1 22 January 2016
2 EMA/CHMP/ICH/11623/2016
3 Committee for Human Medicinal Products

4 ICH guideline E18 on genomic sampling and management 5 of genomic data

6 Step 3

Adopted by CHMP for release for consultation	28 January 2016
Start of public consultation	February 2016
End of consultation (deadline for comments)	May 2016

7

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

8 E18

9 Document History

10

Code	History	Date
E18	Approval by the ICH Assembly under *Step 2* and release for public consultation.	10 December 2015

11

ICH guideline E18 on genomic sampling and management of genomic data

Table of contents

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. Further objectives of this guideline are to increase awareness and provide considerations regarding subject privacy, data protection, informed consent and transparency of findings.

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

There is growing awareness of, and interest in, genomic data generated from clinical studies. In particular, genomic research could be used in all phases of drug development to assess genomic correlates of drug response, disease understanding or mechanism of drug pharmacology. The identification of genomic biomarkers underlying variability in drug response may be valuable to optimize patient therapy, inform drug labelling and to design more efficient studies. 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 the 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 whole-genome approaches, that 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 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 Desoxyribonucleic Acid (DNA) and Ribonuleic Acid (RNA) characteristics defined in ICH E15.

The focus 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 research.

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 principles in this guideline, however, may apply to any genomic research utilising human-derived materials.

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 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 overall quality of these 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 clinical 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 RNAs; 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, the analytes to be assessed and the methodologies used to extract, stabilize and store well-annotated samples for genomic testing. Sample quality and amount can influence the accuracy and reliability of the generated data. Therefore, handling and preparation of the biological samples are critical steps in the process.

Pre-analytical variation should be minimized by developing standardized procedures for genomic sample collection, processing, transport, and storage. Such 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 should be defined, documented and verified prior to implementation. It is important that the timing, method, location and conditions under which samples are collected are recorded. Any deviations in procedures should be well documented in the appropriate inventory database linked to the samples. The chain of custody at all stages of collection, handling and analysis including the timing of each step should be recorded for all samples. Implementation of quality control programs is highly recommended. In general, instructions for collection, processing, transport and storage should be 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 to use 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., whole blood, tissue, buccal swabs, saliva, bone marrow aspirate, urine, feces). Novel sources of tissue-derived nucleic acids (e.g., cell-free DNA and liquid biopsies) 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.

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. While the sequence of germline DNA is relatively stable and does not change with 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 and the need for 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 analytical 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. Handling 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 that conditions of storage and processing are monitored. 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 liquid samples is an issue that requires careful consideration. For example, in pediatric subjects, limited amounts of blood or other tissues may be available and therefore non-invasive alternatives, such as saliva, dried blood spot or skin scrapings (or tape) could be considered. Care should be taken when buccal swabs, saliva or other material is used, as they may bear the risk for contamination with other than host DNA and RNA.

Consideration should be given to the minimum tissue or cell content needed for the intended purposes. 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 or disease aspects represented in a biopsy). As tumor tissue may exhibit molecular heterogeneity (mosaicism), a documented pathological evaluation of the sample may be helpful prior to genomic analysis. In circumstances when paired samples are collected (e.g., tumor versus normal tissue, pre- versus post-treatment samples or 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. 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 different 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. 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 defined downstream genomic testing to be performed, appropriate quality control methods should be applied, such as spectrophotometric Optical Density (OD) 260/280 measurement.

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. The identification of endogenous substances normally present in a specimen type (e.g., hemoglobin from blood or melanin from skin may affect 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 is important to ensure reliable genomic datasets. The effects of potential interferents on 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 samples 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 recommended that samples and extracted nucleic acids are stored as multiple aliquots to avoid repeated freeze and thaw cycles, and potential contamination. 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, 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 time 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, and requests to withdraw samples from the biorepository. Reconciliation of all samples relative to the aforementioned aspects should be performed prior to the use of each sample.

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 and disease tissue 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 such that the generated data

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 validated methods are to be used during data generation. Under exploratory settings genomic data can be generated using research grade reagents and instruments that may not have been validated to support clinical use. When genomic data are to be used for clinical decision making, appropriate level of assay validation should be considered in accordance with local regulations and policies.

For genomic research, the processing and analysis workflow (pipeline) details (e.g., reference genome build, annotation database and parameters) used for mapping purposes should be documented. The use of standard, publicly available annotation (e.g., GenBank, dbSNP) and cross-referencing is highly recommended to enable cross-platform comparisons and integration of genomic and non-genomic (e.g., proteomic) results from different studies. The database version(s) used for annotation should be recorded to allow for data compatibility. In addition, bioinformatic algorithms used for treatment decisions should be documented appropriately.

Sponsors should ensure compliant use of samples and genomic data in alignment with purposeful and permitted use of samples for genomic data generation. The use of the genomic data should be in alignment with the protocol, the consent and, if applicable, legal or regulatory requirements.

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 a raw data file, which is then processed and converted into an analysis-ready format using appropriate Quality Control (QC) procedures, followed by the application of analytical software to generate the results (often referred to as data and analysis pipeline, respectively). 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 pipeline documentation, which should allow for reconstruction of the primary data. These data sources would form the basis to integrate genomic data generated from different technology platforms. Genomic data files should be stored in secured long-term media. 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.

4. Privacy and confidentiality

Processing and handling of genomic samples and data should be conducted in a manner that protects the confidentiality of subjects' individual data. For genomic data, like 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 analysis and storage. Suitable 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, which are single-coded and do not carry any personal identifiers. ICH E15 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, is not

recommended for genomic samples or data, because the process renders the ability to connect previously unlinked genomic data to phenotypic data impossible. In addition, anonymization does not allow for sample destruction pursuant to withdrawal of consent or for long term clinical monitoring.

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. Therefore, strategies and procedures involving systems that ensure strict control of access rights with access logs should be established for all genomic samples and data, similar to that for other clinical data. 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.

5. Informed consent

Informed consent should be obtained in accordance with 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. 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 severe dementia).

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

Ideally, informed consent for the collection and use of genomic samples should permit broad analysis of the samples (e.g., sets of genes, transcriptome, whole genome sequencing) regardless of the timing of analysis. Additional elements might include the possibility to use the samples for assay development, disease research, or pharmacovigilance.

6. Transparency and communication of findings

Subjects, their families and/or healthcare providers may wish to receive their results as related to the intended objectives of the genomic research as with any other clinical study data. Research, including genomic research, may on occasion generate data or reveal findings that are incidental to the main objective of the intended research question, but may be of potential clinical relevance. Some of these incidental findings may also be clinically actionable. For example, *BRCA1* mutations may be identified with whole genome sequencing during research that was not intended to investigate cancer risk.

It is therefore appropriate that research institutions and sponsors who generate genomic data in a study adopt a position regarding return of findings to subjects and their primary healthcare providers. 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) and to whom (subject or in case of children and incapacitated individuals the primary care giver and the primary health care provider) as appropriate. If results are communicated, the applied assay and its level of validation should also be considered. The person(s) responsible for communicating the findings will also need consideration and usually this would be the investigator, with a link to the informed consent. The subject's desire and consent to receive such information or not should be respected. Local and regional considerations as well as guidances may apply.