2023 S.T. Yau High School Science Award

Research Report

The Team

Name of team member: Lei, Kevin Xuan
School: Shanghai High School International Division
City, Country: Shanghai, P.R.China

Name of supervising teacher: Liao, Hui
Job Title: Biology Teacher
School/Institution: Shanghai High School International Division
City, Country: Shanghai, P.R.China

Name of supervising teacher: Zhang, Xin
Job Title: Faculty
School/Institution: Shanghai Jiao Tong University
City, Country: Shanghai, P.R.China

Title of Research Report

A double-edged sword: a shared mechanism for high stemness maintenance in both human germ cells and cancer cells

Date

[2023/09/12]

A double-edged sword: a shared mechanism for high stemness maintenance in both human germ cells and cancer cells

Lei, Kevin Xuan

Abstract

The abnormally high stemness in cancer cells is one of the major factors that make cancer cells resistant to treatment. Despite the importance of cancer stem cells in tumor progression, the underlying mechanism of their stemness maintenance remains poorly understood. For more than three decades, a group of primate specific cancer-testis-antigens (CTAs) have been known to be prevalent in both human testes and cancer cells. Given that primates have a much enhanced spermatogonia stem cell (SSC) system than lower mammalian species, we hypothesize these primate specific CTAs act as a double-edged sword: in human testes, these genes play essential roles in protecting germline cells from environmental stress to guarantee progeny persistence, whereas in cancer cells, these genes' ability may have been hijacked to maintain the viability of cancer cells with corrupted genome, providing high stemness that makes them difficult to kill.

Here, we have carried out a bioinformatics screening using the Ensemble database API to obtain a panel of prospective human CTA genes that only evolved after the Common Marmoset, the most primitive primate in which the advanced SSC progenitor buffer system first appeared. Experimentally, we knocked down these CTAs in both cancer cells and SSCs and found that several novel CTA genes are indispensable for the spheroid formation—a golden standard for high stemness—in both cancer cells and SSCs. Intriguingly, patients with higher level expression of these CTAs had significantly worse prognosis and shorter survival time. In summary, this study unveils a fundamental mechanistic connection between human physiology and pathology that CTAs are crucial to sustaining high stemness in both human SSCs and cancer cells, providing attractive targets for novel cancer diagnosis and therapy.

Keywords: cancer stem cell, spermatogonia stem cell, stemness, cancer-testis-antigen, evolution, primate

CONTENTS

I. INTRODUCTION

Humans have been combating cancer for centuries. Despite great advances in cancer research and clinical practices, malignant tumors are still among the most challenging diseases for the human being. People have been trying to reveal the underlying mechanism of cancer, aiming for a more profound understanding of its onset and progression, as well as the development of effective strategies for cancer treatments. Over the past four decades, it has become evident that mutations in oncogenes and tumor suppressor genes are prevalent in malignant tumors with complex patterns (*1, 2*). Attempts to uncover a fundamental principle that govern the cancer biology has been all in vain, mostly because cancer cells exhibit disordered genomes and epigenomes, making it impossible for one or a few mutation spectra to fully reflect the situation of all cancers (*3, 4*).

Notwithstanding these difficulties, persistent efforts are aimed at identifying a fundamental principle for cancers, and the concept of cancer stem cell (CSC) represents such an attempt (*5-7*). It is well known that tissues within the human body are often supplied and supported by a small group of long-living adult stem cells, which bear the ability to self-renew to maintain a constant supply of short-living cells that populate tissues to perform specialized functions (*8-10*). The parallelism between adult stem cells and tumorigenesis has given rise to the cancer stem cell (CSC) concept nearly three decades ago (*11, 12*). The CSC theory states that like how adult stem cells maintain a tissue of differentiated cells, CSCs fuel the tumor growth by maintaining a group of high-stemness tumor stem cells hidden in cancers (*5, 6*). This explains why cancer patients are prone to almost inevitable recurrence of tumors even after extensive mainstream treatments, making CSC a hot research topic and a compelling target for cancer therapeutics. However, it has gradually become clear that CSCs with different tissue origins are highly heterogeneous with

distinct cellular properties and are maintained in dedicated, distinct microenvironments, so that the underlying mechanism of CSCs has not been as obvious as was initially hoped (*13, 14*). Amidst this complexity, though, there exists a group of genes, cancer testis antigens (CTAs), whose expression normally highly restricted in testicular germ cells, was adversely turned on in malignant tumor cells (*15, 16*). Although CTA genes have been used as biomarkers for CSCs and considered as a group of promising immunotherapy targets, their fundamental functions in CSCs are yet to be understood (*17, 18*).

Spermatogonia stem cells (SSCs) are vital for male fertility as they not only self-renew to maintain a long-term stem cell pool, but also are capable of differentiating to generate a large body of sperms (*19, 20*). Of note, many CTAs are primate-specific, fast-evolving genes that are predominantly expressed in SSCs, suggestive of their fundamental importance in primate spermatogenesis (*21*). Higher primate species starting from New World Monkeys (NWMs) have evolved a unique progenitor-buffered SSC system, whereas other mammals employ a direct SSC system (*22, 23*). The direct SSC system contains a small number of SSCs that go through multiple cycles of mitosis to generate enough differentiated spermatogonia for the production of millions of sperms. In sharp contrast, the progenitor-buffered system in primates maintains a large pool of stem cells. Introducing the progenitor-buffered SSC system in primates protects the male germline from undergoing many mitotic divisions, decreasing the risk for mutations and ultimately contributing to the extended lifespans of primates (*23*). The underlying mechanism of the progenitor-buffered SSC system in higher primates and whether fast-evolving CTA genes play a vital role in this system is yet to be unveiled.

The parallelism in high-stemness between SSCs and CSCs leads us to hypothesize that CTAs might play an important role in maintaining a stable high-stemness state not only in

testicular SSCs but also in CSCs of malignant tumors, making CSCs extremely deadly and difficult to eradicate. Thus, CTAs' influence on cancers and the male reproductive system is like a "double-edged sword" in that primate specific CTAs are essential in protecting germline cells from environmental stress to guarantee progeny persistence in human testes, whereas ironically in cancer cells, these genes' ability may have been hijacked to maintain the viability of CSCs with a corrupted genome, providing high stemness that makes them difficult to kill.

II. MATERIALS AND METHODS

A. Materials

shRNA oligos, Sangon Biotech, Shanghai, China; T4 DNA ligase, Thermo Fisher; DH5α competent cell, Sangon Biotech, Shanghai; LB broth, Sigma Aldrich; Plasmid extraction kit, Axygen; BamHI, EcoRI restriction enzymes, Thermo Fisher; Transfection reagents, Lipo 3000 from Invitrigen, Thermo Fisher; StemPro™-34 SFM, RPMI1640, MEM, DMEM, DMEM/F12 cell culture media, 0.25% Trypsin, Insulin-Transferrin-Sodium Selenite, Fetal bovine serum, MEM Vitamin Solution, Non-Essential Amino Acids, GlutaMAX™, Pyruvic acid, 2-Mercaptoethanol, Anti-Anti, from Gibco; Human FGF2, Human EGF, from Peprotech; Ascorbic Acid, DL-Lactic Acid, d-Biotin, β-Estradiol, Progesterone, from Sigma; human Activin A, human GDNF, human LIF, from R&D Systems; Puromycin, Gibco; cell culture dish and plates, pipettes, from Biofil, China; Matrigel, from Corning; A549, KYSE30, A375, NCIH4006, PLC/PRF/5, CACO-2 cancer cell lines, from Chinese ATCC, Shanghai. All sequencing service are provided by Sangon Biotech, Shanghai, China.

B. Equipment

PCR machine, Eppendorf; Microwave, Media; Cell culture biosafety cabin, ESCO from Thermo Fisher; Automated Cell Counters by Thermo Fisher. FACS sorting machine, Beckman Coulter; Microscope, Olympus IX83, ZEISS Axio Zoom V16; Cell culture incubator, Thermo Fisher; Incucyte® s3 live-cell analysis system, Sartorius; Livecyte® Kinetic Cytometer, phasefocus.

C. Methods

Phylogenetic dating analysis

To perform the phylogenetic data analysis, 16 species were selected from chicken to chimpanzee and grouped into 10 evolutionary clusters from Aves to Apes (Table 1). First, individual dataset annotated by human gene names (GRCh38.p14) with homology information of a specific species was automatically downloaded from the BioMart database (*24*) on the Ensembl website using a self-developed script based on the pybiomart library (http://asia.ensembl.org/biomart/martview/f432ee8c9741d261df31172b35506d71). All datasets were combined into a large Pandas data frame with only the coding genes and their corresponding chromosome annotations. Second, the build-in orthologue identification method on the website provided a binary index called Orthology Confidence (OC) to score the validity of identified orthologs (low confidence: 0; high confidence:1). The earliest appearance of a human coding gene was defined in an evolutionary cluster when its ortholog was found for the first time with the OC value equal to one. The whole procedure was conducted by cluster one after another following the evolutionary timeline, so that a gene was immediately removed once its earliest appearance was defined. Lastly, the species tree was built using TimeTree5 (*25*) and the bar and

line plots were graphed using R. All figures were further edited for easy visualization in Adobe Illustrator.

Table 1. 16 species in 10 evolutionary groups with the corresponding Latin names in Ensembl.

Evolution	Species	Latin name
Aves	Chicken	*Gallus gallus*
Monotremes	Platypus	*Ornithorhynchus anatinus*
Marsupials	Gray Short-tailed Opossum	*Monodelphis domestica*
Laurasiatheria	Pig	*Sus scrofa*
	Cattle	*Bos taurus*
	Domestic cat	*Felis catus*
Glires	Mouse	*Mus musculus*
	Rat	*Rattus norvegicus*
Prosimians	Mouse lemur	*Microcebus murinus*
New-World Monkeys	White-tufted-ear	*Callithrix jacchus*
	Ma's night monkey	*Aotus nancymaae*
Old-World Monkeys	Crab-eating macaque	*Macaca fascicularis*
	Rhesus macaque	*Macaca mulatta*
Apes	Gibbon	*Nomascus leucogenys*
	Gorilla	*Gorilla gorilla*
	Chimpanzee	*Pan troglodytes*
Humans	Human	*Homo sapiens*

Bulk RNAseq organ specificity analysis of human coding genes and testicular scRNAseq analysis of testis-specific human coding genes emerged in NWMs

The tissue specificity analysis was carried out by investigating the online dataset of bulk RNA-seq across 27 different tissues from 95 human individuals (*26*). The original FASTQ files were downloaded from the website (https://www.ebi.ac.uk/gxa/experiments/E-MTAB-1733/Results)

and preprocessed using fastp (v0.23.1) (*27*) with the parameter settings "-n 10 -l 75 -g". The sequenced reads were mapped to human genome GRCh38 using STAR (v2.7.7a) (*28*). The expected gene-level counts were quantified using RSEM (v1.3.1) (*29*). The single cell RNA-seq data of human testis were used to study expression pattern of testis specific genes across the testicular germ cells (*30*). The normalized expression matrix with the corresponding annotation files including cell type information were downloaded from GEO database (GSE106487: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106487). After loading all the data as one Seurat object, nonlinear dimensionality reduction was performed using t-SNE algorithm in Seurat (v3.2.2) (*31*) for the low dimensional visualization of cell populations. All heatmaps were plotted using the R package pheatmap.

