PharmaSUG China 2022 - Paper SR-145

Time to Get in the Genomics Findings (GF) Domain

Xuejie Zhou, Liming Xie, Qiuan Hu and Shanshan Ma, BeiGene, Inc.

ABSTRACT

Genomics Findings (GF) domain is first introduced in Study Data Tabulation Model Implementation Guide (SDTMIG) v3.4 (published 2021-11-29), replaces PF domain from the provisional SDTMIG for Pharmacogenomics and Pharmacogenetics (SDTMIG-PGx, originally published 2015-05-26), which has been deprecated. Even if you're 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. However, GF domain as a new domain is hard to understand and implement with limited instructions, since it requires the basic understanding of genetics and more effort to learn how to map into proper variables. In this paper, we would like to share our experience on implementing GF domain and explain how to successfully create GF domain.

USING PF DOMAIN OR GF DOMAIN?

As our understanding of human diseases evolves and as technologies advance, genetic testing (also known as biomarker testing) has become an important detection for disease diagnosis, selecting standard of care, and disease prognosis (National Cancer Institute, 2021). Genetic testing and assessment results have been collected more and more frequently in human clinical trials, especially for the trials targeting in cancers, e.g., epidermal growth factor receptor (EGFR) gene located on human chromosome 7 for non-small cell lung cancer (NSCLC) (Habets GG, 1992; Francesco Passiglia, 2017).

To handle the genetic data collected in clinical trials, in 2015-05-26, CDISC first introduced the provisional version 1.0 of SDTMIG-PGx, which was intended to provide guidance on the implementation of the SDTM for biospecimen and genetics-related data, e.g., biospecimen collection and handling, genetic mutation, viral genetics, etc. SDTMIG-PGx focuses on primarily on genetic variation data for human and viral studies, and gene expression, provides a comprehensive solution for handling genetic data in SDTM. However, SDTMIG-PGx stands on its own like SDTMIG for Medical Devices (SDTMIG-MD), does not have the genetics domains being incorporated into either SDTMIG or SENDING.

In 2021-11-29, CDISC released SDTMIG v3.4 and replaced PF domain with a new domain – Genomics Findings (GF). Comparing with PF domain mainly capturing results for both genetic variation and gene expression, GF domain includes but is not limited to assessments and results of subject and non-host organism genomic material of interest for genetic variation and transcription, and summary measures derived from these assessments. GF domain can be used for findings from characteristics assessed from nucleic acids and may include subsequent inferences and/or predictions about related proteins/amino acids, excluding direct assessments of proteins (e.g., assessments of amino acids).

Since SDTMIG v3.4 has not been included in United States Food and Drug Administration (FDA) Data Standards Catalog v8.0 (February 16, 2022), most of companies in the pharmaceuticals industry are still following SDTMIG v3.2 or v3.3 and haven't switched to SDTM v3.4 yet. In this case, you might be wondering whether PF domain or GF domain should be used for your study. Our suggestion is to get in the GF domain, due to the following reasons:

1. Both PF domain and GF domain are treated as custom domains if you're following SDTMIG v3.2 or v3.3, which would need extra explanation in study data reviewers' guide. Considering SDTMIG v3.4 will be supported by FDA in upcoming years, it's always better to save for a rainy day.
2. GF domain extends the scope of implementation beyond genetic variation and gene expression, which provides more feasibility and flexibility for the implementation.
3. GF domain demonstrates the most recent interpretation from CDISC and includes more detailed instructions on how to implement it, comparing with PF domain, which should be more helpful and comprehensive to be followed.

However, since genetic data is complex with lots of concepts and terms, especially for the users who are

unfamiliar with the genetic concepts, GF domain is hard to understand and implement with limited instructions. We would like to share our experience on implementing GF domain and explain how to successfully create GF domain through:

1. Comprehensive genetics knowledge background introduction to help readers have a whole picture of genetic concepts.
2. Comparison of GF domain with other generic specimen-based lab findings domains to help readers grasp the main differences.
3. Intensive introduction of interpretation genetic testing results in GF domain to help readers better implement the domain and variables.

