Challenges and Solutions in Chromosome Aberration Data Mapping to Genomics Findings (GF) Domain for Oncology Studies

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

ABSTRACT

Genomic Finding (GF) domain has been introduced in CDISC SDTMIG v3.4 as a new domain to include assessments and results for genetic variation, transcription, and summary measures from these assessments. SDTMIG v3.4 provides the detailed guidance on mapping gene variation data, e.g., single nucleotide and copy number variation. However, there is almost no instruction on chromosome aberrations data, which becomes a major challenge to implement GF domain for oncology studies collecting such information. In this paper, we would like to share our experience on implementing GF domain to map chromosome aberrations and propose some possible solutions before CDISC release upcoming IG document. In addition to the GF domain, we will also briefly discuss two alternative domains – Microscopic Findings (MI) and Cell Phenotype Findings (CP), as potential candidates for mapping chromosome aberrations data.

INTRODUCTION

With the advancements in genome-wide analysis technologies and our knowledge of human diseases, genetic testing has become increasingly integral in disease diagnosis, the selection of treatment options, and disease prognosis. In human clinical trials, particularly in the field of oncology, genetic testing and assessment results are commonly gathered as part of the individual's disease history information.

In human clinical trials, it was initially suggested to organize the genetic data using the PF domain, based on the provisional version 1.0 of SDTMIG-PGx introduced by CDISC on May 26, 2015. On November 29, 2021, CDISC released SDTMIG v3.4 and replaced PF domain with a new domain – Genomics Findings (GF). GF domain is used in capturing the result from both subject and non-host organism genomic material of interest, including assessment in the structure, function, evolution, mapping, and editing. Specifically, GF domain will be used in collecting finding from nucleic acid and may include the subsequent inferences and/or prediction about the related proteins/amino acids. The domain scope includes but not limited to assessments for genetic variation and transcription, for example, chromosome aberration.

In the context of cancer cytogenetics, chromosome aberration refers to changes in the number or structure of chromosomes, which occurs during the formation of cancer cells. However, there are significant challenges in mapping this type of data to the GF domain. These challenges arise from two main sources:

Firstly, the SDTMIG v3.4 examples and accompanying Controlled Terminology do not provide an extensive list of terms specifically for chromosome aberrations. This lack of predefined terminology makes it difficult to accurately categorize and map chromosome aberration data to the GF domain.

Secondly, the broader definition and assessment related to chromosome aberrations make it challenging to properly identify and map them to the GF domain. In this paper, two alternative options, the MI and CP domains, are suggested as potential candidates for mapping chromosome aberration data, in addition to the GF domain.

CHROMOSOME ABERRATION BACKGROUND

Chromosomes are stick-shaped thread-like structures found in the middle (nucleus) of each cell that carry genetic information in the form of DNA in the body (Figure 1). Chromosome aberration, also known as chromosome abnormality, refers to any structural or numerical abnormality in the chromosomes of an organism. Typically, it is a disorder characterized by a morphological or numerical alteration in single or multiple chromosomes, affecting autosomes, sex chromosomes, or both.

