



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

1 September 2017
EMA/CHMP/ICH/11623/2016
Committee for Human Medicinal Products

ICH guideline E18 on genomic sampling and management of genomic data

Step 5

Transmission to CHMP	28 January 2016
Transmission to interested parties	1 February 2016
Deadline for comments	31 May 2016
Final adoption by CHMP	14 September 2017
Date for coming into effect	28 February 2018



E18

Document History

Code	History	Date
E18	Approval by the ICH Assembly under <i>Step 2</i> and release for public consultation.	10 December 2015
E18	Adoption by the Regulatory Members of the ICH Assembly under <i>Step 4</i> and recommendation for adoption to the ICH regulatory bodies.	TBC

ICH guideline E18 on genomic sampling and management of genomic data

Table of contents

1. Introduction	4
1.1. Objectives of the guideline	4
1.2. Background	4
1.3. Scope of the guideline.....	4
1.4. General principles.....	5
2. Genomic sampling	5
2.1. Collection and processing of samples	6
2.1.1. Specimen type	6
2.1.2. Timing of specimen collection.....	6
2.1.3. Specimen preservation conditions.....	6
2.1.4. Specimen stability and degradation.....	7
2.1.5. Specimen volume and composition	7
2.1.6. Parameters influencing genomic sample quality.....	7
2.1.7. Sources of interference	7
2.2. Transport and storage of samples.....	8
2.2.1. Transport of samples	8
2.2.2. Storage of samples	8
2.2.3. Curation of sample inventory	8
3. Genomic data	9
3.1. Generation of genomic data.....	9
3.2. Handling and storage of genomic data	10
4. Privacy and confidentiality	10
4.1. Coding of samples and data.....	10
4.2. Access to genomic samples and data	10
5. Informed consent.....	11
6. Communication of findings	11

1. Introduction

1.1. Objectives of the guideline

The main objective of this guideline is to provide harmonised principles of genomic sampling and management of genomic data in clinical studies. This guideline will facilitate the implementation of genomic studies by enabling a common understanding of critical parameters for the unbiased collection, storage, and optimal use of genomic samples and data. This guideline also intends to increase awareness and provide a reminder regarding subjects' privacy, protection of the data generated, the need to obtain suitable informed consent, and the need to consider transparency of findings in line with local legislation and regulations. This guideline is intended to foster interactions amongst stakeholders, including drug developers, investigators and regulators, and to encourage genomic research within clinical studies.

1.2. Background

Awareness of, and interest in, genomic data obtained from clinical studies are growing. In particular, genomic research could be used in all phases of drug development to assess genomic correlates of drug response, and to understand mechanisms of disease or drug pharmacology. The identification of genomic biomarkers underlying variability in drug response may be valuable to optimize patient therapy, design more efficient studies, and inform drug labelling. Furthermore, the generation and interpretation of genomic data, both within and across clinical studies and drug development programs, allow for a better understanding of pharmacological and pathological mechanisms and enable the identification of new drug targets.

Regulatory agencies in the ICH regions have independently published guidelines encouraging genomic sample collection throughout the life cycle of a drug. The lack of a harmonised ICH guideline on genomic sampling and data management from clinical studies makes it difficult for sponsors and researchers to collect genomic samples and conduct genomic research in a consistent manner in global clinical studies. Genomic samples may be used for a variety of analyses, including single genes, sets of genes and the whole genome, which may or may not be pre-specified in the clinical study objectives at the time of collection.

1.3. Scope of the guideline

The scope of this guideline pertains to genomic sampling and management of genomic data obtained from interventional and non-interventional clinical studies. Genomic research can be conducted during or after a clinical study. It may or may not be pre-specified in the clinical protocol. This document addresses use of genomic samples and data irrespective of the timing of analyses and both pre-specified and non-pre-specified use. Genomic samples and data described in this guideline are consistent with the deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) characteristics defined in ICH E15¹.

The focus of this guideline is on the general principles of collection, processing, transport, storage, and disposition of genomic samples or data, within the scope of an informed consent. Technical aspects are also discussed when appropriate, recognizing the rapidly evolving technological advances in genomic sampling and data generation.