GENETICS KNOWLEDGE BACKGROUND

Chromosomes are thread-like structures located inside the nucleus of animal and plant cells. Each chromosome is made of protein (histone) and deoxyribonucleic acid (DNA). DNA is a double-stranded superhelix structure composed of deoxynucleotides, which is consists of a deoxyribose, phosphoric acid, and a nitrogenous base. Genes are the segments of DNA that determine our traits. The unique structure of chromosomes keeps DNA tightly wrapped around histones and each DNA contains many genes. The bases for DNA are adenine (A), cytosine (C), guanine (G), and thymine (T). The bases pair up with each other in DNA (A with T, C with G) to form units known as base pairs (Figure 1). The human body has two sets of 23 chromosomes and 20,000~25,000 genes, half of which are inherited from our biological mothers and the other half from our biological fathers. The process, by which the information encoded in a gene is used to either make ribonucleic acid (RNA) molecules that code for proteins or to make non-coding RNA molecules that serve other functions, is call as gene expression, which includes gene transcription and gene translation (Figure 2).

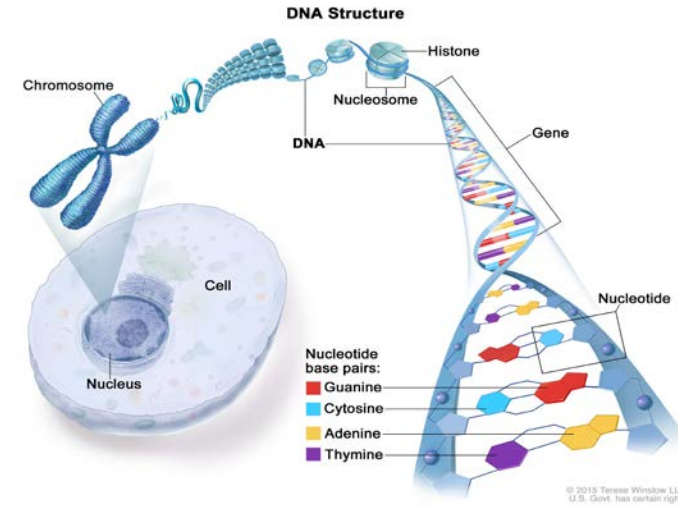


Figure 1. Chromosomes, DNA and Genes

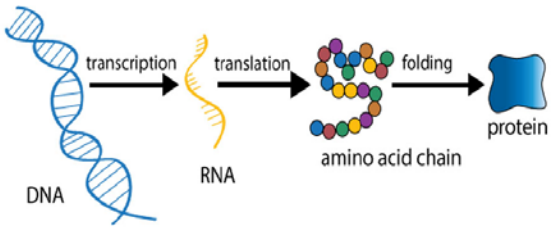


Figure 2. Gene Expression

Genetic variation refers to the differences in the genetic makeup of individuals in a population and it is important to the processes of natural selection and the cause of diseases. Genetic variation that occurs in eggs and sperm can be passed on to offspring, while somatic mutation that occur in body cells are not passed on. The changes in the sequences of genes in DNA and chromosome variation is the important source of genetic variation. According to Human Genome Variation Society (HGVS) recommendations for the description of sequence variants 2016 update (den Dunnen JT, 2016), a letter prefix is mandatory to indicate the type of reference sequence, e.g., “c.” for a coding DNA reference sequence, “g.” for a linear genomic reference sequence, “p.” for a protein reference sequence, “r.” for an RNA reference sequence (transcript) and “chr” for a chromosome.

