
Guidance for Industry

E15 Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories

**U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)**

**April 2008
ICH**

Guidance for Industry

E15 Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories

Additional copies are available from:

*Office of Training and Communication
Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 51, Room 2201
Silver Spring, MD 20993-0002
(Tel) 301-796-3400
<http://www.fda.gov/cder/guidance/index.htm>*

*Office of Communication, Training and
Manufacturers Assistance, HFM-40
Center for Biologics Evaluation and Research
Food and Drug Administration
1401 Rockville Pike, Rockville, MD 20852-1448
(Tel) 1-800-835-4709 or 301-827-1800
<http://www.fda.gov/cber/guidelines.htm>.*

**U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)**

**April 2008
ICH**

TABLE OF CONTENTS

I.	INTRODUCTION (1, 1.1, 1.3)	1
II.	BACKGROUND (1.2)	2
III.	GUIDANCE (2)	2
A.	Genomic Biomarker (2.1)	2
	1. <i>Definition (2.1.1)</i>	2
	2. <i>Additional Information (2.1.2)</i>	2
B.	Pharmacogenomics and Pharmacogenetics (2.2)	3
	1. <i>Definitions (2.2.1)</i>	3
	2. <i>Additional Information (2.2.2)</i>	4
C.	Categories for Genomic Data and Samples Coding (2.3)	4
	1. <i>Identified Data and Samples (2.3.1)</i>	4
	2. <i>Coded Data and Samples (2.3.2)</i>	4
	3. <i>Anonymized Data and Samples (2.3.3)</i>	5
	4. <i>Anonymous Data and Samples (2.3.4)</i>	6
	5. <i>Additional Information (2.3.5)</i>	6
	Table 1: Summary of Genomic Data and Sample Coding Categories	7

Guidance for Industry¹

E15 Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories

This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

I. INTRODUCTION (1, 1.1, 1.3)²

This guidance contains definitions of key terms in the discipline of pharmacogenomics and pharmacogenetics, namely genomic biomarkers, pharmacogenomics, pharmacogenetics, and genomic data and sample coding categories. In the effort to develop harmonized 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). This guidance on definitions is intended to facilitate the integration of the discipline of pharmacogenomics and pharmacogenetics into global drug development and approval processes. As new scientific knowledge in the discipline of pharmacogenomics and pharmacogenetics emerges, the current guidance will be reviewed and expanded if appropriate.

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 guidance.

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are

¹ This guidance was developed within the Expert Working Group (Efficacy) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) and has been subject to consultation by the regulatory parties, in accordance with the ICH process. This document has been endorsed by the ICH Steering Committee at *Step 4* of the ICH process, November 2007. At *Step 4* of the process, the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan, and the United States.

² Arabic numbers reflect the organizational breakdown of the document endorsed by the ICH Steering Committee at *Step 4* of the ICH process, November 2007.

Contains Nonbinding Recommendations

cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. BACKGROUND (1.2)

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 for 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.

III. GUIDANCE (2)

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 guidance 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 guidance might be applicable to proteomics, metabolomics, and other related disciplines.

A. Genomic Biomarker (2.1)

1. Definition (2.1.1)

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. Additional Information (2.1.2)

a. 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

Contains Nonbinding Recommendations

- b. A genomic biomarker can consist of one or more deoxyribonucleic acid (DNA) and/or ribonucleic acid (RNA) characteristics.
- c. 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
- d. RNA characteristics include, but are not limited to:
 - RNA sequences
 - RNA expression levels
 - RNA processing, e.g., splicing and editing
 - microRNA levels
- e. 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 nonclinical and/or toxicological studies.
- f. The definition of a genomic biomarker does not include the measurement and characterization of proteins or low molecular weight metabolites.

B. Pharmacogenomics and Pharmacogenetics (2.2)

1. Definitions (2.2.1)

a. Pharmacogenomics (2.2.1.1)

Pharmacogenomics (PGx) is defined as:

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

b. Pharmacogenetics (2.2.1.2)

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

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

Contains Nonbinding Recommendations

2. *Additional Information (2.2.2)*

- a. The term *drug* should be considered synonymous with investigational (medicinal) product, medicinal product, medicine, and pharmaceutical product (including vaccines and other biological products).
- b. PGx and PGt are applicable to activities such as drug discovery, drug development, and clinical practice.
- c. 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).
- d. The definitions of PGx and PGt do not include other disciplines such as proteomics and metabolomics.

C. *Categories for Genomic Data and Samples Coding (2.3)*

PGx and PGt research depends on the use of biological samples to generate data. A harmonized 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, anonymized, 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 summarized in Table 1.

1. *Identified Data and Samples (2.3.1)*

Identified data and samples are labeled 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 health care confidentiality in everyday medical practice. Identified data and samples are generally not considered appropriate for purposes of clinical trials in drug development.

2. *Coded Data and Samples (2.3.2)*

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

Contains Nonbinding Recommendations

a. Single-Coded Data and Samples (2.3.2.1)

Single-coded data and samples are usually labeled 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 health care confidentiality and privacy protection in everyday medical practice.

b. Double-Coded Data and Samples (2.3.2.2)

Double-coded data and samples are initially labeled with a single specific code and do not carry any personal identifiers. The data and samples are then relabeled 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.

3. *Anonymized Data and Samples (2.3.3)*

Anonymized 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). Anonymization is intended to prevent subject re-identification. As anonymized 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 anonymized 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).

Contains Nonbinding Recommendations

4. Anonymous Data and Samples (2.3.4)

Anonymous data and samples are never labeled 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.

5. Additional Information (2.3.5)

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

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.

Contains Nonbinding Recommendations

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
<i>Identified</i>		Yes (direct) Allows for subjects to be identified	Yes	Yes	Similar to general health care confidentiality and privacy
<i>Coded</i>	<i>Single</i>	Yes (indirectly) Allows for subjects to be identified (via single, specific coding key)	Yes	Yes	Standard for clinical research
	<i>Double</i>	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
<i>Anonymized</i>		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
<i>Anonymous</i>		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