INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED TRIPARTITE GUIDELINE

DEFINITIONS FOR GENOMIC BIOMARKERS, PHARMACOGENOMICS, PHARMACOGENETICS, GENOMIC DATA AND SAMPLE CODING CATEGORIES
E15

Current *Step 4* version
dated 1 November 2007

This Guideline has been developed by the appropriate ICH Expert Working Group and has been subject to consultation by the regulatory parties, in accordance with the ICH Process. At Step 4 of the Process the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and USA.

E15
Document History

Code *	History	Date
E15	Approval by the Steering Committee under *Step 2* and release for public consultation.	25 October 2006

Current *Step 4* version

E15	Approval by the Steering Committee under *Step 4* and recommendation for adoption to the three ICH regulatory bodies.	1 November 2007

* *Code as per the new codification system adopted by the ICH Steering Committee in November 2007*

DEFINITIONS FOR GENOMIC BIOMARKERS, PHARMACOGENOMICS, PHARMACOGENETICS, GENOMIC DATA AND SAMPLE CODING CATEGORIES

ICH Harmonised Tripartite Guideline

Having reached *Step 4* of the ICH Process at the ICH Steering Committee meeting
on 1 November 2007, this guideline is recommended for
adoption to the three regulatory parties to ICH

TABLE OF CONTENTS

DEFINITIONS FOR GENOMIC BIOMARKERS, PHARMACOGENOMICS, PHARMACOGENETICS, GENOMIC DATA AND SAMPLE CODING CATEGORIES

1. INTRODUCTION

1.1 Objective of the Guideline

In order to develop harmonised approaches to drug regulation, it is important to ensure that consistent definitions of terminology are being applied across all constituents of the International Conference on Harmonisation (ICH). An agreement on definitions will facilitate the integration of the discipline of pharmacogenomics and pharmacogenetics into global drug development and approval processes.

1.2 Background

Pharmacogenomics and pharmacogenetics have the potential to improve the discovery, development and use of medicines. Each of the ICH regions has published specific pharmacogenomic and pharmacogenetic guidelines, or concept papers, and is in the process of developing others. However, the lack of consistently applied definitions to commonly used terminology raises the potential for either conflicting use of terms in regulatory documentation and guidelines, or, inconsistent interpretation by regulatory authorities, ethics committees and sponsor companies.

1.3 Scope of the Guideline

This guideline contains definitions of key terms in the discipline of pharmacogenomics and pharmacogenetics, namely genomic biomarkers, pharmacogenomics, pharmacogenetics and genomic data and sample coding categories. The validation and qualification processes for genomic biomarkers, evidence for their intended use and acceptance criteria across ICH regions are outside of the scope of this guideline. As new scientific knowledge in the discipline of pharmacogenomics and pharmacogenetics emerges, the current guideline will be reviewed and expanded if appropriate.

2. GUIDELINE

Definitions of a genomic biomarker, pharmacogenomics, pharmacogenetics, and genomic data and sample coding categories are detailed below. The definition of what constitutes a genomic biomarker is key to understanding the definitions of pharmacogenomics and pharmacogenetics and is therefore introduced in this guideline first. Additional information useful to an understanding of aspects covered by each of the definitions is also provided. Some of the principles described in this guideline might be applicable to proteomics, metabalomics and other related disciplines.

2.1 Genomic Biomarker

2.1.1 Definition

A genomic biomarker is defined as follows:

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.

2.1.2 Additional Information

1. A genomic biomarker could, for example, be a measurement of :

   - The expression of a gene;
   - The function of a gene;
   - The regulation of a gene.

2. A genomic biomarker can consist of one or more deoxyribonucleic acid (DNA) and/or ribonucleic acid (RNA) characteristics.

3. DNA characteristics include, but are not limited to:

   - Single nucleotide polymorphisms (SNPs);
   - Variability of short sequence repeats;
   - Haplotypes;
   - DNA modifications, e.g., methylation;
   - Deletions or insertions of (a) single nucleotide(s);
   - Copy number variations;
   - Cytogenetic rearrangements, e.g., translocations, duplications, deletions or inversions.

4. RNA characteristics include, but are not limited to:

   - RNA sequences;
   - RNA expression levels;
   - RNA processing, e.g., splicing and editing;
   - microRNA levels.

5. The definition of a genomic biomarker is not limited to human samples, but includes samples from viruses and infectious agents as well as animal samples, i.e., for the application of genomic biomarkers to non-clinical and/or toxicological studies.

6. The definition of a genomic biomarker does not include the measurement and characterisation of proteins or low molecular weight metabolites.

2.2 Pharmacogenomics and Pharmacogenetics

2.2.1 Definitions

2.2.1.1 Pharmacogenomics

Pharmacogenomics (PGx) is defined as:

The study of variations of DNA and RNA characteristics as related to drug response.

2.2.1.2 Pharmacogenetics

Pharmacogenetics (PGt) is a subset of pharmacogenomics (PGx) and is defined as:

The study of variations in DNA sequence as related to drug response.

2.2.2 Additional Information

1. The term drug should be considered synonymous with investigational (medicinal) product, medicinal product, medicine and pharmaceutical product (including vaccines and other biological products).

2. PGx and PGt are applicable to activities such as drug discovery, drug development, and clinical practice.

3. Drug response includes the processes of drug absorption and disposition (e.g., pharmacokinetics, (PK)), and drug effects (e.g., pharmacodynamics (PD), drug efficacy and adverse effects of drugs).

