

Whole Exome Sequencing Reveals Potential Single Nucleotide Polymorphisms and Copy Number Variations in 26 Sporadic Patients with Immature Teratoma

Yakun Liu, Yajing Jia, Naiyi Du, Yongping Li, Chunyang Wang , Shan Kang

Department of Gynaecology, The Fourth Hospital of Hebei Medical University, Shijiazhuang, Hebei, People's Republic of China

Correspondence: Shan Kang, Email kangshan@hebmh.edu.cn

Background: Immature teratoma is a malignant tumor and accounts for 1%–3% of ovary teratoma. The objective of this study was to find the potential single nucleotide polymorphisms (SNPs), copy number variations (CNVs) and drug targets in the immature teratoma.

Material and Methods: Genomic DNA was extracted from peripheral blood of 26 sporadic immature teratoma patients for whole exome sequencing (WES) analysis to detect germline variants. SNPs and CNVs were identified, followed by the analysis of physical/chemical properties and conserved domain of mutated genes. Functional enrichment and protein–protein interaction (PPI) analysis of mutated genes was performed, followed by pharmacogenomic analysis. Finally, SNPs, CNVs and mRNA expression of mutated genes were investigated in ovarian cancer by using the online databases.

Results: A total of 24 common mutated genes were identified in 26 patients. Among which, 5 common mutated genes with common mutation sites were identified, including 2 frameshift mutant genes (MYPOP and FRG2C) and 3 nonsynonymous mutant genes (CNTNAP3, GPRIN2 and MUC3A). After mutation, molecular weight of MYPOP, FRG2C, CNTNAP3, GPRIN2 and MUC3A changed slightly. In the PPI network, MUC12 (with the highest degree) interacted with GALNT12, MUC3A and FCGBP. Based on the pharmacogenomic analysis, MUC2 was predicted to be a potential drug target of CHEMBL35482, FLUOROURACIL and DECITABINE. According to the functional analysis, MUC3A, MUC2 and MUC12 were involved in biological processes of activation of the innate immune response. The mutation frequency of FCGBP and CNTNAP3 was rare and had a higher amplification frequency in ovarian cancer. In addition, 2 common CNVs (deletion state) were screened out, which involving 6 genes, such as RP11.

Conclusion: This study identified some potential SNPs and CNVs, which may contribute to clarifying the pathogenesis of immature teratoma and provide potential biomarkers and drug targets for this disease.

Keywords: immature teratoma, whole exome sequencing, single nucleotide polymorphism, copy number variation, biological process, diagnose, drug target

Introduction

Teratomas, a subtype of germ cell tumors, originate from 3 germ cell layers containing different types of tissue (such as hair, bone, and muscles).^{1,2} Histopathologically, teratomas are divided into mature teratoma and immature teratoma.³ Immature teratoma is a rare and malignant ovarian germ cell tumors related to a variable low potential of distant metastasis, which depends on the histological grade.⁴ The lungs, liver and brain are the most distant sites of blood-borne metastasis of immature teratoma.⁵ In addition, rare sites of distant spread include abdominal wall musculature and soft tissue of the thigh.^{6,7} Clinically, immature teratoma is a curable disease. Adjuvant treatment with etoposide, bleomycin and cisplatin is the standard in stages greater than stage IA grade I. In addition, operative intervention combined with chemotherapy is required for advanced stage immature teratoma patients.⁴ Interestingly, recurrent immature teratoma is known to undergo pathological retroconversion to contain mature elements following chemotherapeutic treatment, which is a unique characteristic of immature teratoma.⁵

There are many causes of immature teratoma. It is suggested that copy-neutral loss of heterozygosity (resulting from meiotic errors) could be sufficient to generate immature teratomas from germ cells, which are characterized by extensive allelic imbalances, copy number alterations and a paucity of somatic mutations.⁸ Currently, most studies on immature teratoma are retrospective clinical analyses, whereas studies focusing on its genetic and molecular mechanism remain limited. Therefore, investigations of single nucleotide polymorphisms (SNPs), copy number variations (CNVs) and potential drug targets may become key directions in future research. Meanwhile, whole exome sequencing (WES) has been increasingly applied in clinical genetic diagnosis and genomic research. Peripheral blood is a commonly used sample in WES and is easy to obtain. Tumor tissue from immature teratoma is clinically limited and difficult to obtain routinely. Furthermore, several studies on ovarian-related tumors have also been conducted using peripheral blood samples.^{9,10} In view of this, peripheral blood samples from 26 sporadic immature teratoma patients were collected for WES to identify new mutated genes, which may provide clues for the study of function and pathogenesis of immature teratoma.

Materials and Methods

Patients

In the present study, a total of 26 sporadic patients with immature teratoma were enrolled. The peripheral blood samples of these individuals were collected for whole exome sequencing (WES) analysis. Detailed clinical information of 26 patients is shown in Table 1. This study was approved by the ethics committee of the fourth hospital of hebei medical university. Written consent was obtained from enrolled individuals.

Table 1 Clinical Information of 26 Sporadic Patients with Immature Teratoma

Number	Age (Years Old)	Pelvic Mass	Abdominal Pain	Abdominal Distension	Adnexal Mass	Frequent Urination	Excision	Stage
1	30	Yes	No	Yes	No	No	Yes	III
2	46	No	Yes	No	No	Yes	No	II
3	44	No	Yes	No	No	Yes	No	II
4	21	No	No	No	Yes	No	Yes	IA
5	20	No	No	Yes	Yes	No	No	IA
6	77	No	No	Yes	No	No	No	IA
7	30	No	No	No	No	No	Yes	IA
8	73	No	No	No	Yes	No	Yes	IA
9	56	No	No	Yes	No	No	No	IIIC
10	13	Yes	No	No	No	No	Yes	IIB
11	28	No	No	No	Yes	No	Yes	IIIC
12	34	No	Yes	Yes	Yes	No	No	IA
13	28	No	No	No	Yes	No	No	IA
14	24	Yes	No	No	No	No	No	IA
15	35	No	No	No	No	No	Yes	ICI
16	28	No	Yes	Yes	No	No	No	IA
17	34	No	No	No	No	No	Yes	IA
18	15	Yes	No	No	No	No	No	IA
19	16	No	No	No	Yes	No	No	ICI
20	19	No	No	No	No	No	No	IA
21	23	Yes	No	No	No	No	No	IA
22	24	Yes	No	No	No	No	No	IA
23	39	No	No	No	No	Yes	No	IIIC
24	24	No	No	No	Yes	No	No	IA
25	25	Yes	No	No	No	No	No	IC
26	22	No	No	No	No	No	No	IA

Extraction of Genomic DNA

Genomic DNA was extracted from peripheral blood samples. The DNA concentration was determined by microplate analyzer. The integrity and purity of DNA were determined by 1% of agarose gel. 1 µg of genomic DNA was interrupted to 250–300bp by ultrasound and purified by AxyPrep Mag PCR Clean up Kit. The terminal repair reaction system was purified after a period of moderate temperature reaction. The reaction system with “A” at the end and joint connection was prepared. A certain amount of polymerase chain reaction (PCR) products were hybridized and hybridized with Agilent SureSelect Hybridization and Wash Kits. The eluted product was amplified and purified by the AxyPrep Mag PCR clean up Kit. Constructed libraries were tested for quality using Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System. Qualified libraries were sequenced on Xten platform using 150PE sequencing.