Molecular cloning and shRNA construction

The shRNAs against the selected CTA genes were obtained via public resources (https://portals.broadinstitute.org/gpp/public/seq/search) (Table 2). Three shRNAs were designed for each gene to exclude the off-target effects. The detailed sequences are shown in Table 1. These oligos were synthesized by Sangon Biotech and ligated into the pGreenPuro Lentivector. Briefly, sense and antisense DNA oligonucleotides dissolved in sterile ddH_2O were annealed in 16 μL of reaction system containing 8 μL (100 μM) of sense (forward) oligonucleotide, 8 μL (100 μM) of antisense (reverse) oligonucleotide. The reaction mixture was heated at 95°C for 5 min before cooling down to room temperature. The backbone vectors were digested with EcoRI and BamHI at 37°C in water bath for 2 hours. Then the vectors were extracted by respective kit. After ligation with shRNA fragments at 16°C overnight, the mixture was transformed into

Escherichia coli DH5a competent cells. The positive clones were selected and then validated by PCR amplification and DNA sequencing.

Forward primer	5 µl
Reverse primer	5 µl
10 × NEB 2.1 buffer	2 µl
ddH_2O	up to 20 µl

Place the mixture in the beaker, heated in boiling water for 5 min, then cooled down to room temperature.

Ligation mixture:

Vector	50 ng
DNA fragments	1 µl
10 × buffer	1 µl
T4 DNA ligase	0.6 µl
ddH_2O	up to 10 µl

Place the mixture in PCR machine at 16°C overnight.

Table 2. shRNA sequences for 12 CTA genes.

CT45A10	shRNA-1	GTCCAAAGCAAAGAAGCTTAT
	shRNA-2	ACGAGAAATTAATGCTGATAT
	shRNA-3	ACTTGTCCCTGGAGGATTATC
CTAG2	shRNA-1	GTCCGGCAACCTACTGTTTAT
	shRNA-2	TGTCGCCTTTAATGTGATGTT
	shRNA-3	GCAGAAAGCTAGAGATCTCAG
FAM9A	shRNA-1	GAAGCCGCTACTTGAGCAATT
	shRNA-2	CCTCTGTCACTTGTCTGTTAA
	shRNA-3	GAACACAATTGAACGTGCTTT
GAGE1	shRNA-1	CCTGAAGAAGAGATGAGGTCT
	shRNA-2	CTATGTACAGCCTCCTGAAAT
GAGE2A	shRNA-1	CTACGTAGAGCCTCCTGAAAT
	shRNA-2	TACGTAGAGCCTCCTGAAATG
	shRNA-3	TAGAGCCTCCTGAAATGATTG
PAGE4	shRNA-1	CCTCCGATCGAAGAACGTAAA
	shRNA-2	AGAAGGTGATTGCCAGGAAAT

	shRNA-3	CACCTAATCCTAAGCATGCTA
RHOX2	shRNA-1	CTTTGAGCCAGATTCATATTT
SAGE1	shRNA-1	TGGTCCCACAGGGCTTATTAA
	shRNA-2	CCAAGGAGAAACAAGGACATA
	shRNA-3	CCCAAACTGATAAGGTCATAT
TFDP3	shRNA-1	CGAAGAACTCAAGGTCTTAAT
	shRNA-2	ACTGACGTCCTCTCGCCTTAA
	shRNA-3	GCAGCATCTCCGACGACAAAT
VCX3a	shRNA-1	GACTACCAAGGTGGCCAAGAA
ACTL8	shRNA-1	AGGGTTCCAATAGAAACTTTA
	shRNA-2	TGCAGTACCTCTGGTCATTTG
	shRNA-3	CTATCTCCTCAAGAGTCTCTT
PRAME	shRNA-1	GCTGGACTCTATTGAAGATTT
	shRNA-2	ATGTTGACTTGAGGAGTTAAT
	shRNA-3	GCACATATCAAATGCTTCATT

Quantitative reverse-transcription polymerase chain reaction (qRT-PCR)

RNA extraction was performed using RNeasy Mini (Qiagen, USA) following the manufacturer's instructions. RNAs prepared by RNeasy Mini kit were treated with DNase (Qiagen, USA). Reverse transcription was performed using Hiscript III Reverse Transcriptase (Vazyme). Quantitative PCR was performed using AceQ qPCR SYBR Green Master Mix (Vazyme) on a QuantStudio™ Real-Time PCR instrument. PCR primers were designed using NCBI Primer-BLAST tool (Table 3).

Table 3. qRT-PCR primers (5'-3').

H-CTAG2-F	TGTCAGTTCGGGACCAGGA
H-CTAG2-R	CTGAGACCTTGTGTTTGGGTG
H-CT45A10-F	GATGGCCCTGTTGGCAAGGA
H-CT45A10-R	CGTTTCCTCCCACAGGTGCA
H-FAM9A-F	TGCCAAAGCTCAGTTGGAAG
H-FAM9A-R	TCCTGGGAAGTTAGAGGCGAT
H-GAGE1-F	CCTAGACCAAGGCGCTATGT
H-GAGE1-R	TCCCCTTCTTCAGGTGTTGC

H-GAGE2A-F	AGGGGAACCAGCAACTCAAC
H-GAGE2A-R	TCAGGCGTTTTCACCTCCTC
H-PAGE4-F	GCTCCCGGTGAATCTCAGC
H-PAGE4-R	GATCGGAGGTGTTCCTTCTCT
H-RHOXF2-F	CCCTTGTTCATCAGCGGGAT
H-RHOXF2-R	CAAGGACGGAGGAGGAAAGG
H-SAGE1-F	ACTTCAAACGAGTCAACCAACT
H-SAGE1-R	TCTAACCACGAGGACATACTCTT
H-TFDP3-F	CCTGTGGTAGGAAGCCCAAA
H-TFDP3-R	TCATGGAAAGACGGCACAGG
H-VCX3A-F	AAGAAAGGGGCTGCGACAAA
H-VCX3A-R	GACTCAGGGGGTCGTGCTG
H-ACTL8-F	GAACATCGTGAACTACCTACCG
H-ACTL8-R	CAAGGGTGTCTCCGTGATGAT
H-PRAME-F	AGCCTTTGACGGGAGACAC
H-PRAME-R	GAGTTCTTCCGTAAATCCAGCA

Human SSC cell culture

The 10-cm dish was coated with Laminin at a concentration of 1μg/cm^2 at 37°C overnight. The spermatogonia stem cell (SSC) medium contained supplemented Stem-Pro-34 with nonessential amino acids, 2 mM GlutaMAX, 1 × Anti-Anti, 1 mM sodium pyruvate, 1 × insulin, Transferrin, selenium solution (ITS-G), 0.1 mM 2-mercaptoethanol, 10μl/mL minimal essential medium (MEM) vitamin solution (all from Invitrogen), 6 mg/mL D-(+)-glucose, 30 mg/mL pyruvic acid, 10 μL/mL DL-lactic acid (all from Sigma-Aldrich), 10-4 M ascorbic acid, 10 mg/mL D-biotin, 30 ng/mL β-estradiol, 60 ng/mL progesterone, (20 ng/mL) recombinant human EGF, (10 ng/mL) Human FGF2, 10 ng/mL recombinant human GDNF (all from Sigma-Aldrich), 10 ng/mL recombinant human LIF (Chemicon International Inc., USA), 10 ng/mL recombinant human activin A protein (R&D Systems) and 1% fetal bovine serum (Gibco). The SSCs were cultured in SSC medium on Laminin-coated 10-cm dish at a cell density of 50,000 cells/cm^2 at 37°C with 7%

CO_2 and 5% O_2. SSCs were passaged by short-term trypsination every 7 days in one or several laminin-coated new dishes.

Lentivirus preparation and stable cancer cell lines establishment

Lentivirus was prepared to stably knockdown specific CTA genes as previously described (*32*). Briefly, the backbone plasmid carrying the respective shRNA sequence (16 μg) was mixed with packaging plasmids pCMV-dR8.9 (12 μg) or p-VSVG (8 μg) together with Lipofectamine 3000 (Life Technologies). The detailed sequences are shown in Table 1. The mixture was added to 70% confluent 293T cells in 15-cm dish. The lentivirus was harvested and enriched at 60 hours post transfection. Finally, virus pellet was resuspended with 1 mL medium. To establish stable cancer cell lines, puromycin (Gibco) at 0.5 mg mL^{-1} was used to select transfected cancer cells before purification by FACS (Beckman Coulter).

Lentivirus infection of human SSC

The lentiviral solution was rapidly thawed at room temperature. Virus solution (1 mL) was prepared with 9 mL SSC medium and 10 μg/mL polybrene for each virus. 1×10^6 SSCs were resuspended with the virus solution. The cells were incubated with the virus for 24-36 hours. The media was replaced with 10 mL of fresh SSC medium. The infection efficiency was validated on Day 3 by mCherry fluorescence.

Flow cytometry

The infected SSCs were harvested to prepare a single-cell suspension. SSCs were re-suspended at a concentration of 2-5 $\times$ 10^6/ml, and filtered to prevent clogging. A sorting speed of

approximately 18,000-20,000 events per second was maintained at the optimal pressure. Collection medium was supplemented with 10% serum. The transgenic SSCs expressing mCherry tag were collected in FACS tubes.

2D SSC cell proliferation assay

For 2D SSC proliferation assays, PerkinElmer 96-well plate was coated with laminin at a concentration of 1μg per cm^2 at 37°C overnight. 10,000/well SSCs were seeded in 200 μL SSC medium, cultured for 96 h, scanned and imaged every four hours in the Incucyte® S3 system (Sartorius) equipped with an IncuCyte Zoom 10 × Plan Phase objective (Sartorius). Cell index was calculated using phase-contrast images from 0 to 96 h after normalized to 0 h, by Incucyte Zoom software.

3D SSC colony formation assay

Matrigel matrix (Corning, 356321) was thawed at 4°C before use. SSCs were harvested and trypsinized to make a single-cell suspension, and then the cells were pelleted through centrifugation at 125 g for 5 min at room temperature. The cells were re-suspended with SSC medium to the final cell density of 5×10^4 cells/mL. 10 μL prepared cell suspension was added into each 1.5 mL EP tube, and then the cells were pelleted. 500 cells were gently re-suspended with 50 μL Matrigel matrix by using pre-chilled tips. 50 μL mixture of Matrigel matrix and SSCs were added into each well (24-well plate, PerkinElmer), incubated at 37°C for 10 min to form gel. 1 mL SSC medium was gently added to each well. Culture was kept for 14 days and the half medium was changed every 3 days. The colonies were fixed in 4% PFA and stained with Crystal Violet 0.5%. The image of Crystal Violet staining-3D cultures was collected by ZEISS Zoom

V16 microscope. The number of colonies was calculated and analyzed by the Image J software (NIH). All data were presented as means ± standard error of the mean (S.E.M).

Cancer spheroid formation assay

The spheroid formation assay was performed to evaluate the stemness of the cancer cells. Control and stable knockdown cells were digested with 0.25% trypsin (Gibco), neutralized with the complete medium and spun down with 900 rpm centrifuged for 5 minutes. The cells were resuspended with 1mL PBS and the cell number was counted using the Automated Cell Counters (Thermo Fisher). The cell density was then adjusted to 1×10^4/mL by serial dilution. 3000 cells were seeded per well in the Matrigel (Corning), with the droplet (20 μL) carefully placed in the center of the 48 well plates. Finally, respective complete media were added to the wells. The sphere grew in the incubator for 7-10 days.

Microscopy and imaging analyses

Pictures of the spheres were collected by the Olympus IX83 microscope. The size and number of the spheres were calculated and analyzed by the Image J software (NIH). Significance levels for comparisons between groups were determined with unpaired two-tailed Student's t-test using GraphPad Prism v.7 software or Excel (Microsoft Office). All data were presented as means ± S.E.M.

