

DS11: Evolution of Genomic Data Mappings from PF to GF in SDTM v3.4: Concepts, Challenges, and Controlled Terminology with Genetic Variation Examples

3-18-2025

Rama Kudaravalli

(Merck & Co., Inc., Rahway, NJ, USA)



Agenda

- Introduction
- Purpose
- SDTMIG for Genomic data
- Genomic concepts
- Controlled terminology
- Mapping Changes
- Summary



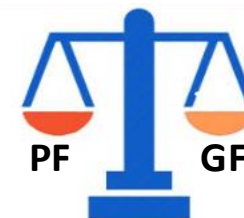
Introduction

- ❖ Genomic Medicine
- ❖ Genomic Findings (GF) in 3.4
 - New Variables
 - Controlled Terminology



Purpose

- ❖ Evolution of genomic based therapies in clinical trials
- ❖ Standardize genomic data
- ❖ Implement changes from PF to GF

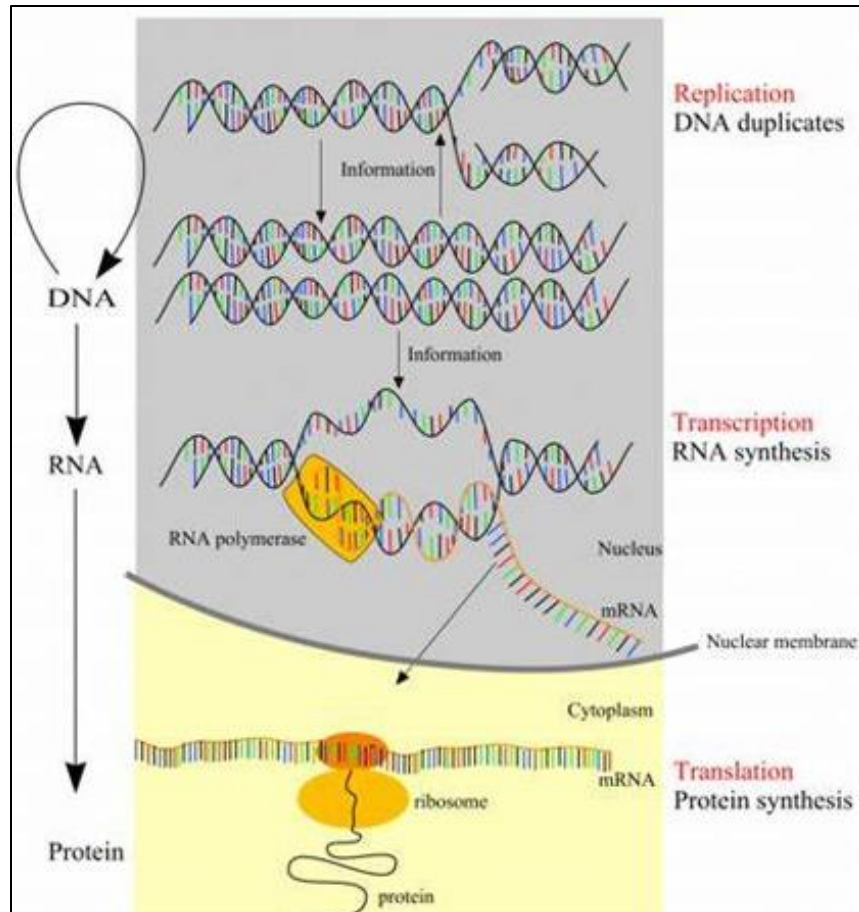


SDTMIG - Genomic Data

SDTMIG	SDTMIG-PGx v1.0	SDTMIG v3.4
CDISC	Provisional guidance May 2015	GF for genomic findings Nov 2021
Controlled Terms	December 2015	Evolving
Specimen Collection	BE, BS, RELSPEC	BE, BS, RELSPEC
Biomarkers	PB, SB	Deprecated
Methods	PG (PGx Methods and Supporting Information)	Deprecated
Genomic data	PF	GF