WES Analysis

The original image data obtained by WES was transformed to sequence data by base calling, which is called raw data or raw reads. Raw reads were quality-controlled with fastQC and SeqPrep. The joint sequence in reads and the reads that have no inserted fragment (due to self-connection of the joint) was removed. The bases with low quality (mass value <20) at the end of the sequence (3' end) were trimmed. Reads with an N ratio of more than 10% were removed. The adapter and sequences whose length is less than 20bp after quality trimming was discarded. The high-quality sequencing sequences (obtained after quality control) were compared with the designated reference genome (version: hg19) by BWA software (<https://github.com/lh3/bwa>). Picard software was used to remove repeated sequences. Through GATK software, the mismatch regions caused by indel were recalculated by BaseRecalibrator module. In addition, base quality values of reads in the BAM file were recalibrated by RealignerTargetCreator module.

Identification of SNPs, Indel and CNVs

SNP is a phenomenon in which most nucleotide sequences of the same chromosome or the same site in different individuals are identical, but only one base is different. Indel is an insertion or deletion of multiple bases. Deletions or insertion at coding sites or splicing sites may alter protein translation. Strelka software was used for mutation detection of SNP and indel. ANNOVAR and tapes software were used to annotate mutation loci, including gene structure, genomic feature, the known mutation database, function of mutation-related gene and hazard prediction of non-synonymous mutation. Refseq and Gencode were used to annotate the gene structure of the mutated loci. The SIFT, PolyPhen, MutationAssessor, LRT and other methods were used to comprehensively evaluate the impact of non-synonymous mutations on the disease. The dbSNP database, SNP database of 1000 genome, cosmic database (involved knowing tumor somatic mutation) and ESP6500 mutation database were used for annotation. Candidate mutations were identified by eliminating mutations with a mutation frequency ≥ 0.01 in 1000 Genome or ExAC databases, synonymous mutations and mutations located in introns. In the hazard prediction, those mutations predicted to be benign/possibly benign were further filtered from ClinVar, InterVar, SIFT and Polyphen databases. In addition, in order to reduce the false positive rate of CNVs, fragments with length ≤ 300 bp were filtered out. High-quality CNVs were filtered through annotation information (prediction quality score/default prediction score > 40). To improve the accuracy of the results, only those genes in the CNVs region were retained.

In Silico Analysis of Mutated Genes

The physical/chemical properties of amino acid after gene mutation were predicted by using the online database protparam (<http://web.expasy.org/protparam/>). Additionally, conserved domain analysis after gene mutation was performed through the Conserved Domain Search Service (<https://www.ncbi.nlm.nih.gov/cdd/?term>).

Functional Enrichment and Protein–Protein Interaction (PPI) Analysis of Mutated Genes

In order to investigate the function of identified mutation genes, Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed. False discovery rate (FDR) <0.05 was the threshold of significantly enriched GO and KEGG terms. Additionally, STRING (<https://string-db.org/>) software was used to perform PPI analysis on the proteins encoded by mutated genes.

Pharmacogenomic Analysis of Mutated Genes

Drug-Gene Interaction Database (DGIdb) is a database of drug–gene interactions that provides information about the association of genes with known or potential drugs. It has more than 14,000 drug–gene interactions involving 2,600 genes and 6,300 drugs that target these genes. It also includes 6,700 other genes. All these genes are likely to be considered potential drug targets. In this study, the interactions between mutated genes and drugs were searched from the database.

Investigation of SNPs and CNVs of Mutated Genes in Ovarian Cancer

In order to explore the SNPs and CNVs of mutated genes in ovarian cancer, SNPs mutation data (involving 620 samples) and CNVs mutation data (involving 436 samples) of mutated genes in ovarian cancer were downloaded from the Cancer Genome Atlas (TCGA) dataset (<https://gdc.xenahubs.net>). In the SNP data of ovarian cancer, the default parameters of the “maftools” in the R package were used to analyze the gene mutations. In addition, the amplification and deletion frequency of mutated genes were calculated based on CNVs data of ovarian cancer.

MRNA Expression of Mutated Genes in Ovarian Cancer and Yolk Sac Tumor

To investigate additionally the mRNA expression of mutated genes in other reproductive-related tumors, GSE26712 and GSE10615 datasets were downloaded from the Gene Expression Omnibus database (GEO) database. GSE26712 dataset and GSE10615 dataset were used the mRNA expression of mutated genes in ovarian cancer and yolk sac tumor, respectively. For the GSE26712 dataset, the Wilcoxon test was used to analyze the statistical significance of differential expression. For the GSE10615 dataset, differential expression analysis was performed using the “limma” in the R package, and the results were visualized with the “ggplot2” in the R package. The mRNA expression of mutated genes was presented by box plot.

Results

Identification of SNPs, Indels and CNVs

After the screening, information of SNPs and indels of 26 sporadic patients with immature teratoma was obtained. Total variation type, type of base change and top 20 mutant genes is shown in Figure 1. Nonsynonymous SNP was the major variation type (Figure 1A). C>T and G>A were the major base change types in SNP class (Figure 1B). It is noted that top 20 mutant genes were identified (Figure 1C), such as mucin 3A, cell surface associated (MUC3A), G protein regulated inducer of neurite outgrowth 2 (GPRIN2), mucin 2, oligomeric mucus/gel-forming (MUC2), Fc gamma binding protein (FCGBP), FSHD region gene 2 family member C (FRG2C). Interestingly, a total of 24 common mutated genes were identified in 26 sporadic immature teratoma patients. Among which, a total of 5 common mutated genes with common mutation sites were identified, including contacting associated protein family member 3 (CNTNAP3), Myb-related transcription factor, partner of profiling (MYPOP), FRG2C, GPRIN2 and MUC3A (Table 2). Mutation type of FRG2C and MYPOP was frameshift_deletion. Mutation type of CNTNAP3, GPRIN2 and MUC3A was nonsynonymous_SNP. In addition, 2 common CNVs (deletion state) were screened out (Table 3), which involving 6 genes, such as pre-mRNA processing factor 31 (RP11, also called PRPF31).

