IDENTIFICATION OF GENOMIC AND TRANSCRIPTOMIC VARIANTS IN CHORDOMA CANCER STEM CELLS

by
Nur Ekimci Gürcan

Submitted to Graduate School of Natural and Applied Sciences
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in
Biotechnology

Yeditepe University
2023

IDENTIFICATION OF GENOMIC AND TRANSCRIPTOMIC VARIANTS IN CHORDOMA CANCER STEM CELLS

APPROVED BY:

Prof. Dr. Ömer Faruk Bayrak ……
(Thesis Supervisor)
(Yeditepe University)

Assoc. Prof. Dr. Ayşegül Kuşkucu ……
(Yeditepe University)

Prof. Dr. Engin Ulukaya ……
(İstinye University)

Prof. Dr. Fikrettin Şahin ……
(Yeditepe University)

Prof. Dr. Semra Demokan ……
(İstanbul University)

DATE OF APPROVAL: …./…./20…

I hereby declare that this thesis is my own work and that all information in this thesis has been obtained and presented in accordance with academic rules and ethical conduct. I have fully cited and referenced all material and results as required by these rules and conduct, and this thesis study does not contain any plagiarism. The necessary permissions have been obtained if any material used in the thesis requires copyright. No material from this thesis has been used for the award of another degree.

I accept all kinds of legal liability that may arise in case contrary to these situations.

Name, Last name Nur Ekimci Gürcan

Signature …..…….……..

ACKNOWLEDGEMENTS

This study would not have been completed without the help and support of many individuals. I would like to gratefully acknowledge the support and guidance of my supervisor Prof Dr. Ömer Faruk BAYRAK for his invaluable advice, support, and knowledge. I am heartily thankful to Assoc. Prof. Dr. Ayşegül KUŞKUCU and Assist. Prof Dr. Esra AYDEMIR for their joy, kind support, and wisdom.

I would also like to thank Assoc. Prof Dr. Emrah NİKEREL and Nehir KIZILİLSOLEY for their valuable contributions and technical support on my study.

I want to introduce my appreciation to my dear Melike BAYINDIR BİLGİÇ for her invaluable friendship and support both in academia and life. This journey wouldn't be completed without my dear lab mates Utku ÖZBEY, Didem TECİMEL, Berfin UZUNKAYA, Gizem İNETAŞ YENGİN, and Dr. Didem SEVEN. I am grateful for their kind support during my Ph.D. journey. I am appreciated to work with Assist. Prof. Dr. Negar TAGHAVI POURIANAZAR, Assist. Prof. Dr. Başak ARU, Dr. Emre Can TÜYSÜZ, Dr. Öznur SUAKAR, Mesut ŞAHİN, Sezin GÜRKAN, Ebru BAKTEMUR.

I would like to express my deepest gratitude to my family for their existence, devotion, confidence, and unconditional support.

Finally, I would like to thank my dearest, Mehmet Buğrahan GÜRCAN. Without his tremendous understanding and encouragement, it would be impossible for me to complete this journey.

I would like to thank TÜBİTAK for their contribution to the scholarship that I received as a TÜBİTAK BİDEB 2211-A scholar.

This work was supported by the Scientific and Technological Research Council of Turkey [Grant number 118S691].

ABSTRACT

IDENTIFICATION OF GENOMIC AND TRANSCRIPTOMIC VARIANTS IN CHORDOMA CANCER STEM CELLS

Chordoma is a rare malignant bone tumor that originates from remnants of the embryonic notochord. Since a reliable treatment for chordoma has not yet been developed, determining the factors that contribute to the pathogenesis of chordoma is of great importance in terms of understanding the disease to develop new targets for treatment. In this dissertation study, prominent genomic variations and molecular pathways in chordoma cancer stem cells were investigated. For this purpose, initially, chordoma tumorspheres, side populations, and $CD133^+$ and $CD15^+$ double-positive cells were enriched in 4 different chordoma cell lines, and the obtained cells were compared in terms of expression changes in stemness and EMT marker genes. Among the techniques that were examined, the tumorsphere enrichment method was determined to be applicable for four chordoma cell lines. Then, genomic variations and differential gene expression profiles of chordoma tumorspheres were determined by next-generation DNA sequencing and total transcriptome microarray techniques. Functional enrichment analyses were used for the evaluation and determination of dysregulated pathways in chordoma tumorspheres in comparison to parental cell lines.

Finally, gene expression of *SEMA5A*, *TPD52L1*, and *GAPLINC* genes, whose expression was found to be upregulated in tumorspheres, were inhibited in two chordoma cell lines to investigate their downstream effects on invasion, migration, and viability. Results showed that siRNA-mediated silencing of *TPD52L1* significantly impairs both the migration and invasion characteristics of chordoma cell lines. Downregulation of *SEMA5A* inhibited migration and invasion of UM-Chor1 cells and impaired the invasive capacity of MUG-Chor1 cells. Additionally, decreased *GAPLINC* expression significantly reduced invasion and migration in UM-Chor1 cells, but did not affect MUG-Chor1 cells.

This study constitutes a first in the literature in terms of examining genomic and transcriptomic changes in tumorsphere-derived chordoma cancer stem cells as well as demonstrating the impact of *TPD52L1* and *SEMA5A* genes on the invasive and migrative capacity of chordoma cells.

ÖZET

KORDOMA KANSER KÖK HÜCRELERINDE GENOMIK VE TRANSKRIPTOMIK VARYASYONLARIN TESPITI

Embriyonik notokord kalıntılarından köken aldığı bilinen kordoma, nadir görülen bir kemik tümörüdür. Kordoma için güvenilir bir tedavi yaklaşımı henüz geliştirilemediğinden, kordoma patogenezinde etkili olan faktörlerin belirlenmesi, hastalığın anlaşılması ve tedavi için yeni hedefler geliştirebilmek açısından büyük önem arz etmektedir.

Bu tez çalışmasında, kordoma kanser kök hücrelerindeki genomik varyantlar moleküler yolaklar incelenmiştir. Bu amaçla ilk olarak 4 farklı kordoma hücre hattında kordoma tümör küreleri, yan popülasyon hücreleri ve CD133+/CD15+ ikili pozitif kordoma hücreleri alınarak köklülük ve epitel-mezenkim geçişi işaretçilerinin ifadesi bakımından karşılaştırılmışlardır. İncelenen teknikler arasında tümörküre kültürlerinin dört hücre hattı için de zenginleştirme için kullanılabileceği tespit edilmiştir. Sonrasında kordoma tümörkürelerinde genomik varyantlar ve diferansiyel gen ifade profilleri yeni nesil DNA dizileme ve total transkriptom mikrodizin teknikleriyle belirlenmiştir. Kontrol hücrelere göre kordoma tümörkürelerinde etkilenmiş yolakların belirlenmesi ve etkilerinin değerlendirilmedi için fonksiyonel zenginleştirme analizleri gerçekleştirilmiştir. Son olarak *SEMA5A*, *TPD52L1* ve *GAPLINC* genlerinin ifadeleri iki kordoma hücre hattında susturularak invazyon, göç ve canlılık üzerine etkileri incelenmiştir. Sonuçlar siRNA ile inhibe edilen *TPD52L1* ifadesinin kordoma hücre hatlarında hem göç hem de invazyonu engellediğini; *SEMA5A* ifadesinin baskılanmasının UM-Chor1 hücrelerinin hem göç hem de invazyon yeteneklerini engellediğini, MUG-Chor1 hücrelerinin ise yalnızca invazyon kapasitesini anlamlı şekilde düşürdüğünü göstermiştir. Ek olarak, *GAPLINC* ifadesinin susturulması ile UM-Chor1 hattında migrasyon ve invazyon anlamlı şekilde azaldığı, ancak MUG-Chor1 hücrelerinin etkilemediği tespit edilmiştir.

Bu çalışma, tümörküre kaynaklı kordoma kanser kök hücrelerinde genomik ve transkriptomik değişikliklerin incelenmesi ve *TPD52L1* ile *SEMA5A* genlerinin kordomaların göç ve invazyon kapasiteleri üzerindeki etkisinin gösterilmesi bakımından literatürde bir ilk teşkil etmektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

cDNA	Complementary Deoxyribonucleic Acid
circRNA	Circular Ribonucleic Acid
CSC	Cancer Stem Cell
DAPI	4′,6-diamidino-2-phenylindole
DPBS	Dulbecco's Phosphate-Buffered Saline
E-cad (*CDH1*)	Epithelial cadherin
EDTA	Ethylene Diamine Tetraacetic Acid
FACS	Fluorescence-activated cell sorting
FBS	Fetal Bovine Serum
GAPLINC	Gastric Adenocarcinoma Associated, Positive CD44 Regulator, Long Intergenic Non-Coding RNA
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
KLF4	Kruppel-like factor 4
lncRNA	Long Non-Coding Ribonucleic Acid
MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium
N-cad (*CDH2*)	Neural cadherin
Oct4 (*POU5F1*)	Octamer-binding transcription factor 4
PCR	Polymerase Chain Reaction
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SEMA5A	Semaphorin-5A
Slug (*SNAI2*)	Snail Family Transcriptional Repressor 2
Snail (*SNAI1*)	Snail Family Transcriptional Repressor 1
SOX2	SRY (sex determining region Y)-box 2
TPD52L1	Tumor Protein D52-Like 1
TWIST	Twist-related protein 1
VIM	Vimentin
ZEB1	Zinc Finger E-Box Binding Homeobox 1

1. INTRODUCTION

Cancer is a sophisticated concept to be considered as a single disease, which covers more than 200 diseases whose common denominator is "abnormal cellular growth".

The term “Cancer” covers more than 200 diseases whose common denominator is "abnormal cell growth" [1]. In 2020 alone, approximately 19 million new cancer cases and 10 million cancer-related deaths were recorded worldwide [2]. It is reported that one out of every five people in the world is diagnosed with cancer during their lifetime, and one in every 11 women and one in every eight men die from cancer. The total number of cancer patients who survive the first five years after diagnosis is estimated to be 50.6 million worldwide [3]. According to the 2019 report of the American National Institutes of Health (NIH), cancer treatment-related expenses in the USA alone are estimated to be more than $21 billion [4].

The stem-like properties of cancer stem cells (CSCs), such as self-renewal, proliferation, ability to maintain a dormant state, and differentiation capacity, enable them to play an active role in the mechanisms of cancer initiation, preservation of tumor heterogeneity, and metastatic spread. CSCs are promising therapeutic targets in current cancer treatment research [5–7]. In chordomas, where en bloc tumor resection with negative margins and radiotherapy are the recommended therapeutic modalities, it is not unusual to see locally recurrent tumors that are resistant to conventional chemotherapy [8,9]. Therefore, advances in targeted molecular therapy approach targeting chordoma cancer stem cells are of critical importance in the treatment of chordoma [10].

2. THEORETICAL BACKGROUND

Cancer is a disease of uncontrolled cell division that occurs as a result of the accumulation of mutations in healthy cells. Malignant features acquired during carcinogenesis confer cancer cells a high adaptation potential and survival capacity during tumor development and metastatic spread. Mutations and epigenetic variations have important roles in the pathogenesis and prognostic character of cancer.

2.1. CHORDOMA

Chordoma is a rare, slow-growing, malignant bone tumor that is thought to originate from embryonic notochord remnants and can occur along the spine, from the sacrum to the clivus. Chordomas are a subtype of sarcomas, a group of malignant bone and soft tissue tumors [11].

Chordoma of clivus was first introduced to the literature by German pathologist Rudolf Virchow in 1857. Virchow thought that the neoplasm originated from cartilage tissue, and due to many cytoplasmic vacuoles observed in chordoma cells he described the tumor as “ecchondrosis physaliphora spheno-occipitalis"; the word physaliphora was Greek for “bubble bearing”. and named these cells “clivus tumors” [12].

Based on the Surveillance, Epidemiology, and End Results (SEER) statistics, the incidence of chordoma in the USA is 1 per million and the 5-year survival rate is approximately 80.3% [13]. Although chordoma can be seen among all age groups and genders, it is usually seen between the ages of 40-70; the incidence of pediatric chordoma diagnosis is approximately 5%. The incidence rate among males and females was found to be 1.8:1 [14–17]

Chordomas are usually seen in the skull base and the sacrum, but can also occur in different parts of the axial skeleton. Although rare, chordoma formations have been reported in unusual areas from the wrist to the foot, in the extra-axial skeleton [18,19]. Chordoma cells which have a very slow division cycle, are highly resistant to chemotherapy [20]. Although radical approaches like en-bloc tumor resection and/or high-

dose radiotherapy give results for sacral chordomas, this practice is not entirely possible for clival chordomas due to the location and jelly-like lesions with the uneven texture of the tumor [21–24].

Due to their localization, the most common initial complaints of chordoma patients are pain and neurological symptoms. In the clinic, chordoma diagnosis is established via non-invasive imaging techniques and pathological evaluation methods based on immunohistochemistry. Histologically, chordomas can be classified into three classes: Classic chordoma, chondroid chordoma, and dedifferentiated chordoma. Chordomas are considered non-anaplastic tumors but increased anaplasticity, as seen in dedifferentiated chordomas, would lead to promoted aggressiveness [25,26].

Chordomas are positive for Cytokeratin 19 (CK19), Epithelial Membrane Antigen (EMA), Vimentin, and S-100; negative for CK7, CK20, and Chromogranin [27,28]. However, markers used in chordoma diagnostics may also be present in other cancers that have similar developmental/histological background to chordoma. For example, the S-100 protein is common in both chordomas and chondrosarcomas; but E-cadherin is positive in chordomas but not in chondrosarcomas.

Brachyury, a nuclear transcription factor expressed by TBXT (T) gene is the gold standard used for the diagnosis of chordoma. The T gene is located on chromosome 6q27 [29]. Brachyury is an important T box gene family member and has a major role in the development of notochord into axial skeleton and bone in embryonic stages. [23,30–33]. For this reason, the origin of the theory that chordomas arise from embryonic notochord remnants is due to the high brachyury expression in chordomas [34].

Chondroitin sulfate proteoglycan 4 (CSPG4) is another protein that is observed in approximately 72% of chordoma cases and is a potential marker for the diagnosis of the disease. Although due to the variable expression levels in different chordomas subtypes, CSPG4 protein is far from being reliable as a diagnostic criterion, it can be considered as a prognostic factor for the metastatic risk and life expectancy [35]

The first chordoma cell line U-CH1 is established in 2001 from a male patient with a sacral/recurrent chordoma tumor by the researchers from the University of Ulm [36]. There are 25 chordoma cell lines accepted by the American Chordoma Foundation [37]. Fifteen of these cell lines are established from the sacrum, nine of them are from the clival region,

and one from the thorax. Ten of those lines are developed from primary neoplasms, 4 from recurrent, 4 from metastatic, and 7 from recurrent and metastatic tumors (see Table 2. 1).

Table 2.1. Current list of chordoma cell lines and their origin and status features

Cell Line	Disease Origin	Disease Status	Age
U-CH1	Sacral	Recurrent	56
U-CH2	Sacral	Recurrent	72
U-CH6	Sacral	Recurrent + Metastatic	56
U-CH7	Sacral	Primary	67
U-CH11	Sacral	Primary	71
U-CH11R	Sacral	Recurrent	75
U-CH12	Sacral	Primary	83
U-CH17P	Sacral	Metastatic	38
U-CH17PII	Sacral	Metastatic	38
U-CH17M	Sacral	Metastatic	38
U-CH17S	Sacral	Metastatic	38
U-CH18	Thoracic	Recurrent + Metastatic	56
U-CH19	Sacral	Primary	77
CH22	Sacral	Recurrent + Metastatic	56
JHC7	Sacral	Primary	61
MUG-Chor1	Sacral	Recurrent	57
MUG-CC1	Clival	Primary	72
U-CH14	Clival	Primary	79
UM-Chor1	Clival	Primary	66
UM-Chor5	Clival	Primary	<20
UM-Chor5C	Clival	Recurrent + Metastatic	<20
UM-Chor5D	Clival	Recurrent + Metastatic	<20
UM-Chor6	Clival	Primary	32
U-CHCF359B	Clival	Recurrent + Metastatic	<20
U-CHCF365	Clival	Recurrent + Metastatic	<20

2.2. GENETIC ALTERATIONS IN CHORDOMA

Mutations that occur during carcinogenesis can result in loss of heterozygosity, leading inactivation of a tumor suppressor gene [38]. The first mutation associated with human cancer is the “Philadelphia Chromosome”, which is a reciprocal translocation between the q arms of chromosome 9 and 22 resulting in uncontrolled division of cells in chronic myeloid leukemia (CML) [39]. A study by Friend and colleagues in 1986, revealed that the deletions in the Rb1 gene, which will be later known as a tumor suppressor gene, are the cause of retinoblastoma [40]. BRCA1 (Breast Cancer Associated 1) gene, whose mutations are identified to cause breast cancer in 1994, has been described as the “gate-keeper” due to its function in the DNA repair mechanism and consequential role in maintaining genomic integrity [41,42]. Similarly, mutations in the Kirsten rat sarcoma viral oncogene homolog (KRAS) gene, which are found in approximately 35-45% of colorectal cancer and 15-30% of lung cancer cases, are also used as a diagnostic marker and therapy response predictor in the clinic. Anti-epidermal growth factor receptor (EGFR) therapy in the clinic to suppress the EGF signal cascade that triggers Ras protein leading phosphorylation of ERK which induces cell cycle and proliferation. However, abnormal –K-Ras protein synthesized by mutated KRAS gene does not require upstream tyrosine kinases initiated by EGFR activation and is unresponsive to anti-EGFR therapy [43,44].

As in chordoma, all cancers are the results of abnormal changes in DNA [45]. The genetic abnormalities leading to chordoma may arise spontaneously, more rarely, be inherited. In most cases, both inherited and sporadic chordomas have been linked to mutations in the T gene. Studies showed that germ-line tandem duplications in the TBXT gene were observed in some familial chordomas, and >20% of sporadic chordomas also acquire an extra copy of the *T* gene during the development of the neoplasm [46]. Comparative genomic hybridization array (array-CgH) studies on patient samples showed several gain mutations on q arms of chromosomes 7, 12, 17, 20, and 22, and copy number losses in 1p, 3, 4, 9, 10, 13, 14, and 18 in sporadic chordomas [36,47,48]. Mutations in the T gene have a dominant negative effect. 97% of familial chordoma patients carry rs2305089 single nucleotide polymorphism (SNP) in at least one allele. Therewithal, SNP rs2305089 in sporadic chordomas is also effective on overall survival [49–51].

EGFR, a receptor tyrosine kinase-expressing gene located on chromosome 7p12 is known to be strongly activated in chordoma cases. Chromosome 7 gains commonly seen in chordomas, or SNP rs1050171 in exon 20 which causes a silent mutation of the tyrosine kinase domain of EGFR, are found effective on this strong EGFR activation [52,53]. PI3K/Akt/mTOR is a known pathway that is activated in the majority of chordomas. A study on chordoma tumor samples showed the presence of activating mutations in PI3KCA and PIK3R1 and PTEN in 16% of the chordoma samples [46]. Another study reveals one copy loss of the chromosome 10q23.31 region which encodes the PTEN gene [47]. 17% of chordoma patients have mutations in SWI/SNF genes, which have a role in chromatin remodeling [46].

2.2.1. Chordoma Treatment

Although there is no treatment approved by the American Food and Drug Administration (FDA) for Chordoma, various clinical studies are in progress at different phases. GI-6301, an *S. cerevisiae*-based cancer vaccine, is in the Phase II stage, which aims to stimulate the activation of T cells in humans by synthesized brachyury protein [54,55]. In another cancer vaccine study, about to be carried on to phase Ib/II, The NANT, contains a combination therapy of Aldoxorubicin Hydrochloride HCI, ALT-803, ETBX-051, ETBX-061, GI-6301, haNK, avelumab, cetuximab, cyclophosphamide and radiotherapy applications [56]. The study includes ETBX-01, a recombinant adenoviral-based brachyury vaccine to be injected under the skin. It is planned to stimulate cytotoxic T lymphocytes with brachyury protein to be synthesized by adenoviruses, that replication incompetent [57]. Similarly, the MVA Ch BN® Brachyury vaccine, another study based on the principle of immune response to the brachyury protein, is being prepared for phase II studies [58]. In a study published by Magnaghi and his team in 2018, the EGFR inhibitor Afatinib in the phase II stage was shown to exhibit an antiproliferative effect on U-CH1 and UM-Chor1 chordoma cell lines by inducing down-regulation of TBXT [59].

2.3. CANCER STEM CELLS

Pluripotent stem cells are undifferentiated cells that have the ability to self-renew, proliferate, and give rise to various cell subtypes during embryogenesis. Depending on the developmental stage of the organism, pluripotency and abundance of the stem cells in that organism decrease correspondingly [60]. Decreased number of stem cells in an adult body, impairs the regeneration capability of the tissues, but their long-term repopulating ability (quality) remained constant [61].

Mutations can cause a series of events leading to alterations in the biochemical pathways of the cell. These alterations could lead to uncontrolled cell divisions, and make cancer cells evade apoptosis and immune response which is called malign progression [62]. It is known that cancer cells may acquire a stem cell-like character, which ensures the continuity of cellular heterogeneity in a cancer cell population, gives metastatic character, the ability to resist chemo and radiotherapy, and may lead recurrence of cancer [63]. During malign progression, the control of differentiation mechanisms in the cell is inevitably lost [64]. De-differentiation of differentiated dormant cells into an oncogenic hyperproliferative state is considered one of the possible origin stories of cancer stem cells [65,66].

Another advantage that the disruption of differentiation mechanisms brings to cancer stem cells is the ability of cancer stem cells to remain in the dormant phase that is insensitive to physiological interventions. Hence, the cancer stem cells, which manage the differentiation process in its favor, can re-enter the proliferative phase after a “successful” chemotherapy and may lead to recurrence at the initial region or formation of metastatic cancer [67].

The theory of cancer stem cells was first introduced in acute myeloid leukemia (AML) patients by Bonnet and Dick who found that the $CD34^+$ and $CD38^-$ subpopulations of AML cells were leukemia stem cells [68]. Many subsequent studies have shown that there are specific surface markers, intracellular proteins, and non-coding RNAs for different cancers (Table 2.2) [69]. In a study by Aydemir et al. in 2012, it was shown that subpopulations with CD133 and CD15 surface markers in the U-CH1 chordoma cell line have cancer stem cell-like characteristics [70].

Table 2.2**.** CSC markers for various tumor types (derived from Zhou et al., 2021, [69])

Classification	**Markers**	**Function and Role**
Isolation markers	CD133	A common CSC marker in various cancers. CD133+Nestin+ is better for CSCs in glioma. CD133+CD44+ is better for CSCs in HCC and colon tumor CD133+ALDH+/ CD133+L1CAM+ is better for CSCs in OC.
	CD44	A common CSC marker in various cancers. CD44+CD133+/CD44+EpCAM+/CD44+ALDH+ is better for CSCs in CRC. CD44+ c-Met+ is better for CSCs in pancreatic cancer; CD44+ALDH+/ CD44v8-10 is better for CSCs in GC.
	EpCAM	A common CSC marker in various cancers. EpCAM+CD44+ CD166+ is better for CSCs in CRC. EpCAM+ CD133+ is better for CSCs in HCC
	ALDH	A common CSC marker in various cancers. EpCAM+CD44+ CD166+ is better for CSCs in CRC. EpCAM+ CD133+ is better for CSCs in HCC
	CD90	A common CSC marker in various cancers. CD90+CD44+ is better for CSCs in lung cancer cells.
Intracellular markers	Oct4	Oct4 is a homeodomain transcription factor by binding to octamers, and regulates the expression of many genes
	Nanog	Nanog is a homeobox transcription factor, And plays a crucial role in the second embryonic cell-fate specification.
	Sox2	Sox2 has an important function in the early development and maintenance of undifferentiated ESCs.
SP fraction	Hoechst	SP cells can be separated by fluorescence

	33342-	screening after the outflow of Hoechst 33342.
	negative population	And SP cells have high homology, self-renewal, and multidirectional differentiation potential.
Noncoding RNAs	Circ008913	Regulate CSC phenotype in nasopharyngeal carcinoma cell line.
	CirGprc5a	Regulate CSC in bladder tumors.
	Circ001680	Promote CSC in CRC and induce irinotecan resistance.
	CircLgr4	Regulate CSC in CRC.
	lncTCF7	Promote CSC by Wnt signaling pathway.
	lnc-β-Catm	Promote CSC together with Wnt.
	H19	Regulate CSC in breast cancer and HCC.

In addition to cancer-specific surface receptors, expression of pluripotency-related oncogenes such as Oct4, Klf4, c-Myc, and Sox2 are increased in CSCs, similar to stem cells [65]. Otherwise, it has been suggested that cancer stem cells provide an invasive and metastatic profile of cancer by driving the epithelial-mesenchymal transition (EMT) status in cancer [71]. Downregulation of E-cadherin and up-regulation of N-cadherin, fibronectin, and matrix metalloproteases leads to an EMT status, which is related to the cancer stem cell phenotype [72]. Additionally, increased expression of Aldehyde Dehydrogenase (ALDH) is another marker for CSCs. Rinner showed increased ALDH expression in the Mug-Chor1 chordoma cell line [73]. Also, it is known that ATB binding cassette (ABC) transporter proteins are highly expressed in CSCs and provide the abilities for drug-resistance which can be targeted in cancer therapy [74].

Numerous studies target cancer stem cells suggesting the dynamic heterogeneity and therapy resistance to cancer. For example, BMS-906024, which has been used as a Notch signaling pathway inhibitor in studies conducted to date, is involved in Phase 1 clinical trials targeting cancer stem cells as well [75]. Phase 3 trials of the inhibitor of the carbonic anhydrase enzyme, which alkalizes the cancer microenvironment, are effective in invasive cancers [76]. It is known that the expression of ABC transporters that bind to ATP is elevated in cancer stem cells [77]. Therefore, clinical trials of ABC transporter inhibitors are ongoing at various stages in the development of therapeutic approaches targeting CSCs [78].

2.3.1. Cancer stem cell enrichment methods

Tumor-sphere cultures and “side population” isolation are the most frequently used methods in studies to detect and enrich cells with stem cell potential *in vitro* [79,80].

Tumorsphere formation assay is an *in vitro* method to enrich CSCs populations based on the ability of cancer stem cells to self-renew and proliferate in serum-free and attachment-independent conditions. The assay also is a highly preferred model for studying molecular mechanisms and treatment approaches targeting cancer stem cells and testing the efficacy of treatments [81,82]. The induction of apoptosis in adherent cells due to the integrin disengagement from the extracellular matrix is called anoikis. Anoikis is a critical control mechanism of the cell cycle that ensures the elimination of detached cells from healthy tissue and an obstacle that cancer cells must overcome to gain metastatic skills [83]. Studies have revealed that cancer stem cells create a niche and protect the non-stem tumor cells from anoikis-induced cell death in breast cancer [84]. In tumorsphere cultures, a hydrogel-coated ultra-low attachment surface that minimizes protein absorption and cellular activation is used. Also, tumorspheres with cancer stem cell characteristics can be formed when cells are prevented from adhering to the surface via an agarose-coated culture plate in serum-free culture conditions [85].