Genomic Concepts



Central dogma of molecular biology

DNA → RNA → Protein

Genetics - Study of genes, genetic variation and heredity in organisms

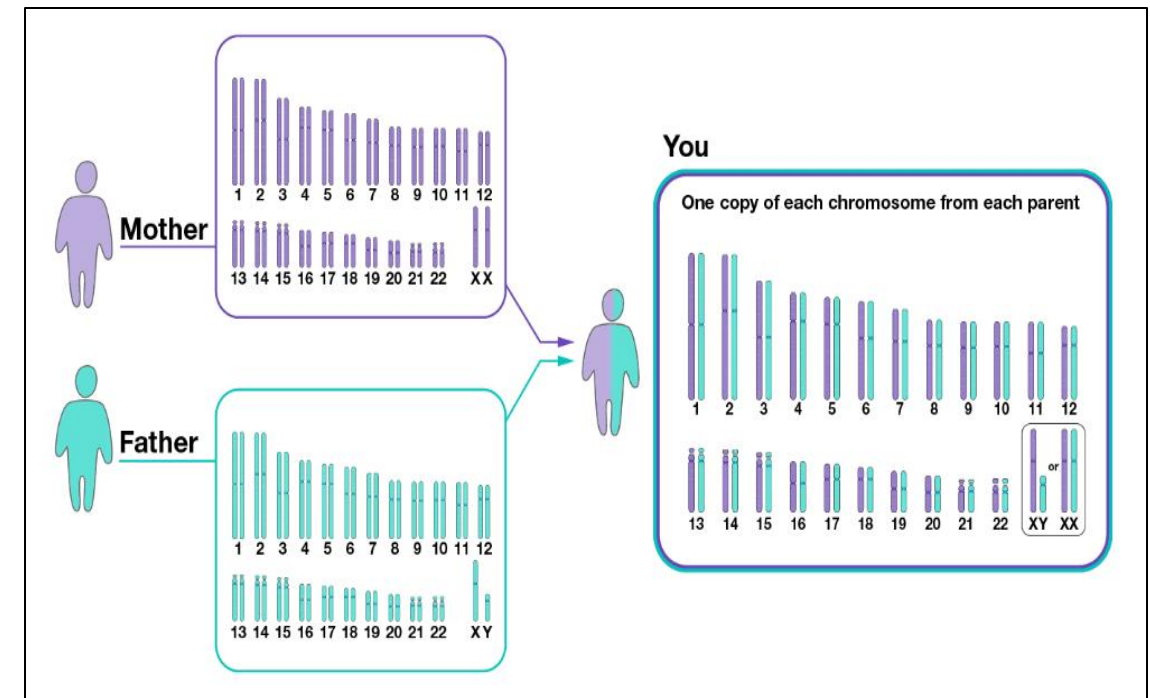
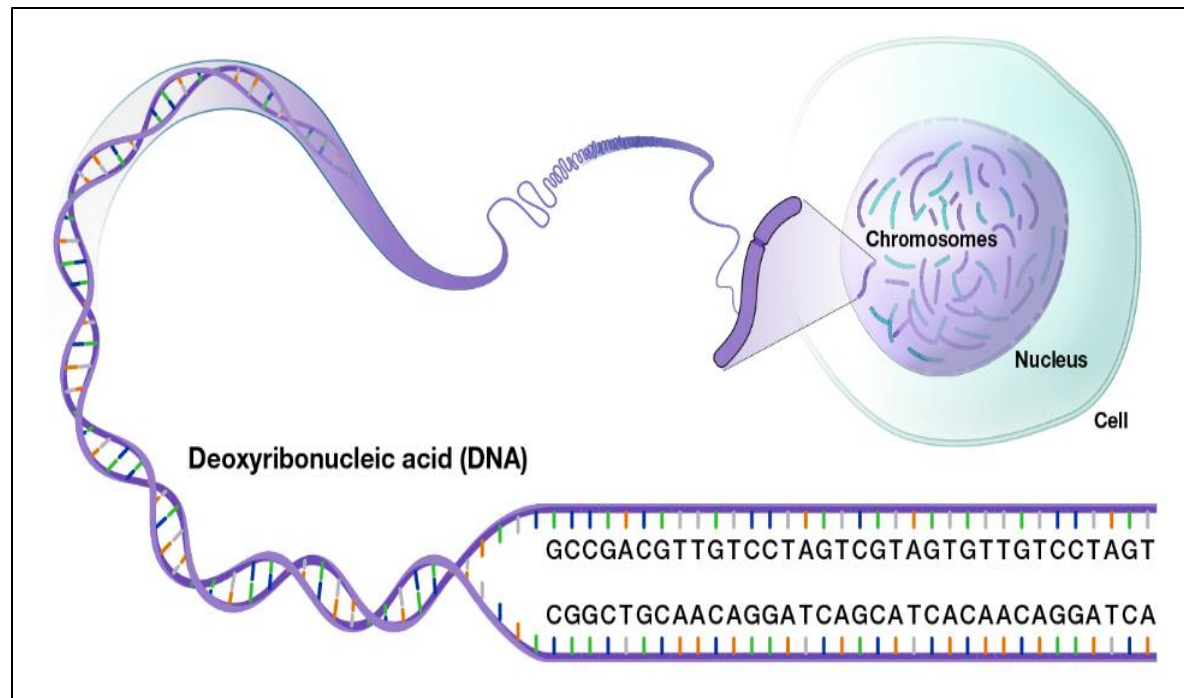
Transcriptomics - Study of the transcriptome, complete set of RNA transcripts that are produced