In Silico Analysis of Mutated Genes

The physical/chemical properties of amino acid and conserved domain after mutations of CNTNAP3, FRG2C, GPRIN2, MUC3A and MYPOP were analyzed. Amino acid number, molecular weight and theoretical PI of CNTNAP3 are 1127,

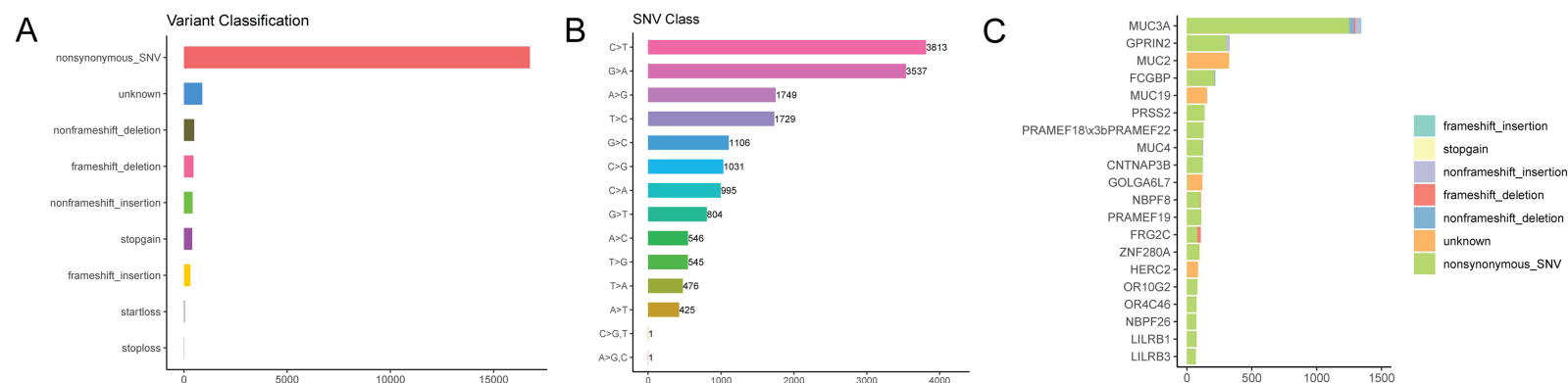


Figure I Total variation type (A), type of base change (B) and top 20 mutant genes (C) in 26 sporadic patients with immature teratoma.

Table 2 5 Common Mutated Genes with Common Mutation Sites in 26 Sporadic Immature Teratoma Patients

Reference Gene	Nucleotide Change	AA Change
CNTNAP3	Chr9- 39088492: T >A	T1051S
FRG2C	Chr3- 75665673: AG >A	R161Sfs*5
GPRIN2	Chr10- 46549538: A >G	L400P
MUC3A	Chr7- 100958510: C>G	T2244S
MUC3A	Chr7- 100957778: C >A	T2000N
MUC3A	Chr7- 100965345: C > A	A3149E
MUC3A	Chr7- 100957777: A >G	T2000A
MYPOP	Chr19- 45890723: C >CG	R367Pfs*83

Table 3 2 Common CNVs in 26 Sporadic Immature Teratoma Patients

Chr	Start_End	Length	Gene	CNV_state
4	111,463,907-111,542,525	78618	PITX2:P ENPEP:P RPRI1-380D23.2:F	Deletion
17	15,546,035-15,603,091	57056	RPRI1-640I15.1:F TRIM16:P ZNF286A:P ZNF29P:F RPRI1-385D13.1:P	Deletion

124,115.80 and 6.67, respectively. In amino acid composition, Leu (9.80%) content is the highest. The total number of negatively and positively charged residues is 107 and 101, respectively. The extinction coefficient is 1.324/1.306. The instability index, fat index and the average hydrophilic are 37.61, 75.09 and -0.281, respectively. After mutation, the molecular weight (124101.78) decreased slightly, and the instability index changed to 38.39. The mutant amino acid is located in the Laminin_G_2 domain (Figure 2A). Amino acid number, molecular weight and theoretical PI of FRG2C are 282, 30,798.11 and 7.09, respectively. In amino acid composition, Ser (12.4%) content is the highest. The total number of negatively and positively charged residues is 35 and 35, respectively. The extinction coefficient is 0.571/0.551. The

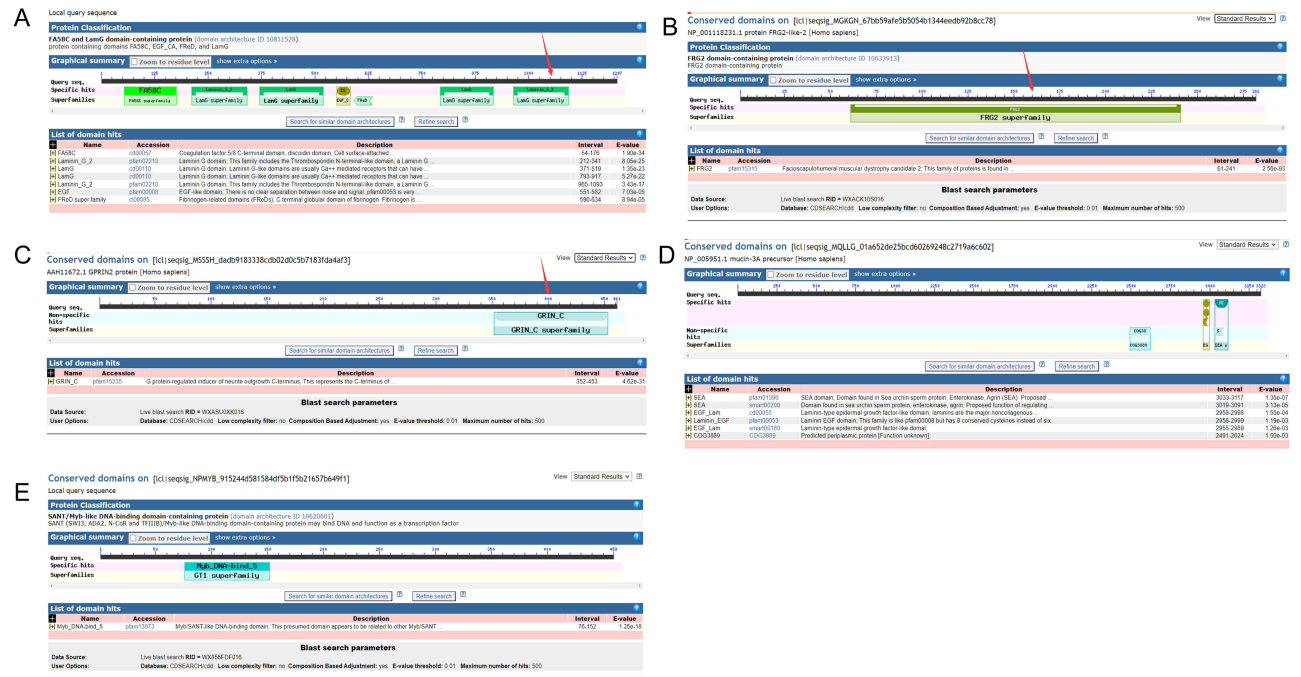


Figure 2 Conserved domain analysis of mutated CNTNAP3 (A), FRG2C (B), GPRIN2 (C), MUC3A (D) and MYPOP (E) in 26 sporadic patients with immature teratoma. The red arrow in (A–C) represents the location of the mutant amino acid within the conserved domain.

