**A framework for Multimodal Biospecimen management in Oncologic Inpatient Cohorts for Multi-Omics based Biobank**

Zhao Sun1#, Xuejiao Xia1#, Rui Deng**1**, Jiahui Li**1**, Guanghui Li**1**, Mingliang Yuan**1**, Sanfei Peng**1**, Xiang-yang Yin**1**, Zhiwei Zhao**1**, Xinhua Dong**1**, Qi Liu**1**,Zhen Yang**1**\*, Yang Fu**1**\*

**1**Department of Gastrointestinal Surgery, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, 450052, China

**\*Correspondence:**

**Yang Fu**, E-mail: [fuyang@zzu.edu.cn](mailto:fuyang@zzu.edu.cn).

**Zhen Yang**, E-mail: 13526685689@163.com

**#** These authors contributed equally to this work

**Keywords:** Biobank, Omics, Clinical Samples, Quality Control, SOP

**Running title:** Bio-sample management for multi-omics biobank.

**Abstract**

**Background:** Omics-based technology has greatly promoted the precision medicine, which relays on the high-quality biobanks for meeting the technological requirements. However, it’s still lacked the universal guidance of end-to-end management of bio-sample collections during patients’ hospitalization, which resulting in the serious biases for further intergraded omics analysis.

**Aim of review:** This review aims to systematically conclude the bio-sample requirements of multi-omics technologies and propose a standardized framework for multimodal biospecimen management.

**Key scientific concepts of review:** This study proposes a framework including: (1). Multimodal biospecimen volume requirement for each omics type; (2). Standard protocols for multimodal biospecimen acquisition, transport and storage including the tissue, serum, saliva, urine and feces; (3). Ethical compliance; (4). Quality control (QC) and data management including metadata acquisition and other processes. This framework make contributed to the standard multimodal biospecimen management, which can accelerate the precision oncology based on omics-technology.

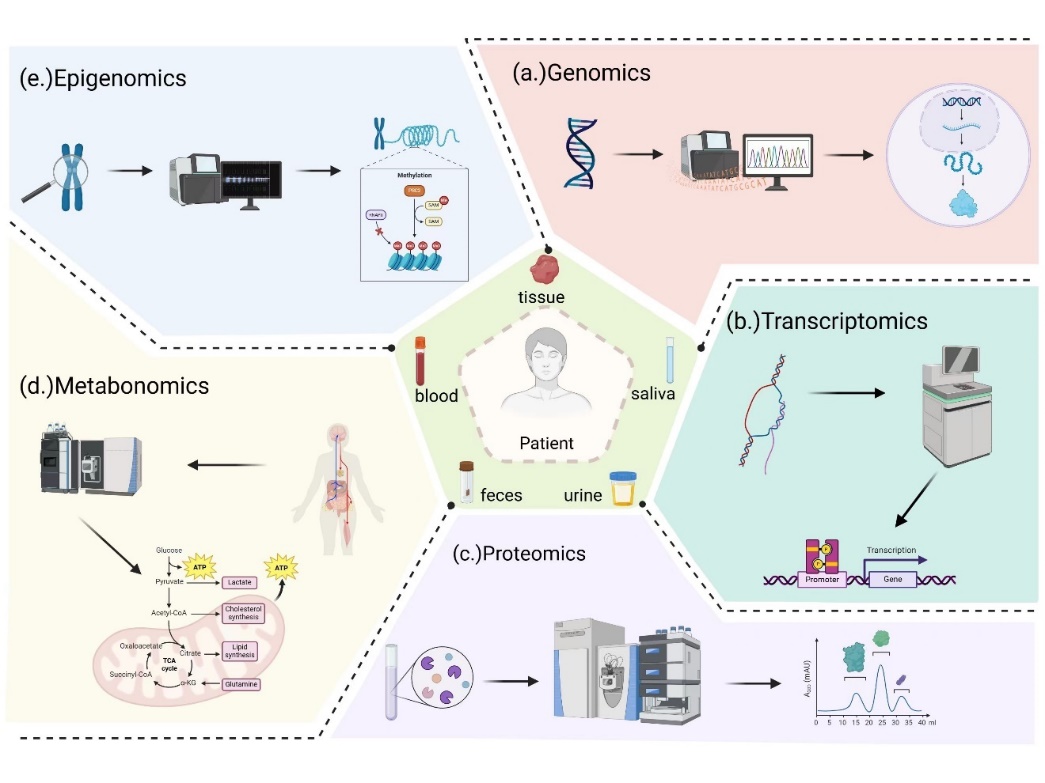
**Introduction**

Biobanks, encompassing programmatic biological materials, associated information, and legal considerations concerning data security and protection [1, 2],serve as cornerstones of biomedical research. Their evolution reflects a transition from disease-specific to multi-dimensional data integration. The 1990s witnessed the emergence of biobanks, primarily disease-oriented and functionally limited [3-5].As biological research advanced, the intricate relationship between diseases and genes became apparent. The Human Genome Project (HGP), initiated to comprehensively explore human genetic information, was successfully completed in 2003. Subsequently, biobanks primarily supporting genomics research emerged[6-8],yet they faced limitations in data dimensionality[9].Breakthroughs in high-throughput technologies facilitated the systematic integration of omics technologies (e.g., genomics, proteomics, metabolomics), enabling comprehensive analysis of cellular components (e.g., RNA, DNA) in their entirety or near-completeness [10].This propelled the transformation of biobanks into multi-omics data hubs. Notable examples include the UK Biobank, which released genomic and phenotypic data from 500,000 individuals[11] , and The Cancer Genome Atlas (TCGA), a multi-omics cancer database[12] , providing crucial resources for precision medicine.

The advent of multi-omics technologies has revolutionized biomedical research by enabling systematic exploration of biological systems. These technologies generate massive datasets that capture genetic, molecular, and functional dynamics, providing unprecedented insights into disease mechanisms and personalized therapies. However, their success hinges on the availability of high-quality biospecimens, highlighting the critical role of biobanks. Currently, biobanks worldwide primarily operate as localized institutions, often established to serve specific research objectives, leading to a fragmented landscape. In this context, constructing multi-omics-oriented biobanks necessitates the integration of multi-dimensional biomarkers and the establishment of a cross-platform standardization framework to support innovation across the entire spectrum from basic research to clinical translation.

**Requirements for Biobank Samples in Multi-Omics Technologies**

In this section, we systemically concluded the bio-samples requirements of major omics types, including genomics, transcriptomics, proteomics, metabolomics and epigenomics.



**Genomics**

Genomics takes the whole genome sequence and DNA variations (single nucleotide variations, insertion deletions, structural variations, etc.) as the object of study, analyses the genome structure, function and evolution, and its carrier genome consists of exons encoding proteins (accounting for 1%) and introns regulating gene expression.[13] .Whole - Genome Sequencing (WGS) involves DNA extraction, library preparation (fragmentation, end - repair, adapter ligation), paired - end 150bp sequencing (e.g., via Illumina platforms), and bioinformatics analysis (FastQC, BWA, GATK) [14, 15]. Its paired - end strategy improves detection of repeats and SVs.Whole - Exome Sequencing (WES), which uses probes to capture exons (1% of the genome but holds 85% of disease - causing mutations), is cost - effective and less data - intensive. Its process resembles WGS, but with exon capture and enrichment in library preparation.[16].For microbial studies, amplicon sequencing amplifies marker genes (16S/18S/ITS) with universal primers for community composition analysis at the genus level and above. Metagenomic sequencing directly assesses all microbial DNA in environmental samples, revealing genetic and metabolic details down to the strain level. Both involve end - repair and adapter ligation in DNA processing[17, 18].

Table 1 Genomics sample size requirements

|  |  |  |  |
| --- | --- | --- | --- |
|  | WGS/WES | Amplicon sequencing | Macrogenome sequencing |
| Sample type | Sample volume | Sample volume | Sample volume |
| Surgical tissue | ≥10mg | / | ≥500mg |
| Biopsy tissue | ≥10mg | / | ≥500mg |
| FFPE tissue | 10-20 sheets, area 1cm2, thickness 5-10µm | / | / |
| Plasma | / | ≥3ml | ≥5ml |
| Whole blood | ≥2ml | / | / |
| Saliva | 1-2ml | / | / |
| Faeces | / | ≥2g | ≥500mg |

**Transcriptomics**

Transcriptomics, the study of all transcribed RNAs (mRNAs, non - coding RNAs, and small RNAs) in a biological organism, reveals gene expression and regulation mechanisms, reflecting a cell's specific state [19].mRNA - seq, a common technique, involves total RNA extraction, mRNA enrichment, double - stranded cDNA synthesis, end - repair, poly - A addition, adapter ligation, PCR enrichment, and library quality assessment, enabling high - throughput sequencing to uncover RNA expression dynamics and structural variations[20].miRNA - seq, focusing on 20 - to 25 - nt small RNAs, includes RNA extraction, adapter addition, reverse transcription, PCR amplification, and library quality assessment, using SE50 sequencing to analyze miRNA expression levels and evolutionary history[21]. LncRNA-seq resolves long non-coding RNA functions by removing rRNA and retaining strand information [22].circRNA - seq constructs libraries by removing ribosomal and linear RNA, employing PE50 sequencing to analyze circular non - coding RNA differential expression and miRNA binding sites, elucidating their functional mechanisms [23, 24]. Macrotranscriptome sequencing takes all RNA as the research object, and after constructing standard libraries, it is able to analyse the transcriptional situation and regulatory laws of the genome of a population in a specific environment. The amount of sequencing data usually reaches 5-10Gb, providing an important tool for the study of gene expression networks in ecosystems. [25].

