the amino acid that is used as a primer for reverse transcription, reviewed in (Jern et al. 2005; Jern and Coffin 2008). For example ERVWE1 endogenous retrovirus is a member of the HERV-W family (Blond et al. 2000). This means that the replication-competent ERVWE1 had used tryptophan (W) as a primer for reverse

transcription. However, this nomenclature does not reflect other HERVs characteristic features or their phylogenic development. The most presice division is based on the comparison of the *pol* genes, which provides information about the phylogenetic deveplopment (Jern et al. 2005). In another division, HERVs are grouped according to their whole sequences. However, this division is complicated by the fact that there are further so called non-canonical or mosaic HERVs. Grouping of these HERVs that came from secondary integrations or recombination events is difficult (Vargiu et al. 2016). Althought the historical division is not very precise, it is still used and for that reason is employed in this disertation thesis.

Integration of endogenous retroviruses shapes the host genome and can be potentially pathogenic. The outcome depends on the integration site. The presence of integrated provirus can influence transcription from adjacent host genes. This is caused by proviral long terminal repeats (LTRs), which can function as promoters, enhancers, repressors, polyA signals, or alternative splicing sites, reviewed in (Grandi and Tramontano 2018). Further, the presence of HERVs in human genome can potentially induce or contribute to a tumor formation. Mechanisms of potential HERV-induced tumorogenesis are following: (1) regulation of the host genes expression by LTRs, (2) coding of potentially oncogenic proteins, such as Rec and Np9 (HERV-K family), (3) chromosomal translocations induced by recombination events between HERVs and (4) immunosupressive activity of Env protein (see Chapter 3.1.2). These mechanisms could contribute to tumor formation and its spreading by suppressing of the immune system of the host organism. The potential HERV-induced tumorogenesis mechanisms are reviewed in (Gonzalez-Cao et al. 2016). Transcription from HERVs was detected in various human cancers; HERV-K was detected in breast cancer (Wang-Johanning et al. 2008; Contreras-Galindo et al. 2008), ovarian cancer (Wang-Johanning et al. 2007), lymphoma (Contreras-Galindo et al. 2008), teratocarcinoma (Löwer et al. 1993), and in melanomas, where HERV-K level correlates with dissease prognosis (Hahn et al. 2008; Krishnamurthy et al. 2015). Further, HERV-E was found in ovarian cancer (Wang-Johanning et al. 2007), Hippel-Lindau-deficient renal carcinoma (Cherkasova et al. 2016), HERV-H was detected in teratocarcinoma (Löwer et al. 1993), and HERV-W was found in breast carcinoma (Schön et al. 2001), cutaneous T-cell lymphoma (Maliniemi et al. 2013), acute myeloid leukemia (Sun et al. 2016), and testicular cancers (Gimenez et al. 2010; Trejbalová et al. 2011). Whether HERVs, except for HERV-K, have some function in the malignancies mentioned-above should be examined in the future. Due to the fact that HERVs transcription (especially HERV-K) can activate both adaptive and innate immunity, they are supposed to be

a promising target for immunotherapy (Hahn et al. 2008; Wang-Johanning et al. 2008; Krishnamurthy et al. 2015; Cherkasova et al. 2016; Bannert et al. 2018).

Usually, as the endogenous retrovirus becomes an integral part of the host genome after the integration event, the lack of selection pressure keeps the proviral open reading frames (ORFs) intact. Therefore, the majority of endogenous retroviral genes contain mutations and STOP-codons that block viral replication. Nevertheless, in some cases, the presence of the viral protein can be advantageous for the host organism, and therefore the ORF remains functional. This is the case of human endogenous retroviruses ERVWE1 and ERVFRDE1. They both express fusogenic envelope glycoproteins in human placenta, which are essential for proper placenta development (Mi et al. 2000; Blaise et al. 2003).

3.1 ERVWE1 endogenous retrovirus

Human endogenous retrovirus ERVWE1 belongs to the D type of retroviruses and is a member of the HERV-W family (Blond et al. 1999). Most of the HERV-Ws (if not all) are replication defective (Patience et al. 1997), but some their genes can be transcribed. However, only few of them can be succesfully translated into functional proteins. This highlights the extraordinary characteristics of ERVWE1 endogenous retrovirus.

3.1.1 *ERVWE1* locus structure and its chromosomal surroundings

ERVWE1 is situated at chromosome 7, in position 7q21.2 (Blond et al. 1999) and has a classical proviral structure. There are two LTRs and three genes – *gag*, *pol* and *env* (Fig. 10). *Gag* codes for the capsid protein and *pol* for polymerase. These two genes contain mutations and STOP-codons that cause frameshifts and premature termination codons without any protein production (Blond et al. 1999). On the other hand, the *env* gene, which codes for the envelope glycoprotein, maintains intact ORF (Blond et al. 1999). Therefore, under appropriate conditions, the fusogenic Env protein (called Syncytin-1) can be translated (Blond et al. 1999).

Analysis of the *ERVWE1* locus polymorphism in different human individuals revealed that the most preserved parts are the 5´LTR and e*nv* ORF (Mallet et al. 2004; de Parseval et al. 2005). In contrast, 3´LTR showed the same level of polymorphism as is characteristic of other human repeated sequences (Mallet et al. 2004). These findings point out the preservation of the *ERVWE1* locus in humans, especially the 5´LTR and *env* ORF, which are important for

functional Syncytin-1 protein production.

ERVWE1 is not present only in the human genome. The ERVWE1 provirus integrated into the germ line of Catarrhini (Old World Monkeys and Hominoids) ancestor 25 - 40 million years ago (Voisset et al. 1999; Bonnaud et al. 2004), probably within the region already enriched in LTR-elements. The *ERVWE1* locus surroundings contains several elements of retroviral origin – *MaLR* (mammalian apparent-LTR retrotransposon), *LTR9b* and *HERV-H* (Fig. 10). Comparative analysis of the *HERVH-ERVWE1* locus in humans and primates suggested that the *HERV-H* locus had integrated earlier (Bonnaud et al. 2005). Thereafter, *ERVWE1* had integrated into the current proximity of the *HERV-H* locus (Bonnaud et al. 2005). The presence of the adjacent retroviral elements can probably play a role in the regulation of *syncytin-1* expression. Specifically, *MaLR* contains TSE (trophoblast specific enhancer), which is crucial for *ERVWE1* transcription regulation (Chapter 3.1.4 in the Literature review).

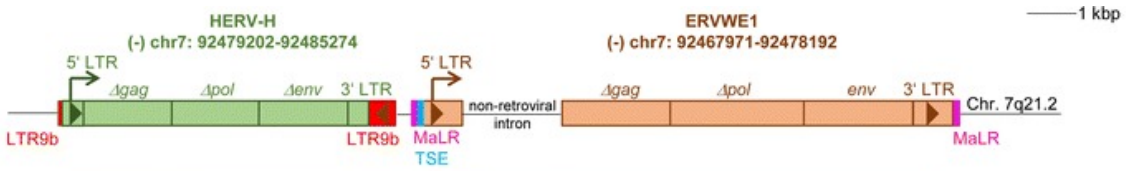

Fig. 10: Schematic representation of ERVWE1 and its chromosomal surroundings. *ERVWE1 is composed of three genes – gag, pol and env, 5´ and 3´ LTR. Between the 5´ LTR and the gag gene, a sequence of non-retroviral origin is depicted. ERVWE1 integrated into the MaLR retroelement. A part of MaLR is represented by TSE, which is important for ERVWE1 expression regulation. Upstream of the ERVWE1, HERV-H is depicted. HERV-H integrated into the LTR9b retroelement. HERV-H has a classical proviral structure with 3´and 5´ LTR, and gag, pol and env genes. (Benešová, Trejbalová, Kovářová, et al. 2017)*

Even though the ancestor *HERVH-ERVWE1* locus was present in the Catarrhini (Bonnaud et al. 2005), the subsequent evolution was different. In the Old world monkeys (e.g., baboon, macaque), a large deletion in the *HERVH-ERVWE1* locus leading to the defective proviral structure occurred. Furthermore, *ERVWE1* in the Old world monkeys contains inactivating mutations in the *env* gene (Bonnaud et al. 2005). On the other hand, in Hominoids (e.g., orangutan, gorilla, human) the ERVWE1 *env* gene remained functional and the Env protein production is preserved (Bonnaud et al. 2004). It seems that *ERVWE1* in Hominoids followed the domestication evolutionary pathway, where possibly *MaLR* could play a role (Bonnaud et al. 2005).

Hence, the exogenous retrovirus turned into a part of the hominoid genome and the

retroviral *env* gene became a bona fide gene involved in cell-cell fusion during placental development (Mallet et al. 2004). The fusogenic ability of Syncytin-1 depends on its expression, protein structure, and on the interaction with the specific receptor.

3.1.2 Syncytin-1, the Env protein

Syncytin-1 protein maturation does not differ from the maturation of Env proteins of infectious retroviruses. Syncytin-1 consists of 584 aminoacids (Blond et al. 1999) and is synthetized as a glycosylated gPr73 protein precursor, which is then cleaved into two subunits – gp50 and gp24 (Cheynet et al. 2005). gp50 serves as a surface subunit (SU) and gp24 as a transmembrane subunit (TM) (Cheynet et al. 2005) (Fig. 11A). SU and TM subunits stay associated by disulphide bonds and the entire Env functions as a homotrimer (Cheynet et al. 2005). SU is essential for the interaction between Syncytin-1 and its specific receptor. TM is needed for the anchorage of the Syncytin-1 protein into the cell membrane (Blond et al. 1999). The TM subunit further contains domains responsible for the fusogenic activity, e.g., the fusion peptide (Blond et al. 1999; Cheynet et al. 2005; Mangeney et al. 2007). The fusion peptide is activated by the SU-TM conformational change induced by binding to the specific receptor (Bonnaud et al. 2004). The activated fusion peptide is inserted into the cytoplasmic membrane of the target cell and initiates the process leading to cell-cell fusion (Cheynet et al. 2005; C.-P. Chen et al. 2008). The result of Syncytin-1-mediated cell-cell fusion is merging of cellular cytoplasms, whereas nuclei stay separated. Multinucleated syncytia having up to 300 nuclei can be produced in this way (Keryer et al. 1998; Blond et al. 2000; Mi et al. 2000; Frendo et al. 2003) (Fig. 11B).

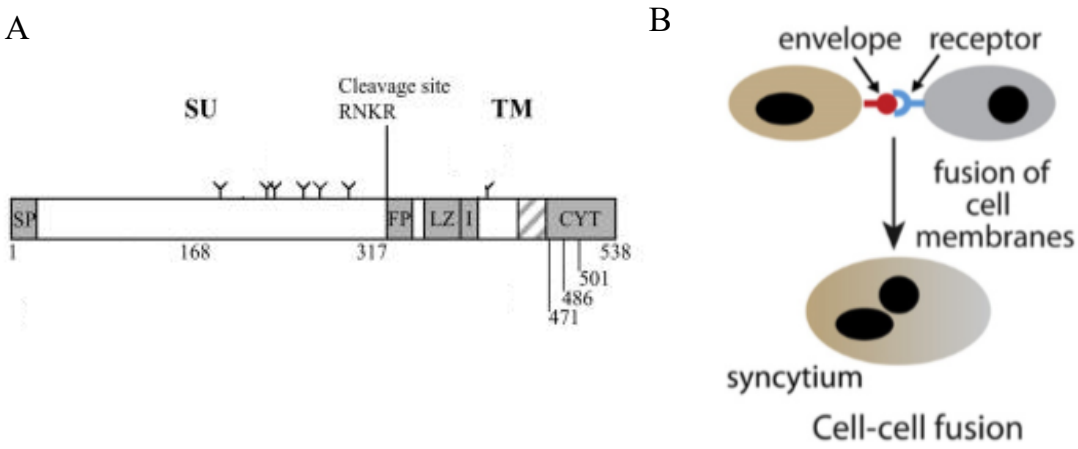

Fig. 11: Schematic representation of Syncytin-1 protein composition (A) and Syncytin-1-induced cell-cell fusion (B). *(A) Syncytin-1 protein is composed of SU (1 – 317 AA) and TM (317 – 538 AA) subunits. Between the two subunits, there is a cleavage site with a characteristic sequence of aminoacids – RNKR (arginine-asparagine-lysine-arginine). The site is recognized by cellular furin-like protease. Y represents putative N-glycosylation sites. Syncytin-1 protein further contains SP (= signal peptide), FP (= fusion peptide), LZ (= leucine zipper, needed for oligomerization), I (= immunosuppresive domain), CYT (= intracytoplasmic tail) and membrane anchorage domain (depicted by the hatched box). The figure was taken and adjusted from (Cheynet et al. 2005). (B) Syncytin-1-induced syncytium formation. The figure was taken from (Dupressoir et al. 2012).*

Similarly to other retroviral Env glycoproteins, the TM subunit of Syncytin-1 further contains an immunosuppressive domain (ISD). Generally, the ISD is characterized by immunosupressive activity (Cianciolo et al. 1985). However, Syncytin-1 ISD contains inactivating mutations that abolish the immunosuppressive activity (Mangeney et al. 2007).

To mediate cell-cell fusion, Syncytin-1 requires interaction with a specific receptor. Two Syncytin-1 receptors have been identified, ASCT2 (SLC1A5) and ASCT1 (SLC1A4). Interestingly, these transmembrane proteins serve as receptors for multiple exogenous and endogenous retroviruses (Marin et al. 2000). ASCT2 and ASCT1 are neutral amino acid transporters (Tailor et al. 1999; Marin et al. 2000) and are expressed in a wide variety of human tissues (Kekuda et al. 1996; Tailor et al. 1999). Therefore, tissue-specific Syncytin-1 expression determines where and when the fusion occurs.

3.1.3 Syncytin-1 expression

3.1.3.1 Syncytin-1 expression and function in the placenta

Under physiological conditions, Syncytin-1 is produced in the placenta, where it mediates fusion of cytotrophoblasts leading to the syncytiotrophoblast layer formation (Mi et al. 2000).

Trophoblasts are the cells forming the outer layer of a blastocyst. Later on, the

trophoblasts establish a large part of the placenta. There are three types of trophoblasts: (i) villous cytotrophoblasts, (ii) syncytiotrophoblast, and (iii) extravillous trophoblasts (Fig. 12). Extravillous trophoblasts are derived from the cytotrophoblasts that cluster in the tips of the fetal villi. They are in direct contact with the maternal uterus and are able to transform maternal spinal arteries to supply the placenta by maternal blood. Villous cytotrophoblasts constitute primary chorionic villi. They can be found as a mononucleated layer under the syncytiotrophoblast. On the contrary, the outer syncytiotrophoblast layer is multinucleated. Syncytiotrophoblast is formed by fusion of the villous cytotrophoblasts (Aplin 1991). Syncytiotrophoblast is in direct contact with the maternal blood and is involved in the implantation of the embryo into the maternal uterus, in the feto-maternal exchange and in the production of pregnancy-specific hormones. Therefore, all three types of trophoblasts are important both for the placenta and fetus formation. Trophoblast types characteristics are reviewed in (Gamage et al. 2016). Syncytin-1 is expressed in all three types of trophoblasts throughout their proliferation and differentiation (Mi et al. 2000; Frendo et al. 2003; Malassiné et al. 2005; Handwerger 2010; Huang et al. 2013); therefore, Syncytin-1 is important not only for placental syncytiotrophoblast formation, but also for its homeostasis. Syncytin-1 can be considered as a marker of the human trophoblast cells (Malassiné et al. 2005).

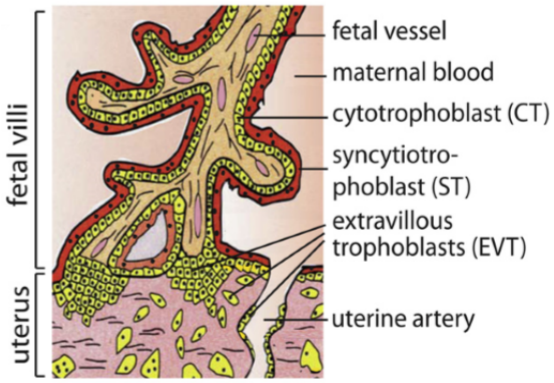

Fig. 12: Schematic representation of the human placenta villus. *There are three types of trophoblast cells: cytotrophoblasts (CT), syncytiotrophoblast (ST) and extravillous trophoblasts (EVT). Extravillous trophoblasts are located in the maternal uterus and around the uterine artery, where they transform arteries to enhance placenta supply by maternal blood. Cytotrophoblasts form a mononucleated layer under the multinucleated syncytiotrophoblast layer. Syncytiotrophoblast is directly washed by maternal blood and gas and nutrient exchange occurs there. The figure was taken and adjusted from (Dupressoir et al. 2012).*

As Syncytin-1 is needed for proper placenta formation (Mi et al. 2000), disrupted Syncytin-1 expression can be problematic. Reduced *syncytin-1* mRNA production was observed in preeclamptic and HELLP (Hemolysis, Elevated liver enzymes, Low platelets)

placentas (Knerr et al. 2002). These pregnancy complications are dangerous for both mother and the fetus. Preeclampsia occurs approximately in 6 % of all pregnancies and HELLP syndrome counts for 10 – 14 % of preeclamptic pregnancies, reviewed in (Ruebner et al. 2013). Preeclampsia is a pregnancy disorder that is characterized by high blood pressure and significant amount of protein in the urine. The HELLP syndrome is a serious preeclampsia associated with hemolysis, trombocytopemia and pathologic liver tests. Decreased placental cell-cell fusion in preeclampsia and HELLP syndrome is connected with Intrauterine Growth Restriction (IUGR) (Ruebner et al. 2013). Even though Syncytin-1 production probably plays a role in these pregnancy complications, it is one of many other factors that together cause these syndromes.

Most of the research in the Syncytin-1 field is done on the BeWo cell line, which originate from choriocarcinoma and sharing some properties with the placenta – such as regions with multinucleated syncytium, upregulated *syncytin-1* expression and elevated production of human chorionic gonadotrophin (hCG) hormone (Pattillo et al. 1968; Mi et al. 2000). Furthermore, *syncytin-1* production was detected in JEG-3, JAR and TCam-2 cell lines. In contrast to BeWo cells, these three cell lines display smaller syncytia. The problematic question of Syncytin-1 research are antibodies. Commercial antibodies for Syncytin-1 protein are not fully specific, because they recognize not only Syncytin-1 but also other envelope glycoproteins from the HERV-W family (unpublished observations from our laboratory, (Sun et al. 2016)). Therefore, specific antibodies for the Syncytin-1 protein should be prepared in the future.

3.1.3.2 Syncytin-1 expression in non-placental tissues

Besides the placenta, expression from the *ERVWE1* locus was reported in the testes. The expression in the testes in comparison with placenta was weak, but still detectable (Mi et al. 2000). However, under normal conditions, Syncytin-1 expression in non-placental tissues should be restricted because its fusogenic ability can disrupt the tissue integrity.

Under pathological conditions, Syncytin-1 production can occur in tissues where it is normally restricted. Expression from the *ERVWE1* locus was observed in malignancies such as testicular cancers (Gimenez et al. 2010; Trejbalová et al. 2011), breast cancer (Bjerregaard et al. 2006), colorectal carcinoma (Larsen et al. 2009), in brains of patients with multiple sclerosis (Antony et al. 2004) and in endometrial carcinomas (Strick et al. 2007). Syncytin-1 involvement in these malignancies has not been fully proved to date.

3.1.3.3 Syncytin-1 expression in physiologically fusing cells in humans

Physiological cell fusions occur in several human cell types. The presence of both Syncytin-1 and its receptor ASCT2 was observed in human gametes; Syncytin-1 is present on spermatozoa, the ASCT-2 receptor on oocytes and on spermatozoa as well. The interaction between Syncytin-1 and ASCT-2 receptor can potentially contribute to fusion between human gametes (Bjerregaard et al. 2014). Moreover, Syncytin-1 and ASCT-2 are supposed to be involved in myoblast fusion (Bjerregard et al. 2014). However, only the presence of Syncytin-1 and its receptor on human gametes and muscle cells is not an evidence for the activity of viral fusogenic proteins in those cells.

Furthermore, Syncytin-1 and ASCT-2 are presumed to participate in the osteoclast fusion (Søe et al. 2011). Syncytin-1 knock-down in osteoclasts revealed untouched fusion of mononucleated osteoclasts, but reduced fusion of multinucleated osteoclasts (Møller et al. 2017). Therefore, Syncytin-1 can be involved in the osteoclast fusion, but the situation is distinct from the trophoblasts fusion, where Syncytin-1 is essential.

3.1.4 Regulation of Syncytin-1 expression

Syncytin-1 production must be strictly regulated, because Syncytin-1-mediated cell-cell fusion in non-placental tissues can disrupt the tissue integrity and functioning. Known regulators of Syncytin-1 expression are specific transcription factors, mRNA splicing, and DNA methylation and histone modifications present on the *ERVWE1* promoter. Probably, other regulators are also involved.

3.1.4.1 Syncytin-1 regulation by transcription factors

Transcription from the *ERVWE1* gene is regulated by the presence of specific transcription factors that bind to the promoter situated in the 5′ LTR. Analysis of the *ERVWE1* locus revealed that the region from +1 to +125 contains placenta-specific binding sites; the region from +125 to +310 is the core of the promoter and is active in every cell type (Prudhomme et al. 2004). Another *ERVWE1* regulator is TSE, which is localized upstream from the *ERVWE1* locus within the *MaLR* region (Chapter 3.1.1 in the Literature review). This 436-bp long element functions as *ERVWE1* enhancer in the placenta, contains binding sites for ubiquitous transcription factors such as AP-2 and Sp-1, and also for trophoblast-specific factor GCM1 (Glial cell missing 1, Chapter GCM1 transcription factor) (Prudhomme

et al. 2004). These data together show that the regulation of *ERVWE1* expression is bipartite – the regulatory elements are 5´ LTR and TSE (Prudhomme et al. 2004).

Both regulatory elements, the *ERVWE1* promoter and TSE, are targets of the signalling pathways that lead to *syncytin-1* expression.

Signaling pathways leading to syncytin-1 transcription

The earliest evidence of the pregnancy beginning is expression of the hCG hormone, which is produced by the placenta. The level of hCG correlates with *syncytin-1* expression (Frendo et al. 2003) and the hCG presence serves as a marker of Syncytin-1 production and syncytiotrophoblast formation (Shi et al. 1993; Frendo et al. 2003). Therefore, the known signaling pathways leading to *syncytin-1* expression begin with hCG production.

While the hCG hormone is present, it binds to the LHCGR (lutheinizing hormone/gonadtrophine receptor) receptor, a trimeric G-protein that activates adenylyl cyclase (Fig. 13). Adenylyl cyclase produces cyclic adenosine monophosphate (cAMP) and stimulates protein kinase A (PKA). Both the presence of cAMP and the activation of PKA lead to *syncytin-1* transcription via GCM1 transcription factor (Keryer et al. 1998; Knerr et al. 2005). Specifically, PKA activates transcription factors such as CREM (cAMP-responsive element modulator) and CREB (cAMP-responsive element binding protein), which have binding sites on the GCM1 promoter (Knerr et al. 2005). Moreover, there is a positive feedback loop between GCM1 expression and hCG production (Cheong et al. 2015) – an upregulated level of GCM1 protein leads to further hCG production. Therefore, GCM1 and hCG stimulates production of each other.

PKA also can be activated by forskolin *in vitro*. Forskolin is a diterpene obtained from the flower *Plectranthus barbatus* and is able to stimulate cAMP production, which leads to *syncytin-1* expression as well (Knöfler et al. 1999).

Moreover hCG, and thus activated PKA, are able to increase the cellular level of Wnt10b protein, which can induce GCM1 production via activation of β-catenin (Malhotra et al. 2017) (Fig. 13). The Wnt/β-catenin system is a well known signaling pathway, which is active in various cellular processes. The Wnt/β-catenin pathway, which leads to GCM1 production and therefore *syncytin-1* expression is called the canonical pathway. Wnt10b production can be triggered by hCG or synthetically by forskolin *in vitro*. The presence of hCG leads to cAMP production, PKA activation, and to Wnt10b upregulation (Malhotra et al. 2015; 2017). Wnt10b ligand binding on its receptor leads to stabilization of β-catenin, which thereafter interacts with BCL9L (B-cell lymphoma 9-like protein) and TCF (T-cell factor)

proteins. Activated transcription factors bind to the two TCF-binding sites on the GCM1 gene and stimulate its expression (Matsuura et al. 2011), which then leads to *syncytin-1* expression.

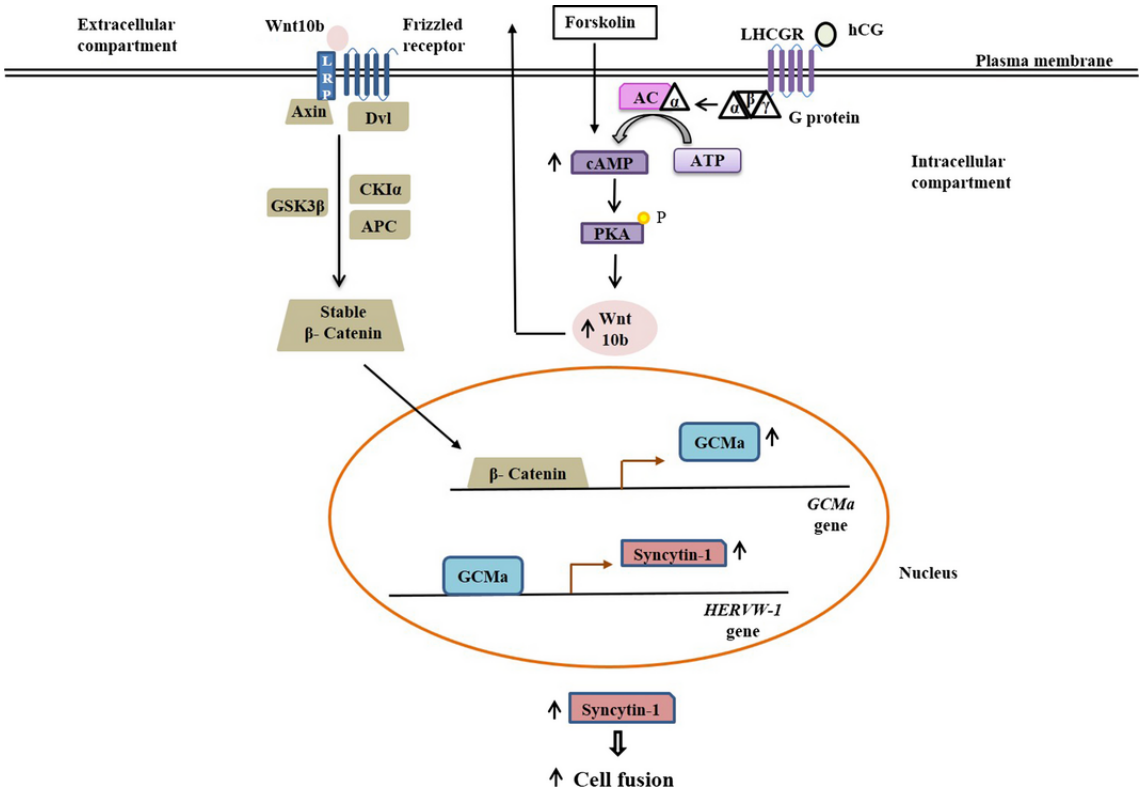

Fig. 13: Summarization of the signaling pathways that lead to syncytin-1 expression. *The canonical β-catenin/Wnt signaling pathway that leads to Syncytin-1-induced cell-cell fusion is depicted in the figure. If β-catenin is not stabilized, it is degraded in the proteasome. The degradation complex of non-stabilized β-catenin consists of APC (adenomatous polyposis coli), axin, CKIα (casein kinase Iα) and GSK-3β (glycogen synthase kinase 3β). When Wnt10b ligand binds to Frizzled receptor, Dvl (disheveled) protein inhibits formation of the degradation complex. Stabilized β-catenin shifts to the nucleus and increases GCMa (= GCM1) expression and subsequently Syncytin-1 production. The Wnt10b ligand level can be upregulated by the cAMP/PKA pathway as well. This cAMP/PKA pathway can be activated also by the upregulated level of hCG or synthetically by forskolin. When hCG is produced, it binds to LHCGR, a trimeric G-protein. After receptor activation, AC (adenylyl cyclase) is activated and produces cAMP with utilization of ATP (adenosine triphosphate). The presence of cAMP leads to PKA activation by phosphorylation. The figure was taken from (Malhotra et al. 2017), information about the signalling pathway was summarized from (Knöfler and Pollheimer 2013; Malhotra et al. 2017).*

GCM1 transcription factor

GCM1, also called GCMa, is a placenta-specific transcription factor (Nait-Oumesmar et al. 2000). GCM1 was found in Drosophila, where it serves as a binary factor that switches between neuronal and glial cell deter,mination (Hosoya et al. 1995; Jones et al. 1995). In humans, GCM1 protein positively influences Syncytin-1 production (C. Yu et al. 2002). *ERVWE1* TSE contains two binding sites for this transcription factor (C. Yu et al. 2002; Prudhomme et al. 2004). Therefore, transactivation of *ERVWE1* promoter is GCM1 dependent

(C. Yu et al. 2002).

GCM1 itself is an unstable protein (Nait-Oumesmar et al. 2000) and posttranslational modifications are needed to increase its stability. CBP (CREB-binding protein) acetylates GCM1 at the transactivation domain, specifically at lysines 367, 406 and 409 (Chang et al. 2005). Moreover, the presence of cAMP activates CaMKI (calcium/calmodulin-dependent protein kinase 1) through Epac1 (exchange protein directly activated by cAMP 1) and Rap1 (ras-related protein 1), which mediates GCM1 phosphorylation at serine 47 (Chang et al. 2011). Thereafter, this phosphorylation leads to GCM1 stabilization by desumoylation (Chang et al. 2011). Further, DUSP23 (dual-specifity phosphatase 23) dephosphorylates GCM1 at serine 23, which prolongs the GCM1 half-life (F. Y. Lin et al. 2011). This dephosphorylation is necessary for subsequent acetylation by CBP (F. Y. Lin et al. 2011). Besides, SENP1 (sentrin-specific protease 1) desumoylation is required for GCM1 stabilization and activation (Chang et al. 2011). The cascade of GCM1 posttranslational modifications that are needed for GCM1 activation and stabilization are summarized in Fig. 14.

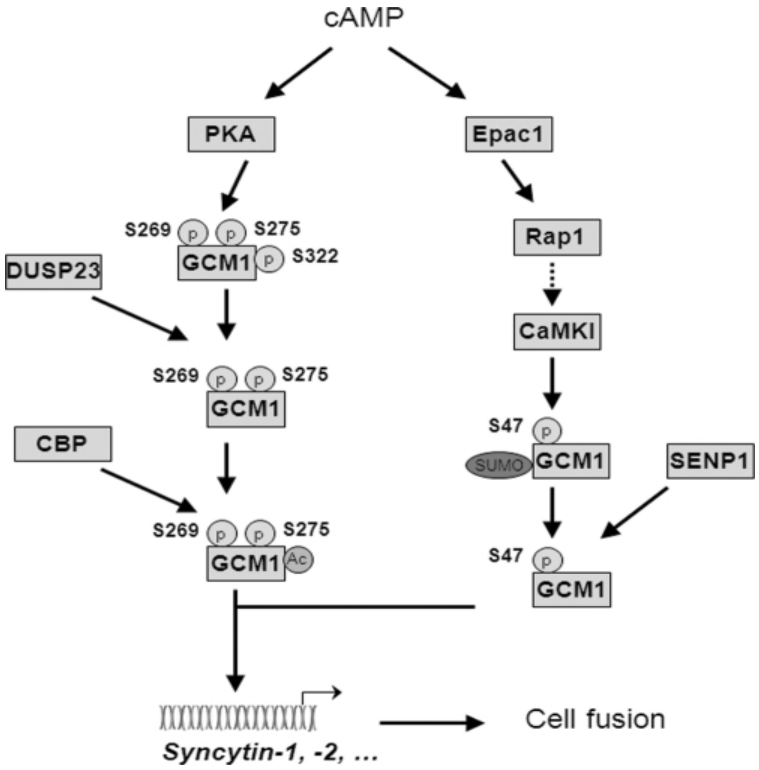

Fig. 14: Schematic representation of GCM1 transcription factor modifications. *The presence of cAMP activates both PKA and Epac1. PKA mediates GCM1 phophorylation at serines 269, 275 and 322. DUSP23 dephosphorylates serine 322 and CBP acetylates GCM1. Epac1 activates CAMKI though Rap1, which leads to GCM1 sumoylation and serine 47 phosphorylation. GCM1 is then desumoylated by SENP1. Active and stabilized GCM1 induces syncytin production and thereafter syncytin-mediated cell-cell fusion. It is supposed that GCM1 affects both Syncytin-1 and Syncytin-2 production. The figure was taken and adjusted from (Chang et al. 2011).*

The activity of GCM1 transcription factor is further regulated by the GLI2 (glioma-associated oncogene 2) protein (C. Tang et al. 2016), which is part of the Hedgehog signalling pathway. GLI2 binds to GCM1 and forms a heterodimer, which stabilizes the GCM1 protein and therefore GCM1 is able to fulfil its function (C. Tang et al. 2016). It is supposed that other GCM1 regulators will be found in the future.

Briefly, GCM1 is a placenta-specific transcription factor, that stimulates *syncytin-1* transcription. GCM1 expression is stimulated by hCG, and the GCM1 protein is further stabilized by posttranslational modifications accomplished by several factors.

The presence of specific cellular transcription factors can influence *syncytin-1* expression. In addition, transcription from *ERVWE1* is regulated by epigenetic modifications.

3.1.4.2 Syncytin-1 regulation by epigenetic modifications

DNA methylation and histone modifications are known regulators of gene expression. These regulatory modifications are present especially at the promotors, which in the case of *ERVWE1* is 5´LTR (Chapter 3.1.1 in the Literature review).

Syncytin-1 regulation by DNA methylation

Analysis of the *ERVWE1* promoter CpG methylation revealed hypermethylation in all examined tissues except for the placenta and BeWo cell line. Moreover, this hypermethylation of the promoter was very stable – demethylation agents did not change the methylation level substantially (Matoušková et al. 2006). Promoter hypermethylation is in good agreement with the lack of *syncytin-1* expression in non-placental tissues.

The placenta and BeWo cells, with a high level of Syncytin-1 production (Mi et al. 2000), show a low level of *ERVWE1* CpG promoter methylation (Matoušková et al. 2006; Gimenez et al. 2009; Trejbalová et al. 2011). 3´LTR is heavily methylated in the placenta, which can prevent local promoter competition between 5´LTR and 3´LTR (Gimenez et al. 2009). The DNA methylation level of *ERVWE1* regulatory locus TSE corresponds to *ERVWE1* 5´LTR methylation both in the placenta and in non-placental tissues (Gimenez et al. 2009). However, the binding affinities of GCM1 to the non-methylated versus methylated TSE has not been determined yet.

During the pregnancy, there are changes in the *ERVWE1* promoter DNA methylation. The level of promoter DNA methylation increases in term placentas in comparison with the first trimester (Gimenez et al. 2009). This is in accordance with the fact that Syncytin-1 protein production decreases at the end of pregnancy (Smallwood et al. 2003).

In placental syndromes, changes in both *ERVWE1* expression and promoter DNA methylation were observed (Langbein et al. 2008; Gao et al. 2012; Ruebner et al. 2013). Syncytin-1 expression was significantly lower in pregnancies with fetal growth restriction (Makaroun and Himes 2018). Furthermore, the level of *ERVWE1* promoter DNA methylation was significantly higher in these pregnancies (Makaroun and Himes 2018). Makaroun and colleagues proposed, that the level of the *ERVWE1* promoter DNA methylation could be used as a marker for distinguishing the fetal growth restriction from the physiologically small fetuses (Makaroun and Himes 2018). Future investigations should be done in this field, because this study is limited by the potential overlap of the studied samples and in the small size of the examined groups.

ERVWE1 expression in the testes and testicular cancers is lower than in the placenta (Mi et al. 2000; Gimenez et al. 2010). Moreover, *ERVWE1* promoter DNA is heavily methylated in the testes and slightly demethylated or hypomethylated in testicular cancers where expression of both *ERVWE1* RNA and spliced mRNA occurs (Gimenez et al. 2010; Trejbalová et al. 2011) (Chapter 3.1.4.3 in the Literature review). The case of the testes and testicular tumors points to the importance of the change in the *ERVWE1* promoter CpG methylation for *syncytin-1* expression, but also shows that the final level of transcribed RNA is influenced by other regulatory factors as well.

Collectively, it was unambiguously shown that *ERVWE1* promoter DNA methylation regulates *syncytin-1* production. Nevertheless, other epigenetic regulators are involved as well.

Syncytin-1 regulation by histone modifications

Histone modifications on the promoters affect the gene expression. H3K9ac (histone 3 lysine 9 acetylation) is a mark of active promoters, while H3K9me3 (histone 3 lysine 9 trimethylation) designates inactive promoters. It was shown that *ERVWE1* 5´LTR in BeWo cells was enriched in H3K9ac. On the other hand, the amount of H3K9me3 on the *ERVWE1* 5 ´LTR was high in HeLa cells. These results corresponded well with the expression differences in both cell types (Trejbalová et al. 2011).

The increased level of H3K36me3 (histone 3 lysine 36 trimethylation) on exons is supposed to be a mark of ongoing transcription (Schwartz et al. 2009). An enriched amount of H3K36me3 on the exon-intron boundary was found in the *ERVWE1* locus in BeWo cells, where *syncytin-1* was expressed (Mi et al. 2000; Trejbalová et al. 2011). The authors associated high levels of H3K36me3 on the exon-intron boundary with the efficient splicing

of *ERVWE1* (Trejbalová et al. 2011)

The above-mentioned data demonstrate that histone modifications of *ERVWE1* influence both the transcription from this gene and the RNA splicing efficiency. The significance of RNA splicing is introduced below.

3.1.4.3 Syncytin-1 regulation at the level of transcription – splicing

Three types of transcripts can be produced from the *ERVWE1* locus (Fig. 15). The longest one, called the full-length or the unspliced transcript, has 8 kbp and contains all three *ERVWE1* genes (Mi et al. 2000; Smallwood et al. 2003). The second transcript is called spliced, has 2.8 kbp and contains only the *env* gene (Mi et al. 2000; Trejbalová et al. 2011). This transcript is spliced from the full-length *ERVWE1* RNA (Blond et al. 2000; Mi et al. 2000; Smallwood et al. 2003). Only from the spliced mRNA, the Syncytin-1 protein can be translated (Mi et al. 2000; Smallwood et al. 2003). Therefore, the efficiency of splicing affects the amount of mRNA available for translation. This is the reason why in ERVWE1 research, both full-length and spliced transcripts should be detected. The third transcript is called pseudospliced and has 7.5 kbp in length (Blond et al. 1999; Trejbalová et al. 2011). Pseudospliced mRNA is produced from the full-length transcript by splicing-out the 2.1 kbp long intron of non-retroviral origin, which is a part of the *ERVWE1* locus (Blond et al. 1999). The same splicing strategy was observed in all apes (Mallet et al. 2004).

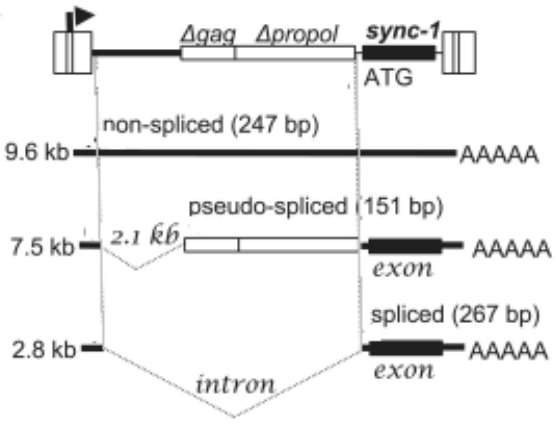

Fig. 15: Schematic representation of ERVWE1 splicing strategy. *Three types of mRNA can be produced from the ERVWE1 locus. Full-length mRNA (unspliced, 8 kbp long), pseudospliced (7.5 kbp, 2.1 kbp long intron of non-retroviral origin is spliced-out) and spliced mRNA (2.8 kbp). Only from the spliced transcript, the Syncytin-1 protein can be translated. The figure was taken and adjusted from* (Trejbalová et al. 2011).

In BeWo cells, the amount of the *ERVWE1* spliced form is six times higher than the level of the full-length form (Trejbalová et al. 2011). Moreover, in HeLa cells, where Syncytin-1 is not produced, the *ERVWE1* full-length form was detected, but the amount of the

spliced form did not reach the detection level (Trejbalová et al. 2011). These results demonstrate the importance of the spliced transcript detection and points out to the regulation of Syncytin-1 by the levels of different transcripts. However, placenta-specific factors involved in high splicing efficiency of *syncytin-1* RNA have not been determined yet.

3.1.5 Regulation of Syncytin-1 activity

Syncytin-1 fusogenic activity can be negatively regulated by the presence of steroid hormones. Steroid hormones stimulate both Syncytin-1 and TGF-β (transforming growth factor-β) production. TGF-β is able to induce cell proliferation, differentiation, migration, and further cell transformation in various cell types, reviewed in (Derynck and Zhang 2003). At the same time, TGF-β is a negative regulator of Syncytin-1-mediated cell-cell fusion. Therefore, the presence of Syncytin-1 together with TGF-β leads to cell proliferation and not the syncytia formation (Strick et al. 2007). This phenomenon was observed in endometrial carcinomas. Further research has to be done to elucidate the TGF-β vs. Syncytin-1 interplay in other tissues.

Furthermore, Syncytin-1 activity can be negatively regulated by the presence of its inhibitor, Supressync (Sugimoto et al. 2013). This inhibitor is a protein product of another endogenous retrovirus, the member of HERV-Fb1 family (Sugimoto et al. 2013). Supresssync is able to bind to Syncytin-1 receptor, ASCT2, and therefore prevent the interaction between Syncytin-1 and ASCT2.

Collectively, Syncytin-1 activity can be post-translationally regulated by specific proteins.

3.1.6 Other endogenous retroviruses in human placenta

Except for Syncytin-1, there are two other fusogenic envelope glycoproteins expressed in the human placenta: ERVFRDE1 and EnvPb (Blaise et al. 2003; 2005). ERVFRDE1 displays placenta-specific expression and fusogenic properties and will be described in more detail below. Endogenous retrovirus EnvPb is expressed in the placenta; however, it is also present in low levels in various other human tissues (Blaise et al. 2005). A placenta-specific expression pattern is also shown by EnvV and ERV-3, but during evolution, their Envs lost the fusion properties (Muir et al. 2004; Blaise et al. 2005). Hypothetically, EnvPb, EnvV and ERV-3 endogenous retroviruses represent ancestral *syncytins* that had entered the primate

germ line before Syncytin-1, but lost their function during evolution (Esnault et al. 2013).

3.2 ERVFRDE1 (Syncytin-2)

Endogenous retrovirus ERVFRDE1 belongs to the HERV-FRD family (Blaise et al. 2003). As ERVFRDE1 shows placenta-specific expression and the Env protein produced from this locus has fusogenic properties, the Env glycoprotein was called Syncytin-2 (Blaise et al. 2003). Syncytin-2 is evolutionarily older than Syncytin-1 and was already present in the ancestor of the New world monkeys and Old world monkeys (Blaise et al. 2003) (Fig. 16).

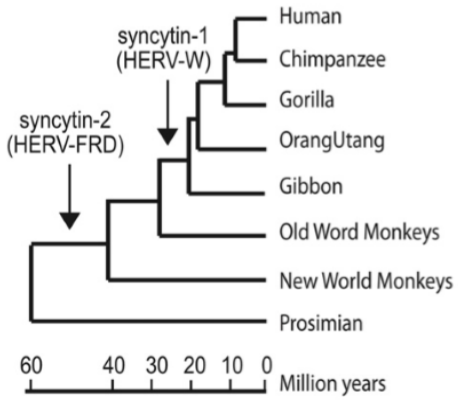

Fig. 16: Schematic representation of syncytin-1 and syncytin-2 integrations. *The figure was taken from (Dupressoir et al. 2012).*

In the placenta, Syncytin-2 is expressed in villous trophoblasts and syncytiotrophoblast, but not in the extravillous trophoblasts (Malassiné et al. 2007). This is in contrast to Syncytin-1, which was found in all three trophoblast types (Mi et al. 2000; Malassiné et al. 2005). Moreover, the level of *ERVFRDE1* expression in the placenta is lower than the level of *ERVWE1* (Trejbalová et al. 2011). The data implies that even though Syncytin-1 and Syncytin-2 are both fusogenic, their functions are not completely redundant (Malassiné et al. 2007). Their involvement in the fusion of human placenta cells is considered to be similar to the situation in the mouse, where two Syncytins are also involved (Chapter 3.3 in the Literature review). Further research has to be done to elucidate the exact activities of Syncytin-1 and Syncytin-2 in human placenta formation.

The SU domain of Syncytin-2 must interact with a specific receptor to induce the cell-cell fusion. The receptor for Syncytin-2 is MFSD2, a carbohydrate transporter (Esnault et al. 2008). The TM domain of Syncytin-2 contains the fusion peptide and ISD. In contrast to Syncytin-1, Syncytin-2 has functional ISD (Mangeney et al. 2007). Nevertheless, the

immunosuppressive role of Syncytin-2 protein in the fetomaternal tolerance still remains to be proved.

It seems that the *ERVFRDE1* expression regulation includes similar mechanisms to that of *ERVWE1*, the regulation of splicing and expression regulation by means of DNA methylation.

The splicing strategy of *ERVFRDE1* corresponds to the classical proviral splicing system. Two forms of transcripts are produced (Fig. 17) – the full-length and the spliced (Vargas et al. 2009). The full-length transcript has 9.1 kbp and contains all three coding genes. The spliced form has 3.1 kbp in length and contains only the *env* gene. Only from the spliced mRNA, the Syncytin-2 protein can be translated (Vargas et al. 2009).

