BPR MCS Bulletin of Pioneering Researches of Medical and Clinical Science

Bulletin of Pioneering Researches of Medical and Clinical Science

Available online: https://bprmcs.com
2021 | Volume 1 | Issue 1 | Page: 131-142

Detection of Ovarian High-Grade Serous Carcinoma through Mitochondrial Gene Variation

Wei Li[1]*, Chen Hao[2], Min Zhang[3]

[1]Department of Clinical Medicine, Peking University Health Science Center, Beijing, China.
[2]Department of Medical Sciences, University of Hong Kong, Hong Kong.
[3]Department of Clinical Sciences, National University of Singapore, Singapore.

Abstract

Women diagnosed with ovarian cancer at advanced stages have significantly poorer survival outcomes compared to those diagnosed at early stages, yet early detection remains a major clinical challenge. Recent evidence suggests that genetic variations may serve as potential biomarkers for the early identification of various cancers. In this pilot observational retrospective study, we investigated whether mitochondrial DNA (mtDNA) variations could distinguish the most common ovarian cancer subtype, high-grade serous carcinoma (HGSC), from normal tissue. mtDNA variations were analyzed in twenty whole-exome sequenced (WES) HGSC samples and fourteen control fallopian tube samples following established genome sequencing protocols. Using these variants, we developed predictive models for HGSC, achieving strong performance with an area under the curve (AUC) of 0.88 (CI: 0.74–1.00). The variants included in the optimal model were further correlated with gene expression to explore potential functional implications. Validation using the Cancer Genome Atlas (TCGA) dataset, encompassing over 420 samples, yielded moderate predictive performance (AUC 0.63–0.71). Overall, our study identified a set of mtDNA variations capable of distinguishing HGSC with high accuracy, with MT-CYB gene variants notably increasing HGSC risk by over 30 percent and exhibiting significantly reduced expression in affected patients. These findings suggest that mtDNA-based predictive models could be integrated into liquid biopsy approaches for early detection of ovarian cancer, paralleling advances in other malignancies.

Keywords: Ovarian cancer, Genetic variation, Whole-exome sequencing—WES, RNA sequencing—RNAseq, Prediction model

Corresponding author: Wei Li
E-mail: weili@163.com

How to Cite This Article: Li W, Hao C, Zhang M. Detection of Ovarian High-Grade Serous Carcinoma through Mitochondrial Gene Variation. Bull Pioneer Res Med Clin Sci. 2021;1(1):131-42. https://doi.org/10.51847/i2gdl8Vpw9

Introduction

Early detection of ovarian cancer remains challenging. While patients diagnosed at early stages have a 5-year survival rate exceeding 90%, over 70 percent are identified at advanced stages, where the 5-year survival drops to approximately 40% [1, 2]. Early diagnosis is therefore crucial for improving prognosis and survival, yet no current screening methods reliably detect ovarian cancer at an early stage [3]. Recent studies have suggested that genomic variations, including those in mitochondrial DNA (mtDNA), could aid in cancer detection [4–7]. Although mtDNA variations may not directly affect transcription or translation, they can be leveraged to develop models that differentiate cancerous from normal cells. Cell-free DNA (cfDNA), present in bodily fluids and reflecting the genetic profile of the tissue of origin, enables non-invasive tumor profiling through liquid biopsy [8]. cfDNA variations have already been employed to detect cancers at early stages [9, 10], guide prognosis, and monitor treatment response in lung cancer [11, 12], colon cancer with KRAS mutations [13], and aggressive breast

cancers such as triple-negative subtypes [14]. These findings indicate that mtDNA variations in cfDNA might also be useful for early ovarian cancer detection. To explore this, we utilized a well-characterized biobank of ovarian cancer and normal fallopian tissue samples to identify and compare mtDNA variations.

The primary aim of this pilot study was to determine whether mtDNA variations could distinguish HGSC from normal tissue. Prediction models were validated across different platforms and using the independent TCGA HGSC dataset. Significant mtDNA variants were further analyzed for correlations with HGSC gene expression and clinical outcomes.

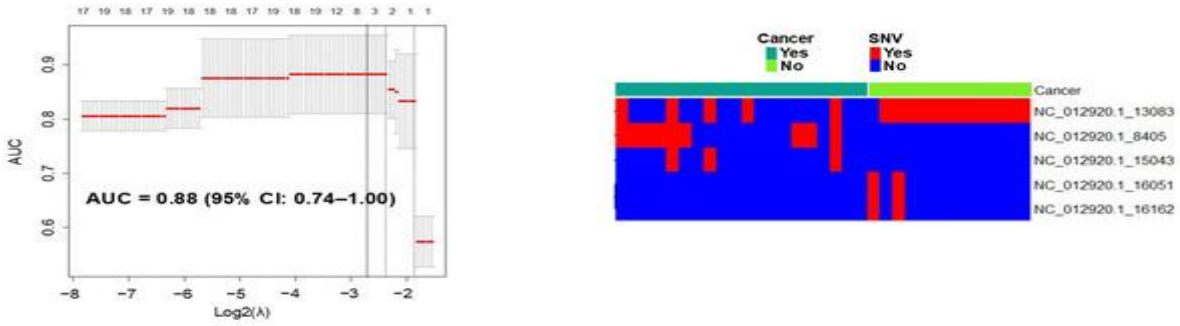
Results and Discussion

mtDNA single nucleotide variations (SNVs)

Whole-exome sequencing (WES) of DNA from HGSC cases (N = 20) and control fallopian tube samples (N = 14) identified 393 variants across 37 mitochondrial genes. In the discovery phase, all mtDNA SNVs were included in a multivariable LASSO regression analysis. Bootstrapping was used to determine the optimal penalty parameter (lambda, λ) to minimize overfitting, resulting in λ = 0.136, which was lower than the standard lambda.min (0.176) and lambda.1se (0.255) recommended by the glmnet package (**(Figure 1b)**, dotted lines). Using this parameter, the five most informative variants for HGSC prediction were selected **(Figure 1a)**. Variants in MT-ATP8, MT-ND5, and MT-CYB were associated with increased HGSC risk (OR > 1), whereas variants in MT-TP and MT-CO1 were protective (OR < 1). The predictive model achieved an AUC of 0.88 (95 percent CI: 0.74–1.00) **(Figure 1b)**.