The side population (SP) technique, which is known to be used for screening the efficacy of therapeutic agents on cancer stem cells, was first developed for the detection of hematopoietic stem cells in mouse bone marrow [86]. The technique is based on the cell's ability to efflux the fluorescent DNA-binding dye Hoechst 33342 via ABCG2 protein, one of the ATP-binding cassette (ABC) transporters, which has also been associated with multidrug resistance mechanisms [87]. The side population cells can be identified and isolated from the main population by their characteristic appearance in flow cytometry analysis. A control group was used to detect the side population that cannot efflux Hoechst 33342 dye which is obtained by inhibiting the ABC transporter proteins with Verapamil before the Hoechst staining. Cells treated with verapamil are identified as “Hoechst 33342 positive” in flow cytometry data. In the group not treated with verapamil, cells in the Hoechst 33342 negative quadrant (appeared) are considered as the side population.

Within the scope of the project, the aforementioned tumorsphere culture and side population stem cell enrichment techniques were used to obtain stem cell-enriched populations in chordoma cells. Details of both methods are given in the Materials and Methods section (3.2).

3. METHODOLOGY

3.1. CELL CULTURE MAINTENANCE

The experiments were carried out using four chordoma cell lines: UM-Chor1 (Primary-Clival), MUG-Chor1(Primary-Sacral), U-CH1 (Recurrent-Sacral), and CH22 (Recurrent/metastatic, Sacral). Chordoma cell lines were maintained at 37ºC with 5% CO_2 culture conditions, in Iscove's Modified Dulbecco's Medium (IMDM) and Roswell Park Memorial Institute (RPMI) 1640 Medium as a combination of a 4 to 1 ratio, which was supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin (PS) antibiotics. The culture flasks used in chordoma culture were covered with 0.1% gelatin. Maintenance of the chordoma cell culture requires a medium change every two days, and the cells were passaged when confluency reached 80%, by using a %0.25 Trypsin-EDTA solution.

3.2. CANCER STEM CELL ENRICHMENT APPROACHES

There are several cancer stem cell enrichment methods in the literature. In this study, side population selection via Hoechst staining, isolation of $CD15^+/CD133^+$ expressing chordoma cells via FACS, and tumorsphere formation methods were used.

3.2.1. Isolation of $CD133^+$/ $CD15^+$ Population via FACS

Previous studies conducted by our group showed for the first time in literature that, $CD133^+$ and $CD15^+$ subpopulations in chordoma exhibit a cancer stem-like character [70]. Based on these findings, $CD15^+/CD133^+$ cell populations in chordoma cell lines were isolated via Fluorescent Assisted Cell Sorting (FACS) protocol by using BD FACS Aria III device.

Immuno-Fluorescent Staining for FACS

1- Cells were trypsinized with 0.25% trypsin-EDTA and the pellet was washed with 5mL Ca^{++} and Mg^{++} free phosphate buffered saline (DPBS).

2- The washed cell pellet was resuspended in ~100uL DPBS, and 5µl human FCR blocker was added to the cell suspension for approximately $1x10^6$ cells and incubated on +4ºC ice.
3- Florescent-conjugated CD133 and CD15 antibodies were added to the cells and incubated on ice for 30 min in the dark.
4- Cells were washed with DPBS twice and resuspended in 500µL PBS with 4',6-diamidino-2-phenylindole (DAPI).
5- The same protocol will be repeated for the unstained (negative control) group w/o antibody addition but with the DAPI step. Since DAPI dye is effluxed in living cells immediately, DAPI negative population in unstained samples was used to determine the gate for living cells in FACS analysis.
6- $CD133^+$ $CD15^+$ double-positive cell populations were selected and sorted. Collected cells were kept at -80ºC for further RNA isolation.

3.2.2. Side Population Selection via Hoechst staining

The Hoechst side population method is based on the selection of cell populations with high efflux capacity to DNA-binding dye Hoechst 33342 through ABC transporters which are enriched in stem cells. Based on the literature, the control of the side population was evaluated via verapamil pre-treatment before the Hoechst staining, which is an ABC transporter inhibitor. Side population cells observed in conditions without the verapamil pre-treatment should have disappeared in the “Hoechst negative quadrant” in verapamil pre-treated groups.

Verapamil treatment and Hoechst staining

1. Detached chordoma cells were washed with DPBS, and the supernatant was removed.
2. For verapamil pre-treated control groups:
   - The cell pellet was dissolved in cold culture medium as 1.10^6 cells/mL. Verapamil was added to the control cells with a final concentration of 50uM.
   - Cells were incubated at dark for 30 minutes in a 37°C water bath. Cells were mixed by gentle finger vortexing in every 15 minutes.
3. Cells were stained with 10uL/mL Hoechst 33342 in PBS and incubated for 90 min at 37ºC. The tubes were mixed by gentle finger vortexing in every 15 minutes.

4. After the incubation, cells were washed with cold DPBS, the supernatant was removed and the pellet was dissolved in 2% FBS containing DPBS.
5. 7AAD stain was used to detect dead cells and only living cells were gated in FACS.
6. Side populations were detected under DAPI A vs. DAPI B channels and sorted. Collected cells were kept at -80ºC for further RNA isolation.
    - The cell pellet was dissolved in cold culture medium as 1.10^6 cells/mL. Verapamil was added to the control cells with a final concentration of 50uM.
    - Cells were incubated at dark for 30 minutes in a 37°C water bath. Cells were mixed by gentle finger vortexing in every 15 minutes.

3.2.3. Tumorsphere Formation

In this assay, higher survival capability in non-adherent and serum-free conditions of stem cells is used as a cancer stem cell enrichment method [81]. The tumorsphere formation was performed as follows, detached cells were dissolved in a defined sphere medium (see table 3. 1) and seeded on an ultra-low attachment surface plate.

Table 3.1**.** Tumorsphere medium content

Material	Final concentration
Antibiotics (PS, 100X)	1X
EGF (100 µg/ml)	20 ng/ml
bFGF(100 µg/ml)	20 ng/ml
N2 (100X)	1X
B27 (50X)	1X

The culture medium was changed once a week. For the medium change process, medium containing tumorspheres were collected from wells, and let sit for 30 minutes at 37ºC culture incubator, the supernatant was removed carefully, and settled tumorspheres were gently dissolved in 2 mL fresh sphere medium. If any clump formation occurred, clumpings were dissolved by gentle pipetting. Three days after each medium change, 1mL fresh sphere medium was added to the wells.

For sub-culturing, sphere-forming cells were dispersed and seeded again to form spheres. TrypLE^{TM} express reagent, a recombinant cell-dissociation enzyme that cleaves bonds on

the C-terminal residues of Lysine and Arginine amino acids, was used as a substitute for Trypsin. Similar to the regular trypsinization procedure, the sphere-bearing culture medium is collected and incubated for approximately 30 minutes at 37°C in a CO_2 incubator. The supernatant was removed carefully and 2-5 mL of TrypLE was added to the sphere pellet and incubated for ~10 minutes. Pellet was finger vortexed every 2 minutes and repeated until the spheres become invisible. To collect sphere cells, the cell suspension was centrifuged for 5 min at 10000rpm. Pellet was re-dissolved in a fresh sphere medium and seeded onto an ultra-low attachment culture plate.

The sub-culturing protocol was applied for all 4 chordoma cell lines U-CH1, MUG-Chor1, UM-Chor1, and CH22. Cells were seeded as 100K cell/well for U-CH1, MUG-Chor1, and UM-Chor1 lines and 50K/well for CH22 cell line. Observations during the experiments showed that the number of viable cells of U-CH1 and UM-Chor1 cell lines reduced dramatically and could not form spheres after the second passaging. Therefore, for U-CH1 and UM-Chor1 cell lines, only the 2^{nd} generation spheres (passaged 1 time) of these two lines were used for further experiments. On the other hand, MUG-Chor1 and CH22 tumorosphere-derived cells successfully continued to form spheres during the sub-culturing; so these lines were passaged 3 times and the 4^{th} generation tumorspheres were used for these cell lines.

3.3. DETERMINATION OF TRANSCRIPTOMIC CHANGES IN CHORDOMAS

3.3.1. Total RNA isolation and Quantification

Total RNAs from chordoma tumorspheres were isolated via PureLink™ RNA Mini Kit (Thermo Fisher Scientific, cat no: 12183020) based on the manufacturer's protocol. Isolated RNA samples were used for Clariom™ D Total Transcriptome microarray analysis to determine both coding and (long) non-coding transcriptome and the expression confirmation qPCR. For gene expression alteration studies, total RNA samples were isolated with easy-BLUE™ Total RNA Extraction Kit (iNtRON Biotechnology, cat no: 17061) according to the manufacturer's instructions. RNA grade Glycogen (Thermo Scientific™, cat no: FERR0551) was used for a final concentration of 0.1µg/µL as an RNA carrier. RNA concentrations and purity were determined via UV spectrometry

(Epoch Microplate Spectrophotometer) and calculated via BioTek Gen5 Data Analysis Software based on OD_{260}:OD_{280} ratio.

3.3.2. cDNA synthesis and Real-Time qPCR

cDNA synthesis was performed using the "High-Capacity cDNA Reverse Transcription Kit" (Applied Biosystems®, Massachusetts, USA; Cat. no: 4368814). Gene expression levels were measured via quantitative Real-Time Polymerase Chain Reaction (RT-qPCR) technique, by using gene-specific TaqMan™ primers (table 3.2), and "TaqMan™ Universal Master Mix II with UNG" (Applied Biosystems®, Massachusetts, USA; Cat. no: 4440042) and StepOne™ Real-Time PCR Systems (Applied Biosystems®, Massachusetts, USA). The relative expression levels between control and experimental groups (tumorspheres vs parental cell lines and silenced samples vs control siRNA transfected samples) for the selected genes were calculated based on the $2^{-\Delta\Delta CT}$ formula, and GAPDH was used as housekeeping.

Table 3.2. Taqman catalog numbers of the selected genes in the study

Gene	**TaqMan™ Gene Expression Assay**
OCT4	hs04260367_gh
SOX2	hs00415716_m1
NANOG	hs04260366_g1
KLF4	hs00358836_m1
CDH1	hs01023894_m1
CDH2	hs00983056_m1
VIM	hs00185584_m1
SLUG	hs00950344_m1
SNAIL	hs00195591_m1
TWIST	hs01675818_s1
ZEB1	hs01566408_m1
GAPLINC	hs00418805_m1
SEMA5A	hs01549381_m1
TPD52L1	hs00914766_m1

3.3.3. Clariom D Microarray

Clariom D total transcriptome microarray enables expression analysis of more than 540.000 coding and (long) non-coding sequences with alternative splicing events by maximizing exon coverage. The assay was performed based on the manufacturer's guidance with isolated total RNA samples, by using Clariom D human assay (Thermofisher Cat no: 902922) via Affymetrix GeneChip™ 3000 microarray system. The microarray experiments were repeated 3 times in 4 different parental chordoma cell lines and the tumorspheres were derived from the parental cell lines. The basic scheme for transcriptome microarray was presented in Figure 3.1.

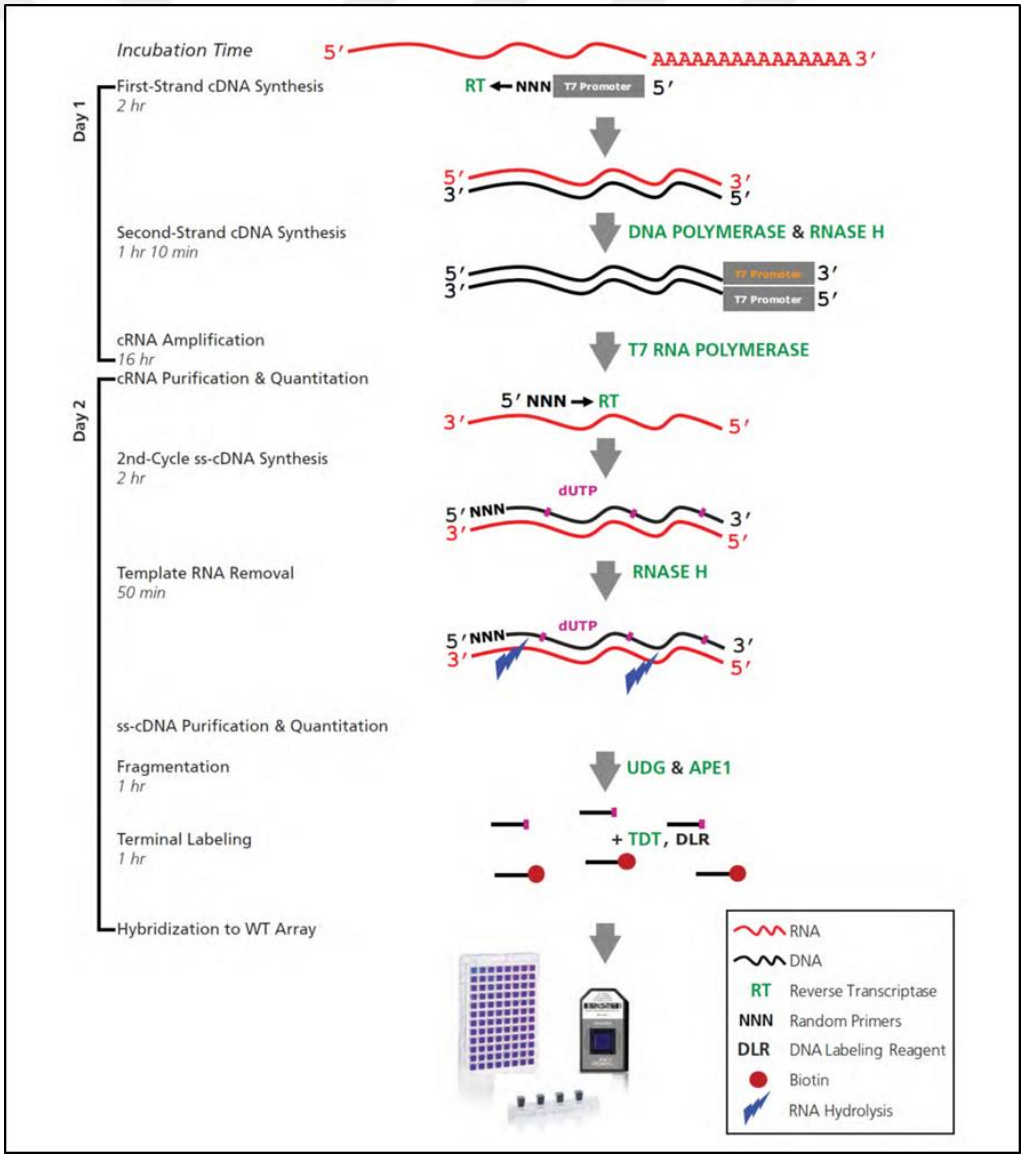

Figure 3.1. Clariom D amplification and labeling protocol scheme.

3.3.4. Differential Expression and Functional Annotation Analysis

Affymetrix® Transcriptome Analysis Console (TAC) software version 4.0.2 (Applied Biosystems) was used for the differential expression analysis among the groups. Raw expression data were background corrected, normalized, and summarized by Robust Multichip Average (RMA) method; differentially expressed genes were analyzed using *Limma* Bioconductor package, based on an empirical Bayesian (eBayes) approach via TAC software [88–90]. Both gene and exon level p values smaller than 0.05 and absolute fold changes greater than 2 are considered significant. The PCA map and hierarchical clustering dendrograms are retrieved from the TAC interface. The Interactivenn tool was used to create Venn diagrams describing common and significant DEGs among cell line groups [91].

All functional enrichment analysis outputs were visualized via R (version 4.2.2) [92]. The DEGs that are significantly altered in tumorspheres in comparison to parental cell lines were used for further functional enrichment analysis. The enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) terms were comparatively analyzed, annotated, and visualized via *clusterProfiler*, *org.Hs.eg.db*, *enrichplot*, and *ggplot2* R packages [93–100]. Enrichment analysis results based on WikiPathways curation were accessed via TAC software and visualized using *ggplot2* and *viridis* packages [101,102]. Gene set enrichment analysis of GO, and KEGG pathways were calculated based on ENTREZ IDs, and MSigDB datasets were based on official gene symbols of DEGs. The “core analysis” provided by Ingenuity Pathway Analysis (IPA, Qiagen) was used for the identification of “analysis ready molecules” commonly dysregulated in chordoma cell line-derived tumorspheres for the assignation of genes that will be used for further functional analysis [103].

3.3.5. Generating a Reference Gene List Using GEO Datasets

A reference differential gene expression data for “chordoma tissue versus healthy control” was prepared for further analysis of DEGs obtained from microarray results of “tumorsphere versus parental cell lines”. Datasets with Affymetrix HG-U133_Plus_2 microarray transcriptome chip (GPL570) were retrieved from the Gene Expression

Omnibus GEO database. Raw expression data of 28 Nucleus Pulposus and 4 chordoma samples in .CEL file format were preprocessed by using MAS5 algorithm, annotated to genome version hg19 using HG-U133_Plus_2.na36.annot.csv annotation file via TAC software [104,105]. Statistical comparison was conducted by using Anova (eBayes), values with $p<0.05$ and absolute fold change >2 are considered statistically significant [90].

Table 3.3. GEO accession numbers of nucleus pulposus and chordoma samples

Sample type	Dataset ID	GEO accession number
Chordoma	GSE94321	GSM2473096
		GSM2473098
		GSM2473100
	GSE108088	GSM2889389
Nucleus pulposus	GSE70362	GSM1725781
		GSM1725783
		GSM1725785
		GSM1725787
		GSM1725789
		GSM1725791
		GSM1725793
		GSM1725795
		GSM1725797
		GSM1725799
		GSM1725801
		GSM1725803
		GSM1725805
		GSM1725807
		GSM1725809
		GSM1725811
		GSM1725812
		GSM1725814
		GSM1725816
		GSM1725818
		GSM1725820
		GSM1725822
		GSM1725824
		GSM1725826
	GSE147383	GSM4429808
		GSM4429810
		GSM4429812
		GSM4429814

The differentially expressed genes list retrieved from the chordoma versus nucleus pulposus sample comparison was used as a reference to check if the differential transcriptome of tumorsphere samples is also compatible with these tissue-based results.

3.4. DETERMINATION OF GENOMIC ALTERATIONS IN CHORDOMA TUMORSPHERES

3.4.1. Genomic DNA Isolation

Genomic DNA is isolated from biological replicates of each chordoma cell line using QIAamp DNA Mini Kit (Cat. no:51304) according to the instructions provided by the manufacturer. DNA concentrations for each genomic DNA sample are determined using a Qubit™ dsDNA HS assay kit (cat no: Q32854) and Qubit™ 4 fluorometer. For library preparation of the OncoPlus DNA-seq panel., 500ng/µL genomic DNA is used.

3.4.2. OncoPlus Next-Generation DNA Sequencing

OncoPlus DNA-seq Research panel v2.0 (Invitrotek cat no: 1001112E) was used to detect genomic variations that occurred during chordoma tumorsphere formation and cancer stem cell enrichment. The kit, which is an oncology-specific panel developed for solid tumors, is a comprehensive next-generation DNA-seq panel containing 565 genes and 14 microsatellite regions frequently associated with cancer in the literature. Panel coverage details are presented in the appendices as tables A.1., A.2., and A.3.

The MGI DNBSEQ next-generation sequencing protocol consists of fragmentation, size selection, end repair, A-tailing, adapter ligation, and PCR, circularization by using splint-oligo library amplification, DNA nanoball (DNB) formation, loading the DNBs into flowcell, and sequencing. The workflow is summarized in figure 3.2.

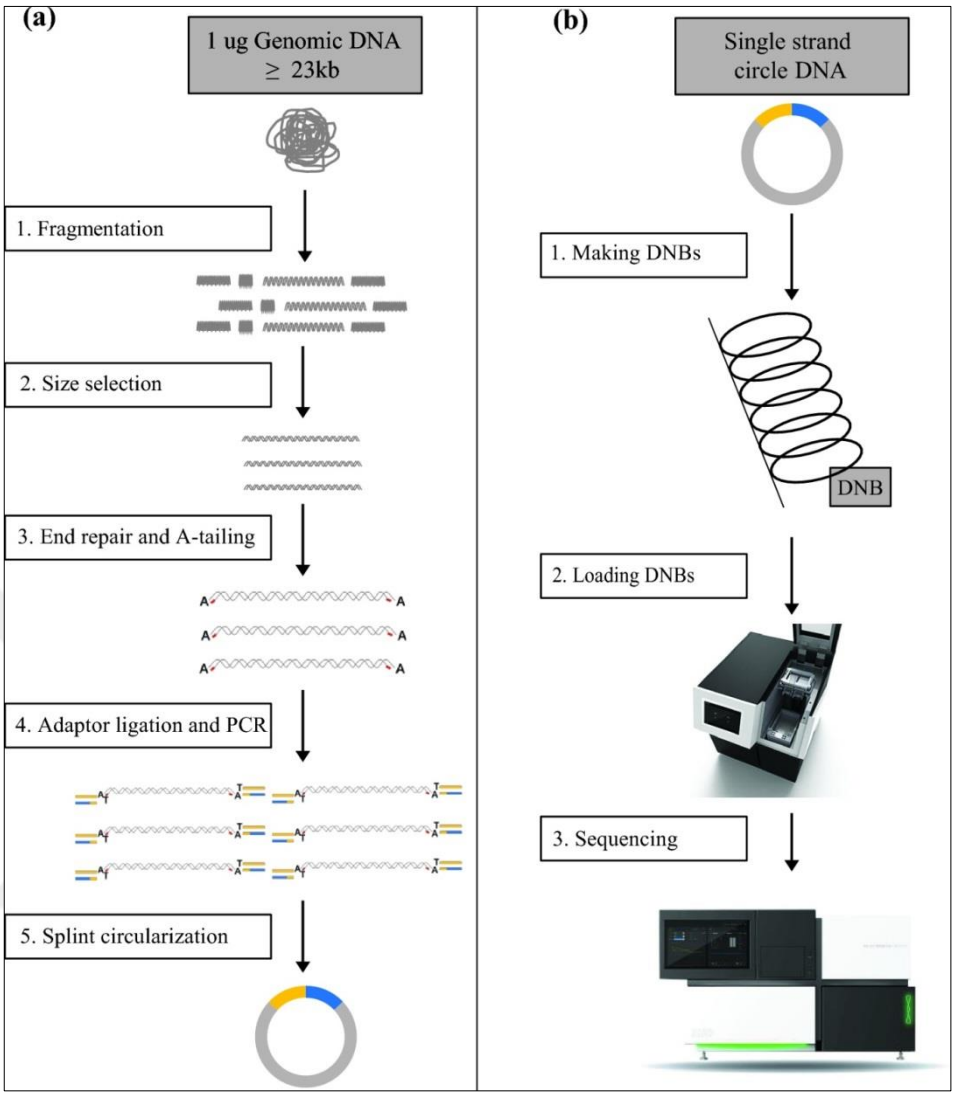

Figure 3.2. OncoPlus DNB-sequencing workflow

3.4.3. DNA-seq variant analysis pipeline

The quality controls of raw .fastq files are analyzed with the program FASTQC to confirm 100bp paired-end reads (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were filtered and trimmed based on the quality parameters of a minimum 80 base length, window width of 4, and minimum quality of 20, along with the removal of adapter sequences via the *Trimmomatic* tool [106].

The trimmed reads were mapped to GRCh37 *Homo sapiens* genome to create .SAM files using HISAT2 tool [107]. SAMtools library was used to convert alignments into BAM

format, and then the mapped reads were sorted and indexed. For variant calling, the GATK HaplotypeCaller or FreeBayes algorithms were used, and the results were saved in .vcf file format [108,109]. The VCF files of each sample were filtered out by excluding the variants covered by fewer than a read depth of 50 with BCFtools to focus on the genome coordinates covered by the OncoPlus panel [108]. The Ensembl Variant Effect Predictor (VEP) tool was used to annotate variants on GRCh37, which was then converted into a .maf file format using the *vcf2maf* tool[110]. *MAFTools* R package is used to merge .maf files into a single file and for the visualization of somatic variants [111,112]. The variant summary graphs obtained with *MAFTools* take the individual .maf files of each sample combined into a single file as input. The plots acquired represent the statistical outcomes of the genes with the highest number of variants and the most frequent variants among all the samples represented in the input file. Additionally, the top 20 frequently mutated genes in the samples were also represented as oncoplots, which display the variant classes in a heatmap format.

3.5. FUNCTIONAL EXPERIMENTS

3.5.1. Gene silencing by using siRNA transfection

Up-regulated genes were silenced via siRNA treatment. *SEMA5A*, *TPD52L1*, *GAPLINC*, control siRNA sequences, and X-tremeGENE™ siRNA Transfection Reagent (Sigma, Cat No:4476093001) were purchased commercially. UM-Chor1 and MUG-Chor1 cells were seeded 60k/well in 6 well plates 24 hours before the experiment. For starvation, the cells were washed with serum-free medium and incubated in 2ml of serum-free medium for 2 hours before the transfection. *SEMA5A*, *TPD52L1*, *GAPLINC* control siRNAs (25pmol, ~0,33ug), and X-treme reagent were separately diluted in 100uL of OptiMEM and incubated for 5 min at room temperature. Then, each siRNA group is mixed with X-tremeGENE reagent to a final ratio of 5vol:1mass (reagent: siRNA) and incubated for 20 min at RT. The mixture was added onto the cells dropwise by gently swirling the plate and incubating at 37C. Six to eight hours later, the medium on the cells was replaced with an antibiotic-free fresh culture medium containing 10% FBS. Cell pellets were collected at the 24^{th}, 48^{th}, and 72^{nd} hours.

Table 3.4.. siRNAs used in the study

Gene	Ambion™ siRNA	Cat. No.
GAPLINC	n550868	4392421
SEMA5A	107284	AM16704
TPD52L1	s14328	4392421
Control siRNA	Negative Control #1	AM4635

3.5.2. Migration and Invasion Assays

Migration and invasion capacities of tumorsphere derived chordoma CSCs and in the later phases the cells in which selective gene expressions were altered were evaluated via Boyden chamber assay with hanging transwell inserts. siRNA-transfected cells were used at the 48th hour of the transfection. The inserts that were used were Millicell® cell culture inserts for 24 well plates, with 8.0um pore diameter (Merck, cat no: MCEP24H48). Due to decreased diffusion rate, the inner cells of a sphere are exposed to more hypoxic conditions, which may cause cell death. Likewise, alterations in gene expression might also cause cell death. Therefore, single-cell suspensions were stained with trypan blue, and living cells were taken into account while migration and invasion capabilities are assayed.

1- Considering the cell size of different chordoma cell lines 40.000 to 80.000 cell/insert were suspended in 200uL serum-free chordoma culture medium (4:1 IMDM:RPMI, with 1% PS, without FBS) and seeded into the transwells.
2- The transwells were placed in a well-containing chordoma culture medium with 20%FBS which the FBS gradient was used as a chemoattractant driving factor for migration. Cells were incubated for 24-48 hours at 37ºC in regular culture conditions.
3- The cells that passed through the pores of the transwells from a serum-deprived condition to a high serum condition are fixed and stained via crystal violet staining protocol and counted.

For the invasion assay protocol, in addition to the migration assay steps, a matrigel layer is covered into the transwell inserts to mimic invasion conditions *in vitro.* As a preparation, 100uL of 300 µg/mL "Basement Membrane Matrigel" (BD, Kat No:356234) was added inside of the transwell and transferred on an empty 24well plate well, and incubated overnight at regular culture conditions.

