Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Feasibility of whole-genome sequencing using cytology samples for cancer genomic analysis: a pilot study with xenograft and clinical samples

Yoshiki Shinomiya
Hokkaido University Hospital
Kanako C Hatanaka
Hokkaido University Hospital
Kotoe Katayama
The University of Tokyo
Asami Okumura
Hokkaido University Hospital
Ayae Nange
Hokkaido University Hospital
Hiroyuki Yamazaki
Hokkaido University
Hiroshi Asano
Hokkaido University
Toru Kondo
Hokkaido University
Ichiro Kinoshita
Hokkaido University
Hidemichi Watari
Hokkaido University
Seiya Imoto
The University of Tokyo
Yutaka Hatanaka
yhatanaka@huhp.hokudai.ac.jp

Hokkaido University Hospital

Research Article

Keywords: Cytology, Whole-Genome Sequencing, Genomics, RNA sequencing, Xenograft

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11226628/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. Kanako C. Hatanaka has received grants or contracts from Eli Lilly, NEC Corporation, Sakura Finetek Japan Co., Ltd., and Chugai Pharmaceutical Co., Ltd. Ichiro Kinoshita has received grants or contracts from Konica Minolta REALM, Eisai, Daiichi Sankyo, and Bayer. He has also received payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing, or educational events from MSD Pharmaceutical, Bayer, Takeda Pharmaceutical, Konica Minolta REALM, Chugai Pharmaceutical Co., Ltd., Novartis Pharma, Guardant Health Japan, and Pfizer Japan. Yutaka Hatanaka has received grants or contracts from Eli Lilly, NEC Corporation, CURED, Konica Minolta REALM, Daiichi Sankyo, BostonGene, Chugai Pharmaceutical Co., Ltd., and biomy Inc. He has also received payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing, or educational events from AstraZeneca and Daiichi Sankyo. All other authors declare no conflicts of interest.

Feasibility of whole-genome sequencing using cytology samples for cancer genomic analysis: a pilot study with xenograft and clinical samples

Authors' names: Yoshiki Shinomiya[1], Kanako C Hatanaka[1,2], Kotoe Katayama[3], Asami Okumura[1], Ayae Nange[1], Hiroyuki Yamazaki[4], Hiroshi Asano[4], Toru Kondo[5], Ichiro Kinoshita[6], Hidemichi Watari[4], Seiya Imoto[3,7], Yutaka Hatanaka[1]*

Authors' affiliations:

1 Center for Development of Advanced Diagnostics, Hokkaido University Hospital, Sapporo, Japan

2 Department of Surgical Pathology, Hokkaido University Hospital, Sapporo, Japan

3 Laboratory of Sequence Analysis, Human Genome Center, the Institute of Medical Science, the University of Tokyo, Tokyo, Japan

4 Department of Obstetrics and Gynecology, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Sapporo, Japan

5 Division of Stem Cell Biology, Institute for Genetic Medicine, Hokkaido University, Sapporo, Japan

6 Division of Clinical Cancer Genomics, Hokkaido University Hospital, Sapporo, Japan

7 Division of Health Medical Intelligence, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Corresponding author: Yutaka Hatanaka, Center for Development of Advanced Diagnostics, Hokkaido University, N15W7, Kita-ku, Sapporo, Hokkaido 060-8638, Japan, Tel: +81-11-716-1161

E-mail: yhatanaka@huhp.hokudai.ac.jp

Abstract

Background: Whole-genome sequencing (WGS) offers comprehensive genomic profiling but generally requires high-quality nucleic acids from fresh-frozen (FF) tissue. For broader application of genome-scale cancer analysis, more readily available sample types are needed. Cytology samples may provide better nucleic acid quality than formalin-fixed paraffin-embedded tissue and therefore represent a potential alternative source for WGS. We evaluated the feasibility of WGS using cytology samples by comparison with matched FF samples.

Methods: WGS and RNA sequencing were performed using cytology and FF tumor samples from a HER2-positive BT-474 xenograft model and a patient with endometrioid carcinoma. Concordance of SNV, CNA, and SV findings between cytology and FF samples was assessed.

Results: For SNVs, filtering with variant allele frequency ≧0.1 yielded nearly 90% concordance between patient FF and cytology samples. For CNAs, ERBB2 amplification was detected in xenograft-derived cytology samples, supporting the utility of cytology samples for copy number analysis, although no clinically relevant CNAs were identified in the clinical sample. Structural variant analysis showed limited concordance.

Conclusions: Cytology samples may represent a feasible alternative source for WGS-based short-variant analysis when fresh-frozen tissue is unavailable, whereas CNA and especially structural variant detection require further optimization and complementary approaches.

Keywords: Cytology, Whole-Genome Sequencing, Genomics, RNA sequencing, Xenograft

Introduction

Whole-genome sequencing (WGS) enables comprehensive analysis of genomic alterations across the entire genome and is being considered for broader clinical implementation. In addition to coding regions, WGS can detect alterations in non-coding regions, which may contribute to tumor biology and therapeutic stratification. WGS therefore has substantial potential not only for research applications, such as the identification of novel disease mechanisms, but also for clinical use through the detection of a wider range of therapeutically relevant alterations [1-4]. Furthermore, the prospective accumulation of whole-genome data from clinical cases may expand future treatment opportunities and deepen our understanding of cancer biology.

Although formalin-fixed paraffin-embedded (FFPE) tissue is the standard sample type for current panel and comprehensive genomic profiling (CGP) testing in clinical settings, fresh-frozen (FF) tissue remains the gold standard for WGS because genome-wide analysis requires higher-quality nucleic acids. Panel and CGP assays focus on predefined genomic regions and sequence them at high depth, allowing successful analysis even with partially degraded nucleic acids. In contrast, applying similarly high coverage across the entire genome is not currently practical for routine WGS, making high nucleic acid quality essential. However, FF tissue is difficult to handle and store in routine clinical practice. Therefore, for the clinical implementation of WGS, it would be advantageous if more commonly available FFPE tissue could also be used.

Cytology samples represent one such candidate material. Most cytology samples are alcohol-fixed, which reduces the nucleic acid damage associated with formalin fixation [5, 6]. As a result, cytology samples generally yield higher-quality nucleic acids than FFPE [7-9]. Previous studies have reported favorable nucleic acid quality in cytology samples and their successful application in NGS.[10-13] In some clinical settings in Japan, cytology samples are already used as alternative materials for multiplex-CDx based NGS [14-17]. However, evidence regarding their applicability to WGS remains limited.

To address this gap, we conducted a pilot study to investigate the feasibility of WGS using cytology samples. By comparing WGS results obtained from cytology samples with those from matched FF tissue, we aimed to

assess whether cytology samples could serve as a feasible sample type for comprehensive genome-wide analysis.

Material and Methods

Sample preparation and extraction of nucleic acids

A HER2-positive BT-474 (RRID: CVCL_0291) xenograft model and a tumor specimen from a patient with endometrioid carcinoma were included in this study. The cell line was purchased from American Type Culture Collection (ATCC, Manassas, VA, USA). The xenograft model was established using 6-week-old NOD/Shi-scid, IL-2RγKO Jic mice purchased from In-Vivo Science Inc (Tokyo, Japan). The mice were euthanized by cervical dislocation without anesthesia by trained and experienced personnel, in accordance with the institutional animal experimentation protocol. In the xenograft model, WGS data from the parental BT-474 cell line were used as the reference. Whole blood was used as a normal sample for the patient case. Materials handled under two different conditions were prepared from each resected tumor: 1) immediately flash-frozen samples (FF samples); and 2) imprint cytology samples fixed in 95% ethanol. To examine the internal structure of the tumor, the resected mass was sectioned to expose the cut surface. The exposed surface was macroscopically observed, and imprint cytology samples were obtained by gently touching and scraping the tumor surface with a glass slide, followed by immediate fixation in 95% ethanol. Subsequently, the tumor-containing area was excised and rapidly frozen at –80°C in acetone using Histo-Tek PINO device (Sakura Finetek Japan Co., Ltd., Tokyo, Japan). For FF samples, a pathologist confirmed the presence of sufficient tumor cells on hematoxylin and eosin-stained frozen sections. For cytology samples, tumor cell adequacy was confirmed on Papanicolaou-stained slides in collaboration with a cytotechnologist.

Nucleic acid extraction from tumor samples was performed using the AllPrep DNA/RNA Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. For normal samples derived from whole

blood, nucleic acids were extracted using the QIAsymphony system (QIAGEN) with the DSP DNA Midi Kit (QIAGEN). The concentration of nucleic acids was measured by both NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA) and Qubit (Thermo Fisher Scientific). TapeStation (Agilent Technologies, Santa Clara, CA, USA) was used to check nucleic acid yields, using DNA integrity number (DIN) for DNA and DV200 for RNA as an indicator of degradation.

Library preparation and sequencing

For WGS, DNA libraries were prepared from 300 ng of input DNA using the Illumina DNA PCR-Free Library Prep, Tagmentation kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. For RNA sequencing (RNA-seq), libraries were prepared from 100 ng of DNase-treated total RNA using the Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus kit (Illumina). Libraries were quantified using the Qubit ssDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) or the KAPA Library Quantification Kit (Roche Diagnostics, Basel, Switzerland) and sequenced on the NovaSeq 6000 system (Illumina) with 150-bp paired-end reads. For WGS, target coverage was 120× for tumor samples and 30× for matched normal samples.