CHROMOSOMES VARIATION

It includes deficiencies in chromosome structure and changes in chromosome number. Deficiencies in chromosome refer to the structure change of genetic information on a chromosome, such as deletions, duplications, rearranging the order of genes on a chromosome (inversions), or moving a gene from one chromosome to another chromosome (translocations). Variations in chromosome number refers to the changes in the total number of chromosomes sets or chromosomes within a set. Number changes or deficiencies in chromosome may result in birth defects, mental retardation and increased risk for infertility or pregnancy loss. For example, a rare genetic disease – Cri-du-chat syndrome, also known as 5p- (5p minus) syndrome or cat cry syndrome, is a genetic condition that is caused by the deletion of genetic material on the small arm (the p arm) of chromosome 5 during meiosis (Genetic and Rare Diseases Information Center, 2021).

GENE VARIATION

Gene variation is a permanent change in the gene sequence in DNA. This type of genetic change used to be known as a gene mutation, but because changes in DNA do not always cause disease, it is thought that gene variation is a more accurate term. Gene variation can occur at a single base position in DNA, such as single nucleotide variant (SNV) or single nucleotide polymorphisms (SNP), and it can also involve a larger-scale variation with a stretch of DNA of hundreds or even thousands of base pairs, such as copy number variation (CNV).

Single Nucleotide Variation (SNV) or Single Nucleotide Polymorphisms (SNP)

SNV is the variation of a single nucleotide that occurs at a specific genomic position, and it can be a substitution (e.g., g.76>T), deletion (e.g., g.76delA), duplication (e.g., g.76dupA), inversion (e.g., g.76_83inv) or insertion (e.g., g.76_77insG) in a single-base nucleotide of a DNA sequence (Table 1). Like SNV, single-nucleotide polymorphism (SNP) is also a single nucleotide, but it must be present in at least 1% of the population. SNV or SNP is the most common type of genetic variants (The Human Genome Structural Variation Working Group, 2007).

Table 1. Nomenclature Definitions with Example Variant Descriptions

Variant Types	Example	Remarks
Substitution (>)	g.1318G>T	A change where one nucleotide is replaced by one other nucleotide
Deletion (del)	g.3661_3706del	A change where one or more nucleotides are not present (deleted)
Inversion (inv)	g.495_499inv	A change where more than one nucleotide replaces the original sequence and is the reverse-complement of the original sequence (e.g., CTCGA to TCGAG)
Duplication (dup)	g.3661_3706dup	A change where a copy of one or more nucleotides are inserted directly 3’ of the original copy of that sequence
Insertion (ins)	g.7339_7340insTAGG	A change where one or more nucleotides are inserted in a sequence and where the insertion is not a copy of a sequence immediately 5’
Conversion (con)	g.333_590con1844_2101	A specific type of deletion-insertion where a range of nucleotides replacing the original sequence are a copy of a sequence from another site in the genome

Deletion-insertion (delins/indel)	g.112_117delinsTG	A change where one or more nucleotides are replaced by one or more other nucleotides, and which is not a substitution, inversion, or conversion

Copy Number Variation (CNV)

CNV refers to a circumstance in which the number of copies of a specific segment of DNA varies among different individuals' genomes. It is submicroscopic gains or losses of chromosomal material and often understood as structural variations in the genome larger than 1 kilobase (kb) in size. CNV plays important roles both in human disease and drug response.

DISTINGUISH GF DOMAIN VS. OTHER SPECIMEN-BASED FINDINGS DOMAINS

Individual domains (e.g., GF, IS, LB, MB) for laboratory measurements, tests, or examinations performed on collected biological specimens (e.g., blood, urine, tumor tissue) are grouped together in Specimen-based Findings Domains section of SDTMIG v3.4. To distinguish GF domain and other Specimen-based Findings Domains, and to find the most appropriate domain for laboratory findings, this part represents summary introduction for mentioned domains in Figure 3 and brings comparison among Specimen-based Findings Domains.