Table 2 Transcriptomics sample size requirements

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | mRNA-seq | miRNA-seq | LncRNA-seq | Circ RNA-seq | Macrotranscriptome |
| Sample type | Sample volume | Sample volume | Sample volume | Sample volume | Sample volume |
| Surgical tissue | ≥25mg | ≥50mg | ≥25mg | ≥500mg | ≥500mg |
| Biopsy tissue | ≥25mg | ≥50mg | ≥25mg | ≥500mg | ≥500mg |
| FFPE tissue | 5-10 sheets, thickness 5-10μm, 25mm2 | 5-10 sheets, thickness 5-10μm, 25mm2 | / | / | / |
| Whole Blood | ≥500μl | ≥500μl | ≥500μl | / | / |
| Faeces | / | / | / | / | 1-3g |

**Proteomics**

Proteomics, which uses high - throughput methods to analyze protein types, quantities, modifications, and interactions in cells or tissues, works with genomics and transcriptomics to uncover dynamic gene expression processes, crucial for understanding physiological and disease mechanisms[26].Unlabeled techniques like DIA generate secondary spectra by spectral library construction and integrate first - level signals. Combined with LC - MS/MS and tools like Spectronaut, it improves detection reliability. 4D - DIA adds ion mobility separation, enhancing scanning speed, sensitivity, and quantification accuracy[27, 28].Label free uses LC - MS/MS for large - scale protein identification by analyzing proteolytic peptides. TMT enables simultaneous relative quantification of 8 or 18 samples using isotope tags, involving protein extraction, digestion, labeling, chromatographic fractionation, and LC - MS/MS detection [29, 30].PRM, after pre - experimental optimization, targets specific peptide segments for accurate quantification. [31].Metaproteomics combines label - free quantification, chromatographic fractionation, and mass spectrometry, using Unipept and Krona software to analyze microbial community protein composition and species abundance [32].

Post - translational modification research covers glycosylation, ubiquitination, methylation, acylation, etc. Glycosylation is divided into N - and O - linked types. Technical routes include lectin - enriched N - glycopeptides (label - free combined with heavy water deglycosylation), motif antibody - enriched O - GlcNAc (4D label - free analysis), and ZIC - HILIC - enriched intact N - glycopeptides for glycan composition analysis [33-35]. The first involves a label-free approach, where N-glycopeptides are enriched using lectin affinity chromatography following enzymatic digestion. Subsequent deglycosylation with heavy water induces a mass shift, enabling detection via mass spectrometry. O-GlcNAc glycosylation analysis involves enzymatic digestion, enrichment with motif antibodies, and 4D-label-free mass spectrometry, followed by qualitative and quantitative analysis of modification sites. The second strategy involves intact N-glycopeptide analysis, where peptides are enriched using ZIC-HILIC chromatography after enzymatic digestion. This is followed by database searching and data analysis, allowing for the identification of modification sites, glycan compositions, and large-scale qualitative and quantitative analysis.Ubiquitination uses motif antibodies (K - ε - GG) to enrich peptides, with 4D label - free and mass spectrometry detecting a 114.1Da mass shift for site - specific analysis[36]. Methylation relies on motif antibody enrichment and 4D label - free analysis of molecular weight changes to precisely locate arginine/lysine methylation sites [37].Acylation uses antibody enrichment followed by TMT labeling (for acetylation) or 4D label - free detection of molecular weight shifts to analyze acylated peptides and their functional mechanisms [38].

Table 3 Proteomics sample size requirements

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | DIA/4D DIA | | Label free/TMT | | PRM | Macroproteomics |
| Sample type | Sample size (not de-hyperabundance) | Sample size (de-abundance) | Sample size (not de-hyperabundance) | Sample size (de-abundance) | Sample volume | Sample volume |
| Surgical tissue | ≥30mg | / | ≥30mg | / | ≥100mg | / |
| Biopsy tissue | ≥30mg | / | ≥30mg | / | ≥100mg | / |
| Puncture Tissue | ≥2 stitches, size of a grain of rice visible to the naked eye | / | / | / | / | / |
| FFPE tissue | ≥20 pieces (5-10μm thick, 50mm2 size) | / | / | / | / | / |
| Serum | ≥20μl | ≥100μl | ≥20μl | ≥100μl | ≥50μl | / |
| Plasma | ≥20μl | ≥100μl | ≥20μl | ≥100μl | ≥50μl | / |
| Urine | ≥500μl | / | ≥500μl | / | ≥500μl | / |
| Saliva | ≥200μl | / | ≥200μl | / | ≥200μl | / |
| Faeces | / | / | / | / | / | ≥2g |

Table 4 Modified proteomics sample size requirements

|  |  |  |  |
| --- | --- | --- | --- |
|  | Phosphorylation | Glycosylation | Ubiquitination、Acylation、Methylation |
| Sample type | Sample volume | Sample volume | Sample volume |
| Surgical tissue | ≥50mg | ≥50mg | ≥75mg |
| Biopsy tissue | ≥50mg | ≥250mg | ≥75mg |
| FFPE tissue | ≥20 tablets (5-10μm, 50mm2 size) | —— | ≥20 tablets (5-10μm, 50mm2 size) |
| Serum | ≥20μl | ≥20μl | ≥50μl |
| Plasma | ≥20μl | ≥20μl | ≥50μl |
| Urine | ≥25ml | ≥30ml | ≥25ml |
| Saliva | ≥200μl | ≥500μl | ≥200μl |
| Faeces | / | / | / |

**Metabolomics**

Metabolomics, a burgeoning "omics" field following genomics, transcriptomics, and proteomics, serves as an extension of the latter two, offering a more direct and precise reflection of an organism's physiological state. It constitutes a crucial component of systems biology. By investigating alterations in metabolites, metabolomics elucidates the roles and impacts of proteins within metabolic pathways, thereby reflecting the outcomes of gene expression and regulation[39].

The primary technical workflow in metabolomics research begins with the separation of target compounds from the sample. Subsequently, appropriate methods are employed for metabolite extraction; for instance, global metabolomics often utilizes a combined extraction approach for hydrophilic metabolites and lipids. Prior to instrumental analysis, quality control (QC) measures are implemented to ensure the accuracy and reliability of the instrument and the detection process. Following this, gas chromatography (GC) or liquid chromatography (LC) separation techniques are used to finely separate the complex sample mixture, providing distinct sample components for subsequent mass spectrometry (MS) analysis. Mass spectrometry, with its high sensitivity and specificity, then precisely identifies and quantifies each component. The acquired data undergoes filtering to eliminate invalid or interfering data, ensuring data quality and reliability, followed by qualitative and quantitative analysis[40].

Table 5 Metabolomics sample size requirements

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Non-targeted metabolomics | Targeted metabolomics | Non-targeted lipid metabolomics | Lipid Metabolomics | Full-Spectrum Metabolome |
| Sample type | Sample volume | Sample volume | Sample volume | Sample volume | Sample volume |
| Surgical tissue | ≥50mg | ≥50mg | ≥60mg | ≥60mg | ≥60mg |
| Biopsy tissue | ≥50mg | ≥50mg | ≥60mg | ≥60mg | ≥60mg |
| serum | ≥100μl | ≥300μl | ≥100μl | ≥300μl | ≥100μl |
| Plasma | ≥100μl | ≥300μl | ≥100μl | ≥300μl | ≥100μl |
| Urine | ≥100μl | ≥1ml | ≥100μl | ≥500μl | ≥100μl |
| Saliva | ≥100μl | ≥200μl | ≥100μl | ≥1ml | ≥100μl |
| Faeces | ≥50mg | ≥100mg | / | ≥200mg | ≥50mg |

**Epigenomics**

Epigenomics explores heritable genome modifications like DNA methylation, chromatin accessibility, and RNA methylation that regulate gene expression without altering the DNA sequence, revealing the dual mechanisms of genetic information storage in both sequences and modifications [41].For DNA methylation, Whole - Genome Bisulfite Sequencing (WGBS) uses bisulfite to convert unmethylated cytosines to uracil, enabling single - base methylation mapping via high - throughput sequencing. The process includes DNA extraction, fragmentation, bisulfite conversion, and library preparation[42, 43].Reduced Representation Bisulfite Sequencing (RRBS) enriches CCGG site - containing fragments via MspI digestion, covering 12% of the genome's methylation sites (over 80% of promoter regions) after bisulfite treatment, ideal for clinical CpG island methylation studies[44].Bisulfite Amplicon Sequencing (BSAS) allows targeted PCR amplification of specific genes or CpG islands, enabling absolute methylation quantification with as little as 1ng of sample [45].For chromatin accessibility and RNA methylation, ATAC - seq uses Tn5 transposase to cleave open chromatin regions, creating sequencing libraries to map nuclear chromatin accessibility, which can be combined with RNA - seq to study gene regulatory networks[46].MeRIP - seq enriches m6A - methylated RNA fragments (around 100nt) using m6A antibodies to profile m6A modification distribution across the transcriptome[47].ChIP - seq enriches DNA regions with histone modifications or transcription factor binding via chromatin immunoprecipitation, detecting genome - wide interaction sites[48, 49].