instability index, fat index and the average hydrophilic are 58.06, 54.40 and -0.886 , respectively. After mutation, molecular weight (30729.00) and theoretical PI (6.75) decreased, and instability index became 57.19. The mutant amino acid was located in the FGR2 domain (Figure 2B). Amino acid number, molecular weight and theoretical PI of GPRIN2 are 458, 47,450.29 and 6.28, respectively. In amino acid composition, Ala (13.80%) content is the highest. The total number of negatively and positively charged residues is 43 and 38, respectively. The extinction coefficient is 1.090/1.075. The instability index, fat index and the average hydrophilic are 63.51, 70.24 and -0.320 , respectively. After mutation, the molecular weight (47434.25) decreased slightly, and the instability index became 63.89. The mutant amino acid was located in the GRIN_C domain (Figure 2C). Amino acid number, molecular weight and theoretical PI of MUC3A are 3323, 345,126.77 and 5.14, respectively. In amino acid composition, Thr (29.40%) content is the highest. The total number of negatively and positively charged residues is 171 and 91, respectively. The extinction coefficient is 0.504/0.500. The instability index, fat index and the average hydrophilic are 43.76, 54.97 and -0.217 , respectively. After mutation, the molecular weight (345154.58) increased slightly. The mutant amino acid was not in its conserved domain (Figure 2D). Amino acid number, molecular weight and theoretical PI of MYPOP are 399, 42,508.45 and 10.08, respectively. In amino acid composition, Pro (20.80%) content is the highest. The total number of negatively and positively charged residues is 39 and 53, respectively. The extinction coefficient is 0.755/0.752. The instability index, fat index and the average hydrophilic are 94.21, 66.17 and -0.677 , respectively. After mutation, the molecular weight (345154.58) increased slightly. The mutant amino acid was not in its conserved domain (Figure 2E).

Functional Enrichment and PPI Analysis of Mutated Genes

Based on GO analysis of 24 mutated genes, golgi lumen was the only one significantly enriched cytological component. A remarkably enriched biological process of mutated genes (involved MUC3A, MUC2 and MUC12) was identified, such as activation of the innate immune response. It is a pity that no enriched molecular function of mutated genes was found. According to KEGG analysis of 24 mutated genes, amoebiasis and gastric cancer were the only two significantly enriched signaling pathways of mutated genes. All significantly enriched biological processes, cytological components and signaling pathways are listed in Table 4. In addition, PPI network was constructed based on 24 mutation genes (Figure 3). The highest degree gene was MUC12. Moreover, MUC12 interacted with polypeptide N-acetylgalactosaminyltransferase 12 (GALNT12), MUC3A and FCGBP.

Pharmacogenomic Analysis of Mutated Genes

Based on the drugs-24 mutated genes interaction analysis, only compounds of CHEMBL35482, FLUOROURACIL, and DECITABINE interacted with MUC2. It is a pity that there was no interaction between drugs and the remaining 23 mutated genes. This suggested that MUC2 could be considered as a potential molecular target of CHEMBL35482, FLUOROURACIL, and DECITABINE in the therapy of immature teratoma.

Investigation of SNPs and CNVs of Mutated Genes in Ovarian Cancer

In order to explore the SNPs and CNVs of mutated genes in ovarian cancer, SNPs mutation data (involving 620 samples) and CNVs mutation data (involving 436 samples) of mutated genes in ovarian cancer were downloaded from TCGA dataset. In the SNPs data of ovarian cancer, 9 genes mutations were analyzed through “maftools” in the R package, including FCGBP, CNTNAP3, GPRIN2, MUC3A, GALNT12, MUC12, FRG2C, MYPOP and MUC2 (Figure 4A). The mutation frequency of FCGBP and CNTNAP3 was, respectively, 2% and 1%, which indicated rare mutation rate in ovarian cancer. In addition, the amplification (gain) and deletion (loss) frequency of FCGBP, CNTNAP3, GPRIN2, MUC3A, GALNT12, MUC12, FRG2C and MYPOP were calculated based on CNVs data of ovarian cancer (Figure 4B). FCGBP, CNTNAP3, GPRIN2, MUC3A and MUC12 had higher amplification frequency of CNV. FRG2C, GALNT12 and MYPOP had higher deletion frequency of CNV.

MRNA Expression of Mutated Genes in Ovarian Cancer and Yolk Sac Tumor

To investigate additionally the mRNA expression of mutated genes in other reproductive-related tumors, GSE26712 dataset and GSE10615 dataset were used to detect mRNA expression of mutated genes in ovarian cancer (Figure 5A) and yolk sac tumor (Figure 5B), respectively. GALNT12 was significantly down-regulated in ovarian cancer compared with

Table 4 Functional Enrichment Analysis of 24 Mutated Genes in 26 Sporadic Immature Teratoma Patients

	ID	Description	P value	FDR	Gene ID	Count
BP	GO:0016266	O-glycan processing	1.12E-06	3.01E-05	MUC3A/MUC2/MUC12	3
	GO:0006493	Protein O-linked glycosylation	5.77E-06	4.54E-05	MUC3A/MUC2/MUC12	3
	GO:0002223	Stimulatory C-type lectin receptor signaling pathway	7.39E-06	4.54E-05	MUC3A/MUC2/MUC12	3
	GO:0002220	Innate immune response activating cell surface receptor signaling pathway	8.20E-06	4.54E-05	MUC3A/MUC2/MUC12	3
	GO:0002758	Innate immune response-activating signal transduction	8.41E-06	4.54E-05	MUC3A/MUC2/MUC12	3
	GO:0002218	Activation of innate immune response	1.62E-05	7.29E-05	MUC3A/MUC2/MUC12	3
	GO:0045089	Positive regulation of innate immune response	5.51E-05	0.000212538	MUC3A/MUC2/MUC12	3
	GO:0006486	Protein glycosylation	7.74E-05	0.000220905	MUC3A/MUC2/MUC12	3
	GO:0043413	Macromolecule glycosylation	7.74E-05	0.000220905	MUC3A/MUC2/MUC12	3
	GO:0002833	Positive regulation of response to biotic stimulus	8.50E-05	0.000220905	MUC3A/MUC2/MUC12	3
	GO:0070085	Glycosylation	9.00E-05	0.000220905	MUC3A/MUC2/MUC12	3
	GO:0045088	Regulation of innate immune response	0.000153637	0.000345683	MUC3A/MUC2/MUC12	3
	GO:0009101	Glycoprotein biosynthetic process	0.000204504	0.000424738	MUC3A/MUC2/MUC12	3
	GO:0031349	Positive regulation of defense response	0.000229828	0.000443239	MUC3A/MUC2/MUC12	3
	GO:0002831	Regulation of response to biotic stimulus	0.000358931	0.000609945	MUC3A/MUC2/MUC12	3
	GO:0009100	Glycoprotein metabolic process	0.000361449	0.000609945	MUC3A/MUC2/MUC12	3
	GO:0002429	Immune response-activating cell surface receptor signaling pathway	0.000534341	0.000801511	MUC3A/MUC2/MUC12	3
	GO:0002757	Immune response-activating signal transduction	0.000534341	0.000801511	MUC3A/MUC2/MUC12	3
	GO:0030277	Maintenance of gastrointestinal epithelium	0.00813734	0.011563589	MUC2	1
	GO:0010669	Epithelial structure maintenance	0.010714562	0.014464659	MUC2	1
	GO:0022600	Digestive system process	0.03617272	0.046507782	MUC2	1
CC KEGG	GO:0005796	Golgi lumen	2.22E-05	4.43E-05	MUC3A/MUC2/MUC12	3
	hsa05146	Amoebiasis	0.012573964	0.01836785	MUC2	1
	hsa05226	Gastric cancer	0.01836785	0.01836785	MUC2	1

Abbreviations: BP, biological processes; CC, cytological component; KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, false discovery rate.