¹ ICH E15 defines a genomic biomarker as a "measurable DNA and/or RNA characteristic that is an indicator of normal biologic processes, pathogenic processes, and/or response to therapeutic or other interventions".

No detailed guidance is included on biobanking regulations or ethical aspects, as these are governed by the principles of the Declaration of Helsinki and national rules and regulations.

The same applies to issues related to privacy/data protection. The principles in this guideline may apply to any genomic research utilizing human-derived materials.

The recommendations in this guideline are principles and they should be interpreted in accordance with legislations, regulations as well as policies in each jurisdiction where genomic research is undertaken.

1.4. General principles

With advances in science and increased awareness of the impact of genomics, there is a need and an opportunity to maximize the value of the collected samples and the data generated from them. Therefore, genomic sample acquisition is strongly encouraged in all phases and studies of clinical development. Moreover, the quality of genomic research is dependent upon unbiased systematic collection and analysis of samples, ideally from all subjects participating in the trial, in order to fully represent the study population.

Maintaining sample integrity is important and has a major impact on the scientific utility of genomic samples. The quality and amount of samples, and technical performance of the assay (e.g., accuracy, precision, sensitivity, specificity, reproducibility) will determine the reliability of genomic data. Establishing standardized practice for handling and processing of genomic samples will foster integration of data from different analytical platforms and facilitate decision making.

Genomic samples and data should be securely stored, maintained, and access-controlled similar to non-genomic samples and health information.

2. Genomic sampling

Genomic research encompasses a wide variety of methods and applications. These may include, but are not limited to, nucleic acid sequencing and genotyping, analysis of various types of RNA, gene expression or regulation, and detection of epigenetic modifications. Ever evolving technological advancements are expected to yield novel applications. The scope of the research will determine the specimen type and the analyte(s) to be assessed, and the methodologies used to extract analytes and to stabilize and store well-annotated samples for genomic testing. Sample quality and amount can influence the accuracy and reliability of the generated data. Therefore, collection, handling, and preparation of the biological samples are critical steps in the research process.

Pre-analytical variation should be minimized by developing and documenting standardized procedures for genomic sample collection, processing, transport, and storage. Procedures and quality monitoring should be tailored to the types of specimens, the analytes, and the tests to be performed. The pre-analytical process for specimen handling and preparation across study sites should be defined, documented, and verified prior to implementation. It is important that the timing, method, location (e.g., study site), and conditions (e.g., storage temperature and duration, fixation time) under which samples are collected are recorded. Any changes and deviations from procedures defined should be well documented over the life time of each sample. The chain of custody at all stages of collection, handling, and analysis including the timing of each step should be recorded for all samples and their aliquots. Implementation of quality control programs is highly recommended. In general, instructions for collection, processing, transport, and storage should be defined and adopted to ensure the stability of the biological samples at each step from the time of acquisition to the time of testing.

2.1. Collection and processing of samples

A number of pre-analytical variables should be considered when developing a strategy for sample collection and processing to ensure suitability of samples for genomic testing. If sites participating in a clinical study use different sample collection and handling procedures, then the subsequent test performance may differ by site. This may affect the interpretability and combinability of the data and may lead to unreliable results. Staff at all participating sites should be properly trained and knowledgeable in the use of standardized procedures.

Specimens should be collected and labelled in accordance with appropriate biosafety practices, subject privacy regulations, and the informed consent.

2.1.1. Specimen type

Nucleic acids may be extracted from a variety of clinical specimen types and matrices (e.g., blood cells, tissue, buccal swab, saliva, bone marrow aspirate, urine, feces). Novel sources of tissue-derived nucleic acids (e.g., cell-free DNA, circulating tumor cells) are emerging and might require distinct isolation methods. The principles detailed herein also apply to these sources. The type of specimens to be collected should be compatible with the intended use. For example, some types of specimens could be used for both DNA and RNA studies while other specimen types may not be suitable for RNA analysis due to the lack of analyte stability.