3.5.2.1. Transwell Crystal Violet Staining Protocol

1. Medium and cells that remained in transwell inserts were discarded, and inserts were washed with DPBS, twice.
2. The attached cells on both the inner and other sides of the inserts were fixed via a 3.7% formalin solution for 4 minutes.
3. Transwells were washed with DPBS twice to remove the remaining formaldehyde.
4. 1mL absolute methanol was added to the transwell and was fixed for 20 minutes at RT.
5. Transwells were washed with PBS twice, again, to remove methanol.
6. %0.1 w/v crystal violet solution was added inside and outside of the transwell and stained for 2 minutes at RT in the dark.
7. The inner part of the insert was gently swiped with a cotton swab to remove cells that cannot achieve to migrate through pores.
8. Migrated/Invased cells were counted by using ImageJ [113] software. Relative migration and invasion capacities were evaluated by comparing the experiment groups with controls.

3.5.3. MTS Cell Viability Assay

The 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) assay was used for the evaluation of cell viability and pluripotency rates of siRNA transfected cells and control cells, which based on reduction of MTS compound into colored formazan crystals that are measurable via photospectrometric devices.

The cells were seeded on a 96well plate as 1000-5000 cells/well (1000 for CH22 and UM-Chor1, 5000 for MUG-Chor1), 24 hours before the siRNA transfection. The same starvation conditions and transfection reagent:siRNA ratios stated above for the 6-well

plate were adapted to 96well plate conditions. The assay was repeated at the 24^{th}, $48^{th,}$ and 72^{nd} hour time points after siRNA transfection. Medium on cells was replenished with MTS reagent (20% in complete culture medium) and incubated for 50 minutes until metabolized formazan crystals create a noticeable color change. The OD values were measured with a UV spectrometer (Epoch Microplate Spectrophotometer) and DeltaOD (OD_{490}-OD_{650}) values were calculated via BioTek Gen5 Data Analysis Software.

3.6. STATISTICAL ANALYSIS

The statistical evaluation of relative gene expression changes (2^-ΔΔCt) of RT-qPCR data was conducted via unpaired *t*-test based on the -ΔΔCt values of the genes of interest. *p* values less than 0.05 was considered statistically significant. Statistical data analyses of RT-qPCR data, migration, invasion and MTS assays were performed using GraphPad Prism version 9.4.1 (San Diego, California USA).

4. RESULTS

4.1. COMPARISON OF DIFFERENT CANCER STEM CELL ENRICHMENT TECHNIQUES ON CHORDOMAS

Here, outcomes of three frequently used cancer stem cell enrichment techniques utilizing stemness and EMT-related gene expression changes are presented.

4.1.1. Isolation and determination of the changes in stemness and EMT-related gene expressions of CD15^{+}/CD133^{+} positive chordoma cells

CD133+/CD15+ dual positive populations, which are known CSC markers for U-CH1 cell line, are detected and isolated within 4 chordoma cell lines via FACS.

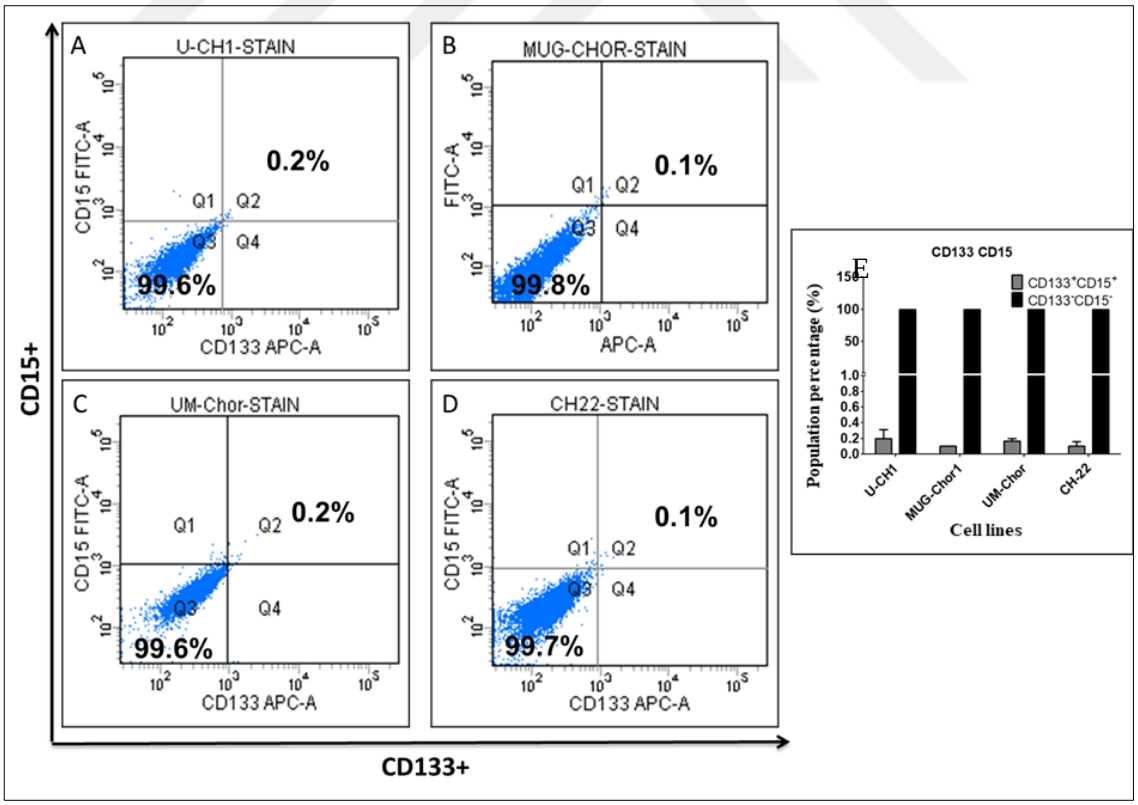

Figure 4.1. Flow cytometry analysis of CD15 and CD133 positive populations in chordoma cell lines. (A)U-CH1, (B)MUG-Chor1, (C)UM-Chor1, (D)CH22 cells, and E) The percentage of positive and negative cells in the population respectively.

Dual immunostaining and flow cytometry analysis revealed that $CD133^{+}/CD15^{+}$ double positive subgroups are approximately 0.1-0.2% of the total cell population in MUG-Chor1, UM-Chor1, U-CH1 in CH22 cell lines (Figure 4.1. A-D). As expected, the majority of the general population, approximately more than 98% were composed of $CD133^{-}/CD15^{-}$ cells (Figure 4.1. E).

In figure 4.2, qPCR results of $CD133^{+}/CD15^{+}$ dual positive subgroups in each cell line are presented. Stemness and EMT-related gene expression levels were evaluated based on the relative expression changes (2^-ΔΔCt) in comparison to parental cell lines.

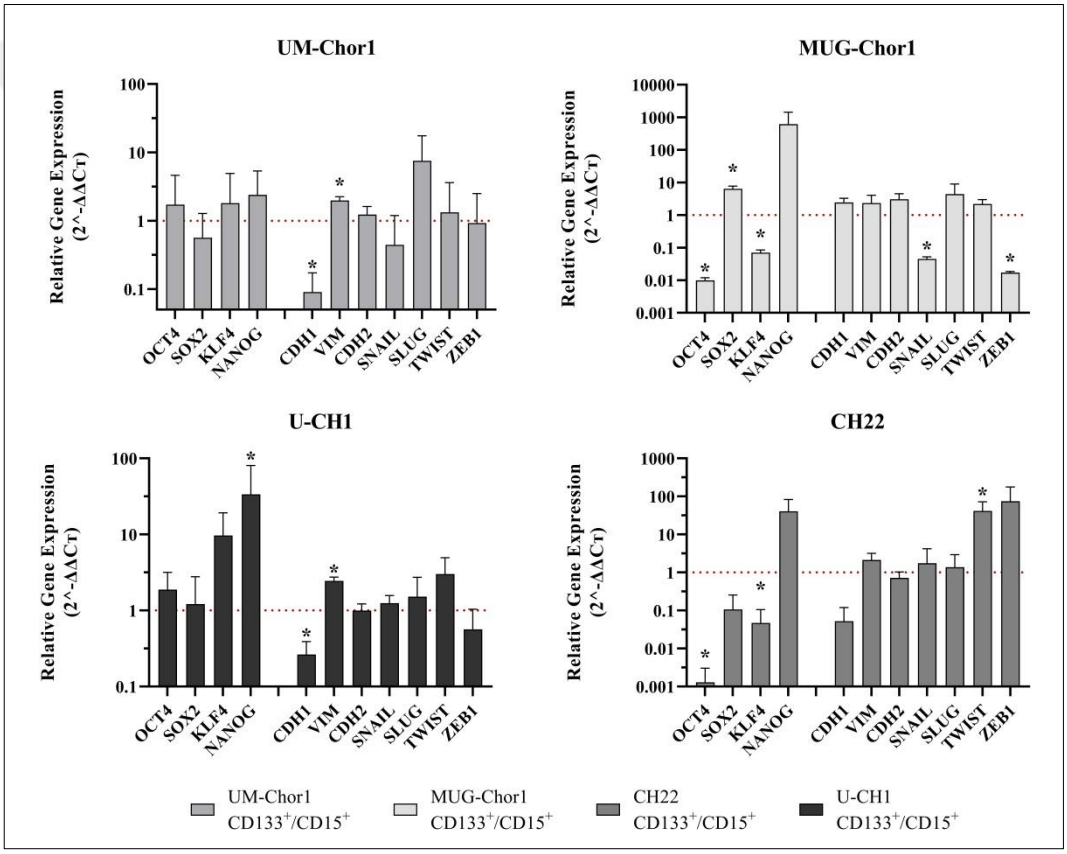

Figure 4.2.Stemness and EMT phenotype-related gene expression assays in CD133+/CD15+ double positive chordoma cells (control groups are represented as red dashed lines). (A) UM-Chor1, (B) MUG-Chor1, (C) U-CH1, (D) CH22 '*'indicates $p<0.05$.

Overall findings demonstrate that $CD133^{+}/CD15^{+}$ subpopulations exhibit various expression profiles depending on the cell line in terms of stemness, epithelial and mesenchymal phenotype. The stemness-related *Nanog* gene and mesenchymal phenotype

indicator *Twist* genes are up-regulated and *Sox2* gene is downregulated in CD133+/CD15+ subpopulations of all 4 cell lines. In UM-Chor1, U-CH1, and CH22 cells, *E-Cad* (*CDH1*) expression is downregulated as an indicator of early EMT process; however, *N-Cad* (*CDH2*) levels did not show any significant changes in these groups (Figure 4.2. A, C, D). U-CH1 and UM-Chor1 derived tumorsphere populations have increased *Klf4, Nanog*, inhibited *E-Cad* expression, and up-regulated *Vimentin*, *Twist*, and *Slug* levels (Figure 4.2. A, C). It is noted that besides the increase in *Nanog* expression, other stemness-related genes either did not significantly change or down-regulated in $CD133^{+/}CD15^{+}$ populations of CH22 (Figure 4.2.D). On the other hand, MUG-Chor1 populations showed increased expression for both epithelial (CDH1) and mesenchymal tissue marker genes that can be interpreted as a hybrid-EMT phenotype (Figure 4.2 B). However, the overall stemness-related profile of $CD133^{+}/CD15^{+}$ populations suggests that these CSC markers may not be applied to the CH22 cell line in comparison to U-CH1, UM-Chor1, and MUG-Chor1 lines.

4.1.2. Isolation and determination of the changes in stemness and EMT-related gene expressions of chordoma side population cells

Side population analysis based on the efflux of Hoechst 33342 stain revealed an irregular pattern for different chordoma cell lines (Figure 4.3). Since both Hoechst and verapamil concentrations should have been determined based on the cell line, and there was no literature finding for chordoma-specific concentrations, an optimization step was required. Chordoma cell suspensions treated with 5µg/mL, 10µg/mL, and 15µg/mL Hoechst 33342 concentrations with 50µM, 75µM, 100µM, or no verapamil combinations were analyzed based on cell viability and the specific "side population". Based on the optimization trials 50µM verapamil and 10ug/mL Hoechst 33342 for 1.10^{6} cells were used in the assay.

In CH22 cell line, a small population in the bottom left quadrant with the verapamil-stained group (Figure 4.3.D, gate P3) decreased by 0.6% in the verapamil-free group. Further FC analysis on U-CH1, MUG-Chor1, and UM-Chor1 groups showed that a slight increase in the P3 gate was caused by the shrinkage which diminished the Hoechst signal and the cells moved slightly to the negative region (Figure 4.3. A-C, gate P3). These findings might not provide a pure side-population but were considered as carrying the "potential" of a side-population-enriched group of cells.

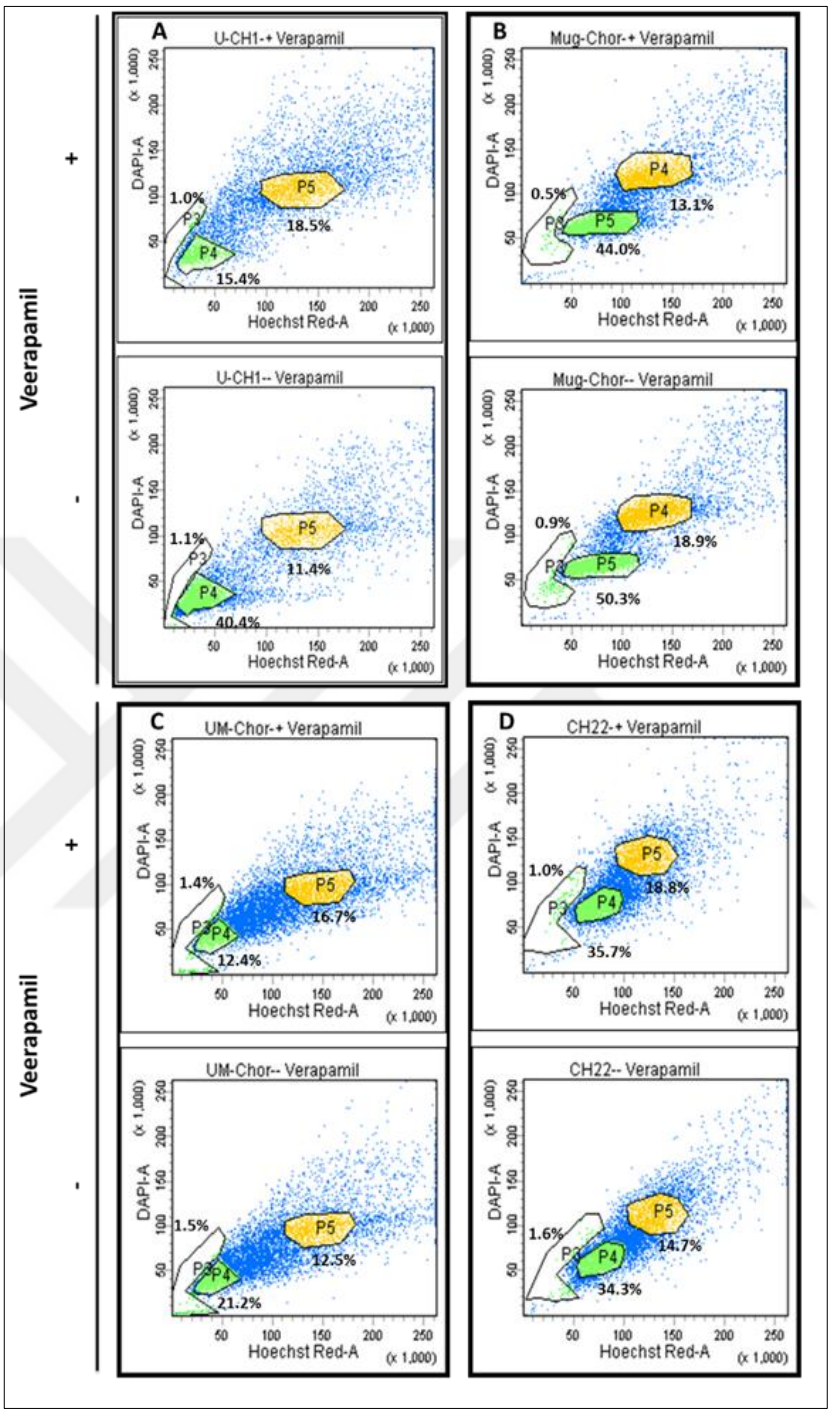

Figure 4.3.Side population analysis of chordoma cell lines. FC analysis of U-CH1, MUG-Chor1, UM-Chor1, CH22 cells

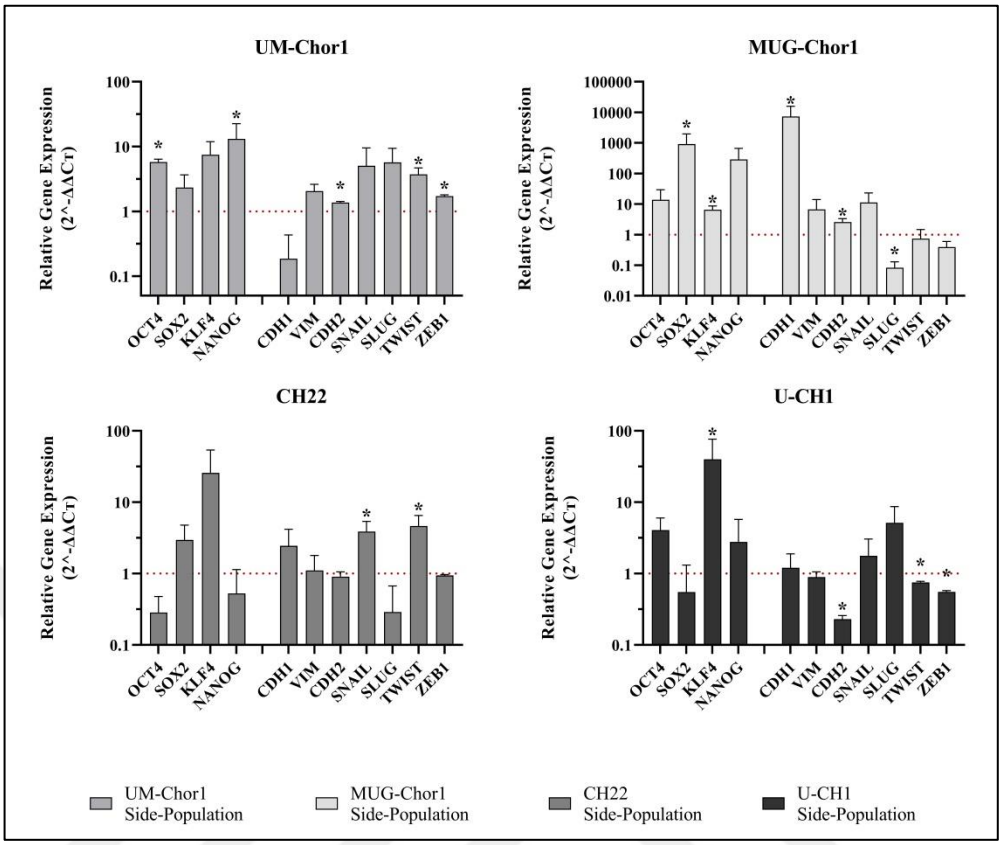

Figure 4.4. Stemness and EMT phenotype-related gene expression assays in Side population cells (control groups are represented as red dashed lines). (A) UM-Chor1, (B) MUG-Chor1, (C) U-CH1, (D) CH22 '*'indicates $p<0.05$.

Based on the qPCR analysis, an overall increase in stemness-related gene expression levels in comparison to parental cell lines was observed (Figure 4.3). Epithelial marker *E-Cad*, which was expected to be down-regulated, was up-regulated in both MUG-Chor1 and CH22, did not change in U-CH1 and down-regulated in UM-Chor1 side population cells. The up-regulated mesenchymal phenotype markers observed in UM-Chor1 were not common among other cell line groups (Figure 4.4.A).

Our findings revealed the mechanism of Hoechst 33342 dye efflux does not occur similarly in all chordoma cell lines. Considering the inhibition of stemness markers in CH22 side populations and the failure to meet the expected change in mesenchymal markers, side population analysis is not found viable as an efficient method of enriching the cancer stem cell population for chordoma cells.

4.1.3. Culture and determination of the changes in stemness and EMT-related gene expressions of tumorsphere-derived chordoma cells

Tumorsphere formation assay is a cell culture model developed to enrich cancer cell sub-populations with high stemness capacity by exposing heterogeneous cancer cell populations to a selective serum-free and attachment impaired- environment.

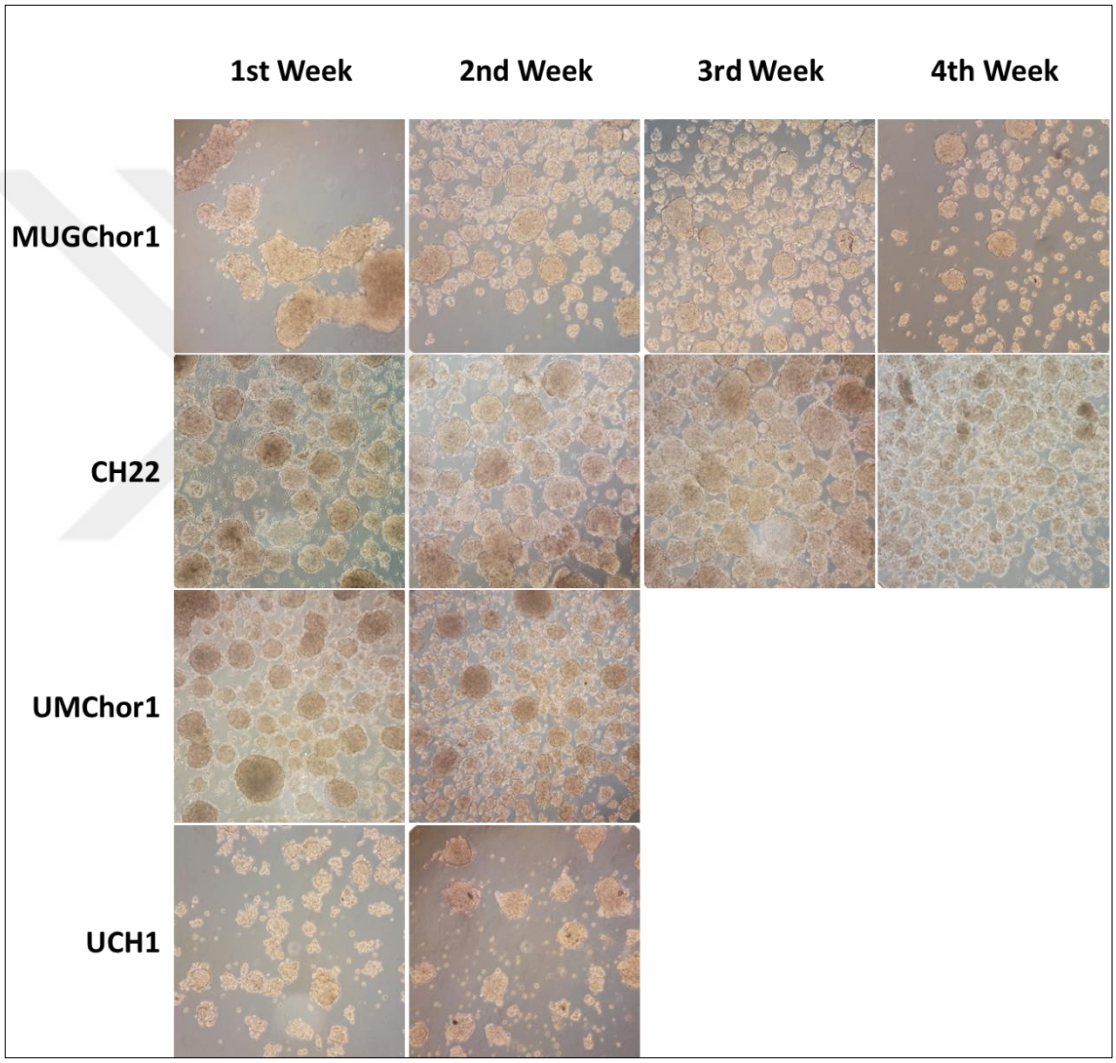

Figure 4.5. Sub-cultured tumorspheres at different time points (top to bottom MUG-Chor1, CH22, UM-Chor1, U-CH1 cell line; left to the right end of the first week, 2nd, 3rd and the 4th respectively.

Since every chordoma cell line has unique stemness features, not all cell lines could manage to form new spheres after sub-culturing more than once. Unlike MUG-Chor1 and CH22 cell lines that formed spheres successfully even after three subsequent passages

(called the 4th generation), U-CH1 and UM-Chor1 cell lines immediately lost their capacity to form spheres after a second passage. Consequently, in the following transcriptome arrays, the 4th generations of tumorspheres for MUG-Chor1 and CH22 (passaged 3 times), and the 2nd generation tumorspheres for UM-Chor1 and U-CH1 lines (passaged one time) were used. Images of the spheres are presented in Figure 4.5; each image was taken at the end of the given week, just before the collecting cells.

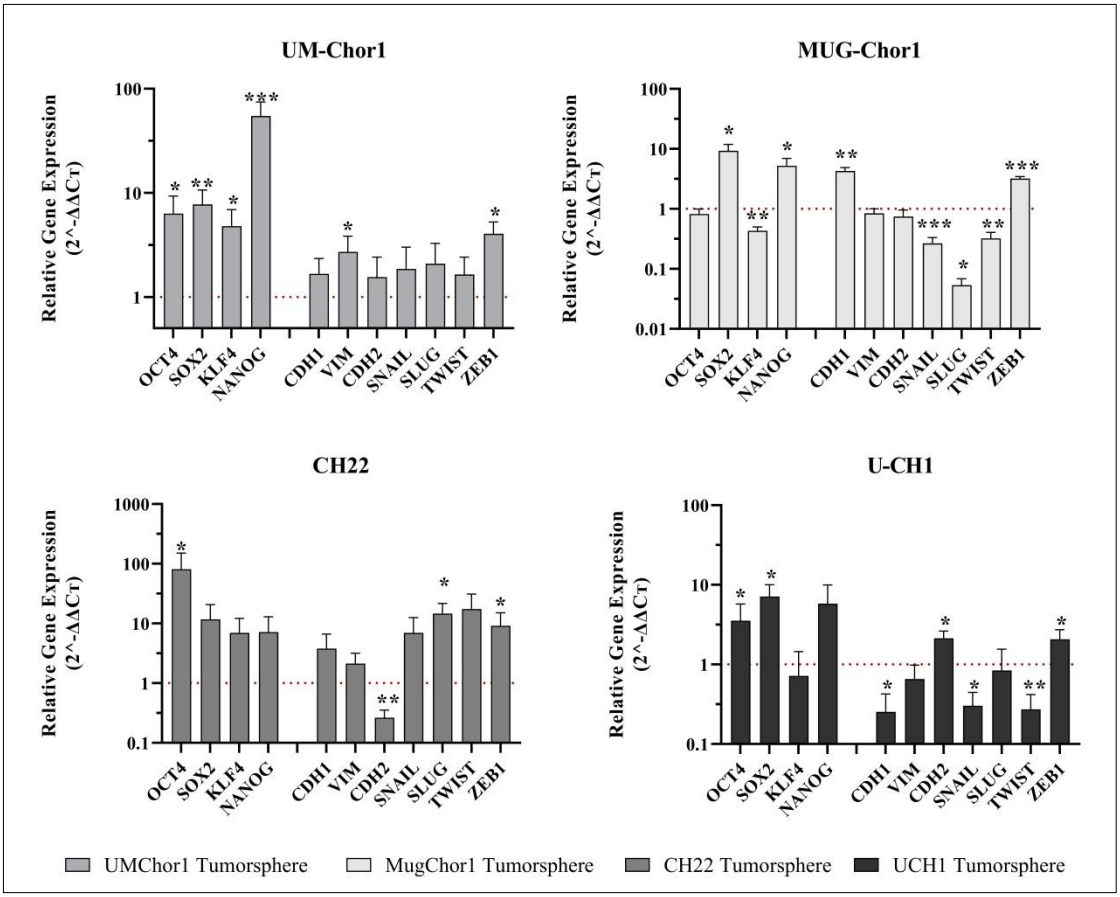

Figure 4.6. Stemness and EMT phenotype-related gene expression assays in chordoma tumorspheres (control groups are represented as red dashed lines). (A) UM-Chor1, (B) MUG-Chor1, (C) U-CH1, (D) CH22 '*'indicates $p<0.05$.