Somatic variant calling and genomic analyses

Sequence reads were processed according to the GATK Best Practices pipeline. After trimming and quality assessment, reads were aligned to the GRCh38 reference genome using BWA-MEM (v0.7.15), followed by duplicate marking with Picard MarkDuplicates and base quality score recalibration (BQSR). Alignment quality and sequencing characteristics were evaluated using Picard CollectMultipleMetrics and CollectWgsMetrics. Somatic single-nucleotide variants (SNVs) and indels were called using Mutect2 in GATK (v4.1.0.0), copy number alterations (CNAs) were analyzed using CNVkit (v0.9.10), and structural variants (SVs) were detected using Manta (v1.6.0). Microsatellite instability (MSI) scores were calculated using mimcall. MSI-high status was classified using a predefined cutoff based on previously reported MSIsensor criteria [18]. For xenograft

samples, tumor-only analysis was performed, whereas tumor–normal matched analysis was used for the clinical sample. Fusion gene detection from RNA-seq data was performed using STAR-Fusion (v1.10.1). Single-base substitution (SBS) signature analysis was conducted using the R package deconstructSigs (v1.9.0), with signatures nature2013 used as the reference.[19] Circos plots were generated using shinyCircos.[20] Sample identity between paired samples was verified using NGSCheckMate. All tools were run with default parameters unless otherwise specified.

Comprehensive genomic profiling results in the clinical case

For the clinical sample, comprehensive genomic profiling was also performed using the GenMineTOP panel (GenMine Labs Corp., Tokyo, Japan), a clinically approved assay used in routine clinical practice in Japan.

Statistical analysis

All statistical analyses in this study were performed using R (v4.3.3) or Python (v3.12.10). Data are presented as mean ± standard deviation (SD) unless otherwise specified. Group comparisons were performed using the appropriate statistical tests as described in each figure legend.

Results

High-quality nucleic acids and comparable WGS metrics in FF and cytology samples

High-quality DNA was successfully extracted from all FF and cytology samples. For FF samples (n=2), DNA concentrations were 490 and 962 ng/µL (as measured by Qubit), with DIN of 8.9 and 8.6. For cytology samples (n=2), DNA concentrations were 9.4 and 86.8 ng/µL (as measured by Qubit), and both samples had a DIN of 6.9, indicating minimal degradation and sufficient quality for WGS. Similarly, high-quality RNA was obtained from all samples. For FF samples (n=2), RNA concentrations were 746.0 and 1272.0 ng/µL, with

DV200 of 96.8% and 95.3%. For cytology samples (n=2), RNA concentrations were 306.0 and 550.0 ng/μL, with DV200 of 92.2% and 96.2%, reflecting excellent RNA integrity (Table 1). Although there were case-specific variations, all samples yielded sufficient nucleic acids for WGS. In terms of quality, all samples were adequate, but cytology samples tended to be slightly lower than FF samples in both patient and xenograft samples.

Sample Type	Condition	DNA conc. (ng/μL): ND	DNA conc. (ng/μL): QB	DIN	RNA conc. (ng/μL): ND	RNA conc. (ng/μL): QB	DV200 (%)
Xenograft	FF	482.5	490.0	8.9	680.4	746.0	96.8
Xenograft	Cytology	17.7	9.4	6.9	320.4	306.0	92.2
Human	FF	1328.7	962.0	8.6	1521.2	1272.0	95.3
Human	Cytology	167.4	86.8	6.9	573.6	550.0	96.2

Table 1. Summary of nucleic acid yield and quality in this study.
Abbreviation: conc., Concentration; DIN, DNA integrity number; FF, Fresh frozen; ND, Nanodrop; QB, Qubit
Note*: "Human" indicates an endometrial carcinoma patient; "Xenograft" refers to a mouse xenograft model established using the human breast cancer cell line BT-474.*

WGS was successfully performed on all samples. All samples showed quality metrics sufficient for downstream analysis. Total read counts were 3.39 ± 0.09 billion, with overall alignment rates of 99.98 ± 0.03% and paired-end alignment rates also above 99.9% in all samples. Insert sizes were smaller in cytology samples (392.2 ± 175 and 404.0 ± 181.8 bp) than in FF samples (429.3 ± 193.7 and 430.2 ± 192.3 bp), with differences of approximately 30–40 bp. Base-level error rates were 0.48 ± 0.02%, mismatch rates were 0.53± 0.03%, and improperly paired reads accounted for 4.97 ± 3.59% of total reads. Chimeric reads were 7.04 ± 3.83%, and soft-clipped reads 4.55 ± 3.14%. Strand balance was consistently close to 0.5 across all samples, indicating minimal sequencing bias. Cytology and FF samples within each case showed highly comparable sequencing metrics. Representative quality metrics are summarized in Table 2.

Cytology samples demonstrated comparable performance to FF in xenograft WGS

In the xenograft model, we analyzed the feasibility of cytology samples for WGS, using cell pellet (CP)–based WGS results as a reference and comparing them with FF samples in a preclinical setting. WGS using CP

identified 154 SNVs in the BT-474 cell line, whereas 230 and 185 SNVs were detected in the FF and cytology samples, respectively. In most variant classifications, the FF sample showed the greatest number of SNV calls (Fig. 1A). Of the identified SNVs, 110 were shared across all three sample types. These included three SNVs (BRCA2 p.S3094*, MAPK1 p.H61Q, TP53 p.E285K) annotated in Cellosaurus (https://www.cellosaurus.org/index.html) as being present in BT-474 and considered important in breast cancer. Unique SNVs numbered 22, 79, and 34 in the CP, FF, and cytology samples, respectively. In addition, 30 SNVs were shared between FF and cytology samples but were not called in CP (Fig. 1B). After filtering SNVs with variant allele frequency (VAF) ≧0.1 and confirming their concordance, 94 SNVs were consistently identified across all 3 samples, whereas 24, 24, and 5 SNVs were uniquely detected in each sample, respectively (Fig. 1C). Using CP-based WGS as the reference, the recall rate was 78.6% for both FF and cytology samples. However, with respect to precision, the FF sample showed 52.6%, whereas the cytology sample exhibited a slightly higher value of 65.4%. The F1 score and overall concordance rate were also better for cytology samples (F1 score: 63.0% vs. 71.4%; concordance rate: 46.0% vs. 55.5%) (Fig. 1D). When restricted to SNVs with VAF ≧0.1, FF sample showed a recall of 72.7%, precision of 77.6%, F1 score of 75.1%, and concordance of 60.1%. In contrast, cytology samples demonstrated a recall of 76.2%, precision of 90.8%, F1 of 82.9%, and concordance of 70.8% (Fig. 1E).

In CNA analysis, 255 genes were consistently identified in all sample types. CNAs unique to CP, FF, and cytology samples numbered 14, 5, and 107 genes, respectively. Notably, the largest number of unique CNAs was observed in cytology samples (Fig. 1F). Using CP WGS as the reference, recall was 81.7% for cytology samples and 78.9% for FF samples. Conversely, precision was 75.2% for cytology samples and 98.4% for FF samples. Consequently, both the F1 score and the overall concordance rate were higher for FF samples (FF vs cytology samples: F1, 87.6% vs 78.3%; concordance, 77.9% vs. 64.4%) (Fig. 1G). BT-474 is a well-known HER2-positive breast cancer cell line. WGS analysis successfully detected ERBB2 (HER2) amplification at a similar copy number level across all sample types (Fig. 1H).

Gene symbol- and location-based analysis confirmed the consistency of WGS-detected rearrangements across the 3 sample types. At the gene level, 6 rearrangement events were commonly detected across all 3 sample types: CP, FF, and cytology samples. In addition, 8 rearrangement events were shared exclusively between FF and cytology samples. 1 rearrangement event was shared between FF and CP samples, and another was shared between cytology and CP samples. In contrast, sample type–specific rearrangement events were identified in FF, cytology, and CP samples, with nine, seven, and one events, respectively (Fig. 1I). In the performance analysis in which rearrangement events were matched based on gene symbols, the recall relative to CP samples was 77.8% for both FF and cytology samples. In contrast, precision was 29.2% for FF samples and 31.8% for cytology samples, and both the F1 score and overall concordance rate were slightly higher for cytology samples (F1 score: 42.4% for FF and 45.2% for cytology; overall concordance rate: 26.9% for FF and 29.2% for cytology) (Fig. 1J).

In contrast, in the analysis based on exact positional matching of rearrangement events, the recall relative to CP samples was 50.0% for both FF and cytology samples. Precision was 18.5% for FF samples and 17.9% for cytology samples, and the F1 score and overall concordance rate were largely comparable between the two sample types, although they were slightly higher for FF samples (F1 score: 27.0% for FF and 26.3% for cytology; overall concordance rate: 15.6% for FF and 15.2% for cytology) (Fig. 1K).

High concordance of SNV profiles between FF and cytology in clinical case

WGS was performed on paired cytology and FF samples obtained from a clinical case diagnosed with endometrioid carcinoma, in order to evaluate the concordance of SNV detection between the two sample types.