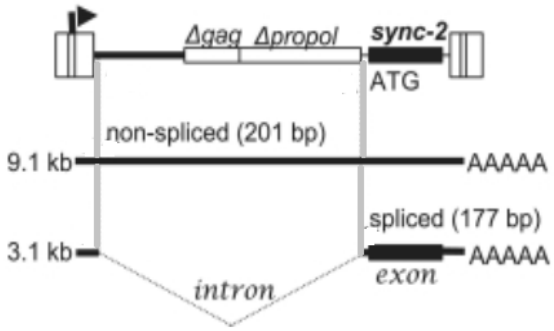

Fig. 17: Schematic representaion of ERVFRDE1 splicing strategy. *There are two types of transcripts produced from the ERVFRDE1 locus. The full-length transcript contains all three coding genes and has 9.1 kbp in length. The spliced mRNA has 3.1 kbp in length and contains only the env gene. Only from the spliced mRNA, the Syncytin-2 protein can be translated. The figure was taken and adjusted from (Trejbalová et al. 2011).*

The level of *ERVFRDE1* promoter DNA methylation is very low in the placenta, where *syncytin-2* expression occurs (Blaise et al. 2003; Gimenez et al. 2009). In contrast to *ERVWE1*, the *ERVFRDE1* promoter retains the same DNA methylation level during the entire pregnancy (Gimenez et al. 2009). In most of the non-plancetal cells, a high level of *ERVFRDE1* promoter DNA methylation corresponding with silenced *syncytin-2* transcription can be found (Gimenez et al. 2009). Interestingly, some of the associated non-trophoblastic cells, specifically fetal cord blood cells and maternal peripheral blood cells, showed a low level of *ERVFRDE1* promoter DNA methylation. Nevertheless *syncytin-2* expression is suppressed in these cells (Gimenez et al. 2009). Therefore, other regulators of the *ERVFRDE1* transcription should be involved, e.g., transcription factors.

Syncytin-2 has been less studied than Syncytin-1, therefore deeper analysis of ERVFRDE1 function has to be performed in the future.

3.3 Syncytins in other placental mammals

Discovery of human Syncytins and their role in human placenta morphogenesis has instigated extensive search for *syncytin* genes in other placental mammals. The following characteristics were required for the candidate genes encoding the retroviral Env proteins: (i) placenta-specific expression, (ii) fusogenic properties, (iii) high level of conservation within the analyzed clade since the integration into the host genome. Indeed, this search was successful and *syncytins* were found in all mammalian placental species that have been explored.

Bioinformatic analysis of the mouse genome led to the discovery of *syncytin-A* and *syncytin-B* genes. Syncytin-A and Syncytin-B are fusogenic Env proteins, which are specifically expressed in the mouse placenta (Dupressoir et al. 2005). They integrated into the murine genome about 20 million years ago. Importantly, *syncytin-A* and *syncytin-B* represent two independent integration events (Dupressoir et al. 2005). Knock-out experiments revealed that Syncytin-A is essential for murine placenta development (Dupressoir et al. 2009). *Syncytin-A* knock-out was lethal for homozygous embryos, which displayed defective cell-cell fusion in placental layers. S*yncytin-B* null embryos were viable; however, the placenta displayed altered structure and unfused cytotrophoblasts (Dupressoir et al. 2011). Double knock-out mouse embryos died even earlier than single *syncytin-A* knock-outs. This knock-out of the mouse *syncytin* genes provided evidence for their critical role in placentation. Knock-out experiments of *syncytins* in other mammals have not been performed due to their technical difficulty.

Other described Syncytins are Syncytin-Ory1 in Langomorpha (Heidmann et al. 2009), Syncytin-Car1 in Carnivora (Cornelis et al. 2012), Syncytin-Rum1 in Ruminantia (Cornelis et al. 2013), Syncytin-Ten1 in Tenrecidaei (Cornelis et al. 2014), Syncytin-Mar1 in Rodentia clade Marmonita (Redelsperger et al. 2014), Syncytin-Opo1 in Didelphimorphanda (Cornelis et al. 2015), Syncytin-Mab1 in viviparian clade of lizards – the Mabuya (Cornelis et al. 2017). All *syncytins* known to date are depicted in Fig. 18. The known *syncytins* described above are not orthologous to each other and moreover they come from independent retroviral infections. The unrelated *syncytins* can contribute to the morphological variability of placentas in different clades. In the future, other *syncytins* in other placental clades are supposed to be found.

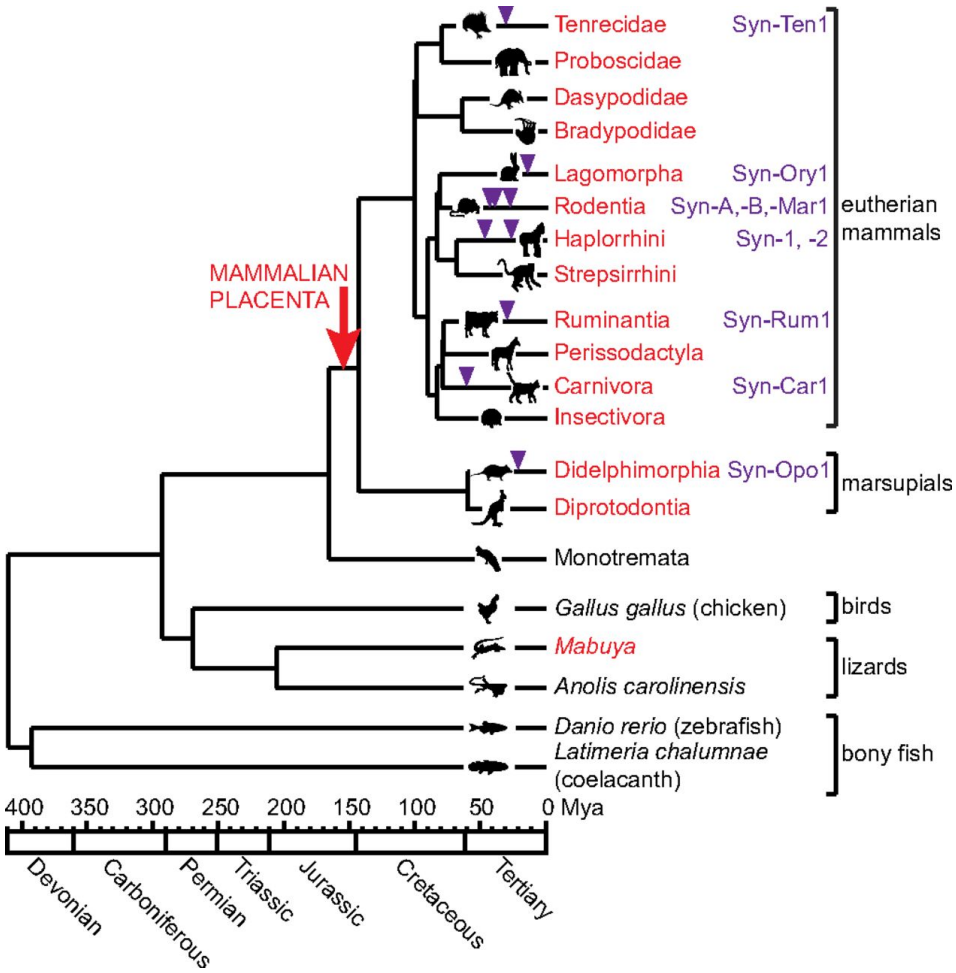

Fig. 18: Schematic representation of the known Syncytins. *The time of integration is represented by the violet triangles. The figure was taken from (Cornelis et al. 2017).*

In Bovinae, which belong to Ruminantia, placenta-specific fusogenic Env called Fematrin-1 was found (Nakaya et al. 2013). It is supposed that Fematrin-1 contributes to the placenta diversity of Ruminantia by fusing binucleated trophoblasts with mononucleated maternal endometrial cells (Nakaya et al. 2013).

In hyenas, which belong to Carnivora, another retroviral *env* gene with placenta-specific expression was found. Therefore, Hyenidae have fusogenic Syncytin-Car1 together with non-fusogenic Hyena-env, which is specifically expressed on the fetomaternal interface (Funk et al. 2018). As Hyenidae, in contrast to other Carnivora, have a hemochorial type of placenta, Hyena-env is supposed to participate in this placenta type transition (Funk et al. 2018). Nevertheless, more research has to be done to elucidate the potential role of Hyena-env in the placenta formation of Hyenidae.

Furthermore, it was hypothesized that *syncytins* had already participated in the emergence of placental mammals. However, it is an open question whether it would be

possible that a fortuitously captured gene starts performing a vital function in the host organism In addition, there is a discrepancy between the time of placenta appearance and the endogenization of *syncytins*. Mammalian placenta evolved 160 – 180 million years ago; nevertheless *syncytins* could be found in mammalian genomes during the last 12 – 80 million years, reviewed in (Imakawa et al. 2015). Two proposed explanations exist for this phenomenon. Imakawa and colleagues offer the so called *baton pass* hypothesis. According to this hypothesis, some other cellular proteins had formed the primary placental syncytiotrophoblast. The subsequently endogenized Syncytins were probably able to induce the trophoblast fusion in a more efficient way, provided its host positive selective advantage and took over this function (Imakawa et al. 2015). According to (Esnault et al. 2013; Lavialle et al. 2013), it was the ancestral retroviral *env* gene that played a decisive role in the formation of primary placental syncytiotrophoblast. This founding *env* gene had been replaced during mammalian radiation by co-optation of new *env* genes from new retroviruses. The old, "degenerating" *syncytin* would progressively lose its function due to the absence of selective pressure. It is possible that EnvPb or ERV-3 represent the "degenerated" *syncytins*. Both hypotheses explain the abundance of placental structures that would be caused by different types of syncytins that were endogenized (Lavialle et al. 2013; Imakawa et al. 2015). The puzzle of the emergence of placental mammals is still waiting for its solution and the evolution of syncytins can help to decipher it.

Material and Methods

1 Ethics statement

The studies presented here were performed in agreement with the principles of WMA Declaration of Helsinki. Research was authorized by the Ethics Committee of the University Hospital Kralovske Vinohrady, under the reference number EK-VP/06/2012. Each patient who participated in the studies had provided an informed written consent. All samples were stored and analyzed at the Institute of Molecular Genetics under the supervision and according to the regulation of the internal Committee of Ethics, Manipulation with Recombinant DNA and Clinical Research of the Institute of Molecular Genetics.

2 Tissue samples

Testicular tissue samples examined in the studies were collected from patients with post-pubertal testicular germ cell tumors. Patients were treated at the Institute of Urology, University Hospital Kralovske Vinohrady, and Third Faculty of Medicine, Charles University. Lymphoma samples were gained from the Department of Otorhinolaryngology, University Hospital Kralovske Vinohrady. Endometrial carcinomas and placental samples were obtained from the Department of Obstetrics and Gynecology, University Hospital Kralovske Vinohrady. Samples were collected between the years 2011 – 2015.

Most of the tissue samples were used in both ‘TET1’ (Overexpression of TET dioxygenases in seminomas associates with low levels of DNA methylation and hydroxymethylation) and ‘Syncytin-1’ (DNA hypomethylation and aberrant expression of the human endogenous retrovirus ERVWE1/syncytin-1 in seminomas) study. Nevertheless, some samples were not incorporated in both mentioned studies because of the insufficient amount of the tissue or low quality of the sample. The described problems are mostly connected with non-GCT testes, due to the fact that most of the samples were obtained from patients with atrophic testes, where the quality of material was variable. Therefore, we were able to use the samples only in some experiments. As gaining the tissue samples, and non-GCT testes in particular, is not very easy, we wanted to use all samples that were available, even though we could not analyze them in both studies. This is the reason why the numbering of samples varies between the studies. In this thesis, the same numbering as had been used in published

articles was applied.

In total, 29 pure seminomas, one scar after the seminoma (only Syncytin-1 study), 18/17 non-seminomas (TET1 and Syncytin-1 study, respectively) were analyzed. Among the non-seminomas, there were 13/12 mixed GCTs (TET1 and Syncytin-1 study, respectively), two pure embryonal carcinomas, two pure teratomas and one pure yolk sac tumor. For most of the tumors, healthy matched tissues were collected. The healthy matched control samples were obtained from the tissues adjacent to the tumor, but macroscopically not affected by the germ cell tumor. In some cases, there were two samples collected from distinct parts of the same tumor or healthy matched tissue. As a negative control, 7/10 non-GCT (non-germ cell tumor) samples (TET1 and Syncytin-1 study, respectively) were used. In our analysis, we also explored three Hodgkin lymphomas, five non-Hodgkin lymphomas and 6/7 endometrial carcinomas (TET1 and Syncytin-1 study, respectively). Furthermore, six healthy terminal placentas and one PBMC sample (peripheral blood mononuclear cells) were analyzed as non-tumoral samples. PBMCs were obtained from Kateřina Trejbalová. Samples were frozen to -80°C immediately after the sampling. All analyzed samples are summarized and characterized in Tables 1 and 2.

Table 1: List of analyzed samples and their characteristics used in the "Overexpression of TET dioxygenases in seminomas associates with low levels of DNA methylation and hydroxymethylation" study. *The nomenclature used in the study is presented. T stands for tumor. When two biopsies were taken from distinct parts of one sample, B stands for the second biopsy. Some tumor samples were pooled with healthy matched controls (depicted by +). Tumor type is described by abbreviation when possible: SE = seminoma, EC = embryonal carcinoma, TE = teratoma, YST = yolk sac tumor and CHC = choriocarcinoma. Samples T18 and T19 were taken from the same patient, who underwent the bilateral orchiectomy (T18 from the left and T19 from the right testis). Composition of the mixed GCTs was determined at the Department of Pathology, Third Faculty of Medicine, Charles University in Prague and is shown in the type of tumor column. Non-Hodgkin lymphomas were classified according to modified Ann Arbor staging classification (Musshoff 1977). Hodgkin lymphomas were classified according to Cotswold revision of the Ann Arbor staging classification of Hodgkin's lymphomas (T. A. Lister et al. 1989). NOS stands for non-otherwise specified. Germ cell tumors were classified according to the WHO classification and TNM staging (Sobin and Compton 2010; Moch et al. 2016). In TNM classification, T characterizes the size of the primary tumor and whether this tumor has invaded the nearby tissue. N describes whether the tumor has invaded lymph nodes. M stand for the possible presence of tumor metastasis, S stands for the elevation of the serum tumor markers.. X indicates the absence of information about the classification.*

Tumor sample	Tumor-matched control	Type of tumor	Age	TNM
T1	+	pure SE	31	T1 N0 M0 S0
T2	+	pure SE	36	T2 N0 M0 S0
T3	+	pure SE	49	T2 N3 M0 S1
T4	+	pure SE	52	T1 N0 M0 S0
T5		pure SE	36	T1 N0 M0 S0
T6	+	pure SE	42	T1 N0 M0 S0
T7		pure SE	36	T2 N0 M0 S0
T8	+	pure SE	67	T1 N0 M0 S0

T9		pure SE	50	T2 N0 M0 S0
T10	+	pure SE	32	T1 N0 M0 S0
T11		pure SE	40	T1 N0 M0 S0
T12	+	pure SE	32	T1 N1 M0 S0
T13	+	pure SE	30	T1 Nx Mx S0
T14	+	pure SE	42	T2 N0 M0 S0
T15	+	pure SE	44	T1 Nx Mx Sx
T16, T16-B		pure SE	29	T1 N0 M0 S0
T17		pure SE	34	T2 N0 M0 S0
T18		pure SE (bilateral orchiectomy – left testis)	31	T2 Nx Mx Sx
T19	+	pure SE (bilateral orchiectomy – right testis)	31	T2 Nx Mx Sx
T20, T20-B		pure SE	46	T1 Nx Mx Sx
T21	+	pure SE	50	T1 N0 M0 S0
T22	+	pure SE	26	T1 Nx Mx Sx
T23	+	pure SE	28	T2 N0 M0 S0
T24	+	pure SE	29	T1 Nx Mx Sx
T25	+	pure SE	43	T1 Nx Mx Sx
T26, T26-B	+	pure SE	39	T3 N0 M0 S0
T27, T27-B	+	pure SE	53	T1 N0 M0 S0
T28, T28-B	+	pure SE	45	T1 N0 M0 S0
T29	+	pure SE	28	T1 N0 M0 S0
T30	+	mixed GCT – 80% SE, 20% EC	51	T1 N0 M0 S0
T31	+	mixed GCT – 60 – 70 % SE, 25 – 30% TE, 15% EC	43	T1 N0 M0 S0
T32	+	mixed GCT – 80% SE, 20% EC	41	T1 N0 M0 S0
T33	+	mixed GCT – 95% SE, 5% EC	32	T1 N0 M0S0
T34	+	mixed GCT – 95% SE, 5% EC	34	T1 N0 M0 S0
T35	+	mixed GCT – 90% EC, 10% SE	28	T1 N2 M1a S0
T36	+	mixed GCT – 40% EC, 25% SE, 25% YST, 5% TE	45	T2 N0 M0 S1
T37, T37-B	+	mixed GCT – 50% EC, 20% TE, 20% SE, 5% CHC	26	T1 N0 M0 S0
T38, T38-B	+	mixed GCT – 60% TE, 30% SE, 10% EC, 1% YST	56	T2 N0 M0 S0
T39	+	mixed GCT – 85% YST, 15% SE	36	T1 N0 M0 S0
T40	+	mixed GCT – 40% YST, 35% EC, 25% TE	36	T1 N1 M0 S1
T41, T41-B	+	mixed GCT – 60% CHC, 30% EC, 10% YST	26	T3 N0 M0 S0
T42, T42-B	+	mixed GCT – 40% CHC, 30% EC, 30% YST	28	T1 N0 M0 S1
T43	+	pure EC	37	T1 Nx Mx Sx
T44	+	pure EC	24	T2 N0 M0 S0
T45	+	pure TE with somatic type malignancy	33	T2 N0 M0 S0
T46		pure TE	47	T1 N0 M0 S0
T47, T47-B	+	pure YST	39	T2 N0 M1b S1
T48		testes without GCT	32	
T49		testes without GCT	52	
T50		testes without GCT	45	
T51		testes without GCT	52	
T52		atrophic testes	78	
T53, T53-B		atrophic testes	21	
T54		testicular ischemia-reperfusion injury	45	
T55		diffuse large B-cell non-Hodgkin lymphoma NOS	61	IIIE
T56		atropic testis due to testicular torsion	78	
T57		atropic undescendent testis	21	

T58		testicular ischemia-reperfusion injury	45	
T59		diffuse large B-cell non-Hodgkin lymphoma NOS	61	IIIE
T60		diffuse large B-cell non-Hodgkin lymphoma NOS	71	IV
T61		Hodgkin lymphoma, nodular sclerosis subtype	26	IIA
T62		Hodgkin lymphoma, nodular sclerosis subtype	78	IVB
T63		Hodgkin pymphoma, nodular sclerosis subtype	61	IIA
T64		diffuse large B-cell non-Hodgkin lymphoma NOS	59	II2
T65		diffuse large B-cell non-Hodgkin lymphoma NOS	80	II2E
T66		follicular B-cell non-Hodgkin lymphoma grade IIIb, transformation to diffuse large B-cell lymphoma	60	IV
T67		endometroid endometrial carcinoma, grade 3	66	T1a N0 M0
T68		endometroid endometrial carcinoma, grade 3	65	T1a N0 M0
placenta 1		endometroid endometrial carcinoma, grade 3	57	T1b N0 M0
placenta 2		endometroid endometrial carcinoma, grade 3		
placenta 3		endometroid endometrial carcinoma with small component of serous carcinoma, grade 3	77	T1a Nx M0
placenta 4		endometroid endometrial carcinoma, grade 3	69	T1b Nx M0
placenta 5		clear cell endometrial carcinoma, grade 3	85	T1a Nx M0
placenta 6		healthy terminal placenta		
PBMC		healthy terminal placenta		

Table 2: List of analyzed samples and their characteristics used in the "DNA hypomethylation and aberrant expression of the human endogenous retrovirus ERVWE1/syncytin-1 in seminomas" study. *The nomenclature used in the study is presented. T stands for tumor. When two biopsies were taken from distinct parts of one sample, B stands for the second biopsy. Some tumor samples were pooled with healthy matched controls (depicted by +). Tumor type is described by abbreviation when possible: SE = seminoma, EC = embryonal carcinoma, TE = teratoma, YST = yolk sac tumor and CHC = choriocarcinoma. Samples T19 and T20 were taken from the same patient, who underwent bilateral orchiectomy (T19 from the left and T20 from the right testis). Composition of the mixed GCTs was determined at the Department of Pathology, Third Faculty of Medicine, Charles University in Prague and is shown in the type of tumor column. Non-Hodgkin lymphomas were classified according to modified Ann Arbor staging classification (Musshoff 1977) Hodgkin lymphomas were classified according to Cotswold revision of the Ann Arbor staging classification of Hodgkin's lymphomas (T. A. Lister et al. 1989). NOS stands for non-otherwise specified. Germ cell tumors were classified according to the WHO classification and TNM staging (Sobin and Compton 2010; Moch et al. 2016). In TNM classification, T characterizes the size of the primary tumor and whether this tumor has invaded the nearby tissue. N describes whether the tumor has invaded the lymph nodes. M stand for the possible presence of tumor metastasis, S stands for the elevation of the serum tumor markers. X indicates the absence of information about the classification.*

Tumor sample	Tumor-matched control	Type of tumor	Age	TNM
T1	+	scar after SE	51	T1 N0 M0 S0
T2	+	pure SE	31	T1 N0 M0 S0
T3	+	pure SE	36	T2 N0 M0 S0
T4	+	pure SE	49	T2 N3 M0 S1
T5	+	pure SE	52	T1 N0 M0 S0
T6		pure SE	36	T1 N0 M0 S0
T7	+	pure SE	42	T1 N0 M0 S0
T8		pure SE	36	T2 N0 M0 S0
T9	+	pure SE	67	T1 N0 M0 S0
T10		pure SE	50	T2 N0 M0 S0
T11	+	pure SE	32	T1 N0 M0 S0
T12		pure SE	40	T1 N0 M0 S0

T13	+	pure SE	32	T1 N1 M0 S0
T14	+	pure SE	30	T1 Nx Mx S0
T15	+	pure SE	42	T2 N0 M0 S0
T16	+	pure SE	44	T1 Nx Mx Sx
T17		pure SE	29	T1 N0 M0 S0
T18		pure SE	34	T2 N0 M0 S0
T19	+	pure SE (bilateral orchiectomy – left testis)	31	T2 Nx Mx Sx
T20	+	pure SE (bilateral orchiectomy – right testis)	31	T2 Nx Mx Sx
T21	+	pure SE	28	T1 N0 M0 S0
T22, T22-B		pure SE	46	T1 Nx Mx Sx
T23	+	pure SE	50	T1 N0 M0 S0
T24	+	pure SE	26	T1 Nx Mx Sx
T25	+	pure SE	28	T2 N0 M0 S0
T26	+	pure SE	29	T1 Nx Mx Sx
T27	+	pure SE	43	T1 Nx Mx Sx
T28	+	pure SE	43	T1 N0 M0 S0
T29, T29-B	+	pure SE	39	T3 N0 M0 S0
T30, T30-B	+	pure SE	53	T1 N0 M0 S0
T31, T31-B	+	pure SE	45	T1 N0 M0 S0
T32	+	mixed GCT – 80% SE, 20% EC	51	T1 N0 M0 S0
T33	+	mixed GCT – 60 – 70% SE, 25 – 30% TE, 15% EC	43	T1 N0 M0 S0
T34	+	mixed GCT – 80% SE, 20% EC	41	T1 N0 M0 S0
T35	+	mixed GCT – 95% SE, 5% EC	34	T1 N0 M0 S0
T36	+	mixed GCT – 90% EC, 10% SE	28	T1 N2 M1a S0
T37	+	mixed GCT – 40% EC, 25% SE, 25% YST, 5% TE	45	T2 N0 M0 S2
T38, T38-B	+	mixed GCT – 50% EC, 20% TE, 20% SE, 5% CHC	26	T1 N0 M0 S0
T39, T39-B	+	mixed GCT – 60% TE, 30% SE, 10% EC, 1% YST	56	T2 N0 M0 S0
T40	+	mixed GCT – 85% YST, 15% SE	36	T1 N0 M0 S0
T41	+	mixed GCT – 40% YST, 35% EC, 25% TE	36	T1 N1 M0 S1
T42, T42-B	+	mixed GCT – 60% CHC, 30% EC, 10% YST	26	T3 N0 M0 S0
T43, T43-B	+	mixed GCT – 40% CHC, 30% EC, 30% YST	28	T1 N0 M0 S1
T44	+	pure EC	37	T1 Nx Mx Sx
T45	+	pure EC	24	T2 N0 M0 S0
T46	+	pure TE with somatic type malignancy	33	T2 N0 M0 S0
T47		pure TE	47	T1 N0 M0 S0
T48, T48-B	+	pure YST	39	T2 N0 M1b S1
T49		atropic testis due to perinatal torsion	32	
T50		atropic testis due to chronic purulent periorchitis	80	
T51		testicular necrosis due to testicular torsion	56	
T52		testes without GCT, testes with normal spermatogenesis	44	
T53		atropic undescendent testis	52	
T54		testicular ischemia-reperfusion injury	45	
T55		testes without GCT	52	
T56		atropic testis due to testicular torsion	78	
T57, T57-B		atropic undescendent testis	21	
T58		testicular ischemia-reperfusion injury	45	
T59		diffuse large B-cell non-Hodgkin lymphoma NOS	61	IIIE
T60		diffuse large B-cell non-Hodgkin lymphoma NOS	71	IV

T61		Hodgkin lymphoma, nodular sclerosis subtype	26	IIA
T62		Hodgkin lymphoma, nodular sclerosis subtype	78	IVB
T63		Hodgkin pymphoma, nodular sclerosis subtype	61	IIA
T64		diffuse large B-cell non-Hodgkin lymphoma NOS	59	II2
T65		diffuse large B-cell non-Hodgkin lymphoma NOS	80	II2E
T66		follicular B-cell non-Hodgkin lymphoma grade IIIb, transformation to diffuse large B-cell lymphoma	60	IV
T67		endometroid endometrial carcinoma, grade 3	66	T1a N0 M0
T68		endometroid endometrial carcinoma, grade 3	65	T1a N0 M0
T69		endometroid endometrial carcinoma, grade 3	57	T1b N0 M0
T70		endometroid endometrial carcinoma, grade 3		
T71		endometroid endometrial carcinoma with small component of serous carcinoma, grade 3	77	T1a Nx M0
T72		endometroid endometrial carcinoma, grade 3	69	T1b Nx M0
T73		clear cell endometrial carcinoma, grade 3	85	T1a Nx M0
placenta 1		healthy terminal placenta		
placenta 2		healthy terminal placenta		
placenta 3		healthy terminal placenta		
placenta 4		healthy terminal placenta		
placenta 5		healthy terminal placenta		
placenta 6		healthy terminal placenta		
PBMC		peripheral blood mononuclear cells		

Each tumor sample was analyzed by immunohistochemistry to determine the appropriate tumor type. Analysis was done at the Department of Pathology, Third Faculty of Medicine, Charles University in Prague. The following markers were used for the tumor identification: alkaline phosphatase, KIT and OCT4 for seminoma, CD30 for embryonal carcinoma, glypican 3 for yolk sac tumor and chorocarcinoma, human chorionic gonadotrophin and placental lactogen for syncytiotrophoblast cells of choriocarcinoma, and p63 for choriocarcinoma cells of choriocarcinoma. For teratoma identification, markers of teratoma differentiated elements were utilized.

3 Cell cultures

In the studies, BeWo (choriocarcinoma derived) and TCam-2 (seminoma derived) cell lines were used. The BeWo cell line was obtained from ATCC cell line collection. TCam-2 was kindly provided from MUDr. Jan Kopecký, DrSc. (Institute of Physiology CAS). BeWo cells were kept in F-12 (Sigma-Aldrich) and MEM-D (Sigma-Aldrich) media mixed 1:1. Medium was supplemented with 1% $NaHCO_3$ and 10% FBS (Gibco). For TCam-2 cells, RPMI-1640 medium (Sigma-Aldrich) with 10% FBS (Gibco) and 0.3 mg/ml L-glutamine was used. A mixture of penicillin and streptomycin antibiotics (0.1 mg/ml each, Sigma-Aldrich)

was added to the media to avoid bacterial contaminations. Cells were kept in 37°C with 5% CO_2. Both BeWo and TCam-2 cell lines were split every 4-5 days using Trypsin-EDTA (Invitrogen™ Life Technologies).

4 Methods

4.1 Isolation of total RNA

For the gene expression analysis, total cellular RNA was isolated. RNA extraction was performed with RNAzol® and BAN (both Molecular Research Centre, INC.). The manufacturer's recommended protocol was followed. Integrity of isolated RNA was controlled by the quality of 28S and 18S rRNA in 1% agarose gel (w/v, Serva – research grade in 1xTAE buffer (Tris (final concentration 0.04 M, Sigma-Aldrich), EDTA pH 8.0 (final concentration 0.001 M, Serva), 5.6% acetic acid (v/v, Penta), water)). For the analysis, 2 μg of total RNA was used. Electrophoresis was run in a Mini Horizontal Submarine Unit HE 33 (AP Czech) device at 95 V.

RNA MicroPrep™ (Zymo Research) kit was employed for RNA isolation from siRNA transfected cells. This method was used due to the low amount of cells in the analyzed samples. The manufacturer's recommended protocol was followed.

4.2 cDNA preparation

cDNA was synthetized from total RNA. The reaction was performed in 50 μl volume using 1 μg of total RNA, Protoscript II reverse transcriptase (100 U, New England Laboratories, NEB) and random hexanucleotides (5 μM, Promega). The protocol recommended by NEB was followed. After the reaction, the volume of 50 μl was diluted to 100 μl by adding water. As a control of the presence of potential residual DNA after the RNA isolation, RT-minus reaction without reverse transcriptase was employed.

4.3 Quantitative RT-PCR (qRT-PCR)

qRT-PCR was performed to examine mRNA expression of the genes of interest. Analysis was run in a CFX96 Touch Real-Time PCR Detection System (BioRad) device with

CFX Manager Software 3.1 (BioRad). Every reaction was done in 20 µl volume, which contained 2 µl of cDNA, forward and reverse primer (250 nM each), and MESA GREEN qRT-PCR Master Mix Plus for SYBR Assay (Eurogentec). The manufacturer's recommended protocol was followed. Each sample was prepared in triplicate. In case the obtained values differed by more than 10 %, new measurement was done. The following protocol was used: 95°C 5 min, 95°C 15 s – X°C 20 s – 72°C 30 s – fluorescence detection – 40 cycles, 10 min 72°C. X °C represents the specific annealing temperature of each pair of primers (Table 3).

Absolute mRNA expression of TET1 (hTET1-FW and RV primers), TET2 (hTET2-FW and RV), TET3 (hTET3-FW and RV), ERVWE1 (ERVWE1-FW, Syncytin-1 spliced-RV and ERVWE1-RV), and ERVFRDE1 (ERVFRDE1-FW, Syncytin-2 spliced-RV, ERVFRDE1-RV), HERV-W chromosome 4 (HERV-W ch4-FW, HERV-W ch4-RV) and HERV-W chromosome 21 (HERV-W ch21-FW, HERV-W ch21-RV) was measured. Gene production was normalized to the expression of housekeeping gene POLR2A (the α-subunit of DNA-directed RNA polymerase II, POLR2A-FW and RV). For all genes of interest and for the housekeeping gene, calibration curves were constructed from ten-fold serial dilutions of a known quantity of target gene molecules. The average amount of the target gene was calculated and then normalized to 100,000 POLR2A molecules. The obtained sum was used for normalization of gene expression and also for standard deviation calculation. One thousand molecules of the gene of interest was determined as 1 % of POLR2A mRNA. To confirm results obtained with POLR2A, another housekeeping gene – TATA Binding Protein (TBP) – was employed (TBP-FW and RV). Sequences of used primers and their annealing temperatures are summarized in Table 3.

The relative mRNA expression of OCT4 (hOCT4-FW and RV), NANOG (hNANOG-FW and RV), SALL4 (hSALL4-FW and RV), GCM1 (hGCM1-FW and RV), ASCT1 (hASCT1-FW and RV), ASCT2 (hASCT2-FW and RV), DNMT1 (hDNMT1-FW and RV), DNMT3A (hDNMT3A-FW and RV), and DNMT3B (hDNMT3B-FW and RV) was quantified. For relative qRT-PCR data analysis, normalization to the relative amount of housekeeping gene POLR2A obtained in the same cDNA sample was employed. Calibration curves were constructed from ten-fold serial dilutions of samples that contained high concentration of target genes as well as POLR2A. Sequences of used primers and their annealing temperatures are summarized in Table 3.

Primers used for qRT-PCR analysis were always chosen in different exons spanning long introns to avoid DNA contaminations. Nevertheless, as a negative control, RT-minus reaction was run (preparation is described in Chapter 4.2 in the Methods). Furthermore, we

employed a non-template negative control, where water instead of cDNA template was loaded, to exclude any DNA contaminations.

4.4 Droplet digital PCR (ddPCR)

Droplet digital PCR was performed to examine both the total amount of expressed RNA and gene copy number. For gene expression analysis, previously prepared cDNA was used. For copy number calculation, chromosomal DNA was employed (chromosomal DNA isolation is described in Chapter 4.5 in the Methods). PCR reaction contained ddPCR™ Supermix for Probes (BioRad), Droplet Generation Oil for Probes (BioRad), FW and RV primers and probe (250 nM each) and 2 µl of analyzed cDNA/20 ng chromosomal DNA. The manufacturer's recommended protocol was followed. The reaction was run in a QX100™ Droplet Digital™ PCR System (BioRad) device. Both gene of interest and gene for normalization were examined in one reaction at the same time. This was possible due to the fact that different types of probe dyes were used. The obtained data were analyzed with QuantaSoft™ software (BioRad). Every reaction was run in duplicate, the measured amount/copy number was averaged and standard deviation was calculated. As a negative control, RT-minus reaction was employed as well as non-template control. The following protocol was performed: 95°C 10 min, 95°C 30 s – 57°C 30 s – 72°C 30 s – 40 cycles.

The absolute mRNA expression of TET1 (hTET1-FW, hTET1-RV and hTET1-probe), HERV-H (HERV-H-FW, HERV-H-RV, HERV-H-probe) was measured. The obtained amount was normalized to the expression of housekeeping gene POLR2A (hPOLR2A-FW, hPOLR2A-RV and hPOLR2A-probe).

ddPCR was further employed for the ERVWE1 copy number analysis (ERVWE1 copy-FW and RV, ERVWE1 copy-probe). RPP30 (Ribonuclease P protein subunit p30; hRPP30-FW, hRPP30-RV and hRPP30-probe) was used for normalization. The following protocol was applied: 95°C 10 min, 95°C 30 s – 60°C 60 s – 72°C 30 s – 40 cycles.

The used primers and probes are summarized in Table 3. Specificity of primers and probes was verified by sequencing of the PCR products.

4.5 Isolation of chromosomal DNA

For bisulfite and oxidative bisulfite experiments, *IDH* mutation and copy number analysis, total chromosomal DNA was isolated. Phenol-chloroform purification was employed. The obtained DNA was dissolved in water. Samples with low amounts of cells were treated with QIAamp DNA Micro KIT (Qiagen). The manufacturer's recommended protocol was followed.

4.6 Bisulfite and oxidative bisulfite sequencing

Cytosine modifications that are present in the promoter can influence transcription from the gene, hence analysis of the 5mC and 5hmC presence at the *ERVWE1* promoter was done. Bisulfite together with oxidative bisulfite sequencing was performed, because bisulfite sequencing is not able to distinguish between 5mC and 5hmC (Chapter 1.3.1 in the Literature review). One μg of genomic DNA was treated with TrueMethyl® Seq Kit (GEGX, Cambridge Epigenetix). The manufacturer's recommended protocol was followed. Converted genomic DNA was then amplified in PCR reaction with primers specific for the *ERVWE1* promoter (ERVWE1-BIS-FW and RV – Table 3). The employed primers were specific only for the converted DNA, because they contained thymine in FW (and adenine in RV) instead of C in non-methylable C (C outside CpG)). Therefore, non-converted DNA could not be amplified. The PCR reaction consisted of 2 μl of bisufite/oxidative bisulfite-treated genomic DNA, $MgCl_2$ (2.5 mM, Invitrogen), HotStart-IT® Binding Protein (2 μg, Affymetrix), FW and RV primers (0.32 μM each), Platinum Taq Polymerase (2 U, Thermo Fisher Scientific), dNTPs (0.2 mM, Promega) and water. The following protocol was applied: 95°C min, 95°C 50s – 58°C 2 min – 68°C 1 min 30 s – 25 cycles, 95°C 45 s – 54°C 2 min – 68°C 1 min 30 s – 15 cycles and 68°C 10 min. Every reaction was performed in triplicate to exclude the amplification starting only from a single molecule. Several non-template controls that contained water instead of the DNA template were used in every round of PCR reaction. The result of the PCR reaction was then analyzed in 1.5% ultrapure agarose gel (w/v, Bioline – Molecular grade, in 1x TAE buffer (protocol described in Chapter 4.1 in the Methods)). PCR products of the correct length (538 bp) were isolated by QIAGEN II Gel Extraction Kit (Qiagen). The manufacturer's recommended protocol was followed. PCR fragments were cloned into pGEM®-T Easy vector (Promega). The manufacturer's recommended protocol

was followed. The ligation mixture was transformed into transformation competent *Escherichia Coli XL-1 Blue* bacterial strain (the protocol is described in Chapter 4.7 in the Methods). The ligation reaction and 50 µl of transformation competent *Escherichia Coli XL-1 Blue* bacteria suspension were mixed together and incubated on ice for 40 min. Heat-shock transformation was applied in a water bath (1 min and 30 s 42°C). Transformed bacteria were immediately placed into 500 µl of LB medium and shaken for 1 h in 37°C. Bacteria were then seeded to agar plates with ampicillin (1 µg/ml, Sigma-Aldrich), IPTG and S-gal (3 mg/ml and 30 mg/ml respectively, Thermo Fisher Scientific). Plates with bacteria were incubated overnight in 37°C. Due to the presence of IPTG, S-gal and beta-galactosidase gene within the pGEM®-T Easy vector, blue-white selection was used. Blue colonies were bacteria with pGEM®-T Easy vector without insert and white colonies were those with pGEM®-T Easy vector with insert. White colonies were transported into 20 µl LB medium with ampicillin (1 µg/ml, Sigma-Aldrich) and thereafter analyzed by PCR. For the PCR reaction, OneTaq Master Mix (NEB) was employed. The manufacturer's recommended protocol was followed. One and half µl of bacterial suspension together with M13 FW and RV universal primers were applied in one reaction (Table 3). The PCR products were analyzed using 1% agarose gel (gel preparation and electrophoresis process are described in Chapter 4.1 in the Methods). Positive colonies, which contained a 780 bp long PCR product, were sent to the GATC company (https://www.eurofinsgenomics.eu/en/) for Sanger sequencing using M13 RV universal primers (Table 3). The gained sequences were analyzed by QUMA software (Quantification Tool for Methylation Analysis, http://quma.cdb. riken.jp/). Only PCR clones, that had at least 95% conversion were taken for the analysis. In case there were more clones with identical sequences, only one was used. We wanted to avoid the possibility that some molecules could be amplified preferentially, which could disturb the overall result.

The efficiency of 5hmC to 5fC oxidation was examined with a 5hmC synthetic probe. The probe was prepared by amplification of the *ERVWE1* promoter sequence with ERVWE1-nonBIS-FW and RV primers (Table 3) with nucleotides containing dhmCTP instead of dCTP. The PCR reaction was done with OneTaq polymerase (NEB) using the same protocol as for the analysis of the genomic DNA (this chapter, paragraph above). The PCR product was isolated from 1.5% ultrapure agarose gel (this chapter, paragraph above) by QIAGEN II Gel Extraction Kit (Qiagen). The manufacturer's recommended protocol was followed. 1,000 ng of isolated PCR product was treated with TrueMethyl® Seq Kit (GEGX, Cambridge Epigenetix). Manufacturer´s recommended protocol was followed. Afterwards, the PCR reaction with ERVWE1-BIS-FW and RV primers was performed, PCR product was purified

from ultrapure agarose gel, ligated into pGEM®-T Easy vector (Promega), transformed into *Escherichia Coli XL-1 Blue* bacteria, and sequenced (protocols are described in the paragraph above and in the Chapter 4.7 in the Methods). Total efficiency of oxidation was calculated using the QUMA software (Quantification tool for Methylation Analysis, http://quma.cdb.riken.jp/).

4.7 Transformation competent *E. Coli XL-1 Blue* bacterial strain preparation

Escherichia Coli XL-1 Blue bacteria were incubated in 5 ml of LB medium at 37°C overnight. The next day, bacteria were inoculated into 100 ml of LB medium and left at 37°C until bacterial suspension reached the optical density of 0.4 (measured with $\lambda = 590$ nm). Bacteria were then kept for 10 min on ice and centrifuged (7 min, 4°C, acceleration 6, break 5). The pellet was resuspended in TFBI buffer (100 mM RbCl, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2$, 15% glycerol (v/v), pH 5.8 – with usage of acetic acid). Bacteria were centrifuged (7 min, 4°C, acceleration 6, break 5) and the pellet was resuspended in TBFII buffer (10 mM MOPS pH 7.0, 10 mM RbCl, 75 mM $CaCl_2$, 15% glycerol (v/v), pH 6.8 – with usage of NaOH). The bacteria were aliquoted (50 µl each) and stored in -80°C for further use.

4.8 Detection of *IDH1* and *IDH2* mutations

Chosen samples were sequenced for the presence of selected *IDH1* and *IDH2* mutations (IDH1R132, IDH2R140 and IDH2R172). The PCR reaction of 25 µl volume contained Taq polymerase (1 U, Thermo Fisher Scientific), $MgCl_2$ (1.5 mM, Thermo Fisher Scientific) and primers (250 nM each). The following protocol was applied: 95°C 1 min 30s, 95°C 30 s – 60°C 30 s – 72°C 25 s – 25 cycles, 72°C 10 min. Primer sequences and their annealing temperatures used in this analysis (hIDH1-FW, RV and hIDH2-FW, RV) are summarized in Table 3.

Amplified PCR products (IDH1 and IDH2) were purified from 1.5% ultrapure agarose gel (gel preparation and electrophoresis process are described in Chapter 4.7). PCR products of expected length (198 bp for IDH1 and 214 bp for IDH2) were isolated by QIAGEN II Gel Extraction Kit (Qiagen). The manufacturer's recommended protocol was followed.

Purified PCR products were inserted into pGEM®-T Easy vector (Promega). The manufacturer's recommended protocol was followed. Prepared vectors were transformed into

the *Escherichia Coli XL-1 Blue* bacterial strain and blue-white selection was performed (protocols are described in Chapter 4.6 in the Methods). White colonies were transported to 20 µl LB medium with ampicillin (1 µg/ml, Sigma-Aldrich) and then analyzed by PCR. For the PCR reaction, OneTaq Master Mix (NEB) was employed. The manufacturer's recommended protocol was followed. For one reaction, 1.5 µl of bacterial suspension was used together with M13 FW and RV universal primers (Table 3). PCR products were analyzed in 1% agarose gel (gel preparation and electrophoresis process are described in Chapter 4.6 in the Methods). White colonies that contained the PCR product of suspected length (457 bp for IDH1 and 476 bp for IDH2) were sent for Sanger sequencing to the SeqMe company (https://www.seqme.eu/cs/). For sequencing, M13 RV universal primer was used. The protocol recommended by SeqMe was followed.

4.9 Immunohistochemistry analysis

To analyze the presence of 5hmC and TET1 in our samples, we performed immunohistochemical staining. The analysis was done at the Department of Pathology (Third Faculty of Medicine, Charles University in Prague). Samples for immunohistochemistry were embedded into paraffin blocks and cut into 5-micrometer thick sections. Sample sections then underwent heat-induced epitope retrieval with antigen retrieval solution (pH 9.0, Dako), 30 min at 91°C. For 5hmC staining, 5hmC antibody (Active Motif, no. 39769, rabbit polyclonal antibody) was used. This antibody was diluted 1:400 in ChemMate antibody diluent (Dako) and applied to samples for 1 hour at room temperature. For TET1 staining, two TET1 antibodies (Abcam, nos. 121587 and 191698, both rabbit polyclonal antibodies) were employed. For both TET1 antibodies, dilution 1:100 in ChemMate antibody diluent (Dako) was employed. Antibodies were incubated with samples overnight at 4°C. Primary antibodies were visualized using Envision Kit (Dako). The manufacturer's recommended protocol was followed.

In addition, the examined samples were permanently stained with chromogen 3,3-diaminobenzidine (Sigma-Aldrich) and Mayer's hematoxylin (Sigma-Aldrich) to demonstrate the cellular structures and components. Chromogen 3,3-diaminobenzidine was used for visualization of proteins and nucleic acids, Mayer's hematoxylin highlighted nuclear and cytoplasmic structures. Optimal staining conditions were determined using negative and positive controls (sample with normal spermatogenesis and GCNIS sample, respectively).

Samples were considered as immuno-positive if more than 10 % of neoplastic cells in

the analyzed cut-off field showed positive. For every sample, eight to ten non-overlapping random cut-off fields were analyzed. The percentage of positive cells was graded in the following way: (0) negative, (+) weak ($10 \leq 25$ % of the stained cells), (++) moderate ($25 \leq 75$ %), (+++) strong (more than 75 % of positive cells). Immunoreactivity H-score was calculated using the following formula: 3x percentage of strongly staining cells + 2x percentage of moderately staining cells + percentage of weakly stained cells. The H-score ranged from 0 to 300.