The tumorspheres were compared with parental cell lines to evaluate the changes in stemness and EMT-related gene expression profiles (Figure 4.6). RT-qPCR results showed that all tumorsphere forming chordoma cells exhibit increased *Sox2, Nanog,* and *Zeb1* expression, which indicates induced stemness and a mesenchymal phenotype. UM-Chor, MUG-Chor1, and CH22–derived tumorspheres are enriched in cell populations with increased epithelial cadherin levels, while U-CH1–derived tumorspheres exhibit lower

expression. N-cadherin levels were downregulated in CH22 spheres, consistent with the EMT phenotype; however, increased in U-CH1, while remaining unaffected in UM-Chor1 and MUG-Chor1-derived tumorspheres. EMT-regulating transcription factors *Snail*, *Slug*, and *Twist* genes are down-regulated in U-CH1 and MUG-Chor1 tumorspheres (Figure 4.6.B, D). A mesenchymal marker, *Vimentin*, is up-regulated in CH22 and UM-Chor1 tumorspheres; however, its expression was not altered in MUG-Chor1 and U-CH1 tumorspheres.

Mean relative expression levels of all stemness, epithelial, and mesenchymal phenotype-related genes in tumorsphere forming UM-Chor1 and CH22 cells showed a general up-regulation which signifies an enriched hybrid-EMT phenotype (Figure 4.6.A, C). In an overall evaluation of tumorsphere formation in terms of stemness-related expression profile, the method was found applicable for all four chordoma cell lines. Additionally, it was noticed that tumorsphere formation leads to a mutual mean increase in epithelial cadherin levels in sub-cultured tumorspheres.

On the whole, there was no significantly superior enrichment technique among the three chordoma cancer stem cell enrichment methods we evaluated in terms of increased mesenchymal transcription factors and marker genes following epithelial phenotype loss with perfect stemness increase. Therefore, tumorsphere-forming cells were chosen for further research into chordoma cancer stem cells due to their significantly higher mean stemness-related gene expression profile, which includes at least one significant up-regulated stemness marker and a mesenchymal transcription factor.

4.2. DETERMINATION OF TRANSCRIPTOMIC VARIATIONS IN CHORDOMA TUMORSPHERES

The total transcriptomes of 4 chordoma cell lines (U-CH1, MUG-Chor1, UMChor1, and CH22) were determined via microarray analysis as 3 replicates of tumorspheres and parental groups by using Clariom D human total transcriptome array. In this array, more than 135000 transcripts were assayed, and differentially expressed transcripts in each experimental group were categorized based on up-/down-regulation and coding/non-coding features. The distribution of down and up-regulated transcripts for each cell line subgroup were presented in Figure 4.7. The differences in the volcano plots of the tumorspheres

reveal that the tumorsphere enrichment method produces a different transcriptomic profile for each cell type.

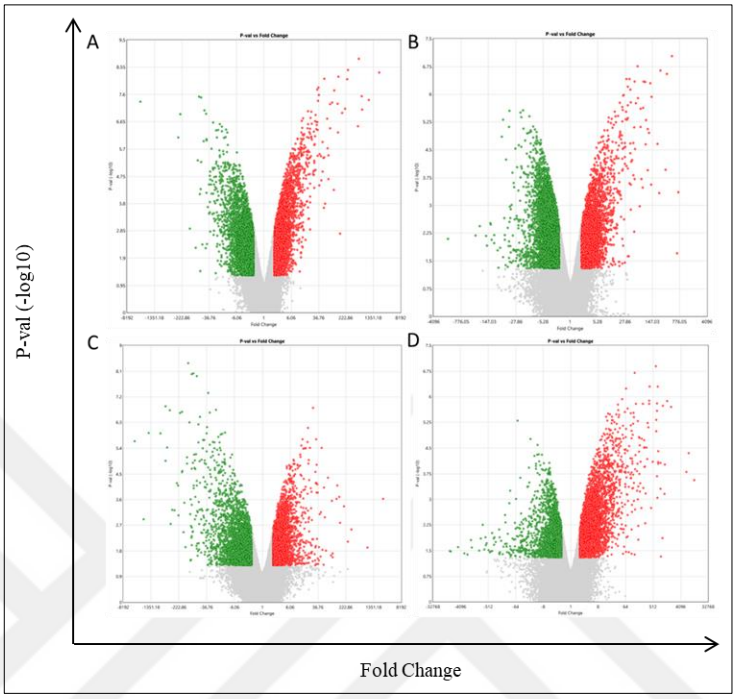

Figure 4.7. Volcano plots of DE transcripts in tumorspheres in comparison to parental cell lines for each chordoma cell line. A) UM-Chor1, B) MUG-Chor1, C) CH22, D) U-CH1. Red dots represent statistically up-regulated, green dots represent down-regulated, and grey dots are non-significant transcripts. $p < 0.05$ and $|FC| > 2$ values are considered significant.

Clariom D is a comprehensive total transcriptome array that covers coding and (mostly long) non-coding transcripts including genes with unassigned features, structural RNAs, and pseudogenes, allowing the formation of competing endogenous RNA (ceRNA) interaction networks and detection of unassigned transcripts with a potential of regulatory functions. Table 4.1 and the pie charts in 4.8 describes the variations of differentially expressed transcript groups by their functional subtypes.

Among the significantly changed transcripts, 4211 were up-regulated, and 6054 were down-regulated transcripts in UM-Chor1; 4197 were up-regulated, 2874 were down-regulated transcripts in UCH-1; 3897 up-regulated, and 7552 were down-regulated transcripts in MUG-Chor1; 3364 up-regulated and 3413 down-regulated transcripts in CH22 derived -tumorspheres are detected in comparison to parental cell lines (Table 4.1)

Table 4.1. Quantitative distribution of differentially expressed transcripts grouped as their biological functions

UM-Chor1 Tumorspheres				
Group	Total	Passed Filter	Up-Regulated	Down-Regulated
Multiple_Complex	29510	3747	2860	887
NonCoding	66845	2583	528	2055
Coding	18858	1767	266	1501
Unassigned	10001	1244	134	1110
Pseudogene	4340	394	331	63
Precursor_microRNA	3297	332	8	324
Small_RNA	2386	188	84	104
Ribosomal	507	6	0	6
tRNA	6	4	0	4
U-CH1 Tumorspheres				
Group	Total	Passed Filter	Up-Regulated	Down-Regulated
Multiple_Complex	29510	3230	2690	540
NonCoding	66845	1691	571	1120
Coding	18858	928	383	545
Unassigned	10001	755	257	498
Pseudogene	4340	243	207	36
Small_RNA	2386	134	65	69
Precursor_microRNA	3297	77	23	54
Ribosomal	507	13	1	12
tRNA	6	0	0	0
MUG-Chor1 Tumorspheres				
Group	Total	Passed Filter	Up-Regulated	Down-Regulated
NonCoding	66845	3521	255	3266
Multiple_Complex	29510	3267	2869	398
Coding	18858	2069	281	1788
Unassigned	10001	1277	46	1231
Precursor_microRNA	3297	686	8	678
Pseudogene	4340	433	377	56
Small_RNA	2386	178	54	124
Ribosomal	507	13	7	6
tRNA	6	5	0	5
CH22 Tumorspheres				
Group	Total	Passed Filter	Up-Regulated	Down-Regulated
Multiple_Complex	29510	2951	1424	1527
NonCoding	66845	1629	825	804
Coding	18858	843	333	510
Unassigned	10001	653	306	347
Pseudogene	4340	348	214	134
Small_RNA	2386	190	166	24
Precursor_microRNA	3297	121	58	63
Ribosomal	507	40	38	2
tRNA	6	2	0	2

The term “NonCoding” in Table 4.1 stands for only non-protein-coding transcripts. Usually describes long non-coding transcripts unless can be determined as small non-coding. “Ribosomal” term is only for ribosomal RNAs(rRNA). Small_RNA term covers sRNA, snRNA, snoRNA, scaRNA, Y RNA, guide RNA, vault RNA (large ribonucleoprotein particles), and ribozymes. The “Multiple Complex” term contains more than one locus type covered in this chip. Unassigned means the detected transcript could not be identified based on source data.

Detailed examination of transcriptome profiles of chordoma tumorspheres in terms of functional subtypes showed that “multiple complex” transcripts constitute the highest percentage, which could be a confounder since it contains common loci for more than one transcript type, for all 4 cell groups. Except for the multiple complexes, the highest differential expression changes were observed at the levels of non-coding transcripts (Figure 4.8).

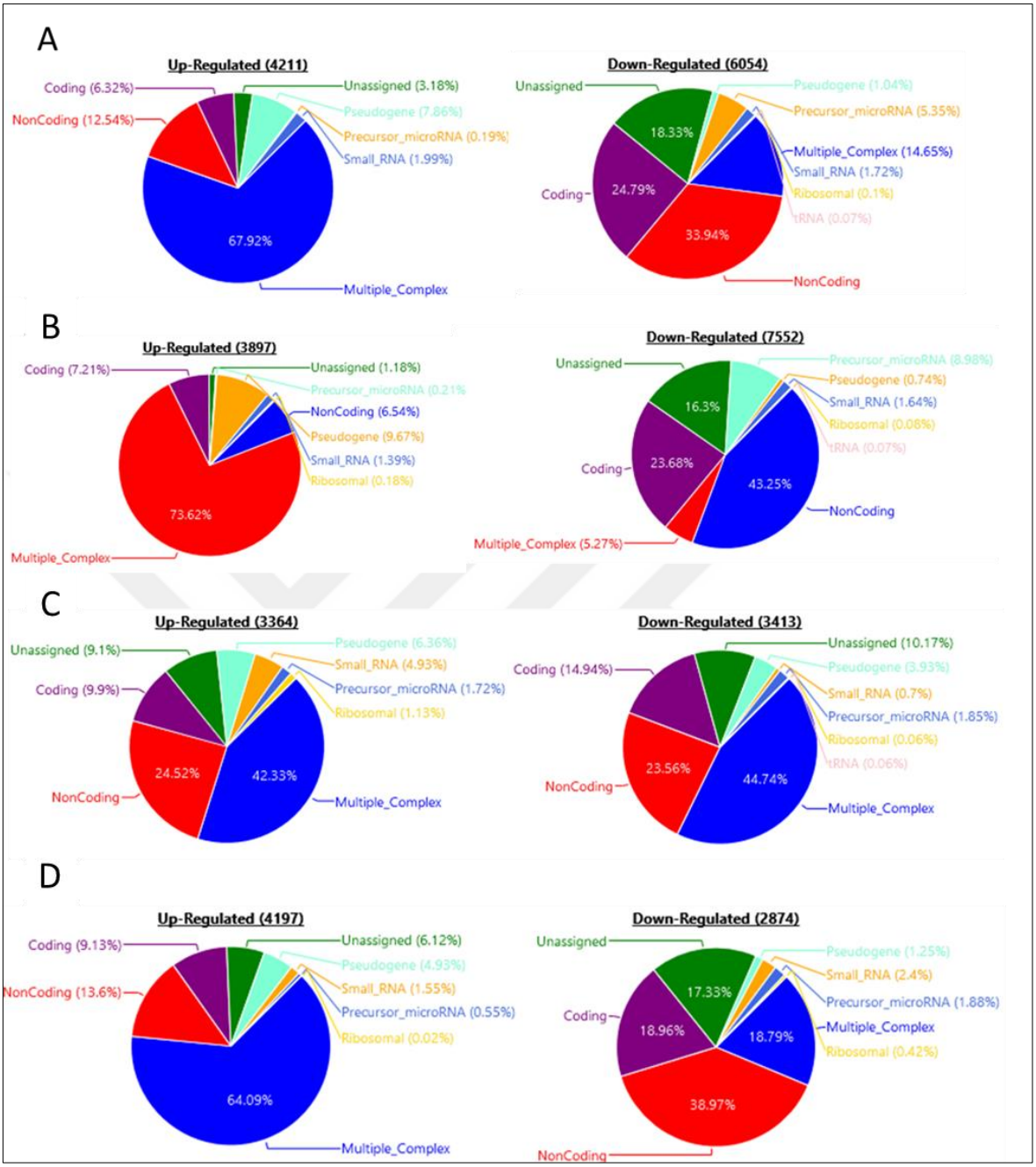

Figure 4.8. Percentage distribution of differentially expressed transcripts in total transcriptome grouped as their biological function. (A) UM-Chor1, (B) MUG-Chor1, (D) CH22, (D) U-CH1.

The evaluation of transcriptomic profile changes identified 256 common DEGs among all tumorsphere-forming cells (Figure 4.9).

Figure 4.9. Unique and common DEGSs of chordoma tumorspheres

All tumorsphere forming and parental chordoma cell lines were also examined for similarity in terms of expression profile changes via hierarchical clustering (Figure 4.10). Cell line-based normalized expression values for each group were used in the clustering analysis. The results showed the clear discrimination of tumorsphere and control groups in each cell line (Figure 4.10 B-E). Noticeably, UM-Chor1, U-CH1, and MUG-Chor1 cell lines are seen to be clustered distinctly as “parental versus control” groups, but CH22 samples were grouped closer to the parental samples of other cell lines in the dendrogram (Figure 4.10 A). Due to these findings, it was decided to proceed with MUG-Chor1 and UM-Chor1 cell lines for further functional experiments which were planned to continue with only two cell lines.

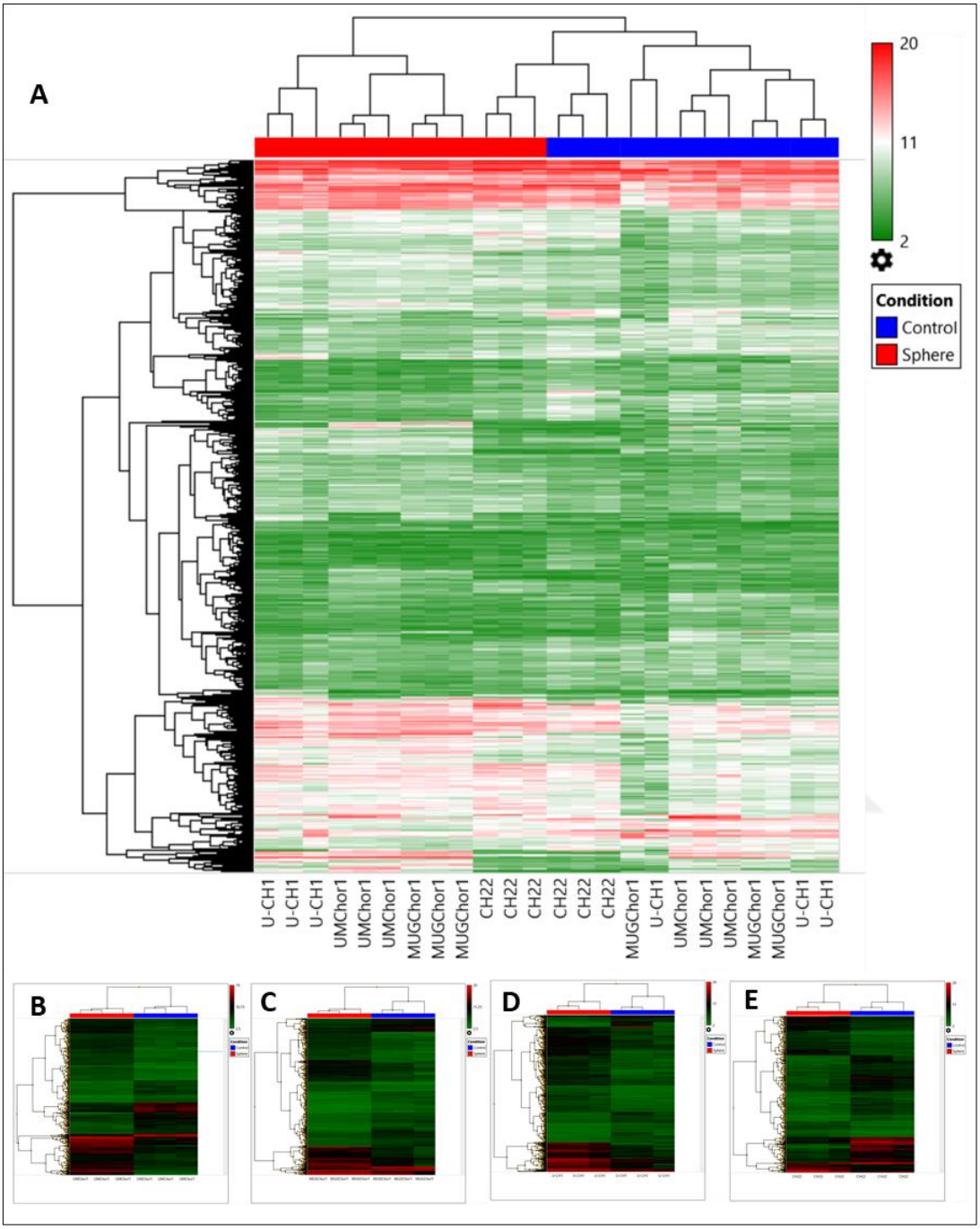

Figure 4.10. Clustering analysis of chordoma tumorspheres and parental cell lines. A-E) Heatplots and dendrograms of clustering analysis of pooled samples, UM-Chor1, MUG-Chor1, U-CH1, and CH22 samples, respectively.

4.2.1. Selection of the target genes for further functional analysis

The “core analysis” provided by QIAGEN-IPA indicated 6 “analysis ready molecules” shared among tumorspheres of all 4 chordoma cell lines. These transcripts were TSC22

Domain Family Member 1 (*TSC22D1*), Src Kinase Associated Phosphoprotein 2 (*SKAP2*), Testis Expressed 9 (*TEX9*), Rho GTPase activating protein 12(*ARHGAP12*), Tumor protein D53 (a.k.a. Tumor protein D52 like 1, *TPD52L1*) and Semaphorin 5A (*SEMA5A*) were also determined among as intersecting DEGs (Figure 4.9).

4.2.1.1. Differential expression analysis of Chordoma and Nucleus pulposus tissue samples

Datasets with Affymetrix HG-U133_Plus_2 microarray transcriptome chip retrieved from the GEO database are presented in table 3.2. The PCA map and hierarchical clustering dendrogram showed that the NP and chordoma samples are distinguished as distinct clusters (figure 4.11 A-B). In the differential expression analysis, 8737 up-regulated, and 5466 down-regulated DEGs were detected in chordoma tissue samples, in comparison to nucleus pulposus tissues (Figure 4.11.C).

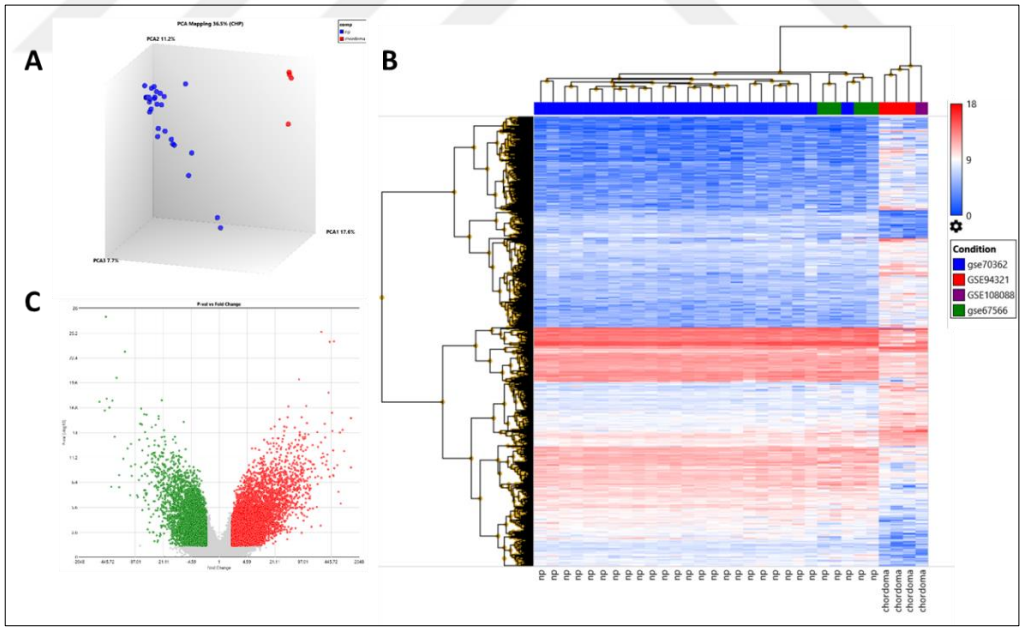

Figure 4.11. Transcriptome analysis of chordoma and Nucleus pulposus tissue samples A) PCA map of Nucleus pulposus and chordoma samples, B) Hierarchical Clustering heatmap C) Scatter plot of DEGs (Fold change vs -log10 p Value; Red dots represent up-regulated, green dots represents down-regulated genes in chordoma samples in comparison to nucleus pulposus samples. Grey dots are representing non-significant transcripts)

Comparison of chordoma tissue samples (chordoma tissue vs NP) with tumorsphere (tumorsphere vs control) revealed 83 common differentially expressed genes. Among those genes, *SEMA5A* and *TPD52L1* were chosen for further functional experiments. A literature review revealed that they are already associated with cancer, but no studies on chordoma have been published, yet. Similarly, Gastric Adenocarcinoma Associated Long Intergenic Non-Coding RNA (*GAPLINC*), which is also differentially expressed in all chordoma tumorspheres, was chosen as one of the targets for further functional experiments.

4.2.2. Functional enrichment analysis of differentially expressed genes in chordoma tumorspheres

Besides the fact that genes with the highest absolute fold change value have a great potential to be a marker or regulatory element, enrichment of gene sets allows us to obtain a wider perspective on altered cellular physiology. Since the initial objective of this study is to unveil the differential role of CSCs in chordoma pathophysiology, identifying the prominent pathways would provide a systematic perspective. For a comprehensive evaluation, several functional enrichment approaches were reviewed.

Initially, for specific insight into the differentially expressed transcripts of chordoma tumorspheres from 4 different cell line sources, Gene Ontology (GO) terms for biological processes (BP), cellular components (CC), and molecular functions (MF) were investigated comparatively (Figure 4.12). The enrichment analysis was conducted via *clusterProfiler* R package; plots were created based on the 5 most enriched terms for each ontology with a p-value cutoff of 0.05 and the highest number of genes involved for each cell line group.

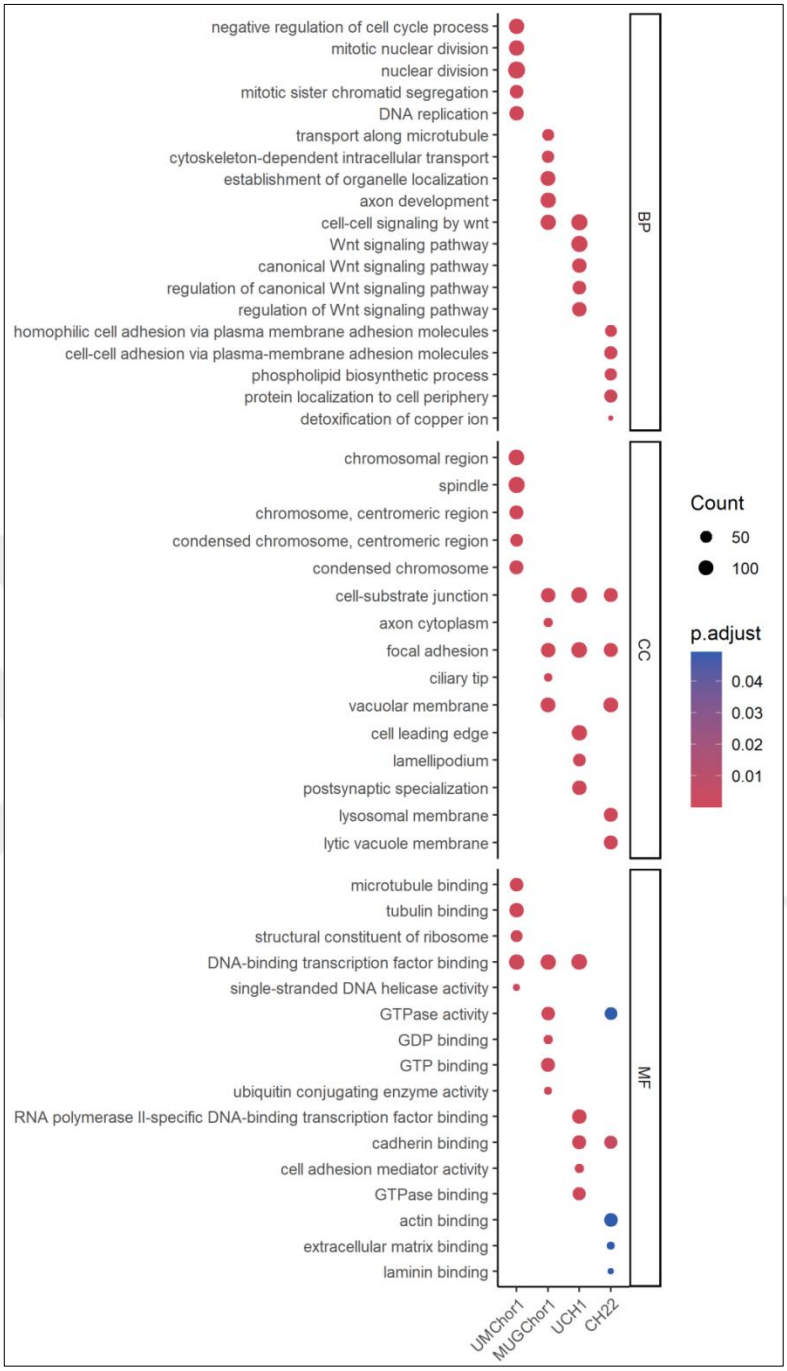

Figure 4.12. Enriched Gene Ontology terms in chordoma tumorspheres (p-value <0.05). Count: number of genes associated in each GO-term, represented as size.

Among the top 5 enriched GO-terms, no common term was detected that was shared by all tumorsphere groups. However, cellular components terms "Focal adhesion" and "cell-substrate junction", in addition to the molecular function term "DNA-binding transcription factor binding" are shared by at least 3 tumorsphere groups.

WikiPathways is a monthly updated curation of biological pathways that also is accessible via the TAC software interface. The top 15 dysregulated WikiPathways with the highest number of differentially expressed genes in chordoma tumorspheres are presented in Figure 4.13. Significance levels of dysregulated pathways were calculated using Fisher's Exact Test considering the number of dysregulated genes involved.