III. RESULTS

A. Identification of prospective "double-edged sword" CTA genes

To test the "double-edged sword" hypothesis, we first set out to search for prospective human testis-specific genes that could play important roles in maintaining high-stemness for both CSCs and SSCs. Because higher primate species starting from New World Monkeys (NWMs) have evolved a progenitor-buffered SSC system with a large, undifferentiated stem cell pool (*33*), we hypothesized that some novel, fast-evolving CTA genes might be the genetic underpinnings for the newly evolved system. Such genes would be present in all higher primates, but absent in rodents and other non-primate mammals. We performed a phylogenetic dating analysis of all human coding genes and grouped them into ten clusters based on their earliest emergence in 16 species ranging from birds to all major linages of mammals and primates (Figure 1A). The result revealed that the majority of human coding genes evolved before the emergence of primates; about two thirds (67.3%) are from Aves (S1), while another 26.3% are from Monotremes (S2, 8.6%), Marsupials (S3, 4.3%), Laurasiatheria (S4, 12.8%), and Glires (S5, 0.6%) (Figure 1B). Despite the fact that only 6.4% of human coding genes evolved in primates (S6-S10) (Figure 1B), more than one third of them emerged in the Last Common Ancestor (hereafter referred to as LCA-S7, 2.3%) of New World Monkeys (NWMs), Old World Monkeys (OWMs), Apes, and humans (Figure 1B). Given that LCA-S7's emergence coincided with the evolution of the progenitor-buffered SSC system, genes that collectively emerged in LCA-S7 may be those that underlie this primate-specific system (Figure 1A-B).

We analyzed the organ-expression pattern of each evolutionary cluster of human coding genes based on existing bulk RNA-sequencing data of various human organs (Figure 2) (*34*). It is most notable that the percentiles of testis-specific genes in the evolutionary clusters increased

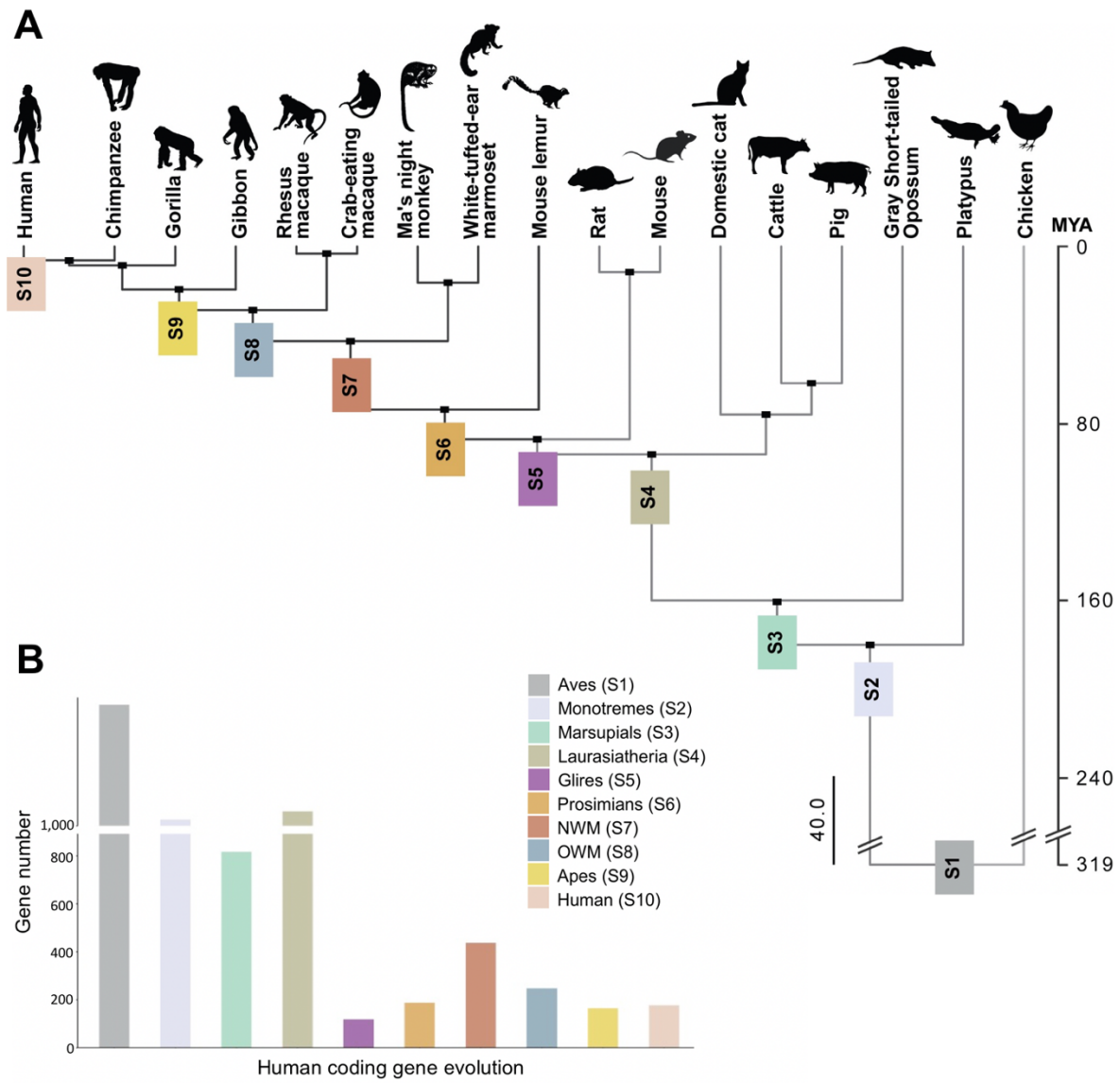
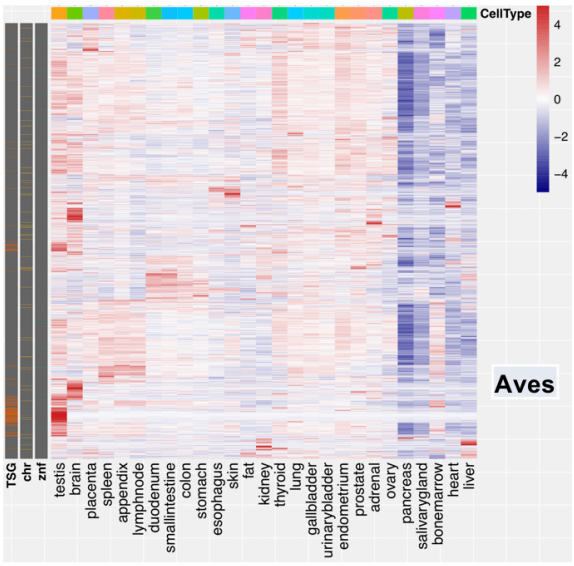
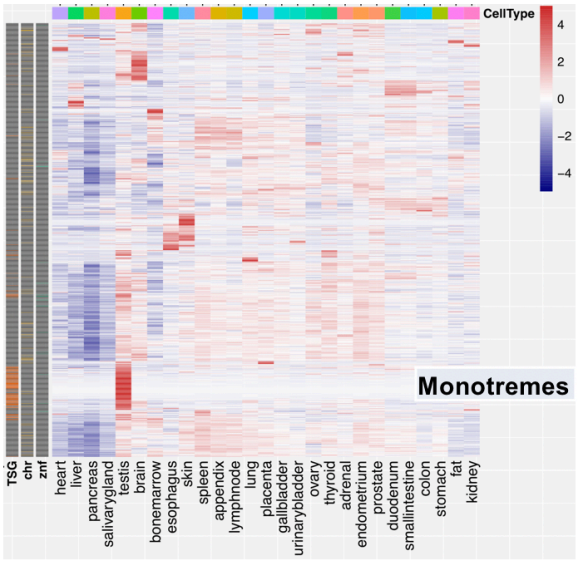
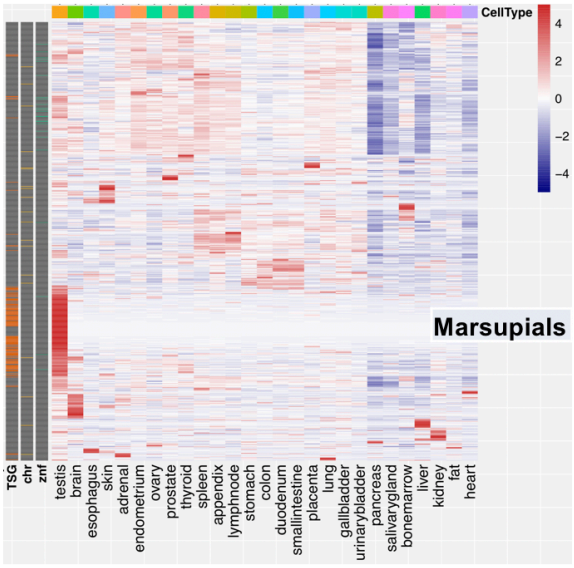


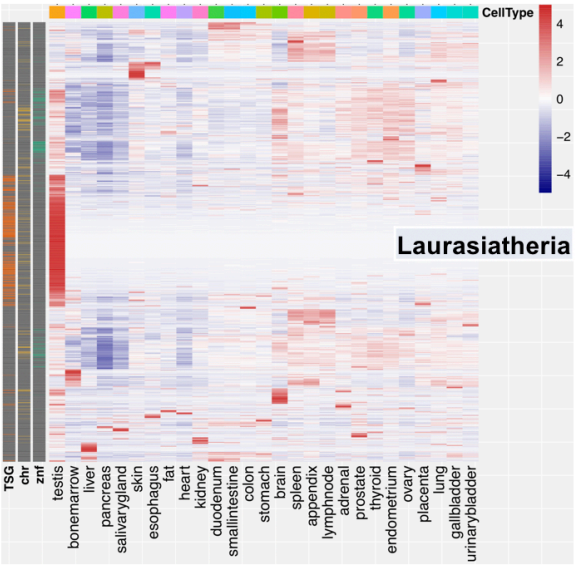
Figure 1. Phylogenetical dating analysis of human coding genes. **A.** Evolutionary tree covering 16 representative species in 10 evolutionary clusters (S1-S10). **B.** Human coding gene numbers in each evolutionary cluster.

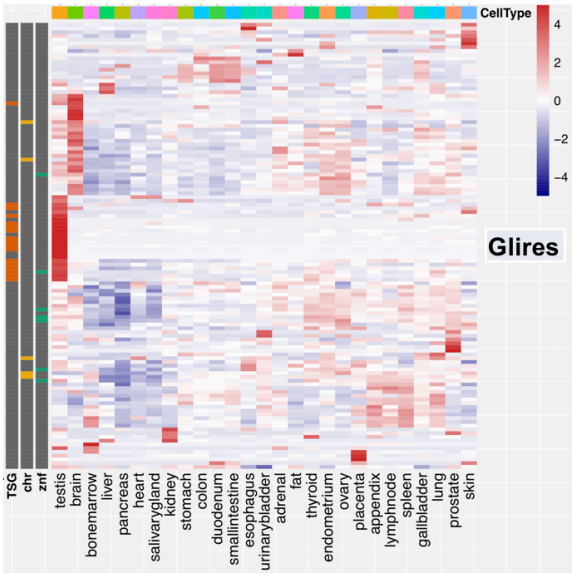
about 4-5 folds from ~6% in Aves to ~25-35% in primates (Figure 2 and 3A) (*35*). In addition to testis, other reproductive organs as well as brain and bone marrow were also found to possess small blocks of organ-specific genes, but the percentiles of these gene blocks do not exhibit trends of changes along the evolutionary trajectory (Figure 2), supporting the notion that testis specific genes evolved late during evolution.