	REF	ALT	VAF (HGSC)	OR	p-value*	95% CI*	MT Gene	SNPs	Description
NC_012920.1_8405	A	G	0.225	2.01	0.013	1.36, 4.48	MT-ATP8	rs1603221462	mitochondrially encoded ATP synthase 8
NC_012920.1_13083	A	C	0.125	1.08	0.775	0.69, 1.69	MT-ND5	N/A	NADH dehydrogenase, subunit 5 (complex I)
NC_012920.1_15043	G	A	0.15	1.32	0.656	0.45, 4.08	MT-CYB	rs193302985	mitochondrially encoded cytochrome b
NC_012920.1_16051	A	G	0	0.84	0.984	0.31, 3.15	MT-TP	rs117565943	mitochondrially encoded tRNA proline
NC_012920.1_16162	A	G	0	0.99	0.637	0.44, 4.44	MT-CO1	rs41466049	mitochondrially encoded cytochrome c oxidase

a)

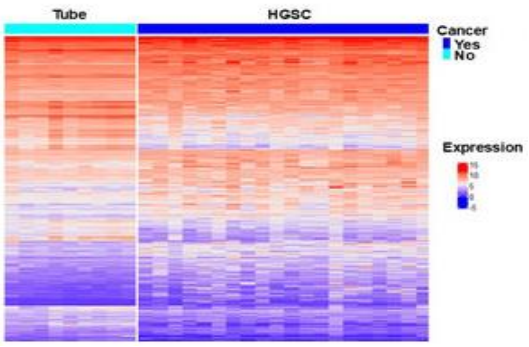


b) c)

Figure 1. Prediction of HGSC based on mtDNA alterations. (a) Out of 393 identified mtDNA variants, five were selected as the most informative for predicting ovarian cancer, achieving an AUC of 0.88 (95% CI: 0.74–1.00). REF indicates the reference allele; ALT represents the alternative allele detected; VAF denotes the variant allele frequency in cancer samples; OR reflects the odds ratio comparing alleles in cancer versus control samples; MT Gene specifies the mitochondrial gene name; SNP refers to the known single nucleotide polymorphism identifier. *The LASSO regression approach prioritizes predictive variable selection rather than formal statistical inference; variant effect sizes and confidence intervals were estimated using the R package selectiveInference (v. 1.2.5), though potential overfitting may influence results. (b) Visualization of LASSO regression with 95% confidence intervals: the top axis shows the number of SNVs selected by the model, the left axis indicates the corresponding AUC, and the bottom axis depicts the log2-transformed lambda tuning parameter chosen through bootstrapping and cross-validation. Dotted lines indicate reference lambda values: lambda.1se (right), lambda.min (center), and the selected λ (left). (c) Heatmap of mtDNA variants included in the model per sample, with red representing variant presence and blue indicating absence; N/A indicates not applicable.

Linking significant mtDNA variants to gene expression

Gene expression profiles were compared between 112 HGSC tumors and 12 control samples. RNA sequencing analysis revealed 3,382 transcripts showing significant differential expression out of 61,851 tested, using an adjusted p-value < 0.005 to correct for multiple testing **(Figure 2a)**. Among the 37 mitochondrial genes analyzed, 13 displayed notable expression differences between cancerous and normal tissues: MT-RNR1, MT-TS1, MT-TM, MT-ND3, MT-TP, MT-TK, MT-TS2, MT-TV, MT-CO2, MT-CO1, MT-TL1, MT-CYB, and MT-TY. These 13 genes were then subjected to multivariable regression, identifying four as independently linked to increased ovarian cancer risk **(Figure 2b)**.



a)

	Gene	OR	95% CI	p-value
ENSG00000210112.1	MT-TM	1.3×10^{7}	294.9, 1.9×10^{14}	0.014
ENSG00000198840.2	MT-ND3	2.3×10^{-5}	9.8×10^{-11}, 0.027	0.024
ENSG00000210156.1	MT-TK	1.7×10^{-4}	5.1×10^{-6}, 0.016	0.005
ENSG00000198727.2	MT-CYB	5.1×10^{-5}	3.9×10^{-10}, 0.067	0.034

b)

	Symbol	Gene type	r	adjusted p-value
ENSG00000266110.1	MIR4423	mRNA	−0.80	<0.001
ENSG00000248884.2	CTC-537E7.3	lncRNA	−0.79	<0.001
ENSG00000248359.2	AC010280.1	lncRNA	−0.80	<0.001
ENSG00000249588.1	AC010280.3	lncRNA	−0.80	<0.001
ENSG00000260711.2	AL121839.2	lncRNA	−0.79	<0.001
ENSG00000128918.15	ALDH1A2	protein_coding	−0.80	<0.001
ENSG00000274719.1	AC012653.2	lncRNA	−0.79	<0.001
ENSG00000140450.9	ARRDC4	protein_coding	−0.80	<0.001
ENSG00000189221.11	MAOA	protein_coding	−0.80	<0.001
ENSG00000204116.12	CHIC1	protein_coding	−0.79	<0.001
ENSG00000187801.15	ZFP69B	protein_coding	0.78	<0.001
ENSG00000117262.19	GPR89A	protein_coding	0.78	<0.001
ENSG00000226963.1	AC078883.2	lncRNA	−0.78	<0.001
ENSG00000186329.9	TMEM212	protein_coding	−0.78	<0.001
ENSG00000164530.15	PI16	protein_coding	−0.78	<0.001
ENSG00000235399.1	AL135902.2	lncRNA	−0.78	<0.001
ENSG00000284471.1	CR769775.2	TEC	−0.78	<0.001
ENSG00000107554.17	DNMBP	protein_coding	−0.78	<0.001
ENSG00000257057.3	C11orf97	protein_coding	−0.78	<0.001
ENSG00000151715.8	TMEM45B	protein_coding	−0.78	<0.001

c)

	Symbol	Gene type	r	adjusted p-value
ENSG00000210144.1	MT-TY	Mt_tRNA	−0.51	<0.001
ENSG00000154678.18	PDE1C	protein_coding	0.50	<0.001
ENSG00000177990.12	DPY19L2	protein_coding	0.47	<0.001
ENSG00000100902.11	PSMA6	protein_coding	−0.48	<0.001
ENSG00000115207.15	GTF3C2	protein_coding	−0.43	<0.001
ENSG00000163798.14	SLC4A1AP	protein_coding	−0.43	<0.001
ENSG00000138035.15	PNPT1	protein_coding	−0.42	<0.001
ENSG00000146278.11	PNRC1	protein_coding	0.42	<0.001
ENSG00000136943.12	CTSV	protein_coding	−0.42	<0.001
ENSG00000172915.20	NBEA	protein_coding	0.42	<0.001
ENSG00000201155.1	RNU1-24P	snRNA	0.42	<0.001
ENSG00000102053.13	ZC3H12B	protein_coding	0.44	<0.001
ENSG00000242525.1	OR7E109P	unprocessed_pseudogene	0.41	<0.001
ENSG00000238000.1	AC116347.1	processed_pseudogene	−0.42	<0.001
ENSG00000172336.5	POP7	protein_coding	−0.42	<0.001

d)

Figure 2. Differential expression analysis and association with mtDNA variants in HGSC versus normal fallopian tube samples. (a) Comparison of gene expression between HGSC tumors and control tubal tissue identified 3,382 transcripts (out of 61,851) as significantly altered, using an FDR-adjusted p-value < 0.005 to account for multiple testing. (b) The 13 mitochondrial genes found significant in univariable analysis were included in a multivariable regression model, revealing four genes independently associated with HGSC. Among these, MT-CYB (highlighted in red) harbored SNVs characteristic of HGSC and was incorporated into the cancer prediction model **(Figure 1)**. (c) Correlation between MT-CYB mtDNA genotype and gene expression showed that 1,728 of the 3,382 differentially expressed genes were significantly associated (FDR-adjusted $p < 0.005$), with the top 20 displayed. (d) Correlation of MT-CYB gene expression with the same set of 3,382 significant transcripts identified 364 genes with significant association (FDR-adjusted $p < 0.005$), with the top 15 shown.

MT-CYB emerged as a key predictor in the ovarian cancer model, conferring elevated HGSC risk, and its expression was markedly reduced in patients **(Figure 2b)**. To explore the broader impact of MT-CYB alterations on gene regulation and biological processes, we first identified genes significantly influenced by these changes and then performed pathway enrichment analysis.