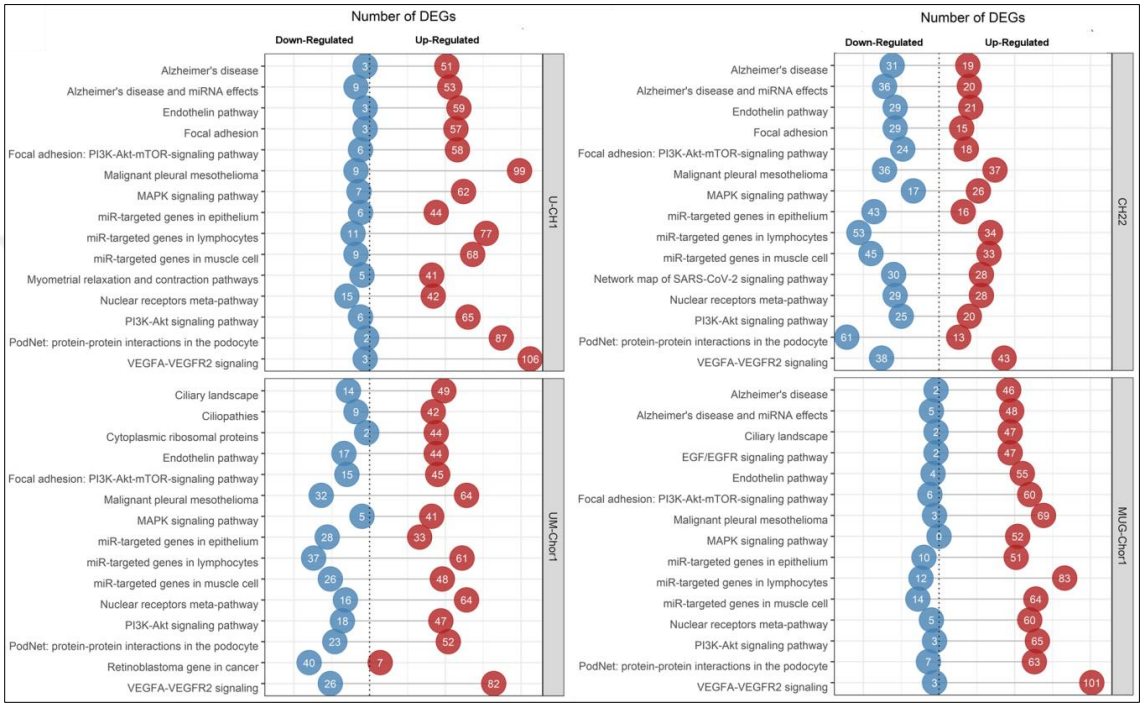

Figure 4.13. Top fifteen enriched WikiPathways in chordoma tumorspheres. Enrichment results are shown as individual grids for each cell line. The color of the spheres defines whether the enriched genes are –up/-down regulated in the aforementioned pathway cluster, and the number of the dysregulated genes based on expression level are stated inside de spheres.

Results showed that among those 15 biological processes, 11 of them are commonly dysregulated in all 4 cell lines: Endothelin pathway, Focal adhesion: PI3K-Akt-mTOR-signaling pathway, PI3K-Akt signaling pathway, Malignant pleural mesothelioma, MAPK signaling pathway, VEGFA-VEGFR2 signaling, miR-targeted genes in epithelium, miR-targeted genes in lymphocytes, miR-targeted genes in muscle cell, Nuclear receptors meta-pathway, PodNet: protein-protein interactions in the podocyte. Those pathways are noted as candidate regulatory/affected processes in chordoma tumorspheres.

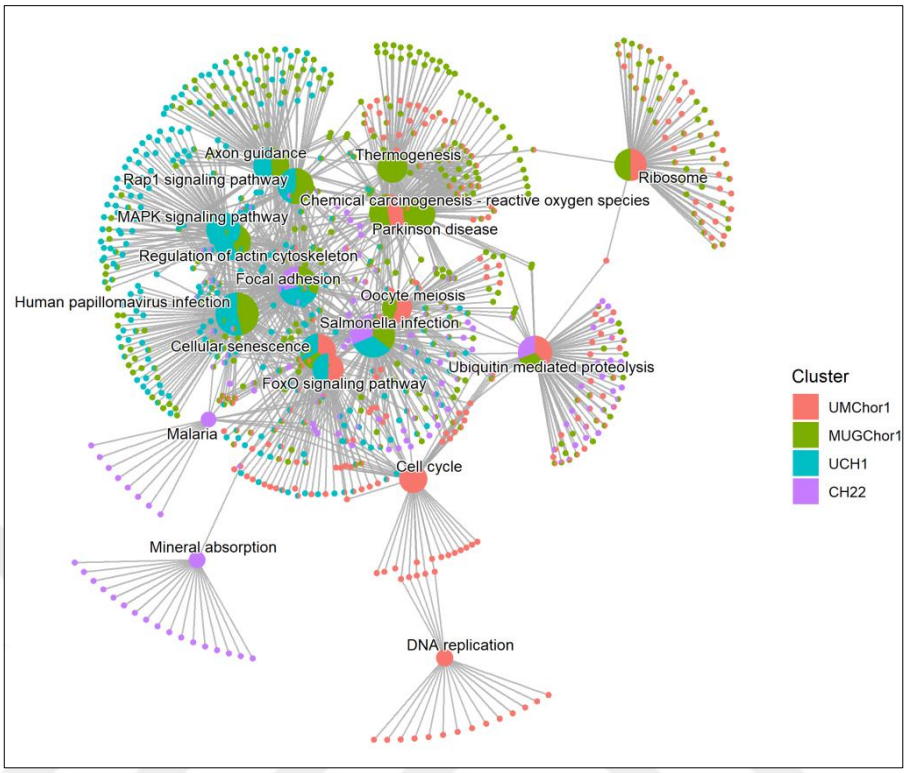

Figure 4.14. The gene-concept network of enriched KEGG pathways for the chordoma tumorsphere DEGs. Outer smaller nodes represent genes, and the central nodes represent enriched pathways composed of gene clusters. The colors of the nodes indicate the cell lines they belong. The number of genes related to each term is indicated by different sizes

Next, the KEGG knowledgebase was used for gene set enrichment of differentially expressed gene sets of tumorspheres. The gene-concept network plot describes both co-expressed genes and the clusters of these genes enriched distinctive several pathways (Figure 4.14). The top 5 significantly enriched KEGG analyses showed no common pathway among all 4 chordoma cell lines (p-value cut off 0.05). Nonetheless, cellular senescence and ubiquitin-mediated proteolysis pathways were common among UM-Chor1, MUG-Chor1, and U-CH1; and Focal adhesion was enriched for MUG-Chor1 and U-CH1and CH22 – derived tumorspheres.

4.3. IDENTIFICATION OF PROMINENT GENOMIC VARIANTS IN CHORDOMA TUMORSPHERES

In this section of the study, somatic alterations of chordoma tumorsphere and parental samples were investigated via the OncoPlus Next-Generation DNA sequencing panel. Twenty-two samples that passed filtering criteria based on read quality metrics and depth were used for further analysis.

To create a general genomic chordoma profile, the genes in which genomic variation events are most frequently occurring in various chordoma cell lines (regardless of the tumorsphere formation) were identified (Figure 4.16). The Oncoplot was created to visualize ranks the top 20 most frequently mutated genes among the samples and depicts the variant variety in a heatmap format, with each sample represented in a column. The percentage of the samples that have variations of a gene is shown in the bar graph on the right; the upper bar graph shows the total mutation burden of the samples.

The oncoplot showed that somatic variants of 6 genes are commonly seen among all chordoma samples. These are BRCA1-associated RING domain 1(*BARD1*), notch receptor 2(*NOTCH2*), RecQ-like helicase 4(*RECQL4*), lysine methyltransferase 2C (*KMT2C*), ERCC excision repair 5, endonuclease (*ERCC5*), aurora kinase A (AURKA) (Figure 4.16). Detailed investigation of these variants showed, five SNPs in *KMT2C* (rs10454320, rs2479172, rs28522267, rs56850341, rs77735469); an SNP leading missense mutation in RECQL4 (rs61788900); 3 SNPs and one deletion mutation in BARD1 (rs1129804, rs2070093, rs2229571, and rs56130510); an SNP in *AURKA* (rs1047972); 2 SNPs in *ERCC5* (rs9514067, rs9514066) are detected in all parental chordoma samples. Additionally, SNP variant rs17655 in *ERCC5* is also detected in all parental cell lines, except MUG-Chor1. An SNP-causing missense mutation in *NOTCH2* (rs61788900) is detected in all parental cell lines, except one sample of the CH22 cell line.

The great majority of the genes altered in the samples had more than one variation detected (Figure 4.16.A, represented as black cells). Since no "healthy" control groups were used in the mutation annotation analysis to exclude germ-line variants, potential benign variations were also captured. This issue was overcome by differential mutation profiling, which was obtained by comparing tumorsphere groups and control groups.

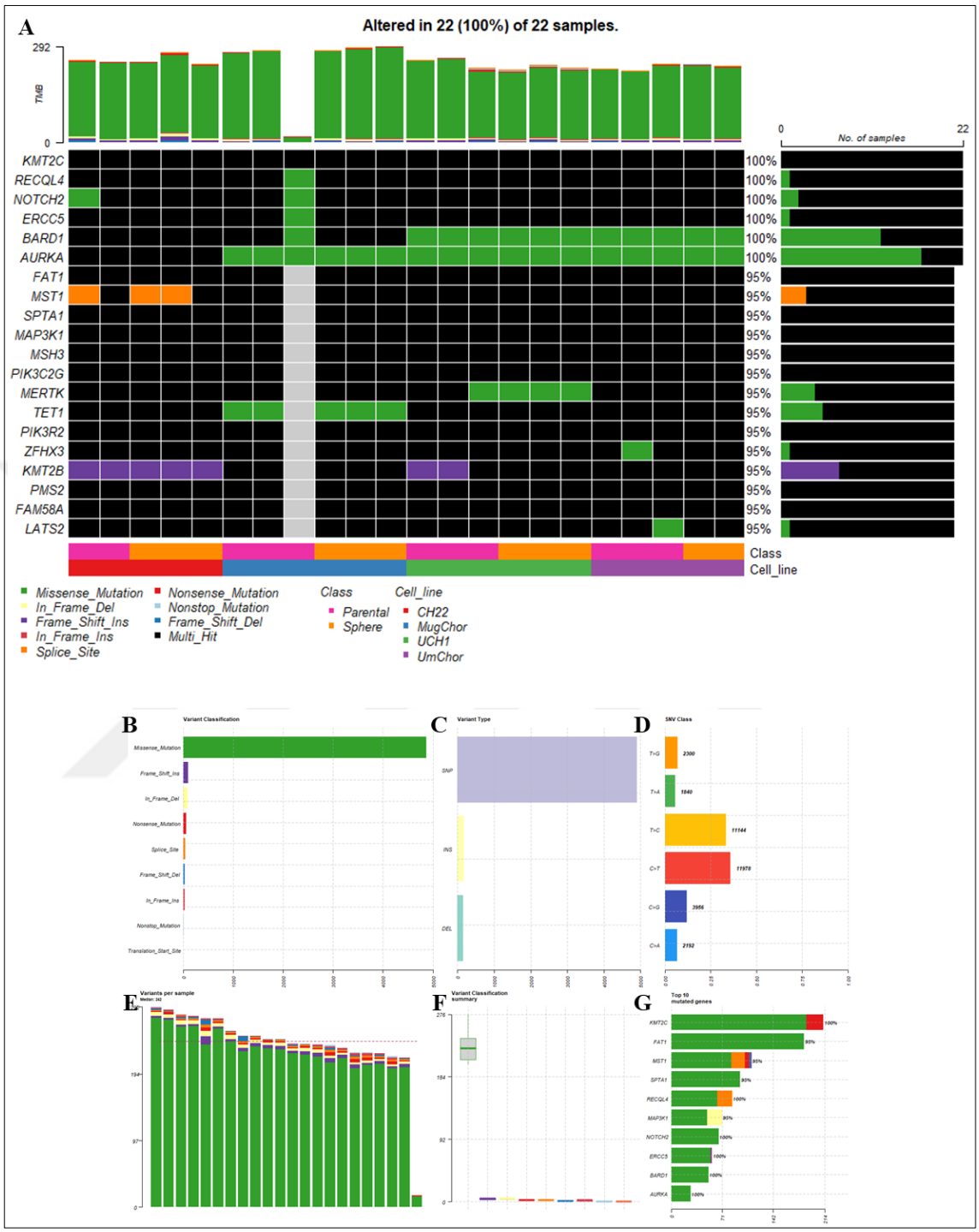

Figure 4.15. Somatic variant profile of chordoma cell lines A) Oncoplot of top 20 genes with the highest mutation burden, B) Variant classification, C) Variant type, D) SNV classes, E) Variants per sample, F) Variant classification summary, G) Variant class details of the top 10 mutated genes

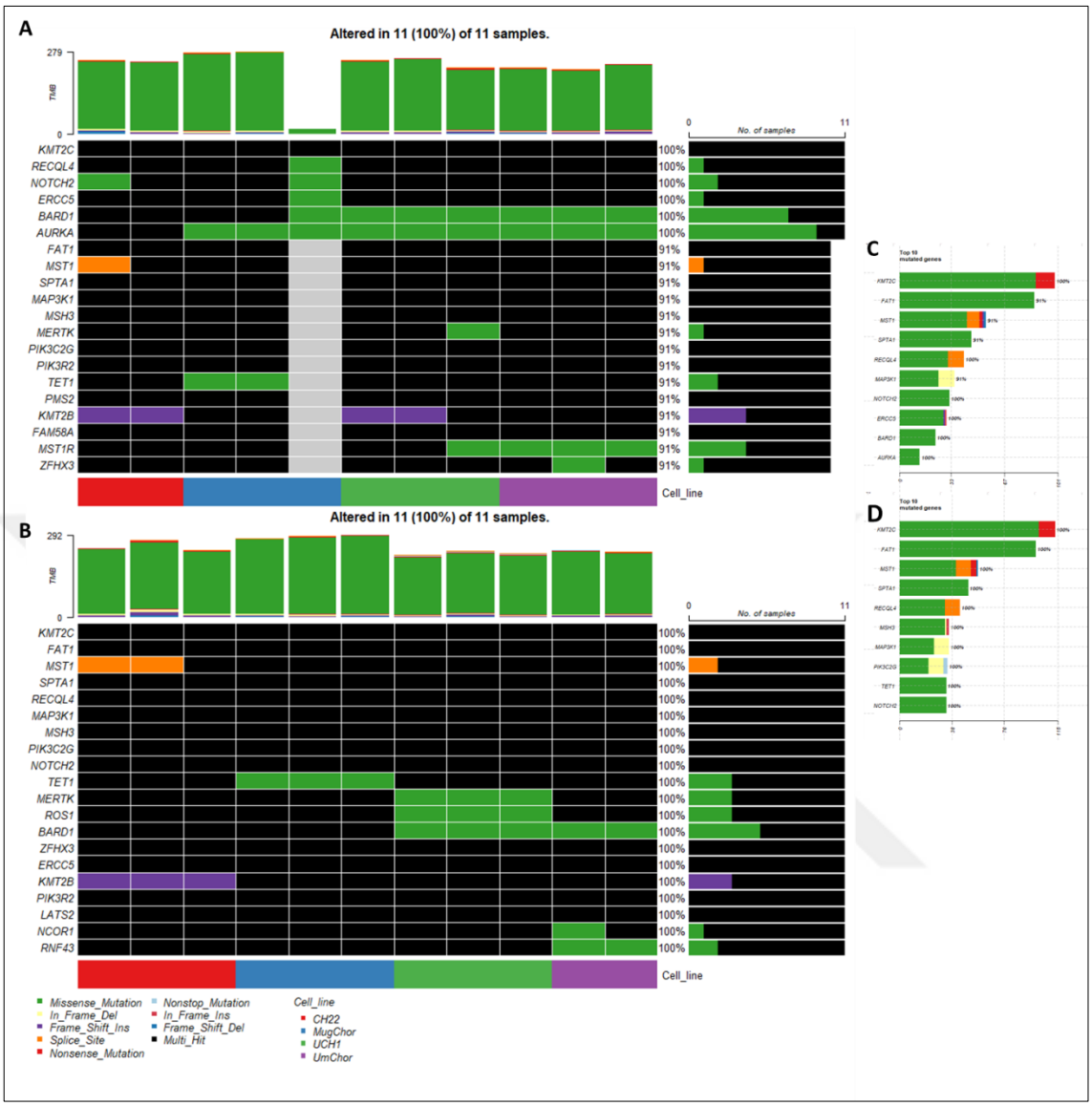

Figure 4.16. The top 10 highly mutated genes in parental (A, C) and tumorsphere (B, D) groups. A, B) Oncoplots are showing the tumor mutation burden of each chordoma sample. C, D) plots are showing the top 10 genes with the highest number of genomic alterations detected.

The comparison showed six common genes among parental samples with somatic variations detected (Figure 4.17 A). However, as a result of tumorsphere formation, 60 additional genes with somatic variations have been found in all sphere samples (Figure 4.17 B) (The complete set of data is not shown). Also noticed top 5 genes with the highest mutation event observed were not changed but the remaining 5 have changed, indicating sphere formation made prominent some of the variants (Figure 4.17 C, D).

A comparison of all tumorspheres and controls as two distinct pools showed no parental or tumorsphere-specific mutation detected. Nevertheless, five mutations were found in more than 60% (7/11) of the parental samples and less than 25% (3/11) of the tumorsphere samples. These are, a deletion in *FLI1* (rs35352545), an insertion in *ERBB4* (rs57466272), SNP variants in *RECQL* (rs577735412), *KMT2C* (rs62478357), and *PYROXD1* (rs577735412) genes, and a novel SNP is detected in the intronic region of *RSPO2* gene (chr8:109095386, c.-170+39N>C). Also, an indel variant rs758227657 is detected in the *PBRM1* gene in more than 60% of tumorsphere samples and less than 20% of parental samples. No polyphen score indicates harmful alterations were reported on variant effect predictor (VEP) analysis for these events.

Further in the analysis, the mutations become in which prominent despite their rarity in parental cells during the tumorsphere-mediated cancer stem cell enrichment protocol, or *de novo* mutations occurred in the process due to the characteristic genomic instability of cancer cells were detected. Tumorsphere samples and their parental controls were grouped for each cell line separately. Because the tumorspheres were derived from parent cell lines and we used cells with only as few number passage differences as possible in the study, the detection of a genomic variant present in parental samples but not in the spheres was not expected. Additionally, no high-impact variation was detected between any tumorsphere and its parental control as a result of mutation annotation and variant effector analyses.

Tumorsphere cultures derived from CH22 cells showed no differential genomic alteration with high impact, or possible damaging outcome, comparing parental CH22 cells. An insertion mutation in *NOTCH2* (rs377249634), a deletion in *KDM5A* (rs60377454), and novel SNVs in *FLI1* (NM_001271010.1: c.655+9865N>C), *MET* (ENST00000318493: c.-14-12124N>G), and *RANBP2* (NM_006267.4: c.72+20N>C) genes with no expected pathogenic outcome were detected in CH22 tumorspheres, but not in parental samples.

Possibly damaging SNVs in *TBX3* (rs1592846474) and *KMT2C* (rs201834857, COSV51275250) genes were identified in UCH1 spheres, but not in parental U-CH1 cells.

An SNP in *HLA-DQB1* (rs1063318) and 2 SNPs in *HLA-DQA2* causing missense mutation with the benign outcome (polyphen score: 0.001, SIFT: score 1) were detected in MUG-Chor1 tumorspheres but not in parental cells.

Several prominent genomic variants have been identified as a result of the tumorsphere culture of UM-Chor1 cells. Based on the VEP analysis neither of the variants was reported to lead to pathogenic outcomes (either low or no polyphen scores). An SNP in *HLA-DRB5* (rs144532965) causing missense mutation, and an insertion mutation in *FLI1* (rs35352545) causing intron variant with modifier impact were detected in UM-Chor1 tumorspheres, but not in parental controls. Additionally, novel SNPs in *CDK12* (NM_015083.1 (NM_016507.2): c.2149N>G), *RTEL1* (ENST00000508582:c.2072N>G), and *KDM5C* (NM_004187.3: c.*141N>C) causing missense mutation with a modifier impact; and a novel SNP in *SMAD3* (NM_005902.3:c.1155-18N>C) causing intronic variant were detected in UM-Chor1-derived tumorspheres.

4.4. CO-EVALUATION OF GENOME AND GENE EXPRESSION-BASED CHANGES IN CHORDOMA TUMORSPHERES

Based on comparisons between chordoma tumorspheres and control samples, potential projections of the genomic variations on transcriptomic changes and, consequently, downstream pathways, were assessed.

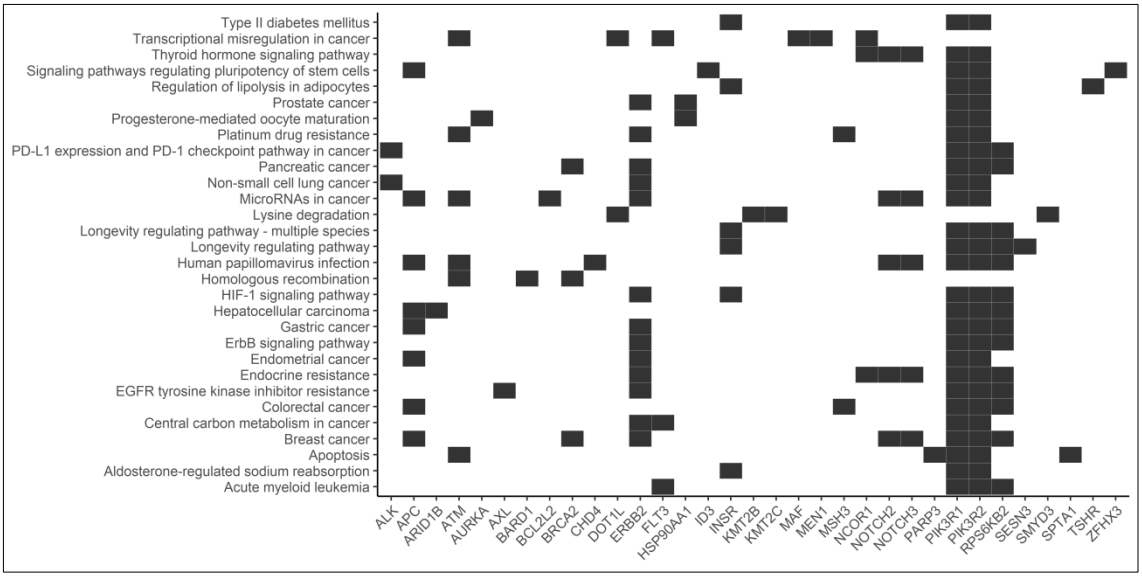

Figure 4.17. GSEA based on KEGG pathways associated with the highly mutated genes on chordoma samples showed deregulations associated with the functions of these genes.

The set of commonly mutated genes in all sphere samples, which also includes the genes detected in parental samples, was used for functional annotation. Enriched KEGG pathways showed many overlaps related to cancer such as “transcriptional misregulation in cancer”, “PD-L1 expression and PD-1 checkpoint pathway in cancer”, “signaling pathways regulating pluripotency of stem cells”, “EGFR tyrosine kinase inhibitor resistance”, in addition to prostate, pancreatic, gastric, endometrial, colorectal, breast cancers and acute myeloid leukemia (Figure 4.17).

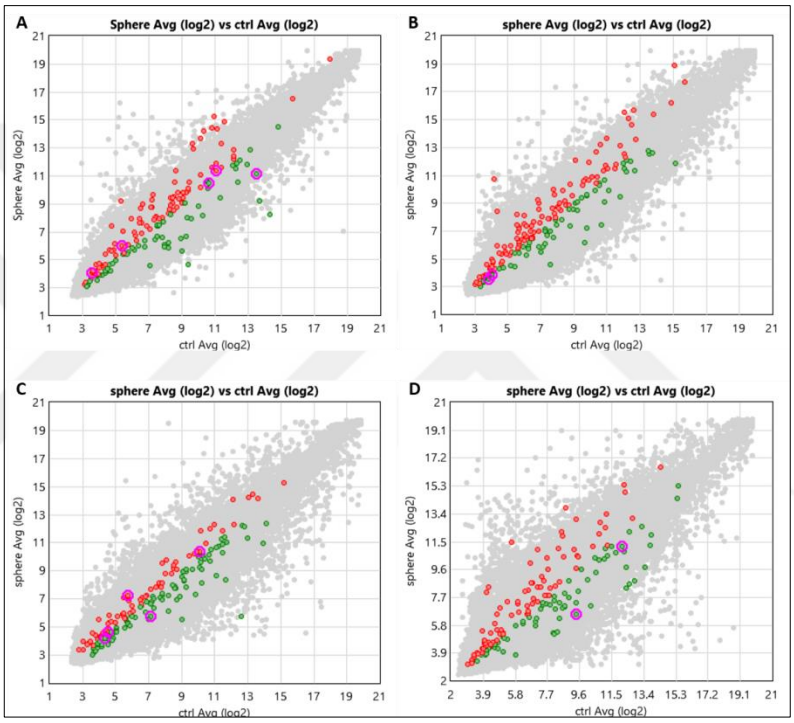

Figure 4.18. Expression changes of genes whose genetic variants have been identified in at least two samples in a cell line group. Scatter plots A–D are UM–Chor1, MUG–Chor1, CH22, and U–Chor1, respectively.

Transcriptome array results and genomic variations detected in at least 2 samples for each cell line based on the oncoplots were co-evaluated (Figure 4.18). Scatter plots showing all outcomes without a p-value or fold change criteria Pink dots describe the genes with differential genomic variation detected in tumorsphere samples but not in parental samples. Gray dots: all genes screened in the transcriptome panel. Colored dots: genes with both differential expression and genomic variants; red dots for upregulation and green dots for downregulation.

Co-evaluation results revealed significant downregulation in the *SMAD3* gene, which was identified to carry a novel SNP variant (NM_005902.3: c.1155-18N>C) in tumorsphere samples. Additionally, a differential expression increase in the *CDK12* gene, in which a novel SNP (NM_015083.1 (NM_016507.2): c.2149N>G) was identified, was detected in UM-Chor1 tumorspheres (Figure 4.18). Because these detected variants are known to occur in intronic regions of genes, it is not anticipated that they will have a significant impact on gene expression. However, because of their positioning, they might result in intron retention or cassette exon alternative splicing events, depending on their region, which might impact downstream pathways. This detailed research will be the subject of our future work.

4.5. EFFECTS OF ALTERING DIFFERENTIALLY EXPRESSED GENES IN TUMORSPHERES ON VIABILITY, MIGRATION, AND INVASION CAPACITIES OF CHORDOMA CELLS

Expressions of protein-coding *SEMA5A* and *TPD52L1* genes and non-coding *GAPLINC* gene, which were found to be up-regulated in tumorspheres compared to the parental group, were targeted in UM-Chor1 and MUG-Chor1 cells as a result of bioinformatic analysis. Relative expression levels of the *SEMA5A*, *TPD52L1*, and *GAPLINC* genes in chordoma tumorspheres and parental lines were determined by qRT PCR for the confirmation of microarray results (Figure 4.19).