In pediatric subjects, only limited amounts of blood or other tissues may be available and therefore non-invasive alternatives, such as saliva, dried blood spot or skin scrapings could be considered. For certain types of sample (e.g., blood, muscle biopsy) attention should be paid to aseptic collection. Care should be taken when biological materials that may bear the risk for contamination with other than host DNA and RNA are used (e.g., buccal swabs, saliva).

2.1.2. Timing of specimen collection

Inter- and intra-subject variability should be considered in the context of the clinical study objectives when defining the sample collection strategy. For example, diurnal variation or administered treatments can influence gene expression and should be considered when selecting sampling time points. Epigenetics such as DNA methylation may also change over time (e.g., subject age). While the sequence of germline DNA is relatively stable over time, information obtained from tumor DNA and RNA can be affected by the source, method and/or timing of the sample collection.

2.1.3. Specimen preservation conditions

The collection container, the need for and the type of an additive, stabilizing agent or preservative will depend upon the nucleic acid target, the specimen type, the size or volume of sample required, and the potential assay and technology. For example, blood or bone marrow aspirate specimens are collected in tubes containing anticoagulants or additives appropriate for the intended nucleic acid type. Tissue samples may be snap-frozen in liquid nitrogen or placed in an appropriate preservative.

Tissues are often fixed for long-term storage. Parameters that should be carefully considered for tissue fixation are the type of fixative, fixation time, humidity, oxygenation, and temperature, as well as the compatibility with the downstream nucleic acid extraction method. It is recommended to evaluate the impact of fixation and additives on the analytes of interest and the types of tests to be carried out prior to sample collection in a clinical study. In addition, the specimen tissue type and volume may affect the optimal duration of fixation and therefore should be taken into account. Sample handling methods subsequent to initial fixation could also impact the integrity of the specimens.

2.1.4. Specimen stability and degradation

Appropriate handling measures should be taken to prevent nucleic acid degradation and genomic profile alterations during sample collection and processing. Nucleic acid fragmentation and apparent changes in gene expression can occur and are dependent on conditions related to pH, hypoxia, the presence of endonucleases, and/or other tissue specific parameters. In addition, the time from specimen collection to freezing, fixation, or processing, as well as the storage time, should be optimized as needed. The parameters employed should be documented in sample collection and handling instructions, training materials and the sample reports. It is recommended to monitor conditions of storage and processing. For example, the temperature should be monitored for possible variations and documented to ensure consistency across samples.

2.1.5. Specimen volume and composition

Collection volume for samples is an issue that requires careful consideration. Consideration should be given to the minimum tissue or cell content needed for the intended purposes (e.g., analytical methodology) to minimize burden on subjects. The optimal amount of tissue may be dependent upon the cellularity of the tissue (e.g., smaller amounts may be sufficient for highly cellular tissue types) and the relative proportion of particular cell types in the entire specimen (e.g., tumor area and/or other aspects of a disease as represented in a biopsy). In the event that only a limited amount of tissue is available, alternative biological material may be considered for collection (see also section 2.1.1). As tumor tissue may exhibit molecular heterogeneity (mosaicism) and tumor biopsies often consist in part of normal tissue, a documented pathological evaluation of the sample prior to genomic analysis may be helpful. When paired samples are collected (e.g., tumor versus normal tissue, pre-versus post-treatment samples, prenatal versus maternal specimens), additional considerations (e.g., matched samples, cell types) may be needed to allow comparison.

2.1.6. Parameters influencing genomic sample quality

The quality and yield of the extracted nucleic acids are affected by the quality of the source specimens amongst other factors (see also section 2.1.5). As a result, the extraction procedures should be defined and validated for the handling conditions and the specimen type to be used. Specimen types have diverse characteristics and components that can affect the recovery of nucleic acids, and these should be considered when selecting a methodology for nucleic acid extraction. For example, the procedures for cell lysis may vary for different tissue and body fluid specimens. The process for removing specific cell constituents may also differ depending on the composition of the specimens. It should be noted that such processes may affect gene expression and lead to unintended spurious results. If both DNA and RNA will be extracted from the same specimen it should be determined whether extraction is best performed simultaneously or if the tissue specimen should be divided at the time of collection. Due to the labile nature of RNA compared to DNA, additional precautions are needed when isolating RNA, such as the use of RNase-free equipment and reagents. Repeated freezing and thawing of specimens prior to nucleic acid extraction can affect genomic sample integrity and should be avoided when possible or otherwise evaluated. To determine if the quality and quantity of the extracted nucleic acid targets are adequate for the intended downstream genomic testing, appropriate quality control methods should be applied relative to the analyte being measured.