Proteomics - The large-scale study of proteomes, which are sets of proteins produced



Genomic Concepts - Genome

- DNA - 4 nucleotides (A, T, G, C) - **GFSPEC**
- CHROMOSOMES - **GFCHROM**
- GENE - **GFSYM, GFSYMTYP, GFGENLOC**





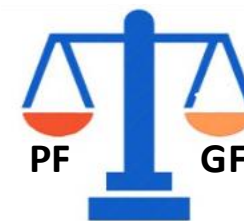
Genomic Concepts – Allele

Genotype	Phenotype
<i>homozygous brown</i> 	<i>brown eyes</i> 
<i>homozygous blue</i> 	<i>blue eyes</i> 
<i>heterozygous brown</i> 	<i>brown eyes</i> 

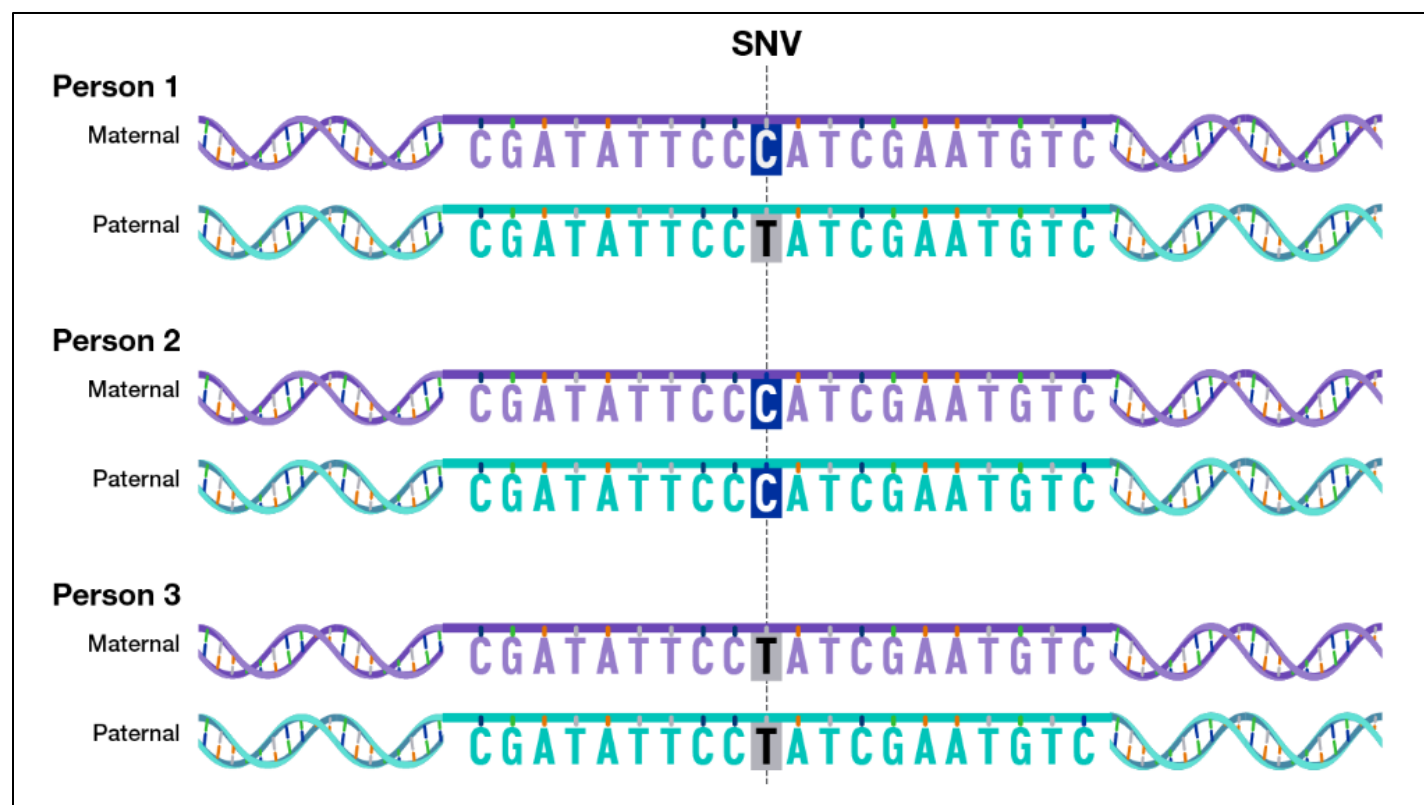
➤ EYE COLOR

- Brown (dominant)
- Blue (recessive)

