Paper PP15

Genomics Findings (GF)

Ruba Sivasamy, Sensan Biosciences, Chennai, India

ABSTRACT

A findings domain that contains data related to the structure, function, evolution, mapping, and editing of subject and non-host organism genomic material of interest. This domain includes but is not limited to assessments and results for genetic variation and transcription, and summary measures derived from these assessments. The GF domain is used for findings from characteristics assessed from nucleic acids and may include subsequent inferences and/or predictions about related proteins/amino acids.

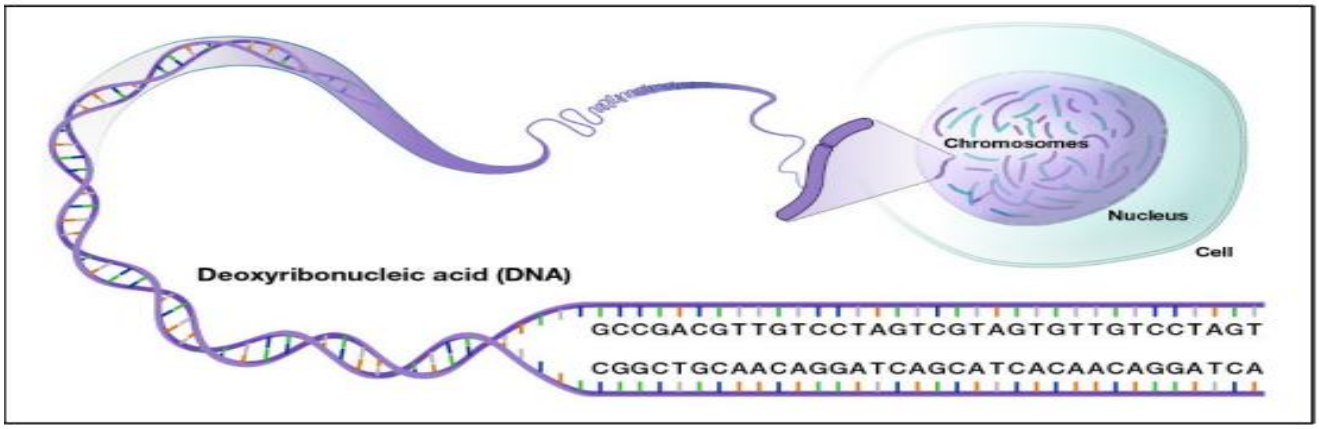
Genomics Findings (GF) domain is introduced in (SDTMIG) v3.4 which replaces PF domain from the previous version.

If we are using SDTMIG v3.2 or v3.3, GF domain can be included as a custom domain, which would be a promising way to handle genetic testing and assessment results.

This paper covered a small intro about Genome and list of some important variables, some GFTESTCD values from GF domain

Introduction to Genome

Genome is complete set of DNA instructions found in every cell. Genome's DNA is packed into structures called chromosomes. DNA is made up of four different chemicals (called nucleotides or bases) represented by : adenine (A), thymine (T), cytosine (C) and guanine (G).
Gene is considered as the basic unit of inheritance and passed from parents to offspring and contains the information needed to specify physical and biological traits.



Standardized variables for Genomic data:

- GFTESTCD / GFTEST : NAME OF GENOMIC TEST
- GFTSTDTL : DETAILS OF THE TEST
- GFSYM : GENOMIC SYMBOL (FROM HGNC)
- GFINHERT : INHERITABILITY OF THE VARIATION
- GFCHROM : CHROMOSOME IDENTIFIER
- GENLOC : GENETIC LOCATION
- GFGENSR : GENETIC SUB_REGION
- GFGENREF: GENOME REFERENCE USED
- GFMETHOD : METHOD OF TEST OR EXAMINATION
- GFSEQID: SEQUENCE IDENTIFIER
- GFPRVID: PUBLISHED VARIANT IDENTIFIER

GFTESTCD AND GFTEST:

GFTESTCD and GFTEST should not include genomic names or symbols, including but not limited to official gene symbols. It's a high level or general description of the assessment. 'VAR' or 'V' will be used as the suffix fragment in GFTESTCD to denote 'Variation' in the GFTEST.