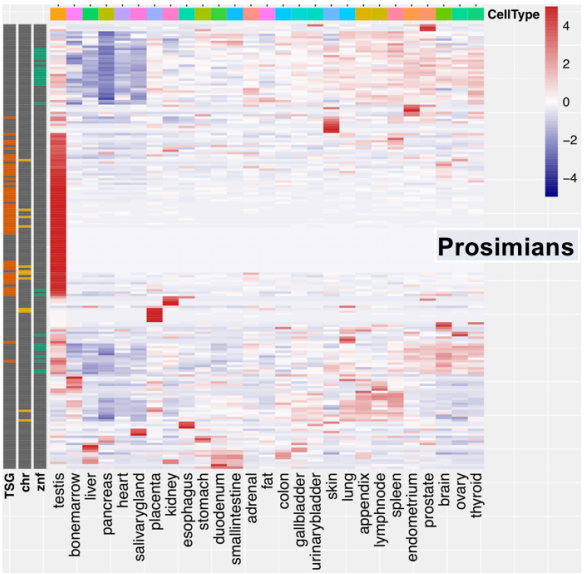

CellType
4
2
0
−2
−4
TSG
chr
znf
testis
brain
placenta
spleen
appendix
lymphnode
duodenum
smallintestine
colon
stomach
esophagus
skin
fat
kidney
thyroid
lung
gallbladder
urinarybladder
endometrium
prostate
adrenal
ovary
pancreas
salivarygland
bonemarrow
heart
liver
Aves


CellType
4
2
0
−2
−4
TSG
chr
znf
heart
liver
pancreas
salivarygland
testis
brain
bonemarrow
esophagus
skin
spleen
appendix
lymphnode
lung
placenta
gallbladder
urinarybladder
ovary
thyroid
adrenal
endometrium
prostate
duodenum
smallintestine
colon
stomach
fat
kidney
Monotremes


CellType
4
2
0
−2
−4
TSG
chr
znf
testis
brain
esophagus
skin
adrenal
endometrium
ovary
prostate
thyroid
spleen
appendix
lymphnode
stomach
colon
duodenum
smallintestine
placenta
lung
gallbladder
urinarybladder
pancreas
salivarygland
bonemarrow
liver
kidney
fat
heart
Marsupials


CellType
4
2
0
−2
−4
TSG
chr
znf
testis
bonemarrow
liver
pancreas
salivarygland
skin
esophagus
fat
heart
kidney
duodenum
smallintestine
colon
stomach
brain
spleen
appendix
lymphnode
adrenal
prostate
thyroid
endometrium
ovary
placenta
lung
gallbladder
urinarybladder
Laurasiatheria


CellType
4
2
0
−2
−4
TSG
chr
znf
testis
brain
bonemarrow
liver
pancreas
heart
salivarygland
kidney
stomach
colon
duodenum
smallintestine
esophagus
urinarybladder
adrenal
fat
thyroid
endometrium
ovary
placenta
appendix
lymphnode
spleen
gallbladder
lung
prostate
skin
Glires


CellType
4
2
0
−2
−4
TSG
chr
znf
testis
bonemarrow
liver
pancreas
heart
salivarygland
placenta
kidney
esophagus
stomach
duodenum
smallintestine
adrenal
fat
colon
gallbladder
urinarybladder
skin
lung
appendix
lymphnode
spleen
endometrium
prostate
brain
ovary
thyroid
Prosimians

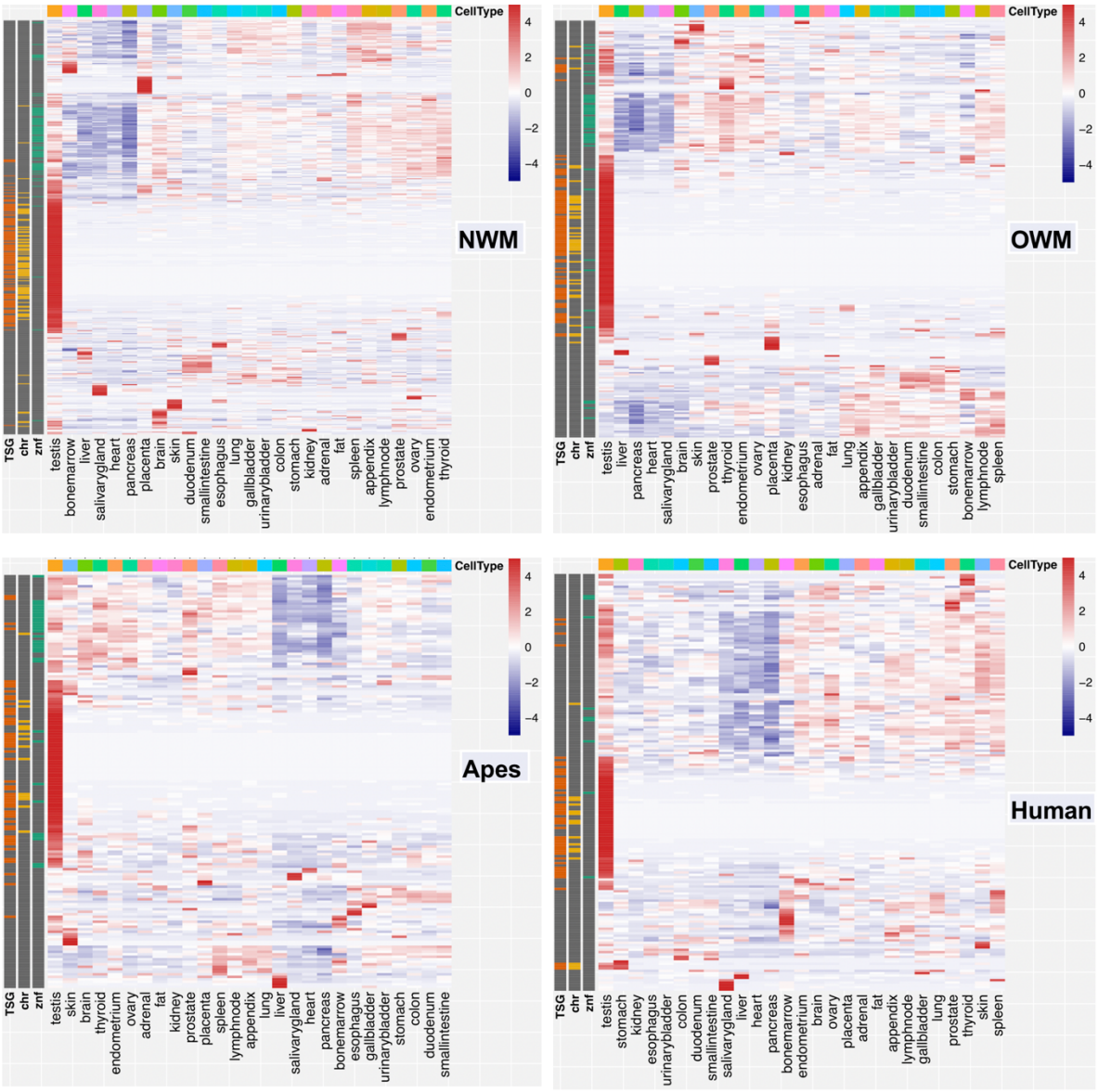


Figure 2. Heatmaps of the organ-expression pattern of human coding genes in each evolutionary cluster. Genes and organs are horizontally and vertically arranged, respectively. Testis-specific genes, X-chromosome genes, and ZNF zinc finger genes are correspondingly marked with orange, yellow and green lines in the "TSG", "chr" and "znf" columns on the left.

Classical evolutionary genetics has converged on a single explanation for the evolution of sex chromosomes in the "Ohno's law", which predicted that X chromosomes would be the gene museums that have preserved gene contents of ancestral autosomes (*36*). Genetic conservation

on the X chromosome has been affirmed via comparing numerous genomic data from multiple placental mammals (*37*). However, recent genomic sequences have led to discoveries that violate "Ohno's law": Some regions of the X chromosome, in particular genes expressed predominantly in the testis, are rapidly evolving (*37*). These male-biased genes are the sources of gene innovation on the X chromosome especially in the primate lineage. Consistent with this notion, a close examination of human coding genes' chromosomal locus and classification in the evolutionary clusters showed high correlations between newly emerged testis-specific genes and their positions on the X-chromosome in the primate lineage (Figure 2 and 3B). Specifically, about 40% of newly emerged testis-specific genes during the evolution of NWMs belong to the X chromosome, with most of these new genes being CTAs that have limited information about their functions in spermatogenesis (Figure 4). It is noteworthy that these testis-specific genes in NWM emerged alongside with the progenitor buffered SSC system, indicating that these newly evolved CTA genes could play important roles in maintaining high-stemness throughout the SSC progenitor buffer in primates.

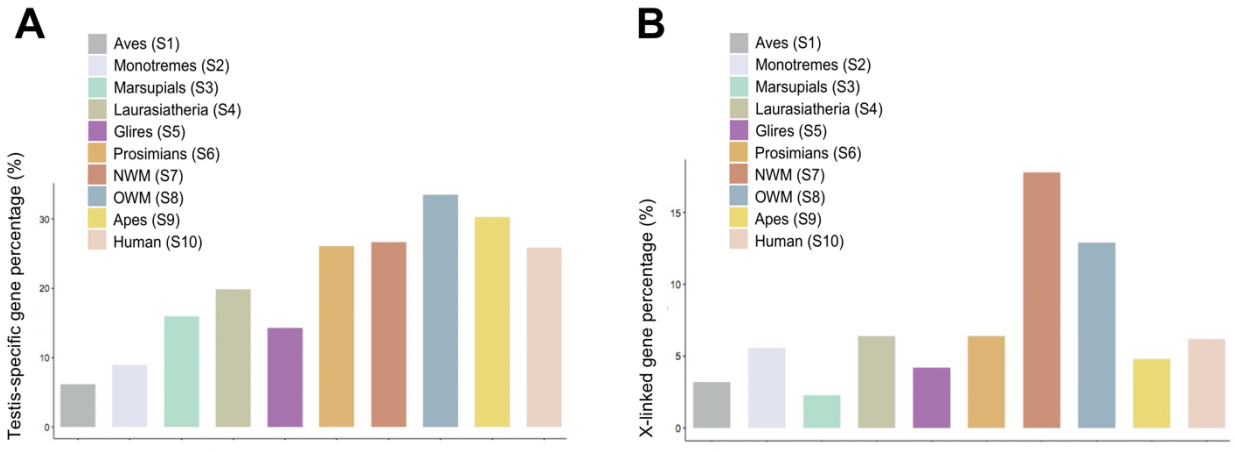


Figure 3. A. Testis-specific gene percentiles in all evolutionary clusters (S1-S10). **B.** X-linked gene percentiles in all evolutionary clusters (S1-S10).

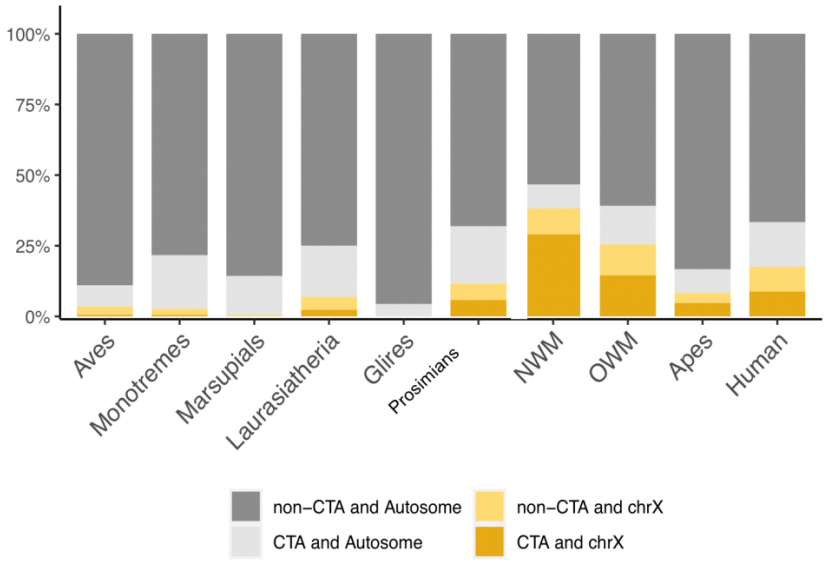


Figure 4. Percentiles of Chromosome X and autosomal CTAs among testis-specific genes in all evolutionary clusters (S1-S10).