GFTSTDTL



Genomic Concepts - Genomic Variations



❖ Single Nucleotide Variants

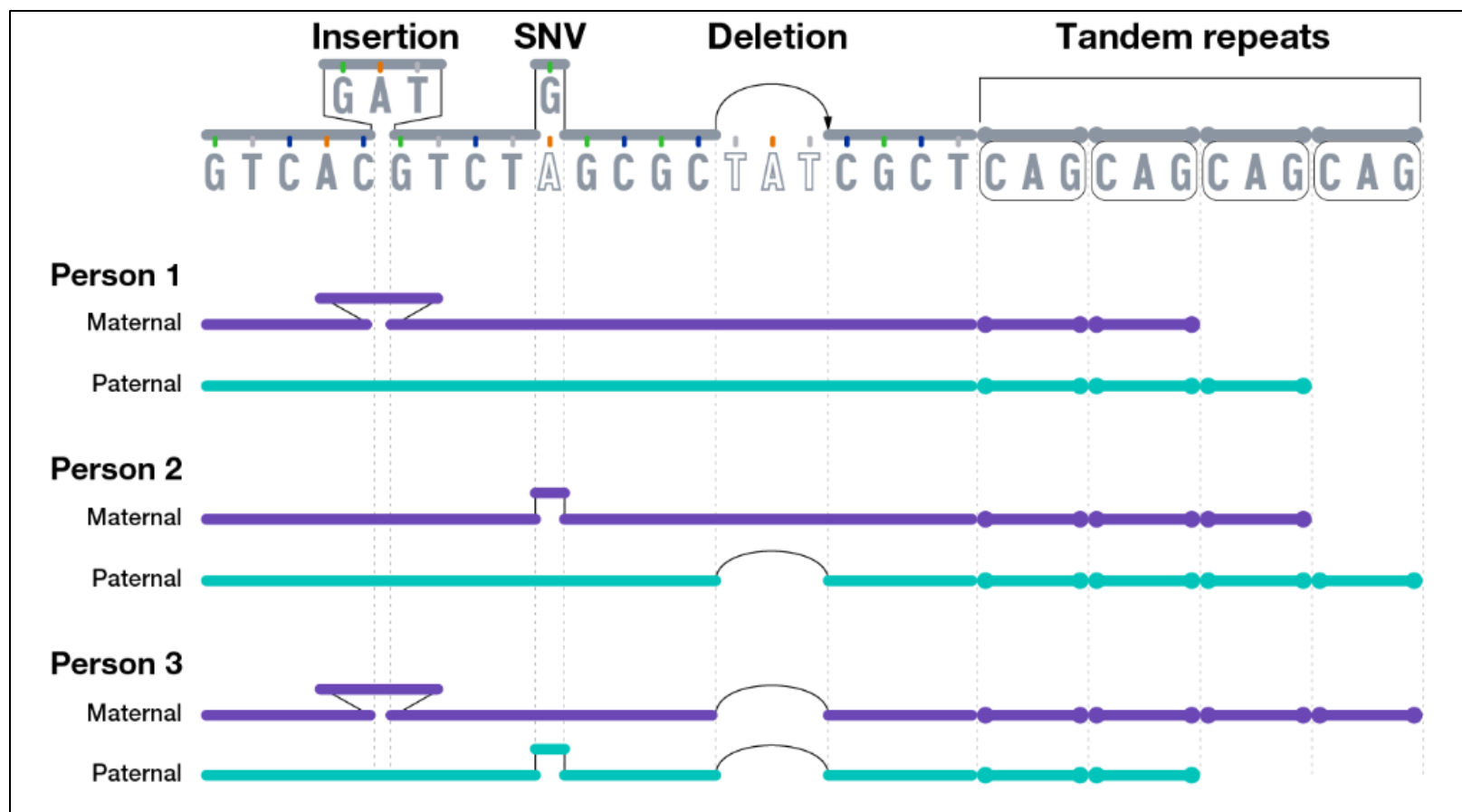
➤ **Synonymous (silent)**

➤ **Nonsynonymous**

- Missense Mutation