4.10 Quantification of 5mC and 5hmC by liquid chromatography-mass spectrometry

The presence of both 5mC and 5hmC was analyzed in a panel of selected samples by LC/MS. The method was performed at the Laboratory of Reprogramming and Chromatin (MCR London Institute of Medical Sciences, London, UK). For this experiment, the protocol of (Amouroux et al. 2016) was followed. The genomic DNA of interest was digested to single nucleotides by enzymatic mixture for at least 8 h at 37°C. The enzymatic mixture used for one sample contained: benzonase (1 U, Novagen), phosphodiesterase I (0.5 mM, Sigma-Aldrich), alkaline phosphatase (200 mU) in 20 mM Tris-HCl pH 7.9 and 4 mM $MgCl_2$. Enzymes were then precipitated by acetonitrile incubation performed overnight at -20°C.

Synthetic nucleotides were used as standards for the calculation of various nucleotides in the samples. The following compounds were used: 2-deoxycytidine as dC, 2-deoxyguanosine as dG, 5-hydrocymethyl-2-deoxycytidin as 5hmC (all Berry & Associate) and C5-methyl-2-deoxycytidine as 5mC (R. I. Chemicals). The standard curve prepared from the synthetic nucleotides was measured in the digestion mix, simirarly as for all analyzed samples.

Nucleotides were separated in an Agilent RRHD Eclipse Plus C18 2.1 x 100 mm 1.8 u column using the HPLC 1290 system (Agilent) and analyzed with an Agilent 6490 quadrupole mass spectrometer. Standard synthetic nucleotide peak response was used to convert the peak-area values to the concentration of the analyzed nucleotide. The detection threshold was above 10 and was calculated by the peak-to-peak method.

4.11 siRNA transfection

To analyze the impact of TET1 in seminomas, TET1 knock-down in the TCam-2 seminoma cell line was performed. For this experiment, specific ON-TARGETplus SMARTpool against TET1 (Dharmacon, GE Healthcare Life Sciences) was applied. Transfection was run with Nucleofector™ V solution (Lonza) and Nucleofector™ program X-005 in a Nucleofector™ 2b Device (Lonza). The manufacturer's recommended protocol was followed. For the determination of transfection efficiency, pmaxGFP® plasmid (Lonza) was used. As a negative control, ON-TARGETplus Non-targeting siRNA (Dharmacon, GE Healthcare Life Sciences) was employed.

To trypsinized TCam-2 cells (2 x 10^6 cells), 100 pmol of siRNA TET1 pool or non-targeting siRNA was added. Two rounds of transfection with the same protocol were performed – on day 0 and day 3. Transfected cells were harvested, while cell density reached between 120 – 140 cells per mm^2, because cell density has an impact on TET1 expression (Nettersheim et al. 2013). Transfection efficiency was measured using a BD LSRII flow cytometer (BD Biosciences, the Centre of Light Microscopy and Flow Cytometry at the Institute of Molecular Genetics of the ASCR, v.v.i.). Cells were collected for ddPCR and analysis of the 5hmC and 5mC content by liquid chromatography-mass spectrometry three or six days after the transfection.

5 Statistical analysis

For statistical analysis of the intergroup specificity, GraphPad Prism software (version 5.04) with two-tailed Mann-Whitney test was employed. Calculated P-values were depicted as follows: * < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001.

6 Primers used in studies

Table 3: Table of primers used in the studies. *Name of primers, primer sequences, and annealing temperatures are indicated. A = adenine, C = cytosine, G = guanine, T = thymine, FAM = 6-flourescein amidite, HEX = hexachloro-fluorescein, BHQ-1 = black hole quencher-1.*

Designation	Sequence	Annealing temperature
hTET1-FW	5′ GAACAGCCATCAGATCTGTAAGAA 3′	63°C qRT-PCR 57°C qRT-ddPCR
hTET1-RV	5′ GTAGTCCATGGATTCTGACTTGGG 3′	
hTET1-probe	5′FAM-ACCATCTGTTGTTGTGCCTCTGGA-BHQ1 3′	57°C
hTET2-FW	5′ TATACCCATCTAGGACAGGTC 3′	62°C
hTET2-RV	5′ CCACTTAGCAATAGGACATCCC 3′	
hTET3-FW	5′ CCAGCATAACCTCTACAATGGG 3′	63°C
hTET3-RV	5′ GTTCTCCTCGCTACCAAACTCATC 3′	
hOCT4-FW	5′ GCAATTTGCCAAGCTCCTGAAGCA 3′	62°C
hOCT4-RV	5′ GGCAGATGGTCGTTTGGCTGAATA 3′	
hNANOG-FW	5′ AAGGTCCCGGTCAAGAAACAG 3′	57°C
hNANOG-RV	5′ CTTCTGCGTCACACCATTGC 3′	
hSALL4-FW	5′ TGTGCAACATTTGTGGGCGAGCTT 3′	61°C
hSALL4-RV	5′ TTCTCGATGGCCAACTTCCTTCCA 3′	
hPOLR2A-FW	5′ GCACCACGTCCAATGACAT 3′	55 - 63°C
hPOLR2A-RV	5′ GTGCGGCTGCTTCCATAA 3′	
hPOLR2A-probe	5′ HEX-TACCACGTCATCTCCTTTGATGGC TCCTA-BHQ1 3′	55 - 63°C
hDNMT1-FW	5′ GAAGGGAGACGTGGAGATGC 3′	60°C
hDNMT1-RV	5′ ATGGAGCGCTTGAAGGAGAC 3′	
hDNMT3A-FW	5′ CCTGTGATGATTGATGCCAAA 3′	63°C
hDNMT3A-RV	5′ GCCCTGCTTTATGGAGTTTGA 3′	
hDNMT3B-FW	5′ GCCGGCTCTTCTTCGAATTT 3′	60°C
hDNMT3B-RV	5′ GAAGTATCGGGCCCTGTGAG 3′	
hIDH1-FW	5′ GATGAGAAGAGGGTTGAGGAGTTC 3′	60°C
hIDH1-RV	5′ GCCAACATGACTTACTTGATCCC 3′	
hIDH2-FW	5′ CTCTGTCCTCACAGAGTTCAAG 3′	60°C
hIDH2-RV	5′ GTGCCCAGGTCAGTGGAT 3′	
universal M13-FW	5′ TGTAAAACGACGGCCAGT 3′	58 - 62°C
universal M13-RV	5′ CAGGAAACAGCTATGAC 3′	
ERVWE1-FW	5′ ACATTTTGGCAACCACGAAC 3′	55°C qRT-PCR 57°C ddPCR
Syncytin-1 spliced-RV	5′ GGCCATGGGGATTTATGATT 3′	
ERVWE1 non-spliced-RV	5′ AAAGTGGAAGCTGGCTTGAG 3′	

ERVWE1 non-spliced-probe	5′ FAM-AGACTCAGGTGTGAGGCTATCTG GG-BHQ1 3′	57°C
ERVFRDE1-FW	5′ CAAGTCAAGGGCTGAACAGG 3′	55°C
Syncytin-2 spliced-RV	5′ CGGTAGGCTGCTAGTGAAGG 3′	
ERVFRDE1 non-spliced-RV	5′ CAGAGCCACTGTGGTTGAGA 3′	
hGCM1-FW	5′ GCCAGATTCCTATGCCAAAC 3′	58°C
hGCM1-RV	5′ CAAGGGATGAGCTTCAGAGG 3′	
hASCT1-FW	5′ TTTGCGACAGCATTTGCTAC 3′	55°C
hASCT1-RV	5′ CGCTGTGGCAGTCACTAGAA 3′	
hASCT2-FW	5′ CATCATCCTCGAAGCAGTCA 3′	57°C
hASCT2-RV	5′ AGTGTTTGAGGAGGGGGTTT 3′	
HERVH-FW	5′ CGTGGCTGCAGTACAAACTTGATAA 3′	57°C
HERVH-RV	5′ GGTCTGTAGCAAAGGAGGATTCAAAG 3′	
HERVH-probe	5′ FAM-CTTTAAATGGCCAGAATATGGCA CTTTCA-BHQ1 3′	
HERVW-FW ch4	5′ GAGCTTTGGTCTGCCTGGAA 3′	60°C
HERVW-RV ch4	5′ TGCGCCACGATCTCAACTGT 3′	
HERVW-FW ch21	5′ AGGCAGAAAGCTGTCGTCCG 3′	60°C
HERVW-RV ch21	5′ GGGGCTACACTTTCAAGAAAGTCATC 3′	
RPP30-FW	5′ GATTTGGACCTGCGAGCG 3′	55 - 60°C
RPP30-RV	5′ GCGGCTGTCTCCACAAGT 3′	
RPP30-probe	5′ HEX-TTCTGACCTGAAGGCTCTG CGC-BHQ1 3′	
ERVWE1 copy-FW	5′ CGCCTGGAGATACAGCAATTA 3′	60°C
ERVWE1 copy-RV	5′ CAGCTAGGCTTAGGGATTCTTAG 3′	
ERVWE1 copy-probe	5′ FAM-TGAGAGACAGGACTAGCTGGATTT CCT-BHQ1 3′	
ERVWE1-BIS-FW	5′ GGAGATATAGTAATTATTTTGTAATTGAG AGAT 3′	54 - 58°C
ERVWE1-BIS-RV	5′ AACAATAACAAACCTTTAACCCAATTA 3′	
ERVWE1 non-BIS-FW	5′ GGAGATACAGCAATTATCTTGCAACTGAG AGAC 3′	60°C
ERVWE1 non-BIS-RV	5′ AACAATGGCAAGCCTTTAGCCCAATTG 3′	
TBP FW	5′ GAGCTGTGATGTGAAGTTTCC 3′	60°C
TBP RV	5′ TCTGGGTTTGATCATTCTGTAG 3′	

Results

PGCs are embryonic precursors of gametes. DNA modifications are crucial for the maintenance of their characteristic features and for their further differentiation (Hajkova et al. 2002; Seisenberger et al. 2012). Tumors derived from the germ cells can be established from PGCs with delayed or blocked maturation (Skakkebék et al. 1998; Oosterhuis and Looijenga 2005; Rajpert-De Meyts 2006; Kristensen et al. 2013). In our first study (Benešová, Trejbalová, Kučerová, et al. 2017), we mainly focused on selected epigenetic features that can explain the previously observed expression of human endogenous retroviruses *ERVWE1* and *ERVFRDE1* (Gimenez et al. 2010; Trejbalová et al. 2011). The expression of these endogenous retroviruses was not specified for the exact GCT type. Moreover, expression from *ERVWE1* and *ERVFRDE1* loci is known to be regulated by DNA methylation (Matoušková et al. 2006; Gimenez et al. 2010; Trejbalová et al. 2011). These facts led us to the exploration of *ERVWE1* production in GCTs. The results are summarized in our second study (Benešová, Trejbalová, Kovářová, et al. 2017).

1 Overexpression of TET dioxygenases in seminomas associates with low levels of DNA methylation and hydroxymethylation

1.1 Expression of pluripotency markers in GCTs

Germ cell tumors (GCTs) originate from PGCs (Oosterhuis and Looijenga 2005) and therefore are characterized by the expression of pluripotency markers (Surani et al. 2007). The mRNA expression levels of some pluripotency factors (OCT4, NANOG and SALL4) were quantified in selected subsets of samples by qRT-PCR.

In seminomas, mRNA expression of *OCT4*, *NANOG* and *SALL4* was significantly higher than in the matched controls ($p < 0.01$, $p < 0.001$, $p < 0.001$, respectively, Fig. 19A, B and C). Comparison between the seminomas and non-GCT controls revealed significantly elevated expression of both *OCT4* and *SALL4* mRNA in the seminomas ($p < 0.01$ and $p < 0.05$, respectively, Fig. 19A and B). In the non-seminomas, we found significantly higher levels of both *OCT4* and *SALL4* mRNA in comparison to the non-GCT controls ($p < 0.01$, $p < 0.05$, respectively, Fig. 19A and C). We did not observe any of the analyzed markers to be

significantly upregulated in non-seminomas in comparison to the matched controls. Comparison of *SALL4* mRNA expression in the seminomas and the non-seminomas showed that the level in the seminomas was significantly higher ($p < 0.05$, Fig. 19C). The TCam-2 cell line, which is derived from a seminoma tumor, displayed elevated mRNA expression of the examined pluripotency markers.

Overall, we confirmed that seminomas retain some characteristics of pluripotency. The non-seminomas expressed only some of the analyzed pluripotency markers, and to a much lesser extent, when compared with the seminomas. Therefore, we confirmed the more differentiated state of the non-seminomas. Non-GCT samples did not show increased expression of the tested pluripotency factors, which is in accordance with their characteristics.

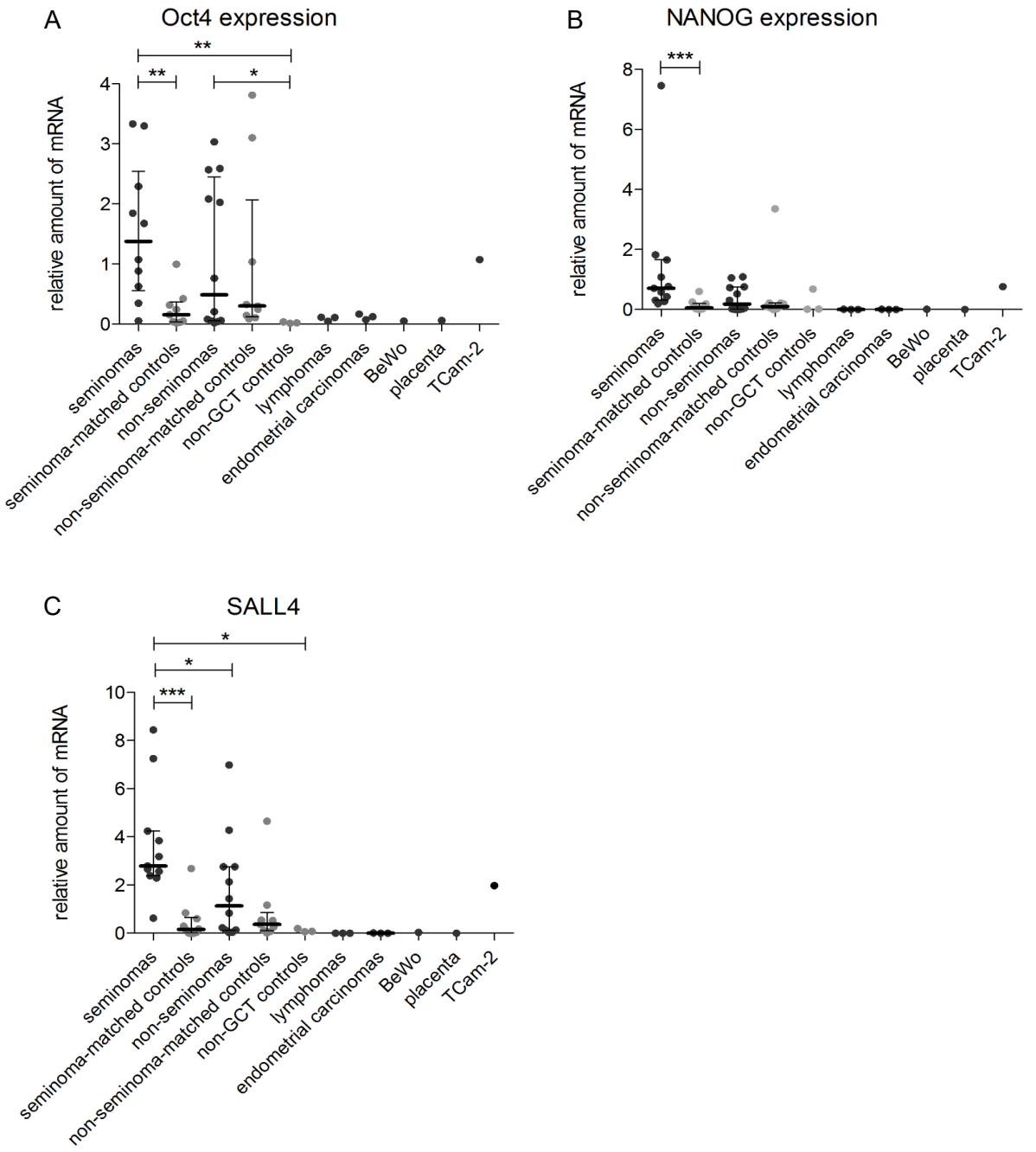

Fig. 19: Relative mRNA expression of OCT4, NANOG and SALL4 pluripotency factors. *(A) Relative expression of OCT4 mRNA. (B) Relative expression of NANOG mRNA. (C) Relative expression of SALL4 mRNA. Relative mRNA expression was measured by qRT-PCR and normalized to the relative expression of the POLR2A (α-subunit of DNA-directed RNA polymerase II) housekeeping gene. Every sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, the placenta, and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the relative amount of the mRNA. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kučerová, et al. 2017)*

1.2 *TET1* mRNA expression in GCTs

All three TET proteins are highly produced in PGCs; however TET1 was reported to have the highest expression there (Yamaguchi et al. 2012; 2013; Vincent et al. 2013; Hill et al. 2018). This was the reason, why we first analyzed the expression of *TET1* mRNA in the panel of our samples. For the analysis, the qRT-PCR method was applied.

We found that the level of *TET1* mRNA was highly elevated in the seminomas in comparison with both the matched and non-GCT controls ($p < 0.0001$ and $p < 0.0001$, respectively, Fig. 20A and B). When we compared *TET1* mRNA expression in the seminomas and non-seminomas, we observed *TET1* mRNA to be significantly higher in the seminomas ($p < 0.05$, Fig. 20B). In the non-seminomas, the level of *TET1* mRNA was significantly elevated in comparison to the non-GCT controls ($p < 0.001$, Fig. 20B). Thereafter, we took a deeper look into the non-seminomas. Two examined embryonal carcinomas showed variable levels of *TET1* mRNA expression (Fig. 20A). Two analyzed teratomas and two yolk sac tumors displayed low *TET1* mRNA production (Fig. 20A). Both tested cell lines, TCam-2 and BeWo, showed high levels of *TET1* mRNA (Fig. 20A). The lymphomas, endometrial carcinomas and PBMCs expressed low amounts of *TET1* mRNA, that were comparable with the non-GCT controls (Fig. 20A).

To summarize, we found *TET1* mRNA to be highly overexpressed in the seminomas in comparison with both types of controls. The non-seminomas manifested upregulated levels of *TET1* mRNA, but only when compared with the non-GCT controls. Furthermore, *TET1* mRNA production observed in the non-seminomas was much lower than in the seminomas.

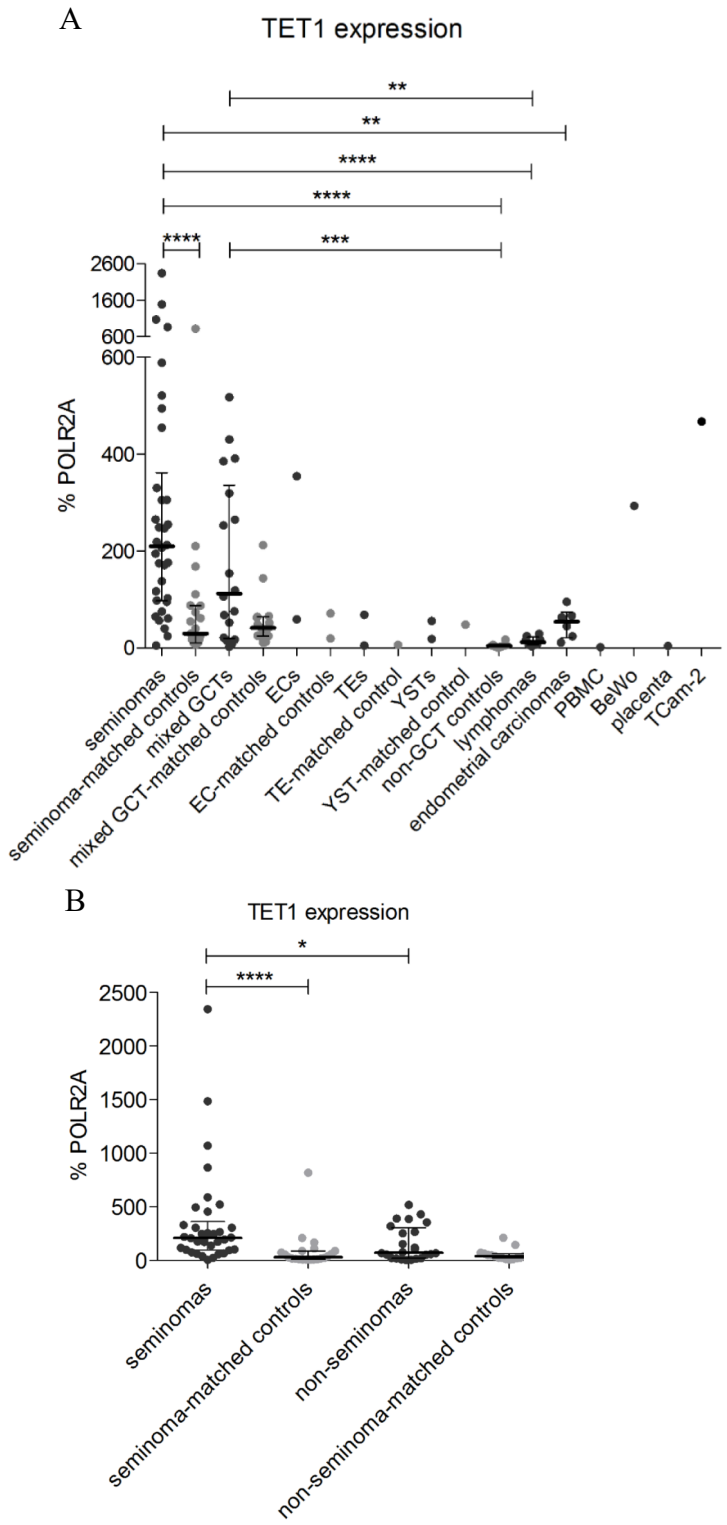

Fig. 20: TET1 mRNA expression in analyzed samples. *(A) Expression of TET1 mRNA in analyzed samples. (B) Comparison of TET1 mRNA expression in the seminomas, non-seminomas and the matched controls. mRNA expression was measured by qRT-PCR and normalized to the percentage of POLR2A expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, the placenta, and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the relative amount of mRNA. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kučerová, et al. 2017)*

1.3 Immunohistochemical analysis of TET1 protein in GCTs

As we described above, *TET1* mRNA production is elevated in the seminomas and to a much lesser extent in some of the non-seminomas. The amount of translated protein does not necessarily correspond to the level of mRNA expressed from the gene. Therefore, we analyzed the TET1 protein presence in our samples by immunohistochemistry (IHC).

In this experiment, a selected sub-set of samples was explored. IHC staining was performed on the whole tissue sections and H-score was determined (Table 4) to quantify the intensity of TET1 staining in our samples. In the mixed GCTs, the H-score was calculated in different parts of the examined sample to avoid the confusion by overestimation of some component.

In the analyzed sections with normal spermatogenesis, we observed strong nuclear and weak granular cytoplasmic staining of gonocytes, Leydig and Sertoli cells (normal testis, Fig. 21A). On the other hand, in most of the interstitial elements, TET1 staining was negative (normal testis, Fig. 21A). When we looked at the GCNIS, we found weak TET1 positivity (GCNIS is indicated by black arrow, Fig. 21A).

Overall, we performed TET1 IHC staining in eight seminoma samples. Six of them – T1, T2, T4, T7, T18, and T22 manifested strong to moderate TET1 positivity (Fig. 21B and Table 4). However, in two other seminomas – T3 and T15 – we observed weak TET1 positivity (Table 4). The examined seminomas displayed TET1 staining in tumor cells, but not in infiltrated lymphocytes. In the mixed GCTs, we observed variable TET1 positivity in the seminoma component. The choriocarcinoma and embryonal carcinoma components displayed stronger TET1 staining than the teratoma and yolk sac tumor sections of GCT. Two representative microphotographs of mixed GCTs (T40 and T41) are depicted in Fig. 22A and B. Two pure embryonal carcinomas (T43 and T44) showed strong TET1 positivity (Fig. 21C, Table 4). Pure teratoma (T46) and pure yolk sac tumor (T47) displayed moderate TET1 staining (Fig. 22C, D and Table 4). Comparison of TET1 staining between the seminomas and non-seminomas did not manifest any significant difference (Fig. 23). Tumor cells showed both nuclear and cytoplasmic staining. All tumor-matched controls analyzed in this study revealed weaker TET1 staining than was observed in the tumor samples (Fig. 21 and 22). It should be noted that the above-mentioned scoring was performed in parallel with a different TET1 antibody, revealing the same result (Fig. 21C).

In summary, we found that TET1 is overexpressed at the protein level in both seminomas and non-seminomas. We did not find any significant difference in TET1 staining

between the seminomas and non-seminomas as a whole group of samples. The highest TET1 staining intensity was observed in the embryonal carcinomas.

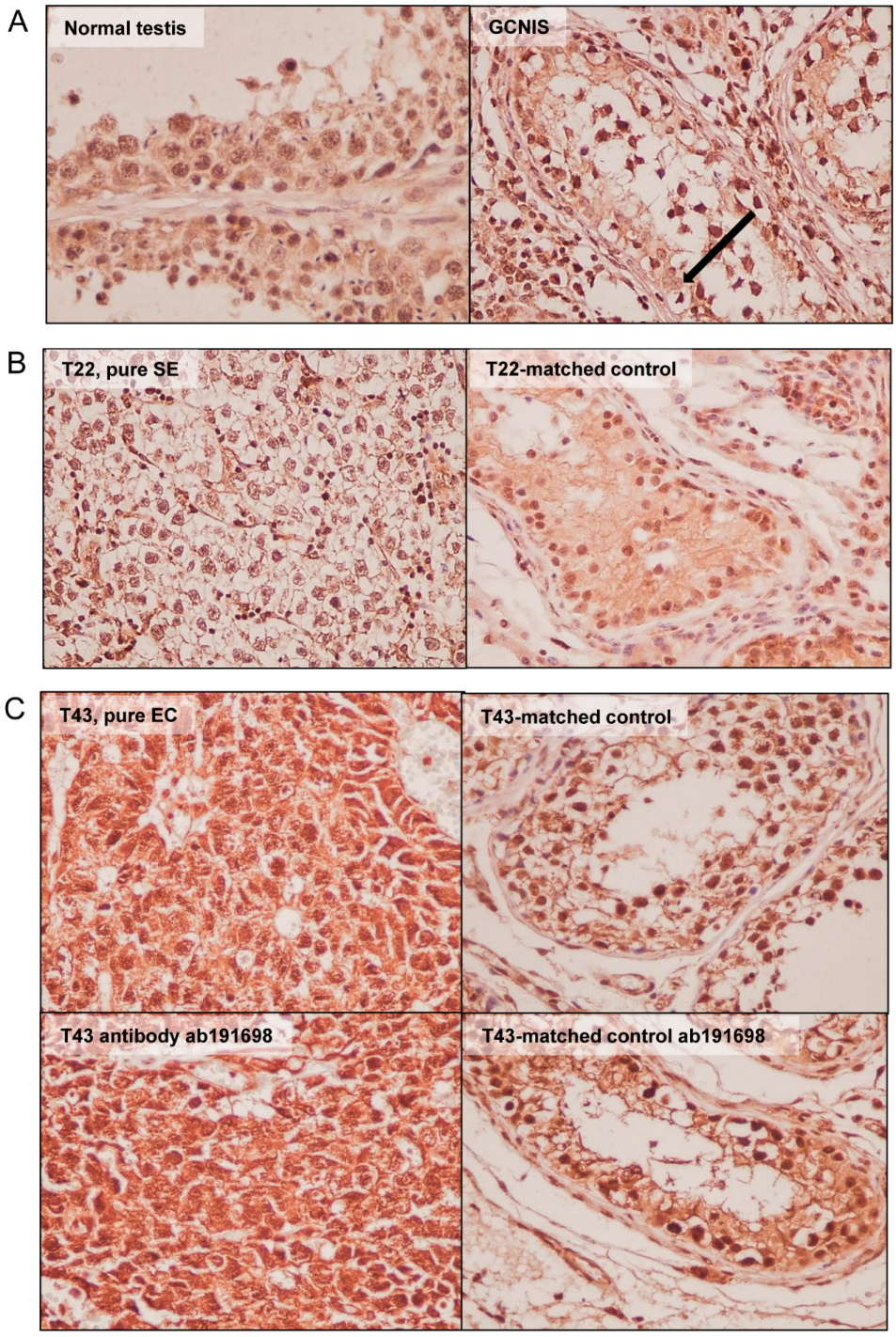

Fig. 21: TET1 IHC staining in selected GCTs. *(A) TET1 IHC staining in the non-GCT testes and GCNIS. Original magnification was 200x in the non-GCT testes and 400x in the GCNIS. (B) TET1 IHC staining in the T22 sample (pure seminoma). Original magnification was 400x. (C) TET1 IHC staining in the T43 sample (pure embryonal carcinoma). In the upper panels, TET1 ab121587 antibody was applied. In the lower panels, TET1 ab191698 antibody was employed. Original magnification was 400x. (Benešová, Trejbalová, Kučerová, et al. 2017)*

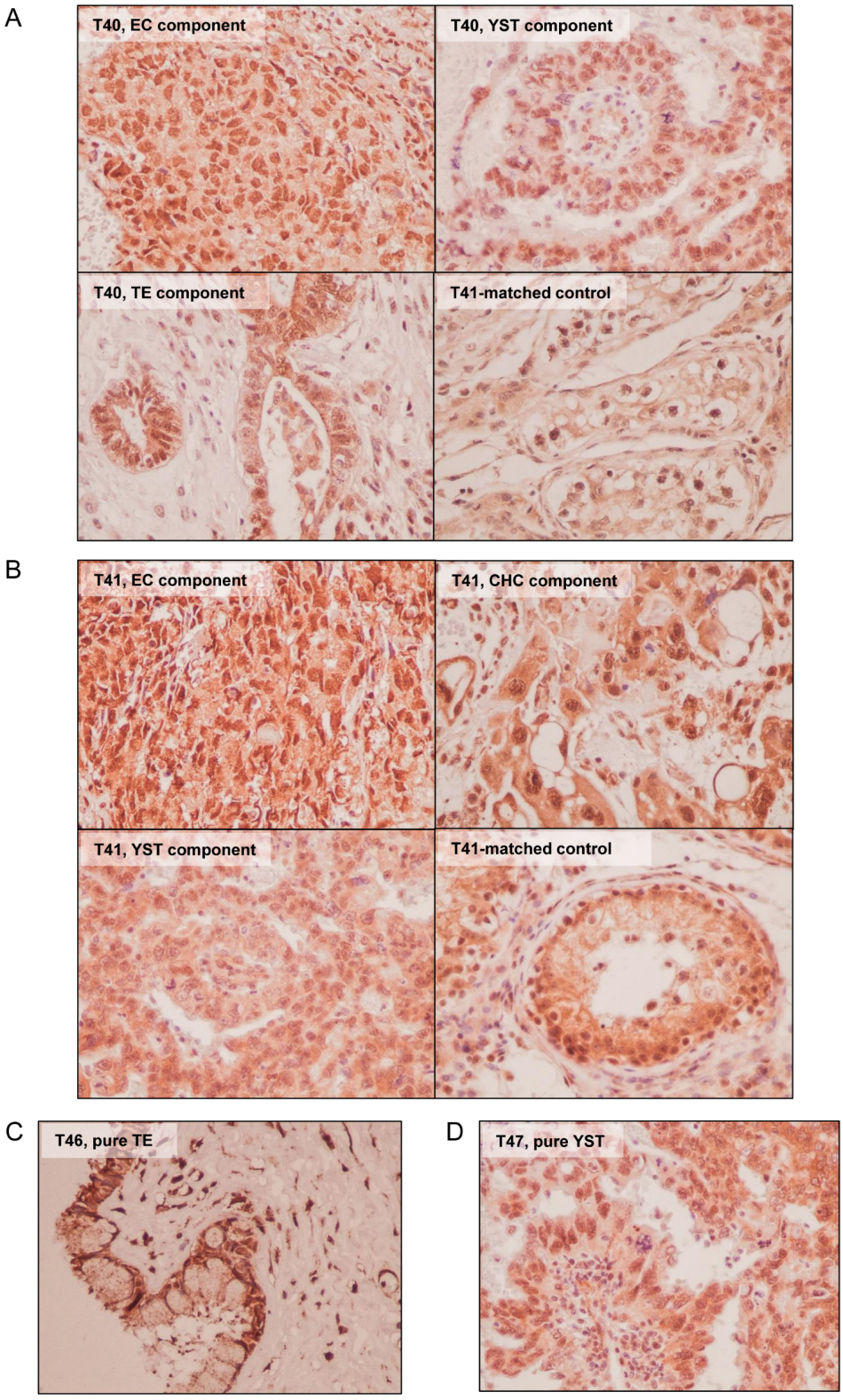

Fig. 22: TET1 IHC staining in selected GCTs. *(A) TET1 IHC staining in the T40 sample (mixed GCT - 40% YST (yolk sac tumor), 35% EC (embryonal carcinoma), 25% TE (teratoma)). Original magnification was 400x. (B) TET1 IHC staining in the T41 sample (mixed GCT – 60% CHC (choriocarcinoma), 30 % EC, 10% YST). Original magnification was 400x. (C) TET1 IHC staining in the T46 sample (pure TE). Original magnification was 400x. (D) TET1 IHC staining in the T47 sample (pure YST). Original magnification was 400x. (Benešová, Trejbalová, Kučerová, et al. 2017)*

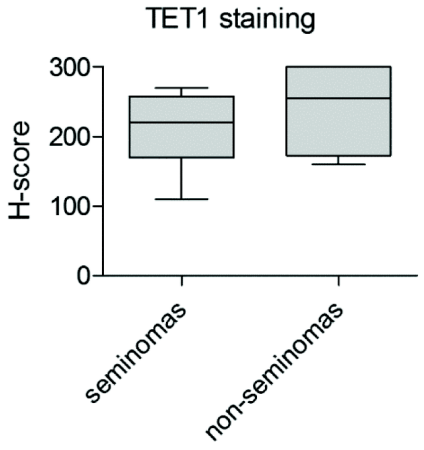

Fig. 23: Comparison of TET1 IHC staining intensity between the seminomas and non-seminomas expressed by H-score. *For the description of TET1 staining intensity, H-score was used (Y-axis). On the X-axis, the analyzed group is indicated. Medians with minimal and maximal values are depicted in the plot. (Benešová, Trejbalová, Kučerová, et al. 2017)*

Table 4: TET1 IHC in selected a sub-set of samples. *The TET1 staining intensity is designated as: 0 (negative), + (weak), ++ (moderate), +++ (strong). (Benešová, Trejbalová, Kučerová, et al. 2017)*

Tumor sample	Type of tumor	Nuclear TET1 staining intensity
T1	pure SE	++/+++
T2	pure SE	++
T3	pure SE	+/++
T4	pure SE	++
T7	pure SE	++/+++
T15	pure SE	+/++
T18	pure SE	++/+++
T22	pure SE	++/+++
T34	mixed GCT – 95% SE, 5% EC	+++
T40	mixed GCT – 40% YST, 35% EC, 25% TE	++/+++
T41	mixed GCT – 60% CHC, 30% EC, 10% YST	+++
T42	mixed GCT – 40% CHC, 30% EC, 30% YST	+/++
T43	pure EC	+++
T44	pure EC	+++
T46	pure TE	++
T47	pure YST	++
T55	non-Hodgkin lymphoma	+

1.4 *TET2* and *TET3* mRNA expression in GCTs

As we found out *TET1* mRNA to be overexpressed in the seminomas and to a lesser extent in non-seminomas and all three members of the TET family (TET1, TET2 and TET3) are able to oxidize 5mC to 5hmC (Tahiliani et al. 2009; Ito et al. 2010), we further analyzed *TET2* and *TET3* mRNA production in our samples. For the expression analysis, the qRT-PCR method was applied.

We found both *TET2* and *TET3* mRNA to be overexpressed in the seminomas in comparison to the matched controls ($p < 0.001$ and $p < 0.0001$, respectively, Fig. 24A and B, Fig. 25A and B) and further with the non-GCT controls ($p < 0.0001$ for both, Fig. 24A and B). The mixed GCTs displayed significantly higher *TET2* mRNA expression when compared with the non-GCT controls ($p < 0.01$, Fig. 24A). *TET3* mRNA production was significantly upregulated in the mixed GCTs in contrast to both the matched and non-GCT controls ($p < 0.05$ and $p < 0.001$, respectively, Fig. 24B). Moreover, *TET2* as well as *TET3* mRNA expression was higher in the seminomas in comparison to the non-seminomas ($p < 0.05$ and $p < 0.01$, respectively, Fig. 25A and B). TCam-2 and BeWo cell lines displayed high levels of both *TET2* and *TET3* mRNA (Fig. 24A and B). Interestingly, the endometrial carcinomas, lymphomas, the placenta and PBMC showed comparable or even higher levels of *TET2* and *TET3* mRNA in comparison to both the seminomas and mixed GCTs (Fig. 24A and B).

Overall, we observed elevated *TET2* and *TET3* mRNA production in both the seminomas and non-seminomas when compared with controls. However, *TET2* and *TET3* mRNA expression was much lower than that of *TET1* in the GCTs. On the other hand, both *TET2* and *TET3* mRNA production was elevated in the non-GCTs when compared with *TET1* mRNA expression. As *TET1* mRNA was highest in GCTs we focused on TET1 dioxygenase in further experiments.

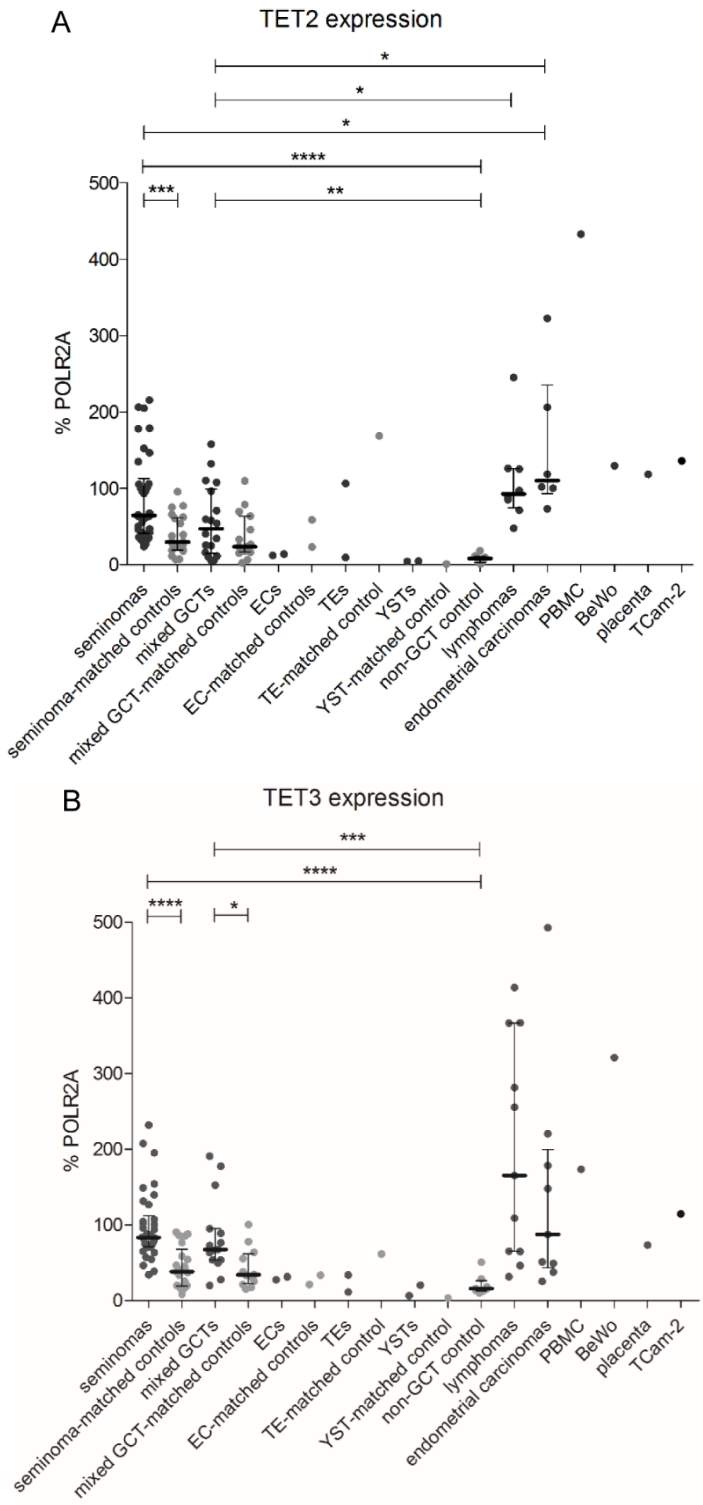

Fig. 24: TET2 and TET3 mRNA expression in analyzed samples. *(A) TET2 and (B) TET3 mRNA expression was measured by qRT-PCR and normalized to the percentage of POLR2A expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, the placenta, and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the absolute amount of mRNA normalized to the percentage of POLR2A. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kučerová, et al. 2017)*

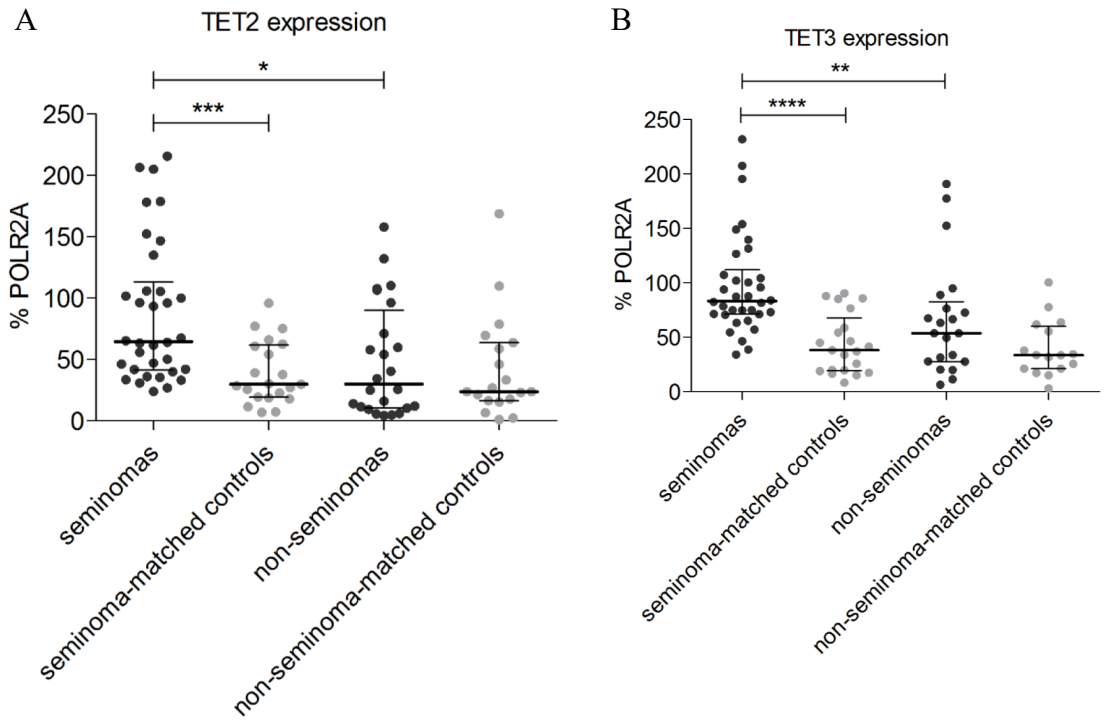

Fig. 25: Comparison of TET2 and TET3 mRNA expression in the seminomas, the non-seminomas and the matched controls. *(A) TET2 and (B) TET3 mRNA expression is shown. The level of mRNA was measured by qRT-PCR and normalized to the percentage of POLR2A expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. The Y-axis represents the absolute amount of the mRNA normalized to the percentage of POLR2A. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kučerová, et al. 2017)*

1.5 *IDH1* and *IDH2* mutations in seminomas

The presence of specific mutations in *IDH1* and *IDH2* genes (Yan et al. 2009) leads to the production of 2-hydroxyglutarate (W. Xu et al. 2011). This oncometabolite is structurally very similar to α-ketoglutarate, an essential cofactor of TET enzymes (Dang et al. 2009; Tahiliani et al. 2009). 2-Hydroxyglutarate is a weak antagonist of α-ketoglutarate, which means that is able to block the function of TET enzymes. To exclude the possibility that TET proteins can be negatively regulated by the presence of 2-hydroxyglutarate in our samples, we analyzed the most common *IDH1* and *IDH2* mutations in a selected sub-set of seminomas (*IDH1R132*, *IDH2R140* and *IDH2R172* mutations in seminomas T1 to T4). We did not observe any of the above-mentioned *IDH1* or *IDH2* mutations in the examined samples (data not shown).

1.6 *DNMT1, DNMT3A* and *DNMT3B* mRNA expression in GCTs

The level of DNA methylation is regulated by the interplay between TET demethylation enzymes and DNMTs. To understand the processes of both DNA methylation level establishment and erasure more deeply, we analyzed mRNA expression of DNMTs (*DNMT1*, *DNMT3A* and *DNMT3B)* in a selected sub-set of samples. For the analysis, the qRT-PCR method was applied.

We observed similar levels of *DNMT1* mRNA expression both in the analyzed GCTs and in controls (Fig. 26A). The investigation of *DNMT3A* mRNA production revealed significantly elevated levels in the seminomas and non-seminomas in comparison to the matched controls ($p < 0.05$ and $p < 0.01$, respectively, Fig. 26B). Furthermore, *DNMT3A* mRNA expression was upregulated in the seminomas in comparison to both non-GCT controls and non-seminomas ($p < 0.05$ and $p < 0.05$, respectively, Fig. 26B). Next, we analyzed *DNMT3B* mRNA production. We observed significantly elevated *DNMT3B* mRNA expression in the seminomas in comparison to the matched controls ($p < 0.05$, Fig. 26C). The highest level of *DNMT3B* mRNA was observed in two pure embryonal carcinomas (T44 and T45) and BeWo cell line (Fig. 26C).