MT-CYB expression was correlated with all 3,382 transcripts that were differentially expressed between HGSC and control tissues to exclude genes with non-significant changes. This identified 1,728 transcripts significantly associated with MT-CYB variation (FDR-adjusted $p < 0.005$), with the top 20 illustrated in **Figure 2c**. Next, MT-CYB gene expression itself (from the multivariable model) was correlated with the same set of 3,382 genes, yielding 364 significantly associated transcripts (FDR-adjusted $p < 0.005$), with the top 15 shown in **Figure 2d**. A total of 202 genes were common to both analyses, showing significant correlation with both MT-CYB variation and expression.

Pathway enrichment of MT-CYB-associated genes

KEGG pathway enrichment analysis was performed using the 1,890 unique genes whose expression correlated significantly with MT-CYB variation and/or expression. The pathways with significant enrichment are summarized in **Table 1**.

Table 1. KEGG pathway enrichment analysis for genes significantly associated with MT-CYB variants and/or expression. A total of 3,382 transcripts were identified as differentially expressed between HGSC and control samples. Among these, 1,890 genes showed significant correlation with MT-CYB genetic variation and/or its expression (FDR-adjusted $p < 0.005$). These genes were analyzed for pathway enrichment, revealing significant involvement in energy metabolism, cancer-related pathways, neurodegenerative disorders, and other cellular processes (FDR-adjusted $p < 0.01$).

KEGG ID	Pathway Description	Pathway Category	Adjusted p-value	Enrichment Fold
hsa00190	Oxidative Phosphorylation	Energy Metabolism	<0.001	2.84
hsa05415	Diabetic Cardiomyopathy	Cardiovascular Disease	<0.001	2.29
hsa04714	Thermogenesis	Environmental Adaptation	<0.001	2.17

hsa05016	Huntington's Disease	Neurodegenerative Disease	0.001	1.89
hsa05014	Amyotrophic Lateral Sclerosis	Neurodegenerative Disease	0.001	1.80
hsa05012	Parkinson's Disease	Neurodegenerative Disease	0.001	1.88
hsa04932	Non-Alcoholic Fatty Liver Disease	Endocrine & Metabolic Disease	0.003	2.12
hsa05020	Prion Disease	Neurodegenerative Disease	0.004	1.76
hsa05208	Chemical Carcinogenesis – Reactive Oxygen Species	Cancer	0.006	1.81
hsa05010	Alzheimer's Disease	Neurodegenerative Disease	0.009	1.55
hsa05022	Multiple Neurodegeneration Pathways	Neurodegenerative Disease	0.014	1.46
hsa00510	N-Glycan Biosynthesis	Glycan Biosynthesis & Metabolism	0.017	2.59
hsa00600	Sphingolipid Metabolism	Lipid Metabolism	0.019	2.54
hsa04723	Retrograde Endocannabinoid Signaling	Nervous System	0.020	1.84

The pathway with the highest enrichment was oxidative phosphorylation (OXPHOS), a central component of cellular energy metabolism. This pathway involves a series of enzymes that oxidize nutrients to generate energy in the form of adenosine triphosphate (ATP). OXPHOS comprises five protein complexes (I–V) and occurs within the mitochondria **(Figure 3)**.

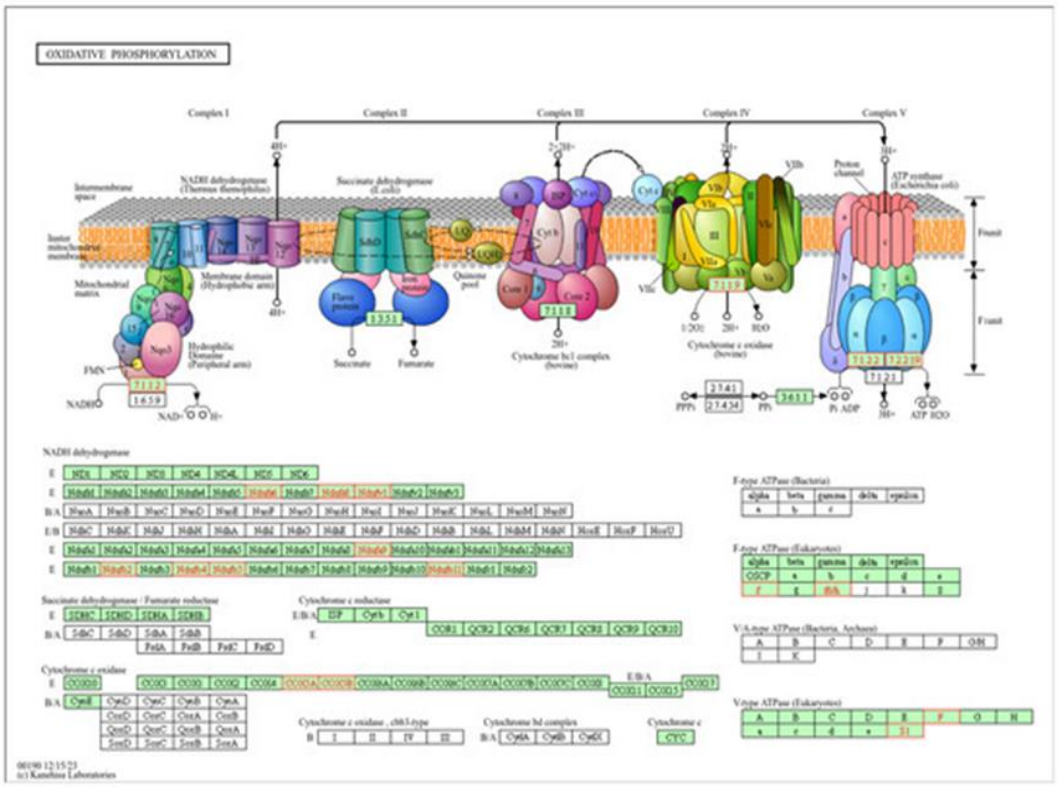


Figure 3. Overview of the oxidative phosphorylation (OXPHOS) pathway. Identified as the most enriched pathway in our analysis (FDR-adjusted $p < 0.001$), OXPHOS is illustrated in the upper panel adapted from KEGG (with permission), showing complexes I–IV embedded in the mitochondrial membrane. The lower panel provides a more detailed breakdown of pathway components. Components highlighted in red indicate those significantly associated with both MT-CYB variants and gene expression. Complex I (NADH dehydrogenase) transfers electrons from NADH while actively pumping protons across the inner mitochondrial membrane. Complex II (succinate dehydrogenase) channels electrons from succinate but does not directly contribute to proton transport. Complex III (ubiquinol-cytochrome c reductase, bc1 complex) shuttles electrons from ubiquinone and simultaneously pumps protons. Complex IV (cytochrome c oxidase) transfers electrons from cytochrome c to molecular oxygen, forming water and continuing proton translocation. ATP synthase (Complex V) harnesses the proton gradient generated by the electron transport chain to produce ATP from ADP and inorganic phosphate.

Development and evaluation of the predictive model

The five mtDNA single nucleotide variants identified during the discovery phase via multivariable regression were used to construct predictive models on two different platforms: (1) LASSO regression implemented in R (v4.4.1) **(Figures 4a and 4b)** and (2) MATLAB

(vR2023b) **(Figures 4c and 4d)**. Using the University of Iowa dataset, the LASSO-based model achieved an AUC of 0.91 (95% CI: 0.82–1.00) **(Figure 4a)**, whereas the MATLAB model reached an AUC of 0.95 (95% CI: 0.87–1.00) **(Figure 4e)**. The slightly higher AUC in the validation phase compared to the initial discovery dataset **(Figure 1b)** reflects that these analyses were performed using preselected variants from the same dataset.