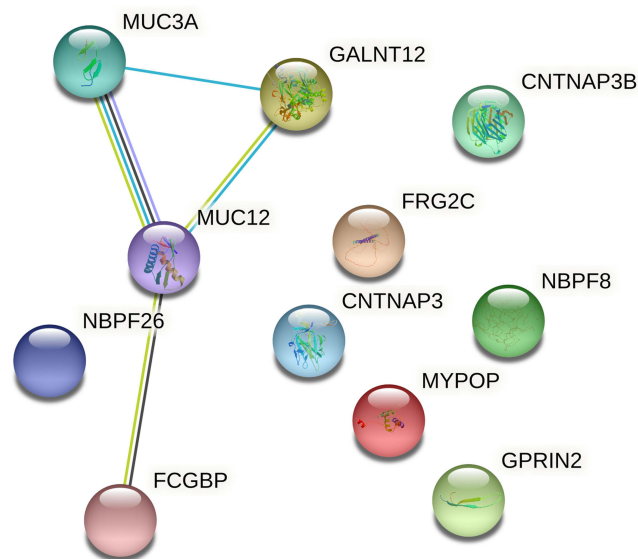


Figure 3 PPI network in 26 sporadic patients with immature teratoma. Light blue lines: Known interactions derived from curated databases.; Yellow lines: Associations supported by textmining evidence; Black lines: Associations derived from gene co-expression data; Light purple lines: Associations based on protein homology evidence. Nodes are connected if at least one line of evidence supports a functional or physical association between the corresponding proteins; the absence of an edge between two nodes indicates that no significant evidence for such an association was found in the analyzed database resources.

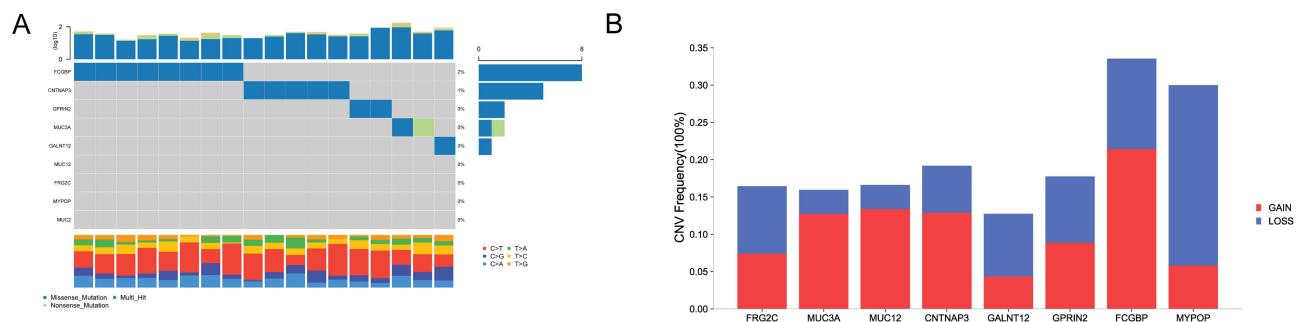


Figure 4 Investigation of SNP (A) and CNV (B) of mutated genes in ovarian cancer.

para-carcinoma tissue. MUC2 had an up-regulated tendency, and FCGBP and GPRIN2 had a down-regulated tendency in ovarian cancer without statistical difference. In addition, GALNT12, MUC2, FCGBP and GPRIN2 had an up-regulated tendency in yolk sac tumor compared with seminoma.

Discussion

Previous multiregion exome sequencing study focused on somatic mutations and clonal origins in tumor tissues of immature teratoma.⁸ However, germline variations from peripheral blood remain unclear. Our study was designed to fill this research gap. In 26 sporadic patients with immature teratoma, nonsynonymous SNP and C>T and G>A was the major variation type and base change types, respectively. A total of 24 common mutated genes were identified in 26 sporadic immature teratoma patients. Among which, 5 common mutated genes with common mutation sites were identified, including 2 frameshift mutant genes (MYPOP and FRG2C) and 3 nonsynonymous mutant genes (CNTNAP3, GPRIN2 and MUC3A). Moreover, molecular weight of these genes changed slightly after mutation. However, it is worth noting that this study is a preliminary exploratory investigation, and the results of this study still require further in-depth research and experimental verification in subsequent work to further confirm their reliability.

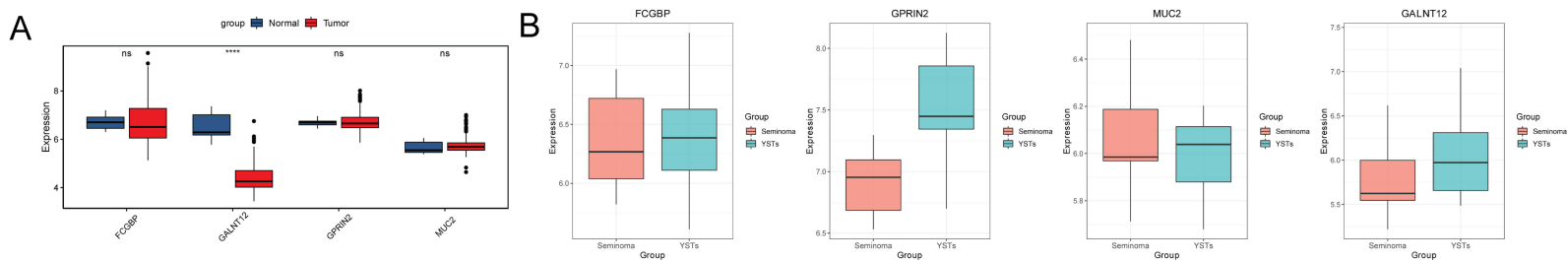


Figure 5 MRNA expression of mutated genes in ovarian cancer (**A**) and yolk sac tumor (**B**). ****, $p < 0.0001$.
Abbreviations: ns, no statistical significance; YST, yolk sac tumor.

It is found that MYPOP can regulate hematopoiesis and tumorigenesis.¹¹ In addition, a differentially methylated DNAm probe site of MYPOP is identified in endometrium.¹² FRG2C is more frequently mutated in estrogen receptor-negative breast cancer.¹³ Additionally, mutation of FRG2C (c.A497G; p.E166G) is found in gestational choriocarcinoma.¹⁴ Herein, frameshift mutations of MYPOP (R367Pfs*83) and FRG2C (R161Sfs*5) were found in immature teratoma. Interestingly, MYPOP and FRG2C had a higher deletion frequency of CNV in ovarian cancer. It is indicated that frameshift mutations in MYPOP (R367Pfs*83) and FRG2C (R161Sfs*5) may be associated with the tumorigenesis of immature teratoma.