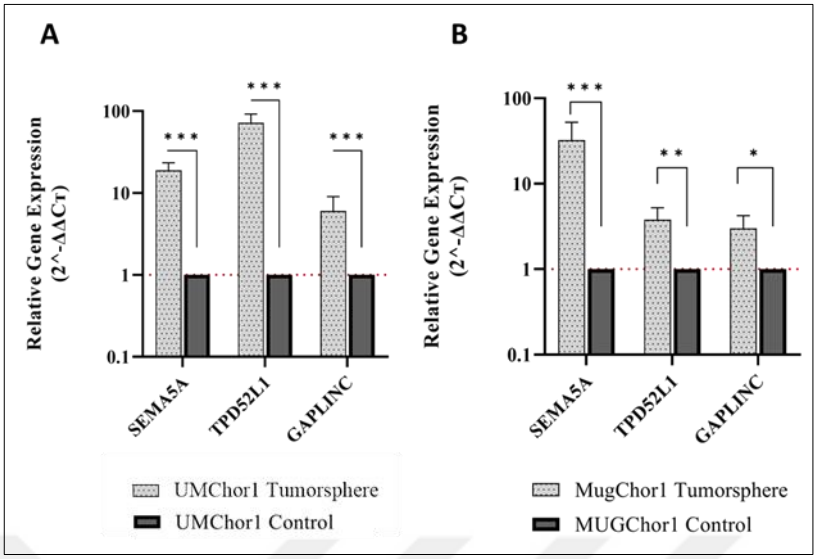

Figure 4.19. Relative gene expression levels of *SEMA5A*, *TPD52L1*, and *GAPLINC* genes in chordoma tumorspheres (A) UM-Chor1, (B) MUG-Chor1 '*' indicates p<0.05, '**' indicates p<0.001, '***' indicates p<0.0001.

Commercially designed targeted siRNAs were used to alter up-regulated expressions of *SEMA5A*, *TPD52L1*, and *GAPLINC* genes.

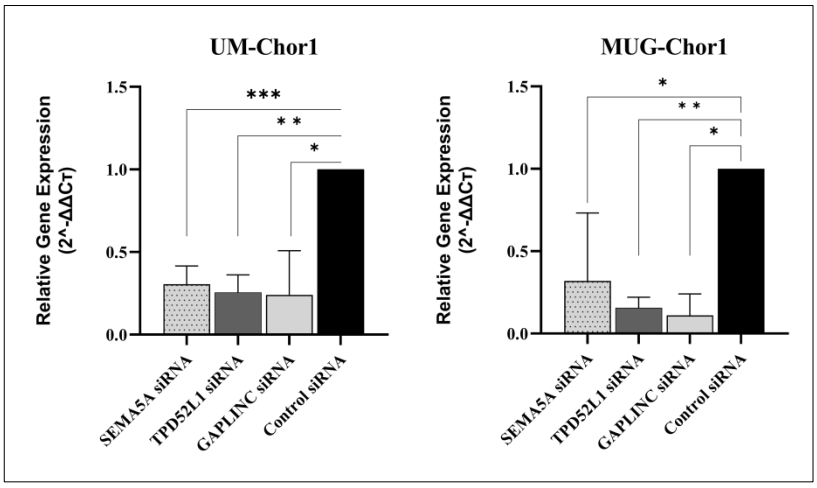

Figure 4.20. Relative gene expression levels of *SEMA5A*, *TPD52L1*, and *GAPLINC* genes after siRNA transfection (A) UM-Chor1, (B) MUG-Chor1 '***' indicates p<0.001.

qPCR results showed that targeted siRNA transfections successfully decreased the expressions of SEME5A, *TPD52L1*, and *GAPLINC* genes (Figure 4.20).

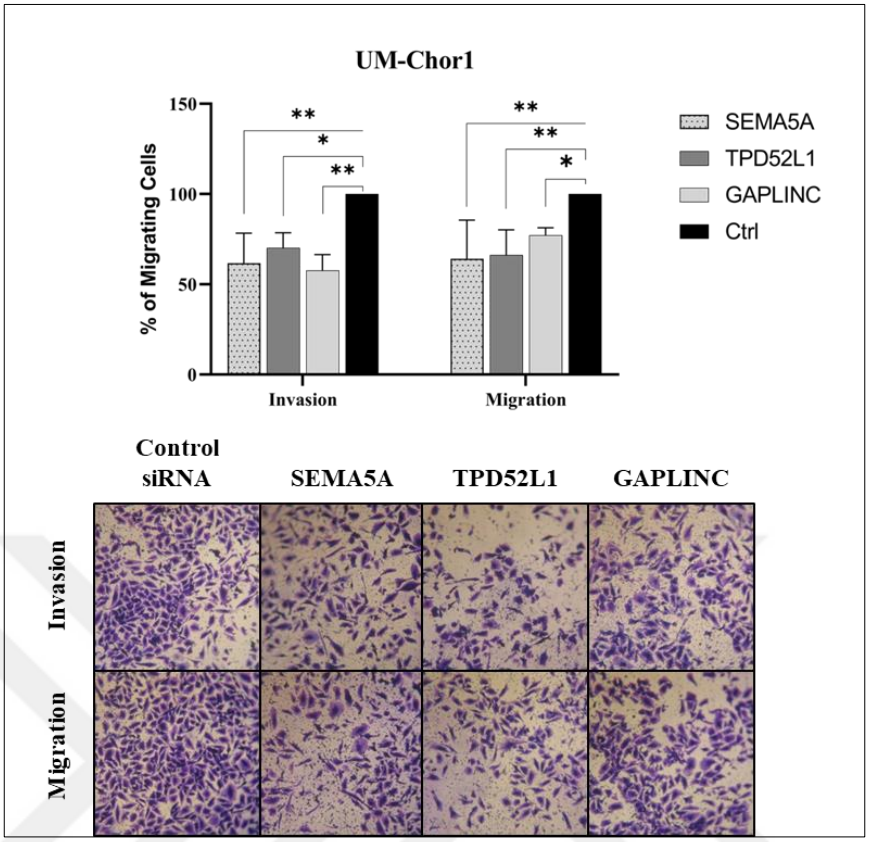

Figure 4.21. The effects of *SEMA5A*, *TPD52L1*, and *GAPLINC* silencing on migration and invasion ability of UM-Chor1 cells '*' indicates $p<0.05$, '**' indicates $p<0.001$

Downstream effects of silencing on migration and invasion capacities of chordoma cells were analyzed via Boyden chamber assay. Results showed that *SEMA5A*, *TPD52L1*, and *GAPLINC* down-regulation significantly inhibited both migration and invasion capacity of the UM-Chor1 cell line (Figure 4.21).

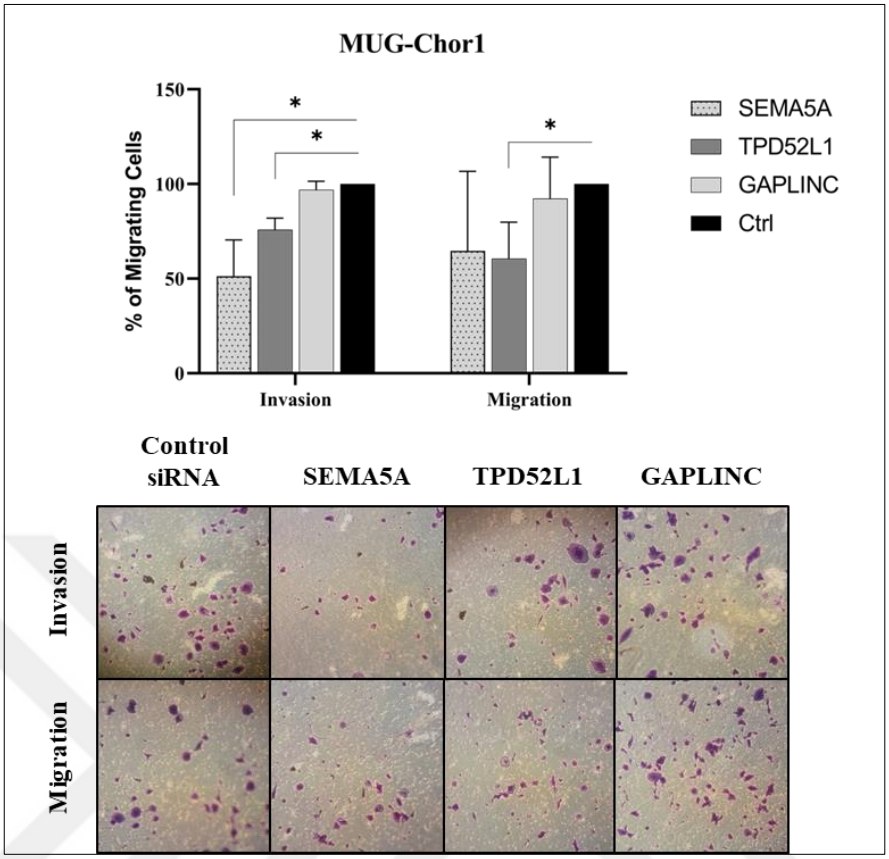

Figure 4.22. The effects of *SEMA5A*, *TPD52L1*, and *GAPLINC* silencing on migration and invasion ability of MUG-Chor1 cells '*' indicates $p<0.05$

Down-regulation of *TPD52L1* expression impaired both the invasion and migration capacities of MUG-Chor1 cells. However, *SEMA5A* inhibition decreased invasion capacity, and *GAPLINC* down-regulation showed no significant effect on either the migration or invasion potential of MUG-Chor1 cells (Figure 4.22).

Later, downstream effects of silencing *SEMA5A*, *TPD52L1*, and *GAPLINC* genes individually on the viability of chordoma cells were analyzed via MTS assay. No discernible difference in cell viability was observed in the UM-Chor1 or MUG-Chor1 cell lines according to MTS results after the *SEMA5A*, *TPD52L1*, and *GAPLINC* genes were silenced. (Figure 4.23).

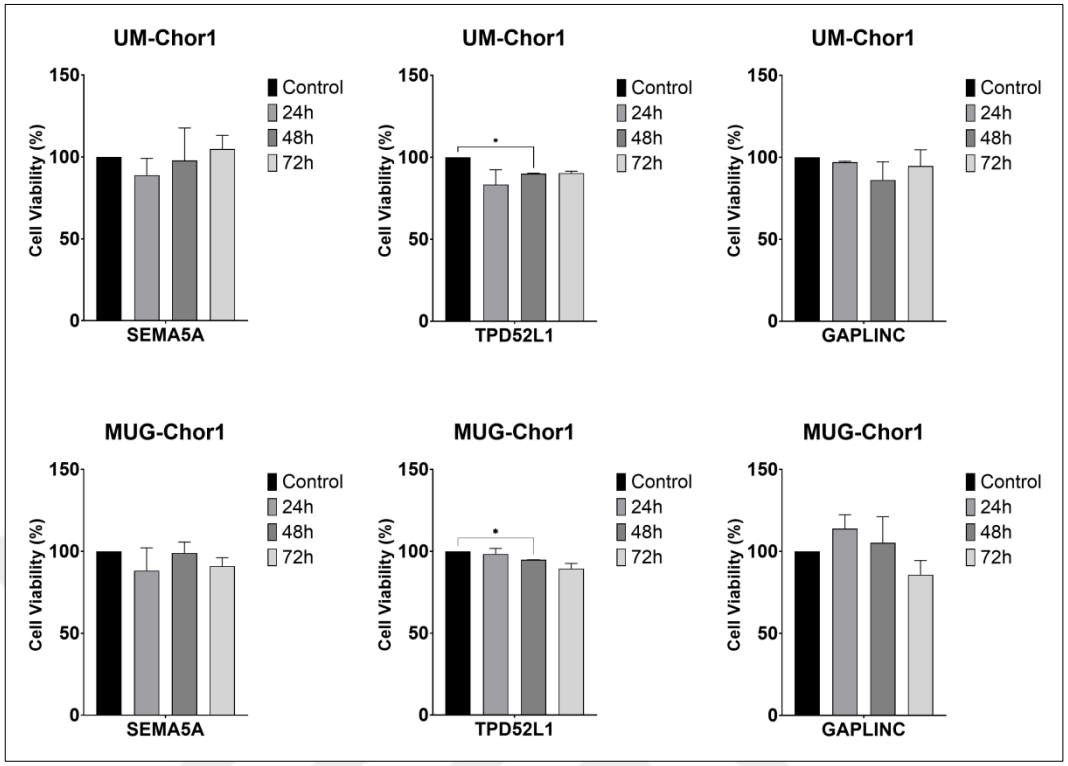

Figure 4.23. The effects of *SEMA5A, TPD52L1*, and *GAPLINC* silencing on the viability of chordoma cells

5. DISCUSSION

In this present dissertation study, prominent genomic variations and transcriptomic differences in chordoma cancer stem cells compared to the general chordoma cell population were investigated. CSCs share common characteristics with embryonic stem cells in terms of both increased expressions of the transcription factors *Nanog, Oct4, c-Myc*, *Sox2,* and *Klf4*, which controls networks regulating pluripotency, as well as features such as self-renewal, proliferation, and differentiation [114]. Researches propose that metastasis in aggressive cancers is initialized by the stimulation of the "cancer type Epithelial-Mesenchymal Transition (EMT, Type 3)"; and the process is driven by cancer stem cell subpopulations in tumors [115–117]. Based on this knowledge, initially, three frequently used cancer stem cell enrichment methods in the literature, side population cells, $CD133^+/CD15^+$ cell populations, and tumorsphere-forming chordoma cells were isolated from 4 different chordoma cells and compared in terms of stemness and EMT-related gene expression changes.

Aydemir et al showed that $CD133^+/CD15^+$ sub-populations in UCH-1 chordoma cell line exhibit stem-like characteristics [118]. Accordingly, $CD133^+/CD15^+$ cells of four different chordoma cell lines were isolated via immunostaining and FACS, and relative stemness and EMT-related gene expression levels in comparison to heterogenous parental cell lines were evaluated. Studies have shown that CSC sub-populations can constitute up to 0.02-25% of the tumor mass, and their abundance in the population may increase especially in undifferentiated cancers [119]. Correlatively, $CD133^+/CD15^+$ dual positive cells are approximately 0.1-0.3% of the general chordoma population.

Relative gene expression profiles of $CD133^+/CD15^+$ chordoma populations cells revealed divergent characteristics for each cell line. Elevated pluripotency indicator Nanog and mesenchymal phenotype marker Twist levels were observed among all $CD133^+/CD15^+$ chordoma cell populations. Similarly, increased Nanog expression in $CD133^+$ enriched cells was also reported in intestinal stem cells, and aggressive endometrial carcinoma samples showed both high *Nanog* and CD133 expression [120,121]. Also, the up-regulation of Twist expression is observed in parallel with the increase in the $CD133^+$ population in triple-negative breast cancer cells [122]. However, there is no direct relation

describing the association between those pluripotency markers. Among 4 cell lines, only $CD133^{+}/CD15^{+}$ subpopulations of U-CH1 and UM-Chor1 exhibit a general increase in stemness-related genes, following the loss of epithelial-cadherin expression. Inhibition of e-cadherin expression is considered an early indicator of the E/M transition, and the initial step of gaining mobility by loss of cell-ECM connections [115]. Accordingly, co-evaluation of the e-cadherin loss and the increase in the levels of Twist, Slug, and Vimentin, which are indicators of mesenchymal phenotype, showed that $CD133^{+}/CD15^{+}$ dual positive U-CH1 cells exhibit a cancer stem-like profile that in line with the literature findings [118].

Partial/hybrid EMT phenotype acquired by the increase in the expression of genes specific to both epithelial and mesenchymal phenotypes provide an advantage since cancer cells of epithelial origin require mesenchymal phenotype to survive in the bloodstream by leaving the primary cancer region, and epithelial character for enhanced metastatic colonization [115,123]. In this context, it is possible to interpret that the MUG-Chor1 $CD133^{+}/CD15^{+}$ cell population with increased expression of both E-cadherin and n-cadherin in addition to mesenchymal markers vimentin, slug, and twist reflects the hybrid EMT character. However, a general downregulation of stemness-related genes in the $CD133^{+}/CD15^{+}$ CH22 cells indicates that $CD133^{+}/CD15^{+}$ markers are not a viable stemness approach for CH22 cells, despite the profile conforming to the EMT phenotype.

In the next phase of the study, chordoma side population cells, as a different CSC enrichment approach, were isolated. The assay is based on the principle that putative stem cell groups can efflux the Hoechst 33342 Dye, which can traverse through the undisrupted cell membrane, through membrane drug transporters [124]. Afterward, chordoma “side population” cells were evaluated for indicators of stemness, and epithelial and mesenchymal phenotype. Similar to $CD133^{+}/CD15^{+}$ dual positive cells of MUG-Chor1, up-regulation of both epithelial and mesenchymal initiators *E-Cad* and *N-Cad*, and the EMT effector genes vimentin, and snail detected in MUG-Chor1 side population cells provide a gene expression profile that can be correlated with the hybrid EMT phenotype [123,125]. Contrary to expectations, CH22 cells with the most “shifting” population in flow cytometry findings compared to verapamil (+) groups do not express an increased stemness profile. Although it has been shown that cancer stem cell profile and cell drug excretion mechanisms are directly related, some studies show side populations do not

always contain stem cell populations [124,126,127]. Some researchers suggest that alternative efflux mechanisms and proteins other than ABC transporters that are responsible for multidrug resistance in stem cells may be the motive behind these different outcomes[128,129].

Tumorsphere culture is an *in vitro* method developed to enrich CSCs with a high stemness feature in an attachment-free and serum-free culture environment that is enriched with supplements that keep cells in an undifferentiated stage [130,131]. Our results showed that sphere formation capacities and spheroid sizes after successive subcultures for each chordoma cell line varied. U-CH1 cells failed to form spheroids after the second sub-culture, and the number of spheroid-forming cells in UM-Chor1 cells was drastically reduced after the second passage; on the other hand, MUG-Chor1 and CH22 cells were found to form tumorspheres even after the third passage. Because of the dramatic reduction in the number of cells capable of forming spheres in U-CH1 and UM-Chor1 cells after the second passage, previous passages of these cells before the potency was decreased were preferred.

The evaluation of the relative gene expression changes of all sphere-derived cells compared to their own 2D cultured (parental) cell controls showed a mean increase in *E-cad* expression in UM-Chor1, MUG-Chor1, and CH22-derived spheres, which is an epithelial phenotype indicator and was expected to be down-regulated in the EMT phenotype. Contrarily, *E-cad* expression appears to be significantly reduced in U-CH1-derived tumorspheres. Previous studies on the sphere formation capacity of breast, ovary, and lung cancer cells reported that *E-cad* expression is a determining factor, and more and larger spheres were observed in cells with E-ad expression, which are consistent with our findings on E-cad expression levels and sphere continuity characteristics of chordoma tumorspheres [132–134].

It is a known fact that the necessary genes for each cancer group may vary because the cancer cell has to be in a dynamic state of change to meet the requirements of various stages of cancer progression, which is a complex and long process [135,136]. Our results revealed that chordoma tumorspheres are generally enriched in stemness markers and all samples are significantly over-expressing EMT-inducing transcription factor *Zeb1*, which is consistent with the general view on cancer stem cells in the literature [114].

Nevertheless, there are several studies reporting cancer stem-like subpopulations that do not exhibit the EMT phenotype. Wu et al showed that *TGF-β1* induced metastasis in cervical cancer cells that promote CSC properties did not alter EMT-related phenotype [137]. Circulating tumor cells (CTCs) as *real-life* metastatic cancer cells are known to endure fluid shear stress (FSS) conditions. The study investigating the effect of FSS on the CSC phenotype of breast cancer cells *in vitro* showed the enrichment of CSC populations without a significant change in EMT expression [138].

Depending on the enrichment method chosen, putative cancer stem cell populations to be detected may yield different outcomes. CTCs of epithelial tumor cells are characterized by ubiquitous epithelial cell adhesion molecule (EpCAM) expression; on the other hand, CTCs having undergone EMT would not be detected using epithelial marker-based methods [139]. In addition to this, some studies indicate that circulating cancer cells prefer to migrate through lymphatic and blood circulation by cohesive epithelial migration instead of following the EMT/MET process [140]. Hence, to overcome this obstacle, recent CTC detection technologies are mainly based on biomarker-mediated isolation or size-sensitive selection techniques [141].

Regarding our findings, tumorsphere cultures were found to be more applicable for all four chordoma cell lines among the three different cancer stem cell enrichment techniques. Thus, the Clariom-D total transcriptome microarray was used in the subsequent stage of the study to identify altered gene expressions in chordoma cell lines-derived tumorspheres in comparison to parental lines. The goal of analyzing transcriptomic changes in chordoma tumorspheres was to identify biomarkers that may be effective in chordoma pathogenesis or genes with regulatory effects that may contribute to the process.

lncRNAs are the most populated class of transcripts in the mammalian genomes and are expressed at more than 50,000 loci in humans, which is almost twice the number of protein-coding genes. However, they are known to be tissue-specific, less conserved evolutionarily, and expressed at significantly lower levels than “protein-coding” mRNAs [142–144]. The most important function of long non-coding RNAs is their active role in the regulation of gene expression, which could act via epigenetic, transcriptional, post-transcriptional, or post-translational mechanisms [145–147]. Non-coding RNAs, which differ significantly in expression compared to parental cells, appear to form a large percentage of the differential transcriptome in chordoma tumorspheres. This recognition

was interpreted as significant differences between tumorspheres and parental cells, not only in terms of protein-coding genes but also in terms of metabolism.

The PI3K-Akt signaling pathway plays an important role in many cellular and pathological mechanisms, such as apoptosis, protein synthesis, metabolism, and cell cycle [148]. The "Focal adhesion: PI3K-Akt-mTOR signaling" WikiPathway, in which many genes are shared with "PI3K-Akt signaling", represents a series of signaling cascades associated with cell-matrix adhesions, such as cell motility, proliferation, differentiation, regulation of gene expression, and survival [149]. The PI3K/Akt/mTOR pathway, whose activation is known to suppress apoptosis and stimulate cell growth and division, is active in many chordomas and is described to have critical importance in pathogenesis [150,151]. Targeting the PI3K/Akt/mTOR pathway in chordomas has been the subject of many clinical and preclinical studies, and promising results have been obtained [150,152–155]. These pathways, which were observed to be affected in chordoma tumorspheres, are interpreted as reflecting the aggressive potential represented by the cancer stem cell profile.

Functional annotation analyses showed that focal adhesion-related cellular components (focal adhesion (GO:0005925) is a cell-substrate junction (GO:0030055)) and KEGG pathways are dysregulated in tumorspheres in comparison to parental controls. Furthermore, WikiPathways enrichment revealed that many genes that regulate focal adhesion are also linked to the deregulation of the PI3K-AKT-mTOR pathway in tumorspheres, and also studies show the active role of PI3K signaling in the regulation of the actin cytoskeleton which might explain our observation of chordoma tumorspheres [141].

Noticeably, while the majority of the differentially expressed genes in “focal adhesion: PI3K-Akt-mTOR-signaling pathway” are up-regulated, many up and down-regulated genes in CH22-derived tumorspheres were found to be close to each other. Nonetheless, the canonical outcome of an up-regulated or down-regulated gene might be specific to a particular pathway and be determined by its inhibitor or activator roles in the pathway. Therefore, focal adhesion and PI3K/Akt signaling mechanisms are affected in chordoma tumorspheres, but predicting the outcomes of this interaction requires further investigation.

The VEGFA-VEGFR2 signaling pathway is in charge of regulating the process of neo-angiogenesis, and also it plays a significant role in the malignant transformation of cancer

and is associated with cell proliferation, migration, and survival [62,156]. Recurrent chordomas have increased VEGFR2 expression, and studies have shown that patients with high VEGFR2 expression have significantly lower 10-year survival rates. However, even when VEGFR2 is not expressed, there are cases with a poor prognosis [157,158]. Our functional enrichment analysis revealed that all chordoma groups had altered VEGFA-VEGFR2 signaling. Notably, *VEGFR2* expression was not directly elevated in the tumorsphere groups, but the majority of the deregulated genes in this pathway were elevated, which might have similar downstream effects in the relevant signaling cascade.

Senescence is a protective negative feedback mechanism that contributes to embryological development, wound healing, and aging in healthy tissues. The mechanism is also known as an autonomous tumor suppressor and can be induced by extensive oncogenic signaling, loss of tumor suppressors, or anticancer therapies [1]. Oncogene-induced senescence is often driven by the inhibitory activity of the RAS/MAPK/PI3K pathways as a negative feedback mechanism [159]. However, recent studies revealed the diverse range of autonomous and non-autonomous features of senescence with tumor suppressor and promoting effects on adjacent cancer cells and other stromal cells in the tumor microenvironment. One of the tumor-promoting features of senescence is the creation of a cancer cell phenotype with transient and reversible cell cycle arrest. Cancer cells in transient senescence can escape from their non-proliferative senescence-associated secretory phenotype (SASP) and exhibit a dormancy state that avoids therapeutic targeting of proliferating cancer cells, which is consistent with our KEGG pathway results showing enriched cellular senescence in chordoma tumorspheres [160].

The next part of the research involved using next-generation DNA sequencing to identify prominent genetic variations detected in chordoma tumorspheres and parental cells. Enrichment analysis of highly mutated genes in most of the tumorsphere samples (10/11) revealed mechanisms related to chordoma pathogenesis such as “PD-L1 expression and PD-1 checkpoint pathway in cancer”, “signaling pathways regulating pluripotency of stem cells”, “EGFR tyrosine kinase inhibitor resistance”, which are also associated with the offerings of many research in chordoma literature, and the cumulative role of our particular gene set should be investigated in details [10,161–165].

A WES study showed *KMT2C* is highly mutated (11 in 27 tumors) in diffuse-type gastric adenocarcinoma patients which indicates worse overall survival and promotes epithelial-

to-mesenchymal transition *in vitro*[166]. Additionally, in prostate cancer, the oncogenic role of KMT2C in tumor growth has been proposed. In our study, among multiple hits, two SNPs in KMT2C (rs10454320, rs2479172) with expected damaging outcomes (polyphen scores are 0.469-tolerated; and 0.999- deleterious respectively) were identified in all parental chordoma cell lines. Furthermore, considering that another missense variant, rs201834857, was identified in U-CH1 spheres but not in parental cells, our data suggest that it is advisable to research the oncogenic function of KMT2C in chordomas.

A comprehensive meta-analysis covering 14358 patients investigating the genetic susceptibility of bone and tissue sarcomas identified that the rs17655 variant in ERCC5 is significantly associated with sarcoma risk, which is also identified in our study in U-CH1-derived tumorsphere samples [167]. Considering another report of copy number loss of ERCC5 in chordoma patients function of ERCC5in chordoma pathogenesis deserves further investigation [168].

Demonstrating that the pathways known to be effective in the pathogenesis of chordoma are also active in chordoma tumor spheres supports our hypothesis. Furthermore, common differentially expressed genes and genomic variants across tumorsphere groups offer the possibility for the potential of chordoma aggressiveness genes that have yet to be discovered. In the final part of the study, three candidate genes that hold promise for a role in chordoma pathogenesis were examined *in vitro* for functional evaluation. The candidate genes were selected based on the similarity of the transcriptomic profile of tumorspheres and chordoma tissue samples.