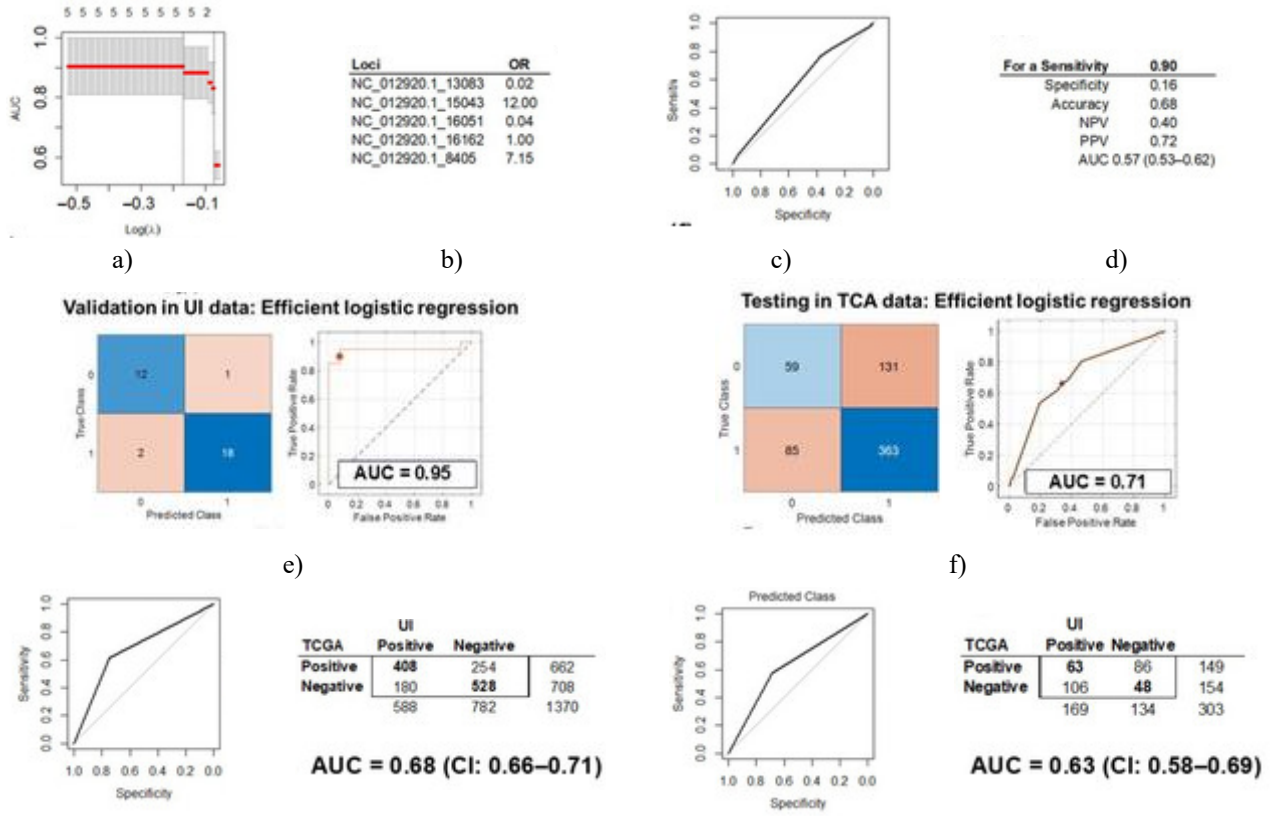


Figure 4. Model training, validation, and testing. (a) The LASSO model for predicting HGSC was validated on UI data using only five key mtDNA variants, achieving an AUC of 91% (CI: 82–100%). (b) The odds ratios of variants in the LASSO model indicate increased risk for HGSC when OR >1 and protective effects when OR <1. (c) ROC curve of the model tested on TCGA data using the pROC package. (d) Model performance on TCGA data, optimized for 0.9 sensitivity, reached an AUC of 0.57 (CI: 0.53–0.62). (e) Validation in UI data with MATLAB, using efficient logistic regression, shows the confusion matrix (left) and ROC curve (right), yielding an AUC of 0.95. (f) Testing on TCGA data produced a confusion matrix (left) and ROC curve (right) with an AUC of 0.71. (g) Correlation analysis of the MT-CYB variant with differential gene expression in HGSC versus controls indicated moderate agreement between UI and TCGA data (AUC = 0.68, CI: 0.66–0.71). (h) A 2 × 2 matrix from **Figure 4g** illustrates that 1370 genes were shared between UI and TCGA datasets. (i) Correlation of MT-CYB expression with gene expression revealed moderate concordance between the datasets (AUC = 0.63, CI: 0.58–0.69). (j) The 2 × 2 matrix for **Figure 4i** shows 303 genes common to both datasets.

After extracting and processing WES data from TCGA HGSC samples, the same five mtDNA variants used in the trained UI model were selected for testing. Using the LASSO model with TCGA data and pROC (v1.18.5) resulted in an AUC of 0.57 (CI: 0.53–0.62; **(Figure 4d)**). MATLAB-based testing on TCGA achieved better performance, with an AUC of 0.71 (95% CI: 0.53–0.88), partially overlapping the CIs from the UI validation **(Figure 4f)**.

Validation of correlation analysis

RNAseq data from TCGA HGSC samples were processed. Of the 3382 genes differentially expressed between HGSC and controls, 2716 transcripts were available in TCGA (423 samples). Among 1728 genes correlated with MT-CYB variation in the UI dataset, 1370 transcripts were present in TCGA, showing fair agreement (AUC = 0.68, CI: 0.66–0.71, **(Figure 4g)**), with 936 genes (68.3%) showing correlations in the same direction (r > 0 positive

or r < 0 negative; **(Figure 4h)**). From 364 genes significantly correlated with MT-CYB expression in UI, 303 were represented in TCGA, with lower concordance (AUC = 0.63, CI: 0.58–0.69, **(Figure 4i)**) and 111 genes (36.6%) showing directional agreement **(Figure 4j)**.

This pilot investigation aimed to identify mitochondrial genetic variations capable of distinguishing HGSC from normal tubal tissue. WES analyses revealed mtDNA variants that were used to build predictive models. A model incorporating five SNVs from distinct mitochondrial genes achieved an AUC of 0.88 (95% CI: 0.74–1.00). Validation across different analytical platforms using UI data yielded consistent results (AUCs: 0.91 and 0.95). When evaluated on TCGA samples with alternative machine learning approaches, AUCs reached up to 0.71 (95% CI: 0.53–0.88), partially overlapping the initial model's CI. While performance is fair [15] and insufficient alone for clinical diagnostics, integration with additional predictive models [16–18] could enhance ovarian cancer prediction. Reduced performance in TCGA testing may be influenced by the origin of control samples, which were taken from the same patients' normal tissues rather than independent normal tubes as in UI. The effect of control source on model accuracy is unclear, especially given mtDNA's tendency for spontaneous mutations in normal tissue. Differences in genetic substructure between TCGA and UI populations may also affect model performance [19]. Comprehensive comparisons require variant analyses across large, diverse populations with similar genetic backgrounds.