To identify CTA genes that could be crucial for the SSC progenitor buffer in primates and the CSCs in malignant tumors, we utilized existing single-cell RNA sequencing (scRNAseq) data of human testicular germ cells (*38*) and analyzed the expression pattern of testis-specific genes emerging in NWMs (S7) (*35*). Our analysis presented a clear pattern: X-linked testis-specific genes (mostly CTAs) are predominantly expressed in spermatogonia (SPG), a majority of which are undifferentiated SSC progenitors (Figure 5A-B) (*33*). Upon further examination, we found that these genes belong to several families with arrays of duplicated genes, including the CT45A-SAGE1, CTAG, FAM9, GAGE, PAGE, RHOX, and VCX families (Figure 5B). It is noteworthy that none of these families have been previously reported to play roles in the primate progenitor buffered SSC system. One or two genes from each of these families were selected to test their importance in human SSCs and CSCs (Figure 5C). To broaden the scope, we also included an X-chromosome single-copy CTA gene TFDP3 and two autosomal CTA genes, ACTL8 from Chr-1 and PRAME from Chr-22, in the experiments. Notably, high expression of seven selected CTAs displayed strong correlations with shorter pan-cancer survival, implying

that their expression is associated with the advanced stage of cancers and poor prognosis (Figure 6).

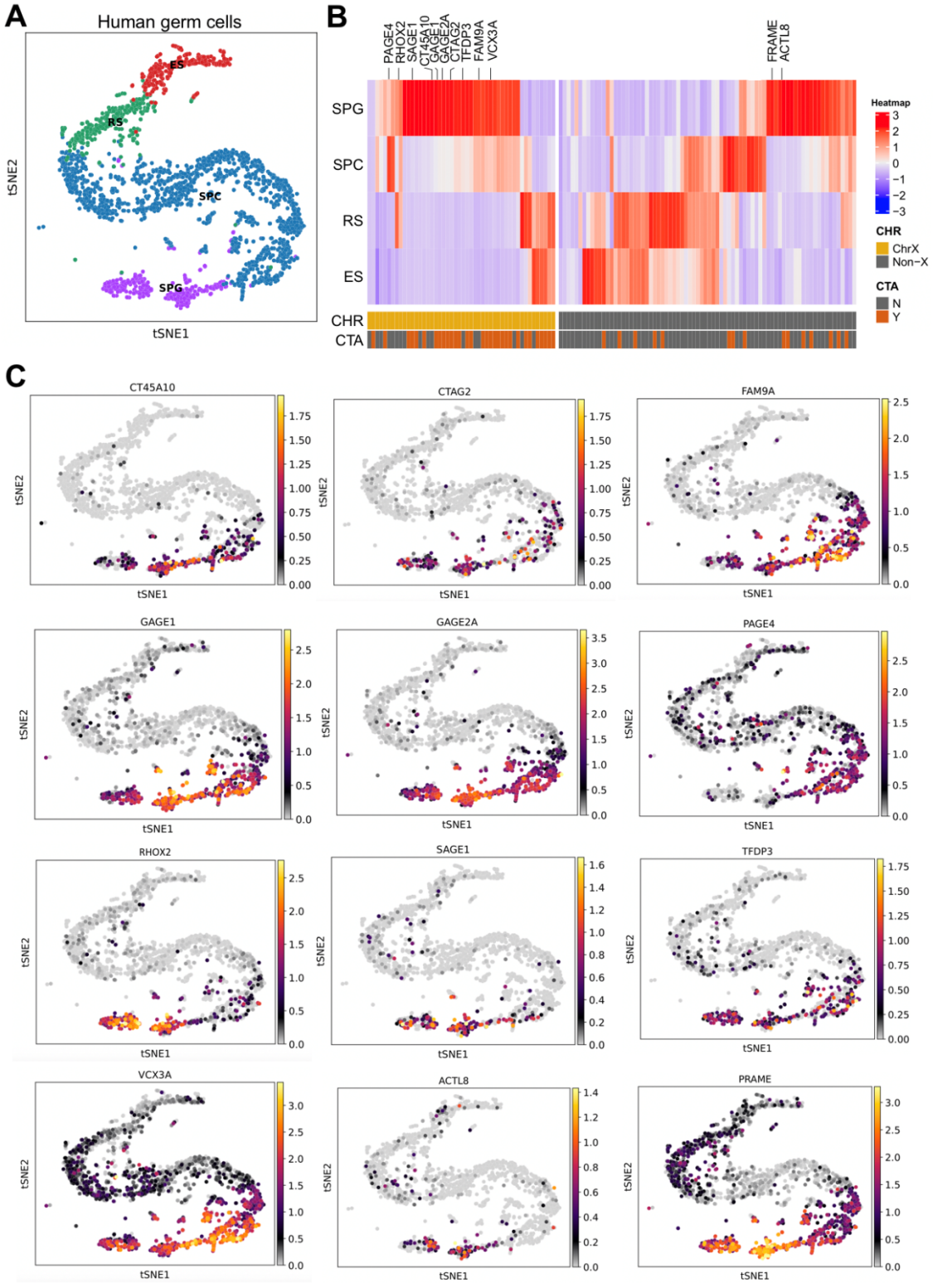


Figure 5. Expression pattern analysis of testis-specific genes emerged from NWMs (S7) based on scRNAseq data of human testicular germ cells. **A.** Clustering of human testicular germ cells reveals four subtypes, as shown in the t-distributed stochastic neighbor embedding (t-SNE) plot. SPG: spermatogonia, SPC: spermatocyte, RS: round spermatid, ES: elongated spermatid. **B.** Heatmap of expression pattern of testis-specific genes evolved in NWMs. X-chromosome genes and CTA genes are respectively marked with yellow and orange lines in the "CHR" and "CTA" columns at the bottom. Selected CTA genes are labeled in the heatmap. **C.** Per-cell expression level of each selected CTA gene.

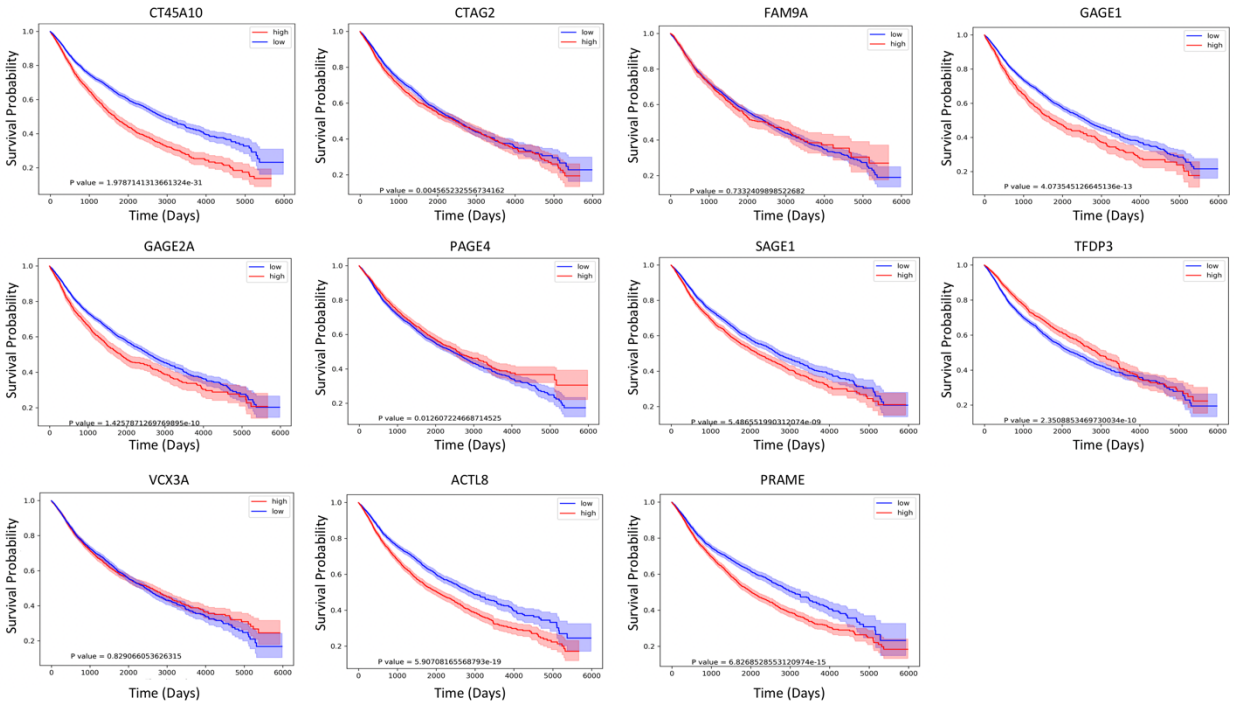
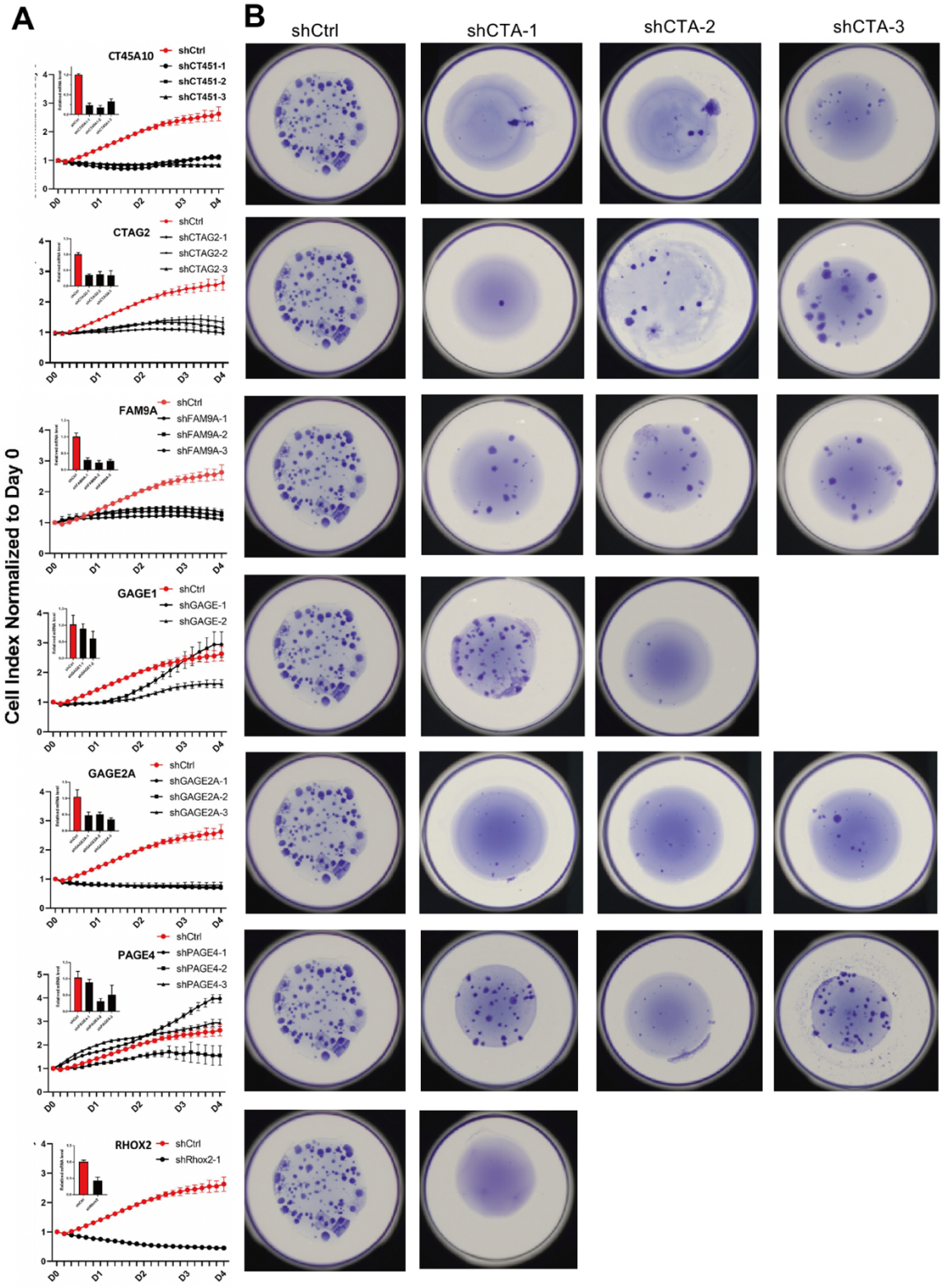


Figure 6. Pan-cancer Kaplan-Meier overall survival curves for patients with high and low expression of selected CTAs (RHOX2 is not included). The *p*-values were calculated with log-rank test.