Table 6 Epigenomics sample size requirements

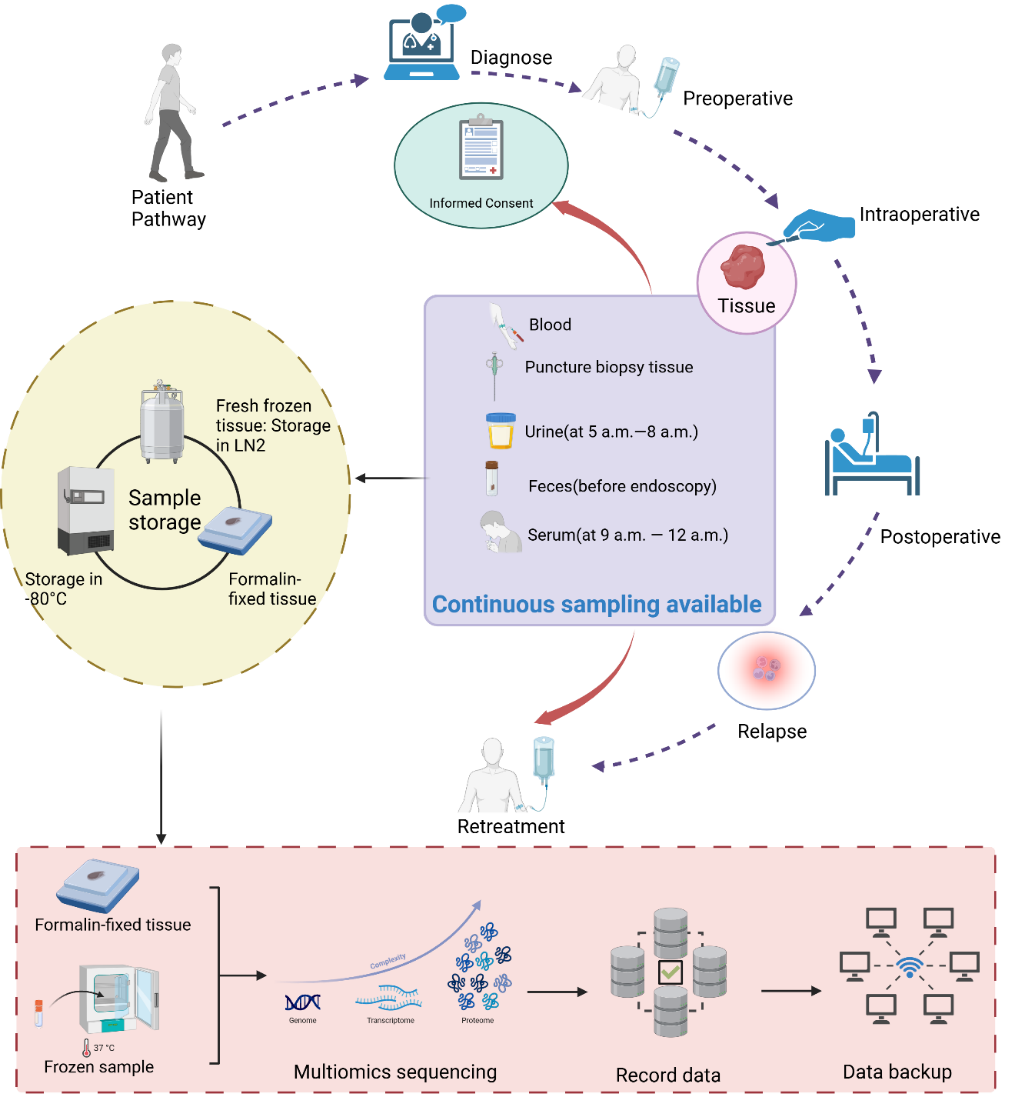
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | WGBS | PRBS | BSAS | ATAC-seq | MeRIP-seq | ChIP-seq |
| Sample type | Sample volume | Sample volume | Sample volume | Sample volume | Sample volume | Sample volume |
| Surgical tissue | ≥50mg | ≥50mg | ≥50mg | ≥50mg | ≥50mg | ≥50mg |
| Biopsy tissue | ≥50mg | ≥50mg | ≥50mg | ≥50mg | ≥50mg | ≥50mg |
| FFPE tissue | 5-10 sheets, thickness 4-5 μm, 0.6mm³ | ≥10 pieces, thickness 5μm-10μm, area 200mm2 | ≥10 pieces, thickness 5μm-10μm, area 200mm2 | / | / | / |
| Serum | / | ≥4ml | ≥4ml | / | / | / |
| Plasma | / | ≥4ml | ≥4ml | / | / | / |
| Whole Blood | ≥1ml | ≥2ml | ≥2ml | / | / | / |
| Faeces | ≥1ml | / | / | / | / | / |

**Quality Management in Biobanking**

**Multimodal Biospecimen Collections and Storage**

The pre-processing of biological samples is a key component in the construction of high-quality biospecimen repositories, and its standardised implementation faces a number of challenges. The first problem is that potential bias in sample collection may lead to systematic errors, especially when multimodal samples are collected from the same patient, which may aggravate the physical and mental burden of the patient and lead to a decrease in the sample acquisition rate. In addition, the time-dependent degradation of endogenous biomolecules in isolated samples severely restricts the reliability of downstream analyses, and it requires some quality control methods to ensure the sample quality. To this end, this study constructs a clinical diagnosis and treatment-sample bank synergistic framework to break through the bottleneck of quality control through systematic strategies.

The standardised procedure based on the clinical pathway is as follows: after diagnosis and admission to the hospital, patients should sign the standardised ‘Informed Consent for Biological Sample Donation’ voluntarily on the premise of being fully informed. The sample collection time window must strictly follow the diagnosis and treatment sequence. In the First Affiliated Hospital of Zhengzhou University, for example, after the patient is diagnosed and before any therapeutic interventions are carried out, the hospital will collect the first biological sample to be sent to the Laboratory Department for testing, and the biospecimen bank can collect the samples required by the bank synchronously, which effectively avoids the trouble caused by the second collection to the patient. Throughout the treatment cycle, the biospecimen repository needs to work closely with the clinical collection to collect dynamic samples, covering the entire spectrum of biospecimens from the initial diagnosis to the end of treatment. The collection of various types of samples needs to be prepared and pre-treated according to the type of downstream experiments. After obtaining clinical samples, it is recommended that the samples be de-identified and labelled in order to protect patients' personal information. In addition, it is recommended to collect 2-3 additional samples for distribution according to the sample volume requirements of the above mentioned groups. Regarding sample transport, it needs to be carried out under temperature monitoring according to environmental conditions and experimental requirements. In addition, all biological samples need to undergo quality testing, including tissue sample quality, DNA quality, RNA quality and protein quality, before depositing into the biosample bank. This well-thought-out sample management programme provides a reliable guarantee for the research and maximizes the effective use of the samples.



**Tissue**

To minimize tissue warm ischemia time, tissue samples should be collected and processed within 30 minutes of gross specimen excision [50, 51] . Processing involves rinsing the excised tissue samples with sterile, nuclease-free saline or PBS to remove surface blood. The tissue should then be dissected into smaller pieces and aliquoted for storage to minimize freeze-thaw cycles. It is important to note that these tissue samples may be derived from "residual tissue" from surgical specimens, or from biopsies obtained via minor surgery, endoscopy, or ultrasound guidance. The order of collection should be as follows: normal tissue (N: >5 cm from the tumor margin), adjacent tissue (P: 2-3 cm from the tumor margin), and tumor tissue (T)[52-55].

For fresh frozen samples, provide an adequate amount of tissue while ensuring sufficient material for pathological examination. Processed tissue samples should be stored in 2 mL RNase-free screw-cap cryovials, rapidly frozen in liquid nitrogen for 0.5 h, and then transferred to -80°C or liquid nitrogen for long-term storage. If immediate processing is not possible, flash-freezing should be performed within approximately 3 hours of storage at 4°C. Certain omics applications require the use of preservation solutions. For example, when using RNAlater, add a volume five times that of the tissue, store at 4°C overnight, and then transfer to -80°C for applications such as transcriptomic sequencing. For preservation with Miltenyi tissue storage solution, add a volume five times that of the tissue and transport to the laboratory within 24 hours at 4°C for applications such as single-cell sequencing. For paraffin-embedded samples, fixation in 10% buffered formalin (preferably neutral buffered formalin) should occur within approximately 3 hours of surgical resection. The fixation time is typically 24-48 hours. Embedded tissues can be stored at room temperature or 4°C.

**Blood**

Blood samples should be processed within 2 hours whenever possible. If immediate processing is not feasible, samples should be stored at 4°C for less than 24 hours[56, 57], . For serum or plasma, processing (centrifugation, separation, aliquoting, and freezing) must occur within 4 hours of collection. It is recommended to store processed blood samples in multiple aliquots to avoid quality degradation due to repeated freeze-thaw cycles. After collection, anticoagulant blood should be gently inverted to mix, centrifuged at 2000g for 10 minutes, and the plasma aliquoted into cryovials, flash-frozen in liquid nitrogen for 0.5 hours, and stored at -80°C. After collection of procoagulant blood, allow it to clot at room temperature (15-25°C) for 0.5 hours [58] , centrifuge at 2000g for 10 minutes[59] , and collect the supernatant (serum). Aliquot the serum into cryovials, flash-freeze in liquid nitrogen for 0.5 hours, and store at -80°C long-term. Discard the final 500 μl to avoid platelet and cellular contamination[60] . Hemolysis, the presence of insoluble flocculent material, or turbidity in the blood sample are considered non-compliant, and re-sampling is recommended.