To summarize, our research showed elevated mRNA expression of both *DNMT3A* and *DNMT3B* (*de novo* methyltransferases) but not *DNMT1* in the seminomas in comparison to the matched controls. The non-seminomas displayed upregulated levels of *DNMT3A* mRNA when compared with the matched controls. Our results manifested that both TET enzymes and DNMTs (DNMT3A and DNMT3B) are possibly able to shape the level of DNA methylation in the seminomas.

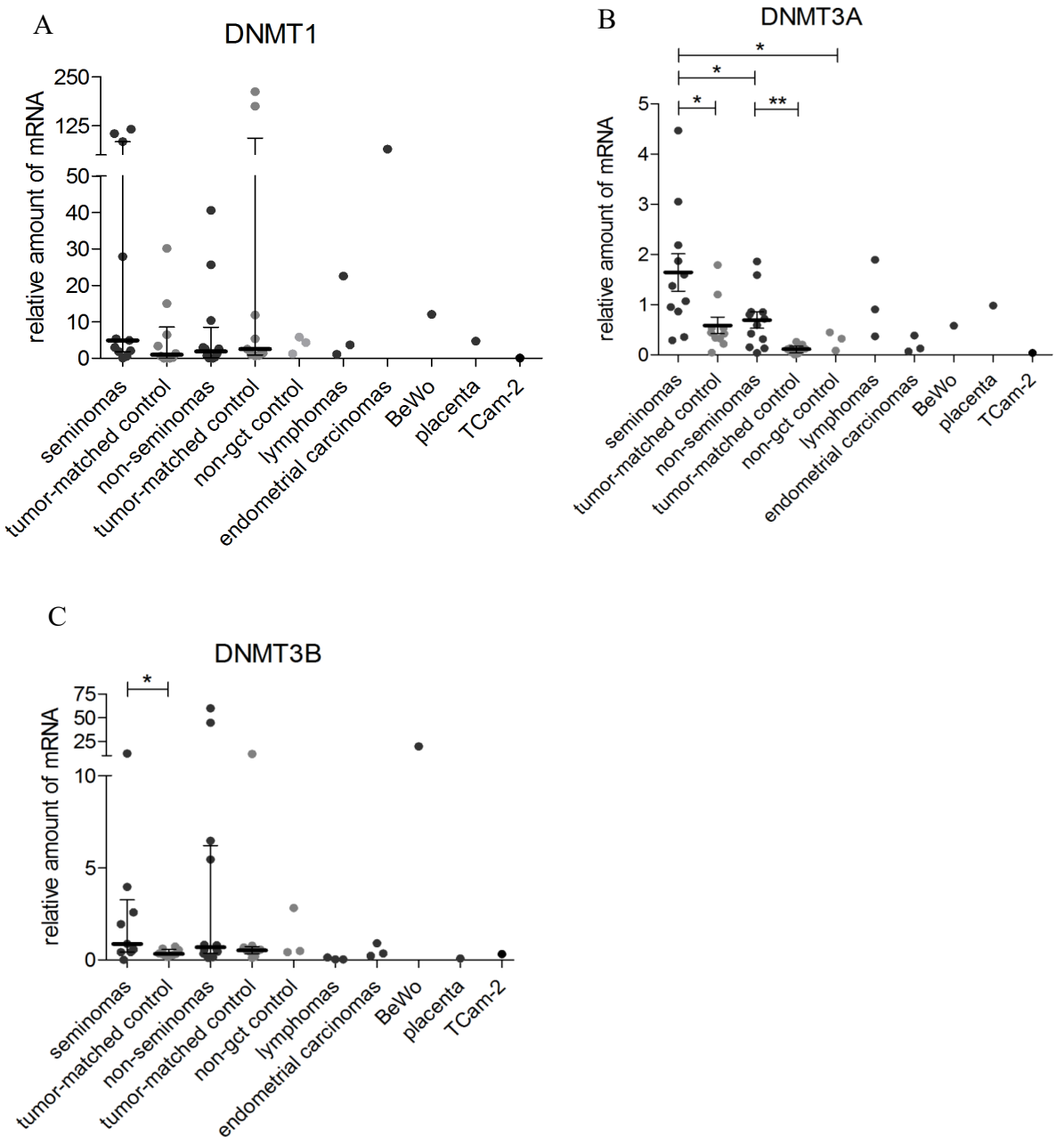

Fig. 26: DNMT1, DNMT3A and DNMTB mRNA expression in a selected sub-set of samples. *The relative amount of (A) DNMT1, (B) DNMT3A and (C) DNMT3B mRNA was measured by qRT-PCR. The amount was normalized to POLR2A expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, the placenta and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the relative amount of mRNA. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kučerová, et al. 2017)*

1.7 Global levels of 5mC and 5hmC in GCTs

TET enzymes are able to transform 5mC to 5hmC (Tahiliani et al. 2009; Ito et al. 2010). Therefore, we further explored the global level of both 5mC and 5hmC using the LC/SM technique in a selected sub-set of samples.

We found that seminomas were significantly hypomethylated (5mC) in comparison to the matched controls (median 1.5 % and 3.5 % of dG, respectively, $p < 0.05$ and $p < 0.05$, respectively, Fig. 27A). We did not observe any significant difference in the 5mC level in mixed GCTs when compared with the matched controls (median 2.2 % and 3.1 % of dG respectively, Fig. 27A). Both the seminomas and mixed GCTs displayed significantly lower global levels of 5hmC in comparison to the matched controls ($p < 0.05$, Fig. 27B). The mixed GCTs showed a slightly elevated level of 5hmC than the seminomas (Fig. 27B). Interestingly, that two of analyzed mixed GCTs (T40 and T42), which showed the highest level of both 5mC and 5hmC, did not contain the seminoma component (Fig. 28). The analyzed lymphoma sample (T55), TCam-2, BeWo and HeLa cell lines displayed relatively high levels of 5mC (Fig. 27A and Fig. 28), but relatively low levels of 5hmC (Fig. 27B and Fig. 28). The placenta sample showed high 5mC and 5hmC levels (Fig. 27A, B and Fig. 28). In all analyzed tumor-matched controls, we observed high levels of both 5mC and 5hmC (Fig. 27A, B and Fig. 28).

Collectively, we showed that the seminomas manifested low 5mC and 5hmC levels. Moreover, the two mixed GCT samples with the highest level of 5mC and 5hmC did not contain the seminoma component. The low level of analyzed cytosine modifications is one characteristic of the seminoma tumor.

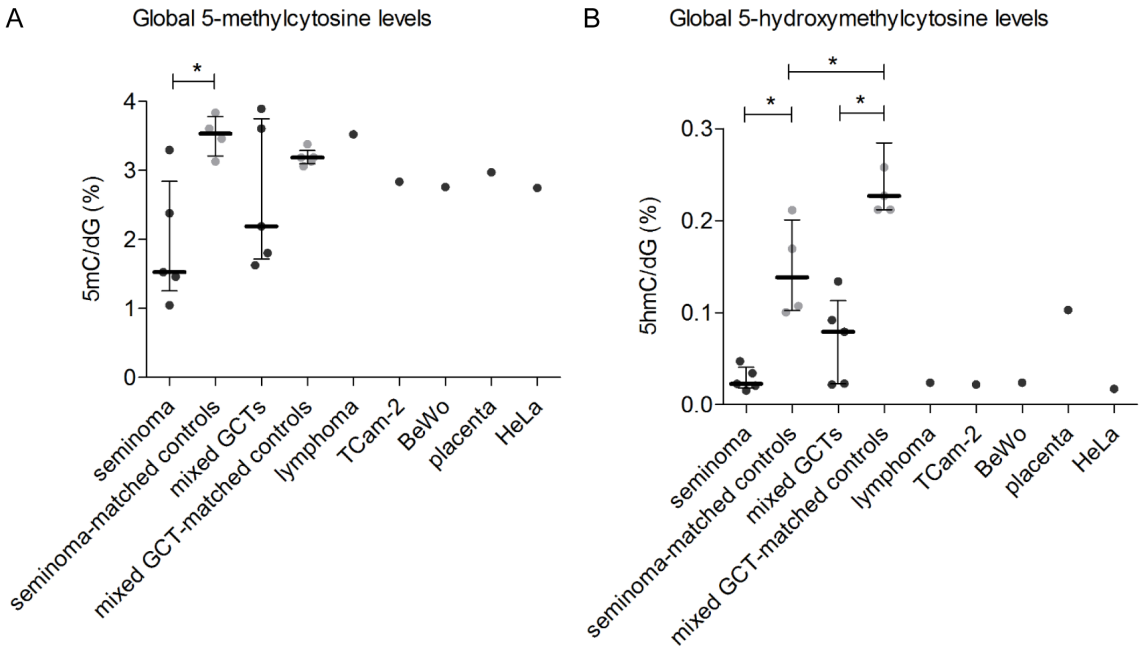

Fig. 27: Global level of 5mC and 5hmC in a selected sub-sets of samples. *The analysis was performed by the LC/MS technique. We measured the percentage of both (A) 5mC and (B) 5hmC as a portion of dG in the samples. On the Y-axis, the percentage 5mC/dG and the percentage 5hmC/dG (respectively) are depicted. On the X-axis, the analyzed group of tumors/samples is presented. Each sample was measured in technical duplicate, averaged, and median values with the interquartile range are depicted. Each sample is represented by a dot in the dot-plot. Significance (P-value) between the analyzed groups/samples is represented by bars with asterick(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kučerová, et al. 2017)*

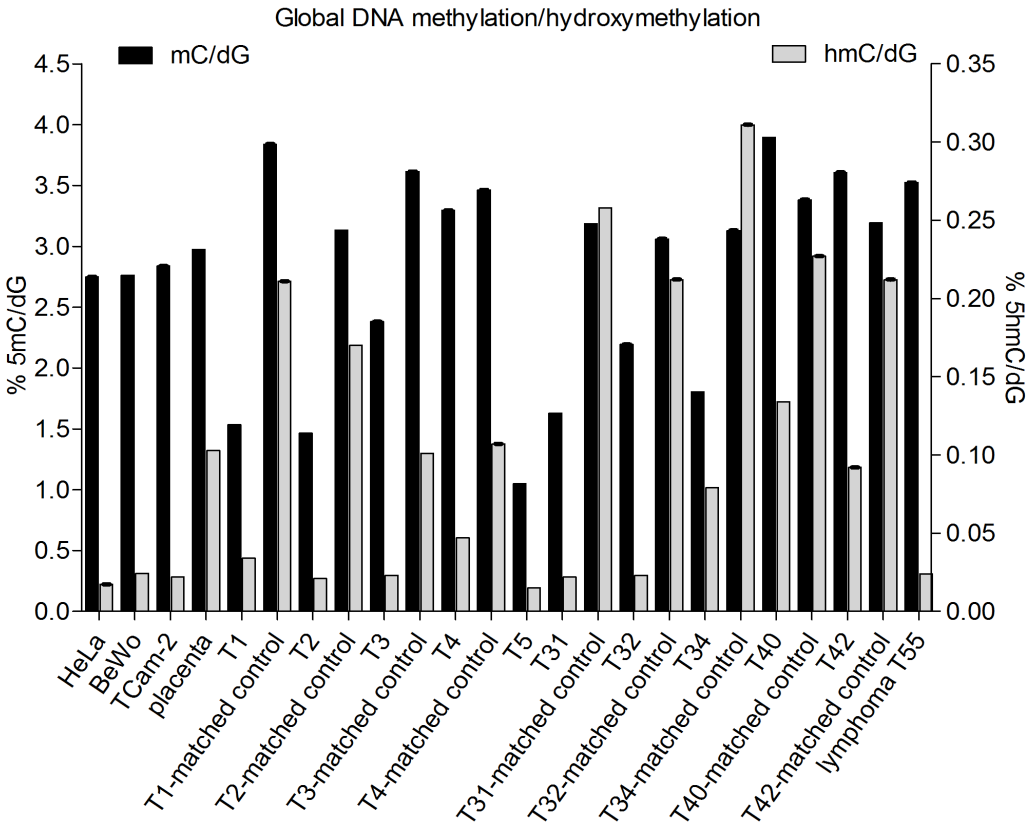

Fig. 28: Global level of 5mC and 5hmC in a selected sub-set of samples. *We analyzed the global level of 5mC and 5hmC in the seminomas (T1 to T5), mixed GCTs (T31, T32, T34, T40, T42), and in the matched controls, the placenta sample, lymphoma (T55), HeLa, BeWo and TCam-2 cell lines. Mixed GCT samples T40 and T42 did not contain the seminoma component; T31, T32 and T34 had variable amounts of the seminoma component. The Y-axis represents the percentage of 5mC/dG (left Y-axis, black bars) and 5hmC/dG (right Y-axis, grey bars). On the X-axis, the analyzed samples are depicted. (Benešová, Trejbalová, Kučerová, et al. 2017)*

1.8 5hmC IHC analysis

In our previous experiments, we used the LC/MS technique to analyze the overall amount of 5hmC in our specimens. To explore the presence of 5hmC in tissue sections, we employed 5hmC IHC in a selected sub-set of samples.

We found that Sertoli cells manifested strong 5hmC staining during normal spermatogenesis. In GCNIS, we observed weak to negative staining (GCNIS is indicated by the black arrow, Fig. 29A). Overall, we analyzed eight seminoma samples (T1, T2, T3, T7, T10, T15, T18 and T22). Seven of them displayed weak or negative 5hmC staining intensity, but seminoma sample T22 showed moderate staining (representative microphotographs, Fig. 29 B, C and Table 5). On the other hand, four analyzed mixed GCTs (T34, T40, T41 and

T42) had moderate to strong 5hmC positivity in different tumor components. 5hmC staining in these mixed GCTs was higher or comparable with the matched controls (Fig. 30A, B and Table 5). The mixed GCTs with the seminomatous component displayed variable level of 5hmC staining. Two of them, T30 and T34 manifested strong 5hmC positivity (Table 5). In contrast, other three analyzed mixed GCTs (T31, T32 and T33) displayed weak 5hmC staining (Table 5). Two pure embryonal carcinomas (T43 and T44) showed strong 5hmC positivity, with weak staining in the matched controls (Fig. 29D). Pure teratoma sample (T46) had moderate to strong 5hmC positivity (Fig. 30C). Pure yolk sac tumor (T47) manifested strong 5hmC positivity (Fig. 29E). Weak 5hmC staining was found in one analyzed lymphoma sample (T55) (Fig. 31). 5hmC staining described above applied only to the tumor cells of the examined samples. The lymphoid infiltrate, that is present in seminomas displayed high 5hmC staining in the nuclei (representative microphotograph, Fig. 29C).

5hmC IHC staining revealed that the seminomas displayed lower levels of 5hmC when compared with the matched controls. This is in accordance with the data obtained by the LC/MS technique (Chapter 1.7 in the Results). On the other hand, the mixed GCTs displayed elevated level of 5hmC, when compared with the matched controls. The seminomas showed significantly lower levels of 5hmC staining in comparison to the non-seminomas as a whole group ($p < 0.01$, Fig. 32).

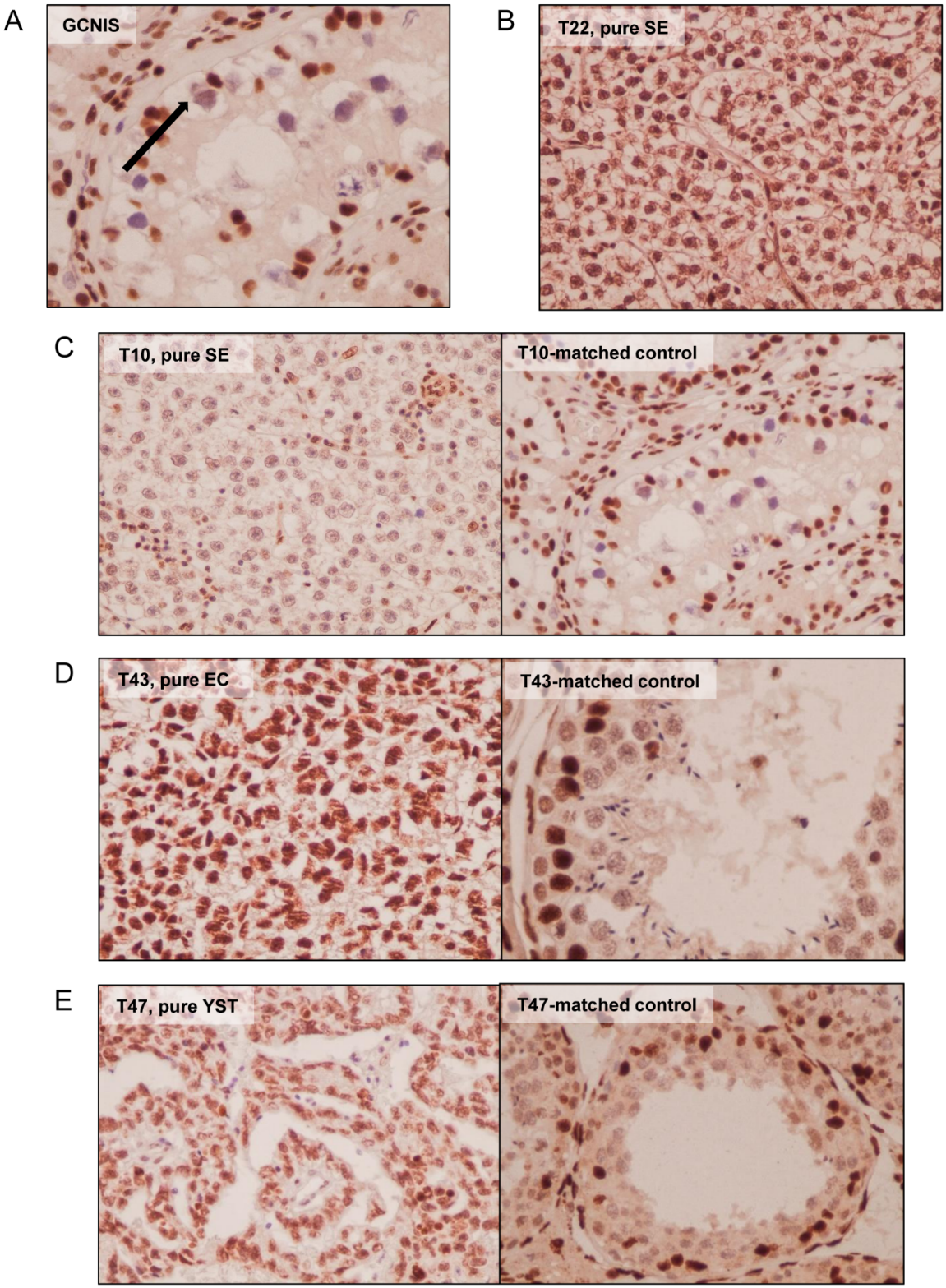

Fig. 29: 5hmC IHC staining in a selected sub-set of samples. *(A) 5hmC staining in GCNIS. Original magnification was 600x. (B) Seminoma sample T22. Original magnification was 400x. (C) Seminoma sample T10 and the matched control. Original magnification was 400x. (D) Pure embryonal carcinoma T43 and the matched control. Original magnification was 400x for T43 and 600x for T43-matched control. (E) Pure yolk sac tumor T47 and the matched control. Original magnification was 400x. (Benešová, Trejbalová, Kučerová, et al. 2017)*

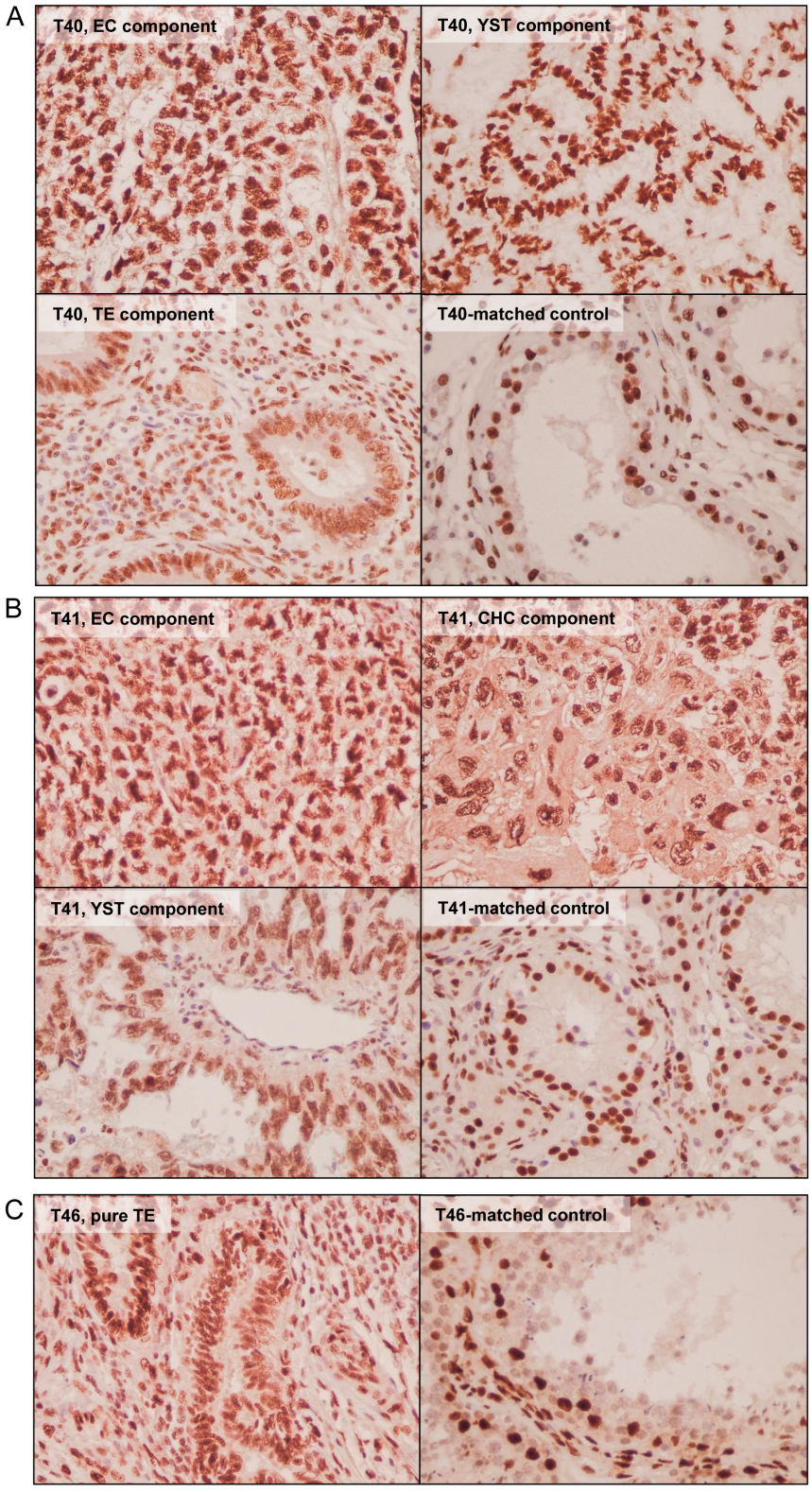

Fig. 30: 5hmC IHC staining in a selected sub-set of samples. *(A) Mixed GCT sample T40. Pictures of all three T40 components (embryonal carcinoma, yolk sac tumor and teratoma) and the matched control are depicted. Original magnification was 400x. (B) Mixed GCT sample T41. Pictures of all three T41 components (embryonal carcinoma, choriocarcinoma and yolk sac tumor) and the matched control are depicted. Original magnification was 400x. (C) Pure teratoma sample T46 and the matched control. Original magnification was 400x. (Benešová, Trejbalová, Kučerová, et al. 2017)*

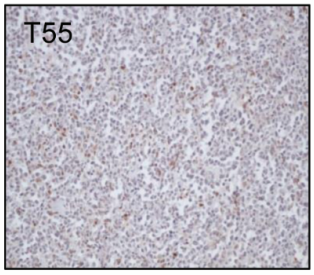

Fig. 31: 5hmC IHC staining in lymphoma T55. *Original magnification was 200x. (Benešová, Trejbalová, Kučerová, et al. 2017)*

Table 5: 5hmC IHC staining in a selected sub-set of tumor samples. *The table shows the name of the analyzed tumor, tumor description, and 5hmC staining intensity: 0 (negative), + (weak), ++ (moderate), +++ (strong). (Benešová, Trejbalová, Kučerová, et al. 2017)*

Tumor sample	Type of tumor	5hmC staining
T1	pure SE	+
T2	pure SE	0
T3	pure SE	+
T4	pure SE	+
T7	pure SE	0
T10	pure SE	+
T15	pure SE	+
T18	pure SE	0/+
T22	pure SE	++
T30	mixed GCT – 80% SE, 20% EC	+++
T31	mixed GCT – 60 – 70% SE, 25 – 30% TE, 15% EC	+
T32	mixed GCT – 80% SE, 20% EC	0/+
T33	mixed GCT – 95% SE, 5% EC	+/++
T34	mixed GCT – 95% SE, 5% EC	+++
T40	mixed GCT – 40% YST, 35% EC, 25% TE	++/+++
T41	mixed GCT – 60% CHC, 30% EC, 10% YST	+++
T42	mixed GCT – 40% CHC, 30% EC, 30% YST	+++
T43	pure EC	+++
T44	pure EC	+++
T46	pure TE	++/+++
T47	pure YST	++
T55	non-Hodgkin lymphoma	++

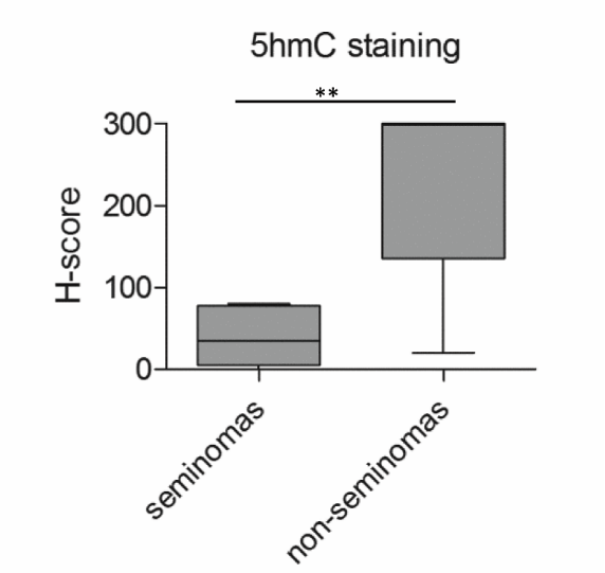

Fig. 32: Comparison of 5hmC IHC staining between the seminomas and non-seminomas expressed by H-score. *The Y-axis represents the H-score, the X-axis shows the analyzed group of samples. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kučerová, et al. 2017)*

1.9 TET1 knock-down in TCam-2 cells

Observed evelated *TET1* mRNA in seminoma samples led us to analysis of the overall impact of TET1 activity in seminomas. We performed TET1 knock-down in the TCam-2 seminoma-derived cell line to explore the TET1 effect on the 5hmC level. For the knock-down, we used the pool of siRNAs targeting TET1 (siTET1) and non-targeting siRNAs (siCTRL) as a control. Two transfections, on day 0 and day 3, were performed. The first harvest of the cells was performed three days post transfection, before the next round of transfection was done. Further on, six days post transfection, the cells were harvested again. The amount of *TET1* mRNA was controlled by ddPCR and thereafter the level of 5mC and 5hmC was measured by the LC/MS technique.

We observed down-regulation of *TET1* mRNA expression both three and six days after the siRNA transfection (26 % and 28 %, respectively) in comparison to siCTRL-transfected cells (Fig. 33A). Analysis of the 5mC and 5hmC content revealed that TET1 knock-down affected especially the level of 5hmC, which was decreased both three and six days after siTET1 transfection (Fig. 33B). The level of 5hmC in siCTRL-transfected TCam-2 cells was comparable to the non-transfected ones (Fig. 3B). The level of 5mC was comparable in all transfected cells (Fig. 33B). To be sure that the decreased level of 5hmC was induced only by TET1 knock-down, we further analyzed the production of *TET2* and *TET3* mRNA by qRT-PCR. In the control experiment, we did not observe any changes in *TET2/TET3* mRNA

expression either in non-transfected TCam-2 cells or in TET1 knock-downs (data not shown).

Collectively, our data showed that TET1 knock-down did not affect the global level of 5mC, but led to a decreased level of 5hmC in the seminoma-derived TCam-2 cell line.

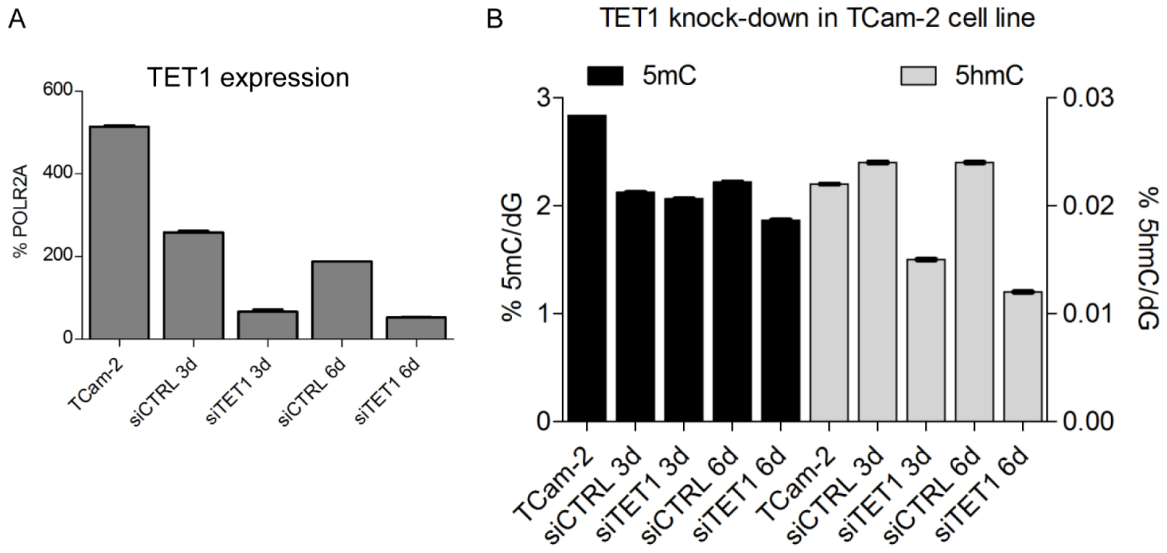

Fig. 33: Analysis of TET1 knock-down in TCam-2 cells. *(A) TET1 mRNA expression measured by RT-ddPCR in non-transfected TCam-2 cells, siCTRL and siTET1-transfected TCam-2 cells. The obtained amount was normalized to the percentage of POLR2A. Each sample was measured in technical duplicate, averaged, and the mean with standard deviation (+/-) was calculated. On the Y-axis, the percentage of POLR2A is depicted. The X-axis represents the analyzed samples. (B) The global level of 5mC (black bars) and 5hmC (grey bars) in transfected TCam-2 cells analyzed by the LC/MS technique. Each sample was measured in technical triplicate, averaged, and the mean value with standard deviation (+/-) was calculated. Left Y-axis represents the percentage of 5mC/dG, right Y-axis shows the percentage of 5hmC/dG. On the X-axis, analyzed samples are depicted. (Benešová, Trejbalová, Kučerová, et al. 2017)*

In the first part of the study, we explored the selected epigenetic features of tumors derived from PGCs, testicular germ cell tumors. We showed that *TET1* mRNA expression is elevated mainly in seminomas, and to a lesser extent in other GCTs. Furthermore, seminomas displayed very low global levels of both 5mC and 5hmC. Our laboratory and others have previously shown that human endogenous retrovirus ERVWE1 is expressed in testicular tumors (Gimenez et al. 2010; Trejbalová et al. 2011). This finding was intriguing, because *ERVWE1* is silenced in most of the human tissues, except for the placenta (Matoušková et al. 2006; Trejbalová et al. 2011). One of the crucial factors that are known to silence *ERVWE1* transcription is DNA methylation (Gimenez et al. 2009; Trejbalová et al. 2011). These facts together with the changes in both *TET1* mRNA expression and the level of 5mC and 5hmC led us to the analysis of *ERVWE1* expression in seminomas and other GCTs.

2 DNA hypomethylation and aberrant expression of the human endogenous retrovirus ERVWE1/syncytin-1 in seminomas

2.1 *ERVWE1* mRNA expression in GCTs

Human endogenous retrovirus ERVWE1 was previously described to be expressed in testicular tumors (Gimenez et al. 2010; Trejbalová et al. 2011). Nevertheless, small panels of samples were used in these studies and further on, the types of testicular tumors were not specified. Therefore, we collected a larger panel of samples and performed *ERVWE1* mRNA expression analysis. There are three types of *ERVWE1* transcripts (Fig. 15 in Chapter 3.1.4.3 in the Literature review) (Blond et al. 1999; Mi et al. 2000; Smallwood et al. 2003; Trejbalová et al. 2011) however, the Syncytin-1 protein can be translated only from the spliced *ERVWE1* mRNA (Mi et al. 2000; Smallwood et al. 2003). We analyzed production of both full-length and spliced forms of *ERVWE1* transcript by the qRT-PCR method.

We found full-length *ERVWE1* RNA to be significantly elevated in the seminomas in comparison to both the matched and non-GCT controls ($p < 0.0001$ and $p < 0.0001$, respectively, Fig. 34A). The level of *ERVWE1* RNA in the seminomas was higher than in TCam-2 and BeWo cell lines (Fig. 34A). Moreover, *ERVWE1* RNA was significantly elevated in the seminomas in comparison to the non-seminomas ($p < 0.001$, Fig. 34A). In the non-seminomas, we observed a different situation. The level of full-length *ERVWE1* RNA was comparable to the matched controls (Fig. 34A). On the other hand, full-length *ERVWE1* RNA expression in the non-seminomas was significantly higher than in the non-GCT controls ($p < 0.05$, Fig. 34A). Among the non-seminoma samples, T43 (40 % choriocarcinoma, 30 % embryonal carcinoma, 30 % yolk sac tumor), T32 (80 % seminomas, 20 % embryonal carcinoma) and T33 (60-70 % seminoma, 25-30 % teratoma, 15 % embryonal carcinoma) displayed the highest level of full-length *ERVWE1* RNA. Endometrial carcinomas and lymphomas showed significantly lower expression of full-length *ERVWE1* RNA than the seminomas ($p < 0.0001$ and $p < 0.05$, respectively, Fig. 34A). Among the lymphoma samples, we analyzed both Hodgkin and non-Hodgkin lymphomas. We did not observe any differences between these two groups of tumors (Fig. 35). As we expected, placentas displayed high levels of *ERVWE1* full-length RNA (Fig. 34A). *ERVWE1* full-length RNA expression was normalized not only to the *POLR2A* housekeeping gene (Fig. 34A), but also to *TBP* (Fig. 34B). We did not observe any differences in *ERVWE1* full-length RNA production.

Overall, we found that the seminomas showed significantly higher levels of *ERVWE1*

full-length RNA than both used controls and non-seminomas as a whole group.

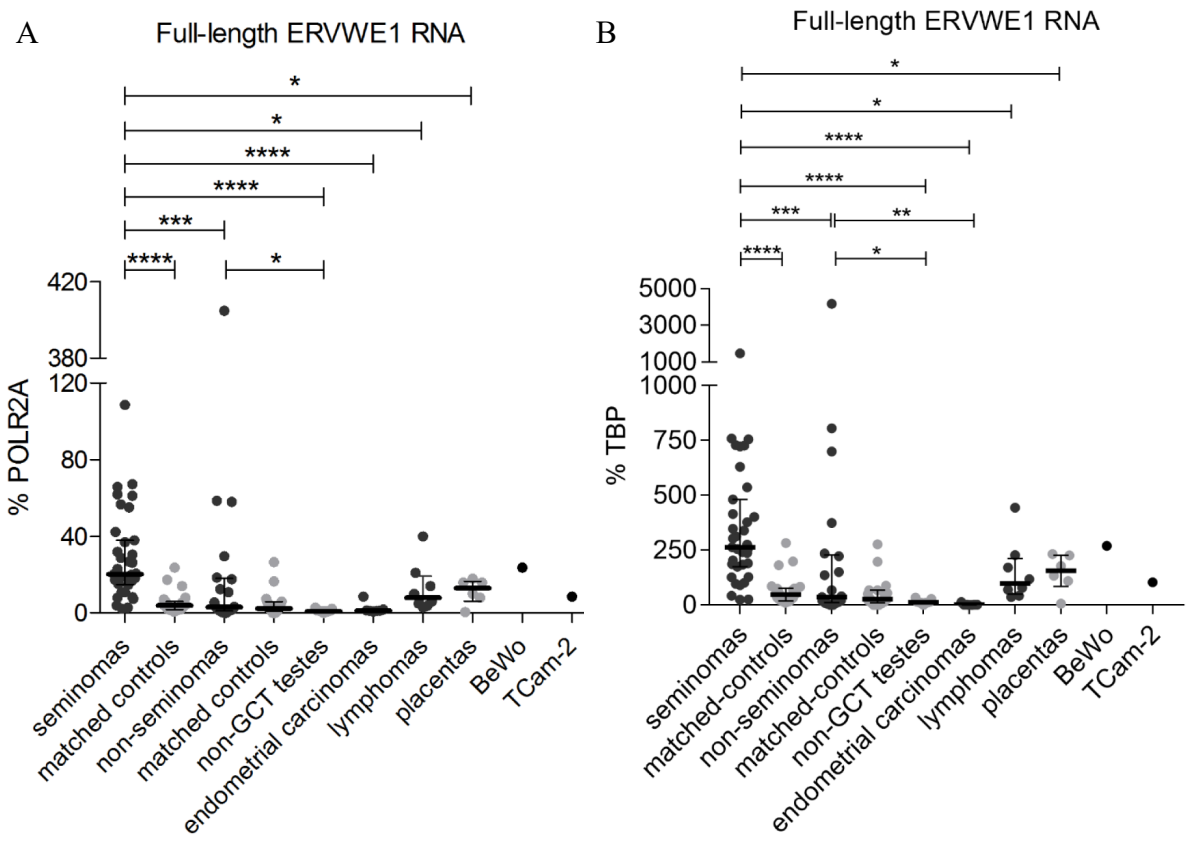

Fig. 34: Analysis of the full-length ERVWE1 RNA expression. *The absolute amount of the full-length ERVWE1 RNA was measured by qRT-PCR. (A) The obtained amount was normalized to the percentage of POLR2A expression. (B) Obtained amount was normalized to to the percentage of TBP housekeeping gene expression. Each sample was measured in technical triplicate, averaged, and then represented as a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, placentas and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the percentage of POLR2A/the percentage of TBP. On the X-axis, the analyzed groups of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

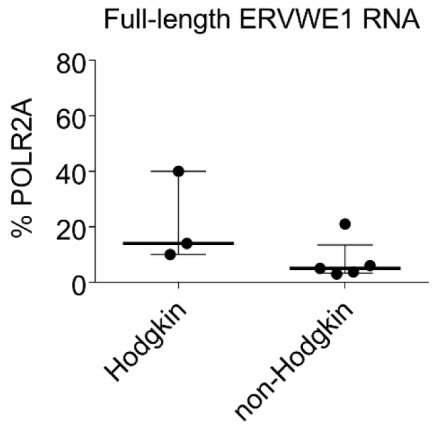

Fig. 35: Expression of the full-length ERVWE1 RNA in Hodgkin and non-Hodgkin lymphomas. *The absolute full-length ERVWE1 RNA production was measured by qRT-PCR. The obtained amount was normalized to the percentage of POLR2A expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. The Y-axis represents the percentage of POLR2A. On the X-axis, the analyzed groups of samples are depicted. (Benešová, Trejbalová, Kovářová, et al. 2017)*

2.2 *ERVWE1* splicing in GCTs

Syncytin-1 protein can be translated only from the spliced form of the *ERVWE1* RNA transcript (Mi et al. 2000; Smallwood et al. 2003), therefore we analyzed the amount of spliced *ERVWE1* mRNA by qRT-PCR.

We observed a significantly elevated level of spliced *ERVWE1* mRNA in the seminomas in comparison to the matched controls, non-GCT controls and non-seminomas ($p < 0.0001$, $p < 0.0001$ and $p < 0.0001$, respectively, Fig. 36A). BeWo cells and the placentas showed high levels of spliced *ERVWE1* mRNA (Fig. 36A). On the other hand, the non-seminomas as a whole group did not display significantly elevated levels of spliced *ERVWE1* mRNA in comparison to both the matched and non-GCT controls (Fig. 36A). However, a few samples from the analyzed non-seminomas manifested high production of spliced *ERVWE1* mRNA (Fig. 36A) – specifically samples T32 and T33 (80 % seminoma, 20 % embryonal carcinoma and 60-70 % seminoma, 25-30 % teratoma, 15 % embryonal carcinoma, respectively). The non-GCT controls, endometrial carcinomas and lymphomas displayed very low levels of spliced *ERVWE1* mRNA (Fig. 36A). Among the lymphoma samples, we did not observe any significant difference between the groups of Hodgkin and non-Hodgkin lymphomas (Fig. 37). The spliced *ERVWE1* mRNA expression was normalized not only to the *POLR2A* (Fig. 36A), but also to *TBP* housekeeping gene (Fig. 36B). We did not observe any differencies in the spliced *ERVWE1* mRNA amount while it was normalized either to *POLR2A* or to *TBP* housekeeping genes.

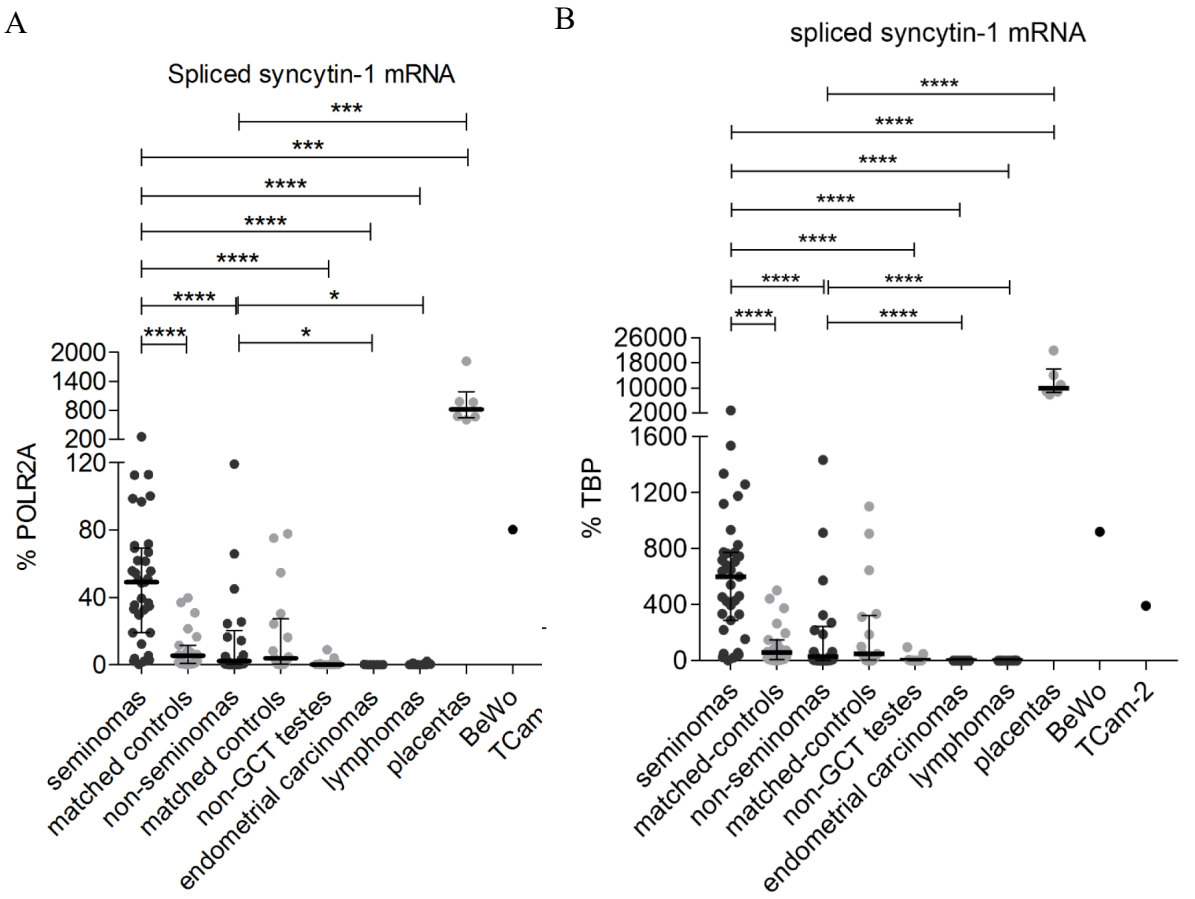

Fig. 36: Expression of the spliced ERVWE1 mRNA. *The absolute amount of the spliced ERVWE1 mRNA is depicted in the figure. (A) mRNA expression was measured by qRT-PCR and normalized (A) to the percentage of POLR2A expression, and (B) to the percentage of TBP expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, placentas and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the percentage of POLR2A/the percentage of TBP. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

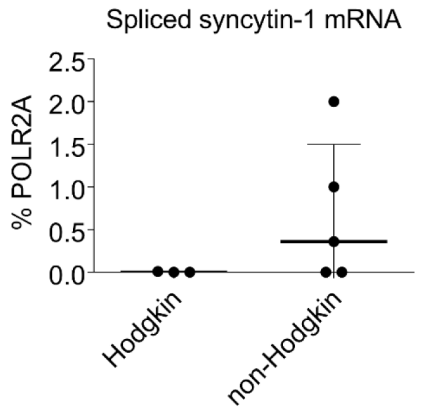

Fig. 37: Expression of the spliced syncytin-1 mRNA in Hodgkin and non-Hodgkin lymphomas. *Spliced syncytin-1 mRNA production was measured by qRT-PCR. The obtained amount was normalized to the percentage of POLR2A expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of the samples. The Y-axis represents the percentage of POLR2A. On the X-axis, the analyzed groups of samples are depicted. (Benešová, Trejbalová, Kovářová, et al. 2017)*

To decipher *ERVWE1* splicing more deeply, we calculated the splicing efficiency as a ratio of spliced *ERVWE1* mRNA to the sum of both spliced and full-length *ERVWE1* RNA in each analyzed sample. Even though we observed higher levels of *ERVWE1* mRNA in the seminomas in comparison to the matched-controls (Fig. 38), the obtained splicing efficiency did not differ significantly. However, the seminomas showed significantly elevated splicing efficiency in comparison to the non-GCT controls ($p < 0.01$, Fig. 38). Furthermore, we found significantly higher splicing efficiency in the seminomas in comparison to the non-seminomas (Fig. 38). The placentas, BeWo and TCam-2 cell lines showed high levels of *ERVWE1* mRNA splicing efficiency (Fig. 38).