4. The definitions of PGx and PGt do not include other disciplines such as proteomics and metabalomics.

2.3 Categories for Genomic Data and Samples Coding

PGx and PGt research depends on the use of biological samples to generate data. A harmonised definition for the coding of these samples and their associated data will facilitate use in research and development of new medicines.

There are four general categories of coding: identified, coded, anonymised and anonymous. Coded data or samples can be single or double coded.

The implications of using a specific data and sample coding category should be considered in the design of PGx and PGt research studies.

Some implications are highlighted in this section and summarised in Table 1.

2.3.1 Identified Data and Samples

Identified data and samples are labelled with personal identifiers such as name or identification numbers (e.g., social security or national insurance number). As the samples and associated data are directly traceable back to the subject, it is possible to undertake actions such as sample withdrawal or the return of individual results in accordance with the subject's request. The use of identified data and samples allows for clinical monitoring, subject follow-up and the addition of new data from the subject. Identified data and samples offer privacy protection comparable to that of general healthcare confidentiality in everyday medical practice. Identified data and samples are generally not considered appropriate for purposes of clinical trials in drug development.

2.3.2 Coded Data and Samples

Coded data and samples are labelled with at least one specific code and do not carry any personal identifiers.

2.3.2.1 Single Coded Data and Samples

Single coded data and samples are usually labelled with a single specific code and do not carry any personal identifiers. It is possible to trace the data or samples back to a given individual with the use of a single coding key. In general, the clinical investigator is responsible for maintaining the coding key. As the samples and associated data are indirectly traceable back to the subject via the coding key, it is possible to undertake actions such as sample withdrawal, or the return of individual results in accordance with the subject's request. The use of single coded data and samples allows for clinical monitoring, subject follow-up or

the addition of new data from the subject. Single coding is the current standard used in clinical research and offers additional safeguards to the subject's identifiers compared to the general healthcare confidentiality and privacy protection in everyday medical practice.

2.3.2.2   Double Coded Data and Samples

Double coded data and samples are initially labelled with a single specific code and do not carry any personal identifiers. The data and samples are then relabelled with a second code, which is linked to the first code via a second coding key. It is possible to trace the data or samples back to the individual by the use of both coding keys. In general, the clinical investigator is responsible for maintaining the first coding key and does not have access to the second coding key. As the samples and associated data can very indirectly be traced back to the subject via the use of both coding keys, it may be possible to undertake actions such as sample withdrawal, or the return of individual results in accordance with the subject's request. However additional electronic or technical processes may be added to further limit the ability to trace back from a genotype result to an individual subject. For example, a specific computer process that allows new subject data to be added but prevents the reconnection of the genotype data back to the individual subject identifier. The use of double coded data and samples allows for clinical monitoring, subject follow-up or the addition of new data from the subject. The use of the second code provides additional confidentiality and privacy protection for subjects over the use of a single code. Access to both coding keys is needed to link any data or samples back to a subject identifier.

2.3.3 Anonymised Data and Samples

Anonymised data and samples are initially single or double coded but where the link between the subjects' identifiers and the unique code(s) is subsequently deleted. Once the link has been deleted it is no longer possible to trace the data and samples back to individual subjects through the coding key(s). Anonymisation is intended to prevent subject re-identification. As anonymised samples and associated data are not traceable back to the subject, it is not possible to undertake actions such as sample withdrawal, or the return of individual results, even at the subject's request. The use of anonymised data and samples does not allow for clinical monitoring, subject follow-up or the addition of new data from the subject. The deletion of the coding key(s) linking the data and samples to a given subject's identifiers provides additional confidentiality and privacy protection over coded data and samples, as it prevents subject re-identification through the use of the coding key(s).

2.3.4 Anonymous Data and Samples

Anonymous data and samples are never labelled with personal identifiers when originally collected, neither is a coding key generated. Therefore there is no potential to trace back genomic data and samples to individual subjects. In some instances only limited clinical data can be associated with anonymous samples (e.g., samples from subjects with diabetes, male, age 50-55, cholesterol>240 mg/dl). As anonymous samples and associated data are not traceable back to subjects, it is not possible to undertake actions such as sample withdrawal, or the return of individual results, even at the subject's request. The use of anonymous data and samples does not allow for clinical monitoring, subject follow-up, or the addition of new data.

2.3.5 Additional Information

The use of a specific coding category in relation to obtaining informed consent from subjects is not within the focus of this guideline and is not addressed herein.

The conditions under which the genomic data can be linked back to a subject's personal identifiers for any purpose, including the return of genomic data to the subject, should be described in research related documents, e.g., the informed consent document.

Table 1: Summary of Genomic Data and Sample Coding Categories

Sample Coding Category		Link Between Subject's Personal Identifiers and Genomic Biomarker Data	Traceability Back to the Subject (Actions possible, including e.g., sample withdrawal or return of individual genomic results at subject's request)	Ability to Perform Clinical Monitoring, Subject Follow-up, or Addition of New Data	Extent of Subject's Confidentiality and Privacy Protection
Identified		Yes (direct) Allows for subjects to be identified	Yes	Yes	Similar to general healthcare confidentiality and privacy
Coded	*Single*	Yes (indirectly) Allows for subjects to be identified (via single, specific coding key)	Yes	Yes	Standard for clinical research
	Double	Yes (very indirectly) Allows for subjects to be identified (via the two specific coding keys)	Yes	Yes	Added privacy and confidentiality protection over single code
Anonymised		No Does not allow for subjects to be re-identified as coding key(s) have been deleted	No	No	Genomic data and samples no longer linked to subject as coding key(s) have been deleted
Anonymous		No Identifiers never collected and coding keys never applied Does not allow for subjects to be identified	No	No	Genomic data and samples never linked to subject