The ovarian cancer prediction model relied on five mtDNA single-nucleotide variants (SNVs), with the MT-CYB variant notably increasing the likelihood of HGSC by more than 30%. Expression analysis revealed a significant downregulation of MT-CYB in HGSC samples relative to controls, a pattern that remained consistent even after accounting for other mitochondrial genes in multivariable models. Importantly, MT-CYB alterations were confined to tumor tissues, suggesting its strong potential as a predictive marker for ovarian cancer. Examination of TCGA datasets showed that under 10% of control samples carried the MT-CYB variant, and in more than half of these, the mutation appeared both in the tumor and matched normal tissue. Cytochrome b, encoded by MT-CYB, is the only mtDNA-derived component of Complex III (ubiquinol:cytochrome c oxidoreductase), which resides in the inner mitochondrial membrane and functions as the second enzyme in oxidative phosphorylation, transferring electrons from ubiquinol to cytochrome c and driving proton translocation across the membrane. This highly conserved, hydrophobic protein contains two heme groups [20, 21]. MT-CYB mutations have been previously linked to ovarian carcinoma [22] and are believed to reprogram mitochondrial metabolism, elevating reactive oxygen species (ROS) production within tumor cells [23], potentially facilitating adaptation to hypoxic microenvironments [24]. Consistent with this, mtDNA mutations accumulate progressively from primary ovarian lesions to metastases, indicating the presence of driver mutations that may confer metastatic advantage [22].

Another notable variant in the predictive model was located in MT-ATP8, a mitochondrial gene recently detected in plasma and exosomes from highly aggressive lung cancers, highlighting its potential for liquid biopsy applications [25]. mtDNA is particularly prone to ROS-induced mutations even in normal tissue, giving rise to multiple coexisting mtDNA variants—a condition termed heteroplasmy [26]. Heteroplasmy refers to the presence of two or more mtDNA types within the same cell, which can occur in both normal and tumor tissues and vary between cells. Consequently, up to 72% of tumor-associated mtDNA variants are also observed in germline cells of healthy individuals [27]. This phenomenon underpinned our decision to include all variants in the ovarian cancer model, irrespective of origin, allowing the model to select variants most predictive of disease.

Multiple classes of mtDNA alterations are clinically relevant in tumorigenesis [28]. Both somatic and germline mtDNA mutations have been implicated in a variety of cancers, including renal, colon, head and neck, pancreatic, breast, ovarian, prostate, and bladder cancers [28, 29]. In prostate cancer, the overall mtDNA variant burden may serve as an indicator of tumorigenicity [30]. Variants in MT-CO1, for example, appear protective against ovarian cancer, consistent with observations in our study [31]. Although the functional impacts of many mtDNA variants remain unclear, analyses integrating mtDNA and nuclear DNA (nDNA) co-expression have consistently highlighted OXPHOS pathways as the most enriched across several cancer types [32]. In line with this, OXPHOS was the top-enriched pathway in our study when comparing HGSC tissues to normal fallopian tubes. Furthermore, genes associated with mtDNA variation were involved in diverse biological functions: some influence neurotransmitter oxidation at the outer mitochondrial membrane (MAOA), potentially affecting behavior [33]; others encode tumor-related proteins (CHIC1) [34]; and some maintain cellular homeostasis, such as GPR89A, which regulates intracellular pH [35].

Comparing gene expression profiles between normal fallopian tubes and HGSC samples appears to be the most reliable strategy for uncovering functional consequences of mtDNA variations, rather than using other normal tissues from the genital tract [36]. Unlike nuclear DNA, mitochondria harbor multiple copies of mtDNA, with replication, transcription, and translation regulated by both mitochondrial-encoded rRNAs and tRNAs and nDNA-

encoded proteins, allowing adaptation to environmental stressors [37]. Cancer cells can exploit these regulatory mechanisms to gain survival advantages, and understanding these processes could guide the development of targeted therapies [38]. The tumor microenvironment may also influence mitochondrial gene expression, as suggested by several studies [38]. In our analyses, some of the top transcripts linked to mtDNA variation corresponded to long non-coding RNAs (lncRNAs) and microRNAs implicated in epigenetic regulation across multiple cancers, including MIR4423, AC010280.1, AC010280.3, AL121839.2, and AC078883.2 [39–43]. Despite these findings, substantial gaps remain regarding how such alterations affect specific cancers in particular contexts. Additionally, numtogenesis—the integration of mtDNA sequences into nDNA—is activated in certain cancers, such as colorectal tumors, and may impact prognosis [44]. Although the precise mechanisms are not fully understood, mtDNA insertions within tumor suppressor genes could disrupt cellular pathways and promote oncogenesis [44], highlighting how subtle mitochondrial changes can profoundly influence nuclear function and cancer progression.

Circulating tumor DNA (ctDNA) and cell-free DNA (cfDNA) can be reliably isolated from the blood of ovarian cancer patients, and tumor-specific alterations have been detected in peripheral blood [45–48]. The identification of tumor-derived genetic changes in blood, termed liquid biopsy, has been applied in several cancers for individualized treatment, monitoring, and early detection, with colorectal and lung cancers recently receiving FDA approval [49, 50]. Since ctDNA is detectable in early-stage ovarian cancer, and our model predicts ovarian cancer with good accuracy, it could serve as an effective tool for early diagnosis. Moreover, this approach may allow detection of recurrent or persistent disease following adjuvant therapy, although the predictive model may require modification to account for mutations present in recurrent tumors. For clinical application in early-stage ovarian cancer, the prediction model would need to be prospectively validated in independent cohorts and include samples representing diverse disease stages to capture the full spectrum of genetic variation.

A major strength of this study is the use of WES to detect mtDNA variants in both HGSC patients and healthy controls without family history of ovarian cancer. Variant analyses followed recommended genome sequencing best practices, including validation, and model testing incorporated sufficient cases and controls from the TCGA HGSC dataset, all matching the tumor histology of the initial cohort. Multiple analytical platforms were employed to validate the prediction model, achieving acceptable performance levels.

However, limitations include the relatively small sample size, which may restrict the diversity of mtDNA variants and exclude potentially discriminative mutations, and may also widen the 95% CI for AUC, potentially inflating model performance. Additionally, the cohort lacked racial diversity, with only one of 20 ovarian cancer patients being Black and the remainder White (one unknown), limiting generalizability. To fully realize the potential of liquid biopsy-based mtDNA prediction, additional studies incorporating larger, racially diverse populations and multiple disease stages are necessary. Including other biomarkers, such as CA125, clinical parameters, or additional genomic data, could further enhance model performance. Prior work has demonstrated that integrating clinical, pathological, and genomic information improves prediction accuracy for chemotherapy response in ovarian cancer [2, 51, 52]. Ultimately, robust prediction models for ovarian cancer will require prospective multi-institutional validation to capture population variability and to establish their utility in early detection, following the precedent set by breast cancer studies [14].

Conclusion

This pilot, retrospective study identifies a set of mtDNA variations capable of distinguishing HGSC with strong performance. These findings represent a potential foundation for developing serum-based detection tools for ovarian cancer, including early-stage disease. Furthermore, the predictive model highlights links between HGSC-associated mtDNA variants and genes involved in OXPHOS pathways, suggesting broader implications for cellular metabolic and biological functions.

