Genomics Findings (GF)

By Ruba Sivasamy

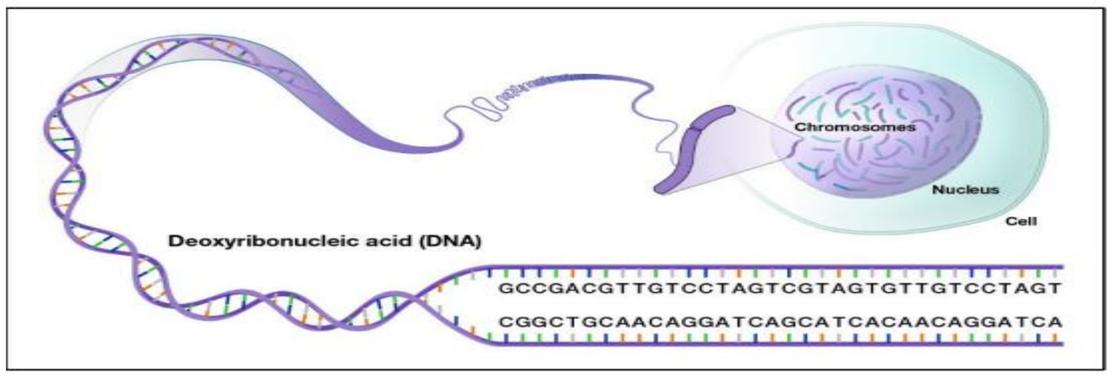
SENSAN BIOSCIENCES

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 (refer SDTM IG 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.

Introduction to Genome

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



Standardized variables for genomic data:

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:

high level or generalized description of the assessment
should not contain gene names or symbols
'VAR' or 'V' will be used as the suffix fragment in GFTESTCD to denote 'Variation' in the GFTEST.
GFTSTDTL →'NORMALIZED', 'INTERPRETATION', 'PREDICTED', 'NUMBER' and 'STATUS'.

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

Refer NCI EVS FTP Portal (Government site) for TEST and TESTCD CT values
https://evs.nci.nih.gov/ftp1/CDISC/SDTM/

Examples

Single nucleotide variation:

Single nucleotide substitutions and small deletions or insertions (1 – 20 bases) account for most mutations in these genes.
BRCA1/BRCA2 mutations in Breast cancer clinical trials .

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

Impact:

Can affect protein function, gene regulation, or be silent.
Example: A change from A to G in the BRCA1 gene might increase cancer risk.

Copy number variation

CNV is a type of structural variation in the genome where sections of DNA are repeated or missing. HER2 gene (also known as ERBB2 – Erythroblastic Leukemia Viral Oncogene Homolog 2 in HGNC database) amplified in Human breast cancer cell lines.

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.

GFSYM and GFSYMTYP:

Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC) database (www.genenames.org)

Symbol report for EGFR Stable symbol

Report | HCOP homology predictions

HGNC data for EGFR

Approved symbol	EGFR
Approved name	epidermal growth factor receptor
Locus type	gene with protein product
HGNC ID	HGNC:3236 GFSYM = EGFR
Symbol status	Approved
Previous symbols	ERBB GFSYMTYP = GENE WITH PROTEIN PRODUCT
Previous names	" epidermal growth factor receptor (avian erythroblastic leukemia viral (v-erb-b) oncogene homolog) "
Alias symbols	ERBB1; ERRP
Alias names	" erythroblastic leukemia viral (v-erb-b) oncogene homolog (avian) " " erb-b2 receptor tyrosine kinase 1 "
Chromosomal location	7p11.2
Gene groups	Erb-b2 receptor tyrosine kinases

GFINHERT:

Identifies whether the variation can be passed to the next generation

Code	Codelist Code	Codelist Extensible (Yes/N	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

Output

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

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

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.
- 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

Contact the author at:
EMAIL: ruba.sivasamy@sensanbio.com
ORGANIZATION: SENSAN BIOSCIENCES, India.
WEBSITE: WWW.SENSANBIO.COM