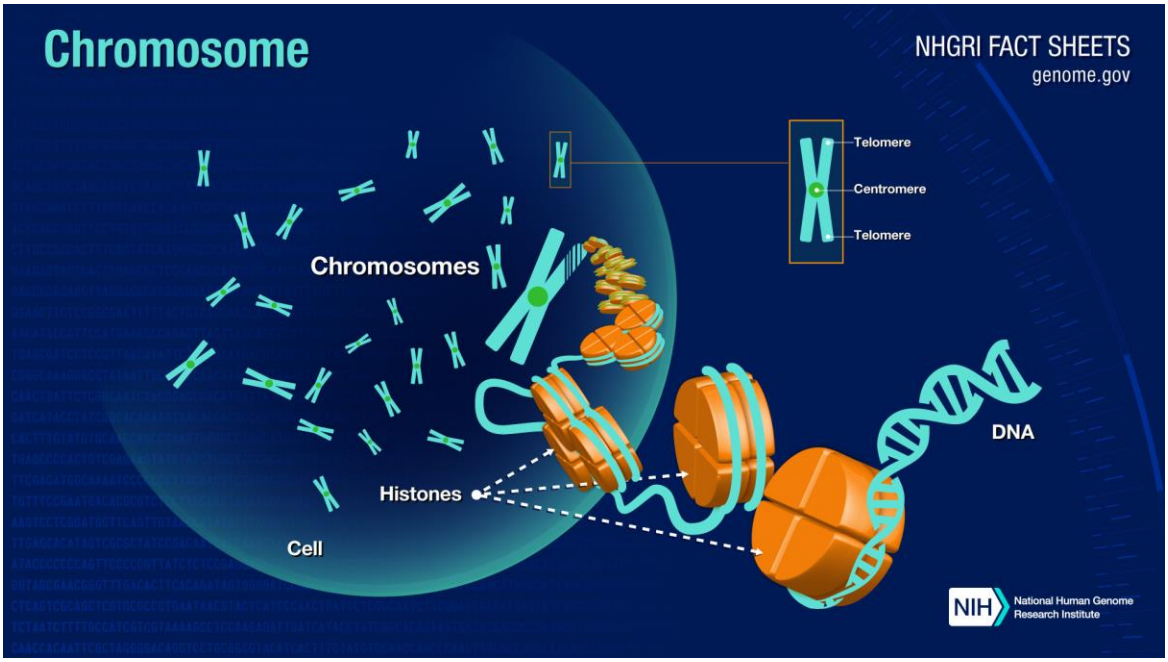

Figure 1. Chromosome, DNA and Genes

Chromosome abnormality was firstly found to be associated with cancer using Cytogenetics by Peter Nowell and David Hungerford in 1960 (Nowell & Hungerford, 1960). Scientists have hypothesized that the primary pathogenetic changes in cancer result from balanced rearrangements, while the secondary changes that occur during cancer progression are from unbalanced changes (Mitelman, 2005). Markers for chromosome aberration are essential in determining the development and advancement of tumor detection, identifying gene targets, and predicting the outcomes. Cancer as a multistep and progressive disease, early chromosome changes can provide the cell with a proliferative advantage. Often, these changes hijack or interfere with the normal cellular control mechanisms by disrupting proto-oncogenes and tumor suppressor genes and allowing additional changes to occur in the genome (Lobo, I. 2008).

When a chromosome is abnormal, it can have significant impacts on an organism's development and health. Abnormal chromosomes most often happen due to an error in DNA replication, recombination, or repair process during cell division. And these chromosome abnormalities often happen due to one or more of these: errors during dividing of sex cells (meiosis); errors during dividing of other cells (mitosis).

The following table summarizes the types of common chromosome aberrations:

Category	Type	Definition
Structural Abnormality	Deletion	Loses part of the chromosome, or part of the DNA code is missing
	Duplication	Presents an extra copy of all or part of a chromosome
	Inversion	Reverses the direction of a chromosome segment
	Translocation	Exchanges of genetic material either within the same chromosome or moves to another chromosome, including balanced translocation and Robertsonian translocation
Numerical Abnormality	Aneuploidy	There are more or fewer chromosomes than the normal number

Table 1. Types of Chromosome Aberrations

Two candidates for Chromosome aberration

Generally, chromosome aberration is mapped to GF domain. Two candidate domains – MI and CP can be used under certain circumstances (mainly based on the method of assessment).

MI domain is a finding domain that contains histopathology findings and microscopic evaluations. When the assessment is performed by microscopic analysis of chromosome and sub-chromosome structure and function but may not need to define this abnormality by structure or functionality, MI domain with the following MITESTCD/MITEST could be a better choice than GF domain.

Code	Codelist Code	Codelist Name	CDISC Submission Value	CDISC Definition	NCI Preferred Term
C18280	C132263	SDTM Microscopic Findings Test Code	CYEXAM	An assessment by microscopic analysis of chromosomal and subchromosomal structure and function.	Cytogenetic Analysis
C18280	C132262	SDTM Microscopic Findings Test Name	Cytogenetic Examination		

Table 2. Possible MITESTCD/MITEST Values from SDTM Controlled Terminology

MI domain to capture chromosome aberration is more focus on what results can we get from microscopic method, and these results are more considering the chromosome and sub-chromosome structure and function as a whole part of cell, and relevant test method could be karyotyping as an example. Comparing to MI domain, GF domain will collect what kind of abnormality this chromosome aberration is, and related abbreviation location in the chromosome will be collected as needed.