Materials and Methods

We conducted a single-center, retrospective, case–control pilot study utilizing tumor specimens collected during cytoreductive surgery from 112 patients with HGSC (cases), compared against benign fallopian tube specimens from 14 women undergoing surgery for non-malignant conditions (controls). DNA and RNA were extracted from all samples. Whole-exome sequencing (WES) was performed on 20 HGSC cases and all 14 controls, while RNA sequencing (RNAseq) was carried out on 112 HGSC cases and 12 control fallopian tube samples.

Specimen collection

HGSC tumor samples and associated clinical data were obtained from the Department of Obstetrics and Gynecology and Gynecologic Oncology Biobank (IRB, ID#200209010), part of the Women's Health Tissue

Repository (WHTR, IRB, ID#201804817). All specimens were originally collected with written informed consent from adult patients under University of Iowa IRB guidelines. Tumor samples were reviewed by a board-certified pathologist, flash-frozen, and diagnosis was confirmed on paraffin-embedded sections from the time of surgery. All experimental protocols were approved by the University of Iowa Biomedical IRB-01.

Fallopian tube specimens were obtained from women undergoing gynecologic procedures for benign indications, primarily sterilization, with no significant cancer history aside from occasional skin squamous cell carcinoma. Fallopian tubes were selected as controls because they represent the most likely site of origin for HGSC [53–55], a strategy previously validated in ovarian cancer research [16]. DNA and RNA were isolated from the epithelial layer at the junction of the ampullary and fimbriated ends. Of the 20 normal fallopian tube samples collected, 12 yielded sufficient RNA for sequencing. RNA from both HGSC and control specimens had been previously extracted and purified [56], and WES was successfully performed on 14 fallopian tube specimens.

DNA sequencing

Genomic DNA (gDNA) was extracted from frozen tumor and fallopian tube tissues using the DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions. DNA yield and purity were evaluated via NanoDrop Model 2000 spectrophotometry and horizontal agarose gel electrophoresis. Whole-exome sequencing was outsourced to GeneWiz (Azenta, Chelmsford, MA, USA). Libraries were prepared with the Agilent SureSelect Human Exome Library V5 kit and sequenced on an Illumina HiSeq 2000 platform (2 × 150 bp) to an average depth of 100×. Raw reads were aligned to the human reference genome (hg38) using the Burrows–Wheeler Aligner [57]. Across samples, the mean Phred quality score was 37.76, and 89.96% of bases had a quality ≥30. Coverage analysis was conducted using GATK v4.6.1.0, and sequencing quality control was performed with FastQC v0.12.1.

RNA sequencing

Total RNA was extracted from HGSC and control specimens stored in the Biobank. The RNA isolation, processing, and sequencing workflow has been described previously [51, 58]. Briefly, RNA was purified using the mirVana kit (Thermo Fisher, Waltham, MA, USA), with quality assessed by Trinean Dropsense 16 spectrophotometry and Agilent 2100 Bioanalyzer. Samples with RNA integrity number (RIN) ≥7 were deemed suitable for sequencing. A total of 500 ng of RNA per sample was quantified using Qubit (Thermo Fisher), converted to cDNA, and ligated with sequencing adaptors using the Illumina TriSeq stranded total RNA library preparation kit (Illumina, San Diego, CA, USA). Sequencing was performed on an Illumina HiSeq 4000 platform using 150 bp paired-end sequencing by synthesis (SBS) chemistry at the Genome Facility of the University of Iowa Institute of Human Genetics (IIHG).

Analysis of single nucleotide variations (SNVs)

WES-derived DNA sequences were aligned to the human mitochondrial reference sequence (Revised Cambridge Reference Sequence [rCRS], GenBank NC_012920, https://www.mitomap.org/MITOMAP/HumanMitoSeq, accessed 20 December 2024) using BWA (v0.7.17-r1188). The resulting BAM files were processed with samtools (v1.19.2) [59], Picard toolkit (v2.27.1), and GATK (v4.6.1.0) [60] to generate Variant Call Format (VCF) files for downstream analyses, following recommended best practices in genome sequencing [61]. A comprehensive table of identified SNVs across all samples was compiled for subsequent analyses.

To determine which mitochondrial SNVs were most predictive of HGSC, multivariable LASSO regression was performed using the glmnet R package (v4.1-8). Internal validation was conducted via k-fold cross-validation, and the regularization parameter (λ) was optimized using bootstrapping to reduce overfitting in the context of a limited sample size [62]. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC) with 95% confidence intervals (CI), where 0.5 indicates no predictive power and 1.0 represents perfect discrimination. Because LASSO prioritizes feature selection rather than inference, variable-specific confidence intervals were estimated using the selectiveInference R package (v1.2.5), recognizing potential limitations due to overfitting.

Correlating mtDNA variants with gene expression and pathway enrichment

RNAseq reads were aligned to the human reference genome (hg38) using STAR (v2.7.11b) [63], generating BAM files for downstream quantification. Gene expression levels were computed with featureCounts [64] and normalized using DESeq2 (v1.38.3) [65], including log2 transformation. Gene annotation and variant identification were performed with ENSEMBL. Univariable analyses were conducted for all 37 mitochondrial genes harboring sequence variants to compare HGSC and control groups. Genes with adjusted p-values < 0.05 were included in multivariable logistic regression to identify those independently associated with HGSC ($p < 0.05$). Whole-genome differential expression analysis employed a stricter false discovery rate (FDR)-adjusted alpha threshold of 0.005.

Spearman's rank correlation was used to examine relationships between mtDNA SNVs and gene expression,

acknowledging that these variables are not fully independent. Statistical significance was assessed via p-values and corrected for multiple comparisons using FDR [66]. Genes showing significant correlation with mtDNA variants were subjected to pathway enrichment analysis using clusterProfiler (v4.3.3) [67], interrogating KEGG pathways (https://www.genome.jp/kegg/pathway.html, accessed 17 December 2024). Only pathways with FDR-corrected p-values < 0.05 were considered significant.

Prediction model evaluation

Validation using TCGA dataset

The HGSC TCGA cohort was employed to test the predictive model. Access to controlled WES data was granted via the Genomic Data Commons Data Portal (dbGaP# 29868). WES data from 448 HGSC tumors and 190 matched normal controls were processed using BWA, samtools, Picard (v2.27.1), and GATK (v4.6.1.0) to identify SNVs and generate VCF files. Informative mtDNA SNVs identified from the UI prediction model were located within the TCGA dataset to test the model's predictive performance. The glmnet R package (v4.1-8) was used to rebuild the model using the most informative SNVs, and pROC (v1.18.5) was applied to assess performance in this independent dataset.

RNAseq BAM files from 423 TCGA HGSC samples were also downloaded. Gene expression was extracted using STAR and featureCounts (v2.0.6), and genes previously identified as significantly correlated in the UI dataset were examined for associations with MT-CYB SNVs and expression changes. Normal tube gene expression from the UI dataset served as controls, and all data were normalized and log2-transformed. Correlation between SNVs and gene expression in TCGA was performed using Spearman's rank test, with FDR correction applied for multiple comparisons [66].

Testing on an independent analytical platform

The HGSC prediction model was further evaluated using MATLAB (v2024b) machine learning tools, employing the classification learner with over 30 different classifier algorithms. The UI-trained model, including only the most informative SNVs, was initially trained and validated. The same SNVs from the TCGA dataset were then input into MATLAB to test the model across this independent analytical platform. Testing accounted for outcome weighting and class imbalance due to the smaller number of controls relative to cases.