In this case, 4,651 SNVs were identified in the cytology sample, and 4,724 SNVs were detected in the corresponding FF sample. The total number of SNVs was comparable between the two samples, and no substantial differences were observed in the types of mutations identified (Fig. 2A). Among all detected

mutations, 3,652 SNVs (63.8%) were shared between cytology and FF samples. In contrast, 1,072 SNVs (18.7%) were detected only in the FF sample, and 999 SNVs (17.5%) were unique to the cytology sample (Fig. 2B). A comparison of variant allele frequencies revealed that SNVs detected exclusively in either the cytology or FF sample had significantly lower VAFs than those detected in both samples (Matched variants:0.46 ± 0.09, FF only variants: 0.08 ± 0.10, Cytology only variants: 0.12 ± 0.14; $p < 0.001$) (Fig. 2C). When the analysis was restricted to SNVs with VAF $\geqq$ 0.1, the number of shared mutations slightly decreased to 3,583; however, the concordance rate increased to 87.7%, indicating that high-confidence variants were more consistently detected between the two sample types (Fig. 2D). SBS signature analysis was also conducted. Signatures 1A, 6, and 15 were detected in both samples, and the relative contributions of each signature were highly similar between cytology and FF samples (Fig. 2E, F). MSI analysis revealed consistent results between sample types. MSI scores were 0.36 in the FF sample and 0.37 in the cytology sample, both exceeding the threshold for MSI-high classification (Table 3). Tumor mutation burden (TMB) was also evaluated using WGS data from paired FF and cytology samples. As shown in Table 3, the TMB values were 141.9 and 139.7 mutations per megabase (Mb) in the FF and cytology samples, respectively, when calculated based on a 33.3-Mb coding region. These values were highly consistent between sample types, and both samples were classified as TMB-high. The TMB values estimated by WGS were slightly higher than those estimated by CGP.

Table3. Comparison of TMB and Mutation Counts Derived from WGS and CGP in FF and Cytological Samples.

	WGS				CGP	
Condition	**MSI score**	**MSI**	**TMB (mut/Mb)**	**TMB**	**TMB (mut/Mb)**	**TMB**
FF	0.36	MSI-High	141.9	TMB-High	121.8	TMB-High
Cytology	0.37	MSI-High	139.7	TMB-High	127.1	TMB-High

Abbreviation: CGP, Comprehensive Genomic Profiling; FF, fresh frozen; MSI, microsatellite instability; TMB, Tumor mutation burden; WGS, whole-genome sequencing.

Pathogenic or likely pathogenic mutations showed strong agreement between clinical FF and cytology samples

To assess the concordance of clinically important mutations, the analysis focused on variants annotated as pathogenic or likely pathogenic in ClinVar (accessed 20211010). Pathogenic or likely pathogenic variants detected by WGS using cytology and FF samples showed an overall concordance of 74.1%. Among all detected pathogenic or likely pathogenic variants, 12.1% were observed only in the FF sample, and 13.8% were found exclusively in the cytology sample (Fig. 3A). When the analysis was limited to pathogenic or likely pathogenic variants with VAF $\geqq$ 0.1, the number of concordant variants remained the same; however, the concordance rate increased to 89.6%, indicating improved agreement for high-confidence mutations (Fig. 3B). All pathogenic or likely pathogenic variants detected in either or both sample types are summarized in Fig. 3C. Of the four variants identified only in cytology samples, three had VAFs around 0.2, which were slightly lower than those of concordant variants.

Comparison with CGP supported the consistency of WGS results between clinical FF and cytology samples

We compared the WGS results with CGP, which is already clinically implemented, and confirmed the consistency of SNVs supported by CGP. Regarding the CGP results, 76.6% showed concordance between FF and cytology samples (Fig. 4A). The recall rate of WGS for SNVs detected by CGP was 80.8% in FF and 78.2% in cytology samples, showing no clear difference (Fig. 4B). Furthermore, when assessing consistency using only the SNVs detected by CGP among those detected by WGS in each sample type, an 82.1% concordance rate was observed between FF and cytology samples (Fig. 4C).

Copy ratio and SV detection differences between FF and cytology samples in clinical case

CNA was not detected in the clinical case analyzed in this study. Upon examining the scatter plot of copy ratios obtained from CNA analysis, it was found that there was variability in the baseline values in cytology samples compared to FF samples (Fig. 5A, B).

In this study, SV analysis focused specifically on rearrangement. WGS-based alteration detection revealed a marked difference in the number of calls between sample types. A total of 8 rearrangements were detected

in the FF sample, whereas 64 alterations were identified in the cytology sample (Fig. 5C). To evaluate the methodological concordance, we also assessed the overlap between rearrangements detected by WGS and fusions detected by RNA-seq as a whole, without considering sample-specific differences. Only a limited number of alterations were commonly detected by both methods, indicating low concordance between WGS and RNA-seq in alteration detection (concordance rate: 1.3%, detected RNA-seq only: 5.3%, detected WGS only: 93.4%) (Fig. 5D). We evaluated the concordance of events identified by WGS and RNA-seq between sample types. The concordance rate for rearrangements detected by WGS was 2.9% (FF: n = 8; cytology: n = 64; shared: n = 2), while the concordance rate for fusion genes detected by RNA-seq was 66.7% (FF: n = 2; cytology: n = 3; shared: n = 2). Concordance between FF and cytology samples was extremely low in WGS, with only a limited subset of rearrangements identified in FF samples also detected in cytology samples. By contrast, RNA-seq showed moderate concordance (Fig. 5E). Circos plots were generated for each condition on the basis of the variant profiles obtained by WGS in this study (Fig. 5F–G). Labeled gene names represent detected pathogenic or likely pathogenic mutations. While SNV profiles, including targetable variants, were largely similar across conditions, rearrangement profiles differed substantially.

RNA-seq supported transcriptomic concordance between FF and cytology samples

Because cytology samples yielded RNA of sufficient quality, we further examined whether transcriptomic profiles were preserved between FF and cytology samples in the clinical case. Whole-transcriptome expression levels determined by RNA-seq showed a very strong correlation between the FF and matched cytology samples ($R = 0.95$, $p < 0.001$) (Fig. 6A). In addition, the expression levels of genes included in a previously reported endometrial carcinoma subtyping signature [21] were comparable between the two sample types (Fig. 6B). These findings suggest that cytology samples may also be suitable for transcriptome-based analyses and may provide complementary information to WGS, particularly when interpretation of copy number alterations or rearrangements requires additional transcriptomic support.

Discussion

In this pilot study, we assessed the feasibility of applying WGS to cytology samples by comparing them with FF samples, the current standard, using both xenograft models and clinical samples. Although cytology samples tended to show slightly lower DIN, shorter insert sizes, and substantially lower DNA and RNA yields than FF samples, nucleic acid quality remained generally acceptable, and no clear increase in metrics associated with artifactual variant calling was observed. These findings suggest that cytology samples can provide analytes of sufficient quality for WGS.

According to the short-variant analysis, most discordant variants between FF and cytology-based WGS concentrated in the low-VAF range, and when the analysis was restricted to SNVs with a VAF of at least 0.1, concordance increased to nearly 90%. As noted above, both sample types showed sufficient DNA quality for WGS analysis; however, cytology samples had slightly lower DIN values and shorter insert lengths than unfixed FF samples. These subtle differences in DNA quality may have partly contributed to the observed differences in SNV concordance, given that previous studies have suggested that DNA fragmentation and shorter insert sizes can affect variant detection [22-24]. However, because many variants uniquely detected in the FF samples were also low-VAF variants, the observed discordance cannot be explained by sample quality alone and was likely also influenced by intratumoral heterogeneity and differences in sampling sites. Several groups have reported the limitations of single-site sampling in the context of intratumoral heterogeneity [25, 26]. In fact, the recall of WGS against CGP did not differ substantially between FF and cytology samples, and even when variants supported by CGP were used as the reference, concordance remained at approximately 76%, indicating that part of the discordance may be attributable to intratumoral heterogeneity and sampling differences rather than sample type itself. Of note, in the xenograft model, mutations considered important in breast cancer were reliably detected from cytology samples, and in the clinical sample, high concordance was observed for pathogenic or likely pathogenic mutations. Taken together, these findings suggest that although cytology samples may have limitations in detecting low-VAF

297 variants, they may provide performance comparable to FF samples for clinically meaningful short variants.
298 Moreover, SBS signature analysis of SNVs derived from the patient's FF and cytology samples demonstrated
299 similar proportions of age-related and DNA mismatch repair deficiency-associated signatures, further
300 supporting the utility of cytology samples. Consistent with this, MSI analysis identified MSI-high in both
301 sample types, an important finding given its clinical significance as a biomarker for treatment eligibility.
302 Pathogenic or likely pathogenic mutations identified in this study included PTEN, PIK3CA, and MSH2, which
303 are frequently observed in endometrial cancer [21, 27]. Importantly, the MSH2 mutation was concordant
304 with both the SBS and MSI results, reinforcing the reliability of the overall findings.

305 With respect to CNAs, analysis using the xenograft model showed that the reproducibility of CNAs relative to
306 those identified in the original cultured cells was broadly similar between FF and cytology samples.
307 However, cytology samples showed a larger number of uniquely detected CNAs, and concordance between
308 FF and cytology samples remained low at approximately 50%. Copy ratio profiles derived from cytology
309 samples showed greater variability than those from FF samples, suggesting that this noise may adversely
310 affect CNA detection. In FFPE samples, optimization of the temperature conditions during nucleic acid
311 extraction has been reported to improve such noise-related issues [28]. However, because cytology samples
312 do not inherently require a decrosslinking process, it remains unclear whether a similar strategy would be
313 effective, and this should be addressed in future studies. On the other hand, the very high correlation
314 observed in RNA-seq expression profiles suggests that transcriptomic data may provide complementary
315 information when interpreting CNA-related alterations, particularly when copy number changes are
316 expected to be accompanied by expression differences. Thus, although the overall results for CNA analysis
317 were less favorable than those for short variants, the reproducibility observed in the xenograft model and
318 the successful detection of ERBB2 amplification in HER2-positive cell lines suggest that cytology samples may
319 still be useful for detecting at least some clinically relevant CNAs.

