INTRODUCTION

Development of the protocol

Developed within the framework of the *Mutographs* project (https://www.cancergrandchallenges.org/mutographs), this protocol supported a large multicancer genomic study led by the Wellcome Sanger Institute and funded by Cancer Research UK. From 2017 to 2023, *Mutographs* aimed to elucidate unexplained global differences in cancer incidence by analyzing mutational signatures through whole-genome sequencing (WGS) of tumor DNA from >6,000 patients with carcinomas of the esophagus, kidney, colorectum, pancreas, head and neck, gallbladder, and urinary bladder, collected from 27 countries across five continents. The International Agency for Research on Cancer (IARC, WHO) coordinated the collection of biological materials and data, performed centralized pathology evaluation of frozen cancer tissues, independent to local diagnoses, and extracted paired tumor–blood DNA for sequencing at the Wellcome Sanger Institute. Despite logistical diversity—large cohort size, multiple infrastructure levels, and variable freezing methods—the workflow was consistently reproducible, enabling multiple high-impact publications [1-4].

The overall design of the *Mutographs* project and its achievements in developing a global cancer biorepository have been previously described [5]. To harmonize practices across 51 participating centers, IARC produced a detailed Standard Operating Procedure (SOP), publicly available on Zenodo (https://zenodo.org/doi/10.5281/zenodo.11836371). **Box 1** summarizes biological material collection, tissue freezing methods and shipping to IARC for *Mutographs project.*

Our protocol is specifically designed for the centralized processing of **frozen tumor tissues** to obtain high-quality DNA for WGS while ensuring rigorous pathology validation. The main objectives were to:

1. Generate reliable morphological and pathology data to enhance interpretation of sequencing results.
2. Align technical feasibility with large-scale, multi-site collection of variable tissue sizes and freezing conditions.
3. Generate reliable morphological and pathology data to enhance interpretation of sequencing results.
4. Align overall SOPs and procedure manual, including robust quality control and quality assurance (QC/QA) methods, to one of the most quality and quantity sensitive analytical method in high-throughput genomic analysis, i.e. WGS.
5. Provide harmonized, traceable digital pathology reporting through a secure pathology data registry program.

The protocol is **pathologist-driven**, integrating diagnostic expertise throughout the molecular pipeline. Based on diagnostic pathology principles but adapted for frozen materials, it addresses the technical challenges inherent to non- formalin-fixed paraffin embedded (FFPE) specimens. Frozen tissues are preferred for genomic studies owing to superior DNA integrity [6-8] and have been applied in other large-scale initiatives [9-11], yet their handling requires specialized expertise. Recent efforts have optimized WGS for clinical workflows [12]. The *Mutographs* workflow uniquely embeds pathology into the sequencing pipeline to maintain diagnostic precision and reproducibility in a large sample size study. It is practical, reproducible, and applicable to various tissue-based research and omics studies. This protocol is now established as the main approach for tissue-based cancer research at IARC, with project specific adaptations when required.

Overview of the procedure

Box 1 summarizes the collection and freezing methods used by recruiting centers.

Figure 1 illustrates the overall workflow, comprising five main parts and 12 sub-parts—from treating of frozen tissues received from contributing centers, to extraction and aliquoting of tumor and germline DNA for sequencing, and is summarized below:

Part 1: Explains embedding in OCT and frozen sectioning for three objectives in three sub-parts: H&E staining for pathology evaluation **(1A)**; thick tissue scrolls for extraction of nucleic acids **(1B)**; and frozen sectioning for Laser Capture Microdissection (LCM) **(1C).**

Part 2: Explains multistep pathology evaluation of Hematoxylin and Eosin (H&E) whole-slide images (WSI), integrating digital annotation, QC, and centralized data registry. Five approaches (**2A to 2E**) were instructed by pathologist and guided the processing of tissues. This part is illustrated in **Figure 2**. Pathology evaluation process was designed to validate inclusion of high-quality tissues and to provide morphological data to be integrated in the multidisciplinary interpretation of sequencing data. Contribution of a team of six external expert pathologists was facilitated by application of digital pathology. A **web-based pathology data registry** harmonized pathologists' reporting and ensured secure data tracking. Various quality control measures were implemented, including blinded evaluation to clinical data and local diagnoses, random dual independent reporting of 20% of cases, and structured resolution of discrepancies between pathologists. The reports and instructions were coordinated by a dedicated data manager, who extracted the pathology reports in a regular weekly basis, informed the responsible technical platforms based on pathologists' instructions, and informed the lead pathologists if any disagreement needed to be resolved.

Part 3: Introduces macrodissection to enrich tumor content when percentage of viable tumor cells (PVTC) is estimated < 50 %. Sequencing platform set the lowest threshold of PVTC as 50%. The workflow was explicitly designed based on the specific requirement for *Mutographs*, however, the protocol is easily adjustable to any other threshold.

Part 4: Defines LCM for low cellularity tumors, notably pancreatic ductal adenocarcinomas (PDAC).

Part 5: Covers DNA extraction and aliquoting, including automated tumor DNA extraction **(5A)**; automated germline (blood) DNA extraction **(5B)**; manual DNA extraction from LCM pieces **(5C)** and DNA aliquoting and preparation for the sequencing platform **(5D)**.

Advantages and limitations

Key advantages of our technique include:

- **Systematic pathologist-led from design to execution**, making it a fully integrated pathology–molecular pipeline linking pathology to WGS.
- **High inter-pathologist reproducibility** (ICC ≥ 0.7 for PVTC estimation).
- **Multi-level WSI generation and systematic evaluation**, allowing robust morphological validation through deeper sectioning to confirm maintenance of eligibility criteria and desired quality and to check the success of tumor enrichment technique.
- **Systematic quality control checkpoints** through multi-step data monitoring by a dedicated data manager working directly with the lead pathologist.
- **Quality control of pathology reports** by being blinded to the diagnostic pathology reports and clinical data, independent dual assessments, and a centralized decision process recorded in a digital registry.

- **Universal and adaptable structure**, able to accommodate different cancer types, tissue characteristics, and project-specific genomic endpoints.
- **Feasible, sustainable, and reproducible in the long term.**
- **Digital pathology registry** enables traceable, harmonized data reporting across pathologists.
- **Efficient resource use**, with exclusion of unsuitable samples before costly WGS.

Limitations include:

- **Technical training**: Frozen sectioning differs from FFPE sectioning, is technically challenging, and requires skilled technicians.
- **Infrastructure**: Frozen tissue handling requires rigorous cold chain logistics and cryostats to preserve tissue quality and to avoid heat shock impacting quality of nucleic acids.
- **Embedding medium (OCT)**: Although compatible with DNA extraction (consider extra washes), it might be unsuitable for proteomic and metabolomic workflows.
- **Safety**: Manual macrodissection introduces extra safety considerations due to using scalpel on frozen tissues.
- **Time and cost: LCM is time-consuming and cost-sensitive and requires specific technical training and regular supervision.**
- **Training: Pathology evaluation of frozen tissues requires training and expertise and is not directly taken from experience in routine practice of pathology.**