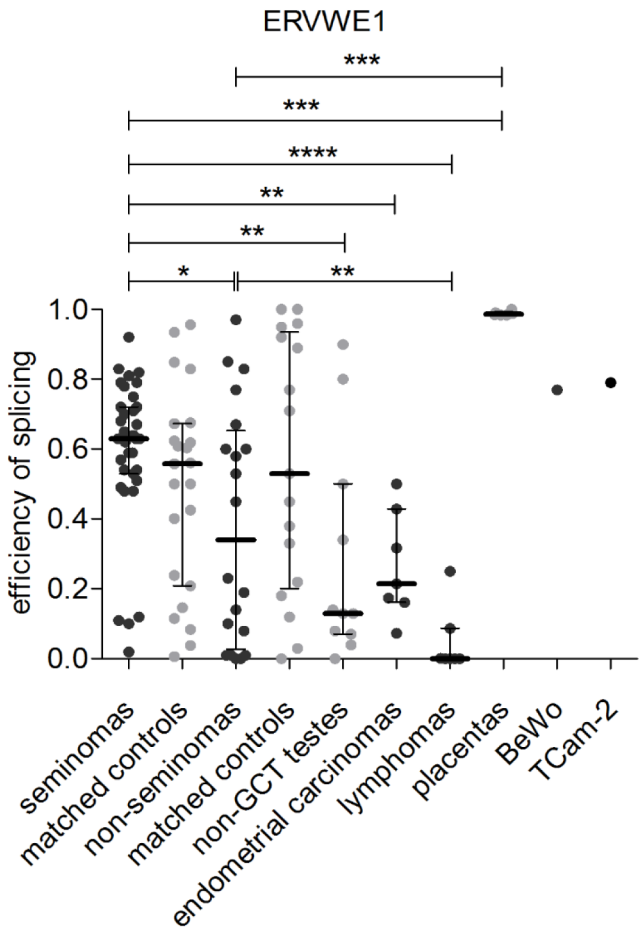

Fig. 38: Splicing efficiency of the ERVWE1 transcript. *Mean values with interquartile ranges were counted. Each sample is represented by a dot in the dot plot. Lymphomas, endometrial carcinomas, placentas and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the splicing efficiency. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

To summarize, we showed that the spliced *ERVWE1* mRNA was significantly elevated in the seminomas in comparison with both used controls and non-seminoma samples. Moreover, the seminomas displayed high splicing efficiency of the *ERVWE1* RNA.

2.3 *ERVFRDE1* expression in GCTs

ERVWE1 is not the only endogenous retrovirus that is able to express functional Env protein in humans. There is also ERVFRDE1, which encodes the fusogenic Syncytin-2 protein (Blaise et al. 2003). Moreover, Syncytin-2, together with Syncytin-1, is produced in the placenta and its participation with Syncytin-1 in the placenta formation is supposed (Mi et al. 2000; Malassiné et al. 2005). Furthermore, the expression from the *ERVFRDE1* gene is

regulated by promoter DNA methylation (Gimenez et al. 2009; Trejbalová et al. 2011). To determine whether *ERVFRDE1* mRNA is expressed in our samples, both *ERVFRDE1* full-length and spliced forms of transcripts were measured by qRT-PCR.

First, we analyzed full-length *ERVFRDE1* RNA expression. The level of full-length *ERVFRDE1* RNA was low in the analyzed samples (Fig. 39A). Nevertheless, expression in the seminomas was still significantly higher in comparison to both the matched and non-GCT controls ($p < 0.0001$ and $p < 0.0001$, respectively, Fig. 39A). Furthermore, the seminomas displayed significantly elevated level of full-length *ERVFRDE1* than the non-seminomas ($p < 0.001$, Fig. 39A). When we looked at the non-seminomas, there was no difference in *ERVFRDE1* full-length RNA production in comparison to the matched controls (Fig. 39A). However, *ERVFRDE1* full-length RNA was significantly higher in the non-seminomas when compared with non-GCT controls ($p < 0.05$, Fig. 39A). The highest expression of full-length *ERVFRDE1* mRNA was detected in the placentas, BeWo and TCam-2 cell lines (Fig. 39A). As other analyzed samples, the lymphomas showed relatively low expression of the full-length *ERVFRDE1* mRNA and we did not observe any difference between the groups of Hodgkin and non-Hodgkin lymphomas (Fig. 40). *ERVFRDE1* full-length RNA expression was normalized not only to the *POLR2A* (Fig. 39A), but also to *TBP* housekeeping gene (Fig. 39B). We did not observe any differences in *ERVFRDE1* full-length RNA production when it was normalized to either *POLR2A* or *TBP* housekeeping genes.

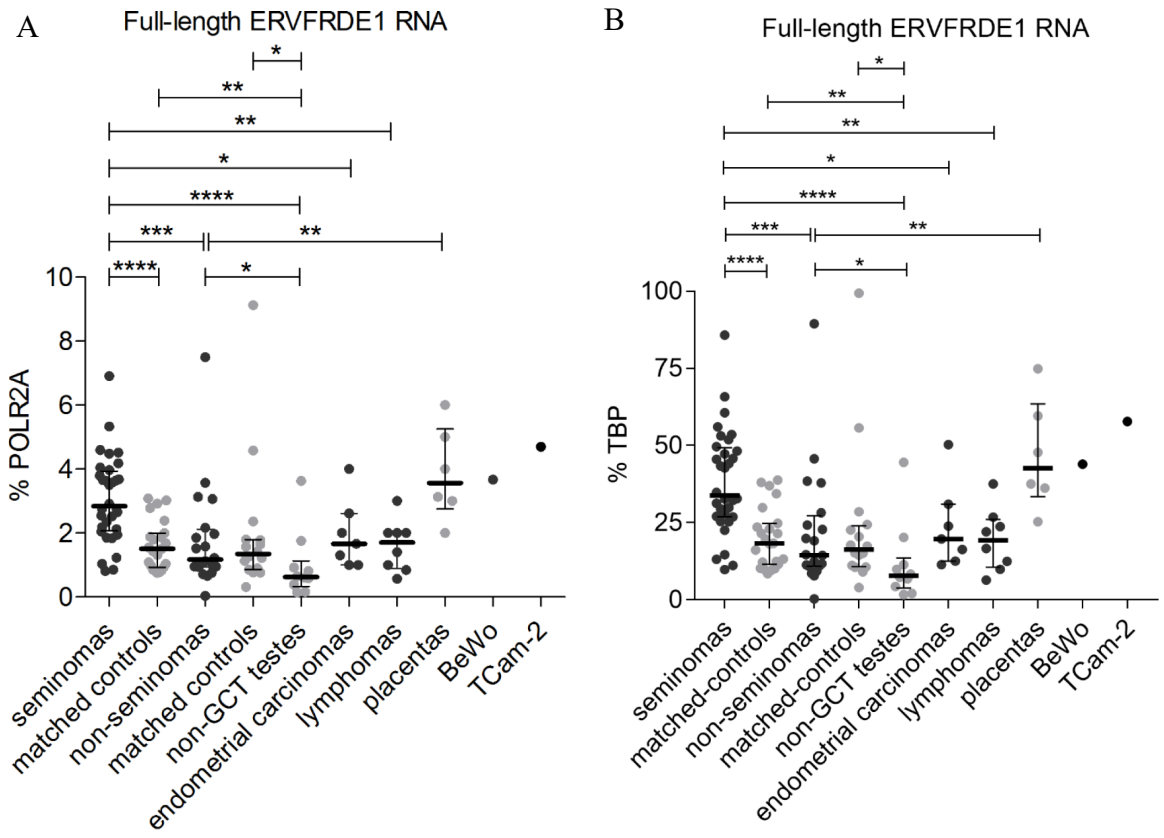

Fig. 39: ERVFRDE1 full-length RNA expression in analyzed samples. *For the analysis, the qRT-PCR method was used. The absolute amount of full-length ERVFRDE1 RNA was normalized (A) to the percentage of POLR2A expression and (B) to the percentage of TBP expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, placentas and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the percentage of POLR2A/the percentage of TBP. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

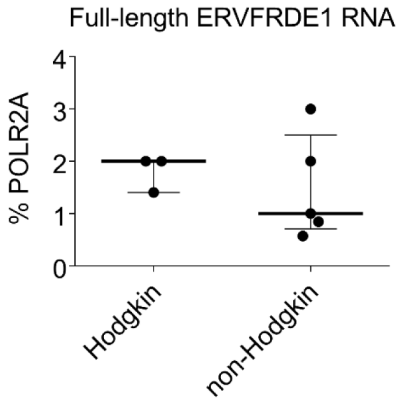

Fig. 40: Expression of the full-length ERVFRDE1 RNA in Hodgkin and non-Hodgkin lymphomas. *For the analysis, the qRT-PCR method was used. The absolute amount was normalized to the percentage of POLR2A expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. The Y-axis represents the percentage of POLR2A. On the X-axis, the analyzed group of samples are depicted. (Benešová, Trejbalová, Kovářová, et al. 2017)*

Further, we analyzed expression of the spliced *ERVFRDE1* mRNA. The highest production was observed in the placental samples (Fig. 41A). The seminomas displayed significantly higher levels of spliced *ERVFRDE1* mRNA in comparison with the non-GCT controls ($p < 0.05$), but not with the matched controls (Fig. 41A). Moreover, the seminomas displayed significantly higher production than the group of non-seminomas ($p < 0.05$, Fig. 41A). The expression measured in the non-seminomas was not significantly different in comparison to both the matched and non-GCT controls (Fig. 41A). Analysis of the lymphomas revealed that there was a difference in spliced *ERVFRDE1* mRNA expression between Hodgkin and non-Hodgkin lymphomas, with higher production in the non-Hodgkin lymphomas (Fig. 42). BeWo and TCam-2 cell lines showed average levels of spliced *ERVFRDE1* mRNA (Fig. 41A). The spliced *ERVFRDE1* mRNA expression was normalized not only to the *POLR2A* (Fig. 41A), but also to *TBP* housekeeping gene (Fig. 41B). We did not observe any differences in *ERVFRDE1* spliced mRNA production when it was normalized to either *POLR2A* or *TBP* housekeeping gene.

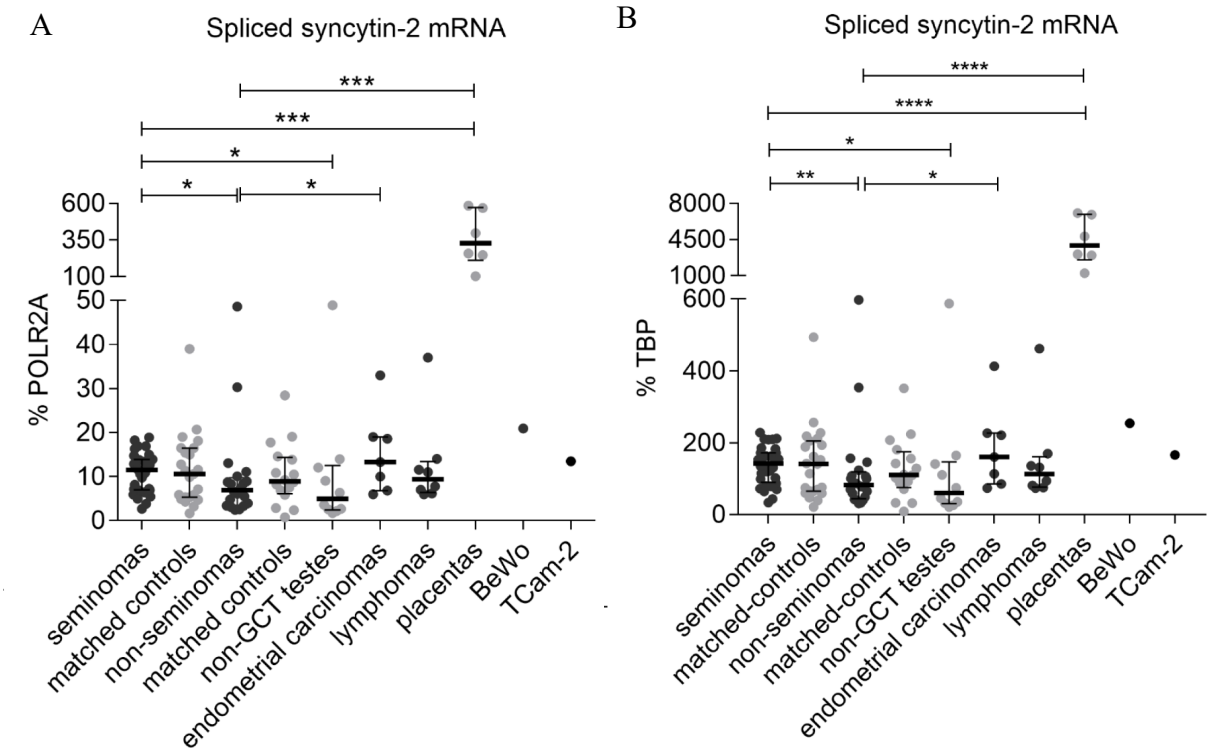

Fig. 41: ERVFRDE1 spliced mRNA expression in analyzed samples. *For the analysis, the qRT-PCR method was used. The absolute amount of spliced ERVFRDE1 mRNA was normalized (A) to the percentage of POLR2A and (B) to the percentage of TBP housekeeping gene expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, placentas and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents to the percentage of POLR2A/to the percentage of TBP. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

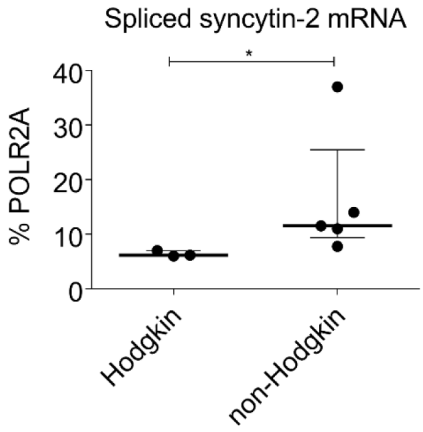

Fig. 42: Expression of the spliced syncytin-2 mRNA in Hodgkin and non-Hodgkin lymphomas. *For the analysis, the qRT-PCR method was used. The obtained amount was normalized to the percentage of POLR2A expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. The Y-axis represents the percentage of POLR2A. On the X-axis, the analyzed groups of samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

Further, we calculated the splicing efficiency of the full-length *ERVFRDE1* transcript as a ratio of spliced *ERVFRDE1* mRNA to the sum of both spliced and full-length *ERVFRDE1* mRNA in each analyzed sample. The seminomas showed significantly higher splicing efficiency in comparison to both the matched and non-GCT controls ($p < 0.001$ and $p < 0.001$, respectively, Fig. 43). On the other hand, the non-seminomas did not display significantly different splicing efficiency when compared with both the matched and non-GCT controls (Fig. 43). The highest level of *ERVFRDE1* splicing efficiency was observed in placental samples (Fig. 43).

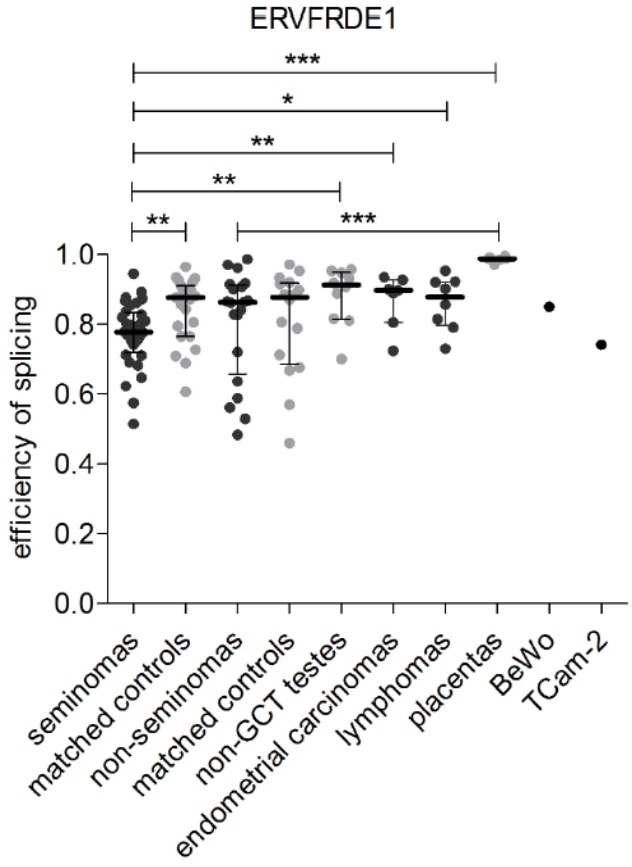

Fig. 43: ERVFRDE1 splicing efficiency in the analyzed groups of samples. *Mean values with interquartile ranges were counted. Each sample is represented by a dot in the dot plot. Lymphomas, endometrial carcinomas, placentas, and BeWo and TCam-2 cell lines were presented for the comparison. The Y-axis represents the splicing efficiency. On the X-axis, the analyzed groups of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

We further compared the splicing efficiency of *ERVWE1* and *ERVFRDE1* RNAs in the seminomas and non-seminomas. In the seminomas, the splicing efficiency of *ERVFRDE1* was

significantly higher than that of *ERVWE1* one ($p < 0.0001$, Fig. 44). Moreover, the same was observed in the non-seminomas ($p < 0.001$, Fig. 44).

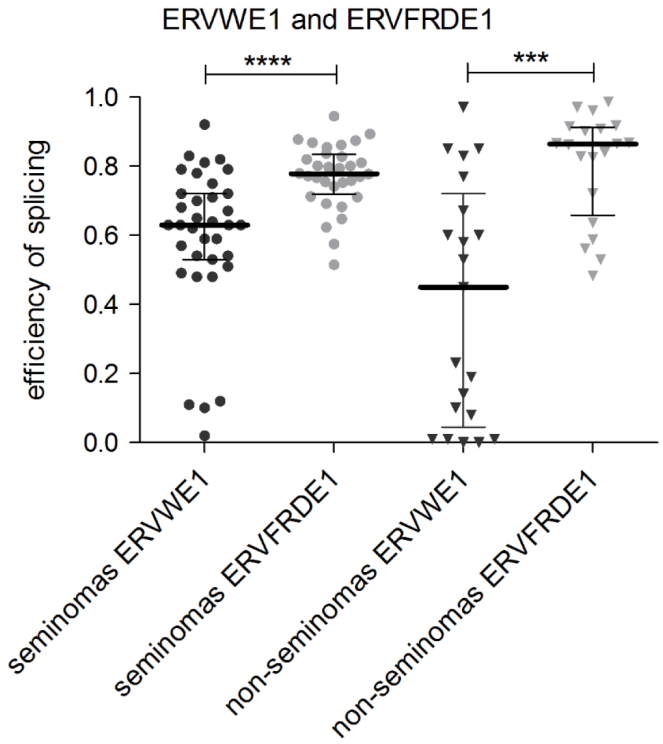

Fig. 44: Comparison of the splicing efficiency of ERVWE1 and ERVFRDE1 in the seminomas and non-seminomas. *Mean values with interquartile ranges were counted. Each sample is represented by a dot in the dot plot. The Y-axis represents the splicing efficiency. On the X-axis, the analyzed groups of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, ** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

Collectively, we showed elevated *ERVFRDE1* expression in the seminomas together with its efficient splicing. Nevertheless, the level of *ERVFRDE1* transcription in the seminomas was much lower than the *ERVWE1* mRNA expression. This is the reason why we stayed focused on the ERVWE1 production in GCTs.

2.4 *ASCT1* and *ASCT2* mRNA expression in GCTs

We observed efficient *ERVWE1* splicing in seminomas, therefore we further wanted to focus on the factors, that are able to affect Syncytin-1 fusogenic potential. One of the prerequisites for sufficient Syncytin-1-induced cell-cell fusion is the presence of its cellular receptors ASCT1 and ASCT2 (Tailor et al. 1999; Marin et al. 2000). We measured *ASCT1* and *ASCT2* mRNA production in selected subsets of samples by the qRT-PCR method.

We analyzed *ASCT1* mRNA production at first. The highest expression of *ASCT1* mRNA was found in the lymphomas (Fig. 45A). We did not observe any significant differences in the *ASCT1* mRNA expression between the groups of Hodgkin and non-Hodgkin

lymphomas (Fig. 45B). We did not observe any significant difference in *ASCT1* mRNA production between the seminomas and the matched or non-GCT controls (Fig. 45A). On the other hand, the non-seminomas displayed significantly lower levels of *ASCT1* mRNA in comparison to the matched controls ($p < 0.05$, Fig. 45A). We did not find any significant difference either between the non-seminomas and the non-GCT controls, or between the groups of the seminomas and non-seminomas (Fig. 45A). The placental samples showed higher expression of *ASCT1* mRNA than BeWo cells, but lower that in the the TCam-2 cell line.

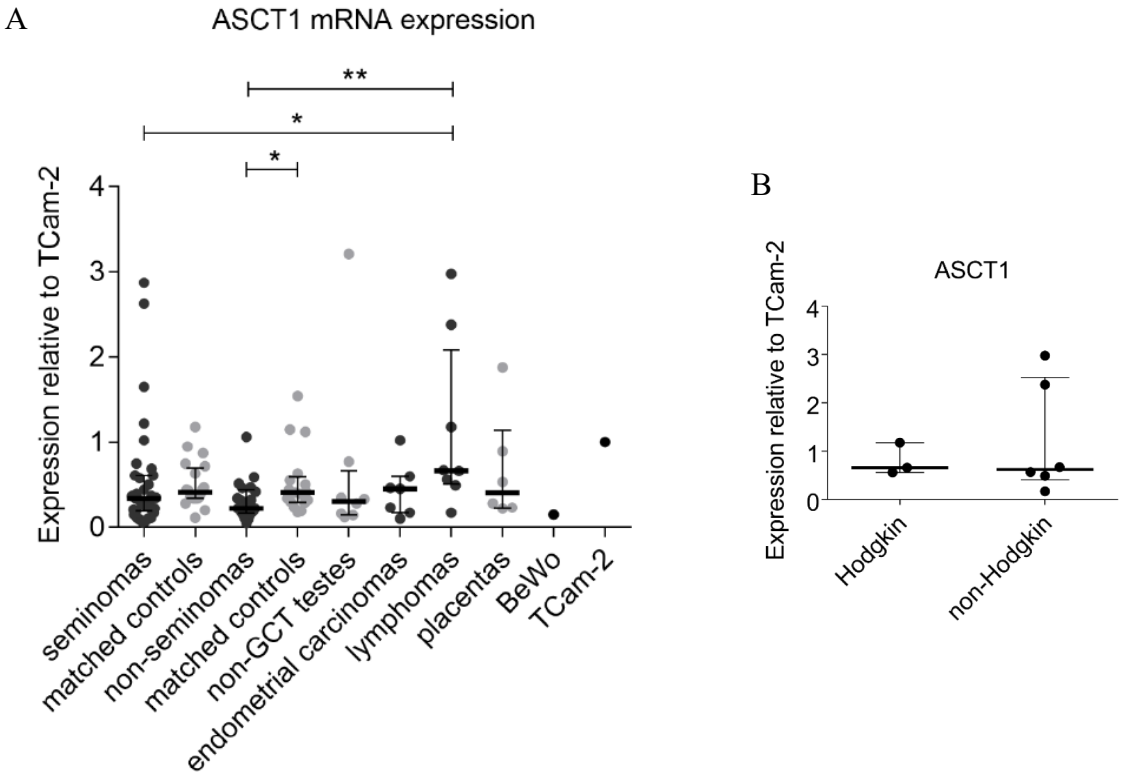

Fig. 45: Relative ASCT1 mRNA expression in a selected sub-set of samples. *(A) Relative ASCT1 mRNA expression. (B) Comparison of the ASCT1 mRNA expression between Hodgkin and non-Hodgkin lymphomas. For the analysis, the qRT-PCR method was used. The obtained amount was normalized to the relative expression observed in the TCam-2 cell line. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, placentas, and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the expression relative to expression in TCam-2 cell line. On the X-axis, the analyzed groups of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

Further, we measured the relative *ASCT2* mRNA expression in a selected sub-set of samples. The seminomas showed significantly higher *ASCT2* mRNA expression in comparison to the non-GCT controls ($p < 0.05$, Fig. 46A), but we did not find any significant difference between the seminomas and the matched controls (Fig. 46A). The non-seminomas

did not prove any significant difference in *ASCT2* mRNA expression in comparison to both types of used controls (Fig. 46A). Lymphomas showed an average level of *ASCT2* mRNA expression, and we did not observe any differences in the *ASCT2* mRNA amount between groups of Hodgkin and non-Hodgkin lymphomas (Fig. 46B). BeWo cells displayed higher *ASCT2* mRNA production than TCam-2 cells and the placental samples (Fig. 46A).

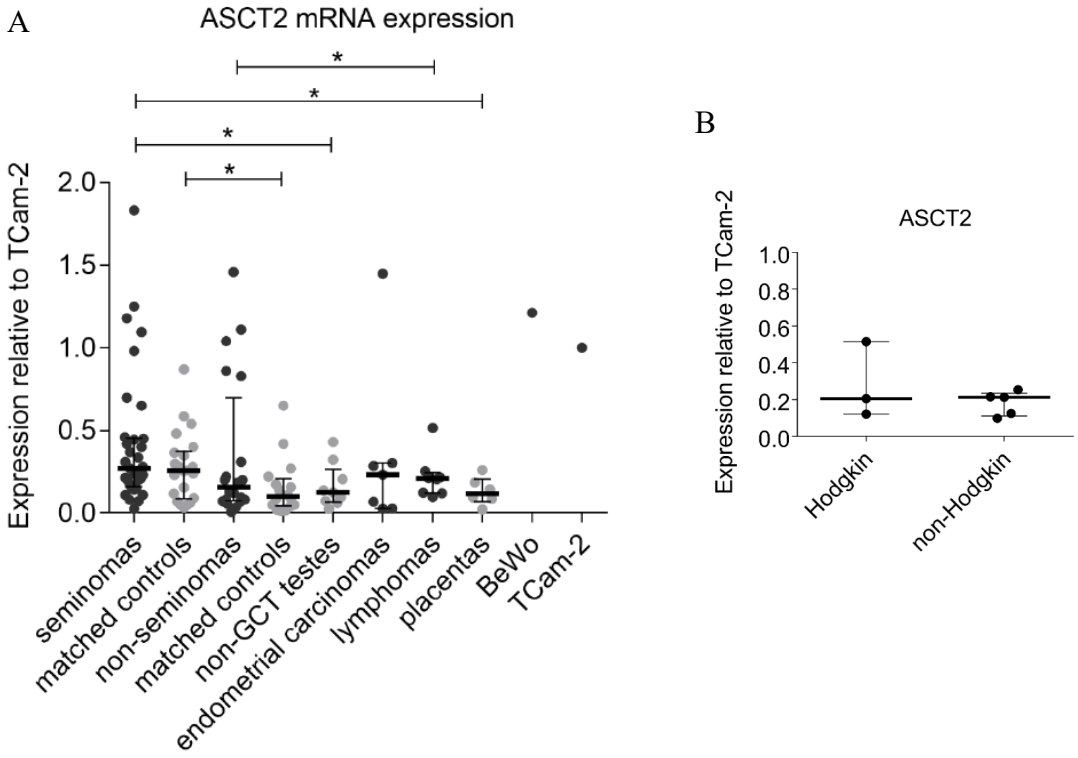

Fig. 46: Relative ASCT2 mRNA expression in a selected sub-set of samples, in Hodgkin and non-Hodgkin lymphomas. *(A) Relative ASCT2 mRNA expression.(B) Comparison of the ASCT2 mRNA expression between Hodgkin and non-Hodgkin lymphomas. For the analysis, the qRT-PCR method was used. The obtained amount was normalized to relative expression obtained in the TCam-2 cell line. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, placentas, and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents expression relative to expression in the TCam-2 cell line. On the X-axis, the analyzed groups of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

To summarize, we did not show any significant difference in the *ASCT1*/*ASCT2* mRNA expression in both the seminomas and non-seminomas when compared to their matched controls. That implies that the fusogenic potential of Syncytin-1 in GCTs should not be influenced by changes in the expression level of its receptors.

2.5 *ERVWE1* copy number in seminomas

To determine whether the high level of *ERVWE1* mRNA expression in the seminomas was not caused by multiplication of the *ERVWE1* gene, we analyzed the *ERVWE1* copy number by ddPCR in a selected sub-set of samples.

The analysis revealed that the matched controls, the non-GCT control and the placenta sample did not show any abnormalities in the *ERVWE1* copy number (Fig. 47). The seminomas displayed various copy numbers, between 0.97 to 1.78 per haploid genome (Fig. 47). Nevertheless, this copy number in the seminomas was not significantly higher than in the matched controls (Fig. 47). On the other hand, the non-seminomas showed a significantly higher copy number when compared with the matched controls ($p < 0.01$, Fig. 47). Even though we observed a difference in the level of both full-length and spliced *ERVWE1* mRNA in the seminomas and in the non-seminomas, their copy number did not differ (median 1.34 per haploid genome for both, Fig. 47). To summarize, we observed increased ERVWE1 copy numbers in some of the analyzed GCTs, but this elevation did not correspond with the expression of both full-length and spliced *ERVWE1* mRNA in these samples (Fig. 48A and Fig. 48B). Collectively, we did not detect elevated *ERVWE1* gene copies in the seminomas. Therefore, differences in the *ERVWE1* mRNA production in our samples should not be caused by changes in the *ERVWE1* gene copy number.

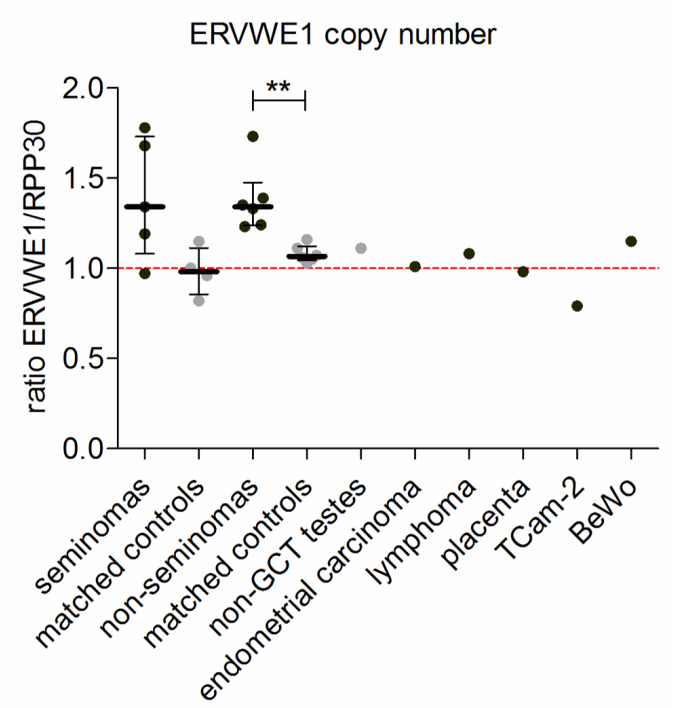

Fig. 47: ERVWE1 copy number in a selected sub-set of samples. *The number of ERVWE1 gene copies was measured relative to the RPP30 gene copy number by ddPCR. Each sample was measured in technical duplicate, averaged, and then represented by a dot in the dot plot. For each group of samples, mean values with interquartile ranges are depicted in the chart. Red interrupted line represents 1 : 1 ERVWE1 : RPP30 ratio. The Y-axis shows the ERVWE1 copy number to RPP30 copy number ratio. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

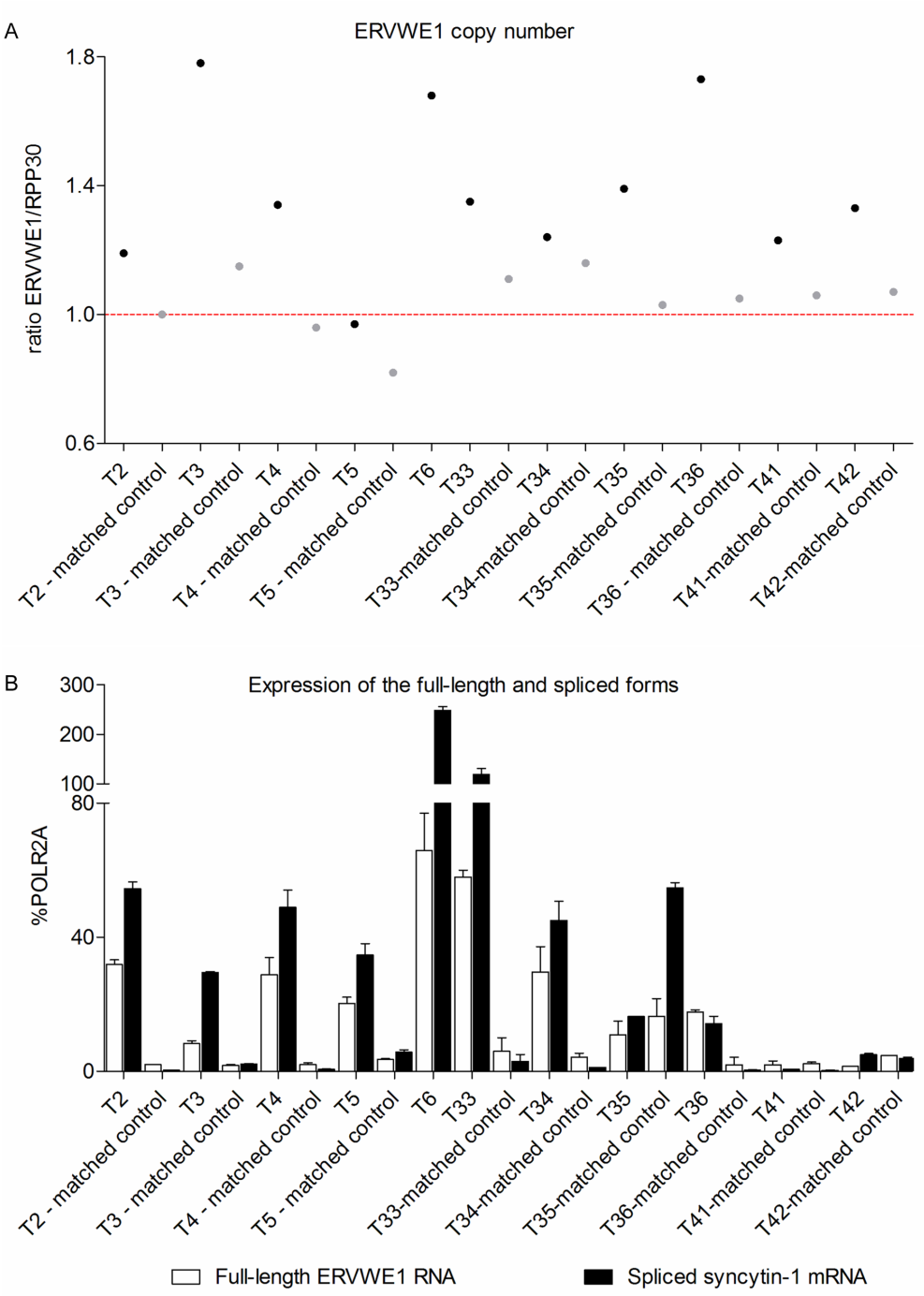

Fig. 48: ERVWE1 copy number and the amount of non-spliced ERVWE1 transcript and spliced mRNA in the tumor samples and matched controls. *(A) ERVWE1 copy number in a selected sub-set of samples and the matched controls. The number of ERVWE1 copies was measured relative to the RPP30 gene copy number by ddPCR. Each sample was measured in technical duplicate, averaged, and then represented by a dot in the dot plot. Red interrupted line represents the ERVWE1 : RPP30 1 : 1 ratio. The Y-axis stands for the ratio of ERVWE1 copy number to RPP30 copy number. On the X-axis, the analyzed samples are depicted. (B) ERVWE1 expression in a selected sub-set of samples and the matched controls. Both full-length and spliced form of ERVWE1 mRNA were detected. The analyzed expression of both types of transcripts was measured relatively to the percentage of POLR2A expression. Production of the full-length ERVWE1 mRNA is depicted by white bars, the spliced mRNA by the black ones. Each sample was measured in technical triplicate and standard deviation with ± was calculated. The Y-axis represents the percentage of POLR2A, the X-axis stands for samples used in the analysis. (Benešová, Trejbalová, Kovářová, et al. 2017)*

2.6 *GCM1* expression in GCTs

Transcription factor GCM1 is known to stimulate transcription from the ERVWE1 gene (C. Yu et al. 2002). To uncover the previously observed elevated *ERVWE1* mRNA expression in our samples, we measured the absolute *GCM1* mRNA expression in selected sub-sets of samples by qRT-PCR.

The highest expression of *GCM1* mRNA was observed in the placental samples and in BeWo cells (Fig. 49A). Furthermore, *GCM1* mRNA production in the placentas was significantly higher than in all other analyzed groups of samples (Fig. 49A). We did not observe any significant differences in *GCM1* mRNA expression between the seminomas and non-seminomas and their matched controls (Fig. 49A) or between the groups of Hodgkin and non-Hodgkin lymphomas (Fig. 49B).

Collectively, we did not show any changes in *GCM1* mRNA expression in the analyzed panel of tumor samples. Therefore, the elevated level of *ERVWE1* expression in seminomas is not caused by an increased level of *GCM1* expression in these tumors.

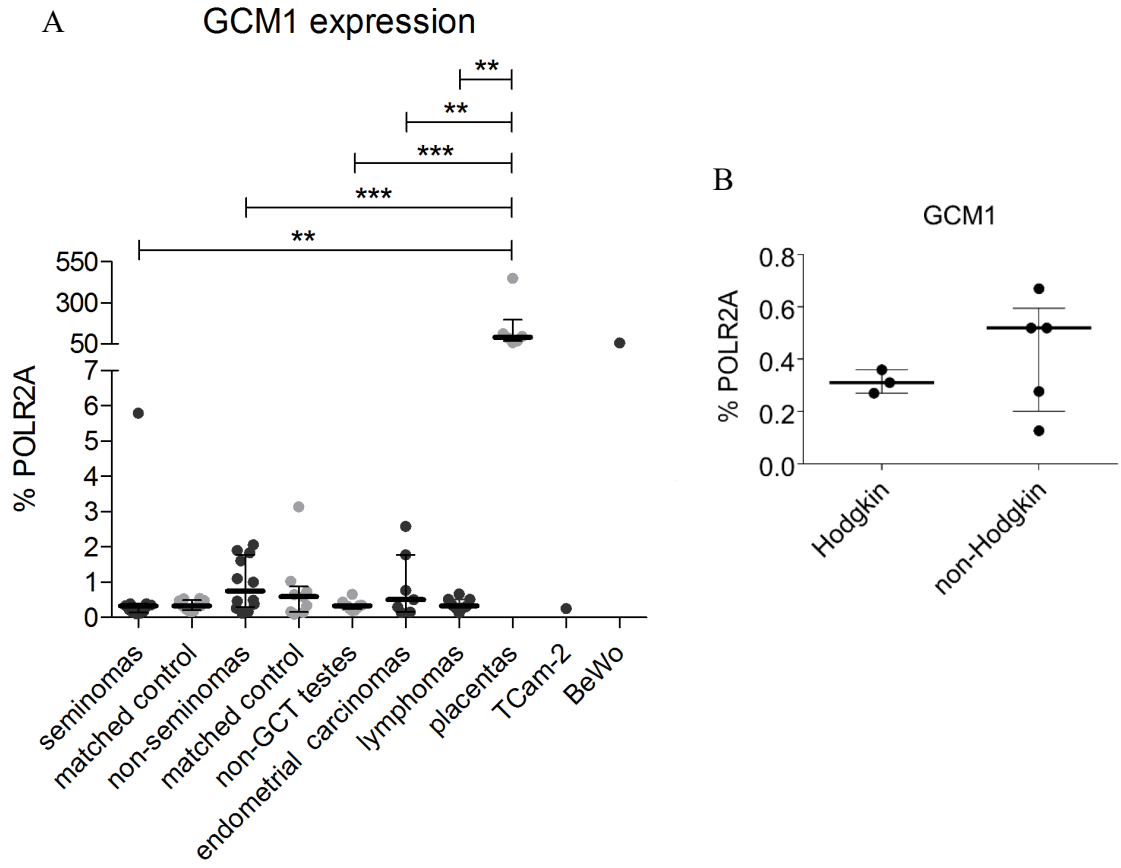

Fig. 49: GCM1 mRNA expression in a selected sub-set of samples. *(A) GCM1 mRNA expression in a selected sub-set of samples. (B) Comparison of the GCM1 mRNA expression between Hodgkin and non-Hodgkin lymphomas. For the analysis, the qRT-PCR method was used. The absolute amount of GCM1 mRNA production was normalized to the percentage of POLR2A expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, placentas, and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the percentage of POLR2A. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asteriskr(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

2.7 *ERVWE1* promoter CpG methylation and hydroxymethylation in seminomas

Another important factor that regulates gene transcription is the extent of promoter CpG methylation. The previous studies done in our laboratory demonstrated unambiguously that *ERVWE1* promoter DNA hypermethylation in non-placental tissues suppresses transcription from this gene (Matoušková et al. 2006; Trejbalová et al. 2011). Furthermore, we have shown that the TET1 DNA demethylation enzyme is upregulated in all tested seminomas

(Benešová, Trejbalová, Kučerová, et al. 2017). To uncover the potential changes in DNA methylation at the *ERVWE1* promoter in our tumor samples, we examined the level of 5mC and 5hmC at this promoter. For the analysis, bisulfite and oxidative bisulfite sequencing were employed.

As the highest *ERVWE1* mRNA expression was observed in the seminomas, we compared *ERVWE1* promoter CpG methylation in a selected sub-set of seminoma samples and the matched controls. We found the *ERVWE1* promoter to have a significantly decreased level of 5mC in the seminomas (Fig. 50). Seminoma sample T6 showed hypomethylated *ERVWE1* promoter (Fig. 50 and Table 6). Unfortunately, sample T6 did not have the matched-control available, and therefore we were not able to compare the level of DNA methylation with the non-tumoral tissue. The examined seminoma samples T3, T4 and T6 displayed either fully methylated or demethylated *ERVWE1* molecules (clones contained all methylated or all demethylated CpGs, Fig. 50). Only the T5 seminoma sample showed a patchy CpG DNA methylation pattern (the same clone contained both methylated and non-methylated CpGs, Fig. 50). All seminoma samples were further analyzed for the presence of 5hmC. We showed that the level of 5hmC in seminomas was very low or undetectable (Fig. 50 and Table 6), despite the fact that the oxidative bisulfite treatment was efficient (Fig. 52). One examined lymphoma sample showed hypermethylation of the *ERVWE1* promoter together with undetectable level of 5hmC (Fig. 50 and Table 6). In accordance with high *ERVWE1* mRNA expression in placental samples and BeWo cell line, we demonstrated *ERVWE*1 promoter hypomethylation together with an undetectable level of 5hmC there (Fig. 50 and Table 6).

We found that the *ERVWE1* promoter DNA was hypomethylated in the analyzed seminomas in comparison to the matched controls. The level of 5hmC on the *ERVWE1* promoter was under the detection limit in the seminomas, even though we observed upregulated mRNA expression of *TET* demethylation enzymes.