SV analysis in the xenograft model showed that reproducibility relative to the cultured-cell reference was 50% at the position-level but exceeded 70% at the gene-level, suggesting limited but meaningful utility. However, detection accuracy relative to the reference dataset was only about 30%, and concordance between FF and cytology samples was also low. Furthermore, in the clinical sample, the number of SVs detected in the cytology sample was approximately eight times higher than that in the FF sample, raising the possibility that at least some of these calls reflected instability or background noise. Similar difficulties in WGS-based SV detection have also been reported in FFPE [29, 30], suggesting that cytology samples may face comparable challenges. At the same time, given that detection performance was also low in FF samples from xenograft models, it is unlikely that these limitations can be explained solely by factors specific to fixed samples. Reports indicate that the detection of SVs, like SNVs, is strongly influenced by biological factors such as intratumoral heterogeneity, tumor purity, and ploidy, as well as sequencing depth; it is possible that these factors also exerted some influence in this study [31, 32]. Although WGS has the advantage of surveying the whole genome, including noncoding regions, and may identify structural variants missed by panel testing or whole-exome sequencing, the present study indicates that WGS remains challenging for accurate SV detection using cytology samples. In contrast, RNA-seq showed relatively consistent performance for fusion detection. Taken together, these findings suggest that rearrangement analysis in cytology samples, as in FF samples, should be complemented by RNA-seq.

In this pilot study, cytology samples demonstrated potential utility for WGS, particularly for SNV/indel detection, where performance was comparable to that of FF samples. Although limited applicability was also suggested for CNA analysis, their utility for SV detection remained less conclusive. This study was limited to a xenograft model and a single clinical case. Moreover, because the clinical case analyzed in this study was MSI-high/TMB-high and harbored a large number of SNVs, the observed concordance may not fully represent performance in tumors with low mutational burden or low tumor purity. Therefore, translation

into broader clinical practice will require validation in larger cohorts, together with more detailed investigations focused specifically on SVs.

Conclusion

Overall, this pilot study supports the technical feasibility of using cytology samples for WGS-based short-variant analysis, with sequencing quality and SNV/indel profiles largely comparable to those of matched FF samples. In contrast, CNA and especially SV detection showed greater variability, indicating that further optimization and validation in larger clinical cohorts are required. These findings provide a basis for future studies evaluating cytology samples as alternative materials for genome-scale cancer analysis.

List of abbreviations

ATCC:	American Type Culture Collection
Bp:	Base pair
BQSR:	Base quality score recalibration
CGP:	Comprehensive genomic profiling
CNA:	Copy number alteration
CP:	Cell pellet
DIN:	DNA integrity number
DNA:	Deoxyribonucleic acid
DNase:	Deoxyribonuclease
FF:	Fresh-frozen
FFPE:	Formalin-fixed paraffin-embedded
GATK:	Genome Analysis Toolkit
HER2:	Human epidermal growth factor receptor 2
Indel:	Insertion and deletion
Mb:	Megabase
MSI:	Microsatellite instability
NGS:	Next-generation sequencing
PCR:	Polymerase chain reaction
RNA:	Ribonucleic acid
RNA-seq:	RNA sequencing

SBS:	Single-base substitution
SD:	Standard deviation
SNV:	Single-nucleotide variant
SV:	Structural variant
TMB:	Tumor mutation burden
VAF:	Variant allele frequency
WGS:	Whole-genome sequencing

Declarations

Ethics approval and consent to participate

Studies involving human samples were approved by the Institutional Review Board of Hokkaido University Hospital (approval no. 022-0154) and were conducted in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants. All animal experiments were performed in accordance with the Japanese Act on Welfare and Management of Animals, the relevant Japanese guidelines for the care and use of laboratory animals, and the institutional guidelines of Hokkaido University. The animal use protocol was reviewed and approved by the Institutional Animal Care and Use Committee of Hokkaido University.

Consent for publication

Not applicable

Availability of data and materials

The sequencing data generated in this study have been deposited in the Japanese Genotype-phenotype Archive (JGA) under accession number JGAS000956 (https://ddbj.nig.ac.jp/search/entry/jga-study/JGAS000956/). Access to these data is controlled to protect participant privacy and requires a data access application and approval in accordance with JGA policies.

Competing interests

Kanako C. Hatanaka has received grants or contracts from Eli Lilly, NEC Corporation, Sakura Finetek Japan Co., Ltd., and Chugai Pharmaceutical Co., Ltd. Ichiro Kinoshita has received grants or contracts from Konica Minolta REALM, Eisai, Daiichi Sankyo, and Bayer. He has also received payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing, or educational events from MSD Pharmaceutical, Bayer, Takeda Pharmaceutical, Konica Minolta REALM, Chugai Pharmaceutical Co., Ltd., Novartis Pharma, Guardant Health Japan, and Pfizer Japan. Yutaka Hatanaka has received grants or contracts from Eli Lilly, NEC Corporation, CURED, Konica Minolta REALM, Daiichi Sankyo, BostonGene, Chugai Pharmaceutical Co., Ltd., and biomy Inc. He has also received payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing, or educational events from AstraZeneca and Daiichi Sankyo. All other authors declare no conflicts of interest.

Funding Information

This work was supported by the Japan Agency for Medical Research and Development (AMED) through the research fund (#21ck0106696h0001) and was supported in part by the Medical Research Support Program (#JP256f0137006).

Author Contributions

Conceptualization, Y.S. and Y.H.; Methodology, Y.H.; Investigation, Y.S., K.C.H., K.K., A.O., A.N., H.Y., H.A., T.K., and S.I.; Data Curation, Y.S., K.C.H., K.K., A.O. and S.I.; Visualization, Y.S.; Writing – Original Draft Preparation, Y.S.; Writing – Review & Editing, Y.S., K.C.H., K.K., I.K., T.K., H.W., S.I. and Y.H.; Supervision, K.C.H., K.K., I.K., H.W., S.I. and Y.H.; Project Administration, S.I. and Y.H.; Funding Acquisition, K.C.H. and Y.H. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

We are grateful to the patients and their families, and to the investigators and study staff who contributed to this study. We also acknowledge the support of AMED.

References

1. Zhang F, Lupski JR: **Non-coding genetic variants in human disease**. *Hum Mol Genet* 2015, **24**(R1):R102-110.

2. Bellone S, Buza N, Choi J, Zammataro L, Gay L, Elvin J, Rimm DL, Liu Y, Ratner ES, Schwartz PE *et al*: **Exceptional Response to Pembrolizumab in a Metastatic, Chemotherapy/Radiation-Resistant Ovarian Cancer Patient Harboring a PD-L1-Genetic Rearrangement**. *Clin Cancer Res* 2018, **24**(14):3282-3291.

3. Hornshoj H, Nielsen MM, Sinnott-Armstrong NA, Switnicki MP, Juul M, Madsen T, Sallari R, Kellis M, Orntoft T, Hobolth A *et al*: **Pan-cancer screen for mutations in non-coding elements with conservation and cancer specificity reveals correlations with expression and survival**. *NPJ Genom Med* 2018, **3**:1.

4. Jung H, Lee KS, Choi JK: **Comprehensive characterisation of intronic mis-splicing mutations in human cancers**. *Oncogene* 2021, **40**(7):1347-1361.

5. Hostein I, Stock N, Soubeyran I, Marty M, De Mascarel I, Bui M, Geneste G, Petersen MC, Coindre JM, Macgrogan G: **Nucleic acid quality preservation by an alcohol-based fixative: comparison with frozen tumors in a routine pathology setting**. *Diagn Mol Pathol* 2011, **20**(1):52-62.

6. Su JM, Perlaky L, Li XN, Leung HC, Antalffy B, Armstrong D, Lau CC: **Comparison of ethanol versus formalin fixation on preservation of histology and RNA in laser capture microdissected brain tissues**. *Brain Pathol* 2004, **14**(2):175-182.

7. Niimi H, Onishi T, Nomura S, Kumazaki A, Masaki Y, Odashima H, Osawa Y, Hattori M: **Assessment of Nucleic Acid Quality in Unstained Cytology Specimens for Cancer Genomic Testing**. *Acta Cytol* 2025:1-8.

8. Sholl LM, Hwang DH: **Expanding the utility of cytology preparations in cancer biomarker testing**. *Cancer Cytopathol* 2021, **129**(5):337-340.

9. Kawahara A, Taira T, Abe H, Watari K, Murakami Y, Fukumitsu C, Takase Y, Yamaguchi T, Azuma K, Akiba J *et al*: **Fixation effect of SurePath preservative fluids using epidermal growth factor receptor mutation-specific antibodies for immunocytochemistry**. *Cancer Cytopathol* 2014, **122**(2):145-152.

10. Fulmer CG, Park K, Dilcher T, Ho M, Mirabelli S, Alperstein S, Hissong EM, Pittman M, Siddiqui M, Heymann JJ *et al*: **Next-generation sequencing of residual cytologic fixative preserved DNA from pancreatic lesions: A pilot study**. *Cancer Cytopathol* 2020, **128**(11):840-851.