- Nonsense Mutation
- Nonstop Mutation

GFTSTDTL

Genomic Concepts - Genomic Variations (INDELS)



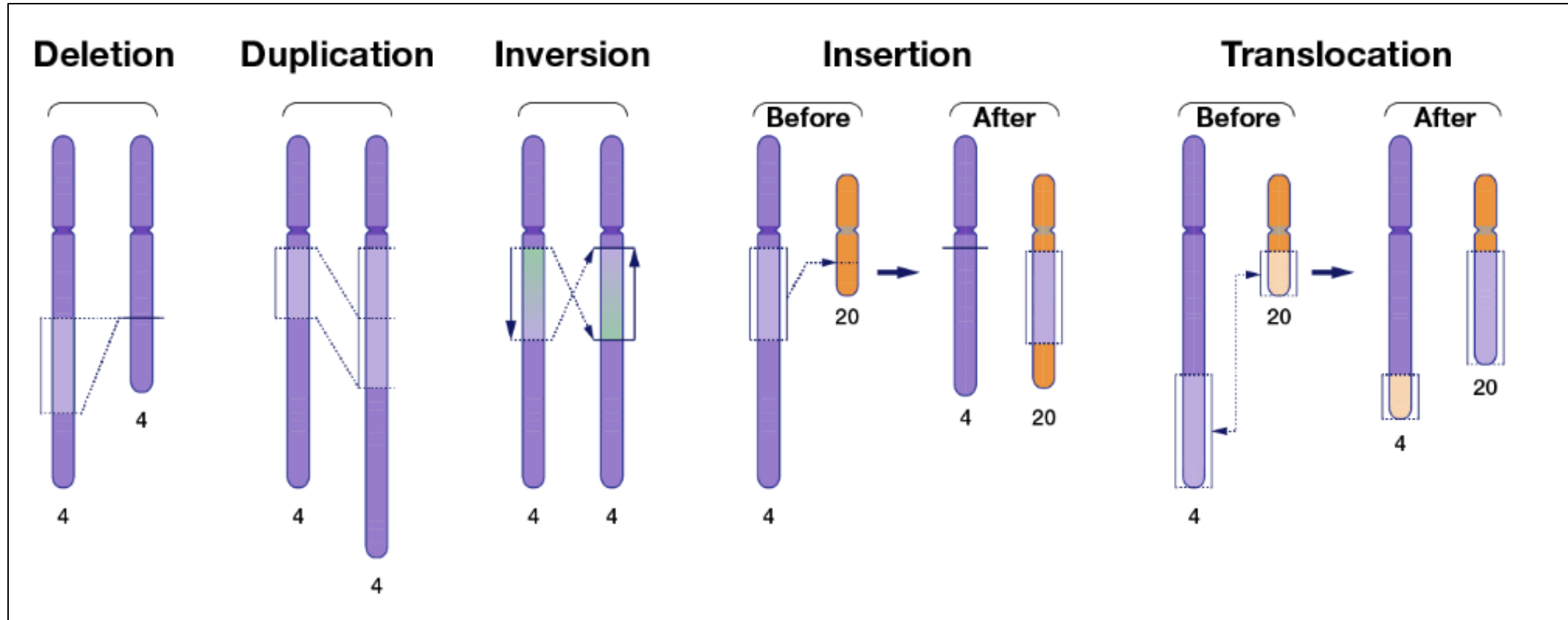
❖ INDELS

- Insertions
- Deletions
- Tandem repeats
(Microsatellites)