The tumor protein 52-like 1 (*TPD52L1*, also known as tumor protein D53, TPD53) gene was detected to be differentially increased in all chordoma tumorspheres compared to parental controls and also found up-regulated in chordoma tumor samples in comparison to nucleus pulposus tissue. Increased *TPD52L1* is associated with poor prognosis in colorectal cancer and in vitro studies showed its upregulation promotes proliferation, colony formation, and wound healing [169]. Similarly, *TPD52L1* overexpression was reported in endometrial cancer [170]; co-amplification of *TPD52L1* and *ERBB2*, was proposed as an indicator of poor prognosis in breast cancer [171]. Somatic intrachromosomal rearrangement of *TPD52L1* with ROS1 was associated with drug resistance in lung adenocarcinomas [172]. Additionally, an *in silico* study investigating downstream targets of Brachyury showed *TPD52L1* is a direct target of *Brachyury* [173].In

accordance with the literature, our results showed that inhibiting *TPD52L1* expression *in vitro* significantly reduced migration and invasion capacities of chordoma cell lines UM-Chor1 and MUG-Chor1, but was not found effective on cellular viability. The prognostic role and regulatory impact on apoptosis and cell cycle require further investigation.

The semaphorin family includes the protein-coding gene semaphorin 5A (*SEMA5A*). Members of this protein family aid in axonal guidance during brain development. *SEMA5A* levels play a regulatory role in carcinogenesis and have been linked to several cancer types, including gliomas, colorectal cancer, breast cancer, pancreatic cancer, and melanomas [174–177]. The function of *SEMA5A* on metastasis and invasion does not appear to have the same effect on all types of cancer. Some studies demonstrate that *SEMA5A* expression promotes metastasis and the invasiveness of pancreatic cancer and melanoma cells [174,177]. On the contrary, *SEMA5A* was also found to inhibit migration and invasion in human glioma cells [178]. Our findings demonstrated that inhibition of *SEMA5A* expression significantly impaired both the invasion and migration capacity of UM-Chor1; significantly reduce the invasion capacity of MUG-Chor1 cells. However, *SEMA5A* silencing had a negligible impact on MUG-Chor1 migration and was found ineffective on cell viability.

Gastric Adenocarcinoma-Associated, Positive CD44 Regulator, Long Intergenic Non-Coding RNA, *GAPLINC*, as its name describes, is a non-coding RNA gene. Its role in cancer progression, invasion, migration, and association with poor prognosis of *GAPLINC* has been demonstrated by various studies[179–183]. Total transcriptome microarray results showed *GAPLINC* is differentially up-regulated in all chordoma tumorspheres. The migration and invasion potential of the UM-Chor1 cell line was hampered by the down-regulation of *GAPLINC*. However, inhibiting *GAPLINC* did not alter the migration or invasion of MUG-Chor1 cells and was found ineffective in cell viability. This was thought to be due to the regulatory role of long non-coding RNAs in cell metabolism. The reasons for this unexpected effect should be investigated via possible lncRNA-mRNA activities in chordomas using various bioinformatics approaches.

6. CONCLUSIONS

In this dissertation study, genomic variations and transcriptomic changes in chordoma tumorspheres were investigated in four different chordoma cell lines. Gene set enrichment and overrepresentation analyses have both revealed the similarities of tumorospheres with pathways known to be involved in the pathogenesis of chordoma and suggested new potential relations. Moreover, the downstream effects of silencing the *TPD52L1, SEMA5A*, and *GAPLINC* genes on chordoma migration, invasion, and viability were investigated. Overall results showed that siRNA-mediated silencing of *TPD52L1* significantly impaired both the migration and invasion characteristics of chordoma cell lines. Downregulation of *SEMA5A* inhibited the migration and invasion of UM-Chor1, and it also impaired the invasive capacity of MUG-Chor1 significantly. Reduced *GAPLINC* expression reduced invasion and migration in UM-Chor1, but did not affect MUG-Chor1.

This study is significant in terms of demonstrating the effect of *TPD52L1* and *SEMA5A* genes on chordoma migration and invasion capacity for the first time in the literature. Additionally, the high-throughput DNA-seq and total transcriptome microarray data produced during this study would also add valuable knowledge to the chordoma literature and be a source for upcoming works. In the future, *TPD52L1*, *SEMA5A*, and *GAPLINC* expression could be silenced using stable shRNA transfection or knocked out to investigate the long-term effects. *In vitro* drug resistance and colony formation potentials, as well as *in vivo* tumorigenicity potentials, can be tested, and effects on cell cycle and apoptosis can be monitored. If they are found effective, prognostic value and impact on the survival of chordoma patients can be investigated.

This study was supported by the Scientific and Technological Research Council of Turkey [Grant number 118S691].

REFERENCES

1. Gürcan NE, Uzunkaya B, Bayrak ÖF. Introduction to cancer. In: Autoimmunity and Cancer. 2022. p. 13–39.

2. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: An overview. Int J Cancer. 2021;149(4):778–89.

3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209–49.

4. Yabroff KR, Mariotto A, Tangka F, Zhao J, Islami F, Sung H, et al. Annual Report to the Nation on the Status of Cancer, Part 2: Patient Economic Burden Associated With Cancer Care. J Natl Cancer Inst. 2021 Oct 26;djab192.

5. Ajani JA, Song S, Hochster HS, Steinberg IB. Cancer stem cells: the promise and the potential. Semin Oncol. 2015 Apr;42 Suppl 1:S3-17.

6. Chang JC. Cancer stem cells: Role in tumor growth, recurrence, metastasis, and treatment resistance. Medicine (Baltimore). 2016 Sep;95(1S):S20.

7. Sun HR, Wang S, Yan SC, Zhang Y, Nelson PJ, Jia HL, et al. Therapeutic Strategies Targeting Cancer Stem Cells and Their Microenvironment. Front Oncol [Internet]. 2019 [cited 2022 Oct 30];9. Available from: https://www.frontiersin.org/articles/10.3389/fonc.2019.01104

8. Barber SM, Sadrameli SS, Lee JJ, Fridley JS, Teh BS, Oyelese AA, et al. Chordoma—Current Understanding and Modern Treatment Paradigms. J Clin Med. 2021 Jan;10(5):1054.

9. Gellner V, Tomazic PV, Lohberger B, Meditz K, Heitzer E, Mokry M, et al. Establishment of clival chordoma cell line MUG-CC1 and lymphoblastoid cells as a model for potential new treatment strategies. Sci Rep. 2016 Apr 13;6(1):24195.

10. Meng T, Jin J, Jiang C, Huang R, Yin H, Song D, et al. Molecular Targeted Therapy in the Treatment of Chordoma: A Systematic Review. Front Oncol [Internet]. 2019 [cited 2022 Oct 30];9. Available from: https://www.frontiersin.org/articles/10.3389/fonc.2019.00030

11. Chugh R, Tawbi H, Lucas DR, Biermann JS, Schuetze SM, Baker LH. Chordoma: the nonsarcoma primary bone tumor. The Oncologist. 2007 Nov;12(11):1344–50.

12. Virchow RLK. Untersuchungen über die entwickelung des schädelgrundes im gesunden und krankhaften zustande: und über den einfluss derselben auf schädelform, gesichtsbildung und gehirnbau. G. Reimer; 1857. 158 p.

13. Chambers KJ, Lin DT, Meier J, Remenschneider A, Herr M, Gray ST. Incidence and survival patterns of cranial chordoma in the United States. The Laryngoscope. 2014 May;124(5):1097–102.

14. Campbell RG, Prevedello DM, Ditzel Filho L, Otto BA, Carrau RL. Contemporary management of clival chordomas. Curr Opin Otolaryngol Head Neck Surg. 2015 Apr;23(2):153–61.

15. Healey JH, Lane JM. Chordoma: a critical review of diagnosis and treatment. Orthop Clin North Am. 1989 Jul;20(3):417–26.

16. McMaster ML, Goldstein AM, Bromley CM, Ishibe N, Parry DM. Chordoma: incidence and survival patterns in the United States, 1973-1995. Cancer Causes Control CCC. 2001 Jan;12(1):1–11.

17. Walcott BP, Nahed BV, Mohyeldin A, Coumans JV, Kahle KT, Ferreira MJ. Chordoma: current concepts, management, and future directions. Lancet Oncol. 2012 Feb;13(2):e69-76.

18. Suzuki H, Yamashiro K, Takeda H, Nojima T, Usui M. Extra-axial soft tissue chordoma of wrist. Pathol - Res Pract. 2011 May 15;207(5):327–31.

19. Tirabosco R, Mangham DC, Rosenberg AE, Vujovic S, Bousdras K, Pizzolitto S, et al. Brachyury expression in extra-axial skeletal and soft tissue chordomas: a marker

that distinguishes chordoma from mixed tumor/myoepithelioma/parachordoma in soft tissue. Am J Surg Pathol. 2008 Apr;32(4):572–80.

20. York JE, Kaczaraj A, Abi-Said D, Fuller GN, Skibber JM, Janjan NA, et al. Sacral chordoma: 40-year experience at a major cancer center. Neurosurgery. 1999 Jan;44(1):74–9; discussion 79-80.

21. Bailey CS, Fisher CG, Boyd MC, Dvorak MFS. En bloc marginal excision of a multilevel cervical chordoma. Case report. J Neurosurg Spine. 2006 May;4(5):409–14.

22. Carrabba G, Dehdashti AR, Gentili F. Surgery for clival lesions: open resection versus the expanded endoscopic endonasal approach. Neurosurg Focus. 2008;25(6):E7.

23. Nibu Y, José-Edwards DS, Di Gregorio A. From Notochord Formation to Hereditary Chordoma: The Many Roles of Brachyury [Internet]. BioMed Research International. 2013 [cited 2018 Sep 20]. Available from: https://www.hindawi.com/journals/bmri/2013/826435/

24. Yang Y, Niu X, Li L, Ding Y. The chordoma arised from ilium: A rare case report. J Bone Oncol. 2015 Oct 1;4(4):103–6.

25. Jambhekar NA, Rekhi B, Thorat K, Dikshit R, Agrawal M, Puri A. Revisiting chordoma with brachyury, a “new age” marker: analysis of a validation study on 51 cases. Arch Pathol Lab Med. 2010 Aug;134(8):1181–7.

26. Mori K, Chano T, Kushima R, Hukuda S, Okabe H. Expression of E-cadherin in chordomas: diagnostic marker and possible role of tumor cell affinity. Virchows Arch Int J Pathol. 2002 Feb;440(2):123–7.

27. Laskin WB, Miettinen M. Epithelial-type and neural-type cadherin expression in malignant noncarcinomatous neoplasms with epithelioid features that involve the soft tissues. Arch Pathol Lab Med. 2002 Apr;126(4):425–31.

28. O'Hara BJ, Paetau A, Miettinen M. Keratin subsets and monoclonal antibody HBME-1 in chordoma: immunohistochemical differential diagnosis between tumors simulating chordoma. Hum Pathol. 1998 Feb;29(2):119–26.

29. TBXT T-box transcription factor T [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2022 Nov 14]. Available from: https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=6862

30. Barresi V, Ieni A, Branca G, Tuccari G. Brachyury: a diagnostic marker for the differential diagnosis of chordoma and hemangioblastoma versus neoplastic histological mimickers. Dis Markers. 2014;2014:514753.

31. Bell AH, DeMonte F, Raza SM, Rhines LD, Tatsui CE, Prieto VG, et al. Transcriptome comparison identifies potential biomarkers of spine and skull base chordomas. Virchows Arch Int J Pathol. 2018 Mar;472(3):489–97.

32. Fehling HJ, Lacaud G, Kubo A, Kennedy M, Robertson S, Keller G, et al. Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. Dev Camb Engl. 2003 Sep;130(17):4217–27.

33. Presneau N, Shalaby A, Ye H, Pillay N, Halai D, Idowu B, et al. Role of the transcription factor T (brachyury) in the pathogenesis of sporadic chordoma: a genetic and functional-based study. J Pathol. 2011 Feb;223(3):327–35.

34. Vujovic S, Henderson S, Presneau N, Odell E, Jacques TS, Tirabosco R, et al. Brachyury, a crucial regulator of notochordal development, is a novel biomarker for chordomas. J Pathol. 2006 Jun;209(2):157–65.

35. Schoenfeld AJ, Wang X, Wang Y, Hornicek FJ, Nielsen GP, Duan Z, et al. CSPG4 as a prognostic biomarker in chordoma. Spine J Off J North Am Spine Soc. 2016 Jun;16(6):722–7.

36. Scheil S, Brüderlein S, Liehr T, Starke H, Herms J, Schulte M, et al. Genome-wide analysis of sixteen chordomas by comparative genomic hybridization and cytogenetics of the first human chordoma cell line, U-CH1. Genes Chromosomes Cancer. 2001 Nov;32(3):203–11.

37. Disease models [Internet]. Chordoma Foundation. 2022 [cited 2022 Oct 25]. Available from: https://www.chordomafoundation.org/research/disease-models/

38. Ryland GL, Doyle MA, Goode D, Boyle SE, Choong DYH, Rowley SM, et al. Loss of heterozygosity: what is it good for? BMC Med Genomics. 2015 Aug 1;8(1):45.

39. Nowell PC. Discovery of the Philadelphia chromosome: a personal perspective. J Clin Invest. 2007 Aug 1;117(8):2033–5.

40. Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, et al. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature. 1986 Oct;323(6089):643–6.

41. Kinzler KW, Vogelstein B. Cancer-susceptibility genes. Gatekeepers and caretakers. Nature. 1997 Apr 24;386(6627):761, 763.

42. Vogelstein B, Kinzler KW. Has the breast cancer gene been found? Cell. 1994 Oct 7;79(1):1–3.

43. Langer CJ. Roles of EGFR and KRAS Mutations in the Treatment Of Patients With Non–Small-Cell Lung Cancer. Pharm Ther. 2011 May;36(5):263–79.

44. van Krieken JHJM, Jung A, Kirchner T, Carneiro F, Seruca R, Bosman FT, et al. KRAS mutation testing for predicting response to anti-EGFR therapy for colorectal carcinoma: proposal for an European quality assurance program. Virchows Arch. 2008 Nov 1;453(5):417–31.

45. Knudson AG. Cancer genetics. Am J Med Genet. 2002;111(1):96–102.

46. Tarpey PS, Behjati S, Young MD, Martincorena I, Alexandrov LB, Farndon SJ, et al. The driver landscape of sporadic chordoma. Nat Commun. 2017 Oct 12;8(1):890.

47. Le LP, Nielsen GP, Rosenberg AE, Thomas D, Batten JM, Deshpande V, et al. Recurrent chromosomal copy number alterations in sporadic chordomas. PloS One. 2011;6(5):e18846.

48. Scheil-Bertram S, Kappler R, von Baer A, Hartwig E, Sarkar M, Serra M, et al. Molecular profiling of chordoma. Int J Oncol. 2014 Apr;44(4):1041–55.

49. Bettegowda C, Yip S, Lo SFL, Fisher CG, Boriani S, Rhines LD, et al. Spinal column chordoma: prognostic significance of clinical variables and T (brachyury) gene SNP rs2305089 for local recurrence and overall survival. Neuro-Oncol. 2017 Mar;19(3):405–13.

50. Pillay N, Plagnol V, Tarpey PS, Lobo SB, Presneau N, Szuhai K, et al. A common single-nucleotide variant in T is strongly associated with chordoma. Nat Genet. 2012 Nov;44(11):1185–7.

51. Sa JK, Lee IH, Hong SD, Kong DS, Nam DH. Genomic and transcriptomic characterization of skull base chordoma. Oncotarget. 2017 Jan 3;8(1):1321–8.

52. Dewaele B, Maggiani F, Floris G, Ampe M, Vanspauwen V, Wozniak A, et al. Frequent activation of EGFR in advanced chordomas. Clin Sarcoma Res. 2011 Jul 25;1(1):4.

53. Shalaby A, Presneau N, Ye H, Halai D, Berisha F, Idowu B, et al. The role of epidermal growth factor receptor in chordoma pathogenesis: a potential therapeutic target. J Pathol. 2011 Feb;223(3):336–46.

54. Heery CR, Singh BH, Rauckhorst M, Marté JL, Donahue RN, Grenga I, et al. Phase I Trial of a Yeast-Based Therapeutic Cancer Vaccine (GI-6301) Targeting the Transcription Factor Brachyury. Cancer Immunol Res. 2015 Nov;3(11):1248–56.

55. QUILT-3.011 Phase 2 Yeast-Brachyury Vaccine Chordoma - NCT02383498 - ClinicalTrials.gov [Internet]. [cited 2018 Sep 20]. Available from: https://clinicaltrials.gov/ct2/show/NCT02383498

56. QUILT-3.091 NANT Chordoma Vaccine vs Radiation in Subjects With Unresectable Chordoma. - NCT03647423 - ClinicalTrials.gov [Internet]. [cited 2018 Sep 20]. Available from: https://clinicaltrials.gov/ct2/show/NCT03647423

57. NCI Drug Dictionary-Adenoviral brachyury vaccine ETBX-051 [Internet]. National Cancer Institute. 2011 [cited 2018 Sep 20]. Available from: https://www.cancer.gov/publications/dictionaries/cancer-drug

58. BN Brachyury and Radiation in Chordoma - NCT03595228 - ClinicalTrials.gov [Internet]. [cited 2018 Sep 20]. Available from: https://clinicaltrials.gov/ct2/show/NCT03595228

59. Magnaghi P, Salom B, Cozzi L, Amboldi N, Ballinari D, Tamborini E, et al. Afatinib Is a New Therapeutic Approach in Chordoma with a Unique Ability to Target EGFR and Brachyury. Mol Cancer Ther. 2018 Mar;17(3):603–13.

60. De Los Angeles A, Ferrari F, Xi R, Fujiwara Y, Benvenisty N, Deng H, et al. Hallmarks of pluripotency. Nature. 2015 Sep 24;525(7570):469–78.

61. Garvin K, Feschuk C, Sharp JG, Berger A. Does the number or quality of pluripotent bone marrow stem cells decrease with age? Clin Orthop. 2007 Dec;465:202–7.

62. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000 Jan 7;100(1):57–70.

63. Kreso A, Dick JE. Evolution of the Cancer Stem Cell Model. Cell Stem Cell. 2014 Mar 6;14(3):275–91.

64. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022 Jan 12;12(1):31–46.

65. Friedmann-Morvinski D, Verma IM. Dedifferentiation and reprogramming: origins of cancer stem cells. EMBO Rep. 2014 Mar;15(3):244–53.

66. Yamada Y, Haga H, Yamada Y. Concise review: dedifferentiation meets cancer development: proof of concept for epigenetic cancer. Stem Cells Transl Med. 2014 Oct;3(10):1182–7.

67. Phan TG, Croucher PI. The dormant cancer cell life cycle. Nat Rev Cancer. 2020 Jul;20(7):398–411.

68. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med. 1997 Jul;3(7):730.

69. Zhou HM, Zhang JG, Zhang X, Li Q. Targeting cancer stem cells for reversing therapy resistance: mechanism, signaling, and prospective agents. Signal Transduct Target Ther. 2021 Feb 15;6(1):1–17.

70. Aydemir E, Bayrak OF, Sahin F, Atalay B, Kose GT, Ozen M, et al. Characterization of cancer stem-like cells in chordoma. J Neurosurg. 2012 Jan 27;116(4):810–20.

71. Liu X, Fan D. The epithelial-mesenchymal transition and cancer stem cells: functional and mechanistic links. Curr Pharm Des. 2015;21(10):1279–91.

72. Agiostratidou G, Hulit J, Phillips GR, Hazan RB. Differential cadherin expression: potential markers for epithelial to mesenchymal transformation during tumor progression. J Mammary Gland Biol Neoplasia. 2007 Sep;12(2–3):127–33.

73. Rinner B, Rinner B, Froehlich EV, Froehlich EV, Buerger K, Buerger K, et al. Establishment and detailed functional and molecular genetic characterisation of a novel sacral chordoma cell line, MUG-Chor1. Int J Oncol. 2012 Feb 1;40(2):443–51.

74. Begicevic RR, Falasca M. ABC Transporters in Cancer Stem Cells: Beyond Chemoresistance. Int J Mol Sci [Internet]. 2017 Nov 8 [cited 2019 Jun 3];18(11). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5713331/

75. Gavai AV, Quesnelle C, Norris D, Han WC, Gill P, Shan W, et al. Discovery of Clinical Candidate BMS-906024: A Potent Pan-Notch Inhibitor for the Treatment of Leukemia and Solid Tumors. ACS Med Chem Lett. 2015 May 14;6(5):523–7.

76. Chamie K, Klöpfer P, Bevan P, Störkel S, Said J, Fall B, et al. Carbonic anhydrase-IX score is a novel biomarker that predicts recurrence and survival for high-risk, nonmetastatic renal cell carcinoma: Data from the phase III ARISER clinical trial. Urol Oncol. 2015 May;33(5):204.e25-33.

77. Harris MA, Yang H, Low BE, Mukherjee J, Guha A, Bronson RT, et al. Cancer stem cells are enriched in the side population cells in a mouse model of glioma. Cancer Res. 2008 Dec 15;68(24):10051–9.

78. Dragu DL, Necula LG, Bleotu C, Diaconu CC, Chivu-Economescu M. Therapies targeting cancer stem cells: Current trends and future challenges. World J Stem Cells. 2015 Oct 26;7(9):1185–201.

79. Pastrana E, Silva-Vargas V, Doetsch F. Eyes wide open: a critical review of sphere-formation as an assay for stem cells. Cell Stem Cell. 2011 May 6;8(5):486–98.

80. Shimoda M, Ota M, Okada Y. Isolation of Cancer Stem Cells by Side Population Method. Methods Mol Biol Clifton NJ. 2018;1692:49–59.

81. Johnson S, Chen H, Lo PK. In vitro Tumorsphere Formation Assays. Bio-Protoc [Internet]. 2013 Feb 5 [cited 2019 Jun 1];3(3). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4972326/

82. Lee CH, Yu CC, Wang BY, Chang WW. Tumorsphere as an effective in vitro platform for screening anti-cancer stem cell drugs. Oncotarget. 2016 Jan 12;7(2):1215–26.

83. Paoli P, Giannoni E, Chiarugi P. Anoikis molecular pathways and its role in cancer progression. Biochim Biophys Acta BBA - Mol Cell Res. 2013 Dec 1;1833(12):3481–98.

84. Kim SY, Hong SH, Basse PH, Wu C, Bartlett DL, Kwon YT, et al. Cancer Stem Cells Protect Non-Stem Cells From Anoikis: Bystander Effects. J Cell Biochem. 2016 Oct;117(10):2289–301.

85. Gulluoglu S, Sahin M, Tuysuz EC, Yaltirik CK, Kuskucu A, Ozkan F, et al. Leukemia Inhibitory Factor Promotes Aggressiveness of Chordoma. Oncol Res. 2017 Aug 7;25(7):1177–88.

86. Challen GA, Little MH. A Side Order of Stem Cells: The SP Phenotype. Stem Cells. 2006 Jan 1;24(1):3–12.

87. Yao CJ, Yeh CT, Lee LM, Chuang SE, Yeh CF, Chao WJ, et al. Elimination of Cancer Stem-Like “Side Population” Cells in Hepatoma Cell Lines by Chinese Herbal Mixture “Tien-Hsien Liquid.” Evid Based Complement Alternat Med. 2012 Oct 14;2012:e617085.

88. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinforma Oxf Engl. 2005 Aug 15;21(16):3439–40.

89. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.

90. Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Stat Appl Genet Mol Biol [Internet]. 2004 Feb 12 [cited 2021 Jul 19];3(1). Available from: https://www.degruyter.com/document/doi/10.2202/1544-6115.1027/html

91. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics. 2015 May 22;16(1):169.

92. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.r-project.org/

93. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25–9.

94. Carlson M. org.Hs.eg.db: Genome wide annotation for Human. 2019.

95. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 2021 Jan 8;49(D1):D325–34.

96. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30.

97. Wickham H. ggplot2 [Internet]. Cham: Springer International Publishing; 2016 [cited 2023 Jan 9]. (Use R!). Available from: http://link.springer.com/10.1007/978-3-319-24277-4

98. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation [Internet]. 2021 Aug 28 [cited 2023 Jan 9];2(3). Available from: https://www.cell.com/the-innovation/abstract/S2666-6758(21)00066-7

99. Yu G. enrichplot: Visualization of Functional Enrichment Result. [Internet]. 2022. Available from: https://yulab-smu.top/biomedical-knowledge-mining-book

100. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J Integr Biol. 2012 May;16(5):284–7.

101. Garnier S, Ross N, Rudis boB, Filipovic-Pierucci A, Galili T, timelyportfolio, et al. viridis: Colorblind-Friendly Color Maps for R [Internet]. Zenodo; 2021 [cited 2023 Jan 9]. (sjmgarnier/viridis: viridis 0.6.0 (pre-CRAN release)). Available from: https://zenodo.org/record/4679424

102. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. Nucleic Acids Res. 2018 Jan 4;46(D1):D661–7.

103. Krämer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinforma Oxf Engl. 2014 Feb 15;30(4):523–30.

104. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013 Jan 1;41(D1):D991–5.

105. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002 Jan 1;30(1):207–10.

106. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

107. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019 Aug;37(8):907–15.

108. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021 Feb 16;10(2):giab008.

109. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297–303.

110. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016 Jun 6;17(1):122.

111. Kandoth C, Gao J, qwangmsk, Mattioni M, Struck A, Boursin Y, et al. mskcc/vcf2maf: vcf2maf v1.6.16 [Internet]. Zenodo; 2018 [cited 2022 Oct 9]. Available from: https://zenodo.org/record/1185418

112. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018 Nov;28(11):1747–56.

113. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012 Jul;9(7):671–5.

114. Zhao W, Li Y, Zhang X. Stemness-Related Markers in Cancer. Cancer Transl Med. 2017;3(3):87–95.

115. Chaffer CL, San Juan BP, Lim E, Weinberg RA. EMT, cell plasticity and metastasis. Cancer Metastasis Rev. 2016 Dec;35(4):645–54.

116. Nisticò P, Bissell MJ, Radisky DC. Epithelial-Mesenchymal Transition: General Principles and Pathological Relevance with Special Emphasis on the Role of Matrix Metalloproteinases. Cold Spring Harb Perspect Biol [Internet]. 2012 Feb [cited 2019 Oct 31];4(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3281569/

117. Tanabe S, Quader S, Cabral H, Ono R. Interplay of EMT and CSC in Cancer and the Potential Therapeutic Strategies. Front Pharmacol [Internet]. 2020 [cited 2022 Oct 30];11. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2020.00904

118. Aydemir E, Bayrak OF, Sahin F, Atalay B, Kose GT, Ozen M, et al. Characterization of cancer stem-like cells in chordoma: Laboratory investigation. J Neurosurg. 2012 Apr 1;116(4):810–20.

119. Toledo-Guzmán ME, Bigoni-Ordóñez GD, Ibáñez Hernández M, Ortiz-Sánchez E. Cancer stem cell impact on clinical oncology. World J Stem Cells. 2018 Dec 26;10(12):183–95.

120. Al-Kaabi M, Noel K, Al-Rubai A jabbar. Evaluation of immunohistochemical expression of stem cell markers (NANOG and CD133) in normal, hyperplastic, and malignant endometrium. J Med Life. 2022 Jan;15(1):117–23.

121. Zhao X, Ma L, Dai L, Zuo D, Li X, Zhu H, et al. TNF-α promotes the malignant transformation of intestinal stem cells through the NF-κB and Wnt/β-catenin signaling pathways. Oncol Rep. 2020 Aug;44(2):577.

122. Zhang D, Sun B, Zhao X, Ma Y, Ji R, Gu Q, et al. Twist1 expression induced by sunitinib accelerates tumor cell vasculogenic mimicry by increasing the population of CD133+ cells in triple-negative breast cancer. Mol Cancer. 2014 Sep 8;13:207.