The choice of blood collection tube type should be based on the experimental objectives, downstream analysis requirements, and sample stability. In metabolomics, serum is considered the gold standard due to the absence of anticoagulant interference, particularly for amino acid metabolite detection. Heparin/EDTA plasma tubes can be used as alternatives, but it is important to note that heparin interferes with DNA studies and EDTA affects NMR signals[61-63] Proteomics studies require special consideration: serum can reflect changes in coagulation-related proteins, while EDTA plasma can inhibit protease activity, which is beneficial for the stability of sensitive proteins such as phosphatases. Although citrate tubes reduce platelet activation, the dilution effect caused by the liquid anticoagulant may affect the accuracy of protein concentrations [64, 65] . For transcriptomics, PAXgene tubes (for rapid RNA stabilization) or EDTA tubes (to inhibit RNase release) are preferred; the latter is more suitable for large-scale clinical studies but requires controlled sample exposure time [66]。

**Urine**

Prior to urine sample collection, donors should be instructed to thoroughly cleanse or disinfect the urethral opening. Following collection, a visual inspection of the urine sample should be performed to assess its suitability, specifically checking for contaminants such as fecal matter. Midstream urine samples should be collected from each donor in the morning before breakfast (5:00-8:00 AM) [67-69] . Samples should be centrifuged at 3000 rpm at 4°C for 10 minutes, snap-frozen in liquid nitrogen for 0.5 hours, and then stored at -80°C. Samples that cannot be processed in a timely manner should also be placed in 4 ℃ <48h, at room temperature (22 ℃ or so) <3h, to avoid ice packs and room temperature transport more than 8h, -20 ℃ <2 weeks, the number of freezing and thawing ≤ 3 times [70-73]

**Saliva**

Saliva samples should be collected between 9:00 AM and 12:00 PM [74-77] . Donors should rinse their mouths with purified water 2 hours before sampling to remove oral debris. Following this, donors should refrain from eating, drinking, smoking, or brushing their teeth. A sterile cotton swab should be placed under the donor's tongue, and the saliva-soaked swab should be removed every 1 minute, repeated four times. Subsequently, the saliva should be extracted from the cotton swab using a disposable syringe. The samples should then be centrifuged at 4°C at 2600g for 15 minutes [78] to remove any debris. The supernatant from each sample should be aliquoted into cryovials, snap-frozen in liquid nitrogen for 0.5 hours, and stored at -80°C until use [79, 80] . Samples should be frozen as soon as possible after sampling, and if freezing conditions cannot be met, they should be <1h at room temperature, <6h at 2-8°C, and ≤4 weeks at -20°C. [81-83]。

**Feces**

Feces were collected within 24 hours prior to endoscopy [84] Fresh specimens are essential. Before collection, participants were instructed to void their bladders and, whenever possible, to use a squat toilet. Fecal matter was collected into a clean, dry fecal collection tube and transferred to a sterile container within 30 minutes. Using a sterile spatula, the middle portion of the fecal sample was collected[85] . Following collection, samples were aliquoted and stored at 4°C. The collected samples were snap-frozen in liquid nitrogen for 0.5 hours and then stored at -80°C. Fecal samples should be stored temporarily at -20°C or 4°C for <8 weeks or <12h, respectively[86, 87] , and at room temperature for <4h [88]。

**Sample Aliquoting**

The collection and processing of multi-omics samples should strictly adhere to established sample volume guidelines. To ensure the smooth progression of the research, we recommend collecting an additional 2-3 aliquots of each sample as backups and storing them using an aliquotting strategy. This backup strategy effectively addresses potential unforeseen circumstances during sample storage, transportation, or usage, such as sample degradation due to temperature fluctuations or leakage caused by improper handling. In the event of such incidents, the intact backup samples can be promptly utilized to prevent delays in research progress caused by a single problematic sample, thereby ensuring the seamless execution of the entire research endeavor. This robust sample management protocol provides a reliable safeguard for the research, minimizing experimental risks.

**Sample transport**

Refrigerated samples (i.e., samples stored at 2-8°C) require the use of biological ice packs to maintain their storage temperature range during transport. For samples that require cryopreservation, dry ice or liquid nitrogen must be used for protected transport. It is recommended to store the samples at 2-8°C before centrifugation and dispensing, and no freezing is allowed [73] . When using dry ice for transport, the amount of dry ice should be reasonably arranged according to the length of the transport: 2 kg of dry ice can maintain the low-temperature environment for ≤6 h, 5 kg of dry ice can maintain ≤24 h, and 8 kg of dry ice can maintain ≤48 h,taking into account that the temperature is high in the summer, the amount of dry ice should be increased to 1.5 times of the original. In order to prevent the dry ice from crushing and damaging the samples, it is recommended to use powdered, rod-shaped dry ice, and brick-shaped dry ice can be smashed with a hammer for transport. In addition, to ensure the safety of the samples, all samples placed in dry ice should be placed in self-sealing bags on the outer packaging and clearly labelled for easy identification and tracking. It should be emphasised that, to ensure the safety and integrity of the samples, it is strictly prohibited to transport samples loose inside the dry ice.

**Quality Control (QC)**

**Sample sampling**

To ensure the reliability of sample quality, it is recommended that strict quality control measures be implemented throughout the entire process of sample collection to use. It is recommended to establish a QC system based on risk analysis with reference to ISO 20387 (ISO 20387:2018 ‘General Requirements for Biological Sample Banks’). Regular quality sampling of samples should be carried out, with small batch sampling (accounting for 2%-5% of the total volume of the sample bank) on a monthly basis and comprehensive sampling (accounting for 10%-15%) every six months; when the sample bank undergoes a major environmental change, introduces a new method or has abnormal results, targeted sampling will be carried out in a timely manner; and the scope of the test covers the quality of proteins, nucleic acids and tissue samples. For samples that fail to meet the standards, they should be destroyed according to ISO 35001 (ISO 35001:2019 ‘Requirements for biosafety management systems’), and destruction records should be completed. Through regular sampling and standardised management, the integrity of the samples and the reliability of the research data can be effectively guaranteed to ensure the smooth implementation of scientific research.

**Tissue sample quality testing**

Cellular activity is often measured by using Taipan blue staining, and the activity score should be >80% to meet the requirements. If the activity score is lower than this standard, it can be considered to use the appropriate technology to remove the dead cells to achieve the enrichment of cells, so as to improve the overall activity level of the sample. However, when the necrosis rate of the tissue is greater than 20%-50%, due to the high percentage of dead cells, it may cause serious interference with subsequent studies or assays, and such tissues can usually be directly discarded [89, 90]。

For the detection of tumour cell content, the percentage of tumour cells in tumour tissues can be effectively assessed using HE staining. When the proportion of tumour cells in the tumour tissue is >65% [91] the sample is judged to be qualified.

**DNA Quality Testing**

DNA quality testing is similar in principle and procedure to RNA quality testing. Firstly, the concentration and purity of DNA is determined by quantitative method based on UV absorbance using a spectrophotometer. The absorbance ratio of OD260/OD280 is the key index to measure the purity of DNA, and when the ratio reaches 1.8, it means that the DNA is in a relatively pure state[90, 92, 93] . Secondly, the integrity of DNA was detected by agarose gel electrophoresis. Intact genomic DNA will show tight bands with high molecular weight characteristics during electrophoresis, and there is basically no or very little low molecular weight trailing phenomenon. The presence of trailing or diffuse bands in the electrophoresis result is an indication that the DNA has been degraded or broken. This result provides a visual basis for determining the suitability of sample handling and storage conditions for downstream molecular experiments to ensure the quality of the DNA used in the experiments.

**RNA Quality Assay**

RNA quality assessment focuses on purity and integrity, both of which determine the usability of RNA samples in biological experiments. Spectrophotometry is the key method in purity detection. Its principle is based on the characteristic absorption of nucleic acids at 260 nm and proteins at 280 nm, and the purity is judged by calculating the OD260/280 ratio. Typically, a ratio between 1.8 and 2.2 indicates a high purity of the RNA sample. However, this method does not directly reflect the integrity of the RNA and makes it difficult to identify other impurities. The choice of buffer for the measurement is equally important; Tris or TE buffers are superior to water in providing a stable OD260/280 ratio. To compensate for the shortcomings of spectrophotometry, there is a trend towards a multi-wavelength OD synthesis judgement strategy covering wavelengths of 240 nm and 320 nm, which are used to capture potential contaminants and background absorption information, respectively. Analysing the changes of OD260/240 and OD260/320 ratios can identify other contaminants in the samples in depth, thus enhancing the accuracy of RNA sample quality assessment and providing reliable support for subsequent biological experiments [90, 93]。

For integrity assessment, the RNA integrity value (RIN) assay is the core tool, which utilises microfluidic capillary electrophoresis to obtain RIN values ranging from 1 (fully degraded) to 10 (fully intact). Typically, a RIN≥7 is considered the critical threshold, reflecting the overall quality of the RNA. RNA extracted from fresh tissues usually exhibits higher RIN values, whereas RIN values of formalin-fixed paraffin-embedded (FFPE) tissues tend to be concentrated in the range of 2-5, which affects the success rate of RNA-seq experiments. Therefore, factors such as sample storage time, environmental conditions and fixation time need to be considered in experimental planning to ensure the accuracy and reliability of experimental results [94-96].

**Protein Quality Assay**

BCA protein assay kit (using bovine serum albumin as the standard) is used to determine the protein concentration, by measuring the absorbance of the sample at a specific wavelength, and calculating the protein concentration according to the standard curve, in the assay can be used to eliminate interfering substances in the BCA protein assay by using the DOC and TCA to precipitate proteins, so as to improve the accuracy and reliability of the quantification of proteins[97] . Using the traditional method of Western blot to detect protein quality, if the specific bands of the target protein can be detected, and the bands are clear and without obvious trailing or dispersion, it indicates that the extracted proteins are not degraded, and the quality of the proteins is good [98]。

**Ethical security**

The long-term development of a biosample bank needs to strike a balance between scientific value and ethical compliance, and all samples and data collection should be carried out after obtaining the informed consent of the donor, in order to protect the rights and interests of both the donor and the collector and avoid legal disputes. The informed consent form should be signed with respect to the donor's wishes, religious beliefs and cultural level, ensuring that it is signed in a sober, fair, uncoerced and fully understood manner, and in a written form that specifies the rights and responsibilities of both parties, and is completed prior to the collection of the samples and permanently stored [99]. Despite the initial establishment of the informed consent framework, there are still pitfalls in its practical application, e.g., is there harm to the patient during the sampling process, the magnitude of that harm, and is there compensation? Does the patient have sufficient informed consent? Is it ethical to collect the sample? Biological sample safety or data security issues and privacy concerns.