2.1.7. Sources of interference

Potential sources of interference and contamination can affect the performance of genomic tests and these include endogenous and exogenous substances. It is important to identify endogenous

substances normally present in a specimen type (e.g., hemoglobin from blood, melanin from skin) that may affect, for example, polymerase chain reaction (PCR) efficiency and exogenous substances (e.g., anticoagulant, other additives, fixative, reagents used for nucleic acid isolation) that interfere with specific testing methods. The effects of potential factors or elements that interfere with assay performance should be addressed during assay development.

2.2. Transport and storage of samples

Transport and storage conditions will vary according to the specimen type and the nucleic acid target. In general, samples should not be exposed to conditions that may affect the stability of the nucleic acid targets during transport and storage.

2.2.1. Transport of samples

The appropriate transport conditions should be established prior to sample shipment. To ensure that specimens and/or extracted analytes are shipped under acceptable conditions, the dates of shipment and receipt should be documented, as well as the approximate temperature of the specimens when received. Where possible, samples should be transported at the intended storage temperature appropriate for the sample type and the analyte of interest. Deviations from the intended shipment parameters should be documented.

2.2.2. Storage of samples

It is highly recommended that samples are stored long-term, i.e., over the course of and beyond a drug development program, to enable re-use and/or future use. The conditions under which specimens or extracted nucleic acids are archived should be suitable for the intended genomic testing application. It is good practice to store samples and extracted nucleic acids as multiple aliquots to avoid repeated freeze/thaw cycles and potential contamination. Storage of aliquots in separate locations avoids simultaneous loss of all samples. If a sample is re-used and undergoes freeze/thaw cycles, then each freeze/thaw cycle, including the temperature and time at each step, should be recorded.

Storage of samples requires a physical infrastructure, as well as a robust laboratory information and data management system. Considerations when depositing samples into biorepositories include adherence to quality assurance and quality control programs, sample tracking systems, information security, and compliance with local legislations and informed consent. It is highly recommended that samples are stored in a physical infrastructure built with appropriate electrical backup systems and disaster plans. It is of the utmost importance that the party responsible for samples is clearly identified at all times and that the chain of custody is documented. Samples should not be stored longer than the allowed total retention period as described in the informed consent document. Furthermore, procedures should be in place to ensure appropriate destruction of the sample(s) when a subject withdraws consent or at the end of the declared retention period.

2.2.3. Curation of sample inventory

Sample inventory should be monitored and curated relative to the following: consent for use of the samples, length of storage relative to the sample retention policy, requests to withdraw samples from the biorepository, and record of sample destruction. Reconciliation of each sample relative to the aforementioned aspects should be performed prior to the use of that sample. Genomic concordance can be used to confirm the expected identities between samples (e.g., sample aliquots, tumor/normal pairs, pre- and post-treatment pairs).

3. Genomic data

Human genomic data can be derived from germline (inherited from parents), somatic (e.g., mutations in tumor tissues) or mitochondrial (e.g., for traceability of maternal lineage) sources. Biological specimens from humans may also include non-human genomic molecules (e.g., microbial DNA or other potentially infectious agents). The type of genomic data generated depends on the analytes and the applied technology platform(s). For comprehensive genomic comparisons it may be appropriate to have multiple DNA or RNA samples collected from a single subject taken from healthy tissue and diseased tissues, different tissue types, and/or at different time points.