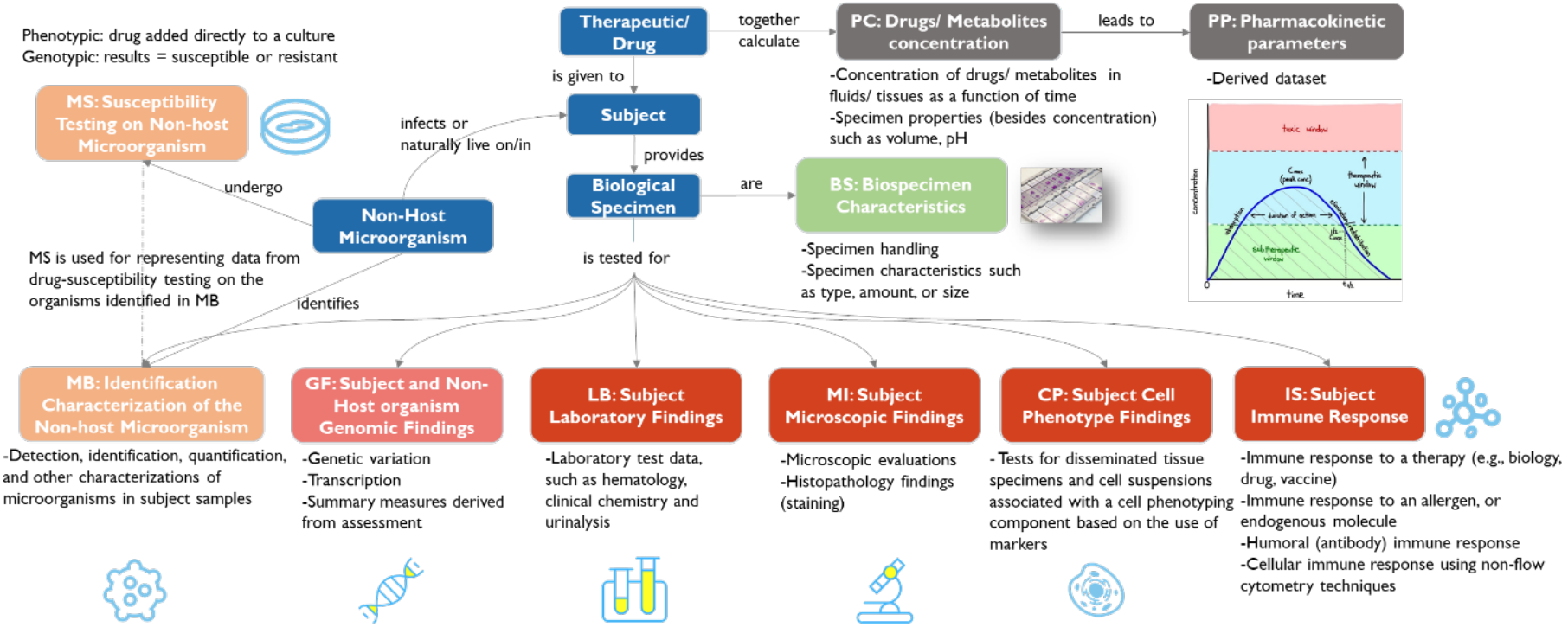


Figure 3. Specimen-based Findings Domains Summary

GF domain focuses on genomic findings from subject or non-host organism, such as assessments and results for gene variation, chromosomes variation, gene rearrangement.

GF vs PC, PP

Pharmacokinetics Domains (PC domain and PP domain) are used for drug and metabolite concentration measurements, and specimen properties (e.g., volume, pH). Compared with GF domain represented genomic information, PC domain and PP domain focus on the study of the time course of drug absorption, distribution, metabolism, and excretion.

GF vs BS

Compared with genomic findings in GF domain, BS domain is used to collect details regarding to specimen handling, the characteristics of biospecimen and extracted samples (e.g., RNA, DNA) such as specimen volume, amount, size, or integrity of then DNA or RNA samples.

GF vs MB, MS

Microbiology Domains (MB domain and MS domain) are used for characterization findings of non-host organisms, including bacteria, viruses, fungi, protozoa, and parasites. MB domain represents findings of detection, identification, quantification, and other characterization of microorganisms. MS domain represents findings of drug susceptibility testing on the microorganisms identified in MB domain. MB, MS, and GF domains are very closely related. For example, the subject is diagnosed with one kind of virus, identification information can be illustrated in MB domain. Then a series of identified virus samples are extracted from the subject and undergo the genotypic susceptibility testing, findings of susceptibility/ resistance data are represented in MS domain, genetic findings such as genetic variation, specific sequence information, are represented in GF domain.