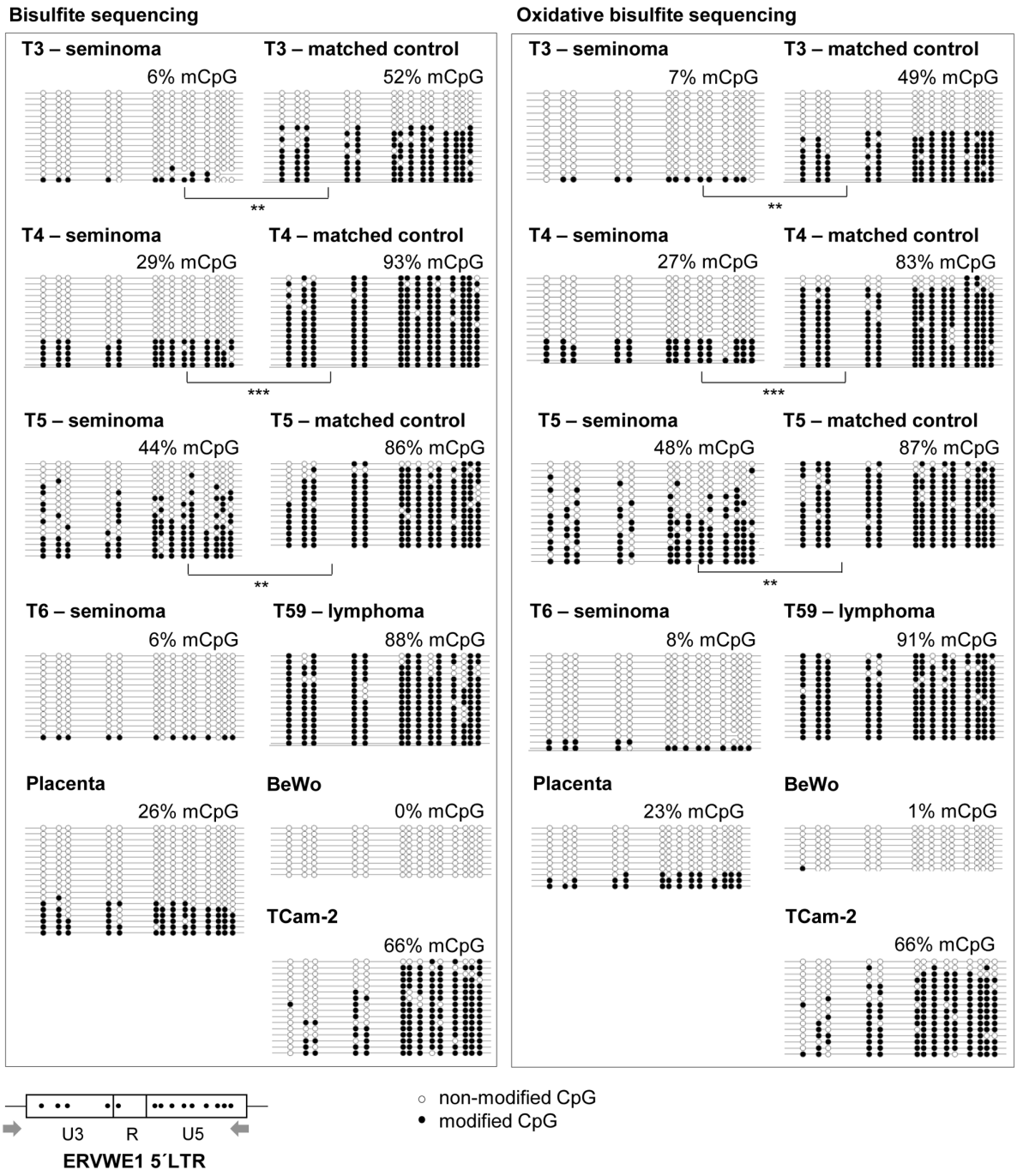

Fig. 50: Bisulfite and oxidative bisulfite sequencing of the ERVWE1 promoter in analyzed samples. *The left panel represents results of bisulfite sequencing (5mC + 5hmC); the right panel displays oxidative bisulfite sequencing (5mC). Each line in the figure represents one analyzed molecule of the ERVWE1 promoter. Each dot represents a CpG dinucleotide in the ERVWE1 promoter sequence. The black dot represents modified cytosine (= methylated or hydroxymethylated) in the CpG dinucleotide; the white dot represents non-modified cytosine in the CpG dinucleotide. Significance (P-value) between the analyzed samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). In the bottom part of the figure, schematic representation of the analyzed ERVWE1 promoter is depicted. Grey arrows represents used primers, black dots the CpG dinucleotides and U3, R and U5 for the parts of the promoter. (Benešová, Trejbalová, Kovářová, et al. 2017)*

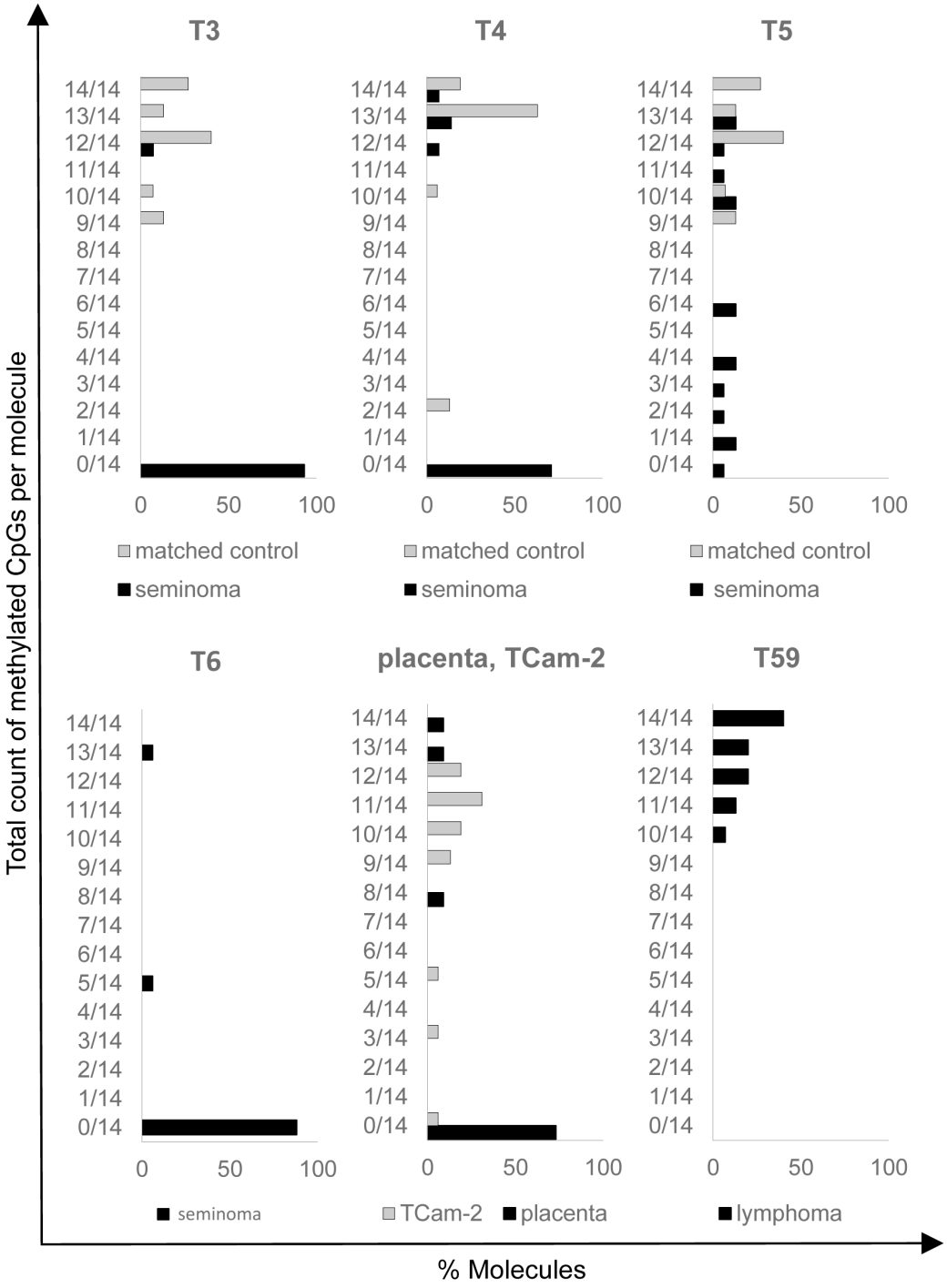

Fig. 51: Distribution of methylated CpGs on the ERVWE1 promoter (5´LTR) in analyzed samples. *Every plot represents the percentage of methylated CpGs in the determined position. The Y-axis describes position of methylated CpG (14 in total), the X-axis shows the percentage of methylated CpG molecules. (Benešová, Trejbalová, Kovářová, et al. 2017)*

Table 6: The level of 5mC and 5hmC on the ERVWE1 promoter in analyzed samples. *The total percentage of methylated CpGs (5mCpG) was obtained by oxidative bisulfite sequencing. The total amount of hydroxymethylated CpGs (5hmCpG) was calculated as a difference between the results obtained by bisulfite and oxidative bisulfite sequencing. (Benešová, Trejbalová, Kovářová, et al. 2017)*

ERVWE1 promoter	Bisulfite sequencing % (total 5mCpG + 5hmCpG)	Oxidative bisulfite sequencing % (5mCpG)	5hmCpG
T2 (seminoma)	6	7	-1
T2-matched control	52	49	3
T3 (seminoma)	29	27	2
T3-matched control	93	83	10
T4 (seminoma)	44	48	-4
T4-matched control	86	87	-1
T5 (seminoma)	6	8	-2
T55 (lymphoma)	88	91	-3
placenta	26	23	3
BeWo	0	1	-1
TCam-2	66	66	0

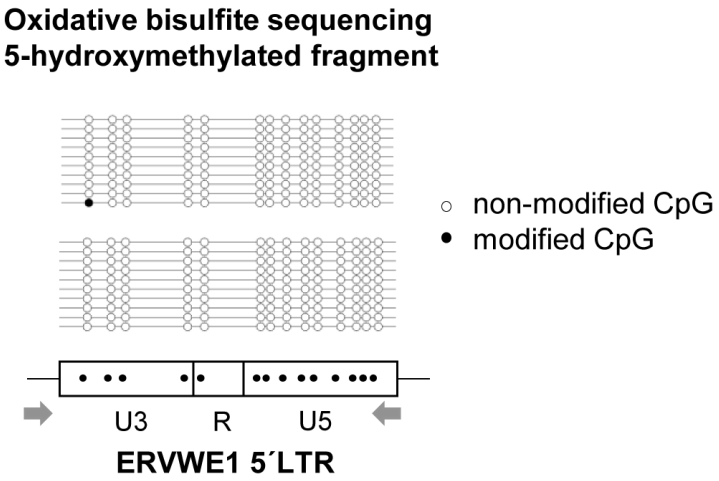

Fig. 52: Oxidative bisulfite sequencing at the 5-hydroxymethylated fragment of the ERVWE1 promoter (5´LTR). *Result of the oxidative bisulfite sequencing of the artificially prepared 5-hydroxymethylated DNA fragment of ERVWE1 5´LTR is depicted. Each line represents one analyzed molecule, each circle stands for a CpG dinucleotide. White circle represents non-methylated CpG, black circle the methylated ones. The plot clearly shows that the oxidation of 5hmC was efficient (97.9 – 99.3 %). In the lower part of the figure, the scheme of the ERVWE1 5´LTR is depicted. Grey arrows represent primers used for oxidative bisulfite sequencing, black circles represents analyzed CpG dinucleotides. ERVWE1 5´LTR consists of three parts – U3, R and U5. (Benešová, Trejbalová, Kučerová, et al. 2017)*

2.8 Expression of other selected endogenous retroviruses in GCTs

Observed derepression of ERVWE1 transcription in GCTs led us to analyzis of mRNA expression of other selected endogenous retroviruses. We measured production from the HERV-H endogenous retrovirus, which is localized at chromosome 7 close to the *ERVWE1*

gene (Fig. 10 in the Literature review). In addition, we looked at two other HERV-Ws that are present at chromosomes 4 and 21. These two *HERV-W* loci were previously shown to be expressed in testicular tumors (Gimenez et al. 2010). The selected sub-set of samples was analyzed by qRT-PCR (HERV-H) and ddPCR (HERV-Ws) methods.

We first measured the expression from HERV-W copies at chromosomes 4 and 21. Both analyzed HERV-Ws showed significantly elevated mRNA expression in the seminomas in comparison to the matched controls ($p < 0.01$ and $p < 0.001$, respectively, Fig. 53A and B), non-GCT controls ($p < 0.001$ and $p < 0.0001$, respectively, Fig. 53A and B) as well as non-seminomas ($p < 0.01$ and $p < 0.01$, respectively, Fig. 53A and B). The non-seminomas displayed significantly upregulated mRNA expression of *HERV-Ws* localized at chromosomes 4 and 21 when compared with the non-GCT controls ($p < 0.05$, Fig. 53A, $p < 0.001$, Fig. 53B, respectively), but not with the matched controls (Fig. 53A and B). The TCam-2 cell line showed high mRNA expression of both analyzed *HERV-Ws* (Fig. 53A and B). We did not observe any significant differencies between the groups of Hodgkin and non-Hodgkin lymphomas (Fig. 55).

Next, we analyzed mRNA expression of HERV-H, which is localized upstream from the *ERVWE1* at chromosome 7. We observed low level of *HERV-H* mRNA in our samples (Fig. 53C). However, we showed significantly upregulated *HERV-H* mRNA expression in both the seminomas and non-seminomas when compared with the placental samples ($p < 0.001$ and $p < 0.001$, respectively, Fig. 53C), but not with the matched or non-GCT controls (Fig. 53C). We did not observe any significant differencies between the groups of Hodgkin and non-Hodgkin lymphomas (Fig. 54).

Further, we compared the mRNA expression of *HERV-Ws* localized at chromosomes 4 and 21 and *HERV-H* at chromosome 7 with previously presented mRNA expression from *ERVWE1* and *ERVFRDE1* loci. The production of *ERVWE1* spliced mRNA was significantly elevated in the seminomas when compared with the mRNA expression from *HERV-W* at chromosome 4, *HERV-W* at chromosome 21 and *HERV-H* at chromosome 7 ($p < 0.0001$, $p < 0.0.01$ and $p < 0.0001$, respectively, Fig. 55A). In the seminomas, the production of *ERVWE1* full-length mRNA was significantly elevated in comparison to the *HERV-H* mRNA expression ($p < 0.05$, Fig. 55A). Moreover, the *ERVWE1* full-length mRNA production was significantly elevated in the non-seminomas when compared with the expression from both *HERV-W* chromosome 4 and 21 loci ($p < 0.01$ and $p < 0.01$, respectively, Fig. 55B). Furthermore, the *ERVFRDE1* full-length mRNA level in the non-seminomas was significantly upregulated when compared with both *HERV-W* chromosome 4 and 21 expression ($p < 0.05$

and $p < 0.05$, respectively, Fig. 55B).

To conclude, we observed upregulated mRNA levels of two analyzed *HERV-Ws* in GCTs, but this expression did not reach the production detected for both *ERVWE1* and *ERVFRDE1*. *HERV-H*, which is localized in the vicinity of *ERVWE1*, showed just low mRNA expression in the analyzed GCTs.

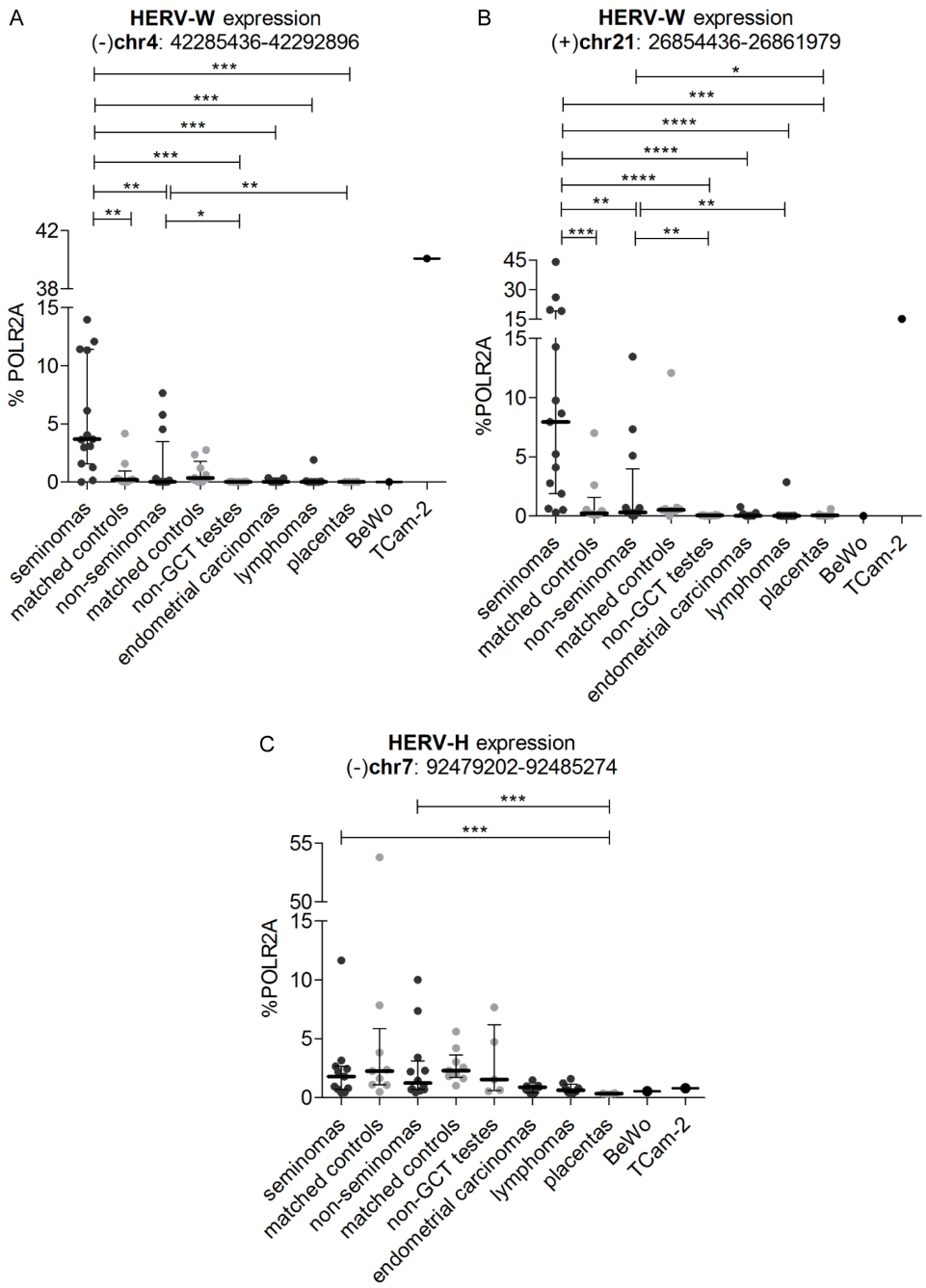

Fig. 53: Expression of the HERV-H mRNA at chromosome 7 and mRNA of HERV-Ws localized at chromosomes 4 and 21 in a selected sub-set of samples. *mRNA expression was analyzed by qRT-PCR (HERV-H) and ddPCR (HERV-Ws). The obtained amount was normalized to the percentage of POLR2A expression. (A) Expression of HERV-W localized at chromosome 4 measured in selected subsets of samples. (B) Expression of HERV-W localized at chromosome 21 measured in selected subsets of samples. (C) Expression of HERV-H measured in selected subsets of samples. In the HERV-H (qRT-PCR) analysis, each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. In the HERV-Ws (ddPCR) analysis, each sample was measured in technical duplicate, averaged, and then represented by a dot in the dot plot. Mean values with interquartile ranges are depicted in the chart for each group of samples. Lymphomas, endometrial carcinomas, placentas and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the percentage fo POLR2A. On the X-axis, the analyzed group of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

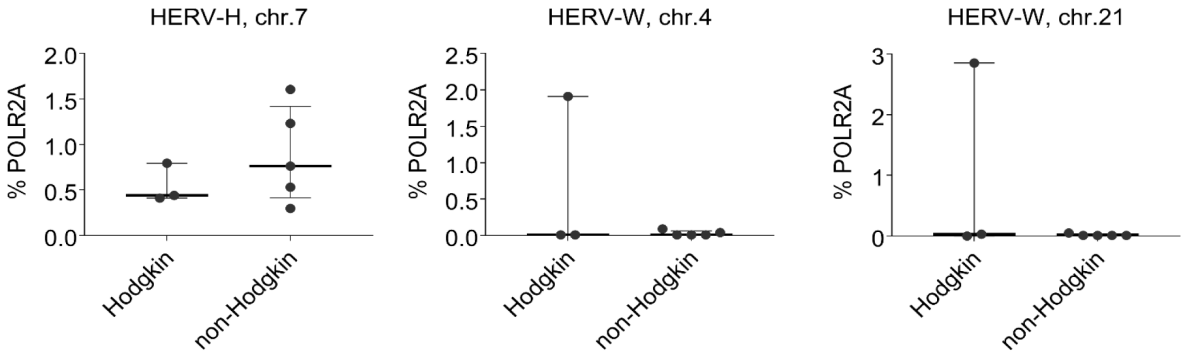

Fig. 54: Expression of the HERV-Ws at chromosomes 4 and 21 and HERV-H at chromosome 7 mRNA in Hodgkin and non-Hodgkin lymphomas. *HERV-W at chromosomes 4 and 21 and HERV-H at chromosome 7 mRNA production was measured by qRT-PCR. The obtained amount was normalized to the percentage of POLR2A expression. Each sample was measured in technical triplicate, averaged, and then represented by a dot in the dot plot. For each group of samples, mean values with interquartile ranges are depicted in the chart. The Y-axis represents the percentage of POLR2A. On the X-axis, the analyzed groups of samples are depicted. (Benešová, Trejbalová, Kovářová, et al. 2017)*

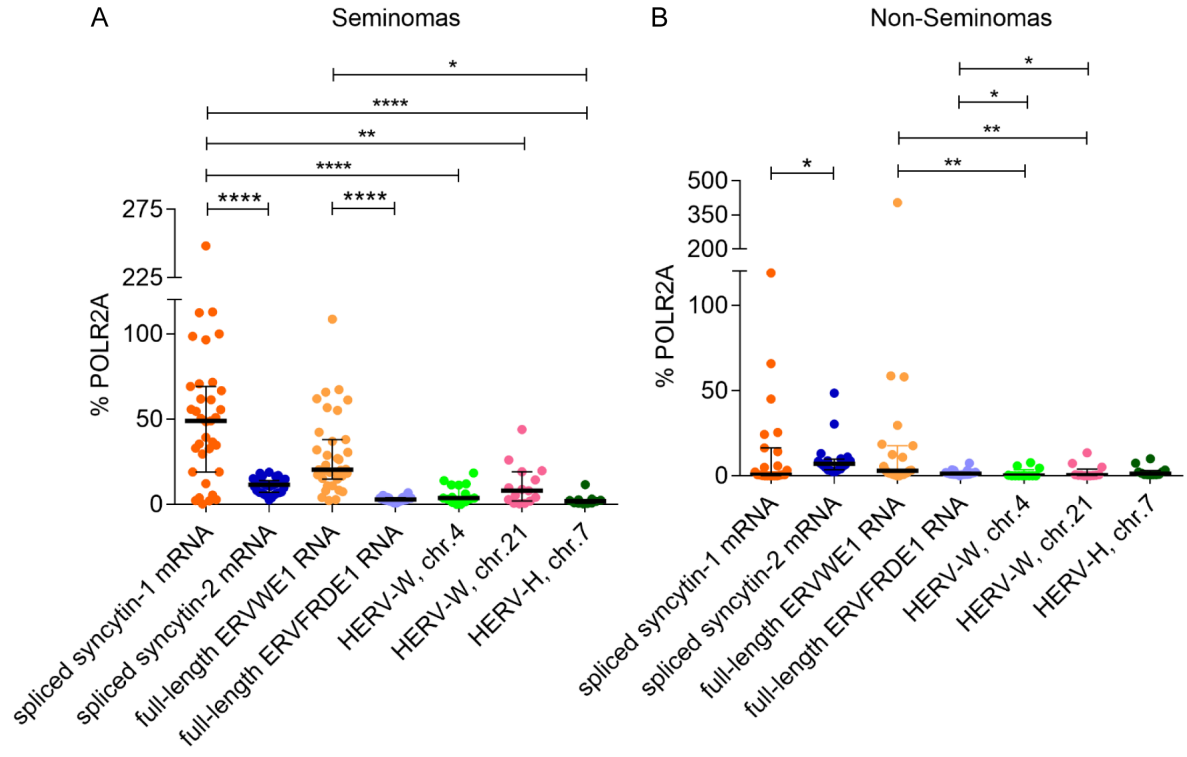

Fig. 55: Comparison of ERVWE1, ERVFRDE1, HERV-W localized at chromosome 4, HERV-W localized at chromosome 21 and HERV-H at chomosome 7 mRNA expression in the seminomas and non-seminomas. *(A) ERVWE1 (spliced and full-length mRNA), ERVFRDE1 (spliced and full-length mRNA), HERV-W chromosome 4, HERV-W chromosome 21 and HERV-H chromosome 7 mRNA expression in the seminomas. (B) ERVWE1 (spliced and full-length mRNA), ERVFRDE1 (spliced and full-length mRNA), HERV-W chromosome 4, HERV-W chromosome 21 and HERV-H chromosome 7 mRNA expression in the non-seminomas. Mean values with interquartile ranges are depicted in the chart. Lymphomas, endometrial carcinomas, placentas and BeWo and TCam-2 cell lines are presented for the comparison. The Y-axis represents the percentage of POLR2A. On the X-axis, the analyzed groups of samples/samples are depicted. Significance (P-value) between the analyzed groups/samples is represented by bars with asterisk(s) above (* < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001). (Benešová, Trejbalová, Kovářová, et al. 2017)*

In the second part of the study, we showed that seminomas are characterized not only by elevated expression of the *TET1* mRNA, but also by upregulated production of both full-length and spliced forms of *ERVWE1* transcripts. Moreover, this expression is allowed due to the lower level of *ERVWE1* promoter CpG methylation. Our analysis also showed that *ERVWE1* expression in the seminomas is regulated independently of other examined endogenous retroviruses.

Discussion

We explored selected features of human GCTs in both presented studies. Specifically, we extended the panel of pluripotency factors that are expressed in GCTs. Moreover, we described the expression pattern of selected human endogenous retroviruses in GCTs.

GCTs are derived from GCNIS, which originates from the gonocyte or PGC with delayed or blocked maturation (Skakkebék et al. 1998; Oosterhuis and Looijenga 2005; Rajpert-De Meyts 2006; Kristensen et al. 2013). GCNIS gives rise to seminoma or embryonal carcinoma (Skakkebék et al. 1998; Oosterhuis and Looijenga 2005; Rajpert-De Meyts 2006; Kristensen et al. 2013). Other non-seminomas (yolk sac tumor, choriocarcinoma and teratoma) originate from the embryonal carcinoma (Oosterhuis and Looijenga 2005). It was proposed that embryonal carcinomas can develop directly from GCNIS or from the existing seminoma tumor (Oosterhuis and Looijenga 2005). Both gonocytes and PGCs express markers characteristic of pluripotent cells (Oosterhuis and Looijenga 2005; Gillis et al. 2011). Therefore, it is not surprising that seminomas and embryonal carcinomas also express these factors and in addition display epigenetic features of pluripotent cells – such as low level of DNA methylation and the presence of H3K9me3 (Looijenga et al. 2003; Hoei-Hansen et al. 2005; Cao et al. 2009; Almstrup et al. 2010; Gillis et al. 2011; Kristensen et al. 2014). On the other hand, yolk sac tumor, choriocarcinoma and teratoma are characterized by a higher level of DNA methylation together with lower expression of the pluripotency factors (Smiraglia et al. 2002; Looijenga et al. 2003; Almstrup et al. 2004; Kehler et al. 2004; Rajpert-De Meyts et al. 2004; Hoei-Hansen et al. 2005; Netto et al. 2008; Cao et al. 2009; Gillis et al. 2011; Kristensen et al. 2014). GCTs formation and their characteristic features are summarized in Fig. 56. The unique status of seminomas is also shown. From the medical point of view, it is very important to distinguish between particular GCT types because of even contradictory ways of their treatment and aggressiveness (Einhorn 1997; Oosterhuis and Looijenga 2005; Wermann et al. 2010). Previously described expression of human endogenous retroviruses *ERVWE1* and *ERVFRDE1* in GCTs (Gimenez et al. 2010; Trejbalová et al. 2011) led us to searching for specific features that enable expression from these endogenous retroviruses. Furthermore, above-mentioned studies did not specify *ERVWE1* and *ERVFRDE1* expression in GCT types. Therefore, we aimed to specify (if possible) the exact GCT type where *ERVWE1* and *ERVFRDE1* are expressed.

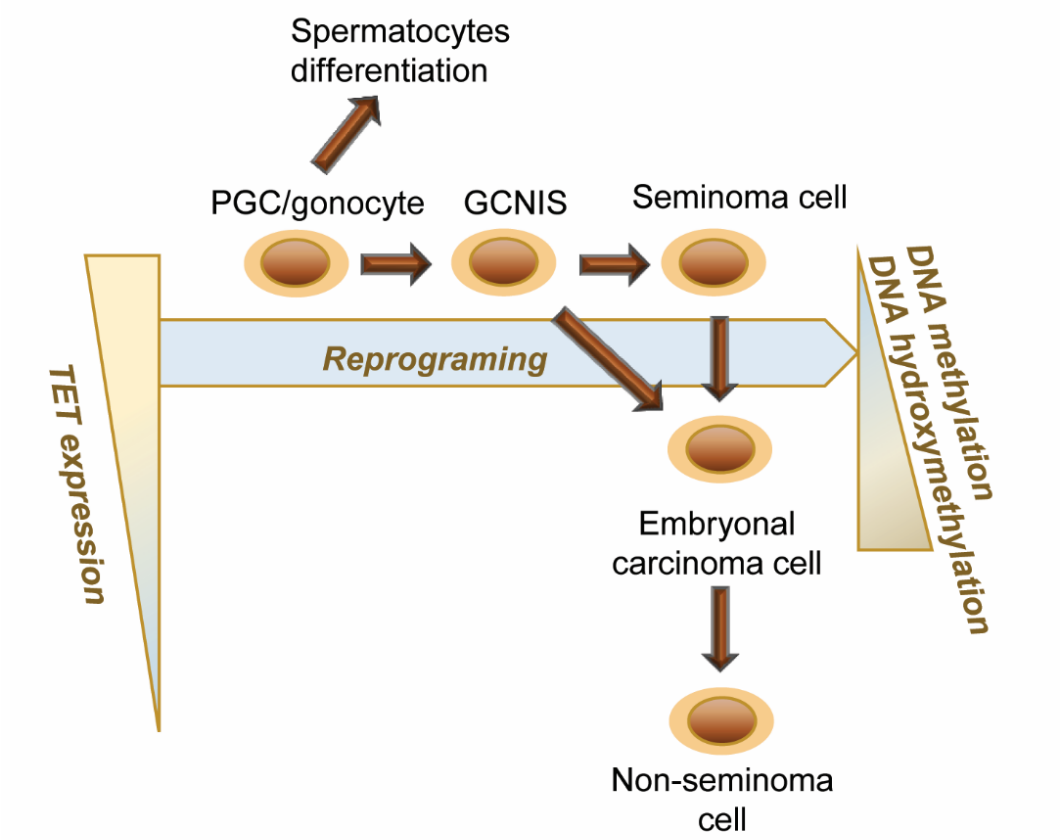

Fig. 56: Schematic representation of the GCTs formation and changes in their characteristic epigenetic features. *GCT types differentiation is mainly accompanied by changes in DNA methylation (from hypomethylated GCNIS/seminoma to more methylated embryonal carcinoma and further to hypermethylated other non-seminomas). The increased level of 5mC/5hmC is accompanied by downregulated expression of TET demethylation enzymes. The level of 5mC/5hmC and TET expression is represented by triangles with color scale score. (Benešová, Trejbalová, Kučerová, et al. 2017)*

To confirm the differentiation status of our samples, we examined the mRNA expression of pluripotency factors. Seminomas showed elevated mRNA expression of *OCT4*, *NANOG* and *SALL4*. Non-seminomas expressed some of the selected pluripotency factors as well, but to a lesser extent. The exception were teratomas, where expression of pluripotency markers was not detected at all. This is in agreement with the fact that non-seminomas are more differentiated than seminomas (Looijenga et al. 2003; Almstrup et al. 2004; Kehler et al. 2004; Rajpert-De Meyts et al. 2004; Hoei-Hansen et al. 2005; Cao et al. 2009; Gillis et al. 2011; Kristensen et al. 2014). The highest expression of pluripotency factors in the seminomas highlights their germ line origin and points to the importance of the change in pluripotency genes transcription during the seminoma to non-seminoma transition.

The differentiation of GCNIS from the PGC/gonocyte and further to GCT is associated with changes in both gene expression and the level of DNA methylation (Skakkebék et al. 1998; Smiraglia et al. 2002; Oosterhuis and Looijenga 2005; Rajpert-De

Meyts 2006; Netto et al. 2008; Gillis et al. 2011; Kristensen et al. 2013; 2014) (Fig. 56). As the family of TET enzymes converts 5mC to 5hmC (Tahiliani et al. 2009; Ito et al. 2010), and therefore might participate in the process of DNA demethylation in GCTs, we analyzed transcription of all three TET enzymes in the panel of our samples. We found *TET1* mRNA expression to be significantly elevated in the seminomas in comparison with both used controls and further with the non-seminomas. The other two TETs (*TET2* and *TET3*) displayed lower mRNA expression than *TET1*, but showed the same expression pattern. Collectively, our data proved overexpression of *TET* dioxygenases in GCTs, especially in seminomas (Fig. 56). These data suggest that the TET1 enzyme is probably the principal dioxygenase in GCTs, but the other two TETs (TET2 and TET3) also contribute to the overall TET activity.

We confirmed TET1 overexpression in GCTs at the protein level as well. Specifically, we observed elevated TET1 protein presence in both the seminomas and non-seminomas in comparison with the matched controls. Nevertheless, we did not observe any significant differences between the groups of seminomas and non-seminomas. To understand the activity of TET enzymes in GCTs more deeply, the total TET enzymatic activity assay should be done in the future.

As TET proteins are DNA-demethylating enzymes (Tahiliani et al. 2009; Ito et al. 2010; 2011; He et al. 2011), the global level of 5mC and 5hmC was analyzed in GCTs. In the seminomas, we observed low levels of both examined cytosine modifications. The low amount of 5mC is in agreement with the fact that seminomas originate from pluripotent cells and hence share some of their properties (Netto et al. 2008; Kristensen et al. 2013). Very low levels of 5hmC in the seminomas can be surprising. Nevertheless, 5hmC can be produced only from the already existing 5mC. Moreover, 5hmC is further transformed to 5fC, 5caC and C by the base excision repair mechanism (Tahiliani et al. 2009; Ito et al. 2010; 2011; He et al. 2011). The dynamics of the 5hmC conversion is supposed to be very fast, but the precise timing of the process remains unknown. In contrast to the seminomas, the non-seminomas displayed higher levels of 5hmC. This can be connected with the higher overall amount of 5mC in non-seminomas. Therefore, the global level of cytosine modifications in different GCT types is a result of the interplay between the activity of DNA methylation/demethylation enzymes and the dynamics of the cytosine modifications conversion. In the future, correlation of the TET1 activity and 5hmC levels in distinct stages of germ cell tumors have to be elucidated.

Described low levels of 5mC and 5hmC together with high expression of TET

dioxygenases in seminomas is in direct contrast to the characteristics of CIMP tumors. Therefore seminomas can be called as a tumors with the “antimethylator” phenotype.

Described levels of 5mC in different GCT types are further supported by the data of (Costa et al. 2018), who proposed the method for distinguishing between GCT types. This method is based on simultaneous comparative analysis of the promoter DNA methylation of five different genes (*CRIPTO, HOXA9, MGMT, RASSF1A* and *SCGB3A1*) (Costa et al. 2018).

The activity of TET enzymes can be negatively influenced by the presence of specific mutations in *IDH1* and *IDH2* genes, which can lead to the production of TET antagonist 2-hydroxyglutarate (Yan et al. 2009; W. Xu et al. 2011). Sequence analysis of the selected sub-sets of samples did not show any of the examined inhibiting mutations to be present. Therefore, the activity of TET enzymes was not disrupted by the presence of 2-hydroxyglutarate in our samples.

The dynamics of 5mC and 5hmC modifications is affected not only by the TET proteins, but also by oppositely acting enzymes – DNMTs. We observed elevated *DNMT3A* mRNA expression in GCTs in comparison with the controls. This finding is in accordance with the fact, that DNMT3A suppressor miR-199-3P was shown to be depleted in GCTs (B. F. Chen et al. 2014). We further observed overexpression of *DNMT3B* mRNA in the seminomas, but not in the non-seminomas, when compared with the controls. The production of *DNMT1* mRNA was comparable in all analyzed groups of samples. The overall level of cytosine modifications in different types of GCTs is regulated by both TET enzymes and *de novo* DNMTs. However, the amount of 5mC and 5hmC in most of the analyzed samples was low, and therefore it is very difficult to assess the final result of the TET-DNMT interplay in GCTs. Knockout of different *TET* and *DNMT* genes would provide better insight into this issue.

To confirm our data obtained for the *TET* and *DNMT* genes, we analyzed RNA seq (NCBI) and microarray (Oncomine) data that are publicly available. RNAseq data showed high *TET1* mRNA expression in the TCam-2 cell line and in PGCs and was accompanied by low expression of *DNMT1*, *DNMT3A* and *DNMT3B* (Fig. 57A). Microarray data displayed upregulated expression of TET enzymes in all analyzed samples, particulary of TET1. TET1 expression was highest in seminoma sample, but was also elevated in embryonal carcinoma, yolk sac tumor and in mixed GCT samples (Fig. 57B). Moreover, TET2 and TET3 proteins were upregulated in seminoma, embryonal carcinoma, yolk sac tumor and mixed GCT samples (Fig. 57B). DNMT1 was not overexpressed in any of the analyzed GCTs (Fig. 57B). DNMT3A was just slightly elevated in all examined GCTs (Fig. 57B). DNMT3B displayed the highest expression in embryonal carcinoma, but was low in seminoma samples (Fig. 57B).

The data taken from the databases are in accordance with the results presented in this study.

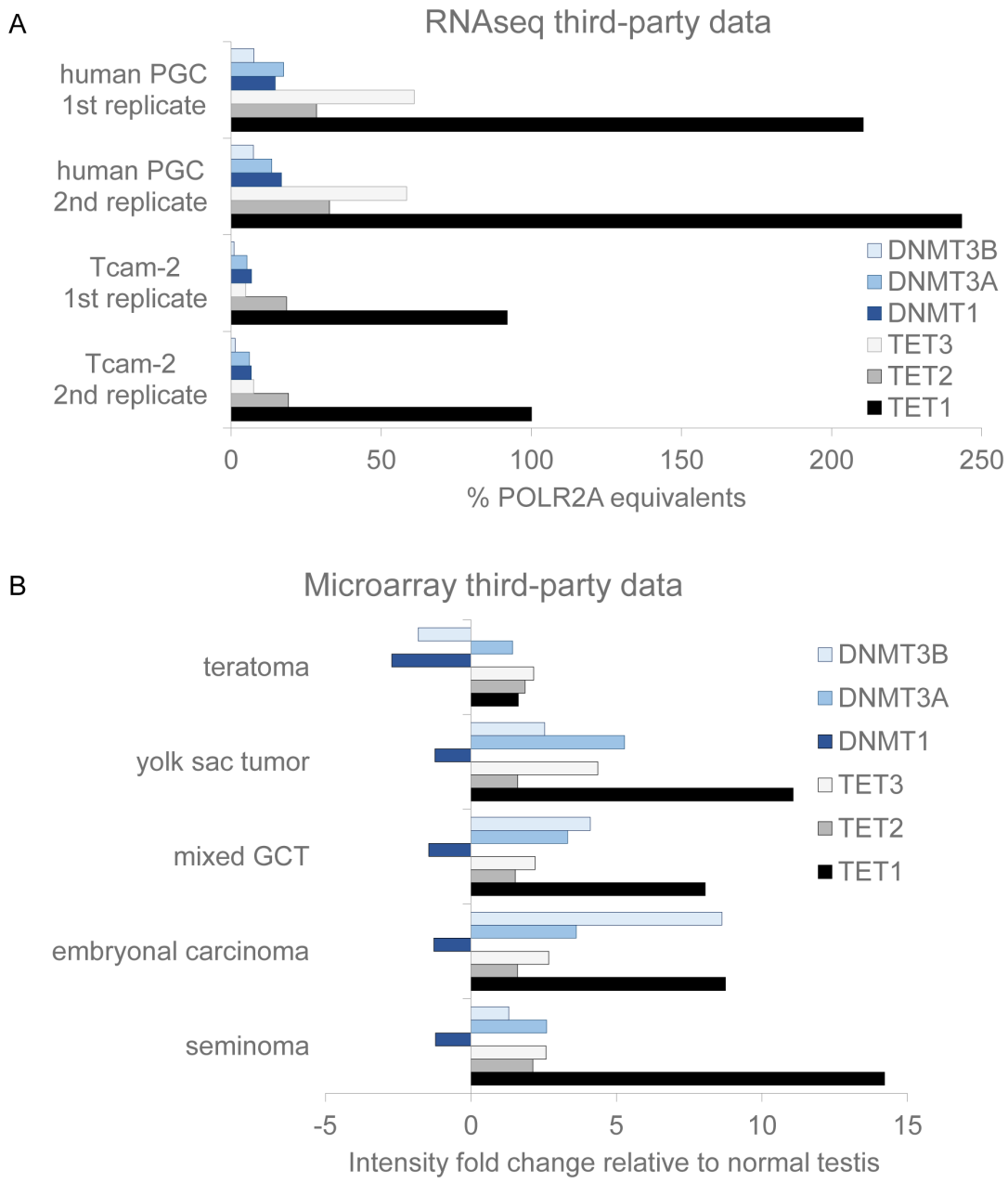

Fig. 57: RNAseq (A) and microarray (B) data for DNMT and TET genes from publicly available databases in selected samples. *RNAseq (NCBI) and microarray (Oncomine) data of DNMT and TET genes from publicly available databases are depicted. (A) DNMTs and TETs RNA seq data from human PGCs and seminoma-derived cell line TCam-2. Gene transcription was normalized to the percentage of POLR2A expression. DNMT transcription is depicted by the bars colored in the shades of blue, TET expression in grey. The Y-axis showes examined samples and genes. The X-axis represents the percentage of POLR2A. (B) DNMT and TET microarray data in testicular tumors (teratoma, yolk sac tumor, embryonal carcinoma, mixed GCT and seminoma). The Y-axis represents the type of analyzed sample, the X-axis the intensity fold change relative to normal testis. DNMT gene production is represented by the bars colored in the shades of blue, TET genes in grey. (Benešová, Trejbalová, Kučerová, et al. 2017)*

TCam-2, a seminoma-derived cell line is known to overexpress the TET1 dioxygenase (Nettersheim et al. 2013). Therefore we performed TCam-2 TET1 knock-down to examine the TET1 effect on the 5hmC level. We observed a decreased level of 5hmC after TET1 depletion. This is in contrast to seminoma tumors, where we showed a high level of TET1 expression together with a low amount of 5hmC. We propose two possible hypothese why we and others (Nettersheim et al. 2013) did not observe an increased level of 5hmC in seminomas. (1) The level of 5hmC in the seminomas is transformed very quickly to 5fC and further to 5caC and C. The kinetics of this process is not precisely understood and needs to be elucidated in the future. Moreover, only a slightly elevated level of *de novo* DNMTs in seminomas did not lead to new 5mC establishment. Nevertheless, the unchanged levels of 5mC after TET1 knock-down in the TCam-2 cell line argues against this hypothesis. (2) The existing 5mC can be protected from the oxidation to 5hmC. Also, this 5mC protection can be TCam-2-specific. The protection could possibly be caused by histone posttranslational modifications. More experiments should be done in this field to examine the hypothesis. Moreover, it would be interesting to know, whether the depletion of 5hmC after TET1 knock-down occurs predominantly in gene bodies or at transcription starts, or at both of these sites. In the TET1 knock-down experiment, we did not observe increased 5mC levels. In contrast, the increased 5mC level in seminomas can be connected with the reprogramming to embryonal carcinoma. This finding indicates, that the sole down-regulation of TET1 production in seminomas is not sufficient for the transition to embryonal carcinoma and other changes occur.

As we described the TET1 and 5hmC presence in the tumor samples, it is important to discuss the exact occurrence of TET1 and 5hmC in non-GCT controls as well. We observed TET1 staining in the germ line cells that were localized at the basal line, in the lumen of seminiferous tubules, in Sertoli cells, and in Leydig cells in the interstitium. IHC analysis of the 5hmC presence showed decreasing signal from the basal lamina to the tubular lumen. These results are in accordance with (Nettersheim et al. 2013) and correspond to the TET1 and 5hmC staining in non-tumoral testes.

It should also be mentioned that non-GCT control samples were obtained from the patients of higher age (median 45 years, Table 1) than the matched controls (median 36 years, Table 2). As spermatogenic epithelium decline during the ageing process, the stage of epithelium can differ between the groups of matching controls and non-GCT samples. Therefore, the level of TET1 and 5hmC can vary between these two groups. Investigation with a wider panel of non-GCT controls should be performed in the future.

The hypomethylated state of seminomas is associated with increased activity of mobile elements. It is known that members of HERV-K, HERV-FRD and HERV-W endogenous retrovirus families are expressed in GCTs (Herbst et al. 1998; Galli et al. 2005; Trejbalová et al. 2011). In our laboratory, we previously described expression of HERV-W member ERVWE1 and HERV-FRD member ERVFRDE1 in GCTs (Trejbalová et al. 2011). However, in this study, only a very small panel of testicular tumors was tested. Therefore, we decided to analyze *ERVWE1* and *ERVFRDE1* mRNA expression in a wider panel of GCT samples and controls to characterize this phenomenon more precisely.

The expression from both *ERVWE1* and *ERVFRDE1* loci are regulated by RNA splicing (Blond et al. 1999; Mi et al. 2000; Smallwood et al. 2003; Vargas et al. 2009) and the level of promoter DNA methylation (Matoušková et al. 2006; Gimenez et al. 2009; 2010; Trejbalová et al. 2011). Moreover, *ERVWE1* transcription in placental trophoblasts is regulated by a trophoblast-specific enhancer, which is localized in *MaLR* LTR (Prudhomme et al. 2004) and binds placenta-specific transcription factor GCM1 (C. Yu et al. 2002; Muroi et al. 2009). Analysis of *GCM1* transcription in our panel of samples did not reveal expression comparable to the BeWo cell line or placenta samples. Therefore, it seems that the GCM1 presence is not the main regulator of ERVWE1 expression in GCTs.