CNTNAP3 encodes the protein related to cell-recognition process within the nervous system.¹⁵ CNTNAP3 is down-regulated in women with adenomyosis.¹⁶ Chaiworapongsa T et al found that CNTNAP3 was up-regulated in late-onset preeclampsia patients.¹⁷ In peripheral blood samples of testicular germ cell tumors, loss of 9p13.1-p12 of CNTNAP3 was the only one novel CNV.¹⁸ In this study, the nonsynonymous mutation of CNTNAP3 (T1051S) was identified in immature teratoma. In addition, CNTNAP3 had higher amplification frequency of CNV and had rare mutation frequency (1%) in ovarian cancer. It is suggested that nonsynonymous mutation of CNTNAP3 (T1051S) may be associated with germ cell development in immature teratoma.

GPRIN2, encodes a G-protein regulated inducer of neurite overgrowth, is involved in formation and extension of neurite-like processes.^{19,20} The mutation of GPRIN2 (rs112620425) is found in the breast cancer.²¹ It is noted that CNV amplification of GPRIN2 is detected in fetuses with congenital malformations. In the present study, the nonsynonymous mutation of GPRIN2 (L400P) was identified in immature teratoma. GPRIN2 had a higher amplification frequency of CNV in ovarian cancer. In addition, mRNA expression levels of GPRIN2 were down-regulated and up-regulated in ovarian cancer and yolk sac tumor, respectively. It is assumed that GPRIN2 is involved in the embryonic neurodevelopment, the nonsynonymous mutation of GPRIN2 (L400P) may be associated with immature teratoma. Heinzelmann-Schwarz VA et al found that MUC3A was up-regulated in mucinous ovarian carcinoma.²² In this study, nonsynonymous mutations of MUC3A (T2244S, T2000N, A3149E and T2000A) were identified in immature teratoma. Furthermore, MUC3A had a higher amplification frequency of CNV in ovarian cancer. These nonsynonymous mutations of MUC3A may be associated with the process of immature teratoma.

In the PPI network, the highest degree gene was MUC12. Furthermore, MUC12 interacted with FCGBP and GALNT12, except MUC3A. In breast cancer, MUC12 is primarily connected with the modulation and function of Janus kinase (JAKs).²³ FCGBP, located in the mucosa of endodermal-derived tissues, plays a crucial role in mucosal immunological defenses.^{24,25} In ovarian cancer, FCGBP is significantly associated with overall survival.²⁶ GALNT12 is involved in embryonic development, organ development and organ morphology.²⁷ In healthy women, GALNT12 is associated with endometrial receptivity.²⁸ Herein, mutations of MUC12, FCGBP and GALNT12 were found in immature teratoma. Moreover, MUC12 and FCGBP had a higher amplification frequency of CNV, and GALNT12 had a higher deletion frequency of CNV in ovarian cancer. In addition, FCGBP was down-regulated and up-regulated in ovarian cancer and yolk sac tumor. These results indicated that MUC12, FCGBP, GALNT12 and MUC3A may act synergistically during the development of immature teratoma.

Based on the drugs-24 mutated genes interaction analysis, only compounds of CHEMBL35482, FLUOROURACIL, and DECITABINE interacted with MUC2. Moreover, mRNA expression levels of MUC2 were up-regulated in both ovarian cancer and yolk sac tumor. MUC2 is expressed in villous goblet cells and immature crypt cells.^{29–31} In ovarian cancer, MUC2 is a tumor promoter and is inversely associated with patient survival.³² It is worth mentioning that MUC2 is expressed in ovarian teratoma.³³ In ovarian cancer, CHEMBL35482 is a potential drug interacted with cyclin-dependent kinase inhibitor 1B (CDKN1B).³⁴ It is found that the production of MUC2 can be blocked by CHEMBL35482 in pulmonary epithelial cells.³⁵ It is found that the adapted chemotherapy based on histology, such as FLUOROURACIL-based regimens for adenocarcinoma transformation, has recently been advocated to improve patient outcomes.^{36,37} A patient with metastatic colonic-type adenocarcinoma arising in a teratoma has achieved response from FLUOROURACIL-based chemotherapy.³⁸ DECITABINE, a specific DNA methyltransferase inhibitor, can exert anti-tumor effects by re-expressing tumor suppressor genes. The early phase clinical trials in epithelial ovarian cancer have showed mixed results of the combination of DECITABINE and carboplatinum.^{39–41} Good antitumor effect in combination DECITABINE with belinostat has been found in ovarian cancer.⁴² It is suggested that MUC2 may be associated with

the pathogenesis of immature teratoma and could be potentially modulated by CHEMBL35482, FLUOROURACIL, and DECITABINE in the treatment of immature teratoma.

According to the GO analysis of 24 mutated genes, MUC3A, MUC2 and MUC12 were significantly enriched in activation of the innate immune response. The diminished innate immune response can help pluripotent stem cells maintain the rapid rate of cell proliferation, which is essential for the early embryo development.⁴³ In teratoma, targeting the immune response significantly inhibits the evolution of tumors progression.⁴⁴ Thus, it can be seen that the innate immune response may contribute to the pathological mechanism of immature teratoma by regulating the activity of MUC3A, MUC2 and MUC12. Moreover, previous studies have shown that MUC2 and MUC3A may regulate disease progression via the PI3K/Akt/mTOR and NF-kappaB signaling pathways, which are also involved in modulating macrophage-mediated innate immune responses.^{45–48} MUC12 exerts an oncogenic role via the c-Jun/TGF- β signaling pathway, and TGF- β is involved in the regulation of innate immunity.^{49,50} These studies suggest that the aberrations of MUC2, MUC3A and MUC12 identified in the study may participate in the development of immature teratoma by regulating the aforementioned signaling pathways and influencing innate immune responses. This provides a theoretical framework for elucidating the pathological mechanism of the disease, and also points out the potential direction for subsequent related research. In addition, 2 common CNVs (deletion state) were screened out in immature teratoma, which involving 6 genes, such as RP11. The expression of RP11 is decreased in human embryo competence and increased in ovarian germ cell tumor.^{51,52} Maybe, CNVs deletion of RP11 is associated with the development of immature teratoma.

In conclusion, this WES study based on peripheral blood was performed to explore germline variations in 26 patients with sporadic immature teratoma. We identified 24 common mutated genes. Among which, 5 common mutated genes with common mutation sites were identified, including 2 frameshift mutant genes (MYPOP and FRG2C) and 3 nonsynonymous mutant genes (CNTNAP3, GPRIN2 and MUC3A). In the PPI network, the highest degree gene was MUC12. MUC2 may be a potential target for CHEMBL3582, FLUOROURACIL and DECITABINE. Functional analysis showed that MUC3A, MUC2 and MUC12 were enriched in innate immune response. Two common deletion CNVs were also identified. These germline SNPs and CNVs may be associated with the development of immature teratoma and provide potential biomarkers and therapeutic targets. However, there are some limitations in this study. Firstly, the number of enrolled immature teratoma patients is small, larger pedigrees or sporadic cases are needed. Secondly, all the identified SNPs should be validated by Sanger sequencing. Finally, the investigation of the underlying molecular mechanism of SNPs and CNVs is further needed in animal model or cell experiments.

Data Sharing Statement

The datasets generated and analyzed during the current study are available in the SRA repository (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA1370473?reviewer=q5p1569l959dtcucctbfce3lji>).