B. CTA knockdown suppresses the proliferation of human SSC

We took advantage of the *in vitro* long-term human SSC culture system established in Shanghai Jiao Tong University School of Medicine to explore how CTAs contribute to SSC stemness. In this system, SSCs are cultured in a feeder free, low-attachment well and maintained in customized medium with supplemented growth factors, which endow cells with a distinct self-renewing state. The cells grow in a semi-suspended manner, loosely connected with the culture

vessel's surface, in two-dimensional (2D) cultures (Figure 7A), or are transferred into low attachment wells with Matrigel to form colonies in three-dimensional (3D) cultures (Figure 7B).



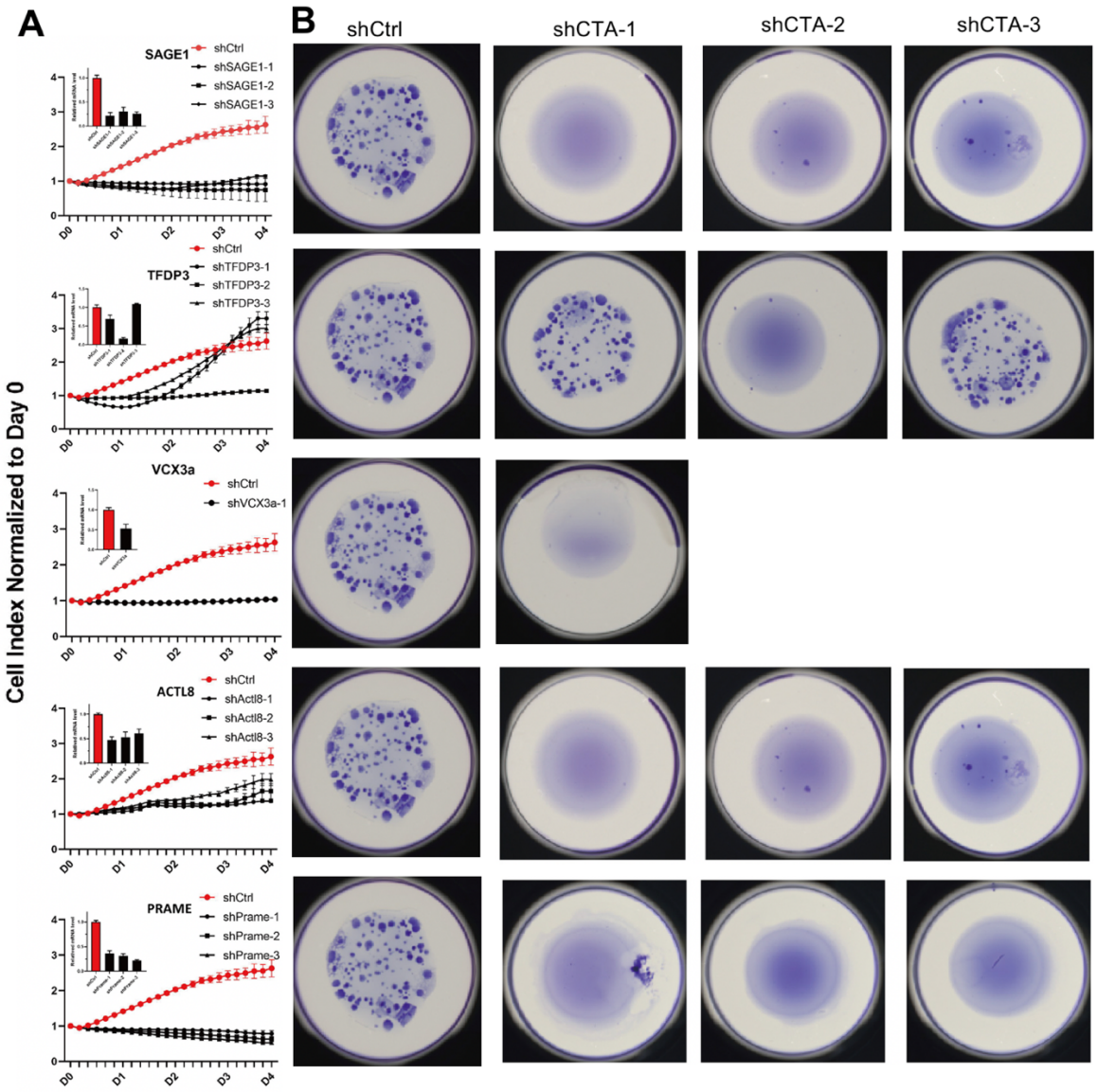


Figure 7. SSC proliferation assays to evaluate the selected CTAs' functions. **A.** SSCs were infected with shRNA against serials of CTA genes. Cell growth in 2D cultures was assessed by the IncuCyte S3 Live-Cell Analysis System and calculated with normalization to Day 0. The bar inserted shows the effects of each shRNA on CTA mRNA levels measured by qRT-PCR. **B.** Colonies of shCtrl and shCTA SSCs in 3D Matrigel were stained with crystal violet. Representative images are shown.

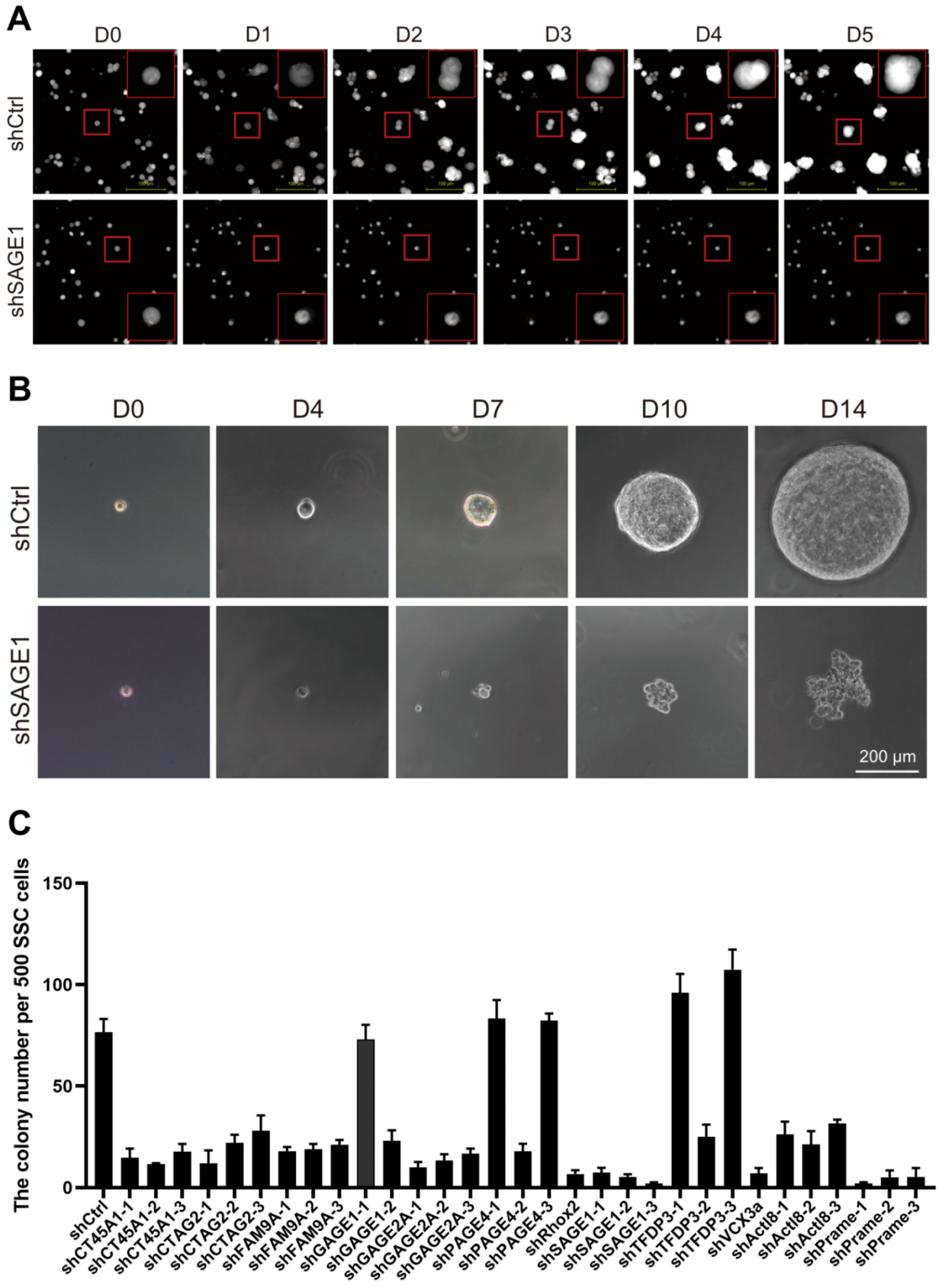

A
D0
D1
D2
D3
D4
D5
shCtrl
shSAGE1
B
D0
D4
D7
D10
D14
shCtrl
shSAGE1
200 μm
C
The colony number per 500 SSC cells
150
100
50
0
shCtrl
shCT45A1-1
shCT45A1-2
shCT45A1-3
shCTAG2-1
shCTAG2-2
shCTAG2-3
shFAM9A-1
shFAM9A-2
shFAM9A-3
shGAGE1-1
shGAGE1-2
shGAGE2A-1
shGAGE2A-2
shGAGE2A-3
shPAGE4-1
shPAGE4-2
shPAGE4-3
shRhox2
shSAGE1-1
shSAGE1-2
shSAGE1-3
shTFDP3-1
shTFDP3-2
shTFDP3-3
shVCX3a
shActl8-1
shActl8-2
shActl8-3
shPrame-1
shPrame-2
shPrame-3

Figure 8. 2D and 3D proliferation assays of the selected CTA genes **A.** Growth of shCtrl and shSAGE1 SSCs in 2D culture was recorded using the Livecyte's ptychography technology. Representative phase images from Day 0 to Day 5 are shown. **B.** Growth of shCtrl and shSAGE1 SSCs in Matrigel was recorded by an inverted microscope. Representative phase images at Day 0, 4, 7, 10, 14 are shown. **C.** SSCs were infected with shRNA against selected CTA genes and cultured in Matrigel. Colony number per 500 cells was calculated to reflect colony formation efficiency.