According to the definition of harm in Munshi's study [100] , the sampling and testing operations of the biobank did not involve behaviours that could cause harm to patients, and all related costs were borne by the biobank, without infringing on the economic interests of patients, and secondly, after the implementation of the standardised synergistic operation plan between clinical treatment and the biobank, the psychological and physical burden of secondary sampling on patients was also avoided. Secondly, after the implementation of the standardised synergistic operation programme between clinical treatment and biobanks, the psychological and physical burden of secondary sampling on patients is avoided. For this kind of social welfare behaviour of donating samples to biobanks, although the final results of the study are contributed to the society, the patients cannot directly benefit from the relevant research, so the biobanks should still give appropriate compensation for this kind of public welfare behaviour, which can help to attract more people to participate in the donation[101] . When obtaining patients' informed consent for biobanking, differences in individual comprehension and age need to be taken into account to ensure effective communication, especially for vulnerable groups, biobanks should develop easy-to-understand protocols, such as using colloquial expressions and avoiding jargon. For patients who do not have the right to make autonomous decisions (e.g. children), their legal guardians may sign on their behalf[102-104]. The information kept in biospecimen repositories covers clinical profile data, biomaterial data and demographic data. When dealing with data or samples that contain personal privacy, donors' concerns about the exposure of private information may affect their willingness to participate. Therefore, before obtaining informed consent, it is necessary to explain in detail to the target individuals the de-identification protection measures of the biospecimen repository for their personal information in order to eliminate their concerns and enhance their willingness to participate [105, 106]. To ensure the safety of biological samples and related data, all activities of the biospecimen repository should be carried out under the strict review and supervision of the ethics committee. Especially when it comes to the use of biological samples, such as the export of samples collected in China to foreign countries, there must be justifiable reasons to ensure that all processes are legal and compliant, and the entire process is subject to the supervision of the ethics committee[107, 108].

**Integrated management of biospecimen repositories**

**Biobank Management**

Biobank management covers the whole life cycle of samples, and its core process includes standardised inbound and outbound management mechanism. In the admission process, the applicant (clinical department/project leader) should first submit the ‘Application for Biological Sample Admission’ and supporting materials such as research proposal, and then enter the approval process after the sample bank comprehensively evaluates the scientific value, ethical compliance and feasibility of the project; for the approved project, the applicant should further submit the standardised collection plan (including the details of collection, transportation and handling), and sign an ‘Agreement on Admission’ with the sample bank to clarify the implementation of informed consent and the process of consumables receipt; in the final process, both parties will jointly verify the samples and sign a standardised admission and discharge mechanism. The applicant shall further submit a standardised collection plan (including collection and transportation and handling rules) and sign the ‘Agreement on Accession’ with the sample bank to specify the implementation of informed consent and the process of consumables collection; in the final stage, both parties shall jointly verify the sample information and complete the ‘Registration Form for Accession’, which shall be digitally archived according to the coding rules of ‘Hospital Code + Organ Code + Water Number + Sample Type + Tube Section Number’, and the samples which do not pass the audit will be issued with the ‘Notice of Rejection’, which will state the specific reasons. With regard to the process of release, the applicant should submit the ‘Release Application Form’ with departmental approvals and ethical approvals 14 working days in advance, and after reviewing the significance of the research and compliance by the sample bank, the project that passes the audit needs to sign the ‘Release Agreement’ to define the relationship between rights and responsibilities, and then the staff locates and verifies the target samples and collaborates with the applicant to confirm the information in the inventory and simultaneously completes the signing of the ‘Release Registration Form’ and the handover of the entity. In order to guarantee the effectiveness of sample use tracking, the applicant is required to submit a sample application progress report 1/3/6 months after the release of samples, forming a closed-loop management to continuously optimise the allocation of resources in the database.

**Metadata Management**

Metadata management integrates patients' clinical data (case system collection and follow-up records) and histological data through a unified identifier system to ensure the uniqueness of the association between the biospecimen bank and the database. In the process of data accession, firstly, the data demand department submits the ‘Application for Data Accession to the Biospecimen Bank’ (the data associated with the already accessed samples can be exempted from application, but the application should be submitted for the case of signing the informed consent form without actually donating the samples); subsequently, the management personnel of the biospecimen bank conducts a comprehensive assessment of the scientific research value of the project, the reasonableness of the data, and the accessibility of the data, and issues the ‘Notice of Rejection of Data Accession’ with an explanation of the reasons if the data passes the audit. If the data do not pass the audit, the ‘Notice of Rejection of Data Entry’ will be issued and the reasons will be explained; the data that pass the audit will be entered into the database in strict accordance with the specifications, and the data will be sorted and filed and filled in the ‘Biological Sample Database Data Entry Registration Form’ after the completion of the audit, and the whole process will be carried out through the review of the legality of data and ethical compliance. The management of data release requires the user department to submit the ‘Application for Data Release from the Biological Sample Bank’ with ethical approvals, and in the process of audit, it is necessary to go through the double checking of the department of data source and the biological sample bank (including the matching of data and samples, and the assessment of the research value and feasibility); if the data do not pass the audit, the ‘Notice of Rejection of Data Release’ will be issued; after the audit is passed, the staff will organise the data according to the requirements and archive the data, and the data will be ultimately transferred to the database through on-site copying or encrypted transmission by the personnel designated by the applicant. After passing the audit, the staff will arrange the data according to the requirements and archive them, and finally the applicant will designate the personnel to obtain the data through on-site copying or encrypted transmission, and at the same time, fill in the Data Release Registration Form to complete the handover.

**Discussion**

Driven by the remarkable progress in multi - omics technologies, disease research has shifted from conventional clinical observation and empirical medicine to in - depth analysis at the genomic, molecular, and genetic levels. Against this backdrop, biobanks, serving as a key bridge between basic research and clinical application, have gained increasing strategic importance. However, the rapid development of omics technologies has led to an exponential increase in multi - dimensional data, posing great challenges to traditional biobanks in terms of storage, management, and data integration. To achieve efficient sharing of samples and in - depth analysis of omics data, traditional biobanks need to transform into digital ones. [109, 110]. This study makes a primary contribution by comprehensively and systematically clarifying the diverse needs of multi - omics technologies in biobanks, thus providing a theoretical basis for personalized services. It also thoroughly analyzes the entire sample - collection process from patient admission to discharge and devises scientific post - collection processing methods based on sample characteristics and research requirements, ensuring high - quality samples. These efforts lay a solid foundation for the scientific management and effective use of biobanks.

Prior to sequencing analysis, sample preparation is a critical step in multi-omics analysis to ensure data accuracy. However, this process may introduce systematic errors due to operational bias, which in turn affects the quality of downstream data and the reliability of biological conclusions[111] , therefore, customised processing protocols are required for the detection of different sample matrices (blood, tissue, urine, etc.) and molecule types (DNA, proteins, metabolites, etc.). Taking gastric cancer research as an example, the research team found, based on proteomics studies, that through in-depth analyses of plasma and tissue samples from gastric cancer patients and healthy populations, the differentially expressed protein markers in the two types of samples showed significant stratification characteristics, and in the blood circulatory system, circulating proteins such as apolipoproteins A-I, C-I, C-III, and prothrombin light chain A were identified as potential biomarkers for gastric cancer In the blood circulatory system, circulating proteins such as apolipoprotein A-I, C-I, C-III and thrombin light chain A were identified as potential gastric cancer biomarkers, whose abnormal expression was closely related to the remodelling of the tumour microenvironment; while in tumour tissues, tissue-specific proteins with a clear oncogenic mechanism such as HER2 and CLIC1 were identified, which not only participated in the regulation of tumour proliferation, but were also important biomarkers to guide the clinical targeting of therapies[112] , which confirmed the core principle of the ‘molecular mapping of samples is determined by sample matrices’. The synergistic application of histological technologies has further promoted the precision of disease diagnosis and treatment. For example, based on the integrated multi-omics research of TCGA and other international consortiums, gastric cancer has been systematically classified into four molecular subtypes, which not only correlate with clinical phenotypes (e.g., metastatic tendency, response to treatment), but also provide a molecular basis for the development of personalised treatment strategies. transformation (EMT) and metabolic reprogramming features, proteomic classification reveals prognostic heterogeneity, and DNA methylation and metabolome integration models elucidate the remodelling of the tumour microenvironment in terms of epigenetic regulatory network dimensions [113] [114] . This case highlights the basic principle of ‘sample as experiment’ in multi-omics research, and only through the establishment of a three-dimensional standardised system covering sample types, molecular species and detection techniques can the biological validity of multi-omics data be fundamentally ensured. It is worth noting that there are significant differences in the sample size requirements of different genomics technologies: for example, whole genome sequencing (WGS) requires only≥10 mg of tissue to complete genome-wide coverage, while mRNA-seq requires≥25 mg of tissue to meet the sensitivity requirements of transcript abundance detection. By systematically combing the technical characteristics of genomics, transcriptomics, proteomics, metabolomics and epigenomics, this study has established for the first time a standardised reference system for the starting dosage of biomass samples in multiple groups (Tables 1-6), which provides an important basis for the design and resource optimisation of cross-omics studies.