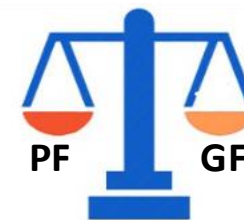
GFTSTDTL



Genomic Concepts - Structural Variations



GFTEST/GFTSTDTL



Genomic Concepts – Reference Sequence

- ❖ Reference Genome (Reference Assembly)
 - **GFGENREF** - Genome Reference (**GRCh38.p13**)
- ❖ Browser for reference genome – Ensembl or UCSC
 - **GFSEQID** - Sequence Identifier (**ENSG00000182533**)
- ❖ Variant characterized in an external database
 - **GFPVRID** - Published Variant Identifier (**rs2231142**)

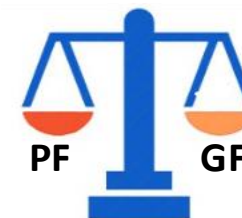


Controlled Terminology

Variable Name	Variable Label	Controlled Terms
GFTESTCD	Short Name of Genomic Measurement	(GFTESTCD)
GFTEST	Name of Genomic Measurement	(GFTEST)
GFTSTDTL	Measurement, Test, or Examination Detail	(GFTSDDL)
GFINHERT	Inheritability	(INHERTGF)
GFSYM	Genomic Symbol	HGNC database
GFSYMTYP	Genomic Symbol Type	(SYMTYPGF)
GFANMETH	Analysis Method	(GFANMET)

CDISC-CT

HGNC (HUGO Gene Nomenclature Committee)



Controlled Terminology - HGNC

Report **HCOP homology predictions**

HGNC data for ABCG2

Approved symbol ? ABCG2 **GFSYM**

Approved name ? ATP binding cassette subfamily G member 2 (JR blood group)

Locus type ? gene with protein product **GFSYMTYP**

HGNC ID ? HGNC:74

Symbol status ? Approved

Previous names ? "ATP-binding cassette, sub-family G (WHITE), member 2 "
"ATP-binding cassette, sub-family G (WHITE), member 2 (Junior blood group) "
"ATP binding cassette subfamily G member 2 (Junior blood group) "

Alias symbols ? EST157481; MXR; BCRP; ABCP; CD338

Chromosomal location ? 4q22.1 **GFGENLOC**

Gene groups ? ATP binding cassette subfamily G
CD molecules
Blood group antigens

Gene symbol report :
HUGO Gene
Nomenclature
Committee (HGNC)

Report **HCOP homology predictions**

Human

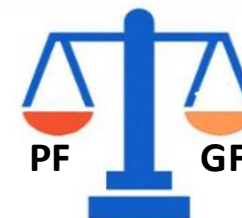
Approved symbol **ABCG2** ⓘ

Approved name ATP binding cassette subfamily G member 2 (Junior blood group) ⓘ

Locus type gene with protein product ⓘ

Chromosomal location 4q22.1

Gene resources ENSG00000118777 9429 **GFSEQID**



Controlled Terminology

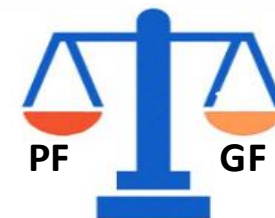
CDISC-CT - GFTESTCD, GFTEST, GFSTDTL, GFINHRT, GFSYMTYP

HGNC (HUGO Gene Nomenclature Committee) - GFSYM

EXTERNAL - GFGENREF, GFSEQID, GFPRVID

STUDYID	DOMAIN	USUBJID	GFSEQ	GFREFID	GFTESTCD	GFTEST	GFSTDTL	GFORRES	GFORREF	GFSTRESC	GFSTREFC
ABC-006	GF	ABC-006-00003	1	NA18537	SNV	Single Nucleotide Variation	GENOTYPE	T/T	G/G	T/T	G/G
ABC-006	GF	ABC-006-00004	2	NA07000	SNV	Single Nucleotide Variation	GENOTYPE	G/T	G/G	G/T	G/G
ABC-006	GF	ABC-006-00005	3	NA00131	SNV	Single Nucleotide Variation	GENOTYPE	G/G	G/G	G/G	G/G