We set out to investigate the potential function of CTAs selected from the bioinformatics analyses by examining CTA-knockdown phenotypes. Several short hairpin RNAs (shRNA) targeting each CTA gene were designed, and cloned into lentivirus vectors with a red fluorescent protein (RFP) tag (Table 2). Lentivirus carrying a CTA gene's shRNA was transfected into cultured SSCs, and cells expressing shRNA were enriched by the co-expressed RFP. The knockdown effect of CTA genes was examined at the mRNA level using quantitative reverse-transcription polymerase chain reaction (qRT-PCR) assay to demonstrate the efficiency of shRNA sequences in knocking down mRNA levels of targeted CTA genes (Figure 7A (insets in the growth curve diagrams) and 8A) (Table 3).

The 2D culture results showed that, except for PAGE4, the other selected CTAs all contributed to the proliferation of cultured-SSCs, and the degree of inhibition on SSC growth corresponded well with the suppression of CTA mRNA levels demonstrated by qRT-PCR data (Figure 7A and 8A). In 3D culture results, the spheroid of shCtrl SSCs exponentially amplified with a clone diameter doubling time of about 2.5-3 days, while cells treated with validated shRNAs refrained from amplifying three-dimensionally up to 14 days (Figure 7B and 8B). Again, 3D growth data completely mirrored the corresponding 2D growth curves and the decreases of mRNA levels by shRNA knockdown, highlighting the consistency between the 2D and 3D growth experiments (Figure 7A-B and 8C). Taken together, our analyses demonstrate

that both X chromosome and autosomal CTA genes evolved in NWMs likely play important roles in the self-renewal and long-term maintenance of a large, undifferentiated SSC progenitor buffer in higher primates.

C. CTA knockdown leads to decreased spheroid formation capacity in cancer cells

To investigate the influence of the selected CTAs on CSC stemness, we performed the *in vitro* spheroid formation assay—a standard functional method for assessing the stemness of CSCs through observing their colony-forming units (CFUs) in a 3D culture environment (*39, 40*). We first made use of the Cancer Cell Line Encyclopedia (CCLE) database (https://sites.broadinstitute.org/ccle/) to examine the expression levels of selected CTAs in cancer cell lines. Based on the analysis, we chose six cell lines available in the laboratory with

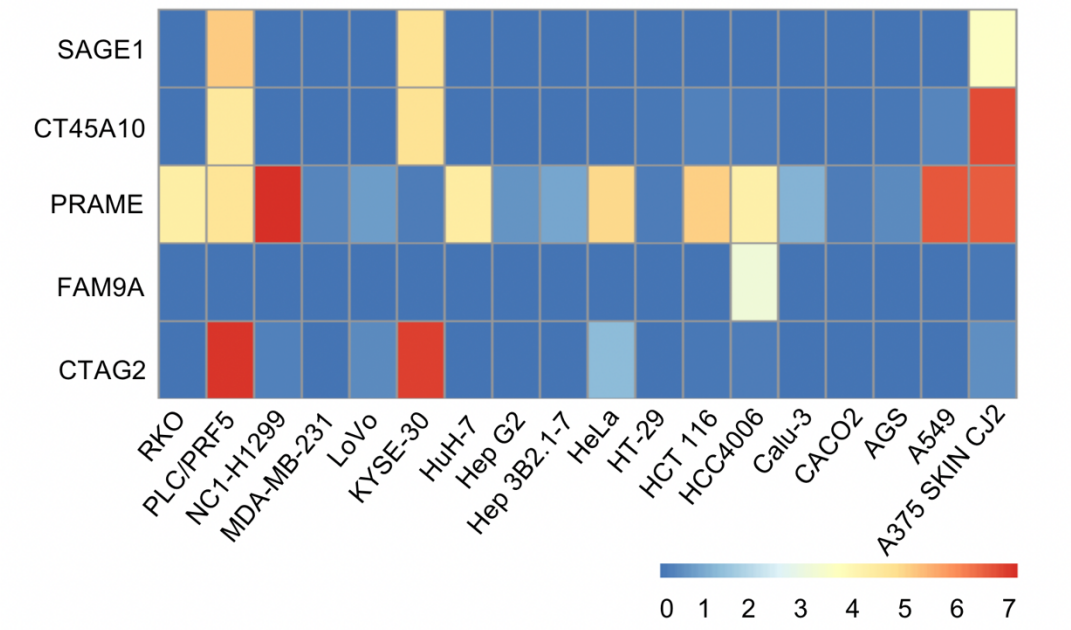


Figure 9. Heatmap of mRNA expression levels (log1p TPM) of target CTAs in different available cancer cell lines from the CCLE database.

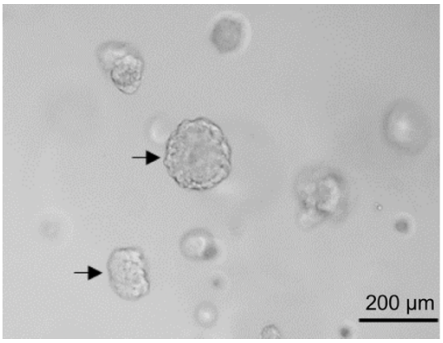


Figure 10. Typical tumor-spheres derived from single-cell of A375 larger than 50 µm cutoff size (as indicated by the black arrows). Spheres smaller than 50 µm are believed negative herein.

high expression of CTAs as evaluated in the SSC proliferation assays (Figure 7 and 8) to study their functions in CSC stemness (Figure 9). These cancer cell lines are A375 for SAGE1, KYSE-30 for CT45A10, A549 for PRAME, PLC/PRF/5 for CTAG2, HCC4006 for FAM9A, and MDA-MB-231 for ACTL8 (Figure 9). A controlled number of cells were initially seeded in the plates, and the number and size of spheroids were measured about 10 days post seeding. Spheres larger than 50 µm shown in Figure 10 were used as a cutoff size. The results indicated that five out of six selected CTAs (SAGE1, CT45A10, PRMAE, CTAG2, and FAM9A) are crucial to the spheroid formation in cancer cells (Figure 11). The number and size of the spheres significantly decreased upon CTA knockdown, suggesting that these CTAs play key roles in CSC proliferation (Figure 10). At the same time, it is worth pointing out that knockdown of ACTL8 in cancer cells exhibited a stimulatory effect on proliferation, opposite to its effect seen on SSC proliferation (Figure 11). This is probably due to different roles of ACTL8 in SSC and MDA-MB-231 cancer cells, which requires further investigation. Nonetheless, our observation that knocking down the expression of CTAs consistently causes proliferation suppression in both SSCs and CSCs supports the "double-edged sword" hypothesis: CSCs adversely hijack primate-specific CTAs' ability of maintaining high-stemness proliferation that is otherwise only restricted to SSCs under normal physiological conditions.

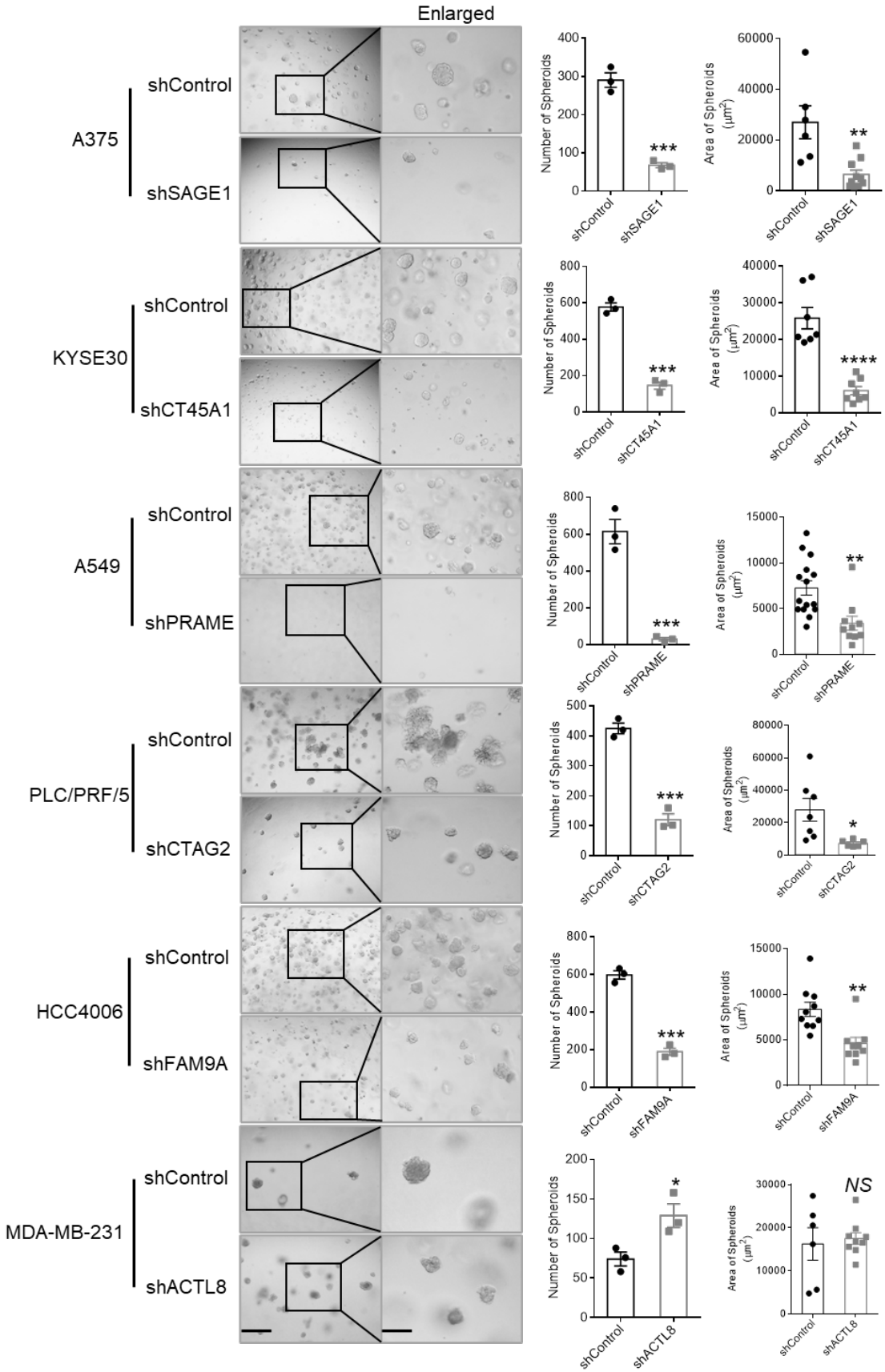

Enlarged
A375
shControl
shSAGE1
KYSE30
shControl
shCT45A1
A549
shControl
shPRAME
PLC/PRF/5
shControl
shCTAG2
HCC4006
shControl
shFAM9A
MDA-MB-231
shControl
shACTL8
Number of Spheroids
Area of Spheroids (μm²)
***
**
****
*
NS

Figure 11. 3D tumor spheroid formation assay shows CSC tumorigenic capacity to be dramatically decreased upon knocking down CTA genes SAGE1, CT45A1, PRAME, CTAG2 and FAM9A. Representative images are shown. Scale bars: 500 μm and 200 μm for low and high magnifications, respectively. Data were shown as mean ± s.e.m. of independent experiments. *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$, ****: $P < 0.0001$, Paired Student's t test for the two groups.