The effectiveness of a biospecimen repository largely depends on the systematic integration of clinical pathways with the sample collection process, which requires pre-planning of critical nodes of sample collection and the establishment of a quality control system covering the entire process through close collaboration of a multidisciplinary team (as shown in Figure 1). In this process, each critical node contains quality control points that require special attention.For example, surgically resected tissues need to be dispensed and pre-treated within 30 minutes of removal from the body[50, 51] and heat-sensitive samples (e.g., RNA) need to be immediately fixed with RNAlater to maintain the integrity of the transcriptome; blood samples need to be centrifuged and dispensed within 2 hours of collection, and blood collection tubes need to be selected according to the needs of downstream omics[61-66] , and the frozen samples need to be transported with sufficient dry ice to maintain -80℃ environment to avoid damage caused by temperature fluctuations. The transport of frozen samples requires sufficient dry ice to maintain the -80℃ environment to avoid damage caused by temperature fluctuations, so the amount of ice should be arranged according to the needs, and in order to ensure the quality of the samples, biological samples need to be processed through the quality control process before warehousing to ensure the accuracy of the subsequent experiments. It should be noted that all sample collection and personal information collection should be carried out after the patients sign an informed consent form, and the clinical data and biological sample information should be separated by de-identification technology to ensure compliance with privacy protection norms[99, 105, 106] . In addition, dynamic digital platforms (e.g., electronic medical records linked to sample banking systems) allow real-time tracking of specimen status and reduce human error. This process integration not only improves specimen collection efficiency, but also ensures data comparability through standardized operations.

By building a multi-omics quality control framework, optimizing the specimen collection process and improving ethical management, this study has significantly improved the resource utilization and data reliability of the biospecimen repository. However, future improvements are still needed in the following directions: 1. to promote the popularity of laboratory-level automation tools based on the existing QC standards[115, 116] ; 2. to establish cross-institutional data sharing protocols to promote the in-depth integration of multiorgan omics data; and 3. to adopt the dynamic consent model and the joint application of blockchain technology to further enhance the ethical compliance[117] . Through continuous technological innovation and interdisciplinary cooperation, biospecimen repositories are expected to become the core engine of precision medicine and translational research.

Reference

1. Kinkorová J: **Biobanks in the era of personalized medicine: objectives, challenges, and innovation: Overview**. *The EPMA journal* 2015, **7**(1):4.

2. Greely HT: **The uneasy ethical and legal underpinnings of large-scale genomic biobanks**. *Annual review of genomics and human genetics* 2007, **8**:343-364.

3. Nørgaard‐Pedersen B, Simonsen HJAp: **Biological specimen banks in neonatal screening**. 1999, **88**:106-109.

4. Shepherd J, Blauw GJ, Murphy MB, Cobbe SM, Bollen EL, Buckley BM, Ford I, Jukema JW, Hyland M, Gaw AJTAjoc: **The design of a prospective study of pravastatin in the elderly at risk (PROSPER)**. 1999, **84**(10):1192-1197.

5. Loft S, Poulsen HEJJomm: **Cancer risk and oxidative DNA damage in man**. 1996, **74**:297-312.

6. Kang B, Park J, Cho S, Lee M, Kim N, Min H, Lee S, Park O, Han B: **Current status, challenges, policies, and bioethics of biobanks**. *Genomics & informatics* 2013, **11**(4):211-217.

7. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L: **New goals for the U.S. Human Genome Project: 1998-2003**. *Science (New York, NY)* 1998, **282**(5389):682-689.

8. Cambon-Thomsen A, Ducournau P, Gourraud PA, Pontille D: **Biobanks for genomics and genomics for biobanks**. *Comparative and functional genomics* 2003, **4**(6):628-634.

9. Pontén F, Schwenk JM, Asplund A, Edqvist PHJJoim: **The Human Protein Atlas as a proteomic resource for biomarker discovery**. 2011, **270**(5):428-446.

10. Aardema MJ, MacGregor JT: **Toxicology and genetic toxicology in the new era of "toxicogenomics": impact of "-omics" technologies**. *Mutation research* 2002, **499**(1):13-25.

11. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J *et al*: **The UK Biobank resource with deep phenotyping and genomic data**. *Nature* 2018, **562**(7726):203-209.

12. Tomczak K, Czerwińska P, Wiznerowicz M: **The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge**. *Contemporary oncology (Poznan, Poland)* 2015, **19**(1a):A68-77.

13. Babu M, Snyder M: **Multi-Omics Profiling for Health**. *Molecular & cellular proteomics : MCP* 2023, **22**(6):100561.

14. Balloux F, Brynildsrud OB, Van Dorp L, Shaw LP, Chen H, Harris KA, Wang H, Eldholm VJTim: **From theory to practice: translating whole-genome sequencing (WGS) into the clinic**. 2018, **26**(12):1035-1048.

15. Meienberg J, Bruggmann R, Oexle K, Matyas GJHg: **Clinical sequencing: is WGS the better WES?** 2016, **135**:359-362.

16. Alfares A, Aloraini T, Alissa A, Al Qudsi A, Alahmad A, Al Mutairi F, Alswaid A, Alothaim A, Eyaid W, Albalwi MJGiM: **Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing**. 2018, **20**(11):1328-1333.

17. Saarenpää S, Shalev O, Ashkenazy H, Carlos V, Lundberg DS, Weigel D, Giacomello SJNb: **Spatial metatranscriptomics resolves host–bacteria–fungi interactomes**. 2024, **42**(9):1384-1393.

18. Forward AL, LRS L-RS: **Full-Length 16S/18S/ITS Sequencing for Advanced Microbial Diversity Analysis**.

19. Jiang L, Goh DX, Koh JHZ, Chan X, Brennan NA, Barathi VA, Hoang QV: **Applications of Genomics and Transcriptomics in Precision Medicine for Myopia Control or Prevention**. *Biomolecules* 2023, **13**(3).

20. Sengupta S, Bolin JM, Ruotti V, Nguyen BK, Thomson JA, Elwell AL, Stewart R: **Single read and paired end mRNA-Seq Illumina libraries from 10 nanograms total RNA**. *Journal of visualized experiments : JoVE* 2011(56):e3340.

21. Tam S, Tsao M-S, McPherson JDJBib: **Optimization of miRNA-seq data preprocessing**. 2015, **16**(6):950-963.

22. Huang S, Feng C, Chen L, Huang Z, Zhou X, Li B, Wang LL, Chen W, Lv FQ, Li TS: **Identification of Potential Key Long Non-Coding RNAs and Target Genes Associated with Pneumonia Using Long Non-Coding RNA Sequencing (lncRNA-Seq): A Preliminary Study**. *Medical science monitor : international medical journal of experimental and clinical research* 2016, **22**:3394-3408.

23. Xie Y, Li J, Li P, Li N, Zhang Y, Binang H, Zhao Y, Duan W, Chen Y, Wang YJFio: **RNA-Seq profiling of serum exosomal circular RNAs reveals Circ-PNN as a potential biomarker for human colorectal cancer**. 2020, **10**:982.

24. Xu H, Wang C, Song H, Xu Y, Ji GJMc: **RNA-Seq profiling of circular RNAs in human colorectal Cancer liver metastasis and the potential biomarkers**. 2019, **18**:1-6.

25. Tong Y, Gao H, Qi Q, Liu X, Li J, Gao J, Li P, Wang Y, Du L, Wang CJT: **High fat diet, gut microbiome and gastrointestinal cancer**. 2021, **11**(12):5889.

26. Liu Y, Beyer A, Aebersold R: **On the Dependency of Cellular Protein Levels on mRNA Abundance**. *Cell* 2016, **165**(3):535-550.

27. Yin H, Xie J, Xing S, Lu X, Yu Y, Ren Y, Tao J, He G, Zhang L, Yuan XJCRM: **Machine learning-based analysis identifies and validates serum exosomal proteomic signatures for the diagnosis of colorectal cancer**. 2024, **5**(8).

28. Hu A, Noble WS, Wolf-Yadlin A: **Technical advances in proteomics: new developments in data-independent acquisition**. *F1000Research* 2016, **5**.

29. Zhu W, Smith JW, Huang C-MJBri: **Mass spectrometry‐based label‐free quantitative proteomics**. 2010, **2010**(1):840518.

30. Megger DA, Bracht T, Meyer HE, Sitek BJBeBA-P, Proteomics: **Label-free quantification in clinical proteomics**. 2013, **1834**(8):1581-1590.

31. Cui Y, He L, Yang C-Y, Ye QJCP: **iTRAQ and PRM-based quantitative proteomics in early recurrent spontaneous abortion: biomarkers discovery**. 2019, **16**:1-15.

32. Si W, Zhang Y, Li X, Du Y, Xu QJN: **Understanding the functional activity of polyphenols using omics-based approaches**. 2021, **13**(11):3953.

33. Temporini C, Calleri E, Massolini G, Caccialanza GJMsr: **Integrated analytical strategies for the study of phosphorylation and glycosylation in proteins**. 2008, **27**(3):207-236.

34. Mann M, Jensen ONJNb: **Proteomic analysis of post-translational modifications**. 2003, **21**(3):255-261.

35. Peter‐Katalinić JJMie: **Methods in enzymology: O‐glycosylation of proteins**. 2005, **405**:139-171.

36. Komander DJBst: **The emerging complexity of protein ubiquitination**. 2009, **37**(5):937-953.

37. Bremang M, Cuomo A, Agresta AM, Stugiewicz M, Spadotto V, Bonaldi TJMb: **Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome**. 2013, **9**(9):2231-2247.