GFGENREF	GFCHROM	GFSYM	GFGENLOC	GFSEQID	GFPRVID	GFNAM	GFSPEC	GFMETHOD	VISITNUM	VISIT	GFDTCT	GFDTY
GRCh38.p13	4	ABCG2	4:88131171	ENSG00000118777	rs2231142	ACME LABS	DNA	REAL-TIME POLYMERASE CHAIN REACTION	1	SCREENING	2020-06-25	-3
GRCh38.p13	4	ABCG2	4:88131171	ENSG00000118777	rs2231142	ACME LABS	DNA	REAL-TIME POLYMERASE CHAIN REACTION	1	SCREENING	2020-06-25	-3
GRCh38.p13	4	ABCG2	4:88131171	ENSG00000118777	rs2231142	ACME LABS	DNA	REAL-TIME POLYMERASE CHAIN REACTION	1	SCREENING	2020-06-25	-3



Mapping Changes - Variables added/replaced/retired

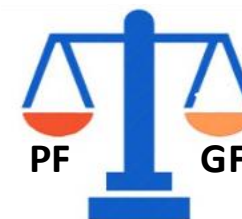
Variable Name	Variable Label	Changes
GFLNKGRP	Link Group ID	Added
NHOID	Non-Host Organism Identifier	Non-host Species (PFNPSCES) and Non-host Strain (PFNSTRN) have been replaced
GFSTREFC	Reference Result in Standard Format	Added
GFSTREFN	Numeric Reference Result in Std Units	Added
GFINHERT	Inheritability	PFMUTYP is replaced
GFGENREF	Genome Reference	Added



Mapping Changes - Variables added/replaced/retired

Variable Name	Variable Label	Changes
GFCHROM	Chromosome Identifier	Added
GFSYM	Genomic Symbol	PFGENRI is replaced
GFSYMTYP	Genomic Symbol Type	PFGENTYP is replaced
GFSEQID	Sequence Identifier	PFREFSEQ is replaced
GFPVRID	Published Variant Identifier	PFRSNUM is replaced
GFCOPYID	Copy Identifier	PFALLELC is replaced

PFGENLI and **PFGENTRG** variables were retired.



Mapping changes - Genetic Variation - Variables

STUDYID	DOMAIN	USUBJID	GFSEQ	GFREFID	GFTESTCD	GFTEST	GFTSTDTL
ABC-006	GF	ABC-006-00003	1	NA18537	SNV	Single Nucleotide Variation	GENOTYPE
ABC-006	GF	ABC-006-00004	2	NA07000	SNV	Single Nucleotide Variation	GENOTYPE
ABC-006	GF	ABC-006-00005	3	NA00131	SNV	Single Nucleotide Variation	GENOTYPE

GFORRES	GFORREF	GFSTRESC	GFSTREFC	GFSTREFN
T/T	G/G	T/T	G/G	
G/T	G/G	G/T	G/G	
G/G	G/G	G/G	G/G	

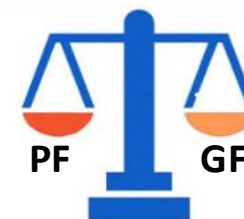
GFORRES - Result or Finding in Original Units

GFORREF - Reference Result in Original Units

GFSTRESC - Result or Finding in Standard Format

GFSTREFC - Reference Result in Standard Format

GFSTREFN - Numeric Reference Result in Std Units



Mapping changes - Genetic Variation - Variables

GFGENREF	GFCHROM	GFSYM	GFSYMTYP	GFGENLOC	GFSEQID	GFPVRID	GFMETHOD
GRCh38.p13	4	ABCG2	GENE WITH PROTEIN PRODUCT	4:88131171	ENSG00000118777	rs2231142	REAL-TIME POLYMERASE CHAIN REACTION
GRCh38.p13	4	ABCG2	GENE WITH PROTEIN PRODUCT	4:88131171	ENSG00000118777	rs2231142	REAL-TIME POLYMERASE CHAIN REACTION
GRCh38.p13	4	ABCG2	GENE WITH PROTEIN PRODUCT	4:88131171	ENSG00000118777	rs2231142	REAL-TIME POLYMERASE CHAIN REACTION