IV. DISCUSSION

In this study, we explored the hypothesis that CTAs act as a "double-edged sword", playing key roles of maintaining stemness and proliferation in CSCs and SSCs. Through bioinformatic analysis, we were able to detect a group of testis-specific genes on the X chromosome, mostly CTAs, during the evolution of NWMs (Figure 2-4). Twelve of the CTAs in the group were selected for wet lab experimentation. 2D and 3D growth experiments on all selected genes were performed within the SSC system, half of which were picked for further experiments in cancer cells based on the availability of cancer cell lines in the laboratory. From the results of our 2D and 3D growth experiments in the two stem cell systems, it can be seen that the majority of knockdown assays strongly inhibited the growth of CSCs and SSCs (Figure 7, 8 and 11). The consistency between the results of the two stem cell systems verified our hypothesis of CTAs being "double-edged swords", demonstrating that these genes are equally important for maintaining stemness and proliferation in SSCs and CSCs. These discoveries will lead to a deeper understanding of the primate specific progenitor buffer SSC system in male reproduction biology, and can provide a new direction for future research regarding tumor malignancy. The restricted cancer-testis expression pattern endows CTAs with the unique opportunity for highly specific cancer therapeutics that could inhibit cancer cell growth with limited toxicities on normal cells.

Our current bioinformatic analyses and experimental approaches have room for improvement. The accuracy of the algorithm we developed for the bioinformatic analysis may be enhanced by more complex analytical methods. Although we implemented original methods to reduce false positives and negatives, these improvements are still based on the binary list of confidence values. For instance, our current algorithm defines a gene's earliest appearance based upon a given list of emergence confidence levels that are either 1.0 or 0.0, which could fail to identify ancestral genes whose sequences contain fragments substantially different from their primate counterparts. More complex methods such as dedicated sequential alignment for all genes could be implemented for optimal accuracy. Moreover, our cancer experiments pale in comparison to those done in the physiological SSC system in that the limited number of cancer cell lines results in an incomplete representation of general CTA functionality in cancer. Our experiment is the first attempt. Further *in vivo* investigations will be needed in the future to fully understand the biological mechanisms behind CTA operation in cancer.

The phylogenetic dating analysis proved to be an effective approach to seek novel evolutionary insights into the underpinning mechanism of the progenitor buffered SSC system in primates. Evolution is the best experiment. It is because of the need for survival that humans and other organisms have evolved over millions of years to acquire powerful physiological functions they possess today. By understanding the process of evolution, we would be able to gain novel insights not only into how these functions are being evolved to improve physiological performance, but also into how they are hijacked by the wrong hands to use against us. Diseases do not create, they merely unleash and misuse powerful physiological functions in places where they should be regulated, causing disruptions within the human body. Pathology is based upon physiology, and physiology is determined by evolution. With many of today's cancer research

being mainly focused on cancer itself, studies of a related physiological system guided by evolutionary insights could provide an alternative approach for therapeutic interventions for cancers.

V. CONCLUSION

Through examining the similarities between human SSCs and CSCs, this paper discusses the important role that CTAs play in maintaining high stemness in these two stem cell systems and the evolutionary timeline behind the CTAs. The effect that these genes have on the stemness of SSCs and CSCs are hypothesized, analyzed using bioinformatic tools performed upon existing evolution and cancer databases, and verified through 2D and 3D cell growth experiments. This study suggests that CTAs have great potential of becoming novel targets for future cancer therapeutics that could specifically inhibit cancer cell growth with limited toxicities on normal cells. It also exemplifies the power of evolution perspective in understanding the physiology which in turn explains the pathology of diseases.

REFERENCES

1. E. Y. Lee, W. J. Muller, Oncogenes and tumor suppressor genes. *Cold Spring Harb Perspect Biol* **2**, a003236 (2010).
2. S. H. Yuspa, A. A. Dlugosz, C. K. Cheng, M. F. Denning, T. Tennenbaum, A. B. Glick, W. C. Weinberg, Role of oncogenes and tumor suppressor genes in multistage carcinogenesis. *J Invest Dermatol* **103**, 90S-95S (1994).
3. D. Hanahan, Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**, 31-46 (2022).
4. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
5. E. Batlle, H. Clevers, Cancer stem cells revisited. *Nat Med* **23**, 1124-1134 (2017).

6. H. Clevers, The cancer stem cell: premises, promises and challenges. *Nat Med* **17**, 313-319 (2011).
7. J. E. Dick, Stem cell concepts renew cancer research. *Blood* **112**, 4793-4807 (2008).
8. H. Clevers, The intestinal crypt, a prototype stem cell compartment. *Cell* **154**, 274-284 (2013).
9. S. Doulatov, F. Notta, E. Laurenti, J. E. Dick, Hematopoiesis: a human perspective. *Cell Stem Cell* **10**, 120-136 (2012).
10. A. Uccelli, L. Moretta, V. Pistoia, Mesenchymal stem cells in health and disease. *Nat Rev Immunol* **8**, 726-736 (2008).
11. G. B. Pierce, W. C. Speers, Tumors as caricatures of the process of tissue renewal: prospects for therapy by directing differentiation. *Cancer Res* **48**, 1996-2004 (1988).
12. P. Dalerba, R. W. Cho, M. F. Clarke, Cancer stem cells: models and concepts. *Annu Rev Med* **58**, 267-284 (2007).
13. P. R. Prasetyanti, J. P. Medema, Intra-tumor heterogeneity from a cancer stem cell perspective. *Mol Cancer* **16**, 41 (2017).
14. D. G. Tang, Understanding cancer stem cell heterogeneity and plasticity. *Cell Res* **22**, 457-472 (2012).
15. P. Kulkarni, T. Shiraishi, K. Rajagopalan, R. Kim, S. M. Mooney, R. H. Getzenberg, Cancer/testis antigens and urological malignancies. *Nat Rev Urol* **9**, 386-396 (2012).
16. F. Meklat, Z. Li, Z. Wang, Y. Zhang, J. Zhang, A. Jewell, S. H. Lim, Cancer-testis antigens in haematological malignancies. *Br J Haematol* **136**, 769-776 (2007).
17. O. Gordeeva, Cancer-testis antigens: Unique cancer stem cell biomarkers and targets for cancer therapy. *Semin Cancer Biol* **53**, 75-89 (2018).
18. D. S. Nin, L. W. Deng, Biology of Cancer-Testis Antigens and Their Therapeutic Implications in Cancer. *Cells* **12**, (2023).
19. B. T. Phillips, K. Gassei, K. E. Orwig, Spermatogonial stem cell regulation and spermatogenesis. *Philos Trans R Soc Lond B Biol Sci* **365**, 1663-1678 (2010).
20. M. Kanatsu-Shinohara, T. Shinohara, Spermatogonial stem cell self-renewal and development. *Annu Rev Cell Dev Biol* **29**, 163-187 (2013).
21. A. J. Simpson, O. L. Caballero, A. Jungbluth, Y. T. Chen, L. J. Old, Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer* **5**, 615-625 (2005).
22. J. Ehmcke, J. Wistuba, S. Schlatt, Spermatogonial stem cells: questions, models and perspectives. *Hum Reprod Update* **12**, 275-282 (2006).
23. S. Sharma, J. Wistuba, T. Pock, S. Schlatt, N. Neuhaus, Spermatogonial stem cells: updates from specification to clinical relevance. *Hum Reprod Update* **25**, 275-297 (2019).
24. D. Smedley, S. Haider, B. Ballester, R. Holland, D. London, G. Thorisson, A. Kasprzyk, BioMart--biological queries made easy. *BMC Genomics* **10**, 22 (2009).
25. S. Kumar, M. Suleski, J. M. Craig, A. E. Kasprowicz, M. Sanderford, M. Li, G. Stecher, S. B. Hedges, TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol Biol Evol* **39**, (2022).
26. L. Fagerberg, B. M. Hallstrom, P. Oksvold, C. Kampf, D. Djureinovic, J. Odeberg, M. Habuka, S. Tahmasebpoor, A. Danielsson, K. Edlund, A. Asplund, E. Sjostedt, E. Lundberg, C. A. Szigyarto, M. Skogs, J. O. Takanen, H. Berling, H. Tegel, J. Mulder, P. Nilsson, J. M. Schwenk, C. Lindskog, F. Danielsson, A. Mardinoglu, A. Sivertsson, K. von Feilitzen, M. Forsberg, M. Zwahlen, I. Olsson, S. Navani, M. Huss, J. Nielsen, F. Ponten, M. Uhlen, Analysis of the human tissue-specific expression by genome-wide

integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* **13**, 397-406 (2014).
27. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018).
28. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
29. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
30. M. Wang, X. Liu, G. Chang, Y. Chen, G. An, L. Yan, S. Gao, Y. Xu, Y. Cui, J. Dong, Y. Chen, X. Fan, Y. Hu, K. Song, X. Zhu, Y. Gao, Z. Yao, S. Bian, Y. Hou, J. Lu, R. Wang, Y. Fan, Y. Lian, W. Tang, Y. Wang, J. Liu, L. Zhao, L. Wang, Z. Liu, R. Yuan, Y. Shi, B. Hu, X. Ren, F. Tang, X. Y. Zhao, J. Qiao, Single-Cell RNA Sequencing Analysis Reveals Sequential Cell Fate Transition during Human Spermatogenesis. *Cell Stem Cell* **23**, 599-614 e594 (2018).
31. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).
32. X. Zhang, X. Wang, Z. Yuan, S. J. Radford, C. Liu, S. K. Libutti, X. F. S. Zheng, Amino acids-Rab1A-mTORC1 signaling controls whole-body glucose homeostasis. *Cell Rep* **34**, 108830 (2021).
33. M. Uhlen, L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Ponten, Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
34. S. Ohno, *Sex Chromosomes and Sex-Linked Genes*. Monographs on Endocrinology (Springer Berlin, Heidelberg, 1966).
35. J. L. Mueller, H. Skaletsky, L. G. Brown, S. Zaghlul, S. Rock, T. Graves, K. Auger, W. C. Warren, R. K. Wilson, D. C. Page, Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* **45**, 1083-1087 (2013).
36. X. L. Ma, Y. F. Sun, B. L. Wang, M. N. Shen, Y. Zhou, J. W. Chen, B. Hu, Z. J. Gong, X. Zhang, Y. Cao, B. S. Pan, J. Zhou, J. Fan, W. Guo, X. R. Yang, Sphere-forming culture enriches liver cancer stem cells and reveals Stearoyl-CoA desaturase 1 as a potential therapeutic target. *BMC Cancer* **19**, 760 (2019).
37. H. F. Bahmad, K. Cheaito, R. M. Chalhoub, O. Hadadeh, A. Monzer, F. Ballout, A. El-Hajj, D. Mukherji, Y. N. Liu, G. Daoud, W. Abou-Kheir, Sphere-Formation Assay: Three-Dimensional in vitro Culturing of Prostate Cancer Stem/Progenitor Sphere-Forming Cells. *Front Oncol* **8**, 347 (2018).

ACKNOWLEDGMENTS

From my school years learning biology, chemistry, and other natural sciences, I have enjoyed solving scientific problems and participated in research-oriented competitions such as iGEM and IMMC. I have always been longing to participate in research of a field on the forefront that people can benefit from and dive in with excitement when such an opportunity arises. Shanghai Jiao Tong University bestowed me upon a precious opportunity to participate in a cancer research project. The most difficult part about fighting against this deadly disease lies in the tumor's relapse even after extensive treatment. The feature of tumorigenesis seemed to me reminiscent of normal stem cell behavior in the human body, which intrigued me to investigate deeper into their correlation. The absence of the target genes in mice, the most common model organism for research, led me to tackle the problem through an evolutionary lens. Not only did I witness how an evolutionary perspective helped unveil a theory that cancer could be a misplacement of a physiological function, but also it taught me that evolution is the best experiment by nature.

I extend my deepest gratitude to my mentors: Teacher Hui Liao, who imparted vital biological knowledge and offered invaluable feedback on this paper, and Dr. Xin Zhang, who furnished me with the SSC culture system and guided me through experiments involving cancer cell lines, as well as bioinformatics procedures.