38. Satchell DJQR, Chemical Society: **An outline of acylation**. 1963, **17**(2):160-203.

39. Idle JR, Gonzalez FJJCm: **Metabolomics**. 2007, **6**(5):348-351.

40. Zhang A, Sun H, Wang P, Han Y, Wang XJA: **Modern analytical techniques in metabolomics analysis**. 2012, **137**(2):293-300.

41. Wang KC, Chang HYJCr: **Epigenomics: technologies and applications**. 2018, **122**(9):1191-1199.

42. Beck D, Ben Maamar M, Skinner MKJE: **Genome-wide CpG density and DNA methylation analysis method (MeDIP, RRBS, and WGBS) comparisons**. 2022, **17**(5):518-530.

43. Olova N, Krueger F, Andrews S, Oxley D, Berrens RV, Branco MR, Reik WJGb: **Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data**. 2018, **19**:1-19.

44. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch RJNar: **Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis**. 2005, **33**(18):5868-5877.

45. Bashtrykov P, Jeltsch AJEEM, Protocols: **DNA methylation analysis by bisulfite conversion coupled to double multiplexed amplicon-based next-generation sequencing (NGS)**. 2018:367-382.

46. Buenrostro JD, Wu B, Chang HY, Greenleaf WJJCpimb: **ATAC‐seq: a method for assaying chromatin accessibility genome‐wide**. 2015, **109**(1):21.29. 21-21.29. 29.

47. Meng J, Lu Z, Liu H, Zhang L, Zhang S, Chen Y, Rao MK, Huang YJM: **A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor package**. 2014, **69**(3):274-281.

48. Park PJJNrg: **ChIP–seq: advantages and challenges of a maturing technology**. 2009, **10**(10):669-680.

49. Mardis ERJNm: **ChIP-seq: welcome to the new frontier**. 2007, **4**(8):613-614.

50. Spruessel A, Steimann G, Jung M, Lee SA, Carr T, Fentz AK, Spangenberg J, Zornig C, Juhl HH, David KA: **Tissue ischemia time affects gene and protein expression patterns within minutes following surgical tumor excision**. *BioTechniques* 2004, **36**(6):1030-1037.

51. Kanai Y, Nishihara H, Miyagi Y, Tsuruyama T, Taguchi K, Katoh H, Takeuchi T, Gotoh M, Kuramoto J, Arai E *et al*: **The Japanese Society of Pathology Guidelines on the handling of pathological tissue samples for genomic research: Standard operating procedures based on empirical analyses**. *Pathology international* 2018, **68**(2):63-90.

52. Li C, Sun YD, Yu GY, Cui JR, Lou Z, Zhang H, Huang Y, Bai CG, Deng LL, Liu P *et al*: **Integrated Omics of Metastatic Colorectal Cancer**. *Cancer cell* 2020, **38**(5):734-747.e739.

53. Wang C, Wang Y, Chen J, Wang Y, Pang C, Liang C, Yuan L, Ma Y: **CLDN18.2 expression and its impact on prognosis and the immune microenvironment in gastric cancer**. *BMC gastroenterology* 2023, **23**(1):283.

54. Liang B, Wang S, Zhu XG, Yu YX, Cui ZR, Yu YZ: **Increased expression of mitogen-activated protein kinase and its upstream regulating signal in human gastric cancer**. *World journal of gastroenterology* 2005, **11**(5):623-628.

55. Zhao BC, Wang ZJ, Mao WZ, Ma HC, Han JG, Zhao B, Xu HM: **CXCR4/SDF-1 axis is involved in lymph node metastasis of gastric carcinoma**. *World journal of gastroenterology* 2011, **17**(19):2389-2396.

56. Stevens VL, Hoover E, Wang Y, Zanetti KAJM: **Pre-analytical factors that affect metabolite stability in human urine, plasma, and serum: a review**. 2019, **9**(8):156.

57. Yang S, McGookey M, Wang Y, Cataland SR, Wu HMJAjocp: **Effect of blood sampling, processing, and storage on the measurement of complement activation biomarkers**. 2015, **143**(4):558-565.

58. Krishnan S, Queiroz AT, Gupta A, Gupte N, Bisson GP, Kumwenda J, Naidoo K, Mohapi L, Mave V, Mngqibisa RJFii: **Integrative multi-omics reveals serum markers of tuberculosis in advanced HIV**. 2021, **12**:676980.

59. Breier M, Wahl S, Prehn C, Fugmann M, Ferrari U, Weise M, Banning F, Seissler J, Grallert H, Adamski JJPo: **Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples**. 2014, **9**(2):e89728.

60. Geyer PE, Voytik E, Treit PV, Doll S, Kleinhempel A, Niu L, Müller JB, Buchholtz ML, Bader JM, Teupser D *et al*: **Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies**. *EMBO molecular medicine* 2019, **11**(11):e10427.

61. López-Bascón MA, Priego-Capote F, Peralbo-Molina A, Calderón-Santiago M, Luque de Castro MD: **Influence of the collection tube on metabolomic changes in serum and plasma**. *Talanta* 2016, **150**:681-689.

62. Cruickshank-Quinn C, Zheng LK, Quinn K, Bowler R, Reisdorph R, Reisdorph N: **Impact of Blood Collection Tubes and Sample Handling Time on Serum and Plasma Metabolome and Lipidome**. *Metabolites* 2018, **8**(4).

63. Sotelo-Orozco J, Chen SY, Hertz-Picciotto I, Slupsky CM: **A Comparison of Serum and Plasma Blood Collection Tubes for the Integration of Epidemiological and Metabolomics Data**. *Frontiers in molecular biosciences* 2021, **8**:682134.

64. Ilies M, Iuga CA, Loghin F, Dhople VM, Thiele T, Völker U, Hammer E: **Impact of blood sample collection methods on blood protein profiling studies**. *Clinica chimica acta; international journal of clinical chemistry* 2017, **471**:128-134.

65. Banks RE, Stanley AJ, Cairns DA, Barrett JH, Clarke P, Thompson D, Selby PJ: **Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry**. *Clinical chemistry* 2005, **51**(9):1637-1649.

66. Leidinger P, Backes C, Rheinheimer S, Keller A, Meese E: **Towards Clinical Applications of Blood-Borne miRNA Signatures: The Influence of the Anticoagulant EDTA on miRNA Abundance**. *PloS one* 2015, **10**(11):e0143321.

67. Xu L, Qu YH, Chu XD, Wang R, Nelson HH, Gao YT, Yuan JM: **Urinary levels of N-nitroso compounds in relation to risk of gastric cancer: findings from the shanghai cohort study**. *PloS one* 2015, **10**(2):e0117326.

68. Chan AW, Mercier P, Schiller D, Bailey R, Robbins S, Eurich DT, Sawyer MB, Broadhurst D: **(1)H-NMR urinary metabolomic profiling for diagnosis of gastric cancer**. *British journal of cancer* 2016, **114**(1):59-62.

69. Chen Y, Zhang J, Guo L, Liu L, Wen J, Xu L, Yan M, Li Z, Zhang X, Nan P *et al*: **A characteristic biosignature for discrimination of gastric cancer from healthy population by high throughput GC-MS analysis**. *Oncotarget* 2016, **7**(52):87496-87510.

70. Herrington W, Illingworth N, Staplin N, Kumar A, Storey B, Hrusecka R, Judge P, Mahmood M, Parish S, Landray M *et al*: **Effect of Processing Delay and Storage Conditions on Urine Albumin-to-Creatinine Ratio**. *Clinical journal of the American Society of Nephrology : CJASN* 2016, **11**(10):1794-1801.

71. Jung CE, Chopyk J, Shin JH, Lukacz ES, Brubaker L, Schwanemann LK, Knight R, Wolfe AJ, Pride DT: **Benchmarking urine storage and collection conditions for evaluating the female urinary microbiome**. *Scientific reports* 2019, **9**(1):13409.

72. Wang X, Gu H, Palma-Duran SA, Fierro A, Jasbi P, Shi X, Bresette W, Tasevska N: **Influence of Storage Conditions and Preservatives on Metabolite Fingerprints in Urine**. *Metabolites* 2019, **9**(10).

73. Emwas AH, Luchinat C, Turano P, Tenori L, Roy R, Salek RM, Ryan D, Merzaban JS, Kaddurah-Daouk R, Zeri AC *et al*: **Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review**. *Metabolomics : Official journal of the Metabolomic Society* 2015, **11**(4):872-894.

74. Koopaie M, Ghafourian M, Manifar S, Younespour S, Davoudi M, Kolahdooz S, Shirkhoda M: **Evaluation of CSTB and DMBT1 expression in saliva of gastric cancer patients and controls**. *BMC cancer* 2022, **22**(1):473.

75. Belstrøm D, Holmstrup P, Bardow A, Kokaras A, Fiehn NE, Paster BJ: **Comparative analysis of bacterial profiles in unstimulated and stimulated saliva samples**. *Journal of oral microbiology* 2016, **8**:30112.

76. Ebersole JL, Schuster JL, Stevens J, Dawson D, 3rd, Kryscio RJ, Lin Y, Thomas MV, Miller CS: **Patterns of salivary analytes provide diagnostic capacity for distinguishing chronic adult periodontitis from health**. *Journal of clinical immunology* 2013, **33**(1):271-279.