GF vs LB, MI, CP, IS

Compared with genomics findings from nucleic acids, or subsequent inferences and/ or predictions about related proteins/ amino acids in GF domain, other domains tested from biology specimen (LB domain, MI domain, CP domain and IS domain) represent findings about laboratory examination of fluid specimens, microscopic examination of tissue samples, tests of disseminated tissue specimens and cell suspensions based on the use of markers, immune response induced by antigen respectively. Laboratory examination data in LB domain include urinalysis, hematology, chemistry, and coagulation, such as findings of 24-hour urine protein test. Microscopic examination data in MI domain include histopathology findings and microscopic evaluations, such as the assessment of HER2 receptor protein amount using Immunohistochemistry method with stain. Based on expression of specific markers, test data modeled in CP domain is for characteristics (cell phenotype, lineage, and function) of disseminated tissue specimens like blood and bone marrow aspirates, and cell suspensions, currently not for evaluation of solid tissue specimens. Since test in CP domain is associate with marker-based phenotyping, it cannot replace LB domain for routine lab hematology, and MI domain for microscopic assessment of cells. Immune response in IS domain including humoral and cell-mediated immune response may be induced by a therapy, allergen, microorganism, or endogenous molecule, for example, the findings of the screening, the confirmation, and the measurement of titer from antidrug antibody (ADA) evaluation should be recorded in IS domain.

HOW TO INTERPRET GENETIC TESTING RESULTS IN GF DOMAIN?

SDTMIG v3.4 provides guidance of findings in genetic variation, transcription and summary measure mapped to GF domain, and gives the assumptions and specification of GF domain to help users implement it. Such as, GFTEST and GFTESTCD used to describe and distinguish assessment (e.g., gene variation, chromosomes variation, transcription), GFSYM and GFCHROM used to qualify genome and chromosome respectively. In this section, we summarize the usage of following key variables in GF domain based on SDTMIG v3.4 and present our understanding through several examples:

Variable Name	Variable Label	CDISC Codelist	CDISC Notes
GFTESTCD	Short Name of Genomic Measurement	(GFTESTCD)	Short name of the measurement, test, or examination described in GFTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in GFTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g., "1TEST" is not valid). GFTESTCD cannot contain characters other than letters, numbers, or underscores.
GFTEST	Name of Genomic Measurement	(GFTEST)	Long name for GFTESTCD. The value in GFTEST cannot be longer than 40 characters.
GFTSTDTL	Measurement, Test, or Examination Detail	(GFTSDTL)	Description of a reportable qualifying the assessment in GFTESTCD and GFTEST.
GFSYM	Genomic Symbol		A published symbol for the portion of the genome serving as a locus for the experiment/test.
GFSYMTYP	Genomic Symbol Type	(SYMTYPGF)	A description of the type of genomic entity that is represented by the published symbol in GFSYM.
GFINHERT	Inheritability		Identifies whether the variation can be passed to the next generation.
GFCHROM	Chromosome Identifier		The designation (name or number) of the chromosome or contig on which the variant or other feature appears (e.g., "17"; "X").
GFGENLOC	Genetic Location		Specifies the location within a sequence for the observed value in GFORRES.
GFGENSR	Genetic Sub-Region		The portion of the locus in which the variation was found. Examples: "Exon 15", "Kinase domain".

Note: CDISC Codelist is based on SDTM Terminology 2022-06-24.

Per SDTMIG v3.4 and CDISC Rules for Genomics, GF Codetable Mapping File, here are the basic usage for the variables:

1. **GFTESTCD, GFTEST and GFTSTDTL:**

   GFTESTCD and GFTEST represent a high level or generalized description of the assessment, which is considered a characteristic finding of the genomic material and should not contain gene names or symbols (see information below for GFSYM), including but not limited to official gene symbols. ‘VAR’ or ‘V’ will be used as the suffix fragment in GFTESTCD to denote ‘Variation’ in the GFTEST.