Acknowledgments: The authors would like to acknowledge the work of the University of Iowa Core laboratories.

Conflict of interest: None

Financial support: This work was supported in part by the NIH 5R01CA99908-18 (K. Leslie PI), Department of Defense OC190352 (K. Leslie PI), and by the Research Fund of the Gynecologic Oncology Division of the University of Iowa Hospitals and Clinics. Also, was supported in part by the American Association of Obstetricians and Gynecologists Foundation (AAOGF) Bridge Funding Award.

Ethics statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of University of Iowa (IRB ID#200209010, approved on 09/19/05; IRB ID#201809807, approved on 04/10/2019). Informed consent was obtained from all subjects involved in the study.

References

1. Elias KM, Guo J, Bast RC Jr. Early detection of ovarian cancer. Hematol Oncol Clin North Am. 2018;32(6):903–14.
2. Gonzalez Bosquet J, Newtson AM, Chung RK, Thiel KW, Ginader T, Goodheart MJ, et al. Prediction of chemo-response in serous ovarian cancer. Mol Cancer. 2016;15(1):66.
3. US Preventive Services Task Force, Grossman DC, Curry SJ, Owens DK, Barry MJ, Davidson KW, et al. Screening for ovarian cancer: US Preventive Services Task Force recommendation statement. JAMA. 2018;319(6):588–94.
4. Lam ET, Bracci PM, Holly EA, Chu C, Poon A, Wan E, et al. Mitochondrial DNA sequence variation and risk of pancreatic cancer. Cancer Res. 2012;72(3):686–95.
5. Kabekkodu SP, Bhat S, Mascarenhas R, Mallya S, Bhat M, Pandey D, et al. Mitochondrial DNA variation analysis in cervical cancer. Mitochondrion. 2014;16:73–82.
6. Nakai T, Sakurada A, Endo T, Kobayashi H, Masuda S, Makishima M, et al. Caution for simple sequence repeat number variation in the mitochondrial DNA D-loop to determine cancer-specific variants. Oncol Lett. 2019;17(2):1883–8.
7. Gentiluomo M, Katzke VA, Kaaks R, Tjonneland A, Severi G, Perduca V, et al. Mitochondrial DNA copy-number variation and pancreatic cancer risk in the prospective EPIC cohort. Cancer Epidemiol Biomarkers Prev. 2020;29(4):681–6.
8. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: Monitoring cancer-genetics in the blood. Nat Rev Clin Oncol. 2013;10(8):472–84.
9. Phallen J, Sausen M, Adleff V, Leal A, Hruban C, White J, et al. Direct detection of early-stage cancers

using circulating tumor DNA. Sci Transl Med. 2017;9(403):eaan2415.
10. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science. 2018;359(6378):926–30.
11. Hou JM, Krebs MG, Lancashire L, Sloane R, Backen A, Swain RK, et al. Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer. J Clin Oncol. 2012;30(5):525–32.
12. Casagrande GMS, Silva MO, Reis RM, Leal LF. Liquid biopsy for lung cancer: Up-to-date and perspectives for screening programs. Int J Mol Sci. 2023;24(3):2505.
13. Zhu G, Pei L, Xia H, Tang Q, Bi F. Role of oncogenic KRAS in the prognosis, diagnosis and treatment of colorectal cancer. Mol Cancer. 2021;20(1):143.
14. Mazzeo R, Sears J, Palmero L, Bolzonello S, Davis AA, Gerratana L, et al. Liquid biopsy in triple-negative breast cancer: Unlocking the potential of precision oncology. ESMO Open. 2024;9:103700.
15. Nahm FS. Receiver operating characteristic curve: Overview and practical use for clinicians. Korean J Anesth. 2022;75(1):25–36.
16. Gonzalez-Bosquet J, Cardillo ND, Reyes HD, Smith BJ, Leslie KK, Bender DP, et al. Using genomic variation to distinguish ovarian high-grade serous carcinoma from benign fallopian tubes. Int J Mol Sci. 2022;23(23):14814.
17. Roy T, Oliveira S, Gonzalez Bosquet J, Wu X. 3D supervised contrastive-learning network for classification of ovarian neoplasms. In: Medical Imaging with Deep Learning 2023; 2023 Jul 10; Nashville, TN, USA.
18. Linder K, Watson R, Ulmer K, Bender D, Goodheart MJ, Devor E, et al. Prediction of ovarian cancer with deep machine learning and alternative splicing. Med Res Arch. 2023;11(11):1–16.
19. Miller MD, Devor EJ, Salinas EA, Newtson AM, Goodheart MJ, Leslie KK, et al. Population substructure has implications in validating next-generation cancer genomics studies with TCGA. Int J Mol Sci. 2019;20(5):1192.
20. Esposti MD, De Vries S, Crimi M, Ghelli A, Patarnello T, Meyer A. Mitochondrial cytochrome b: Evolution and structure of the protein. Biochim Biophys Acta. 1993;1143(3):243–71.
21. Weiss H, Linke P, Haiker H, Leonard K. Structure and function of the mitochondrial ubiquinol: Cytochrome c reductase and NADH: Ubiquinone reductase. Biochem Soc Trans. 1987;15:100–2.
22. Xie F, Guo W, Wang X, Zhou K, Guo S, Liu Y, et al. Mutational profiling of mitochondrial DNA reveals an epithelial ovarian cancer-specific evolutionary pattern contributing to high oxidative metabolism. Clin Transl Med. 2024;14(1):e1523.
23. Hahn A, Zuryn S. Mitochondrial genome (mtDNA) mutations that generate reactive oxygen species. Antioxidants (Basel). 2019;8(10):392.
24. Klimova T, Chandel NS. Mitochondrial complex III regulates hypoxic activation of HIF. Cell Death Differ. 2008;15(4):660–6.
25. Lou C, Ma X, Chen Z, Zhao Y, Yao Q, Zhou C, et al. The mtDNA fragments within exosomes might be novel diagnostic biomarkers of non-small cell lung cancer. Pathol Res Pract. 2023;249:154718.
26. Jimenez-Morales S, Perez-Amado CJ, Langley E, Hidalgo-Miranda A. Overview of mitochondrial germline variants and mutations in human disease: Focus on breast cancer. Int J Oncol. 2018;53(3):923–36.
27. Larman TC, DePalma SR, Hadjipanayis AG, Cancer Genome Atlas Research N, Protopopov A, Zhang J, et al. Spectrum of somatic mitochondrial mutations in five cancers. Proc Natl Acad Sci U S A. 2012;109(35):14087–91.
28. Kopinski PK, Singh LN, Zhang S, Lott MT, Wallace DC. Mitochondrial DNA variation and cancer. Nat Rev Cancer. 2021;21(7):431–45.
29. Wallace DC. Mitochondria and cancer. Nat Rev Cancer. 2012;12(10):685–98.
30. Kalsbeek AM, Chan EF, Grogan J, Petersen DC, Jaratlerdsiri W, Gupta R, et al. Mutational load of the mitochondrial genome predicts pathological features and biochemical recurrence in prostate cancer. Aging (Albany NY). 2016;8(11):2702–12.
31. Permuth-Wey J, Chen YA, Tsai YY, Chen Z, Qu X, Lancaster JM, et al. Inherited variants in mitochondrial biogenesis genes may influence epithelial ovarian cancer risk. Cancer Epidemiol Biomarkers Prev. 2011;20(6):1131–45.
32. Yuan Y, Ju YS, Kim Y, Li J, Wang Y, Yoon CJ, et al. Comprehensive molecular characterization of mitochondrial genomes in human cancers. Nat Genet. 2020;52(3):342–52.
33. Chen K, Holschneider DP, Wu W, Rebrin I, Shih JC. A spontaneous point mutation produces monoamine oxidase A/B knock-out mice with greatly elevated monoamines and anxiety-like behavior. J Biol Chem. 2004;279(38):39645–52.
34. Han W, Shi CT, Chen H, Zhou Q, Ding W, Chen F, et al. Role of LncRNA MIR99AHG in breast cancer: Bioinformatic analysis and preliminary verification. Heliyon. 2023;9(3):e19805.