- **GFCHROM** - Chromosome Identifier
- **GFSYM** - Genomic Symbol
- **GFSYMTYP** - Genomic Symbol Type
- **GFGENLOC** - Genetic Location

- **GFGENREF** - Genome Reference
- **GFSEQID** - Sequence Identifier
- **GFPVRID** - Published Variant Identifier



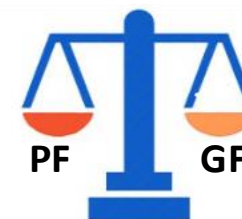
Mapping changes - Genetic Variation - PF to GF

STUDYID	DOMAIN	USUBJID	PFTESTCD	PFTEST	PFTSTDTL	PFGENRI	PFGENTYP	PFCAT
ABC-006	PF	ABC-006-00003	GENVAR	Genetic Variation	INSERTION	CALR	GENE	GENOMIC ABERRATION TESTING
ABC-006	PF	ABC-006-00003	GENVAR	Genetic Variation		JAK2 V617F	GENE	GENOMIC ABERRATION TESTING
ABC-006	PF	ABC-006-00003	GENVAR	Genetic Variation	DELETION	KDM6A	GENE	GENOMIC ABERRATION TESTING
ABC-006	PF	ABC-006-00003	GENVAR	Genetic Variation	SINGLE NUCLEOTIDE VARIANT	KIT	GENE	GENOMIC ABERRATION TESTING

STUDYID	DOMAIN	USUBJID	GFTESTCD	GFTEST	GFTSTDTL	GFSYM	GFSYMTYP	GFCAT
ABC-006	GF	ABC-006-00003	GENVAR	Genetic Variation	INSERTION	CALR	GENE WITH PROTEIN PRODUCT	GENOMIC ABERRATION TESTING
ABC-006	GF	ABC-006-00003	GENVAR	Genetic Variation		JAK2 V617F	GENE WITH PROTEIN PRODUCT	GENOMIC ABERRATION TESTING
ABC-006	GF	ABC-006-00003	GENVAR	Genetic Variation	DELETION	KDM6A	GENE WITH PROTEIN PRODUCT	GENOMIC ABERRATION TESTING
ABC-006	GF	ABC-006-00003	GENVAR	Genetic Variation	SINGLE NUCLEOTIDE VARIANT	KIT	GENE WITH PROTEIN PRODUCT	GENOMIC ABERRATION TESTING

PFGENRI to GFSYM

PFGENTYP to GFSYMTYP



Mapping changes - Genetic Variation - PF to GF

SUPPPF.QVAL

STUDYID	RDOMAIN	USUBJID	QNAM	QLABEL	QVAL	QORIG
ABC-006	PF	ABC-006-00003	PFGENRIO	Genetic Region of Interest Other	ChrX:44732820	CRF
ABC-006	PF	ABC-006-00003	PFGENRIO	Genetic Region of Interest Other	Chr4:55593395	CRF
ABC-006	PF	ABC-006-00003	PFGENRIO	Genetic Region of Interest Other	Chr19:13054627	CRF

GF variables (GFCHROM, GFGENLOC)

GFORRES	GFCHROM	GFGENLOC	GFSTRESC	GFSPEC	GFMETHOD	GFDTCT
DETECTED	19	Chr19:13054627	DETECTED	DNA	NEXT GENERATION SEQUENCING	2021-08-05
NOT DETECTED			NOT DETECTED	DNA	NEXT GENERATION SEQUENCING	2021-08-05
DETECTED	X	ChrX:44732820	DETECTED	DNA	NEXT GENERATION SEQUENCING	2021-08-05
DETECTED	4	Chr4:55593395	DETECTED	DNA	NEXT GENERATION SEQUENCING	2021-08-05



Summary

- **Data Collection and Standardization**
- **Integration and Analysis**
- **Improved Data Quality**
- **Enhanced Research Capabilities**
- **Regulatory Compliance**



Recommended Reading

1. CDISC SDTM PGxIG V1.0
2. CDISC SDTMIG v3.4



Questions



Contact Information:

Rama Kudaravalli

Merck & Co., Inc., Rahway, NJ, USA

rama.kudaravalli@merck.com