77. Scebba F, Salvadori S, Cateni S, Mantellini P, Carozzi F, Bisanzi S, Sani C, Robotti M, Barravecchia I, Martella F *et al*: **Top-Down Proteomics of Human Saliva, Analyzed with Logistic Regression and Machine Learning Methods, Reveal Molecular Signatures of Ovarian Cancer**. *International journal of molecular sciences* 2023, **24**(21).

78. Yang P, Lei H, Fu Y, Chen C, Tang L, Xia S, Guo Y, Chen G, Xie M, Yang J *et al*: **Exosomal miR-151-3p in saliva: A potential non-invasive marker for gastric cancer diagnosis and prognosis modulated by Sijunzi decoction (SJZD) in mice**. *Heliyon* 2024, **10**(7):e29169.

79. Rabe A, Gesell Salazar M, Michalik S, Fuchs S, Welk A, Kocher T, Völker U: **Metaproteomics analysis of microbial diversity of human saliva and tongue dorsum in young healthy individuals**. *Journal of oral microbiology* 2019, **11**(1):1654786.

80. Foratori-Junior GA, Ventura TMO, Grizzo LT, Jesuino BG, Castilho A, Buzalaf MAR, Sales-Peres SHC: **Is There a Difference in the Proteomic Profile of Stimulated and Unstimulated Saliva Samples from Pregnant Women with/without Obesity and Periodontitis?** *Cells* 2023, **12**(10).

81. Song M, Bai H, Zhang P, Zhou X, Ying B: **Promising applications of human-derived saliva biomarker testing in clinical diagnostics**. *International journal of oral science* 2023, **15**(1):2.

82. Yang HJ, Seo SI, Lee J, Huh CW, Kim JS, Park JC, Kim H, Shin H, Shin CM, Park CH *et al*: **Sample Collection Methods in Upper Gastrointestinal Research**. *Journal of Korean medical science* 2023, **38**(32):e255.

83. Duarte D, Castro B, Pereira JL, Marques JF, Costa AL, Gil AM: **Evaluation of Saliva Stability for NMR Metabolomics: Collection and Handling Protocols**. *Metabolites* 2020, **10**(12).

84. Mills RH, Dulai PS, Vázquez-Baeza Y, Sauceda C, Daniel N, Gerner RR, Batachari LE, Malfavon M, Zhu Q, Weldon K *et al*: **Multi-omics analyses of the ulcerative colitis gut microbiome link Bacteroides vulgatus proteases with disease severity**. *Nature microbiology* 2022, **7**(2):262-276.

85. Nechvatal JM, Ram JL, Basson MD, Namprachan P, Niec SR, Badsha KZ, Matherly LH, Majumdar AP, Kato I: **Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces**. *Journal of microbiological methods* 2008, **72**(2):124-132.

86. Bassis CM, Moore NM, Lolans K, Seekatz AM, Weinstein RA, Young VB, Hayden MK: **Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles**. *BMC microbiology* 2017, **17**(1):78.

87. Cunningham JL, Bramstång L, Singh A, Jayarathna S, Rasmusson AJ, Moazzami A, Müller B: **Impact of time and temperature on gut microbiota and SCFA composition in stool samples**. *PloS one* 2020, **15**(8):e0236944.

88. Widjaja F, Rietjens I: **From-Toilet-to-Freezer: A Review on Requirements for an Automatic Protocol to Collect and Store Human Fecal Samples for Research Purposes**. *Biomedicines* 2023, **11**(10).

89. Wu SZ, Roden DL, Al-Eryani G, Bartonicek N, Harvey K, Cazet AS, Chan CL, Junankar S, Hui MN, Millar EA *et al*: **Cryopreservation of human cancers conserves tumour heterogeneity for single-cell multi-omics analysis**. *Genome medicine* 2021, **13**(1):81.

90. Shabihkhani M, Lucey GM, Wei B, Mareninov S, Lou JJ, Vinters HV, Singer EJ, Cloughesy TF, Yong WH: **The procurement, storage, and quality assurance of frozen blood and tissue biospecimens in pathology, biorepository, and biobank settings**. *Clinical biochemistry* 2014, **47**(4-5):258-266.

91. Sandusky GE, Teheny KH, Esterman M, Hanson J, Williams SD: **Quality control of human tissues--experience from the Indiana University Cancer Center-Lilly Research Labs human tissue bank**. *Cell and tissue banking* 2007, **8**(4):287-295.

92. Ghatak S, Muthukumaran RB, Nachimuthu SK: **A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis**. *Journal of biomolecular techniques : JBT* 2013, **24**(4):224-231.

93. Mullegama SV, Alberti MO, Au C, Li Y, Toy T, Tomasian V, Xian RR: **Nucleic Acid Extraction from Human Biological Samples**. *Methods in molecular biology (Clifton, NJ)* 2019, **1897**:359-383.

94. Michalska-Falkowska A, Niklinski J, Juhl H, Sulewska A, Kisluk J, Charkiewicz R, Ciborowski M, Ramlau R, Gryczka R, Piwkowski C *et al*: **Applied Molecular-Based Quality Control of Biobanked Samples for Multi-Omics Approach**. *Cancers* 2023, **15**(14).

95. Sheng Q, Vickers K, Zhao S, Wang J, Samuels DC, Koues O, Shyr Y, Guo Y: **Multi-perspective quality control of Illumina RNA sequencing data analysis**. *Briefings in functional genomics* 2017, **16**(4):194-204.

96. Gaignaux A, Ashton G, Coppola D, De Souza Y, De Wilde A, Eliason J, Grizzle W, Guadagni F, Gunter E, Koppandi I *et al*: **A Biospecimen Proficiency Testing Program for Biobank Accreditation: Four Years of Experience**. *Biopreservation and biobanking* 2016, **14**(5):429-439.

97. Brown RE, Jarvis KL, Hyland KJJAb: **Protein measurement using bicinchoninic acid: elimination of interfering substances**. 1989, **180**(1):136-139.

98. Simard JR, Lee L, Vieux E, Improgo R, Tieu T, Phillips AJ, Fisher SL, Pollock RM, Park EJSDAtSoDD: **High-throughput quantitative assay technologies for accelerating the discovery and optimization of targeted protein degradation therapeutics**. 2021, **26**(4):503-517.

99. Shickle DJSiH, History PoSPCSi, Biological Po, Sciences B: **The consent problem within DNA biobanks**. 2006, **37**(3):503-519.

100. Munshi R, Thatte UJPicr: **Compensation for research related injury**. 2013, **4**(1):61-69.

101. Henry LM, Larkin ME, Pike ERJJoL, Biosciences t: **Just compensation: a no-fault proposal for research-related injuries**. 2015, **2**(3):645-668.

102. Smith-Tyler JJPotATS: **Informed consent, confidentiality, and subject rights in clinical trials**. 2007, **4**(2):189-193.

103. D’Abramo F, Schildmann J, Vollmann JJBme: **Research participants’ perceptions and views on consent for biobank research: a review of empirical data and ethical analysis**. 2015, **16**:1-11.

104. Reilly J, Gallagher L, Chen JL, Leader G, Shen SJMA: **Bio-collections in autism research**. 2017, **8**:1-36.

105. Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KLJTAJoHG: **Public opinion about the importance of privacy in biobank research**. 2009, **85**(5):643-654.

106. Biobank UJUB: **UK Biobank ethics and governance framework**. 2007, **3**.

107. Staunton C, Moodley KJSSAMJ: **Data mining and biological sample exportation from South Africa: A new wave of bioexploitation under the guise of clinical care?** 2016, **106**(2):136-138.

108. Zhang X, Matsui K, Krohmal B, Zeid AA, Muthuswamy V, Koo YM, Kita Y, Lie RKJBme: **Attitudes towards transfers of human tissue samples across borders: an international survey of researchers and policy makers in five countries**. 2010, **11**:1-7.

109. Frascarelli C, Bonizzi G, Musico CR, Mane E, Cassi C, Guerini Rocco E, Farina A, Scarpa A, Lawlor R, Reggiani Bonetti LJJoPM: **Revolutionizing cancer research: the impact of artificial intelligence in digital biobanking**. 2023, **13**(9):1390.

110. Bonizzi G, Zattoni L, Fusco N: **Biobanking in the digital pathology era**. *Oncology research* 2021, **29**(4):229-233.

111. Issaq HJ, Waybright TJ, Veenstra TD: **Cancer biomarker discovery: Opportunities and pitfalls in analytical methods**. *Electrophoresis* 2011, **32**(9):967-975.

112. Lin L-L, Huang H-C, Juan H-FJJop: **Discovery of biomarkers for gastric cancer: a proteomics approach**. 2012, **75**(11):3081-3097.

113. Shi X-J, Wei Y, Ji BJFiMB: **Systems biology of gastric cancer: Perspectives on the omics-based diagnosis and treatment**. 2020, **7**:203.

114. Ma Y, Jiang Z, Pan L, Zhou Y, Xia R, Liu Z, Yuan LJIJoO: **Current development of molecular classifications of gastric cancer based on omics**. 2024, **65**(3):1-22.

115. Fu Q, Murray CI, Karpov OA, Van Eyk JEJMSR: **Automated proteomic sample preparation: The key component for high throughput and quantitative mass spectrometry analysis**. 2023, **42**(2):873-886.

116. More D, Khan N, Tekade RK, Sengupta PJCRiAC: **An update on current trend in sample preparation automation in bioanalysis: Strategies, challenges and future direction**. 2024:1-25.

117. Mamo N, Martin GM, Desira M, Ellul B, Ebejer J-PJEJoHG: **Dwarna: a blockchain solution for dynamic consent in biobanking**. 2020, **28**(5):609-626.

Figure 1. Biobank-Clinical Integration Flowchart

Figure 2. Biobank-Clinical Integration Flowchart