11. Ramani NS, Chen H, Broaddus RR, Lazar AJ, Luthra R, Medeiros LJ, Patel KP, Rashid A, Routbort MJ, Stewart J *et al*: **Utilization of cytology smears improves success rates of RNA-based next-generation sequencing gene fusion assays for clinically relevant predictive biomarkers**. *Cancer Cytopathol* 2021, **129**(5):374-382.

12. Diks J, Tang Z, Altan M, Anderson S, Chen H, Rashid A, Yang RK, Routbort MJ, Patel KP, Toruner GA *et al*: **Detection of clinically actionable gene fusions by next-generation sequencing-based RNA sequencing of non-small cell lung cancer cytology specimens: A single-center experience with comparison to fluorescence in situ hybridization**. *Cancer Cytopathol* 2024, **132**(1):41-49.

13. Akahane T, Yamaguchi T, Kato Y, Yokoyama S, Hamada T, Nishida Y, Higashi M, Nishihara H, Suzuki S, Ueno S *et al*: **Comprehensive validation of liquid-based cytology specimens for next-generation sequencing in cancer genome analysis**. *PLoS One* 2019, **14**(6):e0217724.

14. Fujita K, Arai R, Shoji S, Saito R, Nomura M, Hotta T, Asahina H, Kawakami M, Nakachi I, Hasegawa Y *et al*: **Detection of multiple druggable mutations of lung cancer from cytology specimens by MINtS: An advanced medicine A trial**. *Cancer Sci* 2023, **114**(8):3342-3351.

15. Kunimasa K, Tamiya M, Inoue T, Kawamura T, Miyazaki A, Kojitani Y, Honma K, Nishino K: **Clinical application of the Lung Cancer Compact Panel(TM) using various types of cytological specimens in patients with lung cancer**. *Lung Cancer* 2024, **189**:107498.

16. Shinomiya Y, Ishida Y, Hatanaka KC, Yamamoto M, Nange A, Okumura A, Wada Y, Abiko M, Shimizu T, Watanabe R *et al*: **Utility of a Novel High-Sensitivity Multiplex Companion Diagnostic Test Using Formalin-Fixed Paraffin-Embedded Cell Block Materials of Non-small Cell Lung Cancer**. *Cancer Med* 2025, **14**(13):e71028.

17. Morikawa K, Kida H, Handa H, Inoue T, Saji H, Koike J, Nakamura S, Sato Y, Ueda Y, Suzuki F *et al*: **A Prospective Validation Study of Lung Cancer Gene Panel Testing Using Cytological Specimens**. *Cancers (Basel)* 2022, **14**(15).

18. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, Wendl MC, Ding L: **MSIsensor: microsatellite instability detection using paired tumor-normal sequence data**. *Bioinformatics* 2014, **30**(7):1015-1016.

19. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL *et al*: **Signatures of mutational processes in human cancer**. *Nature* 2013, **500**(7463):415-421.

20. Yu Y, Ouyang Y, Yao W: **shinyCircos: an R/Shiny application for interactive creation of Circos plot**. *Bioinformatics* 2018, **34**(7):1229-1231.

21. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R *et al*: **Integrated genomic characterization of endometrial carcinoma**. *Nature* 2013, **497**(7447):67-73.

22. Bonnet E, Moutet ML, Baulard C, Bacq-Daian D, Sandron F, Mesrob L, Fin B, Delepine M, Palomares MA, Jubin C *et al*: **Performance comparison of three DNA extraction kits on human whole-exome data from formalin-fixed paraffin-embedded normal and tumor samples**. *PLoS One* 2018, **13**(4):e0195471.

23. Oh E, Choi YL, Kwon MJ, Kim RN, Kim YJ, Song JY, Jung KS, Shin YK: **Comparison of Accuracy of Whole-Exome Sequencing with Formalin-Fixed Paraffin-Embedded and Fresh Frozen Tissue Samples**. *PLoS One* 2015, **10**(12):e0144162.

24. Steiert TA, Parra G, Gut M, Arnold N, Trotta JR, Tonda R, Moussy A, Gerber Z, Abuja PM, Zatloukal K *et al*: **A critical spotlight on the paradigms of FFPE-DNA sequencing**. *Nucleic Acids Res* 2023, **51**(14):7143-7162.

25. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A *et al*: **Intratumor heterogeneity and branched evolution revealed by multiregion sequencing**. *N Engl J Med* 2012, **366**(10):883-892.

26. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, Shafi S, Johnson DH, Mitter R, Rosenthal R *et al*: **Tracking the Evolution of Non-Small-Cell Lung Cancer**. *N Engl J Med* 2017, **376**(22):2109-2121.

27. Lopez-Reig R, Fernandez-Serra A, Romero I, Zorrero C, Illueca C, Garcia-Casado Z, Poveda A, Lopez-Guerrero JA: **Prognostic classification of endometrial cancer using a molecular approach based on a twelve-gene NGS panel**. *Sci Rep* 2019, **9**(1):18093.

28. Robbe P, Popitsch N, Knight SJL, Antoniou P, Becq J, He M, Kanapin A, Samsonova A, Vavoulis DV, Ross MT *et al*: **Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project**. *Genet Med* 2018, **20**(10):1196-1205.

29. Wei L, Dugas M, Sandmann S: **SimFFPE and FilterFFPE: improving structural variant calling in FFPE samples**. *Gigascience* 2021, **10**(9).

30. Haile S, Corbett RD, Bilobram S, Bye MH, Kirk H, Pandoh P, Trinh E, MacLeod T, McDonald H, Bala M *et al*: **Sources of erroneous sequences and artifact chimeric reads in next generation sequencing of genomic DNA from formalin-fixed paraffin-embedded samples**. *Nucleic Acids Res* 2019, **47**(2):e12.

31. Griffith M, Miller CA, Griffith OL, Krysiak K, Skidmore ZL, Ramu A, Walker JR, Dang HX, Trani L, Larson DE *et al*: **Optimizing cancer genome sequencing and analysis**. *Cell Syst* 2015, **1**(3):210-223.

32. Talsania K, Shen TW, Chen X, Jaeger E, Li Z, Chen Z, Chen W, Tran B, Kusko R, Wang L *et al*: **Structural variant analysis of a cancer reference cell line sample using multiple sequencing technologies**. *Genome Biol* 2022, **23**(1):255.

Figure legends

Figure 1. Comparison of FF and cytology samples in WGS using a xenograft model.

(A) The variant classifications and numbers of variants detected in CP, FF, and cytology samples are shown in a bar plot. (B) Venn diagram showing concordance of detected short variants across CP, FF, and cytology. The variants indicated by the boxed label are known variants in the BT-474 cell line. (C) Concordance of detected variants after filtering for variants with VAF ≧ 0.1 across CP, FF, and cytology samples. (D) Bar plot showing recall, precision, F1 score, and concordance rate for FF and cytology samples when referencing CP. Scores after filtering for VAF ≧ 0.1 are shown in (E). (F) Venn diagram showing concordance of CNAs detected from CP, FF, and cytology samples. (G) Recall, precision, F1 score, and concordance rate for FF and cytology samples against CP. (H) Copy ratio (log2) of ERBB2 amplification detected in three sample types as a scatter plot. (I) Upset plot showing rearrangements detected in three sample types. (J)(K) Concordance metrics are shown at the gene symbol level in (J) and at the exact breakpoint level in (K).

Figure 2. Concordance of SNV profiles between cytology and FF samples in a clinical case.

(A) Bar plot showing the types and numbers of SNVs detected in FF and cytology samples. (B) Venn diagram showing overlap of variants between FF and cytology samples. (C) Boxplots comparing VAFs across detection categories. Group comparisons were performed using the Kruskal–Wallis test followed by Dunn's post hoc test. ***P < 0.001. (D) Venn diagram showing overlap of variants after filtering for VAF ≧ 0.1. (E) Mutational signatures based on single base substitutions (SBS) in the FF sample. (F) SBS signatures in the cytology sample.

Figure 3. Agreement of pathogenic or likely pathogenic SNVs in clinical cytology and FF samples.

(A) Concordance of detected variants between FF and cytology samples was visualized using a Venn diagram. (B) Venn diagram showing the concordance of SNVs between FF and cytology samples after filtering for variants with VAF ≧ 0.1. (C) All pathogenic or likely pathogenic SNVs detected in FF and cytology samples are presented as a tile map, with color intensity indicating the VAF of each mutation.

Figure 4. Comparison with CGP and consistency of SNVs across FF and cytology samples.

(A) Concordance of SNVs detected by CGP between FF and cytology samples, illustrated using a Venn diagram. (B) Recall rates of WGS for SNVs previously identified by CGP, shown as a bar plot for FF and cytology samples. (C) Venn diagram showing the overlap of variants detected by WGS in FF and cytology samples and also detected by CGP in each sample type.

Figure 5. Genome-wide variant profiles and fusion detection by WGS and RNA-seq in FF and cytology samples.

