



# **Standard Operating Procedure of Biomarker Module**

**SOP 1.3 in iCMDB**

Version 1.0

January 22, 2016

## CONTENTS

Summary.....	2
Introduction and Purpose .....	5
Responsibilities.....	5
Specific Procedure .....	5
Reference .....	8
Internal and External References .....	8
External References.....	8
Change History .....	8

## TABLES

Table 1 Summary of SOP in biomarker module (using “EGFR” in <i>Homo sapiens/human</i> as example) .	2
Table 2 Type and description of gene .....	6
Table 3 Definition of categories .....	7

## FIGURES

Figure 1 Example of biomarker entry .....	4
---	---

## SOP of Biomarker Module

### SUMMARY

The standard operating of procedure (SOP) of biomarker module is summarized in Table 1.

**Table 1 Summary of SOP in biomarker module (using “EGFR” in *Homo sapiens/human* as example)**

Name	Description/Format	Examples
<b>Gene ID*</b>	Unique identifier of gene record in Entrez Gene.  Otherwise, record the cytogenetic location.  Use the numeric ID provided by Entrez Gene in NCBI.  Possible values are numbers, p and q.	1956  1p19q
<b>Biomarker Type</b>	Select based on the type of biomarker	Gene
<b>Gene Symbol/Protein Name*</b>	The name of basic functional unit of DNA sequence.  Follow the official name provided by Entrez Gene in NCBI.	EGFR
<b>Synonyms*</b>	Other names of the gene.  Follow the context in “Also known as” in Entrez Gene. If other common names were reported frequently in references, the name can be included.	ERBB, ERBB1, HER1, NISBD2, PIG61, mENA
<b>Gene Type*</b>	Role of the gene.  Possible values are tRNA, rRNA, snRNA, scRNA, snoRNA, miscRNA, ncRNA, protein coding, pseudo, other, and unknown (Table 2).	Protein coding
<b>Chromosome*</b>	Chromosome of gene.  Possible values are 1~22, X and Y.	7
<b>Strand*</b>	Orientation of the gene transcription.  Possible values are + or - .	+
<b>Start*</b>	Starting position on genome.  Use the number provided in genomic context.	55086678
<b>End*</b>	End position on genome.  Use the number provided in genomic context.	55279262

\*Fields are required for every entry.

## SOP of Biomarker Module

Table 1 (continued)

Name	Description/Format	Examples
<b>Category*</b>	Group biomarker based on the module that used this biomarker	Drug Efficacy Predictive Biomarker
<b>Assembly*</b>	Release version of genomic reference sequence.	GRCh37
<b>Mechanism Description</b>	Descriptive text about the gene, its cellular localization, its function, and its effect on phenotype.	<p><i>About gene</i></p> <p>EGFR encodes a transmembrane glycoprotein which is a member of the protein kinase superfamily. This protein is a receptor for members of the epidermal growth factor family.</p> <p><i>Cellular localization</i></p> <p>EGFR is located in cell surface protein and can bind to epidermal growth factor (Herbst et al., 2004).</p> <p><i>Function</i></p> <p>Binding of the protein to a ligand induces receptor dimerization and tyrosine autophosphorylation (Downward et al., 1984). This autophosphorylation elicits downstream activation and signaling by several other proteins which initiate several signal transduction cascades, principally the MAPK, AKT and JNK pathways, leading to DNA synthesis and cell proliferation (Oda et al., 2005).</p> <p><i>Effect on phenotype</i></p> <p>Mutations affecting EGFR expression or activity could result in cancer (Zhang et al., 2007; Walker et al., 2009).</p>
<b>Reference</b>	<p>References used in "Mechanism Description".</p> <p>Only numeric reference ID is used and separated by comma without space.</p>	5365,5366,5367,5368,5369

\* Fields are required for every entry.

# SOP of Biomarker Module

Gene ID:	1956	Biomarker Type:	gene
Gene Symbol/Protein Name:	EGFR	Synonyms:	ERBB; HER1; mENA; ER
Gene Type:	protein coding	Chromosome:	7
Strand:	+	Start:	55086678
End:	55279262		<button>Retrieve</button>
Assembly:	GRCh37	Category:	Drug Efficacy Predictive Biomarker
Mechanism Description:	<p>EGFR encodes a transmembrane glycoprotein which is a member of the protein kinase superfamily. This protein is a receptor for members of the epidermal growth factor family. EGFR is located in cell surface and can bind to epidermal growth factor (Herbst et al., 2004). Binding of the protein to a ligand induces receptor dimerization and tyrosine autophosphorylation (Downward et al., 1984). This autophosphorylation elicits downstream activation and signaling by several other proteins which initiate several signal transduction cascades, principally the MAPK, AKT and JNK pathways, leading to DNA synthesis and cell proliferation (Qda et al., 2005). Mutations affecting EGFR expression or activity could result in cancer (Zhang et al., 2007; Walker et al., 2009).</p>		
Reference:	260 characters remaining 5365,5366,5367,5368,5369		

Figure 1 Example of biomarker entry

# SOP of Biomarker Module

## INTRODUCTION AND PURPOSE

SOP 1.3 is a standard operating procedure to guarantee the accuracy and conciseness of data for biomarker module in iCMDB. It mainly uses Entrez Gene as reference.

The Purpose of SOP 2.1 is to provide detailed guidelines and standards for data search and data entry of biomarker in biomarker module.