Copy Number Amplification	Copy Number Amplification; Gene Copy Number Amplification
Copy Number Variation	Copy Number Variation; Gene Copy Number Variation
Gene Signature	Gene Expression Signature; Gene Signature
Microsatellite Instability	Microsatellite Instability
Sequence Rearrangement	Sequence Rearrangement
Short Variation	Short Variation
Single Nucleotide Variation	Single Nucleotide Variation
Transcription	Transcription
Tumor Mutation Burden	Tumor Mutation Burden; Tumor Mutational Burden
Variable Number Tandem Repeats	Tandem Repeat Variation; Variable Nucleotide Tandem Repeats; Variable Number Tandem Repeats
Variant Profile	Variant Profile

Single nucleotide variation:

A Single Nucleotide Variation (SNV) is a change in one nucleotide (A, T, C, or G) at a specific position in the DNA sequence. It is the most basic form of genetic variation and can be harmless, beneficial, or disease-causing depending on where it occurs.

Example	GFTESTCD	GFTEST	GFTSTDTL	GFORRES	GFSYM	GFSYMTYP	GFINHERT
Germline *BRCA1* mutation	SNV	Single Nucleotide Variation	DETECTION	DETECTED	BRCA1	GENE WITH PROTEIN PRODUCT	**GERMLINE** VARIATION
Somatic *BRCA1* mutation	SNV	Single Nucleotide Variation	DETECTION	DETECTED	BRCA1	GENE WITH PROTEIN PRODUCT	**SOMATIC** VARIATION
Germline *BRCA2* mutation	SNV	Single Nucleotide Variation	DETECTION	DETECTED	BRCA2	GENE WITH PROTEIN PRODUCT	**GERMLINE** VARIATION
Somatic *BRCA2* mutation	SNV	Single Nucleotide Variation	DETECTION	DETECTED	BRCA2	GENE WITH PROTEIN PRODUCT	**SOMATIC** VARIATION

Example: A change from A to G in the BRCA1 gene might increase cancer risk

Copy number variation:

Copy Number Variation (CNV) is a type of genetic variation where sections of DNA are duplicated or deleted, resulting in individuals having a different number of copies of a particular DNA segment compared to the reference genome.

GFTESTCD	GFTEST	GFTSTDTL	GFORRES	GFSYM	GFSYMTYP
CPNUMVAR	Copy Number Variation	COPY NUMBER ALTERATION INTERPRETATION	AMPLIFICATION	ERBB2	GENE WITH PROTEIN PRODUCT

Impact: Alters gene dosage, potentially leading to disease or developmental disorders.

Example: A deletion in chromosome 22 can cause DiGeorge syndrome.

GFSTDTL

It represents specific reportable form for the assessment described in the GFTESTCD/GFTEST value.

It would be recommended to have GFTSTDTL as GFTESTCD may have multiple explanations according to the perspective of genomic testing needs.

Values of GFSTDTL can be 'NORMALIZED', 'INTERPRETATION', 'PREDICTED', 'NUMBER' and 'STATUS'.

GFSYM and GFSYMTYP:

A published symbol for the portion of the genome serving as a locus for the experiment/test.

For human genetic data, standard nomenclature populated in variable GFSYM must be obtained from the genomic symbol list maintained in the Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC) database (https://www.genenames.org/). CDISC will not control this variable with CDISC Controlled Terminology. GFSYMTYP is controlled by codelist SYMTYPGF

GFSYMTYP values as per CT:

CIRCULAR RNA
ENDOGENOUS RETROVIRUS
FRAGILE SITE
GENE WITH PROTEIN PRODUCT
LONG NON-CODING RNA
MICRO RNA
NONSENSE MEDIATED DECAY
PIWI-INTERACTING RNA

GFINHERIT:
GFINHERT identifies whether the variation can be passed to the next generation.

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C181177		Yes	Genomic Inheritability Type Response	INHERTGF	Genomic Inheritability Type Response	Terminology relevant to whether the variation can be passed to the next generation.	CDISC SDTM Genomic Inheritability Type Response Terminology
C17666	C181177		Genomic Inheritability Type Response	GERMLINE VARIATION		Any inheritable variation in the DNA that is transmitted to the progeny with some frequency.	Germline Variation
C181352	C181177		Genomic Inheritability Type Response	MITOCHONDRIAL VARIATION		Any inheritable variation that appears in mitochondrial DNA that is transmitted to the progeny with some frequency.	Mitochondrial DNA Variation
C18060	C181177		Genomic Inheritability Type Response	SOMATIC VARIATION		Any non-inheritable variation in the DNA that is not transmitted to the progeny but will be maintained during any mitosis within the individual.	Somatic Variation

GFINHERIT is controlled by codelist INHERTGF.

GERMLINE VARIATION:
Germline variation refers to genetic differences that are present in the DNA of reproductive cells (sperm or egg) and therefore can be inherited by offspring. These variations are passed down through generations and are present in every cell of the individual's body.

MITOCHONDRIAL VARIATION:
Mitochondrial variation refers to genetic differences that occur in the DNA of mitochondria — the tiny energy-producing organelles inside cells. Unlike nuclear DNA, mitochondrial DNA (mtDNA) is inherited almost exclusively from the mother.