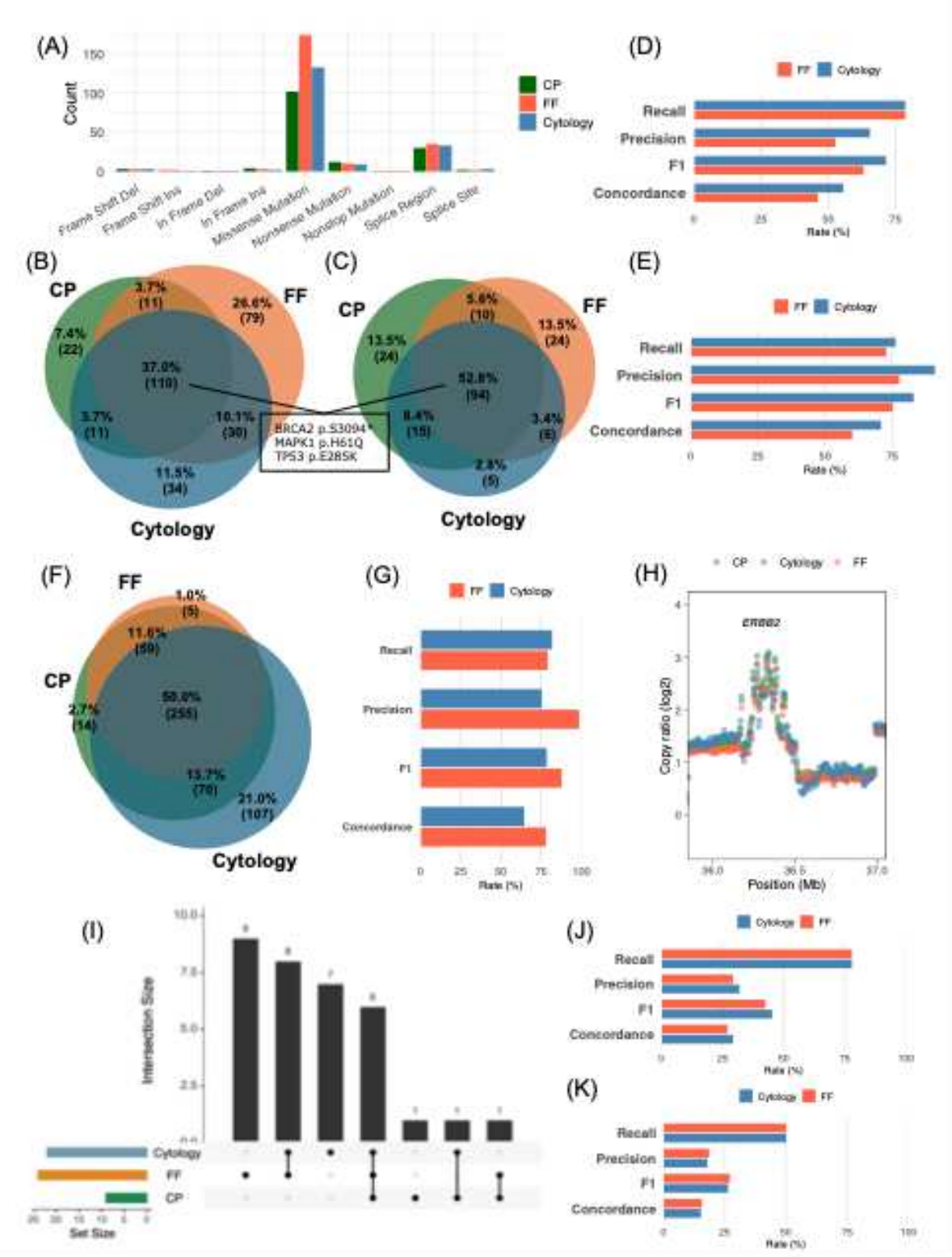
(A, B) Scatter plots of genome-wide log2 copy ratio distributions in FF (A) and cytology (B) samples. (C) Bar plot of rearrangements detected by WGS in FF and cytology samples. (D) Venn diagram showing the overlap between rearrangements detected by WGS and fusion transcripts identified by RNA-seq, irrespective of sample type. (E) Bar plot summarizing fusion concordance between FF and cytology samples by WGS and

RNA-seq. (F, G) Circos plots showing the distribution of single nucleotide variants (SNVs) and structural variants (SVs) in FF (F) and cytology (G) samples. SNV allele frequencies are shown as scatter plots and rearrangements as chromosomal links. Pathogenic or likely pathogenic SNVs are labeled with gene symbols.

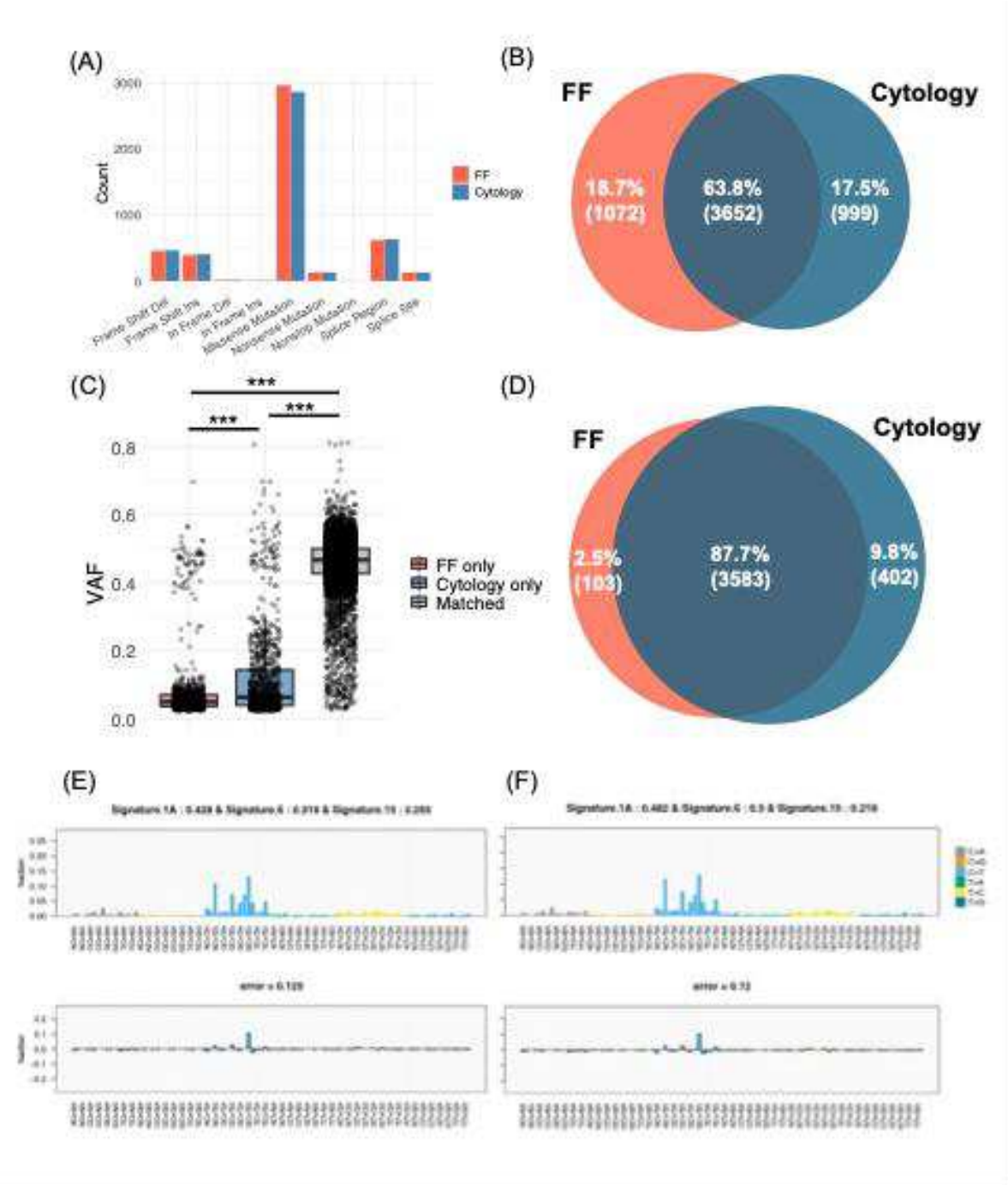
Figure 6. Consistency of RNA expression levels between FF and cytology samples in a clinical case.

Whole-transcriptome expression levels (log2(CPM + 1)) of FF and cytology samples were scatter-plotted in (A) to demonstrate their correlation. Genes reported to be effective for endometrial carcinoma subtyping were extracted, and their expression levels in FF and cytology samples are shown as a heatmap (B).