Ethics Approval and Informed Consent

This study was approved by the ethics committee of the fourth hospital of hebei medical university (2025KS205). Written consent was obtained from enrolled individuals. All patients were informed about the purpose of the study, and this study was conducted in accordance with the Declaration of Helsinki.

Funding

This study was funded by the “2023 Medical Science Research Project Plan of Hebei Province (20230796)”.

Disclosure

The authors declare no competing interests in this work.

References

1. Algahtani HA, Al-Rabia MW, Al-Maghrabi HQ, Kutub HY. Posterior fossa teratoma. *Neurosciences*. 2013;18(4):371–374.
2. Tavassoli F. Pathology and genetics of tumours of the breast and female genital organs. *World Health Organization Classification of Tumours*. 2003.

3. Desai K, Nadkarni T, Muzumdar D, Goel A. Midline posterior fossa teratoma--case report. *Neurol Med Chir.* 2001;41(2):94–96. doi:10.2176/nmc.41.94
4. O'Connor DM, Norris HJ. The influence of grade on the outcome of stage I ovarian immature (malignant) teratomas and the reproducibility of grading. *Int J Gynecol Pathol.* 1994;13(4):283–289. doi:10.1097/00004347-199410000-00001
5. Kurata A, Hirano K, Nagane M, Fujioka Y. Immature teratoma of the ovary with distant metastases: favorable prognosis and insights into chemotherapeutic retroconversion. *Int J Gynecol Pathol.* 2010;29(5):438–444. doi:10.1097/PGP.0b013e3181cef16b
6. Byun JC, Choi JJ, Han MS, Lee SC, Roh MS, Cha MS. Soft tissue metastasis of an immature teratoma of the ovary. *J Obstet Gynaecol Res.* 2011;37(11):1689–1693. doi:10.1111/j.1447-0756.2011.01553.x
7. Swartjes JM, de Blok S, Blaauwgeers JL. Abdominal wall metastases after surgical resection of an immature teratoma of the ovary. *Eur J Obstet Gynecol Reprod Biol.* 1997;74(1):41–43. doi:10.1016/s0301-2115(97)02758-9
8. Heskett MB, Sanborn JZ, Boniface C, et al. Multiregion exome sequencing of ovarian immature teratomas reveals 2N near-diploid genomes, paucity of somatic mutations, and extensive allelic imbalances shared across mature, immature, and disseminated components. *Modern Pathology.* 2020;33(6):1193–1206. doi:10.1038/s41379-019-0446-y
9. Chen J, Li Y, Wu J, Liu Y, Kang S. Whole-exome sequencing reveals potential germline and somatic mutations in 60 malignant ovarian germ cell tumors†. *Biol Reprod.* 2021;105(1):164–178. doi:10.1093/biolre/iaob052
10. Chang TY, Chen SW, Lin WH, et al. Comparison of genetic profiling between primary tumor and circulating tumor cells captured by microfluidics in epithelial ovarian cancer: tumor heterogeneity or allele dropout? *Diagnostics.* 2021;11(6). doi:10.3390/diagnostics11061102
11. Vargova K, Curik N, Burda P, et al. MYB transcriptionally regulates the miR-155 host gene in chronic lymphocytic leukemia. *Blood.* 2011;117(14):3816–3825. doi:10.1182/blood-2010-05-285064
12. Mortlock S, Restuadi R, Levien R, et al. Genetic regulation of methylation in human endometrium and blood and gene targets for reproductive diseases. *Clin Epigenetics.* 2019;11(1):49. doi:10.1186/s13148-019-0648-7
13. Kim JY, Lee E, Park K, et al. Clinical implications of genomic profiles in metastatic breast cancer with a focus on TP53 and PIK3CA, the most frequently mutated genes. *Oncotarget.* 2017;8(17):27997–28007. doi:10.18632/oncotarget.15881
14. Lazare C, Zhi W, Dai J, et al. A pilot study comparing the genetic molecular biology of gestational and non-gestational choriocarcinoma. *Am J Transl Res.* 2019;11(11):7049–7062.
15. Spiegel I, Salomon D, Erne B, Schaeren-Wiemers N, Peles E. Caspr3 and caspr4, two novel members of the caspr family are expressed in the nervous system and interact with PDZ domains. *Mol Cell Neurosci.* 2002;20(2):283–297. doi:10.1006/mcne.2002.1110
16. Prašnikar E, Kunej T, Gorenjak M, Potočnik U, Kovačič B, Knez J. Transcriptomics of receptive endometrium in women with sonographic features of adenomyosis. *Reproduct Biol Endocrinol.* 2022;20(1):2. doi:10.1186/s12958-021-00871-5
17. Chaiworapongsa T, Romero R, Whitten A, et al. Differences and similarities in the transcriptional profile of peripheral whole blood in early and late-onset preeclampsia: insights into the molecular basis of the phenotype of preeclampsia. *J Perinat Med.* 2013;41(5):485–504. doi:10.1515/jpm-2013-0082
18. Silveira SM, da Cunha IW, Marchi FA, Busso AF, Lopes A, Rogatto SR. Genomic screening of testicular germ cell tumors from monozygotic twins. *Orphanet J Rare Dis.* 2014;9:181. doi:10.1186/s13023-014-0181-x
19. Ohtoshi A, Behringer RR. Neonatal lethality, dwarfism, and abnormal brain development in Dmbx1 mutant mice. *Mol Cell Biol.* 2004;24(17):7548–7558. doi:10.1128/mcb.24.17.7548-7558.2004
20. Chen LT, Gilman AG, Kozasa T. A candidate target for G protein action in brain. *J Biol Chem.* 1999;274(38):26931–26938. doi:10.1074/jbc.274.38.26931
21. Lin YY, Gawronski A, Hach F, et al. Computational identification of micro-structural variations and their proteogenomic consequences in cancer. *Bioinformatics.* 2018;34(10):1672–1681. doi:10.1093/bioinformatics/btx807
22. Heinzelmann-Schwarz VA, Gardiner-Garden M, Henshall SM, et al. A distinct molecular profile associated with mucinous epithelial ovarian cancer. *Br J Cancer.* 2006;94(6):904–913. doi:10.1038/sj.bjc.6603003
23. Liu F, Wu H. Identification of prognostic biomarkers and molecular targets among JAK family in breast cancer. *J Inflamm Res.* 2021;14:97–114. doi:10.2147/jir.s284889
24. Kobayashi K, Ogata H, Morikawa M, et al. Distribution and partial characterisation of IgG Fc binding protein in various mucin producing cells and body fluids. *Gut.* 2002;51(2):169–176. doi:10.1136/gut.51.2.169
25. Stamp LA, Braxton DR, Wu J, et al. The GCTM-5 epitope associated with the mucin-like glycoprotein FCGBP marks progenitor cells in tissues of endodermal origin. *Stem Cells.* 2012;30(9):1999–2009. doi:10.1002/stem.1167
26. Wang L, Gao S. Identification of 5-methylcytosine-related signature for predicting prognosis in ovarian cancer. *Biol Res.* 2021;54(1):18. doi:10.1186/s40659-021-00340-8
27. Le Guillou S, Sdassi N, Laubier J, et al. Overexpression of miR-30b in the developing mouse mammary gland causes a lactation defect and delays involution. *PLoS One.* 2012;7(9):e45727. doi:10.1371/journal.pone.0045727
28. Prašnikar E, Kunej T. Determining the molecular background of endometrial receptivity in adenomyosis. *Biomolecules.* 2020;10(9):. doi:10.3390/biom10091311
29. Audie JP, Janin A, Porchet N, Copin MC, Gosselin B, Aubert JP. Expression of human mucin genes in respiratory, digestive, and reproductive tracts ascertained by in situ hybridization. *J Histochem Cytochem.* 1993;41(10):1479–1485. doi:10.1177/41.10.8245407
30. Chang SK, Dohrman AF, Basbaum CB, et al. Localization of mucin (MUC2 and MUC3) messenger RNA and peptide expression in human normal intestine and colon cancer. *Gastroenterology.* 1994;107(1):28–36. doi:10.1016/0016-5085(94)90057-4
31. Buisine MP, Janin A, Maunoury V, et al. Aberrant expression of a human mucin gene (MUC5AC) in rectosigmoid villous adenoma. *Gastroenterology.* 1996;110(1):84–91. doi:10.1053/gast.1996.v110.pm8536891
32. He YF, Zhang MY, Wu X, et al. High MUC2 expression in ovarian cancer is inversely associated with the M1/M2 ratio of tumor-associated macrophages and patient survival time. *PLoS One.* 2013;8(12):e79769. doi:10.1371/journal.pone.0079769
33. Mall AS, Tyler M, Lotz Z, et al. The characterisation of mucin in a mature ovarian teratoma occurring in an eight year old patient. *Int J Med Sci.* 2007;4(2):115–123. doi:10.7150/ijms.4.115