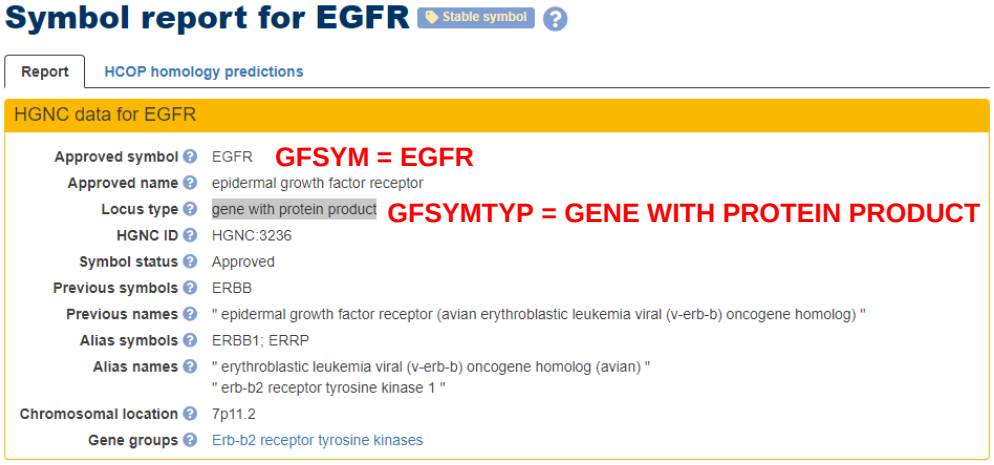
   GFTSTDTL represents the specific reportable for the assessment described in the GFTESTCD/GFTEST value, concepts that are insufficient on their own should contain additional descriptive text in the GFTSTDTL value, such as ‘NORMALIZED’, ‘INTERPRETATION’, ‘PREDICTED’, ‘NUMBER’ and ‘STATUS’. It would be recommended to have GFTSTDTL since

GFTESTCD may have multiple explanation according to the perspective of the genomic testing needs.

For the values of GFTESTCD, GFTEST and GFTSTDTL, it is suggested to check with GF Codetable Mapping File at first and select the appropriate values.

2. GFSYM and GFSYMTYP:

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 (www.genenames.org), as well as GFSYMTYP, see below the example of EGFR from the database. CDISC will not control this variable with CDISC Controlled Terminology.



3. GFINHERT:

GFINHERT identifies whether the variation can be passed to the next generation. In real practice, this variable could be used in Breast cancer clinical trials to capture germline *BRCA* mutation and somatic *BRCA* mutation, possible values from SDTM Terminology 2022-06-24 as below:

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Definition
C181177		Yes	Genomic Inheritability Type Response	INHERTGF	Terminology relevant to whether the variation can be passed to the next generation.
C17666	C181177		Genomic Inheritability Type Response	GERMLINE VARIATION	Any inheritable variation in the DNA that is transmitted to the progeny with some frequency.
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.
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.

4. GFCHROM, GFGENLOC and GFGENSR:

GFCHROM represents the designation (name or number) of the chromosome or contig on which the variant or other feature appears, e.g., "17"; "X". GFGENLOC represents numeric position within the sequence for the observed result as expressed in GFORRES. GFGENSR represents the sub-region within the genetic region of interest in which the observed variation at the position given in GFGENLOC is located, if relevant.

EXAMPLES BASED ON THE RESULTS OF GENETIC VARIATION

1) Single nucleotide variation examples – *BRCA1*/*BRCA2* mutations in Breast cancer clinical trials:

BRCA1/*BRCA2* is a tumor suppressor gene (Yoshida K, 2004) that encodes a protein (called breast cancer type 1 or 2 susceptibility protein) responsible for repairing DNA. When they have certain changes (called harmful variants or mutations), cancer can develop and grow out of control. Single nucleotide substitutions and small deletions or insertions (1 – 20 bases) account for most mutations in these genes. Most of these alterations result in a truncated

form of the breast cancer type 1 or 2 susceptibility protein. Some mutations are known to be associated with increased breast and ovarian cancer (deleterious) (National Cancer Institute, 2020).