35. Maeda Y, Ide T, Koike M, Uchiyama Y, Kinoshita T. GPHR is a novel anion channel critical for acidification and functions of the Golgi apparatus. Nat Cell Biol. 2008;10(9):1135–45.
36. Marquez RT, Baggerly KA, Patterson AP, Liu J, Broaddus R, Frumovitz M, et al. Patterns of gene expression in different histotypes of epithelial ovarian cancer correlate with those in normal fallopian tube, endometrium, and colon. Clin Cancer Res. 2005;11(17):6116–26.
37. Cogliati S, Lorenzi I, Rigoni G, Caicci F, Soriano ME. Regulation of mitochondrial electron transport chain assembly. J Mol Biol. 2018;430(24):4849–73.
38. Berner MJ, Wall SW, Echeverria GV. Deregulation of mitochondrial gene expression in cancer: Mechanisms and therapeutic opportunities. Br J Cancer. 2024;131(10):1415–24.
39. Yi C, Zhang X, Chen X, Huang B, Song J, Ma M, et al. A novel 8-genome instability-associated lncRNAs signature predicting prognosis and drug sensitivity in gastric cancer. Int J Immunopathol Pharmacol. 2022;36:3946320221103195.
40. Jing Z, Guo S, Zhang P, Liang Z. LncRNA-associated ceRNA network reveals novel potential biomarkers of laryngeal squamous cell carcinoma. Technol Cancer Res Treat. 2020;19:1533033820985787.
41. Ferrasi AC, Fernandez GJ, Grotto RMT, Silva GF, Goncalves J, Costa MC, et al. New LncRNAs in chronic hepatitis C progression: From fibrosis to hepatocellular carcinoma. Sci Rep. 2020;10(1):9886.
42. Zhu P, Pei Y, Yu J, Ding W, Yang Y, Liu F, et al. High-throughput sequencing approach for the identification of lncRNA biomarkers in hepatocellular carcinoma and revealing the effect of ZFAS1/miR-150-5p on hepatocellular carcinoma progression. PeerJ. 2023;11:e14891.
43. Perdomo C, Campbell JD, Gerrein J, Tellez CS, Garrison CB, Walser TC, et al. MicroRNA 4423 is a primate-specific regulator of airway epithelial cell differentiation and lung carcinogenesis. Proc Natl Acad Sci U S A. 2013;110(47):18946–51.
44. Srinivasainagendra V, Sandel MW, Singh B, Sundaresan A, Mooga VP, Bajpai P, et al. Migration of mitochondrial DNA in the nuclear genome of colorectal adenocarcinoma. Genome Med. 2017;9(1):31.
45. Martignetti JA, Camacho-Vanegas O, Priedigkeit N, Camacho C, Pereira E, Lin L, et al. Personalized ovarian cancer disease surveillance and detection of candidate therapeutic drug target in circulating tumor DNA. Neoplasia. 2014;16(2):97–103.
46. Oikkonen J, Zhang K, Salminen L, Schulman I, Lavikka K, Andersson N, et al. Prospective longitudinal ctDNA workflow reveals clinically actionable alterations in ovarian cancer. JCO Precis Oncol. 2019;3:1–12.
47. Vanderstichele A, Busschaert P, Smeets D, Landolfo C, Van Nieuwenhuysen E, Leunen K, et al. Chromosomal instability in cell-free DNA as a highly specific biomarker for detection of ovarian cancer in women with adnexal masses. Clin Cancer Res. 2017;23(9):2223–31.
48. Nakabayashi M, Kawashima A, Yasuhara R, Hayakawa Y, Miyamoto S, Iizuka C, et al. Massively parallel sequencing of cell-free DNA in plasma for detecting gynaecological tumour-associated copy number alteration. Sci Rep. 2018;8(1):11205.
49. Gupta R, Othman T, Chen C, Sandhu J, Ouyang C, Fakih M, et al. Guardant360 circulating tumor DNA assay is concordant with FoundationOne next-generation sequencing in detecting actionable driver mutations in anti-EGFR naive metastatic colorectal cancer. Oncologist. 2020;25(3):235–43.
50. Rolfo C, Mack P, Scagliotti GV, Aggarwal C, Arcila ME, Barlesi F, et al. Liquid biopsy for advanced NSCLC: A consensus statement from the International Association for the Study of Lung Cancer. J Thorac Oncol. 2021;16(10):1647–62.
51. Gonzalez Bosquet J, Devor EJ, Newtson AM, Smith BJ, Bender DP, Goodheart MJ, et al. Creation and validation of models to predict response to primary treatment in serous ovarian cancer. Sci Rep. 2021;11(1):5957.
52. Gonzalez Bosquet J, Marchion DC, Chon H, Lancaster JM, Chanock S. Analysis of chemotherapeutic response in ovarian cancers using publically available high-throughput data. Cancer Res. 2014;74(14):3902–12.
53. Erickson BK, Conner MG, Landen CN Jr. The role of the fallopian tube in the origin of ovarian cancer. Am J Obstet Gynecol. 2013;209(5):409–14.
54. Shih IM, Wang Y, Wang TL. The origin of ovarian cancer species and precancerous landscape. Am J Pathol. 2021;191(1):26–39.
55. Ploner A. Heatplus: Heatmaps with row and/or column covariates and colored clusters. Available from: https://github.com/alexploner/Heatplus. Accessed 20 Dec 2025.
56. Reyes HD, Devor EJ, Warrier A, Newtson AM, Mattson J, Wagner V, et al. Differential DNA methylation in high-grade serous ovarian cancer (HGSOC) is associated with tumor behavior. Sci Rep. 2019;9(1):17996.
57. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

58. Miller MD, Salinas EA, Newtson AM, Sharma D, Keeney ME, Warrier A, et al. An integrated prediction model of recurrence in endometrial endometrioid cancers. Cancer Manag Res. 2019;11:5301–15.
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
60. Alganmi N, Abusamra H. Evaluation of an optimized germline exomes pipeline using BWA-MEM2 and Dragen-GATK tools. PLoS ONE. 2023;18(3):e0288371.
61. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinform. 2013;43(1):11.10.1–11.10.33.
62. Riley RD, Snell KIE, Martin GP, Whittle R, Archer L, Sperrin M, et al. Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. J Clin Epidemiol. 2021;132:88–96.
63. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
64. Liao Y, Smyth GK, Shi W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.
65. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:106.
66. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100(16):9440–5.
67. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology. 2012;16(5):284-7.