3.1. Generation of genomic data

Genomic data can be generated by using many different and rapidly evolving technology platforms and methods. Broad genomic profiling of subjects is technologically feasible, and data generated may be stored and used repeatedly over time. It is important to choose the appropriate platform and method in light of the intended purpose of the genomic data. Therefore, it is relevant to understand whether research grade or more validated methods will be used during data generation. When genomic data are to be used for clinical decision making, the appropriate level of assay validation should be considered in accordance with local regulations and policies. Under exploratory settings, genomic data can be generated using research grade reagents and instruments that may not have been validated to support clinical use. However, even under exploratory settings where clinical-grade validation is not required, sufficient analytical validation should be conducted to ensure an adequate level of accuracy and consistency to allow for reliable interpretation of the results.

For genomic research, the processing and analytical workflow for all stages of analysis should be documented in as much depth as would be expected for upstream sample collection and processing. This documentation should include tools, versions, and parameters used at each stage of the analysis, source and version of genomic references and databases utilized (e.g. genome build, transcriptome assembly, variant annotation databases), and the computational environment and resources used to process data. Appropriate quality control (QC) procedures and metric thresholds should be set for all stages of the workflow. Such QC procedures and thresholds should be fixed for each stage of the workflow before proceeding to downstream analyses. Further, where possible QC procedures and thresholds should be fixed for a given type of assay and intended output to ensure consistency between different datasets. Likewise, bioinformatics tools, algorithms and relevant parameters should be selected prior to downstream analyses and as far as possible should be consistent for a given type of assay, analyte, and the output which the experiment is intended to detect.

While informatics methods in genomics change and evolve continuously, and numerous choices for tools, resources, and analysis methods may be available, where “gold-standard” approaches or resources have been defined by the community, these should be considered when designing a workflow. The use of publicly available annotation resources is highly recommended to enable cross-platform comparisons and integration of genomic and non-genomic (e.g., proteomic) results from different studies. Finally, any algorithms (including simple heuristics used by a researcher or clinician in interpretation of results) or statistical tests used to combine genomic data with clinical or biological data for treatment decisions or for research purposes should be documented appropriately.

Sponsors should ensure compliant use of samples and genomic data in accordance with purposeful and permitted use of samples for genomic data generation. The use of the genomic data should be in accordance with the protocol and the consent process in each region/jurisdiction.

3.2. Handling and storage of genomic data

It is important to understand how different types of genomic data are generated, handled, analyzed and stored. In general, an instrument generates one or more raw data files, which are then processed and converted to a format ready for integration with clinical or biological data. In addition to the final processed dataset, it is recommended to retain data files that maintain the complete features of the raw data; these could be either the raw data files or derived analysis-ready files along with workflow documentation, which allows for reconstruction of the primary data. Genomic data files should be stored in secured media with long-term capabilities. In addition, there should be a possibility to link the genomic data to other clinical data to allow for current and future use, as appropriate. Whereas genomic samples may be destroyed upon participant request, destruction of data contradicts the principles of scientific integrity, particularly in the context of clinical studies. Indeed, once data has been analysed and incorporated in the study results, the data cannot be destroyed without jeopardizing the scientific integrity of the clinical study. Therefore it is recommended to retain data once generated and used in a study. Where applicable, procedures may have to be developed to enable desired disposition of genomic data at request of the subject.

4. Privacy and confidentiality

Processing and handling of genomic samples and data should be conducted in a manner that protects the subjects' privacy. For genomic data, as for other clinical data, coding techniques as well as security and access procedures help maintain confidentiality. Appropriate security measures using coding schemata and restriction of access should be implemented at each step of sample collection, transport, analysis and storage. Consideration should also be given to data protection and confidentiality legislation and policies in each jurisdiction.