Our analysis showed significant overexpression of both full-length and spliced *ERVWE1* mRNA in the seminomas and to a lesser extent in the mixed GCTs. Specifically, the seminomas displayed significantly elevated levels of both full-length and spliced *ERVWE1* RNAs in comparison with the controls. On the other hand, the non-seminomas as a whole group showed significantly upregulated levels of both *ERVWE1* RNAs in comparison with non-GCT testes. Moreover, the seminomas showed significantly higher *ERVWE1* splicing efficiency when compared with the non-seminomas. The high level of the spliced *ERVWE1* transcript, especially in the seminomas, reflects to the fact, that there are conditions appropriate for efficient *ERVWE1* mRNA splicing, which is one of the prerequisites needed for the sufficient Syncytin-1 protein production. Therefore, we showed that the presence of full-length and especially spliced *ERVWE1* forms of the transcript is a hallmark of the seminoma tumor.

As *ERVWE1* mRNA expression was previously described in endometrial carcinomas (Strissel et al. 2012), we analyzed *ERVWE1* transcription in our endometrial carcinoma samples as well. However, we did not observe overexpression of either full-length *ERVWE1* RNA or spliced *ERVWE1* mRNA. However, our collection of endometrial carcinomas was smaller (six endometroid and one endometrial carcinoma) than in the study of (Strissel et al.

2012). Therefore, investigation with a wider panel of endometrial samples should be performed in the future to elucidate the level of *ERVWE1* transcription.

Hodgkin lymphomas contain giant cells. These cells are mostly multinucleated. This fact led us to the analysis of lymphoma samples for the *ERVWE1* expression. We observed a slightly elevated level of full-length *ERVWE1* RNA, but only a background level of spliced *ERVWE1* mRNA. Therefore, lymphoma samples showed very low *ERVWE1* splicing efficiency. We did not observe any differences in the expression of both types of *ERVWE1* transcripts between the groups of Hodgkin and non-Hodgkin lymphomas. Therefore, the presence of giant cells in Hodgkin lymphomas should not be connected with Syncytin-1 fusogenic protein.

Previous study that was done in our laboratory revealed not only *ERVWE1* expression in GCTs, but also transcription from the *ERVFRDE1* locus (Trejbalová et al. 2011). Therefore, we analyzed *ERVFRDE1* expression in our panel of samples. The levels of both full-length and spliced *ERVFRDE1* RNAs were lower than the *ERVWE1* RNAs in all analyzed samples. This finding is very similar to the results obtained in BeWo cells and in placental samples (Gimenez et al. 2009; Trejbalová et al. 2011). Despite low *ERVFRDE1* RNA expression in GCTs, the splicing was efficient. Collectively, *ERVFRDE1* is transcribed and efficiently spliced in GCTs, but to a much lesser extent when compared with *ERVWE1*. More research in the ERVFRDE1 field should be done to elucidate its possible role in GCTs.

BeWo, a choriocarcinoma-derived cell line, is characterized by the presence of multinucleated syncytia together with upregulated production of hGC, Syncytin-1, and increased mRNA expression of Syncytin-1 receptors *ASCT1* and *ASCT2* (Pattillo et al. 1968; Kekuda et al. 1996; Mi et al. 2000). The seminoma-derived TCam-2 cell line, where we observed elevated *ERVWE1* transcription, contains syncytia with low numbers of nuclei. In some seminomas and to a lesser extent in mixed GCTs, neoplastic giant cells or multinuclear elements with hCG production were observed (Butcher et al. 1985; von Hochstetter et al. 1985) while the presence of these cells is connected with more aggressive progression of the disease (Butcher et al. 1985). As the mRNA levels of *ASCT1* and *ASCT2* ERVWE1 receptors are unknown in the seminomas, we analyzed the panel of our GCT samples for the presence of *ASCT1* and *ASCT2* mRNA. Even though we did not observe any significant differences between the analyzed groups of samples, in most of them the level of *ASCT1* mRNA was comparable or higher than in BeWo cells. On the other hand, *ASCT2* receptor mRNA was lower in most of our samples than in BeWo cells. Collectively, GCTs contain giant cells and Syncytin-1 receptor is expressed in them. However, the connection between the presence of

giant cells and Syncytin-1 in GCTs should be examined more deeply in the future.

The level of *ERVWE1* promoter DNA methylation is crucial for mRNA expression of the *env* gene. Collectively, the *ERVWE1* promoter was hypomethylated in seminomas. Several analyzed placental and seminoma samples revealed bimodal DNA methylation of the *ERVWE1* promoter. This means that there were either completely demethylated or heavily methylated promoter sequences. The only exception in our analysis was seminoma sample T5, which showed a patchy DNA methylation pattern. These observations point to the importance of *ERVWE1* promoter DNA methylation in transcription derepression. These results further strengthen the importance of the *ERVWE1* promoter DNA methylation for transcription derepression. One of the reasons why we observed this "all or nothing" phenotype might be the fact that the analyzed tumor sample is composed of different types of cells, which can show distinct methylation patterns. The bisulfite sequencing method that we used in our study, was adjusted to avoid preferential amplification of one molecule (Trejbalová et al. 2016).

In our study, we described seminomas as hypomethylated tumors with upregulated production of TET dioxygenases. Changes in the DNA methylation can affect the expression from repetitive elements, as up to 90 % of methylated CpGs can be found in them (Bestor and Tycko 1996). There are reports that showed reactivation of HERVs and retrotransposons in embryonic stem cells (Santoni et al. 2012) and induced pluripotent stem cells (Fuchs et al. 2013; Friedli et al. 2014). Therefore, ERVWE1 and to a lesser extent ERVFRDE1 derepression in the seminoma may be a direct result of the genome DNA demethylation in this tumor. However, the observed transcription derepression does not apply to all HERVs in general. The expression analysis of HERVs either in the close proximity of *ERVWE1* (*HERV-H*) or of those that are situated on different chromosomes (*HERV-Ws* on chromosomes 4 and 21) did not show their expression patterns to be the same as for *ERVWE1* in GCTs. Therefore, *ERVWE1* expression in seminomas is independent on the transcription of other examined endogenous retroviruses.

The fact that we found TET1 dioxygenase to be overexpressed in seminomas points to the role of 5hmC in transcription derepression. However, in our seminoma samples, we observed a very low or almost undetectable level of 5hmC on the *ERVWE1* promoter. This is in agreement with the study of Tang and colleagues, who detected 5hmC in PGCs, but later during the development it become depleted (W. W. C. Tang et al. 2015). Moreover, 5hmC analysis in PGCs showed a low but still detectable level of 5hmC in genic regions, but retroelements displayed much lower amounts (F. Guo et al. 2015). As seminomas share some of the features with PGCs, the very low or undetectable level of 5hmC on the *ERVWE1*

promoter is in agreement with the above-mentioned studies.

Collectively, we detected significant overexpression of *TET1* dioxygenase, particullary in seminomas. Therefore we propound TET1 as another pluripotent marker of GCTs. Furthermore, we described significant overexpression of the *ERVWE1* transcript together with efficient *ERVWE1* splicing in the seminomas. We propose high *ERVWE1* expression with efficient RNA splicing as a marker of seminoma tumor and seminoma component of mixed GCT.

Summary

In our studies, we presented wide panel of histologically well-defined GCT samples and controls. We demonstrated upregulated levels of TET dioxygenases in GCTs, especially TET1 in seminomas. This overexpression was accompanied by decreased levels of 5mC and 5hmC. Therefore, we propose TET1 as another pluripotent marker of GCTs. We hypothesize that the high level of TET expression might affect the low level of DNA methylation in the seminomas and is able to prevent further reprogramming and differentiation of the seminoma tumor (Fig. 56).

In the same set of human GCT samples, we observed upregulated transcription of *ERVWE1* endogenous retrovirus and to a lesser extent *ERVFRDE1* as well. Splicing of both these transcripts was efficient in GCTs. Especially seminomas displayed elevated *ERVWE1* expression together with a low level of promoter CpG methylation and hydroxymethylation. Our study is the first to show all these features together. We propose that high *ERVWE1* expression and efficient splicing can be used as a marker of seminomas and seminoma component in mixed GCTs and therefore can help to distinguish them. Even though we observed *ERVWE1* expression in the seminomas, we did not find increased *ERVWE1* copy numbers or upregulated mRNA expression of the *GCM1* transcription factor. These data highlight the importance of epigenetic control of *ERVWE1* promoter in GCTs.

Collectively, seminomas are characterized by low levels of 5mC and 5hmC, high expression of TET1 dioxygenase, no mutations in *IDH1/IDH2* genes, production of factors that are characteristic of pluripotent germ cells and reactivation of *ERVWE1* transcription. As all these above-mentioned features are in contrast to the CpG island methylator phenotype of many other tumors, we suggest seminomas to be designated as tumors with the "antimethylator" phenotype.

Conclusion

- GCTs displayed elevated expression of TET dioxygenases. Especially TET1 was upregulated in seminomas.
- The overall level of DNA methylation in GCTs can be influenced by the activity of DNA methyltransferases. We observed elevated expression of *de novo* methyltransferases in seminomas.
- Seminomas showed low levels of 5mC and 5hmC. Non-seminomas displayed higher amounts of both examined cytosine modifications.
- Down-regulation of TET1 in a model cell line led to a decreased level of 5hmC and an unaltered amount of 5mC.
- GCTs showed upregulated expression of both *ERVWE1,* and to a much lesser extent of *ERVFRDE1,* together with efficient RNA splicing. Especially seminomas displayed a high level of *ERVWE1* transcription.
- Analysis of known *ERVWE1/ERVFRDE1* regulators highlights the importance of epigenetic control of their promoters in GCTs.
- *ERVWE1/ERVFRDE1* transcription is independent of the expression of other analyzed distal or proximal endogenous retroviruses in GCTs.

In our studies, we propose that the TET1 dioxygenase serves as another pluripotent marker of GCTs. We further propound high *ERVWE1* expression together with efficient RNA splicing as a marker of seminomas and seminoma component of GCT.

References

Almstrup, K., Hoei-Hansen, C. E., Wirkner, U., Blake, J., Schwager, C., Ansorge, W., Nielsen, J. E., Skakkebæk, N. E., Rajpert-De Meyts, E., and Leffers, H. (2004). Embryonic Stem Cell-like Features of Testicular Carcinoma in Situ Revealed by Genome-Wide Gene Expression Profiling. *Cancer Research* 64 (14): 4736–43. https://doi.org/10.1158/0008-5472.CAN-04-0679.

Almstrup, K., Nielsen, J. E., Mlynarska, O., Jansen, M. T., Jørgensen, A., Skakkebæk, N. E., and Rajpert-De Meyts, E. (2010). Carcinoma in Situ Testis Displays Permissive Chromatin Modifications Similar to Immature Foetal Germ Cells. *British Journal of Cancer* 103 (8): 1269–76. https://doi.org/10.1038/sj.bjc.6605880.

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P. W. S., D'Souza, Z., Nakayama, M., et al. (2016). De Novo DNA Methylation Drives 5hmC Accumulation in Mouse Zygotes. *Nature Cell Biology* 18 (2): 225–33. https://doi.org/10.1038/ncb3296.

Antony, J. M., Marle, G. Van, Opii, W., Butterfield, D. A., Mallet, F., Yong, V. W., Wallace, J. L., Deacon, R. M., Warren, K., and Power, C. (2004). Human Endogenous Retrovirus Glycoprotein-Mediated Induction of Redox Reactants Causes Oligodendrocyte Death and Demyelination. *Nature Neuroscience* 7 (10): 1088–95. https://doi.org/10.1038/nn1319.

Aplin, J. D. (1991). Implantation, Trophoblast Differentiation and Haemochorial Placentation: Mechanistic Evidence in Vivo and in Vitro. *Journal of Cell Science* 99 (4): 681–92.

Bannert, N., Hofmann, H., Block, A., and Hohn, O. (2018). HERVs New Role in Cancer: From Accused Perpetrators to Cheerful Protectors. *Frontiers in Microbiology* 9 (178). https://doi.org/10.3389/fmicb.2018.00178.

Bedford, M. T., and Helden, P. D. van. (1987). Hypomethylation of Dna in Pathological Conditions of the Human Prostate. *Cancer Research* 47 (20): 5274–76.

Benešová, M., Trejbalová, K., Kovářová, D., Vernerová, Z., Hron, T., Kučerová, D., and Hejnar, J. (2017). DNA Hypomethylation and Aberrant Expression of the Human Endogenous Retrovirus ERVWE1/Syncytin-1 in Seminomas. *Retrovirology* 14 (1). https://doi.org/10.1186/s12977-017-0342-9.

Benešová, M., Trejbalová, K., Kučerová, D., Vernerová, Z., Hron, T., Szabó, A., Amouroux, R., Klézl, P., Hajkova, P., and Hejnar, J. (2017). Overexpression of TET Dioxygenases in Seminomas Associates with Low Levels of DNA Methylation and Hydroxymethylation. *Molecular Carcinogenesis* 56 (8): 1837–50. https://doi.org/10.1002/mc.22638.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and Sequencing of a CDNA Encoding DNA Methyltransferase of Mouse Cells. The Carboxyl-Terminal Domain of the Mammalian Enzymes Is Related to Bacterial Restriction Methyltransferases. *Journal of Molecular Biology* 203 (4): 971–83. https://doi.org/10.1016/0022-2836(88)90122-2.

Bestor, T., and Tycko, B. (1996). Creation of Genomic Methylation Patterns. *Nature Genetics* 12 (4): 363–67. https://doi.org/10.1038/ng0496-363.

Bian, K., Lenz, S. A. P., Tang, Q., Chen, F., Qi, R., Jost, M., Drennan, C. L., Essigmann, J. M., Wetmore, S. D., and Li, D. (2019). DNA Repair Enzymes ALKBH2, ALKBH3, and AlkB Oxidize 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine and 5-Carboxylcytosine in Vitro. *Nucleic Acids Research* 47 (11): 5522–29. https://doi.org/10.1093/nar/gkz395.

Bjerregaard, B., Holck, S., Christensen, I. J., and Larsson, L. I. (2006). Syncytin Is Involved in Breast Cancer-Endothelial Cell Fusions. *Cellular and Molecular Life Sciences* 63 (16): 1906–11. https://doi.org/10.1007/s00018-006-6201-9.

Bjerregaard, B., Lemmen, J. G., Petersen, M. R., Østrup, E., Iversen, L. H., Almstrup, K., Larsson, L. I., and Ziebe, S. (2014). Syncytin-1 and Its Receptor Is Present in Human Gametes. *Journal of Assisted Reproduction and Genetics* 31 (5): 533–39. https://doi.org/10.1007/s10815-014-0224-1.

Bjerregard, B., Ziomkiewicz, I., Schulz, A., and Larsson, L. I. (2014). Syncytin-1 in Differentiating Human Myoblasts: Relationship to Caveolin-3 and Myogenin. *Cell and Tissue Research* 357 (1): 355–62. https://doi.org/10.1007/s00441-014-1930-9.

Blaise, S., Parseval, N. De, Bénit, L., and Heidmann, T. (2003). Genomewide Screening for Fusogenic Human Endogenous Retrovirus Envelopes Identifies Syncytin 2, a Gene Conserved on Primate Evolution. *Proceedings of the National Academy of Sciences of the United States of America* 100 (22): 13013–18. https://doi.org/10.1073/pnas.2132646100.

Blaise, S., Parseval, N. de, and Heidmann, T. (2005). Functional Characterization of Two Newly Identified Human Endogenous Retrovirus Coding Envelope Genes. *Retrovirology* 2 (19). https://doi.org/10.1186/1742-4690-2-19.

Blond, J. L., Besème, F., Duret, L., Bouton, O., Bedin, F., Perron, H., Mandrand, B., and Mallet, F. (1999). Molecular Characterization and Placental Expression of HERV-W, a New Human Endogenous Retrovirus Family. *Journal of Virology* 73 (2): 1175–85. http://www.ncbi.nlm.nih.gov/pubmed/9882319.

Blond, J. L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., Mandrand, B., Mallet, F., and Cosset, F.-L. (2000). An Envelope Glycoprotein of the Human Endogenous Retrovirus HERV-W Is Expressed in the Human Placenta and Fuses Cells Expressing the Type D Mammalian Retrovirus Receptor. *Journal of Virology* 74 (7): 3321–29. https://doi.org/10.1128/jvi.74.7.3321-3329.2000.

Bonnaud, B., Beliaeff, J., Bouton, O., Oriol, G., Duret, L., and Mallet, F. (2005). Natural History of the ERVWE1 Endogenous Retroviral Locus. *Retrovirology* 2 (57). https://doi.org/10.1186/1742-4690-2-57.

Bonnaud, B., Bouton, O., Oriol, G., Cheynet, V., Duret, L., and Mallet, F. (2004). Evidence of Selection on the Domesticated ERVWE1 Env Retroviral Element Involved in Placentation. *Molecular Biology and Evolution* 21 (10): 1895–1901. https://doi.org/10.1093/molbev/msh206.

Booth, M. J., Branco, M. R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science* 336 (6083): 934–37. https://doi.org/10.1126/science.1220671.

Butcher, D. N., Gregory, W. M., Gunter, P. A., Masters, J. R. W., and Parkinson, M. C. (1985). The Biological and Clinical Significance of HCG-Containing Cells in Seminoma. *British Journal of*

Cancer 51 (4): 473–78. https://doi.org/10.1038/bjc.1985.68.

Cakouros, D., Hemming, S., Gronthos, K., Liu, R., Zannettino, A., Shi, S., and Gronthos, S. (2019). Specific Functions of TET1 and TET2 in Regulating Mesenchymal Cell Lineage Determination. *Epigenetics & Chromatin* 12 (1). https://doi.org/10.1186/s13072-018-0247-4.

Cao, D., Li, J., Guo, C. C., Allan, R. W., and Humphrey, P. A. (2009). SALL4 Is a Novel Diagnostic Marker for Testicular Germ Cell Tumors. *American Journal of Surgical Pathology* 33 (7): 1065–77. https://doi.org/10.1097/PAS.0b013e3181a13eef.

Carella, A., Tejedor, J. R., García, M. G., Urdinguio, R. G., Bayón, G. F., Sierra, M., López, V., et al. (2019). Epigenetic Downregulation of TET3 Reduces Genome-Wide 5hmC Levels and Promotes Glioblastoma Tumorigenesis. *International Journal of Cancer*. https://doi.org/10.1002/ijc.32520.

Chang, C.-W., Chang, G.-D., and Chen, H. (2011). A Novel Cyclic AMP/Epac1/CaMKI Signaling Cascade Promotes GCM1 Desumoylation and Placental Cell Fusion. *Molecular and Cellular Biology* 31 (18): 3820–31. https://doi.org/10.1128/mcb.05582-11.

Chang, C.-W., Chuang, H.-C., Yu, C., Yao, T.-P., and Chen, H. (2005). Stimulation of GCMa Transcriptional Activity by Cyclic AMP/Protein Kinase A Signaling Is Attributed to CBP-Mediated Acetylation of GCMa. *Molecular and Cellular Biology* 25 (19): 8401–14. https://doi.org/10.1128/mcb.25.19.8401-8414.2005.

Chen, B. F., Gu, S., Suen, Y. K., Li, L., and Chan, W. Y. (2014). MicroRNA-199a-3p, DNMT3A, and Aberrant DNA Methylation in Testicular Cancer. *Epigenetics* 9 (1): 119–28. https://doi.org/10.4161/epi.25799.

Chen, C.-P., Chen, L.-F., Yang, S.-R., Chen, C.-Y., Ko, C.-C., Chang, G.-D., and Chen, H. (2008). Functional Characterization of the Human Placental Fusogenic Membrane Protein Syncytin 21. *Biology of Reproduction* 79 (5): 815–23. https://doi.org/10.1095/biolreprod.108.069765.

Chen, C. C., Wang, K. Y., and Shen, C. K. J. (2013). DNA 5-Methylcytosine Demethylation Activities of the Mammalian DNA Methyltransferases. *Journal of Biological Chemistry* 288 (13): 9084–91. https://doi.org/10.1074/jbc.M112.445585.

Chen, D., Maruschke, M., Hakenberg, O., Zimmermann, W., Stief, C. G., and Buchner, A. (2017). TOP2A, HELLS, ATAD2, and TET3 Are Novel Prognostic Markers in Renal Cell Carcinoma. *Urology* 102 (April): 265.e1-265.e7. https://doi.org/10.1016/j.urology.2016.12.050.

Chen, Y. H., Lin, T. T., Wu, Y. P., Li, X. D., Chen, S. H., Xue, X. Y., Wei, Y., Zheng, Q. S., Huang, J. B., and Xu, N. (2019). Identification of Key Genes and Pathways in Seminoma by Bioinformatics Analysis. *OncoTargets and Therapy* 12: 3683–93. https://doi.org/10.2147/OTT.S199115.

Cheng, P., Schmutte, C., Cofer, K. F., Felix, J. C., Yu, M. C., and Dubeau, L. (1997). Alterations in DNA Methylation Are Early, but Not Initial, Events in Ovarian Tumorigenesis. *British Journal of Cancer* 75 (3): 396–402. https://doi.org/10.1038/bjc.1997.64.

Cheong, M.-L., Wang, L.-J., Chuang, P.-Y., Chang, C.-W., Lee, Y.-S., Lo, H.-F., Tsai, M.-S., and Chen, H. (2015). A Positive Feedback Loop between Glial Cells Missing 1 (GCM1) and Human Chorionic Gonadotropin (HCG) Regulates Placental HCGβ Expression and Cell Differentiation. *Molecular and Cellular Biology* 36 (1): 197–209. https://doi.org/10.1128/mcb.00655-15.

Cherkasova, E., Scrivani, C., Doh, S., Weisman, Q., Takahashi, Y., Harashima, N., Yokoyama, H., et al. (2016). Detection of an Immunogenic HERV-E Envelope with Selective Expression in Clear Cell Kidney Cancer. *Cancer Research* 76 (8): 2177–85. https://doi.org/10.1158/0008-5472.CAN-15-3139.

Cheung, H. H., Yang, Y., Lee, T. L., Rennert, O., and Chan, W. Y. (2016). Hypermethylation of Genes in Testicular Embryonal Carcinomas. *British Journal of Cancer* 114 (2): 230–36. https://doi.org/10.1038/bjc.2015.408.

Cheynet, V., Ruggieri, A., Oriol, G., Blond, J.-L., Boson, B., Vachot, L., Verrier, B., Cosset, F.-L., and Mallet, F. (2005). Synthesis, Assembly, and Processing of the Env ERVWE1/Syncytin Human Endogenous Retroviral Envelope. *Journal of Virology* 79 (9): 5585–93. https://doi.org/10.1128/jvi.79.9.5585-5593.2005.

Cianciolo, G. J., Copeland, T. D., Oroszlan, S., and Snyderman, R. (1985). Inhibition of Lymphocyte Proliferation by a Synthetic Peptide Homologous to Retroviral Envelope Proteins. *Science* 230 (4724): 453–55. https://doi.org/10.1126/science.2996136.

Ciesielski, P., Jóźwiak, P., Wójcik-Krowiranda, K., Forma, E., Cwonda, Ł., Szczepaniec, S., Bieńkiewicz, A., Bryś, M., and Krześlak, A. (2017). Differential Expression of Ten-Eleven Translocation Genes in Endometrial Cancers. *Tumor Biology* 39 (3). https://doi.org/10.1177/1010428317695017.

Contreras-Galindo, R., Kaplan, M. H., Leissner, P., Verjat, T., Ferlenghi, I., Bagnoli, F., Giusti, F., et al. (2008). Human Endogenous Retrovirus K (HML-2) Elements in the Plasma of People with Lymphoma and Breast Cancer. *Journal of Virology* 82 (19): 9329–36. https://doi.org/10.1128/jvi.00646-08.

Cornelis, G., Funk, M., Vernochet, C., Leal, F., Tarazona, O. A., Meurice, G., Heidmann, O., et al. (2017). An Endogenous Retroviral Envelope Syncytin and Its Cognate Receptor Identified in the Viviparous Placental Mabuya Lizard. *Proceedings of the National Academy of Sciences of the United States of America* 114 (51): E10991–0. https://doi.org/10.1073/pnas.1714590114.

Cornelis, G., Heidmann, O., Bernard-Stoecklin, S., Reynaud, K., Véron, G., Mulot, B., Dupressoir, A., and Heidmann, T. (2012). Ancestral Capture of Syncytin-Car1, a Fusogenic Endogenous Retroviral Envelope Gene Involved in Placentation and Conserved in Carnivora. *Proceedings of the National Academy of Sciences of the United States of America* 109 (7). https://doi.org/10.1073/pnas.1115346109.

Cornelis, G., Heidmann, O., Degrelle, S. A., Vernochet, C., Lavialle, C., Letzelter, C., Bernard-Stoecklin, S., et al. (2013). Captured Retroviral Envelope Syncytin Gene Associated with the Unique Placental Structure of Higher Ruminants. *Proceedings of the National Academy of Sciences of the United States of America* 110 (9). https://doi.org/10.1073/pnas.1215787110.

Cornelis, G., Vernochet, C., Carradec, Q., Souquere, S., Mulot, B., Catzeflis, F., Nilsson, M. A., et al. (2015). Retroviral Envelope Gene Captures and Syncytin Exaptation for Placentation in Marsupials. *Proceedings of the National Academy of Sciences of the United States of America* 112 (5): E487–96. https://doi.org/10.1073/pnas.1417000112.

Cornelis, G., Vernochet, C., Malicorne, S., Souquere, S., Tzika, A. C., Goodman, S. M., Catzeflis, F., et al. (2014). Retroviral Envelope Syncytin Capture in an Ancestrally Diverged Mammalian

Clade for Placentation in the Primitive Afrotherian Tenrecs. *Proceedings of the National Academy of Sciences of the United States of America* 111 (41): E4332–41. https://doi.org/10.1073/pnas.1412268111.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Coz, M. Le, et al. (2011). Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. *Cell* 146 (1): 67–79. https://doi.org/10.1016/j.cell.2011.06.020.

Costa, A. L., Moreira-Barbosa, C., Lobo, J., Vilela-Salgueiro, B., Cantante, M., Guimarães, R., Lopes, P., et al. (2018). DNA Methylation Profiling as a Tool for Testicular Germ Cell Tumors Subtyping. *Epigenomics* 10 (12): 1511–23. https://doi.org/10.2217/epi-2018-0034.

Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., Fantin, V. R., et al. (2009). Cancer-Associated IDH1 Mutations Produce 2-Hydroxyglutarate. *Nature* 462 (7274): 739–44. https://doi.org/10.1038/nature08617.

Dawlaty, M. M., Breiling, A., Le, T., Barrasa, M. I., Raddatz, G., Gao, Q., Powell, B. E., et al. (2014). Loss of Tet Enzymes Compromises Proper Differentiation of Embryonic Stem Cells. *Developmental Cell* 29 (1): 102–11. https://doi.org/10.1016/j.devcel.2014.03.003.

Dawlaty, M. M., Ganz, K., Powell, B. E., Hu, Y. C., Markoulaki, S., Cheng, A. W., Gao, Q., et al. (2011). Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development. *Cell Stem Cell* 9 (2): 166–75. https://doi.org/10.1016/j.stem.2011.07.010.

Derynck, R., and Zhang, Y. E. (2003). Smad-Dependent and Smad-Independent Pathways in TGF-β Family Signalling. *Nature* 425 (6958): 577–84. https://doi.org/10.1038/nature02006.

Dieckmann, K. P., Radtke, A., Spiekermann, M., Balks, T., Matthies, C., Becker, P., Ruf, C., et al. (2017). Serum Levels of MicroRNA MiR-371a-3p: A Sensitive and Specific New Biomarker for Germ Cell Tumours. *European Urology* 71 (2): 213–20. https://doi.org/10.1016/j.eururo.2016.07.029.

Dupressoir, A., Lavialle, C., and Heidmann, T. (2012). From Ancestral Infectious Retroviruses to Bona Fide Cellular Genes: Role of the Captured Syncytins in Placentation. *Placenta* 33 (9): 663–71. https://doi.org/10.1016/j.placenta.2012.05.005.

Dupressoir, A., Marceau, G., Vernochet, C., Bénit, L., Kanellopoulos, C., Sapin, V., and Heidmann, T. (2005). Syncytin-A and Syncytin-B, Two Fusogenic Placenta-Specific Murine Envelope Genes of Retroviral Origin Conserved in Muridae. *Proceedings of the National Academy of Sciences of the United States of America* 102 (3): 725–30. https://doi.org/10.1073/pnas.0406509102.

Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P., and Heidmann, T. (2009). Syncytin-A Knockout Mice Demonstrate the Critical Role in Placentation of a Fusogenic, Endogenous Retrovirus-Derived, Envelope Gene. *Proceedings of the National Academy of Sciences of the United States of America* 106 (29): 12127–32. https://doi.org/10.1073/pnas.0902925106.

Dupressoir, A., Vernochet, C., Harper, F., Guégan, J., Dessen, P., Pierron, G., and Heidmann, T. (2011). A Pair of Co-Opted Retroviral Envelope Syncytin Genes Is Required for Formation of the Two-Layered Murine Placental Syncytiotrophoblast. *Proceedings of the National Academy of Sciences of the United States of America* 108 (46). https://doi.org/10.1073/pnas.1112304108.

Einhorn, L. H. (1997). Testicular Cancer: An Oncological Success Story. *Clinical Cancer Research* 3 (12 II): 2630–32.

Elhamamsy, A. R. (2017). Role of DNA Methylation in Imprinting Disorders: An Updated Review. *Journal of Assisted Reproduction and Genetics* 34 (5): 549–62. https://doi.org/10.1007/s10815-017-0895-5.

Elleder, D., Kim, O., Padhi, A., Bankert, J. G., Simeonov, I., Schuster, S. C., Wittekindt, N. E., Motameny, S., and Poss, M. (2012). Polymorphic Integrations of an Endogenous Gammaretrovirus in the Mule Deer Genome. *Journal of Virology* 86 (5): 2787–96. https://doi.org/10.1128/jvi.06859-11.

Esnault, C., Cornelis, G., Heidmann, O., and Heidmann, T. (2013). Differential Evolutionary Fate of an Ancestral Primate Endogenous Retrovirus Envelope Gene, the EnvV Syncytin, Captured for a Function in Placentation. *PLoS Genetics* 9 (3). https://doi.org/10.1371/journal.pgen.1003400.

Esnault, C., Priet, S., Ribet, D., Vernochet, C., Bruls, T., Lavialle, C., Weissenbach, J., and Heidmann, T. (2008). A Placenta-Specific Receptor for the Fusogenic, Endogenous Retrovirus-Derived, Human Syncytin-2. *Proceedings of the National Academy of Sciences of the United States of America* 105 (45): 17532–37. https://doi.org/10.1073/pnas.0807413105.

Exner, R., Pulverer, W., Diem, M., Spaller, L., Woltering, L., Schreiber, M., Wolf, B., et al. (2015). Potential of DNA Methylation in Rectal Cancer as Diagnostic and Prognostic Biomarkers. *British Journal of Cancer* 113 (7): 1035–45. https://doi.org/10.1038/bjc.2015.303.

Fan, M., He, X., and Xu, X. (2015). Restored Expression Levels of TET1 Decrease the Proliferation and Migration of Renal Carcinoma Cells. *Molecular Medicine Reports* 12 (4): 4837–42. https://doi.org/10.3892/mmr.2015.4058.

Frendo, J. L., Olivier, D., Cheynet, V., Blond, J.-L., Bouton, O., Vidaud, M., Rabreau, M., Evain-Brion, D., and Mallet, F. (2003). Direct Involvement of HERV-W Env Glycoprotein in Human Trophoblast Cell Fusion and Differentiation. *Molecular and Cellular Biology* 23 (10): 3566–74. https://doi.org/10.1128/mcb.23.10.3566-3574.2003.

Friedli, M., Turelli, P., Kapopoulou, A., Rauwel, B., Castro-Diáz, N., Rowe, H. M., Ecco, G., et al. (2014). Loss of Transcriptional Control over Endogenous Retroelements during Reprogramming to Pluripotency. *Genome Research* 24 (8): 1251–59. https://doi.org/10.1101/gr.172809.114.

Frigola, J., Song, J., Stirzaker, C., Hinshelwood, R. A., Peinado, M. A., and Clark, S. J. (2006). Epigenetic Remodeling in Colorectal Cancer Results in Coordinate Gene Suppression across an Entire Chromosome Band. *Nature Genetics* 38 (5): 540–49. https://doi.org/10.1038/ng1781.

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. (1992). A Genomic Sequencing Protocol That Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands. *Proceedings of the National Academy of Sciences of the United States of America* 89 (5): 1827–31. https://doi.org/10.1073/pnas.89.5.1827.

Fuchs, N. V., Loewer, S., Daley, G. Q., Izsvák, Z., Löwer, J., and Löwer, R. (2013). Human Endogenous Retrovirus K (HML-2) RNA and Protein Expression Is a Marker for Human Embryonic and Induced Pluripotent Stem Cells. *Retrovirology* 10 (1). https://doi.org/10.1186/1742-4690-10-115.

Funk, M., Cornelis, G., Vernochet, C., Heidmann, O., Dupressoir, A., Conley, A., Glickman, S., and Heidmann, T. (2018). Capture of a Hyena-Specific Retroviral Envelope Gene with Placental Expression Associated in Evolution with the Unique Emergence among Carnivorans of Hemochorial Placentation in Hyaenidae. *Journal of Virology* 93 (4). https://doi.org/10.1128/jvi.01811-18.

Galli, U. M., Sauter, M., Lecher, B., Maurer, S., Herbst, H., Roemer, K., and Mueller-Lantzsch, N. (2005). Human Endogenous Retrovirus Rec Interferes with Germ Cell Development in Mice and May Cause Carcinoma in Situ, the Predecessor Lesion of Germ Cell Tumors. *Oncogene* 24 (19): 3223–28. https://doi.org/10.1038/sj.onc.1208543.

Gama-Sosa, M. A., Wang, R. Y. H., Kuo, K. C., Gehrke, C. W., and Ehrlich, M. (1983). The 5-Methylcytosine Content of Highly Repeated Sequences in Human DNA. *Nucleic Acids Research* 11 (10): 3087–95. https://doi.org/10.1093/nar/11.10.3087.

Gamage, T. K. J. B., Chamley, L. W., and James, J. L. (2016). Stem Cell Insights into Human Trophoblast Lineage Differentiation. *Human Reproduction Update* 23 (1): 77–103. https://doi.org/10.1093/humupd/dmw026.

Gao, Y., He, Z., Wang, Z., Luo, Y., Sun, H., Zhou, Y., Huang, L., Li, M., Fang, Q., and Jiang, S. (2012). Increased Expression and Altered Methylation of HERVWE1 in the Human Placentas of Smaller Fetuses from Monozygotic, Dichorionic, Discordant Twins. *PLoS ONE* 7 (3). https://doi.org/10.1371/journal.pone.0033503.

García, M. G., Carella, A., Urdinguio, R. G., Bayón, G. F., Lopez, V., Tejedor, J. R., Sierra, M. I., et al. (2018). Epigenetic Dysregulation of TET2 in Human Glioblastoma. *Oncotarget* 9 (40): 25922–34. https://doi.org/10.18632/oncotarget.25406.

Giehr, P., Kyriakopoulos, C., Lepikhov, K., Wallner, S., Wolf, V., and Walter, J. (2018). Two Are Better than One: HPoxBS - Hairpin Oxidative Bisulfite Sequencing. *Nucleic Acids Research* 46 (15): e88–e88. https://doi.org/10.1093/nar/gky422.

Gillis, A. J. M., Stoop, H., Biermann, K., Gurp, R. J. H. L. M. van, Swartzman, E., Cribbes, S., Ferlinz, A., Shannon, M., Oosterhuis, J. W., and Looijenga, L. H. J. (2011). Expression and Interdependencies of Pluripotency Factors LIN28, OCT3/4, NANOG and SOX2 in Human Testicular Germ Cells and Tumours of the Testis. *International Journal of Andrology* 34 (4 PART 2). https://doi.org/10.1111/j.1365-2605.2011.01148.x.

Gimenez, J., Montgiraud, C., Oriol, G., Pichon, J. P., Ruel, K., Tsatsaris, V., Gerbaud, P., Frendo, J. L., Evain-Brion, D., and Mallet, F. (2009). Comparative Methylation of ERVWE1/Syncytin-1 and Other Human Endogenous Retrovirus LTRs in Placenta Tissues. *DNA Research* 16 (4): 195–211. https://doi.org/10.1093/dnares/dsp011.

Gimenez, J., Montgiraud, C., Pichon, J. P., Bonnaud, B., Arsac, M., Ruel, K., Bouton, O., and Mallet, F. (2010). Custom Human Endogenous Retroviruses Dedicated Microarray Identifies Self-Induced HERV-W Family Elements Reactivated in Testicular Cancer upon Methylation Control. *Nucleic Acids Research* 38 (7): 2229–46. https://doi.org/10.1093/nar/gkp1214.

Goelz, S. E., Vogelstein, B., Hamilton, S. R., and Feinberg, A. P. (1985). Hypomethylation of DNA from Benign and Malignant Human Colon Neoplasms. *Science* 228 (4696): 187–90. https://doi.org/10.1126/science.2579435.

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., Golic, K. G., Jacobsen, S. E., and Bestor, T. H. (2006). Methylation of TRNAAsp by the DNA Methyltransferase Homolog Dnmt2. *Science* 311 (5759): 395–98. https://doi.org/10.1126/science.1120976.

Gonzalez-Cao, M., Iduma, P., Karachaliou, N., Santarpia, M., Blanco, J., and Rosell, R. (2016). Human Endogenous Retroviruses and Cancer. *Cancer Biology and Medicine* 13 (4): 483–88. https://doi.org/10.20892/j.issn.2095-3941.2016.0080.

Grandi, N., and Tramontano, E. (2018). Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses. *Frontiers in Immunology* 9 (2039). https://doi.org/10.3389/fimmu.2018.02039.

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., Xie, Z. G., et al. (2011). The Role of Tet3 DNA Dioxygenase in Epigenetic Reprogramming by Oocytes. *Nature* 477 (7366): 606–12. https://doi.org/10.1038/nature10443.

Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., et al. (2015). The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell* 161 (6): 1437–52. https://doi.org/10.1016/j.cell.2015.05.015.

Guo, J. U., Su, Y., Zhong, C., Ming, G. L., and Song, H. (2011). Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell* 145 (3): 423–34. https://doi.org/10.1016/j.cell.2011.03.022.

Haffner, M. C., Taheri, D., Luidy-Imada, E., Palsgrove, D. N., Eich, M. L., Netto, G. J., Matoso, A., et al. (2018). Hypomethylation, Endogenous Retrovirus Expression, and Interferon Signaling in Testicular Germ Cell Tumors. *Proceedings of the National Academy of Sciences of the United States of America* 115 (37): E8580–82. https://doi.org/10.1073/pnas.1803292115.

Hahn, S., Ugurel, S., Hanschmann, K. M., Strobel, H., Tondera, C., Schadendorf, D., Löwer, J., and Löwer, R. (2008). Serological Response to Human Endogenous Retrovirus K in Melanoma Patients Correlates with Survival Probability. *AIDS Research and Human Retroviruses* 24 (5): 717–23. https://doi.org/10.1089/aid.2007.0286.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M. A. (2002). Epigenetic Reprogramming in Mouse Primordial Germ Cells. *Mechanisms of Development* 117 (1–2): 15–23. https://doi.org/10.1016/S0925-4773(02)00181-8.

Hajkova, P., Jeffries, S. J., Lee, C., Miller, N., Jackson, S. P., and Surani, M. A. (2010). Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway. *Science* 329 (5987): 78–82. https://doi.org/10.1126/science.1187945.

Hamilton-Dutoit, S. J., Lou, H., and Pallesen, G. (1990). The Expression of Placental Alkaline Phosphatase (PLAP) and PLAP-like Enzymes in Normal and Neoplastic Human Tissues. *APMIS* 98 (7–12): 797–811. https://doi.org/10.1111/j.1699-0463.1990.tb05000.x.

Handwerger, S. (2010). New Insights into the Regulation of Human Cytotrophoblast Cell Differentiation. *Molecular and Cellular Endocrinology* 323 (1): 94–104. https://doi.org/10.1016/j.mce.2009.12.015.

Hanger, J. J., Bromham, L. D., McKee, J. J., O'Brien, T. M., and Robinson, W. F. (2000). The Nucleotide Sequence of Koala (Phascolarctos Cinereus) Retrovirus: A Novel Type C Endogenous

Virus Related to Gibbon Ape Leukemia Virus. *Journal of Virology* 74 (9): 4264–72. https://doi.org/10.1128/jvi.74.9.4264-4272.2000.

Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L Cooperates with the Dnmt3 Family of de Novo DNA Methyltransferases to Establish Maternal Imprints in Mice. *Development* 129 (8): 1983–93. https://dev.biologists.org/content/129/8/1983.long.

He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., et al. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* 333 (6047): 1303–7. https://doi.org/10.1126/science.1210944.

Heidmann, O., Vernochet, C., Dupressoir, A., and Heidmann, T. (2009). Identification of an Endogenous Retroviral Envelope Gene with Fusogenic Activity and Placenta-Specific Expression in the Rabbit: A New "Syncytin" in a Third Order of Mammals. *Retrovirology* 6 (107). https://doi.org/10.1186/1742-4690-6-107.

Hendrich, B., and Bird, A. (1998). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Molecular and Cellular Biology* 18 (11): 6538–47. https://doi.org/10.1128/mcb.18.11.6538.

Herbst, H., Sauter, M., Kühler-Obbarius, C., Löning, T., and Mueller-Lantzsch, N. (1998). Human Endogenous Retrovirus (HERV)-K Transcripts in Germ Cell and Trophoblastic Tumours. *APMIS* 106 (1): 216–20. https://doi.org/10.1111/j.1699-0463.1998.tb01338.x.

Hill, P. W. S., Leitch, H. G., Requena, C. E., Sun, Z., Amouroux, R., Roman-Trufero, M., Borkowska, M., et al. (2018). Epigenetic Reprogramming Enables the Transition from Primordial Germ Cell to Gonocyte. *Nature* 555 (7696): 392–96. https://doi.org/10.1038/nature25964.

Hochstetter, A. R. von, Sigg, C., Saremaslani, P., and Hedinger, C. (1985). The Significance of Giant Cells in Human Testicular Seminomas - A Clinico-Pathological Study. *Virchows Archiv A Pathological Anatomy and Histopathology* 407 (3): 309–22. https://doi.org/10.1007/BF00710656.

Hoei-Hansen, C. E., Almstrup, K., Nielsen, J. E., Brask Sonne, S., Graem, N., Skakkebaek, N. E., Leffers, H., and Rajpert-De Meyts, E. (2005). Stem Cell Pluripotency Factor NANOG Is Expressed in Human Fetal Gonocytes, Testicular Carcinoma in Situ and Germ Cell Tumours. *Histopathology* 47 (1): 48–56. https://doi.org/10.1111/j.1365-2559.2005.02182.x.

Hosoya, T., Takizawa, K., Nitta, K., and Hotta, Y. (1995). Glial Cells Missing: A Binary Switch between Neuronal and Glial Determination in Drosophila. *Cell* 82 (6): 1025–36. https://doi.org/10.1016/0092-8674(95)90281-3.

Huang, Q., Li, J., Wang, F., Oliver, M. T., Tipton, T., Gao, Y., and Jiang, S. W. (2013). Syncytin-1 Modulates Placental Trophoblast Cell Proliferation by Promoting G1/S Transition. *Cellular Signalling* 25 (4): 1027–35. https://doi.org/10.1016/j.cellsig.2013.01.008.

Imakawa, K., Nakagawa, S., and Miyazawa, T. (2015). Baton Pass Hypothesis: Successive Incorporation of Unconserved Endogenous Retroviral Genes for Placentation during Mammalian Evolution. *Genes to Cells* 20 (10): 771–88. https://doi.org/10.1111/gtc.12278.

Inoue, A., Shen, L., Matoba, S., and Zhang, Y. (2015). Haploinsufficiency, but Not Defective Paternal 5mC Oxidation, Accounts for the Developmental Defects of Maternal Tet3 Knockouts. *Cell*

Reports 10 (4): 463–70. https://doi.org/10.1016/j.celrep.2014.12.049.

Ito, S., Dalessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., and Zhang, Y. (2010). Role of Tet Proteins in 5mC to 5hmC Conversion, ES-Cell Self-Renewal and Inner Cell Mass Specification. *Nature* 466 (7310): 1129–33. https://doi.org/10.1038/nature09303.

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* 333 (6047): 1300–1303. https://doi.org/10.1126/science.1210597.

Jern, P., and Coffin, J. M. (2008). Effects of Retroviruses on Host Genome Function. *Annual Review of Genetics* 42 (1): 709–32. https://doi.org/10.1146/annurev.genet.42.110807.091501.

Jern, P., Sperber, G. O., and Blomberg, J. (2005). Use of Endogenous Retroviral Sequences (ERVs) and Structural Markers for Retroviral Phylogenetic Inference and Taxonomy. *Retrovirology* 2 (50). https://doi.org/10.1186/1742-4690-2-50.

Jin, S. G., Kadam, S., and Pfeifer, G. P. (2010). Examination of the Specificity of DNA Methylation Profiling Techniques towards 5-Methylcytosine and 5-Hydroxymethylcytosine. *Nucleic Acids Research* 38 (11). https://doi.org/10.1093/nar/gkq223.

Jones, B. W., Fetter, R. D., Tear, G., and Goodman, C. S. (1995). Glial Cells Missing: A Genetic Switch That Controls Glial versus Neuronal Fate. *Cell* 82 (6): 1013–23. https://doi.org/10.1016/0092-8674(95)90280-5.

Kehler, J., Tolkunova, E., Koschorz, B., Pesce, M., Gentile, L., Boiani, M., Lomelí, H., et al. (2004). Oct4 Is Required for Primordial Germ Cell Survival. *EMBO Reports* 5 (11): 1078–83. https://doi.org/10.1038/sj.embor.7400279.