Germline mutation: everyone has two copies of each of these genes - one copy inherited from each parent. Inherited mutations, also called germline mutations, are present from birth in all cells in the body.

Somatic mutation: a harmful variant in *BRCA1/BRCA2* can be inherited from one parent, while a normal copy of that gene would be inherited from the other parent (that's because in most cases, embryos with a harmful variant from each parent cannot develop). But the normal copy can be lost or change in some cells in the body during that person's lifetime, such a change is called a somatic mutation.

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

2) Copy number variation example – human epidermal growth factor receptor 2 (*HER2*) amplification result in Breast cancer clinical trials:

HER2 gene (also known as *ERBB2* – Erythroblastic Leukemia Viral Oncogene Homolog 2 in HGNC database, a member of the *ERBB*-like oncogene family) is important for breast cancer growth, which led to the development of the drug trastuzumab and other targeted treatments that have improved survival for women with HER2-positive breast cancer. This gene has been shown to be amplified in human breast cancer cell lines and is a significant predictor of both overall survival and time to relapse in patients with breast cancer (National Cancer Institute, 2018).

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

3) Other variations:

For other variations, please refer to GF Codetable Mapping File for the appropriate usage. For chromosomes variation, there is no controlled terms of GFTESTCD/GFTEST from SDTM Terminology 2022-06-24, we suggest checking the previous versions of SDTM Terminology for the possible values and submit the CDISC New Term Request via https://ncitermform.nci.nih.gov/ncitermform/. Here is an example of possible value for chromosomes variation from SDTM Terminology 2021-06-25:

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition
C116106		Yes	Pharmacogenomics Findings Test Code	PFTESTCD	Pharmacogenomics Findings Test Code	Terminology relevant to the test codes that describe pharmacogenomics findings.
C18280	C116106		Pharmacogenomics Findings Test Code	CHROMAB	Chromosomal Aberration	The determination of the holistic or specific regional aberrations in a chromosome (insertions, deletions, amplifications, translocations).

EXAMPLES OF SUMMARY MEASURES DERIVED FROM THE RESULTS OF GENETIC VARIATION

1) Microsatellite Instability (MSI) status assessed using a panel of microsatellite markers (an example from SDTMIG v3.4):

MSI status, the condition or state of genomic instability associated with defective DNA mismatch repair in tumors, is assessed by determining the difference in expression of microsatellite sequences within 5 of the Bethesda markers in tumor versus normal tissue (Horvat M, 2011): BAT25, BAT26, D2S123, D5S346, and D17S250 recommended in a 1997 NCI consensus meeting (Boland CR, 1998). Microsatellites are stretches of DNA with a repetitive sequence of nucleotides, e.g., AAAAA or CGCGCGCG, which are particularly susceptible to acquiring errors when the Mismatch Repair (MMR) gene function is impaired (Hegde M, 2014).

A tumor can be classified as follows (Boland CR, 1998) (Hegde M, 2014):

MSI-high: if two or more of the five markers of the core panel show instability or more than 30% of markers show instability in other marker panels.

MSI-low: if one of the five markers in the core panel shows instability or fewer than 30% of markers show instability in other marker panels.

MSI-stable: if 0 (or 0%) of the markers show instability in the core panel or other marker panels.

In colorectal carcinoma, MSI has been associated with the anatomical location of the tumor, poor differentiation, and TNM stage (Nazemalhosseini Mojarad E, 2016).