SOMATIC VARIATION:
A Somatic variation is a mutation or alteration in DNA that arises after conception in body cells (e.g., skin, liver, lung, brain). It can happen due to errors in DNA replication, environmental exposures (radiation, chemicals), or random cellular processes. Unlike germline mutation it cannot be passed to offspring because it does not occur in sperm or egg cells. Somatic mutations are a major driver of cancer. For example, mutations in **TP53** or **KRAS** genes can lead to uncontrolled cell growth.

GFCHROM:
It records the chromosome identifier where the genomic finding (such as a variation, mutation, or structural change) is located.
(e.g., "17"; "X").

GFGENLOC:
It specifies the precise genomic region or locus where the finding (such as a mutation, variation, or structural change) is observed.

GFTESTCD	GFTEST	GFCHROM	GFGENLOC	GFLOC	GFSTART	GFEND
SNV	Single Nucleotide Variation	17	17q21.31 region	BRCA1	43044295	43125482

Here:

- **GFCHROM** = Chromosome 17
- **GFGENLOC** = Cytogenetic band 17q21.31 (broader genomic location)
- **GFLOC** = BRCA1 gene (specific locus)

GFGENSR:
It specifies the portion of the locus in which the variation was found. It is used when you need to be more specific than just the gene or locus — it drills down into the functional sub-region (like an exon, intron, or protein domain) where the variation occurs.

GFGENREF:
An identifier for the genome reference used to generate the reported result. It ensures clarity and reproducibility by stating which reference sequence was used for mapping the variation.
For example, Genome Reference Consortium Human Build 38 patch release 13 may be represented as "GRCh38.p13".

GFMETHOD:
The test method by which the examination is performed by the wet lab in order to yield the result reported in the dataset. GFMETHOD documents how the genomic finding was detected — the sequencing or analysis method. This is critical for reproducibility, interpretation, and regulatory submissions, since different methods have different sensitivity and accuracy.
Examples of values:

- *Next-Generation Sequencing (NGS)*
- *Whole Genome Sequencing (WGS)*
- *Whole Exome Sequencing (WES)*
- *Sanger sequencing*
- *Array Comparative Genomic Hybridization (array CGH)*
- *PCR-based methods*
- *RNA-seq*

Outputs:

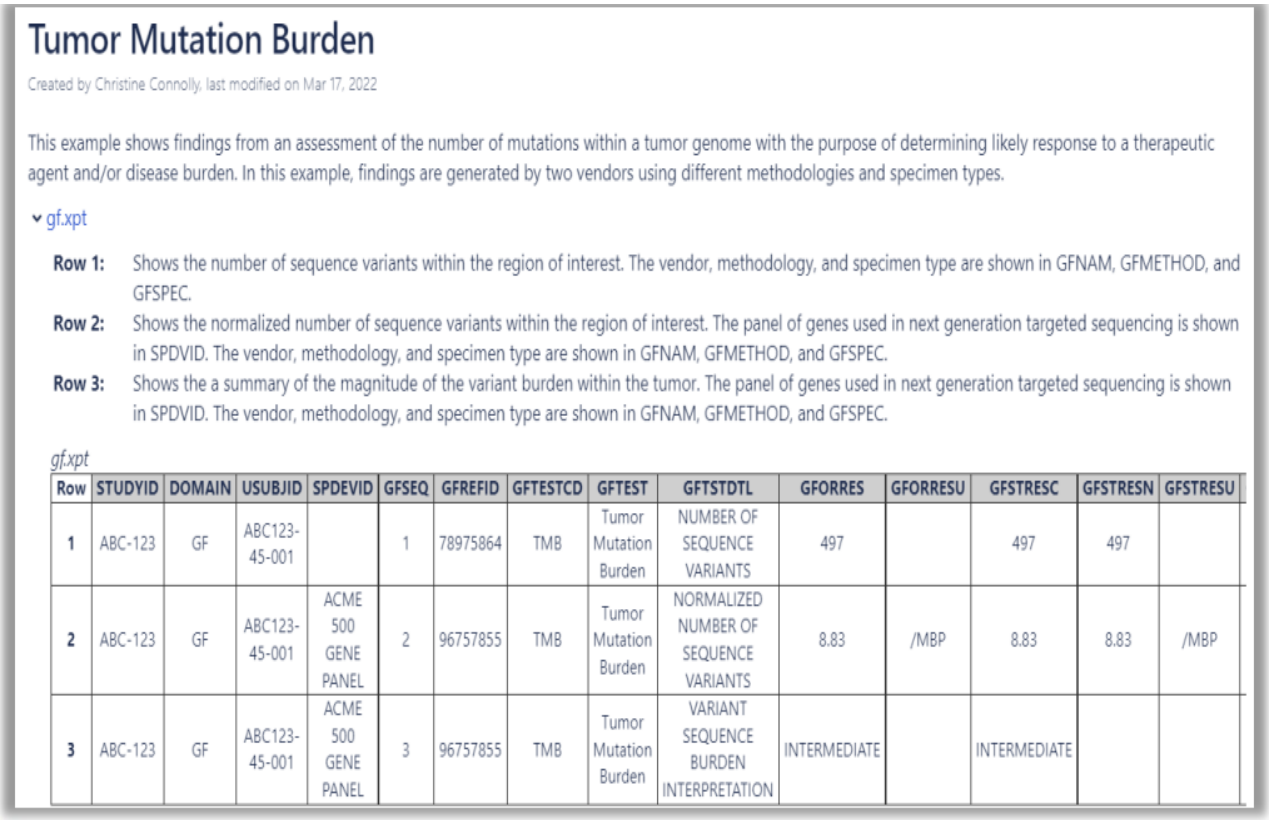
1. Genomic DNA sample obtained from blood is used to detect the SNVs in ABCG2 gene by doing real-time PCR method for three subjects.