Kekuda, R., Prasad, P. D., Fei, Y. J., Torres-Zamorano, V., Sinha, S., Yang-Feng, T. L., Leibach, F. H., and Ganapathy, V. (1996). Cloning of the Sodium-Dependent, Broad-Scope, Neutral Amino Acid Transporter B(o) from a Human Placental Choriocarcinoma Cell Line. *Journal of Biological Chemistry* 271 (31): 18657–61. https://doi.org/10.1074/jbc.271.31.18657.

Keryer, G., Alsat, E., Taskén, K., and Evain-Brion, D. (1998). Cyclic AMP-Dependent Protein Kinases and Human Trophoblast Cell Differentiation in Vitro. *Journal of Cell Science* 111 (7): 995–1004.

Knerr, I., Beinder, E., and Rascher, W. (2002). Syncytin, a Novel Human Endogenous Retroviral Gene in Human Placenta: Evidence for Its Dysregulation in Preeclampsia and HELLP Syndrome. *American Journal of Obstetrics and Gynecology* 186 (2): 210–13. https://doi.org/10.1067/mob.2002.119636.

Knerr, I., Schubert, S. W., Wich, C., Amann, K., Aigner, T., Vogler, T., Jung, R., Dötsch, J., Rascher, W., and Hashemolhosseini, S. (2005). Stimulation of GCMa and Syncytin via CAMP Mediated PKA Signaling in Human Trophoblastic Cells under Normoxic and Hypoxic Conditions. *FEBS Letters* 579 (18): 3991–98. https://doi.org/10.1016/j.febslet.2005.06.029.

Knöfler, M., and Pollheimer, J. (2013). Human Placental Trophoblast Invasion and Differentiation: A Particular Focus on Wnt Signaling. *Frontiers in Genetics* 4 (190). https://doi.org/10.3389/fgene.2013.00190.

Knöfler, M., Saleh, L., Strohmer, H., Husslein, P., and Wolschek, M. F. (1999). Cyclic AMP- and Differentiation-Dependent Regulation of the Proximal AHCG Gene Promoter in Term Villous

Trophoblasts. *Molecular Human Reproduction* 5 (6): 573–80. https://doi.org/10.1093/molehr/5.6.573.

Ko, M., An, J., Bandukwala, H. S., Chavez, L., Äijö, T., Pastor, W. A., Segal, M. F., et al. (2013). Modulation of TET2 Expression and 5-Methylcytosine Oxidation by the CXXC Domain Protein IDAX. *Nature* 497 (7447): 122–26. https://doi.org/10.1038/nature12052.

Koh, K. P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., et al. (2011). Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage Specification in Mouse Embryonic Stem Cells. *Cell Stem Cell* 8 (2): 200–213. https://doi.org/10.1016/j.stem.2011.01.008.

Korkola, J. E., Houldsworth, J., Chadalavada, R. S. V., Olshen, A. B., Dobrzynski, D., Reuter, V. E., Bosl, G. J., and Chaganti, R. S. K. (2006). Down-Regulation of Stem Cell Genes, Including Those in a 200-Kb Gene Cluster at 12p13.31, Is Associated with in Vivo Differentiation of Human Male Germ Cell Tumors. *Cancer Research* 66 (2): 820–27. https://doi.org/10.1158/0008-5472.CAN-05-2445.

Korkola, J. E., Houldsworth, J., Dobrzynski, D., Olshen, A. B., Reuter, V. E., Bosl, G. J., and Chaganti, R. S. K. (2005). Gene Expression-Based Classification of Nonseminomatous Male Germ Cell Tumors. *Oncogene* 24 (32): 5101–7. https://doi.org/10.1038/sj.onc.1208694.

Korkola, J. E., Houldsworth, J., Feldman, D. R., Olshen, A. B., Qin, L. X., Patil, S., Reuter, V. E., Bosl, G. J., and Chaganti, R. S. K. (2009). Identification and Validation of a Gene Expression Signature That Predicts Outcome in Adult Men with Germ Cell Tumors. *Journal of Clinical Oncology* 27 (31): 5240–47. https://doi.org/10.1200/JCO.2008.20.0386.

Kriaucionis, S., and Heintz, N. (2009). The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* 324 (5929): 929–30. https://doi.org/10.1126/science.1169786.

Krishnamurthy, J., Rabinovich, B. A., Mi, T., Switzer, K. C., Olivares, S., Maiti, S. N., Plummer, J. B., et al. (2015). Genetic Engineering of T Cells to Target HERV-K, an Ancient Retrovirus on Melanoma. *Clinical Cancer Research* 21 (14): 3241–51. https://doi.org/10.1158/1078-0432.CCR-14-3197.

Kristensen, D. G., Nielsen, J. E., Jørgensen, A., Skakkebæk, N. E., Rajpert-De Meyts, E., and Almstrup, K. (2014). Evidence That Active Demethylation Mechanisms Maintain the Genome of Carcinoma in Situ Cells Hypomethylated in the Adult Testis. *British Journal of Cancer* 110 (3): 668–78. https://doi.org/10.1038/bjc.2013.727.

Kristensen, D. G., Skakkebæk, N. E., Rajpert-De Meyts, E., and Almstrup, K. (2013). Epigenetic Features of Testicular Germ Cell Tumours in Relation to Epigenetic Characteristics of Foetal Germ Cells. *International Journal of Developmental Biology* 57 (2–4): 309–17. https://doi.org/10.1387/ijdb.130142ka.

Kudo, Y., Tateishi, K., Yamamoto, K., Yamamoto, S., Asaoka, Y., Ijichi, H., Nagae, G., Yoshida, H., Aburatani, H., and Koike, K. (2012). Loss of 5-Hydroxymethylcytosine Is Accompanied with Malignant Cellular Transformation. *Cancer Science* 103 (4): 670–76. https://doi.org/10.1111/j.1349-7006.2012.02213.x.

Lander, S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., et al. (2001).

Initial Sequencing and Analysis of the Human Genome International Human Genome Sequencing Consortium* The Sanger Centre: Beijing Genomics Institute/Human Genome Center. *Nature* 409 (6822): 860–921. https://doi.org/10.1038/35057062.

Langbein, M., Strick, R., Strissel, P. L., Vogt, N., Parsch, H., Beckmann, M. W., and Schild, R. L. (2008). Impaired Cytotrophoblast Cell-Cell Fusion Is Associated with Reduced Syncytin and Increased Apoptosis in Patients with Placental Dysfunction. *Molecular Reproduction and Development* 75 (1): 175–83. https://doi.org/10.1002/mrd.20729.

Langemeijer, S. M. C., Kuiper, R. P., Berends, M., Knops, R., Aslanyan, M. G., Massop, M., Stevens-Linders, E., et al. (2009). Acquired Mutations in TET2 Are Common in Myelodysplastic Syndromes. *Nature Genetics* 41 (7): 838–42. https://doi.org/10.1038/ng.391.

Larsen, J. M., Christensen, I. J., Nielsen, H. J., Hansen, U., Bjerregaard, B., Talts, J. F., and Larsson, L. I. (2009). Syncytin Immunoreactivity in Colorectal Cancer: Potential Prognostic Impact. *Cancer Letters* 280 (1): 44–49. https://doi.org/10.1016/j.canlet.2009.02.008.

Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., and Heidmann, T. (2013). Paleovirology of "Syncytins", Retroviral Env Genes Exapted for a Role in Placentation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1626). https://doi.org/10.1098/rstb.2012.0507.

Leroy, X., Augusto, D., Leteurtre, E., and Gosselin, B. (2002). CD30 and CD117 (c-Kit) Used in Combination Are Useful for Distinguishing Embryonal Carcinoma from Seminoma. *Journal of Histochemistry and Cytochemistry* 50 (2): 283–85. https://doi.org/10.1177/002215540205000216.

Li, Z., Cai, X., Cai, C. L., Wang, J., Zhang, W., Petersen, B. E., Yang, F. C., and Xu, M. (2011). Deletion of Tet2 in Mice Leads to Dysregulated Hematopoietic Stem Cells and Subsequent Development of Myeloid Malignancies. *Blood* 118 (17): 4509–18. https://doi.org/10.1182/blood-2010-12-325241.

Lian, C. G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., et al. (2012). Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma. *Cell* 150 (6): 1135–46. https://doi.org/10.1016/j.cell.2012.07.033.

Lifschitz-Mercer, B., Elliott, D. J., Issakov, J., Leider-Trejo, L., Schreiber, L., Misonzhnik, F., Eisenthal, A., and Maymon, B. (2002). Localization of a Specific Germ Cell Marker, DAZL1, in Testicular Germ Cell Neoplasias. *Virchows Archiv* 440 (4): 387–91. https://doi.org/10.1007/s004280100528.

Lin, C. H., Hsieh, S. Y., Sheen, I. S., Lee, W. C., Chen, T. C., Shyu, W. C., and Liaw, Y. F. (2001). Genome-Wide Hypomethylation in Hepatocellular Carcinogenesis. *Cancer Research* 61 (10): 4238–43.

Lin, F. Y., Chang, C. W., Cheong, M. L., Chen, H. C., Lee, D. Y., Chang, G. D., and Chen, H. (2011). Dual-Specificity Phosphatase 23 Mediates GCM1 Dephosphorylation and Activation. *Nucleic Acids Research* 39 (3): 848–61. https://doi.org/10.1093/nar/gkq838.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., et al. (2009). Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences. *Nature* 462 (7271): 315–22. https://doi.org/10.1038/nature08514.

Lister, T. A., Crowther, D., Sutcliffe, S. B., Glatstein, E., Canellos, G. P., Young, R. C., Rosenberg, S. A., Coltman, C. A., and Tubiana, M. (1989). Report of a Committee Convened to Discuss the Evaluation and Staging of Patients with Hodgkin's Disease: Cotswolds Meeting. *Journal of Clinical Oncology* 7 (11): 1630–36. https://doi.org/10.1200/JCO.1989.7.11.1630.

Looijenga, L. H. J., Stoop, H., Leeuw, H. P. J. C. De, Gouveia Brazao, C. A. De, Gillis, A. J. M., Roozendaal, K. E. P. Van, Zoelen, E. J. J. Van, et al. (2003). POU5F1 (OCT3/4) Identifies Cells with Pluripotent Potential in Human Germ Cell Tumors. *Cancer Research* 63 (9): 2244–50. https://cancerres.aacrjournals.org/content/63/9/2244.long.

Lorsback, R. B., Moore, J., Mathew, S., Raimondi, S. C., Mukatira, S. T., and Downing, J. R. (2003). TET1, a Member of a Novel Protein Family, Is Fused to MLL in Acute Myeloid Leukemia Containing the t(10;11)(Q22;23) [3]. *Leukemia* 17 (3): 637–41. https://doi.org/10.1038/sj.leu.2402834.

Löwer, R., Löwer, J., Tondera-Koch, C., and Kurth, R. (1993). A General Method for the Identification of Transcribed Retrovirus Sequences (R-U5 PCR) Reveals the Expression of the Human Endogenous Retrovirus Loci HERV-H and HERV-K in Teratocarcinoma Cells. *Virology* 192 (2): 501–11. https://doi.org/10.1006/viro.1993.1066.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014). Role of Tet Proteins in Enhancer Activity and Telomere Elongation. *Genes and Development* 28 (19): 2103–19. https://doi.org/10.1101/gad.248005.114.

Maiti, A., and Drohat, A. C. (2011). Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential Implications for Active Demethylation of CpG Sites. *Journal of Biological Chemistry* 286 (41): 35334–38. https://doi.org/10.1074/jbc.C111.284620.

Makaroun, S., and Himes, K. (2018). Differential Methylation of Syncytin-1 and 2 Distinguishes Fetal Growth Restriction from Physiologic Small for Gestational Age. *American Journal of Perinatology Reports* 08 (01): e18–24. https://doi.org/10.1055/s-0038-1627473.

Malassiné, A., Blaise, S., Handschuh, K., Lalucque, H., Dupressoir, A., Evain-Brion, D., and Heidmann, T. (2007). Expression of the Fusogenic HERV-FRD Env Glycoprotein (Syncytin 2) in Human Placenta Is Restricted to Villous Cytotrophoblastic Cells. *Placenta* 28 (2–3): 185–91. https://doi.org/10.1016/j.placenta.2006.03.001.

Malassiné, A., Handschuh, K., Tsatsaris, V., Gerbaud, P., Cheynet, V., Oriol, G., Mallet, F., and Evain-Brion, D. (2005). Expression of HERV-W Env Glycoprotein (Syncytin) in the Extravillous Trophoblast of First Trimester Human Placenta. *Placenta* 26 (7): 556–62. https://doi.org/10.1016/j.placenta.2004.09.002.

Malhotra, S. S., Banerjee, P., Chaudhary, P., Pal, R., and Gupta, S. K. (2017). Relevance of Wnt10b and Activation of β-Catenin/GCMa/Syncytin-1 Pathway in BeWo Cell Fusion. *American Journal of Reproductive Immunology* 78 (4). https://doi.org/10.1111/aji.12676.

Malhotra, S. S., Suman, P., and Gupta, S. K. (2015). Alpha or Beta Human Chorionic Gonadotropin Knockdown Decrease BeWo Cell Fusion by Down-Regulating PKA and CREB Activation. *Scientific Reports* 5 (June). https://doi.org/10.1038/srep11210.

Maliniemi, P., Vincendeau, M., Mayer, J., Frank, O., Hahtola, S., Karenko, L., Carlsson, E., et al. (2013). Expression of Human Endogenous Retrovirus-W Including Syncytin-1 in Cutaneous T-

Cell Lymphoma. *PLoS ONE* 8 (10). https://doi.org/10.1371/journal.pone.0076281.

Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G., Duret, L., and Mandrand, B. (2004). The Endogenous Retroviral Locus ERVWE1 Is a Bona Fide Gene Involved in Hominoid Placental Physiology. *Proceedings of the National Academy of Sciences of the United States of America* 101 (6): 1731–36. https://doi.org/10.1073/pnas.0305763101.

Malone, C. S., Miner, M. D., Doerr, J. R., Jackson, J. P., Jacobsen, S. E., Wall, R., and Teitell, M. (2001). CmC(A/T)GG DNA Methylation in Mature B Cell Lymphoma Gene Silencing. *Proceedings of the National Academy of Sciences of the United States of America* 98 (18): 10404–9. https://doi.org/10.1073/pnas.181206898.

Mangeney, M., Renard, M., Schlecht-Louf, G., Bouallaga, I., Heidmann, O., Letzelter, C., Richaud, A., Ducos, B., and Heidmann, T. (2007). Placental Syncytins: Genetic Disjunction between the Fusogenic and Immunosuppressive Activity of Retroviral Envelope Proteins. *Proceedings of the National Academy of Sciences of the United States of America* 104 (51): 20534–39. https://doi.org/10.1073/pnas.0707873105.

Marin, M., Tailor, C. S., Nouri, A., and Kabat, D. (2000). Sodium-Dependent Neutral Amino Acid Transporter Type 1 Is an Auxiliary Receptor for Baboon Endogenous Retrovirus. *Journal of Virology* 74 (17): 8085–93. https://doi.org/10.1128/jvi.74.17.8085-8093.2000.

Matoušková, M., Blažková, J., Pajer, P., Pavlíček, A., and Hejnar, J. (2006). CpG Methylation Suppresses Transcriptional Activity of Human Syncytin-1 in Non-Placental Tissues. *Experimental Cell Research* 312 (7): 1011–20. https://doi.org/10.1016/j.yexcr.2005.12.010.

Matsuura, K., Jigami, T., Taniue, K., Morishita, Y., Adachi, S., Senda, T., Nonaka, A., Aburatani, H., Nakamura, T., and Akiyama, T. (2011). Identification of a Link between Wnt/I 2-Catenin Signalling and the Cell Fusion Pathway. *Nature Communications* 2 (1). https://doi.org/10.1038/ncomms1551.

Mi, S., Lee, X., Li, X. ping, Veldman, G. M., Finnerty, H., Racie, L., LaVallie, E., et al. (2000). Syncytin Is a Captive Retroviral Envelope Protein Involved in Human Placental Morphogenesis. *Nature* 403 (6771): 785–89. https://doi.org/10.1038/35001608.

Miettinen, M., Wang, Z., McCue, P. A., Sarlomo-Rikala, M., Rys, J., Biernat, W., Lasota, J., and Lee, Y.-S. (2014). SALL4 Expression in Germ Cell and Non-Germ Cell Tumors: A Systematic Immunohistochemical Study of 3215 Cases. *The American Journal of Surgical Pathology* 38 (3): 410–20. https://doi.org/10.1097/PAS.0000000000000116.

Misawa, K., Imai, A., Mochizuki, D., Mima, M., Endo, S., Misawa, Y., Kanazawa, T., and Mineta, H. (2018). Association of TET3 Epigenetic Inactivation with Head and Neck Cancer. *Oncotarget* 9 (36): 24480–93. https://doi.org/10.18632/oncotarget.25333.

Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E., and Ulbright, T. M. (2016). The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part A: Renal, Penile, and Testicular Tumours. *European Urology* 70 (1): 93–105. https://doi.org/10.1016/j.eururo.2016.02.029.

Møller, A. M. J., Delaissé, J. M., and Søe, K. (2017). Osteoclast Fusion: Time-Lapse Reveals Involvement of CD47 and Syncytin-1 at Different Stages of Nuclearity. *Journal of Cellular Physiology* 232 (6): 1396–1403. https://doi.org/10.1002/jcp.25633.

Morgan, H. D., Dean, W., Coker, H. A., Reik, W., and Petersen-Mahrt, S. K. (2004). Activation-Induced Cytidine Deaminase Deaminates 5-Methylcytosine in DNA and Is Expressed in Pluripotent Tissues: Implications for Epigenetic Reprogramming. *Journal of Biological Chemistry* 279 (50): 52353–60. https://doi.org/10.1074/jbc.M407695200.

Muir, A., Lever, A., and Moffett, A. (2004). Expression and Functions of Human Endogenous Retroviruses in the Placenta: An Update. *Placenta* 25 (SUPPL. A): 16–25. https://doi.org/10.1016/j.placenta.2004.01.012.

Murata, A., Baba, Y., Ishimoto, T., Miyake, K., Kosumi, K., Harada, K., Kurashige, J., et al. (2015). TET Family Proteins and 5-Hydroxymethylcytosine in Esophageal Squamous Cell Carcinoma. *Oncotarget* 6 (27): 23372–82. https://doi.org/10.18632/oncotarget.4281.

Muroi, Y., Sakurai, T., Hanashi, A., Kubota, K., Nagaoka, K., and Imakawa, K. (2009). CD9 Regulates Transcription Factor GCM1 and ERVWE1 Expression through the CAMP/Protein Kinase a Signaling Pathway. *Reproduction* 138 (6): 945–51. https://doi.org/10.1530/REP-09-0082.

Musshoff, K. (1977). [Clinical Staging Classification of Non-Hodgkin's Lymphomas (Author's Transl)]. *Strahlentherapie* 153 (4): 218–21. http://www.ncbi.nlm.nih.gov/pubmed/857349.

Nait-Oumesmar, B., Copperman, A. B., and Lazzarini, R. A. (2000). Placental Expression and Chromosomal Localization of the Human Gcm 1 Gene. *Journal of Histochemistry and Cytochemistry* 48 (7): 915–22. https://doi.org/10.1177/002215540004800704.

Nakajima, H., and Kunimoto, H. (2014). TET2 as an Epigenetic Master Regulator for Normal and Malignant Hematopoiesis. *Cancer Science* 105 (9): 1093–99. https://doi.org/10.1111/cas.12484.

Nakaya, Y., Koshi, K., Nakagawa, S., Hashizume, K., and Miyazawa, T. (2013). Fematrin-1 Is Involved in Fetomaternal Cell-to-Cell Fusion in Bovinae Placenta and Has Contributed to Diversity of Ruminant Placentation. *Journal of Virology* 87 (19): 10563–72. https://doi.org/10.1128/jvi.01398-13.

Nettersheim, D., Heukamp, L. C., Fronhoffs, F., Grewe, M. J., Haas, N., Waha, A., Honecker, F., Waha, A., Kristiansen, G., and Schorle, H. (2013). Analysis of TET Expression/Activity and 5mC Oxidation during Normal and Malignant Germ Cell Development. *PLoS ONE* 8 (12). https://doi.org/10.1371/journal.pone.0082881.

Nettersheim, D., Jostes, S., Sharma, R., Schneider, S., Hofmann, A., Ferreira, H. J., Hoffmann, P., Kristiansen, G., Esteller, M. B., and Schorle, H. (2015). BMP Inhibition in Seminomas Initiates Acquisition of Pluripotency via NODAL Signaling Resulting in Reprogramming to an Embryonal Carcinoma. *PLoS Genetics* 11 (7). https://doi.org/10.1371/journal.pgen.1005415.

Netto, G. J., Nakai, Y., Nakayama, M., Jadallah, S., Toubaji, A., Nonomura, N., Albadine, R., et al. (2008). Global DNA Hypomethylation in Intratubular Germ Cell Neoplasia and Seminoma, but Not in Nonseminomatous Male Germ Cell Tumors. *Modern Pathology* 21 (11): 1337–44. https://doi.org/10.1038/modpathol.2008.127.

Nikolaou, M., Valavanis, C., Aravantinos, G., Fountzilas, G., Tamvakis, N., Lekka, I., Arapantoni-Dadioti, P., et al. (2007). Kit Expression in Male Germ Cell Tumors. *Anticancer Research* 27 (3 B): 1685–88.

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA Methyltransferases Dnmt3a and

Dnmt3b Are Essential for de Novo Methylation and Mammalian Development. *Cell* 99 (3): 247–57. https://doi.org/10.1016/S0092-8674(00)81656-6.

Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., and Hayashi, Y. (2002). LCX, Leukemia-Associated Protein with a CXXC Domain, Is Fused to MLL in Acute Myeloid Leukemia with Trilineage Dysplasia Having t(10;11)(Q22;Q23). *Cancer Research* 62 (14): 4075–80.

Oosterhuis, J. W., and Looijenga, L. H. J. (2005). Testicular Germ-Cell Tumours in a Broader Perspective. *Nature Reviews Cancer* 5 (3): 210–22. https://doi.org/10.1038/nrc1568.

Park, J. S., Kim, J., Elghiaty, A., and Ham, W. S. (2018). Recent Global Trends in Testicular Cancer Incidence and Mortality. *Medicine (United States)* 97 (37). https://doi.org/10.1097/MD.0000000000012390.

Parseval, N. de, Diop, G., Blaise, S., Helle, F., Vasilescu, A., Matsuda, F., and Heidmann, T. (2005). Comprehensive Search for Intra- and Inter-Specific Sequence Polymorphisms among Coding Envelope Genes of Retroviral Origin Found in the Human Genome: Genes and Pseudogenes. *BMC Genomics* 6 (117). https://doi.org/10.1186/1471-2164-6-117.

Patience, C., Wilkinson, D. A., and Weiss, R. A. (1997). Our Retroviral Heritage. *Trends in Genetics* 13 (3): 116–20. https://doi.org/10.1016/S0168-9525(97)01057-3.

Pattillo, R. A., Gey, G. O., Delfs, E., and Mattingly, R. F. (1968). Human Hormone Production in Vitro. *Science* 159 (3822): 1467–69. https://doi.org/10.1126/science.159.3822.1467.

Pauls, K., Jäger, R., Weber, S., Wardelmann, E., Koch, A., Büttner, R., and Schorle, H. (2005). Transcription Factor AP-2γ, a Novel Marker of Gonocytes and Seminomatous Germ Cell Tumors. *International Journal of Cancer* 115 (3): 470–77. https://doi.org/10.1002/ijc.20913.

Penn, N. W., Suwalski, R., O'Riley, C., Bojanowski, K., and Yura, R. (1972). The Presence of 5-Hydroxymethylcytosine in Animal Deoxyribonucleic Acid. *The Biochemical Journal* 126 (4): 781–90. https://doi.org/10.1042/bj1260781.

Prabhakar, N., Sethi, B., Nagger, S., and Saxena, A. (2017). Synchronous Seminoma in Abdominopelvic and Inguinal Testes: A Rare Presentation with Unusual Morphology. *Case Reports in Pathology* 2017: 1–6. https://doi.org/10.1155/2017/6179861.

Prudhomme, S., Oriol, G., and Mallet, F. (2004). A Retroviral Promoter and a Cellular Enhancer Define a Bipartite Element Which Controls Env ERVWE1 Placental Expression. *Journal of Virology* 78 (22): 12157–68. https://doi.org/10.1128/jvi.78.22.12157-12168.2004.

Quivoron, C., Couronné, L., Valle, V. Della, Lopez, C. K., Plo, I., Wagner-Ballon, O., Cruzeiro, M. Do, et al. (2011). TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer Cell* 20 (1): 25–38. https://doi.org/10.1016/j.ccr.2011.06.003.

Rai, K., Huggins, I. J., James, S. R., Karpf, A. R., Jones, D. A., and Cairns, B. R. (2008). DNA Demethylation in Zebrafish Involves the Coupling of a Deaminase, a Glycosylase, and Gadd45. *Cell* 135 (7): 1201–12. https://doi.org/10.1016/j.cell.2008.11.042.

Rajpert-De Meyts, E. (2006). Developmental Model for the Pathogenesis of Testicular Carcinoma in Situ: Genetic and Environmental Aspects. *Human Reproduction Update* 12 (3): 303–23. https://doi.org/10.1093/humupd/dmk006.

Rajpert-De Meyts, E., Hanstein, R., Jørgensen, N., Græm, N., Vogt, P. H., and Skakkebæk, N. E. (2004). Developmental Expression of POU5F1 (OCT-3/4) in Normal and Dysgenetic Human Gonads. *Human Reproduction* 19 (6): 1338–44. https://doi.org/10.1093/humrep/deh265.

Ramsawhook, A., Lewis, L., Coyle, B., and Ruzov, A. (2017). Medulloblastoma and Ependymoma Cells Display Increased Levels of 5-Carboxylcytosine and Elevated TET1 Expression. *Clinical Epigenetics* 9 (1). https://doi.org/10.1186/s13148-016-0306-2.

Redelsperger, F., Cornelis, G., Vernochet, C., Tennant, B. C., Catzeflis, F., Mulot, B., Heidmann, O., Heidmann, T., and Dupressoir, A. (2014). Capture of Syncytin-Mar1, a Fusogenic Endogenous Retroviral Envelope Gene Involved in Placentation in the Rodentia Squirrel-Related Clade. *Journal of Virology* 88 (14): 7915–28. https://doi.org/10.1128/jvi.00141-14.

Riggs, A. D. (1975). X Inactivation, Differentiation, and DNA Methylation. *Cytogenetic and Genome Research* 14 (1): 9–25. https://doi.org/10.1159/000130315.

Rijnsoever, M. Van, Elsaleh, H., Joseph, D., McCaul, K., and Iacopetta, B. (2003). CpG Island Methylator Phenotype Is an Independent Predictor of Survival Benefit from 5-Fluorouracil in Stage III Colorectal Cancer. *Clinical Cancer Research* 9 (8): 2898–2903. https://doi.org/10.1016/s0360-3016(02)03228-5.

Rosenberg, C., Gurp, R. J. H. L. M. Van, Geelen, E., Oosterhuis, J. W., and Looijenga, L. H. J. (2000). Overrepresentation of the Short Arm of Chromosome 12 Is Related to Invasive Growth of Human Testicular Seminomas and Nonseminomas. *Oncogene* 19 (51): 5858–62. https://doi.org/10.1038/sj.onc.1203950.

Ruebner, M., Strissel, P. L., Ekici, A. B., Stiegler, E., Dammer, U., Goecke, T. W., Faschingbauer, F., Fahlbusch, F. B., Beckmann, M. W., and Strick, R. (2013). Reduced Syncytin-1 Expression Levels in Placental Syndromes Correlates with Epigenetic Hypermethylation of the ERVW-1 Promoter Region. *PLoS ONE* 8 (2). https://doi.org/10.1371/journal.pone.0056145.

SanMiguel, J. M., Abramowitz, L. K., and Bartolomei, M. S. (2018). Imprinted Gene Dysregulation in a Tet1 Null Mouse Model Is Stochastic and Variable in the Germline and Offspring. *Development (Cambridge)* 145 (7). https://doi.org/10.1242/dev.160622.

Santoni, F. A., Guerra, J., and Luban, J. (2012). HERV-H RNA Is Abundant in Human Embryonic Stem Cells and a Precise Marker for Pluripotency. *Retrovirology* 9 (111). https://doi.org/10.1186/1742-4690-9-111.

Sardina, J. L., Collombet, S., Tian, T. V., Gómez, A., Stefano, B. Di, Berenguer, C., Brumbaugh, J., et al. (2018). Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. *Cell Stem Cell* 23 (5): 727-741.e9. https://doi.org/10.1016/j.stem.2018.08.016.

Schomacher, L., and Niehrs, C. (2017). DNA Repair and Erasure of 5-Methylcytosine in Vertebrates. *BioEssays* 39 (3). https://doi.org/10.1002/bies.201600218.

Schön, U., Seifarth, W., Baust, C., Hohenadl, C., Erfle, V., and Leib-Mösch, C. (2001). Cell Type-Specific Expression and Promoter Activity of Human Endogenous Retroviral Long Terminal Repeats. *Virology* 279 (1): 280–91. https://doi.org/10.1006/viro.2000.0712.

Schüler, P., and Miller, A. K. (2012). Sequencing the Sixth Base (5-Hydroxymethylcytosine): Selective DNA Oxidation Enables Base-Pair Resolution. *Angewandte Chemie - International Edition* 51

(43): 10704–7. https://doi.org/10.1002/anie.201204768.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin Organization Marks Exon-Intron Structure. *Nature Structural and Molecular Biology* 16 (9): 990–95. https://doi.org/10.1038/nsmb.1659.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The Dynamics of Genome-Wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Molecular Cell* 48 (6): 849–62. https://doi.org/10.1016/j.molcel.2012.11.001.

Shen, H., Shih, J., Hollern, D. P., Wang, L., Bowlby, R., Tickoo, S. K., Thorsson, V., et al. (2018). Integrated Molecular Characterization of Testicular Germ Cell Tumors. *Cell Reports* 23 (11): 3392–3406. https://doi.org/10.1016/j.celrep.2018.05.039.

Shi, Q. J., Lei, Z. M., Rao, C. V., and Lin, J. (1993). Novel Role of Human Chorionic Gonadotropin in Differentiation of Human Cytotrophoblasts. *Endocrinology* 132 (3): 1387–95. https://doi.org/10.1210/endo.132.3.7679981.

Skakkebék, N. E., Meyts, E. R.-D., Jørgensen, N., Carlsen, E., Petersen, P. M., Giwercman, A., Andersen, A.-G., Jensen, T. K., Andersson, A.-M., and Müller, J. (1998). Germ Cell Cancer and Disorders of Spermatogenesis: An Environmental Connection? *APMIS* 106 (1–6): 3–12. https://doi.org/10.1111/j.1699-0463.1998.tb01314.x.

Smallwood, A., Papageorghiou, A., Nicolaides, K., Alley, M. K. R., Jim, A., Nargund, G., Ojha, K., Campbell, S., and Banerjee, S. (2003). Temporal Regulation of the Expression of Syncytin (HERV-W), Maternally Imprinted PEG10, and SGCE in Human Placenta1. *Biology of Reproduction* 69 (1): 286–93. https://doi.org/10.1095/biolreprod.102.013078.

Smiraglia, D. J., Szymanska, J., Kraggerud, S. M., Lothe, R. A., Peltomäki, P., and Plass, C. (2002). Distinct Epigenetic Phenotypes in Seminomatous and Nonseminomatous Testicular Germ Cell Tumors. *Oncogene* 21 (24): 3909–16. https://doi.org/10.1038/sj.onc.1205488.

Sobin, L. H., and Compton, C. C. (2010). TNM Seventh Edition: What's New, What's Changed. *Cancer* 116 (22): 5336–39. https://doi.org/10.1002/cncr.25537.

Søe, K., Andersen, T. L., Hobolt-Pedersen, A. S., Bjerregaard Bolette, B., Larsson, L. I., and Delaissé, J. M. (2011). Involvement of Human Endogenous Retroviral Syncytin-1 in Human Osteoclast Fusion. *Bone* 48 (4): 837–46. https://doi.org/10.1016/j.bone.2010.11.011.

Spans, L., Broeck, T. Van den, Smeets, E., Prekovic, S., Thienpont, B., Lambrechts, D., Jeffrey Karnes, R., et al. (2016). Genomic and Epigenomic Analysis of High-Risk Prostate Cancer Reveals Changes in Hydroxymethylation and TET1. *Oncotarget* 7 (17): 24326–38. https://doi.org/10.18632/oncotarget.8220.

Strick, R., Ackermann, S., Langbein, M., Swiatek, J., Schubert, S. W., Hashemolhosseini, S., Koscheck, T., et al. (2007). Proliferation and Cell-Cell Fusion of Endometrial Carcinoma Are Induced by the Human Endogenous Retroviral Syncytin-1 and Regulated by TGF-β. *Journal of Molecular Medicine* 85 (1): 23–38. https://doi.org/10.1007/s00109-006-0104-y.

Strissel, P. L., Ruebner, M., Thiel, F., Wachter, D., Ekici, A. B., Wolf, F., Thieme, F., Ruprecht, K., Beckmann, M. W., and Strick, R. (2012). Reactivation of Codogenic Endogenous Retroviral (ERV) Envelope Genes in Human Endometrial Carcinoma and Prestages: Emergence of New

Molecular Targets. *Oncotarget* 3 (10): 1204–19. https://doi.org/10.18632/oncotarget.679.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction. *Journal of Biological Chemistry* 279 (26): 27816–23. https://doi.org/10.1074/jbc.M400181200.

Sugimoto, J., Sugimoto, M., Bernstein, H., Jinno, Y., and Schust, D. (2013). A Novel Human Endogenous Retroviral Protein Inhibits Cell-Cell Fusion. *Scientific Reports* 3 (1462). https://doi.org/10.1038/srep01462.

Sun, Y., Zhu, H., Song, J., Jiang, Y., Ouyang, H., Huang, R., Zhang, G., et al. (2016). Upregulation of Leukocytic Syncytin-1 in Acute Myeloid Leukemia Patients. *Medical Science Monitor* 22 (July): 2392–2403. https://doi.org/10.12659/MSM.899303.

Surani, M. A., Hayashi, K., and Hajkova, P. (2007). Genetic and Epigenetic Regulators of Pluripotency. *Cell* 128 (4): 747–62. https://doi.org/10.1016/j.cell.2007.02.010.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., et al. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* 324 (5929): 930–35. https://doi.org/10.1126/science.1170116.

Tailor, C. S., Nouri, A., Zhao, Y., Takeuchi, Y., and Kabat, D. (1999). A Sodium-Dependent Neutral-Amino-Acid Transporter Mediates Infections of Feline and Baboon Endogenous Retroviruses and Simian Type D Retroviruses. *Journal of Virology* 73 (5): 4470–74. http://www.ncbi.nlm.nih.gov/pubmed/10196349.

Tang, C., Tang, L., Wu, X., Xiong, W., Ruan, H., Hussain, M., Wu, J., Zou, C., and Wu, X. (2016). Glioma-Associated Oncogene 2 Is Essential for Trophoblastic Fusion by Forming a Transcriptional Complex with Glial Cell Missing-A. *Journal of Biological Chemistry* 291 (11): 5611–22. https://doi.org/10.1074/jbc.M115.700336.

Tang, W. W. C., Dietmann, S., Irie, N., Leitch, H. G., Floros, V. I., Bradshaw, C. R., Hackett, J. A., Chinnery, P. F., and Surani, M. A. (2015). A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell* 161 (6): 1453–67. https://doi.org/10.1016/j.cell.2015.04.053.

Toyota, M., and Issa, J. P. J. (1999). CpG Island Methylator Phenotgpes in Aging and Cancer. *Seminars in Cancer Biology* 9 (5): 349–57. https://doi.org/10.1006/scbi.1999.0135.

Trejbalová, K., Blažková, J., Matoušková, M., Kučerová, D., Pecnová, L., Vernerová, Z., Heráček, J., Hirsch, I., and Hejnar, J. (2011). Epigenetic Regulation of Transcription and Splicing of Syncytins, Fusogenic Glycoproteins of Retroviral Origin. *Nucleic Acids Research* 39 (20): 8728–39. https://doi.org/10.1093/nar/gkr562.

Trejbalová, K., Kovářová, D., Blažková, J., Machala, L., Jilich, D., Weber, J., Kučerová, D., Vencálek, O., Hirsch, I., and Hejnar, J. (2016). Development of 5‘ LTR DNA Methylation of Latent HIV-1 Provirus in Cell Line Models and in Long-Term-Infected Individuals. *Clinical Epigenetics* 8 (1): 1–20. https://doi.org/10.1186/s13148-016-0185-6.

Tsukada, Y. I., Akiyama, T., and Nakayama, K. I. (2015). Maternal TET3 Is Dispensable for Embryonic Development but Is Required for Neonatal Growth. *Scientific Reports* 5 (15876). https://doi.org/10.1038/srep15876.

Vargas, A., Moreau, J., Landry, S., LeBellego, F., Toufaily, C., Rassart, É., Lafond, J., and Barbeau, B. (2009). Syncytin-2 Plays an Important Role in the Fusion of Human Trophoblast Cells. *Journal of Molecular Biology* 392 (2): 301–18. https://doi.org/10.1016/j.jmb.2009.07.025.

Vargiu, L., Rodriguez-Tomé, P., Sperber, G. O., Cadeddu, M., Grandi, N., Blikstad, V., Tramontano, E., and Blomberg, J. (2016). Classification and Characterization of Human Endogenous Retroviruses Mosaic Forms Are Common. *Retrovirology* 13 (1). https://doi.org/10.1186/s12977-015-0232-y.

Verma, N., Pan, H., Doré, L. C., Shukla, A., Li, Q. V., Pelham-Webb, B., Teijeiro, V., et al. (2018). TET Proteins Safeguard Bivalent Promoters from de Novo Methylation in Human Embryonic Stem Cells. *Nature Genetics* 50 (1): 83–95. https://doi.org/10.1038/s41588-017-0002-y.

Vincent, J. J., Huang, Y., Chen, P. Y., Feng, S., Calvopiña, J. H., Nee, K., Lee, S. A., et al. (2013). Stage-Specific Roles for Tet1 and Tet2 in DNA Demethylation in Primordial Germ Cells. *Cell Stem Cell* 12 (4): 470–78. https://doi.org/10.1016/j.stem.2013.01.016.

Voisset, C., Blancher, A., Perron, H., Mandrand, B., Mallet, F., and Paranhos-Baccalà, G. (1999). Phylogeny of a Novel Family of Human Endogenous Retrovirus Sequences, HERV-W, in Humans and Other Primates. *AIDS Research and Human Retroviruses* 15 (17): 1529–33. https://doi.org/10.1089/088922299309810.

Wang-Johanning, F., Liu, J., Rycaj, K., Huang, M., Tsai, K., Rosen, D. G., Chen, D. T., Lu, D. W., Barnhart, K. F., and Johanning, G. L. (2007). Expression of Multiple Human Endogenous Retrovirus Surface Envelope Proteins in Ovarian Cancer. *International Journal of Cancer* 120 (1): 81–90. https://doi.org/10.1002/ijc.22256.

Wang-Johanning, F., Radvanyi, L., Rycaj, K., Plummer, J. B., Yan, P., Sastry, K. J., Piyathilake, C. J., Hunt, K. K., and Johanning, G. L. (2008). Human Endogenous Retrovirus K Triggers an Antigen-Specific Immune Response in Breast Cancer Patients. *Cancer Research* 68 (14): 5869–77. https://doi.org/10.1158/0008-5472.CAN-07-6838.

Wang, P., Yan, Y., Yu, W., and Zhang, H. (2019). Role of Ten-eleven Translocation Proteins and 5-hydroxymethylcytosine in Hepatocellular Carcinoma. *Cell Proliferation* 52 (4). https://doi.org/10.1111/cpr.12626.

Wang, R. Y. H., Gehrke, C. W., and Ehrlich, M. (1980). Comparison of Bisulfite Modification of 5-Methyldeoxycytidine and Deoxycytidine Residues. *Nucleic Acids Research* 8 (20): 4777–90. https://doi.org/10.1093/nar/8.20.4777.

Wegman, S. J., Parwani, A. V., and Zynger, D. L. (2019). Cytokeratin 7, Inhibin, and P63 in Testicular Germ Cell Tumor: Superior Markers of Choriocarcinoma Compared to β-Human Chorionic Gonadotropin. *Human Pathology* 84 (February): 254–61. https://doi.org/10.1016/j.humpath.2018.10.007.

Wermann, H., Stoop, H., Gillis, A. J. M., Honecker, F., Gurp, R. J. H. L. M. Van, Ammerpohl, O., Richter, J., Oosterhuis, J. W., Bokemeyer, C., and Looijenga, L. H. J. (2010). Global DNA Methylation in Fetal Human Germ Cells and Germ Cell Tumours: Association with Differentiation and Cisplatin Resistance. *Journal of Pathology* 221 (4): 433–42. https://doi.org/10.1002/path.2725.

Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E., and Zhang, Y. (2011).

Genome-Wide Analysis of 5-Hydroxymethylcytosine Distribution Reveals Its Dual Function in Transcriptional Regulation in Mouse Embryonic Stem Cells. *Genes and Development* 25 (7): 679–84. https://doi.org/10.1101/gad.2036011.

Wu, H., D'Alessio, A. C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Eve Sun, Y., and Zhang, Y. (2011). Dual Functions of Tet1 in Transcriptional Regulation in Mouse Embryonic Stem Cells. *Nature* 473 (7347): 389–94. https://doi.org/10.1038/nature09934.

Wyatt, G. R., and Cohen, S. S. (1953). The Bases of the Nucleic Acids of Some Bacterial and Animal Viruses: The Occurrence of 5-Hydroxymethylcytosine. *The Biochemical Journal* 55 (5): 774–82. https://doi.org/10.1042/bj0550774.

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S. H., Ito, S., et al. (2011). Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α-Ketoglutarate-Dependent Dioxygenases. *Cancer Cell* 19 (1): 17–30. https://doi.org/10.1016/j.ccr.2010.12.014.

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J. G., Bian, C., Hu, Y., et al. (2012). Tet3 CXXC Domain and Dioxygenase Activity Cooperatively Regulate Key Genes for Xenopus Eye and Neural Development. *Cell* 151 (6): 1200–1213. https://doi.org/10.1016/j.cell.2012.11.014.

Yagi, K., Akagi, K., Hayashi, H., Nagae, G., Tsuji, S., Isagawa, T., Midorikawa, Y., et al. (2010). Three DNA Methylation Epigenotypes in Human Colorectal Cancer. *Clinical Cancer Research* 16 (1): 21–33. https://doi.org/10.1158/1078-0432.CCR-09-2006.

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 Controls Meiosis by Regulating Meiotic Gene Expression. *Nature* 492 (7429): 443–47. https://doi.org/10.1038/nature11709.

Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., and Zhang, Y. (2013). Role of Tet1 in Erasure of Genomic Imprinting. *Nature* 504 (7480): 460–64. https://doi.org/10.1038/nature12805.

Yan, H., Parsons, D. W., Jin, G., McLendon, R., Rasheed, B. A., Yuan, W., Kos, I., et al. (2009). Mutations in Gliomas. *New England Journal of Medicine* 360 (8): 765–73. https://doi.org/10.1056/NEJMoa0808710.

Yang, L., Yu, S. J., Hong, Q., Yang, Y., and Shao, Z. M. (2015). Reduced Expression of TET1, TET2, TET3 and TDG MRNAs Are Associated with Poor Prognosis of Patients with Early Breast Cancer. *PLoS ONE* 10 (7). https://doi.org/10.1371/journal.pone.0133896.

Yoder, J. A., and Bestor, T. H. (1998). A Candidate Mammalian DNA Methyltransferase Related to Pmt1p of Fission Yeast. *Human Molecular Genetics* 7 (2): 279–84. https://doi.org/10.1093/hmg/7.2.279.

Yoshioka, M., Matsutani, T., Hara, A., Hirono, S., Hiwasa, T., Takiguchi, M., and Iwadate, Y. (2018). Real-Time Methylation-Specific PCR for the Evaluation of Methylation Status of MGMT Gene in Glioblastoma. *Oncotarget* 9 (45): 27728–35. https://doi.org/10.18632/oncotarget.25543.

Yu, C., Shen, K., Lin, M., Chen, P., Lin, C., Chang, G. D., and Chen, H. (2002). GCMa Regulates the Syncytin-Mediated Trophoblastic Fusion. *Journal of Biological Chemistry* 277 (51): 50062–68. https://doi.org/10.1074/jbc.M209316200.

Yu, M., Hon, G. C., Szulwach, K. E., Song, C. X., Jin, P., Ren, B., and He, C. (2012). Tet-Assisted Bisulfite Sequencing of 5-Hydroxymethylcytosine. *Nature Protocols* 7 (12): 2159–70.

https://doi.org/10.1038/nprot.2012.137.

Yue, X., Lio, C. W. J., Samaniego-Castruita, D., Li, X., and Rao, A. (2019). Loss of TET2 and TET3 in Regulatory T Cells Unleashes Effector Function. *Nature Communications* 10 (1). https://doi.org/10.1038/s41467-019-09541-y.

Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y. G. (2010). TET1 Is a DNA-Binding Protein That Modulates DNA Methylation and Gene Transcription via Hydroxylation of 5-Methylcytosine. *Cell Research* 20 (12): 1390–93. https://doi.org/10.1038/cr.2010.156.

Zynger, D. L., Dimov, N. D., Luan, C., Tean Teh, B., and Yang, X. J. (2006). Glypican 3: A Novel Marker in Testicular Germ Cell Tumors. *American Journal of Surgical Pathology* 30 (12): 1570–75. https://doi.org/10.1097/01.pas.0000213322.89670.48.