34. Hoshino R, Tanimura S, Watanabe K, Kataoka T, Kohno M. Blockade of the extracellular signal-regulated kinase pathway induces marked G1 cell cycle arrest and apoptosis in tumor cells in which the pathway is constitutively activated: up-regulation of p27(Kip1). *J Biol Chem*. 2001;276(4):2686–2692. doi:10.1074/jbc.M006132200
35. Kim YD, Kwon EJ, Park DW, Song SY, Yoon SK, Baek SH. Interleukin-1 β induces MUC2 and MUC5AC synthesis through cyclooxygenase-2 in NCI-H292 cells. *Mol Pharmacol*. 2002;62(5):1112–1118. doi:10.1124/mol.62.5.1112
36. Khurana A, Mehta A, Kamboj M. Colonic-type adenocarcinoma (somatic-type malignancy) arising in a mediastinal mature cystic teratoma: a case report of a rare entity. *Indian J Pathol Microbiol*. 2011;54(1):199–200. doi:10.4103/0377-4929.77405
37. Donadio AC, Motzer RJ, Bajorin DF, et al. Chemotherapy for teratoma with malignant transformation. *J Clin Oncol*. 2003;21(23):4285–4291. doi:10.1200/jco.2003.01.019
38. Dechaphunkul A, Bigras G, Sawyer M. Response to 5-Fluorouracil-based chemotherapy in a patient with metastatic colonic-type adenocarcinoma arising in a primary mediastinal teratoma. *Case Rep Oncol Med*. 2012;2012:729278. doi:10.1155/2012/729278
39. Fang F, Balch C, Schilder J, et al. A Phase I and pharmacodynamic study of decitabine in combination with carboplatin in patients with recurrent, platinum-resistant, epithelial ovarian cancer. *Cancer*. 2010;116(17):4043–4053. doi:10.1002/cncr.25204
40. Glasspool RM, Brown R, Gore ME, et al. A randomised, Phase II trial of the DNA-hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in combination with carboplatin vs carboplatin alone in patients with recurrent, partially platinum-sensitive ovarian cancer. *Br J Cancer*. 2014;110(8):1923–1929. doi:10.1038/bjc.2014.116
41. Matei D, Fang F, Shen C, et al. Epigenetic resensitization to platinum in ovarian cancer. *Cancer Res*. 2012;72(9):2197–2205. doi:10.1158/0008-5472.can-11-3909
42. Steele N, Finn P, Brown R, Plumb JA. Combined inhibition of DNA methylation and histone acetylation enhances gene re-expression and drug sensitivity in vivo. *Br J Cancer*. 2009;100(5):758–763. doi:10.1038/sj.bjc.6604932
43. Guo YL. The underdeveloped innate immunity in embryonic stem cells: the molecular basis and biological perspectives from early embryogenesis. *Am J Reprod Immunol*. 2019;81(2):e13089. doi:10.1111/aji.13089
44. Wang X, Chen T, Leng L, et al. MIF produced by bone marrow-derived macrophages contributes to teratoma progression after embryonic stem cell transplantation. *Cancer Res*. 2012;72(11):2867–2878. doi:10.1158/0008-5472.can-11-3247
45. Wei J, Chen C, Feng J, et al. Muc2 mucin O-glycosylation interacts with enteropathogenic *Escherichia coli* to influence the development of ulcerative colitis based on the NF- κ B signaling pathway. *J Transl Med*. 2023;21(1):793. doi:10.1186/s12967-023-04687-2
46. Su W, Feng B, Hu L, Guo X, Yu M. MUC3A promotes the progression of colorectal cancer through the PI3K/Akt/mTOR pathway. *BMC Cancer*. 2022;22(1):602. doi:10.1186/s12885-022-09709-8
47. Sun Y, Sun X, You C, et al. MUC3A promotes non-small cell lung cancer progression via activating the NF κ B pathway and attenuates radiosensitivity. *Int J Biol Sci*. 2021;17(10):2523–2536. doi:10.7150/ijbs.59430
48. Wang Y, Liu Y, Zhang XY, et al. Ginsenoside Rg1 regulates innate immune responses in macrophages through differentially modulating the NF- κ B and PI3K/Akt/mTOR pathways. *Int Immunopharmacol*. 2014;23(1):77–84. doi:10.1016/j.intimp.2014.07.028
49. Gao SL, Yin R, Zhang LF, et al. The oncogenic role of MUC12 in RCC progression depends on c-Jun/TGF- β signalling. *J Cell Mol Med*. 2020;24(15):8789–8802. doi:10.1111/jcmm.15515
50. Thomas BJ, Kan OK, Loveland KL. In the shadow of fibrosis: innate immune suppression mediated by transforming growth factor- β . *Ame J Respir Cell Mole Biol*. 2016;55(6):759–766. doi:10.1165/rcmb.2016-0248PS
51. Groff AF, Resetskova N, DiDomenico F, et al. RNA-seq as a tool for evaluating human embryo competence. *Genome Res*. 2019;29(10):1705–1718. doi:10.1101/gr.252981.119
52. Schmidtova S, Dorssers LCJ, Kalavska K, et al. Napabucasin overcomes cisplatin resistance in ovarian germ cell tumor-derived cell line by inhibiting cancer stemness. *Cancer Cell Int*. 2020;20:364. doi:10.1186/s12935-020-01458-7