537 **Figure. 1**


(A)
Count
CP
FF
Cytology
Frame Shift Del
Frame Shift Ins
In Frame Del
In Frame Ins
Missense Mutation
Nonsense Mutation
Nonstop Mutation
Splice Region
Splice Site
(B)
CP
FF
Cytology
7.4% (22)
3.7% (11)
26.6% (79)
37.0% (110)
3.7% (11)
10.1% (30)
11.5% (34)
(C)
CP
FF
Cytology
13.5% (24)
5.6% (10)
13.5% (24)
52.8% (94)
8.4% (15)
3.4% (6)
2.8% (5)
BRCA2 p.S3094*
MAPK1 p.H61Q
TP53 p.E285K
(D)
FF
Cytology
Recall
Precision
F1
Concordance
Rate (%)
(E)
FF
Cytology
Recall
Precision
F1
Concordance
Rate (%)
(F)
FF
CP
Cytology
1.0% (5)
11.6% (59)
50.0% (255)
2.7% (14)
13.7% (70)
21.0% (107)
(G)
FF
Cytology
Recall
Precision
F1
Concordance
Rate (%)
(H)
CP
Cytology
FF
ERBB2
Copy ratio (log2)
Position (Mb)
(I)
Intersection Size
Cytology
FF
CP
Set Size
(J)
Cytology
FF
Recall
Precision
F1
Concordance
Rate (%)
(K)
Cytology
FF
Recall
Precision
F1
Concordance
Rate (%)

538 **Figure. 2**


(A)
Count
3000
2000
1000
0
FF
Cytology
Frame Shift Del
Frame Shift Ins
In Frame Del
In Frame Ins
Missense Mutation
Nonsense Mutation
Nonsnp Mutation
Splice Region
Splice Site
(B)
FF
Cytology
16.7%
(1072)
63.8%
(3652)
17.5%
(999)
(C)
***
***
***
VAF
0.8
0.6
0.4
0.2
0.0
FF only
Cytology only
Matched
(D)
FF
Cytology
2.5%
(103)
87.7%
(3583)
9.8%
(402)
(E)
(F)

539 **Figure. 3**

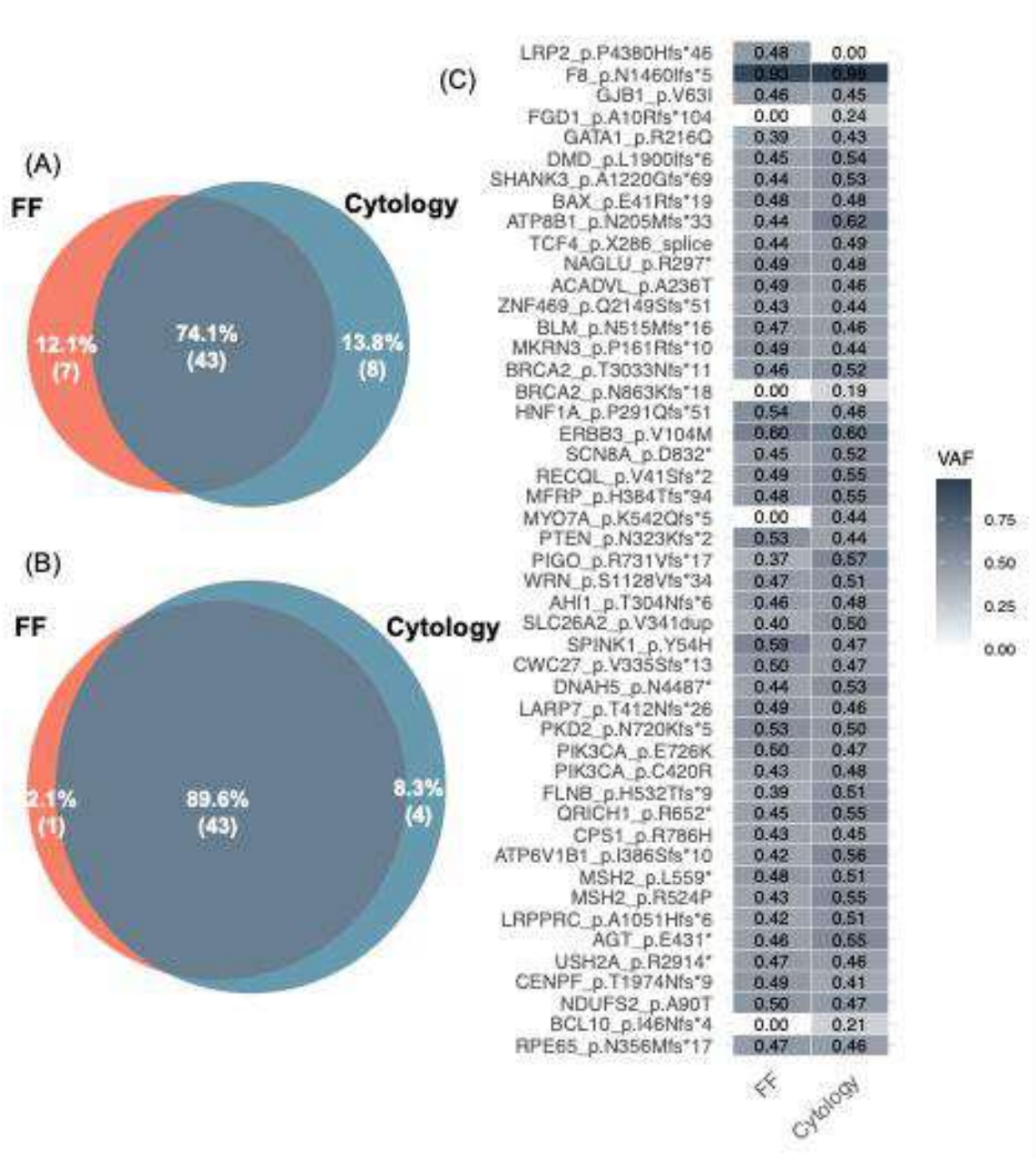
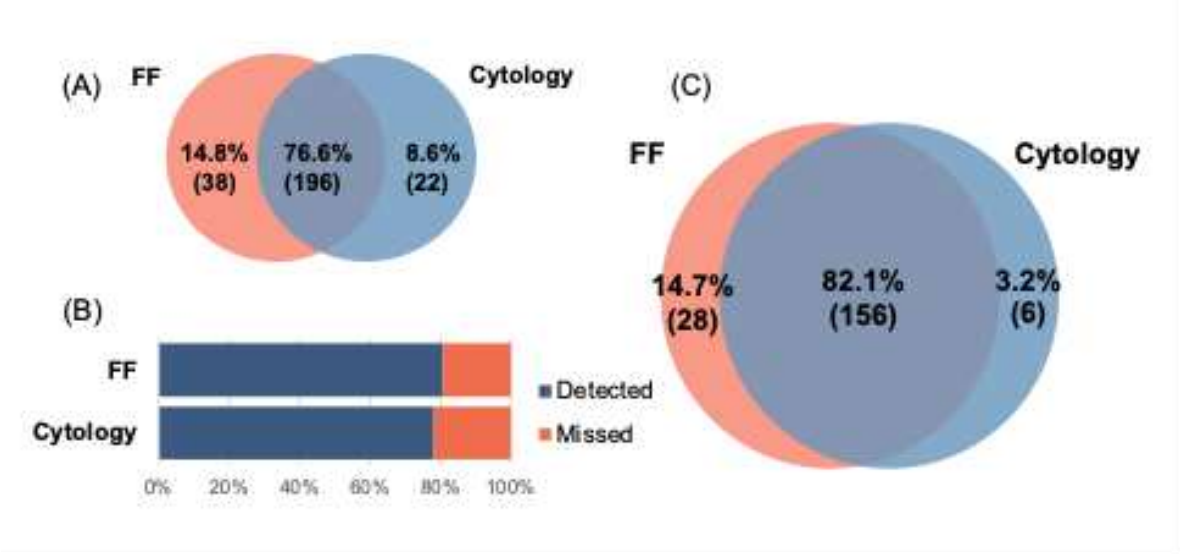
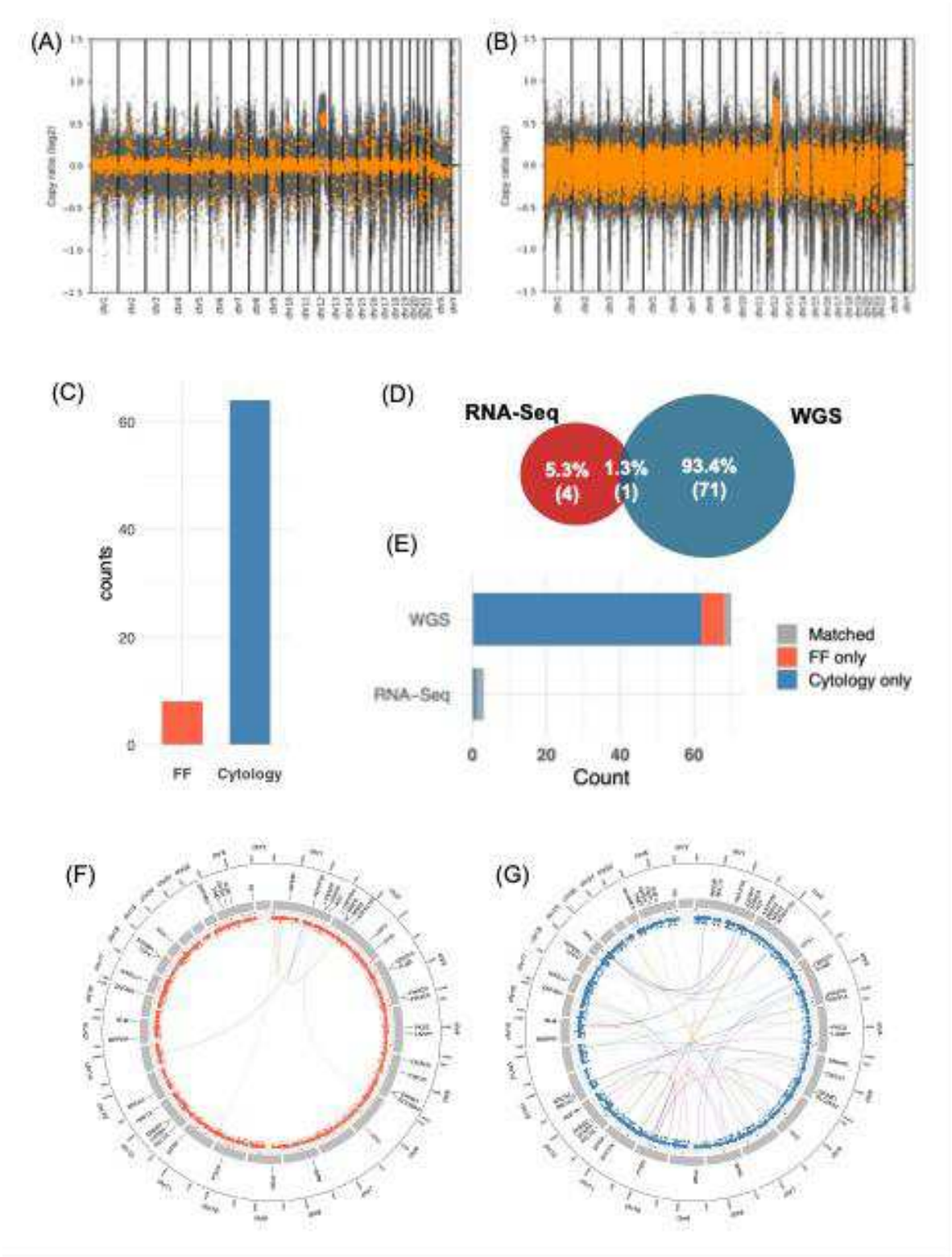