GFTESTCD	GFTEST	GFTSTDTL	GFORRES	GFGENSR
MICRISTB	Microsatellite Instability	DETECTION	NOT DETECTED	BAT-25
MICRISTB	Microsatellite Instability	DETECTION	NOT DETECTED	BAT-26
MICRISTB	Microsatellite Instability	DETECTION	NOT DETECTED	MONO-27
MICRISTB	Microsatellite Instability	DETECTION	NOT DETECTED	NR-21
MICRISTB	Microsatellite Instability	DETECTION	NOT DETECTED	NR-24
MICRISTB	Microsatellite Instability	MICROSATELLITE INSTABILITY OVERALL STATUS	MSI-STABLE	

CONCLUSION

Considering GF domain is quite new for most of us, we hope our experience on implementing this domain could provide extra helpful information beyond SDTMIG v3.4. However, there are some limitations in this paper, e.g., no examples of chromosomes variation due to no available controlled term of GFTESTCD/GFTEST. With the ongoing evolvement of GF domain assumptions, e.g., more detailed instructions and controlled terms, we believe it will be more and more easy for us to implement it successfully in clinical trials.

REFERENCES

Habets GG, van der Kammen RA, Willemsen V, Balemans M, Wiegant J, Collard JG. Sublocalization of an invasion-inducing locus and other genes on human chromosome 7. Cytogenet Cell Genet. 1992;60(3-4):200-5. doi: 10.1159/000133336. PMID: 1505215.

Francesco Passiglia, Angela Listì, Marta Castiglia, Alessandro Perez, Sergio Rizzo, Viviana Bazan, Antonio Russo. EGFR inhibition in NSCLC: New findings…. and opened questions?. Critical Reviews in Oncology/Hematology, Volume 112, 2017: Pages 126-135, ISSN 1040-8428, https://doi.org/10.1016/j.critrevonc.2017.02.009.

Den Dunnen et al. HGVS recommendations for the description of sequence variants: 2016 update. Hum.Mutat. 2016;25: 37: 564-569.

Genetic and Rare Diseases Information Center. “Cri du chat syndrome”. November 8, 2021. Available at https://rarediseases.info.nih.gov/diseases/6213/cri-du-chat-syndrome.

The Human Genome Structural Variation Working Group. Completing the map of human genetic variation. Nature 447, 161–165 (2007).

Yoshida K, Miki Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. Cancer Sci. 2004 Nov;95(11):866-71. doi: 10.1111/j.1349-7006.2004.tb02195.x. PMID: 15546503.

National Cancer Institute. “BRCA Gene Mutations: Cancer Risk and Genetic Testing”. November 19, 2020. Available at https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet.

National Cancer Institute. “HER2’s Genetic Link to Breast Cancer Spurs Development of New Treatments”. April 11, 2018. Available at https://www.cancer.gov/research/progress/discovery/her2.

Horvat M, Stabuc B. Microsatellite instability in colorectal cancer. Radiol Oncol. 2011;45(2):75-81.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, RodriguezBigas MA, Fodde R, Ranzani GN, Srivastava S. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res. 1998;58:5248-57.

Hegde M, Ferber M, Mao R, Samowitz W, Ganguly A, Working Group of the American College of Medical Genetics and Genomics (ACMG) Laboratory Quality Assurance Committee. ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). Genet Med. 2014;16:101-16.

Nazemalhosseini Mojarad E, Kashfi SM, Mirtalebi H, et al. Low level of microsatellite instability correlates with poor clinical prognosis in stage II colorectal cancer patients. J Oncol. 2016;2016:2196703.

ACKNOWLEDGMENTS

Thanks to our managers (Leo Li, Cindy Song and Tony Guo) for their great support on the conference and programming colleagues for their help on preparing the paper.

RECOMMENDED READING

- *CDISC SDTMIG v3.4, SDTMIG-PGx*
- *CDISC Supplemental Files – Rules for Genomics,GF Codetable Mapping File*

CONTACT INFORMATION

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

Xuejie Zhou
BeiGene, Inc.
xuejie.zhou@beigene.com

Shanshan Ma
BeiGene, Inc.
shanshan.ma@beigene.com

Qiuan Hu
BeiGene, Inc.
qiuan.hu@beigene.com

Liming Xie
BeiGene, Inc.
liming.xie@beigene.com