## RESPONSIBILITIES

Whoever creates the record is responsible for the accuracy of the data in biomarker module.

## SPECIFIC PROCEDURE

1. Go to Biomarker module in ICMDDB backend database.
2. Click “Add New” hyperlink near “Search” button in the “Search for Biomarker” Page.
3. Record the “Gene ID” in the left right corner before summary section from Entrez Gene (NCBI), which is the unique identifier of gene record in Entrez Gene. Value in this variable is number only. If the biomarker is a cytogenetic alteration not a gene, record the cytogenetic symbol and select corresponding biomarker type. E.g. 1p19q

**i** Only number, p, q are allowed.

4. Record the “official symbol” in summary section from Entrez Gene (NCBI) as “Gene Symbol/Protein Name” in ICMDDB, which is the name of the gene. It must follow the official name provided by Entrez Gene.
5. Record the “Also known as” in summary section from Entrez Gene (NCBI) as “Synonyms” in ICMDDB, which is a different name of gene. If other common names were reported frequently in references, the name can be included.
6. Record the “Gene type” in summary section from Entrez Gene (NCBI) as “Gene Type” in ICMDDB. Gene type describes the role of gene. Our database iCMDB uses the same types and definitions of gene in Entrez Gene (<http://www.ncbi.nlm.nih.gov/books/NBK3841/#EntrezGene.Properties>). The detailed definition of different roles is described in Table 2.
7. Record the “Assembly” in genomic context from Entrez Gene (NCBI) as “Assembly” in ICMDDB. GRCh37 is preferred in ICMDDB.
8. Record location in genomic context from Entrez Gene (NCBI) as “Chromosome”, “Start” and “End” in ICMDDB. GRCh37 is preferred in ICMDDB.
9. Record the orientation of transcription process in genomic context from Gene Database (NCBI) as “Strand” in ICMDDB. If the transcription process runs from 5’ to 3’, the strand is sense strand (marked as “+”). If the transcription process runs from 3’ to 5’, the strand is antisense strand (marked as “-”).

**i** Alternatively, “retrieve” function can be used to automatically get information mentioned above. However, please confirm the information is current and correct before save.

10. Select the category following the criteria in Table 3**Error! Reference source not found..**
11. Describe gene in four aspects:
  - a) *About gene*  
Describe the basic information of gene in a sentence. E.g.: “Gene encodes xxx, which is a member of xxx.”
  - b) *Cellular localization*

## SOP of Biomarker Module

Describe the cellular localization in a sentence. E.g. “xxx is located/expressed/synthesized/distributed in xxx”.

c) *Function*

Describe the function and involved pathways in a few sentences. E.g. “xxx induces/elicits/initiate xxx, leading to xxx”.

d) *Effect on phenotype*

Describe the effect on phenotype including disease risk and disease predisposition in a sentence. E.g: “xxx result in/lead to xxx”.

**Table 2 Type and description of gene**

Type	Description
tRNA	transfer RNA
rRNA	ribosomal RNA
snRNA	small nuclear RNA
scRNA	small cytoplasmic RNA
snoRNA	small nucleolar RNA
miscRNA	miscellaneous RNA
ncRNA	non-coding RNA; includes all ncRNA classes except for snRNA, snoRNA, and scRNA
protein coding	protein coding
pseudo	pseudogene
other	when the type is known, but there is no specific enumeration for it; includes immunoglobulin and TCR gene segments, repetitive elements, regulatory
unknown	when the type of gene is uncertain

## SOP of Biomarker Module

**Table 3 Definition of categories**

<b>Name</b>	<b>Description</b>
Circulation tumor cell	Circulation tumor cell biomarker
Microsatellite	Microsatellite biomarker
Virus biomarker	Biomarker for infectious disease
Drug toxicity predictive biomarker	Biomarker that has predictive effect of drug toxicity
miRNA cancer diagnosis biomarker	miRNA that used for cancer diagnosis
Disease diagnostic biomarker in blood-early detection	Biomarker used for disease diagnosis in blood-early detection
Disease diagnostic biomarker-disease classification	Biomarker used for disease classification
Mendelian disorder related biomarker	Biomarker for Mendelian disorders
Disease prognostic biomarker	Biomarker that has prognosis indication
Drug efficacy predictive biomarker and drug target	Biomarker that used in Pharmacogenomics module
Drug target biomarker	Biomarker that is the target of drug
Drug efficacy predictive biomarker	Biomarker relevant to drug efficacy



## SOP of Biomarker Module

### REFERENCE

#### Internal and External References

SOP No.	Version	Description
1.4	1.0	SOP for Reference
1.8	1.0	Reference Selection

#### External References

No.	Resource	Description
1	Entrez Gene	Entrez Gene ( <a href="http://www.ncbi.nlm.nih.gov/gene">http://www.ncbi.nlm.nih.gov/gene</a> ) is National Center for Biotechnology Information (NCBI)'s database for gene-specific information. Entrez Gene maintains records from genomes which have been completely sequenced, which have an active research community to submit gene-specific information, or which are scheduled for intense sequence analysis. Records in Entrez Gene are assigned unique, stable and tracked integers as identifiers (PMCID: PMC3013746).

### CHANGE HISTORY

Revision	Date	Significant Changes
0.1	14-Dec-15	Draft of initial version
1.0	22-Jan-16	First release to department of database