STUDYID	DOMAIN	USUBJID	GFSEQ	GFREFID	GFTESTCD	GFTEST	GFTSTDTL	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	GFPVRID	GFNAM	GFSPEC	GFMETHOD	VISITNUM	VISIT	GFDTC	GFDY
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

Each subject has a different genotype at the genetic locus of the interest. The inheritability of the variation is considered to be germline as the sample is DNA. The nucleotide variant is located on chromosome number 4, gene location is provided in GFGENLOC.

2.Tumor Mutation Burden



Tumor Mutation Burden

Created by Christine Connolly, last modified on Mar 17, 2022

This example shows findings from an assessment of the number of mutations within a tumor genome with the purpose of determining likely response to a therapeutic agent and/or disease burden. In this example, findings are generated by two vendors using different methodologies and specimen types.

gf.xpt

Row 1: Shows the number of sequence variants within the region of interest. The vendor, methodology, and specimen type are shown in GFNAM, GFMETHOD, and GFSPEC.

Row 2: Shows the normalized number of sequence variants within the region of interest. The panel of genes used in next generation targeted sequencing is shown in SPDVID. The vendor, methodology, and specimen type are shown in GFNAM, GFMETHOD, and GFSPEC.

Row 3: Shows the a summary of the magnitude of the variant burden within the tumor. The panel of genes used in next generation targeted sequencing is shown in SPDVID. The vendor, methodology, and specimen type are shown in GFNAM, GFMETHOD, and GFSPEC.

gf.xpt

Row	STUDYID	DOMAIN	USUBJID	SPDEVID	GFSEQ	GFREFID	GFTESTCD	GFTEST	GFTSTDTL	GFORRES	GFORRESU	GFSTRESC	GFSTRESN	GFSTRESU
1	ABC-123	GF	ABC123-45-001		1	78975864	TMB	Tumor Mutation Burden	NUMBER OF SEQUENCE VARIANTS	497		497	497	
2	ABC-123	GF	ABC123-45-001	ACME 500 GENE PANEL	2	96757855	TMB	Tumor Mutation Burden	NORMALIZED NUMBER OF SEQUENCE VARIANTS	8.83	/MBP	8.83	8.83	/MBP
3	ABC-123	GF	ABC123-45-001	ACME 500 GENE PANEL	3	96757855	TMB	Tumor Mutation Burden	VARIANT SEQUENCE BURDEN INTERPRETATION	INTERMEDIATE		INTERMEDIATE		

CONCLUSION:

- GF domain provides a standardized way to represent genomic data
- Ensuring consistency and accuracy across studies.
- Facilitates integration of genomic data with other clinical data, enhancing the ability to conduct comprehensive analyses.
- Addition of new variables gives flexibility to map additionally collected information to the core domain and makes it convenient for analysis.
- By using controlled terminology and predefined structures, the GF domain helps improve the quality and reliability of genomic data.
- It aligns with regulatory requirements, making it easier for researchers to submit genomic data to regulatory agencies.
- Overall, the future of the GF domain in SDTM is geared towards enhancing the quality and usability of genomic data in clinical research, ultimately contributing to better patient outcomes and advancements in personalized medicine.

CONTACT INFORMATION:

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Ruba
Company: Sensan Biosciences
Email: ruba.sivasamy@sensanbio.com
WEBSITE: WWW.SENSANBIO.COM