(A)
FF
Cytology
12.1%
(7)
74.1%
(43)
13.8%
(8)
(B)
FF
Cytology
2.1%
(1)
89.6%
(43)
8.3%
(4)
(C)
LRP2_p.P4380Hfs*46 0.48 0.00
F8_p.N1460Ifs*5 0.93 0.99
GJB1_p.V63I 0.46 0.45
FGD1_p.A10Rfs*104 0.00 0.24
GATA1_p.R216Q 0.39 0.43
DMD_p.L1900Ifs*6 0.45 0.54
SHANK3_p.A1220Gfs*69 0.44 0.53
BAX_p.E41Rfs*19 0.48 0.48
ATP8B1_p.N205Mfs*33 0.44 0.62
TCF4_p.X286_splice 0.44 0.49
NAGLU_p.R297* 0.49 0.48
ACADVL_p.A236T 0.49 0.46
ZNF469_p.Q2149Sfs*51 0.43 0.44
BLM_p.N515Mfs*16 0.47 0.46
MKRN3_p.P161Rfs*10 0.49 0.44
BRCA2_p.T3033Nfs*11 0.46 0.52
BRCA2_p.N863Kfs*18 0.00 0.19
HNF1A_p.P291Qfs*51 0.54 0.46
ERBB3_p.V104M 0.60 0.60
SCN8A_p.D832* 0.45 0.52
RECQL_p.V41Sfs*2 0.49 0.55
MFRP_p.H384Tfs*94 0.48 0.55
MYO7A_p.K542Qfs*5 0.00 0.44
PTEN_p.N323Kfs*2 0.53 0.44
PIGO_p.R731Vfs*17 0.37 0.57
WRN_p.S1128Vfs*34 0.47 0.51
AHI1_p.T304Nfs*6 0.46 0.48
SLC26A2_p.V341dup 0.40 0.50
SPINK1_p.Y54H 0.59 0.47
CWC27_p.V335Sfs*13 0.50 0.47
DNAH5_p.N4487* 0.44 0.53
LARP7_p.T412Nfs*26 0.49 0.46
PKD2_p.N720Kfs*5 0.53 0.50
PIK3CA_p.E726K 0.50 0.47
PIK3CA_p.C420R 0.43 0.48
FLNB_p.H532Tfs*9 0.39 0.51
QRICH1_p.R652* 0.45 0.55
CPS1_p.R786H 0.43 0.45
ATP6V1B1_p.I386Sfs*10 0.42 0.56
MSH2_p.L559* 0.48 0.51
MSH2_p.R524P 0.43 0.55
LRPPRC_p.A1051Hfs*6 0.42 0.51
AGT_p.E431* 0.46 0.55
USH2A_p.R2914* 0.47 0.46
CENPF_p.T1974Nfs*9 0.49 0.41
NDUFS2_p.A90T 0.50 0.47
BCL10_p.I46Nfs*4 0.00 0.21
RPE65_p.N356Mfs*17 0.47 0.46
FF
Cytology
VAF
0.75
0.50
0.25
0.00

Figure. 4


(A)
FF
Cytology
14.8%
(38)
76.6%
(196)
8.6%
(22)
(B)
FF
Cytology
Detected
Missed
0%
20%
40%
60%
80%
100%
(C)
FF
Cytology
14.7%
(28)
82.1%
(156)
3.2%
(6)

541 **Figure. 5**


(A)
(B)
(C)
counts
60
40
20
0
FF
Cytology
(D)
RNA-Seq
WGS
5.3%
(4)
1.3%
(1)
93.4%
(71)
(E)
WGS
RNA-Seq
0
20
40
60
Count
Matched
FF only
Cytology only
(F)
(G)

542 **Figure. 6**

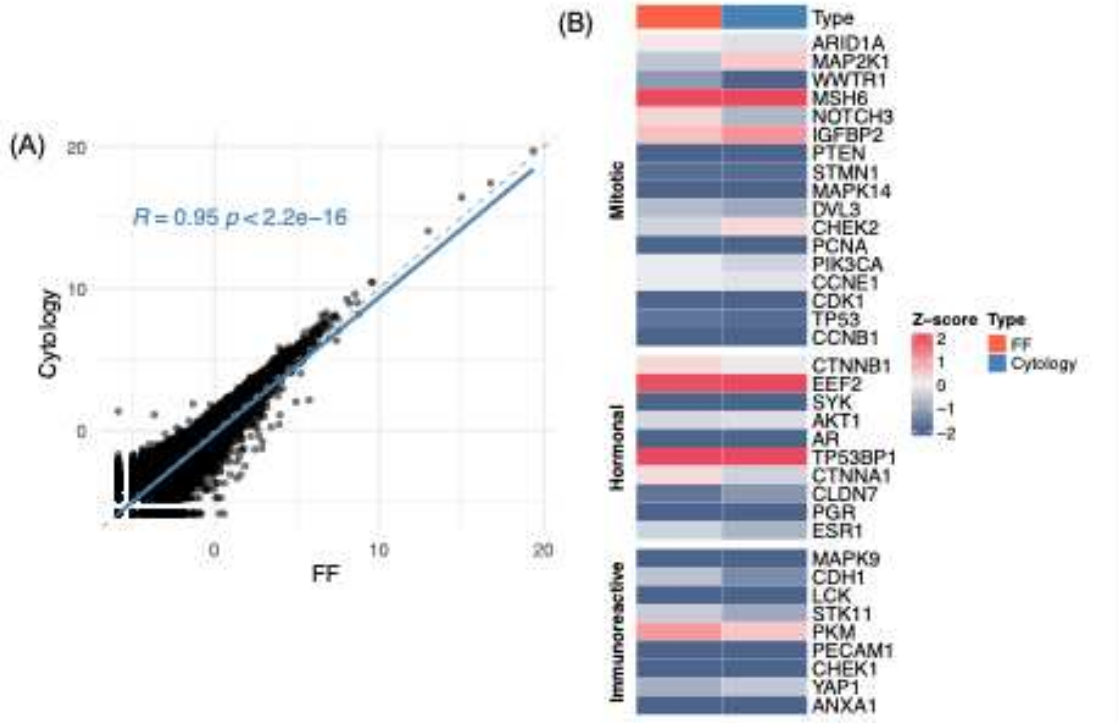

(A)
R = 0.95 p < 2.2e−16
Cytology
FF
20
10
0
0
10
20
(B)
Type
ARID1A
MAP2K1
WWTR1
MSH6
NOTCH3
IGFBP2
PTEN
STMN1
MAPK14
DVL3
CHEK2
PCNA
PIK3CA
CCNE1
CDK1
TP53
CCNB1
CTNNB1
EEF2
SYK
AKT1
AR
TP53BP1
CTNNA1
CLDN7
PGR
ESR1
MAPK9
CDH1
LCK
STK11
PKM
PECAM1
CHEK1
YAP1
ANXA1
Mitotic
Hormonal
Immunoreactive
Z-score
2
1
0
−1
−2
Type
FF
Cytology

Table 2. Representative whole-genome sequencing quality metrics in fresh frozen and cytology samples.

Model	Condition	Insert size (mean±SD)	Total Reads ($\times 10^9$)	Aligned (%)	Mismatch Rate (%)	Error Rate (%)	Paired Reads (%)	Improper Pairs (%)	Chimeras (%)	Softclip (%)	Strand Balance
Xenograft	FF	429.3 ± 193.7	3.43	99.94	0.57	0.50	99.94	9.74	12.08	8.62	0.50
Xenograft	Cytology	392.2 ± 175.7	3.28	99.97	0.54	0.49	99.97	5.73	7.97	5.42	0.50
Human	FF	430.2 ± 192.3	3.38	100	0.52	0.47	100	2.31	4.08	2.00	0.50
Human	Cytology	404.0 ± 181.8	3.48	100	0.49	0.45	100	2.08	4.04	2.17	0.50

Abbreviation: FF, fresh frozen

Note*: "Human" indicates an endometrial carcinoma patient; "Xenograft" refers to a mouse xenograft model established using the human breast cancer cell line BT-474.*