4.1. Coding of samples and data

Genomic data should be treated with the same high standards of confidentiality as other clinical data. ICH E15² describes various ways for coding of genomic samples and data, including single and double coding. To decrease complexity and likelihood of error, single coding is recommended for genomic samples and data, but should be consistent with local regulation or legislation. Anonymization, as defined in ICH E15, in which coding does not allow for subjects to be re-identified as the coding keys have been deleted, has limitations. That is because with the increasing availability of genomic information and analysis methods, it is not always possible to prevent re-identification of each study participant by deleting the link between the subjects' identifiers and the unique code(s). Furthermore, anonymization carries two implications: 1) the process renders the ability to connect previously de-linked genomic data to phenotypic data impossible, and 2) sample destruction pursuant to withdrawal of consent or for long-term clinical monitoring will not be possible. When processing genomic data, investigators and sponsoring organizations should respect the applicable privacy and data protection regulation and legislation.

4.2. Access to genomic samples and data

Use of genomic samples and data may involve repeated access, over time, in accordance with the informed consent (broad consent permitting sharing and distribution is recommended: see section 5). Such access could be within the sponsoring organization, or collaborators under supervision of the sponsoring organization, or external researchers. This would encompass both individual level data

² Refer to Section 2.3.2 in ICH E15

and/or aggregated results. Policies and procedures involving systems to ensure strict control of access rights with user access logs for all genomic samples and data, similar to those for other clinical data, should be developed. Policies and procedures governing access should consider risks of compromising the privacy of individual study participants as well as risks of compromising data quality and interpretation.

Risk-based approaches to access of genomic samples and data should be applied both within the consenting research institution as well as for external institutions with which samples and data are shared. When outsourcing sample storage, genomic analysis or data storage, contractual agreements should specify that the responsible party will supervise the outsourced facility in an appropriate manner to ensure that the samples and/or data are properly safeguarded.

Sharing data and/or samples with external organizations or researchers enables the enhancement of medical science and offers benefits at several levels, as well as maximizes transparency of research findings. Different national and local rules that govern sharing of individual level data with third parties including public databases shall be fully complied with.

5. Informed consent

Informed consent is part of Good Clinical Practices (GCP) considerations per ICH E6. Consent for genomic research may be either included in the consent for the clinical study or obtained separately. Genomic research has to be conducted in accordance with applicable local legislation and within the scope of informed consent, which includes collection and storage of genomic samples and data. The informed consent should describe, in simple language, the type and quantity of biological material to be collected, the collection procedures and the position on returning genomic data. If there are opportunities to obtain genetic counseling, it is desirable to include the information in the informed consent form. Specific considerations should be given to subjects who can only be enrolled in the study with the consent of the subjects' legal representatives or guardians (e.g., minors, subjects with dementia).

Whereas local regulations currently guide informed consent practices, the identification of common and essential elements for a globally acceptable informed consent for the collection and use of genomic samples would greatly enable genomic research.

Informed consent for the collection and use of genomic samples should permit broad analysis of the samples (e.g., sets of genes, transcriptome analysis or whole genome sequencing) regardless of the timing of analysis. Ideally, informed consent should allow for broad use of the samples, such as assay development, disease research, drug response, or pharmacovigilance. Local regulations and policies should be observed and respected.

6. Communication of findings

Genomic research, in the context of clinical studies, aims to assess genomic correlates of drug response, to advance understanding of disease biology and/or to identify mechanisms of drug pharmacology. This research may on occasion generate data incidental to the main objective of the intended research question, but may be of potential clinical relevance. Some of these data may also be clinically actionable. For example, BRCA1 mutations may be identified with whole genome sequencing during research that has not been intended to investigate cancer risk.

Research institutions and sponsors who generate genomic data in a study are encouraged to adopt a position and mechanisms regarding return of data to subjects, as appropriate. The position should articulate whether the intended research findings, incidental findings, neither or both will be

communicated. Ideally, the position would describe the timing of such communication (during or after the clinical study), by whom (e.g., investigator, physician, genetic counselor), and to whom (subjects, or the primary care giver or legal guardian in case subjects are children or subjects have been diagnosed with dementia). When communicating results to subjects, the pertinence of genetic counseling should be evaluated; the impact of results on treatment decisions should be interpreted clinically and discussed with the subject (or the primary care giver or legal guardian). The subject's desire and consent as to whether receiving such information should be respected. In addition, the applied assay and its level of validation should be considered, as this may affect the accuracy and validity of the results.