123. Liao TT, Yang MH. Revisiting epithelial-mesenchymal transition in cancer metastasis: the connection between epithelial plasticity and stemness. Mol Oncol. 2017 Jul;11(7):792–804.

124. Golebiewska A, Brons NHC, Bjerkvig R, Niclou SP. Critical appraisal of the side population assay in stem cell and cancer stem cell research. Cell Stem Cell. 2011 Feb 4;8(2):136–47.

125. Usman S, Waseem NH, Nguyen TKN, Mohsin S, Jamal A, Teh MT, et al. Vimentin Is at the Heart of Epithelial Mesenchymal Transition (EMT) Mediated Metastasis. Cancers. 2021 Jan;13(19):4985.

126. Goodell MA, Rosenzweig M, Kim H, Marks DF, DeMaria M, Paradis G, et al. Dye efflux studies suggest that hematopoietic stem cells expressing low or undetectable levels of CD34 antigen exist in multiple species. Nat Med. 1997 Dec;3(12):1337–45.

127. Mouthon MA, Fouchet P, Mathieu C, Sii-Felice K, Etienne O, Lages CS, et al. Neural stem cells from mouse forebrain are contained in a population distinct from the ‘side population.’ J Neurochem. 2006;99(3):807–17.

128. Moitra K, Lou H, Dean M. Multidrug Efflux Pumps and Cancer Stem Cells: Insights Into Multidrug Resistance and Therapeutic Development. Clin Pharmacol Ther. 2011;89(4):491–502.

129. Richard V, Nair MG, Santhosh Kumar TR, Pillai MR. Side Population Cells as Prototype of Chemoresistant, Tumor-Initiating Cells. BioMed Res Int. 2013 Nov 4;2013:e517237.

130. Deshmukh AP, den Hollander P, Kuburich NA, Vasaikar S, Joseph R, Mani SA. Enrichment of Cancer Stem Cells in a Tumorsphere Assay. In: Kannan N, Beer P, editors. Stem Cell Assays: Methods and Protocols [Internet]. New York, NY: Springer US; 2022 [cited 2022 Oct 30]. p. 501–7. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-0716-1979-7_34

131. García-Rocha R, Monroy-García A, Carrera-Martínez M, Hernández-Montes J, Don-López CA, Weiss-Steider B, et al. Evidence that cervical cancer cells cultured as tumorspheres maintain high CD73 expression and increase their protumor characteristics through TGF-β production. Cell Biochem Funct. 2022 Oct;40(7):760–72.

132. Manuel Iglesias J, Beloqui I, Garcia-Garcia F, Leis O, Vazquez-Martin A, Eguiara A, et al. Mammosphere formation in breast carcinoma cell lines depends upon expression of E-cadherin. PloS One. 2013;8(10):e77281.

133. Powan P, Luanpitpong S, He X, Rojanasakul Y, Chanvorachote P. Detachment-induced E-cadherin expression promotes 3D tumor spheroid formation but inhibits tumor formation and metastasis of lung cancer cells. Am J Physiol Cell Physiol. 2017 Nov 1;313(5):C556–66.

134. Xu S, Yang Y, Dong L, Qiu W, Yang L, Wang X, et al. Construction and characteristics of an E-cadherin-related three-dimensional suspension growth model of ovarian cancer. Sci Rep. 2014 Jul 10;4:5646.

135. Lai X, Li Q, Wu F, Lin J, Chen J, Zheng H, et al. Epithelial-Mesenchymal Transition and Metabolic Switching in Cancer: Lessons From Somatic Cell Reprogramming. Front Cell Dev Biol [Internet]. 2020 [cited 2022 Oct 31];8. Available from: https://www.frontiersin.org/articles/10.3389/fcell.2020.00760

136. Song J, Wang W, Wang Y, Qin Y, Wang Y, Zhou J, et al. Epithelial-mesenchymal transition markers screened in a cell-based model and validated in lung adenocarcinoma. BMC Cancer. 2019 Jul 11;19(1):680.

137. Wu L, Han L, Zhou C, Wei W, Chen X, Yi H, et al. TGF-β1-induced CK17 enhances cancer stem cell-like properties rather than EMT in promoting cervical cancer metastasis via the ERK1/2-MZF1 signaling pathway. FEBS J. 2017;284(18):3000–17.

138. Triantafillu UL, Park S, Klaassen NL, Raddatz AD, Kim Y. Fluid shear stress induces cancer stem cell-like phenotype in MCF7 breast cancer cell line without inducing epithelial to mesenchymal transition. Int J Oncol. 2017 Mar;50(3):993–1001.

139. Raimondi C, Nicolazzo C, Gradilone A. Circulating tumor cells isolation: the "post-EpCAM era." Chin J Cancer Res. 2015 Oct;27(5):461–70.

140. Hamilton G, Rath B. Mesenchymal-Epithelial Transition and Circulating Tumor Cells in Small Cell Lung Cancer. In: Magbanua MJM, Park JW, editors. Isolation and Molecular Characterization of Circulating Tumor Cells [Internet]. Cham: Springer International Publishing; 2017 [cited 2023 Jan 7]. p. 229–45. (Advances in Experimental Medicine and Biology). Available from: https://doi.org/10.1007/978-3-319-55947-6_12

141. Deng Z, Wu S, Wang Y, Shi D. Circulating tumor cell isolation for cancer diagnosis and prognosis. eBioMedicine [Internet]. 2022 Sep 1 [cited 2023 Jan 7];83. Available from: https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(22)00419-4/fulltext

142. Antonov I, Medvedeva Y. Direct Interactions with Nascent Transcripts Is Potentially a Common Targeting Mechanism of Long Non-Coding RNAs. Genes. 2020 Dec 10;11(12):1483.

143. Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJL, Gough J, et al. An atlas of human long non-coding RNAs with accurate 5′ ends. Nature. 2017 Mar 9;543(7644):199–204.

144. Ulitsky I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nat Rev Genet. 2016 Oct;17(10):601–14.

145. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep. 2015 May 19;11(7):1110–22.

146. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015 Mar;47(3):199–208.

147. Sui Y, Han Y, Zhao X, Li D, Li G. Long non-coding RNA Irm enhances myogenic differentiation by interacting with MEF2D. Cell Death Dis. 2019 Feb 21;10(3):181.

148. PI3K-Akt signaling pathway (Homo sapiens) - WikiPathways [Internet]. [cited 2022 Nov 1]. Available from: https://www.wikipathways.org/index.php/Pathway:WP4172

149. Focal adhesion: PI3K-Akt-mTOR-signaling pathway (Homo sapiens) - WikiPathways [Internet]. [cited 2022 Nov 1]. Available from: https://www.wikipathways.org/index.php/Pathway:WP3932

150. Schwab J, Antonescu C, Boland P, Healey J, Rosenberg A, Nielsen P, et al. Combination of PI3K/mTOR inhibition demonstrates efficacy in human chordoma. Anticancer Res. 2009 Jun;29(6):1867–71.

151. Williams BJ, Raper DMS, Godbout E, Bourne TD, Prevedello DM, Kassam AB, et al. Diagnosis and treatment of chordoma. J Natl Compr Cancer Netw JNCCN. 2013 Jun 1;11(6):726–31.

152. Davies JM, Robinson AE, Cowdrey C, Mummaneni PV, Ducker GS, Shokat KM, et al. Generation of a patient-derived chordoma xenograft and characterization of the phosphoproteome in a recurrent chordoma. J Neurosurg. 2014 Feb;120(2):331–6.

153. Han S, Polizzano C, Nielsen GP, Hornicek FJ, Rosenberg AE, Ramesh V. Aberrant hyperactivation of akt and Mammalian target of rapamycin complex 1 signaling in sporadic chordomas. Clin Cancer Res Off J Am Assoc Cancer Res. 2009 Mar 15;15(6):1940–6.

154. Lebellec L, Chauffert B, Blay JY, Le Cesne A, Chevreau C, Bompas E, et al. Advanced chordoma treated by first-line molecular targeted therapies: Outcomes and prognostic factors. A retrospective study of the French Sarcoma Group (GSF/GETO) and the Association des Neuro-Oncologues d'Expression Française (ANOCEF). Eur J Cancer Oxf Engl 1990. 2017 Jul;79:119–28.

155. Stacchiotti S, Morosi C, Lo Vullo S, Casale A, Palassini E, Frezza AM, et al. Imatinib and everolimus in patients with progressing advanced chordoma: A phase 2 clinical study. Cancer. 2018 Oct 15;124(20):4056–63.

156. Koch S, Claesson-Welsh L. Signal transduction by vascular endothelial growth factor receptors. Cold Spring Harb Perspect Med. 2012 Jul;2(7):a006502.

157. Akhavan-Sigari R, Gaab MR, Rohde V, Abili M, Ostertag H. Prognostic significance of immunohistochemical expression of VEGFR2 and iNOS in spinal chordoma. Eur Spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc. 2014 Nov;23(11):2416–22.

158. Morimoto Y, Tamura R, Ohara K, Kosugi K, Oishi Y, Kuranari Y, et al. Prognostic significance of VEGF receptors expression on the tumor cells in skull base chordoma. J Neurooncol. 2019 Aug;144(1):65–77.

159. Hoare M, Narita M. The Power Behind the Throne: Senescence and the Hallmarks of Cancer. Annu Rev Cancer Biol. 2018;2(1):175–94.

160. De Blander H, Morel AP, Senaratne AP, Ouzounova M, Puisieux A. Cellular Plasticity: A Route to Senescence Exit and Tumorigenesis. Cancers. 2021 Sep 11;13(18):4561.

161. Scheipl S, Barnard M, Cottone L, Jorgensen M, Drewry DH, Zuercher WJ, et al. EGFR inhibitors identified as a potential treatment for chordoma in a focused compound screen. J Pathol. 2016 Jul;239(3):320–34.

162. Scheipl S, Barnard M, Lohberger B, Zettl R, Brcic I, Liegl-Atzwanger B, et al. Drug combination screening as a translational approach toward an improved drug therapy for chordoma. Cell Oncol Dordr. 2021 Dec;44(6):1231–42.

163. Feng Y, Shen J, Gao Y, Liao Y, Cote G, Choy E, et al. Expression of programmed cell death ligand 1 (PD-L1) and prevalence of tumor-infiltrating lymphocytes (TILs) in chordoma. Oncotarget. 2015 May 10;6(13):11139–49.

164. Mathios D, Ruzevick J, Jackson CM, Xu H, Shah SR, Taube JM, et al. PD-1, PD-L1, PD-L2 expression in the chordoma microenvironment. J Neurooncol. 2015 Jan;121(2):251–9.

165. Zou MX, Peng AB, Lv GH, Wang XB, Li J, She XL, et al. Expression of programmed death-1 ligand (PD-L1) in tumor-infiltrating lymphocytes is associated with favorable spinal chordoma prognosis. Am J Transl Res. 2016;8(7):3274–87.

166. Cho SJ, Yoon C, Lee JH, Chang KK, Lin J xian, Kim YH, et al. KMT2C mutations in diffuse type gastric adenocarcinoma promote epithelial-to-mesenchymal transition. Clin Cancer Res Off J Am Assoc Cancer Res. 2018 Dec 15;24(24):6556–69.

167. Benna C, Simioni A, Pasquali S, De Boni D, Rajendran S, Spiro G, et al. Genetic susceptibility to bone and soft tissue sarcomas: a field synopsis and meta-analysis. Oncotarget. 2018 Apr 6;9(26):18607–26.

168. Rinner B, Weinhaeusel A, Lohberger B, Froehlich EV, Pulverer W, Fischer C, et al. Chordoma Characterization of Significant Changes of the DNA Methylation Pattern. PLoS ONE. 2013 Mar 22;8(3):e56609.

169. Hong Q, Li B, Cai X, Lv Z, Cai S, Zhong Y, et al. Transcriptomic Analyses of the Adenoma-Carcinoma Sequence Identify Hallmarks Associated With the Onset of Colorectal Cancer. Front Oncol. 2021;11:704531.

170. Mamoor S. Over-expression of tumor protein D52-like 1 in human endometrial cancer. [Internet]. OSF Preprints; 2021 [cited 2023 Jan 15]. Available from: https://osf.io/ba9jy/

171. Kim HY, Choi HJ, Lee JY, Kong G. Cancer Target Gene Screening: a web application for breast cancer target gene screening using multi-omics data analysis. Brief Bioinform. 2020 Mar 23;21(2):663–75.

172. Xu C, Li D, Duan W, Tao M. TPD52L1-ROS1 Rearrangement as a New Acquired Resistance Mechanism to Osimertinib That Responds to Crizotinib in Combination With Osimertinib in Lung Adenocarcinoma. JTO Clin Res Rep. 2020 Mar 14;1(2):100034.

173. Lolas M, Valenzuela PDT, Tjian R, Liu Z. Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. Proc Natl Acad Sci U S A. 2014 Mar 25;111(12):4478–83.

174. D'Aguanno S, Valentini E, Tupone MG, Desideri M, Di Martile M, Spagnuolo M, et al. Semaphorin 5A drives melanoma progression: role of Bcl-2, miR-204 and c-Myb. J Exp Clin Cancer Res CR. 2018 Nov 19;37(1):278.

175. Ko PH, Lenka G, Chen YA, Chuang EY, Tsai MH, Sher YP, et al. Semaphorin 5A suppresses the proliferation and migration of lung adenocarcinoma cells. Int J Oncol. 2020 Jan;56(1):165–77.

176. Li X, Law JWS, Lee AYW. Semaphorin 5A and plexin-B3 regulate human glioma cell motility and morphology through Rac1 and the actin cytoskeleton. Oncogene. 2012 Feb 2;31(5):595–610.

177. Sadanandam A, Varney ML, Singh S, Ashour AE, Moniaux N, Deb S, et al. High Gene Expression of Semaphorin 5A in Pancreatic Cancer is associated with Tumor Growth, Invasion and Metastasis. Int J Cancer J Int Cancer. 2010 Sep 1;127(6):1373–83.

178. Li X, Lee AYW. Semaphorin 5A and plexin-B3 inhibit human glioma cell motility through RhoGDIalpha-mediated inactivation of Rac1 GTPase. J Biol Chem. 2010 Oct 15;285(42):32436–45.

179. Chen HH, Zong J, Wang SJ. LncRNA GAPLINC promotes the growth and metastasis of glioblastoma by sponging miR-331-3p. Eur Rev Med Pharmacol Sci. 2019 Jan;23(1):262–70.

180. Liao S, Zhou S, Wang C. GAPLINC is a predictor of poor prognosis and regulates cell migration and invasion in osteosarcoma. Biosci Rep. 2018 Oct 31;38(5):BSR20181171.

181. Wang S, Yang X, Xie W, Fu S, Chen Q, Li Z, et al. LncRNA GAPLINC Promotes Renal Cell Cancer Tumorigenesis by Targeting the miR-135b-5p/CSF1 Axis. Front Oncol. 2021;11:718532.

182. Zhang W, Yao H, Wu Y. Poor expression of long-chain noncoding RNA GAPLINC inhibits epithelial-mesenchymal transition, and invasion and migration of hepatocellular carcinoma cells. Anticancer Drugs. 2019 Sep;30(8):784–94.

183. Zhao J, Wang C, Liu S, Su X, Ouyang A. TGF-β1 mediates lncRNA GAPLINC expression to promote the migration and invasion of non-small cell lung cancer. OncoTargets Ther. 2019;12:6175–80.

APPENDIX A: ONCOPLUS PANEL CONTENT

Table A.1-A.3 gives detailed lists of genes with coding regions, intronic regions, and microsatellite markers respectively that are presented in the OncoPlus tumor panel.

Table A.1. Genes with full coding regions covered in the OncoPlus panel

ABL1	*ABL2*	*ACVR1*	*ACVR1B*	*AGO2*	*AKT1*	*AKT2*
AKT3	*ALK*	*ALOX12B*	*AMER1*	*ANKRD11*	*APC*	*AR*
ARAF	*ARFRP1*	*ARID1A*	*ARID1B*	*ARID2*	*ARID5B*	*ASXL1*
ASXL2	*ATM*	*ATR*	*ATRX*	*AURKA*	*AURKB*	*AXIN1*
AXIN2	*AXL*	*B2M*	*BABAM1*	*BAP1*	*BARD1*	*BBC3*
BCL10	*BCL2*	*BCL2L1*	*BCL2L11*	*BCL2L2*	*BCL6*	*BCOR*
BCORL1	*BIRC3*	*BLM*	*BMPR1A*	*BRAF*	*BRCA1*	*BRCA2*
BRD4	*BRIP1*	*BTG1*	*BTG2*	*BTK*	*C11orf30*	*CALR*
CARD11	*CARM1*	*CASP8*	*CBFB*	*CBL*	*CCND1*	*CCND2*
CCND3	*CCNE1*	*CD22*	*CD274*	*CD276*	*CD70*	*CD79A*
CD79B	*CDC42*	*CDC73*	*CDH1*	*CDK12*	*CDK4*	*CDK6*
CDK8	*CDKN1A*	*CDKN1B*	*CDKN2A*	*CDKN2B*	*CDKN2C*	*CEBPA*
CENPA	*CHD2*	*CHD4*	*CHEK1*	*CHEK2*	*CIC*	*CREBBP*
CRKL	*CRLF2*	*CSDE1*	*CSF1R*	*CSF3R*	*CTCF*	*CTLA4*
CTNNA1	*CTNNB1*	*CUL3*	*CUL4A*	*CXCR4*	*CYLD*	*CYP17A1*
CYSLTR2	*DAXX*	*DCUN1D1*	*DDR1*	*DDR2*	*DICER1*	*DIS3*
DNAJB1	*DNMT1*	*DNMT3A*	*DNMT3B*	*DOT1L*	*DROSHA*	*DUSP4*
E2F3	*EED*	*EGFL7*	*EGFR*	*EIF1AX*	*EIF4A2*	*EIF4E*
ELF3	*EP300*	*EPAS1*	*EPCAM*	*EPHA3*	*EPHA5*	*EPHA7*
EPHB1	*EPHB4*	*ERBB2*	*ERBB3*	*ERBB4*	*ERCC1*	*ERCC2*
ERCC3	*ERCC4*	*ERCC5*	*ERF*	*ERG*	*ERRFI1*	*ESR1*
ETV1	*ETV6*	*EZH1*	*EZH2*	*FAM175A*	*FAM46C*	*FAM58A*
FANCA	*FANCC*	*FANCD2*	*FANCE*	*FANCF*	*FANCG*	*FANCL*
FAS	*FAT1*	*FBXW7*	*FGF10*	*FGF12*	*FGF14*	*FGF19*

FGF23	*FGF3*	*FGF4*	*FGF6*	*FGFR1*	*FGFR2*	*FGFR3*
FGFR4	*FH*	*FLCN*	*FLT1*	*FLT3*	*FLT4*	*FOXA1*
FOXL2	*FOXO1*	*FOXP1*	*FRS2*	*FUBP1*	*FYN*	*GABRA6*
GATA1	*GATA2*	*GATA3*	*GATA4*	*GATA6*	*GID4*	*GLI1*
GNA11	*GNA13*	*GNAQ*	*GNAS*	*GPR124*	*GPS2*	*GREM1*
GRIN2A	*GRM3*	*GSK3B*	*H3F3A*	*H3F3B*	*H3F3C*	*HDAC1*
HGF	*HIST1H1C*	*HIST1H2BD*	*HIST1H3A*	*HIST1H3B*	*HIST1H3C*	*HIST1H3D*
HIST1H3E	*HIST1H3F*	*HIST1H3G*	*HIST1H3H*	*HIST1H3I*	*HIST1H3J*	*HIST2H3C*
HIST2H3D	*HIST3H3*	*HLA-A*	*HLA-B*	*HLA-C*	*HLA-DMA*	*HLA-DMB*
HLA-DOA	*HLA-DOB*	*HLA-DPA1*	*HLA-DPB1*	*HLA-DQA1*	*HLA-DQA2*	*HLA-DQB1*
HLA-DQB2	*HLA-DRA*	*HLA-DRB1*	*HLA-DRB3*	*HLA-DRB4*	*HLA-DRB5*	*HNF1A*
HOXB13	*HRAS*	*HSD3B1*	*HSP90AA1*	*ICOSLG*	*ID3*	*IDH1*
IDH2	*IFNGR1*	*IGF1*	*IGF1R*	*IGF2*	*IKBKE*	*IKZF1*
IL10	*IL7R*	*INHA*	*INHBA*	*INPP4A*	*INPP4B*	*INPPL1*
INSR	*IRF2*	*IRF4*	*IRS1*	*IRS2*	*JAK1*	*JAK2*
JAK3	*JUN*	*KAT6A*	*KDM5A*	*KDM5C*	*KDM6A*	*KDR*
KEAP1	*KEL*	*KIT*	*KLF4*	*KLHL6*	*KMT2A*	*KMT2B*
KMT2C	*KMT2D*	*KNSTRN*	*KRAS*	*LATS1*	*LATS2*	*LMO1*
LRP1B	*LTK*	*LYN*	*LZTR1*	*MAF*	*MAGI2*	*MALT1*
MAP2K1	*MAP2K2*	*MAP2K4*	*MAP3K1*	*MAP3K13*	*MAP3K14*	*MAPK1*
MAPK3	*MAPKAP1*	*MAX*	*MCL1*	*MDC1*	*MDM2*	*MDM4*
MED12	*MEF2B*	*MEN1*	*MERTK*	*MET*	*MGA*	*MITF*
MKNK1	*MLH1*	*MPL*	*MRE11A*	*MSH2*	*MSH3*	*MSH6*
MSI1	*MSI2*	*MST1*	*MST1R*	*MTAP*	*MTOR*	*MUTYH*
MYC	*MYCL*	*MYCN*	*MYD88*	*MYOD1*	*NBN*	*NCOA3*
NCOR1	*NEGR1*	*NF1*	*NF2*	*NFE2L2*	*NFKBIA*	*NKX2-1*
NKX3-1	*NOTCH1*	*NOTCH2*	*NOTCH3*	*NOTCH4*	*NPM1*	*NRAS*
NRG1	*NSD1*	*NT5C2*	*NTHL1*	*NTRK1*	*NTRK2*	*NTRK3*
NUF2	*NUP93*	*P2RY8*	*PAK1*	*PAK3*	*PAK7*	*PALB2*
PARK2	*PARP1*	*PARP2*	*PARP3*	*PAX5*	*PBRM1*	*PDCD1*
PDCD1LG2	*PDGFRA*	*PDGFRB*	*PDK1*	*PDPK1*	*PGR*	*PHOX2B*
PIK3C2B	*PIK3C2G*	*PIK3C3*	*PIK3CA*	*PIK3CB*	*PIK3CD*	*PIK3CG*

PIK3R1	*PIK3R2*	*PIK3R3*	*PIM1*	*PLCG2*	*PLK2*	*PMAIP1*
PMS1	*PMS2*	*PNRC1*	*POLD1*	*POLE*	*PPARG*	*PPM1D*
PPP2R1A	*PPP2R2A*	*PPP4R2*	*PPP6C*	*PRDM1*	*PRDM14*	*PREX2*
PRKAR1A	*PRKCI*	*PRKD1*	*PRKDC*	*PRSS8*	*PTCH1*	*PTEN*
PTP4A1	*PTPN11*	*PTPRD*	*PTPRO*	*PTPRS*	*PTPRT*	*QKI*
RAB35	*RAC1*	*RAC2*	*RAD21*	*RAD50*	*RAD51*	*RAD51B*
RAD51C	*RAD51D*	*RAD52*	*RAD54L*	*RAF1*	*RANBP2*	*RARA*
RASA1	*RB1*	*RBM10*	*RECQL*	*RECQL4*	*REL*	*RET*
RFWD2	*RHEB*	*RHOA*	*RICTOR*	*RIT1*	*RNF43*	*ROS1*
RPS6KA4	*RPS6KB2*	*RPTOR*	*RRAGC*	*RRAS*	*RRAS2*	*RTEL1*
RUNX1	*RUNX1T1*	*RXRA*	*RYBP*	*SDHA*	*SDHAF2*	*SDHB*
SDHC	*SDHD*	*SESN1*	*SESN2*	*SESN3*	*SETD2*	*SETD8*
SF3B1	*SGK1*	*SH2B3*	*SH2D1A*	*SHOC2*	*SHQ1*	*SLIT2*
SLX4	*SMAD2*	*SMAD3*	*SMAD4*	*SMARCA4*	*SMARCB1*	*SMARCD1*
SMO	*SMYD3*	*SNCAIP*	*SOCS1*	*SOS1*	*SOX10*	*SOX17*
SOX2	*SOX9*	*SPEN*	*SPOP*	*SPRED1*	*SPTA1*	*SRC*
SRSF2	*STAG2*	*STAT3*	*STAT4*	*STAT5A*	*STAT5B*	*STK11*
STK19	*STK40*	*SUFU*	*SUZ12*	*SYK*	*TAF1*	*TAP1*
TAP2	*TBX3*	*TCEB1*	*TCF3*	*TCF7L2*	*TEK*	*TERC*
TERT	*TET1*	*TET2*	*TGFBR1*	*TGFBR2*	*TIPARP*	*TMEM127*
TMPRSS2	*TNFAIP3*	*TNFRSF14*	*TOP1*	*TOP2A*	*TP53*	*TP53BP1*
TP63	*TRAF2*	*TRAF7*	*TSC1*	*TSC2*	*TSHR*	*TYRO3*
U2AF1	*UPF1*	*VEGFA*	*VHL*	*VTCN1*	*WHSC1*	*WHSC1L1*
WISP3	*WT1*	*WWTR1*	*XIAP*	*XPO1*	*XRCC2*	*YAP1*
YES1	*ZBTB2*	*ZFHX3*	*ZNF217*	*ZNF703*		

Table A. 2. Intronic regions covered in the OncoPlus panel

ALK intron 18 - 19	*BCL2* 3'UTR	*BCR* intron 8,13 - 14	*BRAF* intron 7-10	*BRCA1* intron2, 7-8,12,16,19 - 20	*BRCA2* intron 2
CD74 intron 6-8	*EGFR* intron 7, 15, 24-27	*ETV4* intron 5 - 6	*ETV5* intron 6 - 7	*ETV6* intron 5 - 6	*EWSR1* intron 6-13
EZR intron 9-12	*FGFR1* intron 1,5,17	*FGFR2* intron 1,17	*FGFR3* intron 17	*FLI1* intron 3-8	*KIT* intron 16
KMT2A intron 6-11	*MET* intron 1,14	*MSH2* intron 5	*MYB* intron 14	*MYC* intron 1	*NOTCH2* intron 26
NTRK1 intron 8-10	*NTRK2* intron 12,15	*NTRK3* intron 13 - 14	*NUTM1* intron 1	*PDGFB* intron 1	*PDGFRA* intron 7,9,11
RAF1 intron 4-9	*RARA* intron 2	*RET* intron 7-11	*ROS1* intron 31-35	*RSPO2* Upstream, 5'UTR, exon 1- 2, intron 1	*SDC4* intron 2
SLC34A2 intron 4	*TMPRSS2* intron 1-3				

Table A. 3. Microsatellite markers covered in the OncoPlus panel

BAT-25	*BAT-26*	*BAT-40*	*BAT-RII*	*NR-21*	*NR-22*
NR-24	*NR-27*	*MONO-27*	*D2S123*	*D5S346*	*D17S261*
D17S520	*D18S34*				