CP domain is a finding domain that contains data related to the characterization of cell phenotype, linage, and function based on expression of specific markers in single cell or particle suspensions. Cancer biomarker including genes, proteins, and other substances that can be used in testing and reflect the details about patient's disease situation. In real practice, certain biomarker status such as MRD (Minimal residual disease) test can be done by either method of flow cytometry or gene detection. If test result is focusing on genomic finding and results for gene variation, chromosomal variation, gene rearrangement, GF domain should be used. However, if test is focusing on gene-based phenotyping in disseminated tissue specimen and measured by flow cytometry, CP domain should be considered at first.

GF vs MI, CP

According to SDTMIG v3.4, GF domain is utilized for assessments and results of genetic variation and transcription, and summary measures derived from these assessments. On the other hand, CP domain supports tests that utilize markers for cell phenotyping, assisting in the identification and classification of both normal and abnormal cell populations. MI domain is used for microscopic findings, and it is common for biomarkers to be stained for differentiation purposes.

Mapping chromosome aberration into GF domain

Due to the current limitations in the available GFTESTCD and GFTEST information in the current IG and Controlled Terminology, we have developed our own GFTEST and GFTESTCD. These custom codes are illustrated in the following examples. Additionally, we have selected specific key variables based on SDTMIG v3.4 to assist in presenting typical examples related to chromosome aberration.

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.
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").
GFMETHOD	Method of Test or Examination	(METHOD)	The test method by which the examination is Performed by the wet lab to yield the result reported in the dataset.

Table 3. Key Variables from SDTMIG v3.4 GF Domain

According to 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. According to SDTMIG, if the CDISC Controlled Terminology is available, it should be utilized. Alternatively, sponsors are responsible for defining their own controlled list of terms when the CDISC Controlled Terminology is unavailable. Additionally, the GF Codetable Mapping File is provided to aid in selecting appropriate options for GFTEST, GFTESTCD.

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

The custom GFTEST, GFTESTCD and GFTSTDTL are showed in the following table:

Genomic Findings Test Code (GFTESTCD)	Genomic Findings Test Name (GFTEST)	Genomic Findings Test Detail (GFTSDTL)
TRANS	Translocation	OVERALL STATUS
TRANS	Translocation	PERCENT POSITIVE NUCLEI
DEL	Deletion	OVERALL STATUS
DEL	Deletion	PERCENT POSITIVE NUCLEI
INV	Inversion	OVERALL STATUS
INV	Inversion	PERCENT POSITIVE NUCLEI
DUP	Duplication	OVERALL STATUS
DUP	Duplication	PERCENT POSITIVE NUCLEI
AMPLI	Amplification	OVERALL STATUS
AMPLI	Amplification	PERCENT POSITIVE NUCLEI

Table 4. Custom GFTEST/ GFTESTCD/ GFTSTDTL Values

2. GFCHROM:

GFCHROM captures the specific designation, either a name or number, of the chromosome or contig on which the variant or other feature is present. Each chromosome has a centromere located near its center, dividing it into a long arm (q) and a short arm (p). For example, "13q" refers to the long arm (q) of chromosome 13, while "12p2" represents the short arm (p) of chromosome 12 at the labeled band 2. -- LOC would not be used in GF, GFCHROM is used to collect chromosome name where variant or other features are found.

EXAMPLES BASED ON THE CYTOGENETIC TESTINGS FOR CLL/SLL:

For chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL), cytogenetic testing performed with various methodologies (chromosome banding analysis (CBA), chromosome microarray analysis (CMA), fluorescence in situ hybridization (FISH)), has been used to detect the most common cytogenetic abnormalities of CLL/SLL – deletions of the 13q14, 11q22 (*ATM*), 17p13 (*TP53*) regions (Kiefer Y, 2012).

Other less common chromosome changes include an extra copy of chromosome 12 (trisomy 12) or a translocation (swapping of DNA) between chromosomes 11 and 14 [written as t(11;14)].

The loss of part of chromosome 13 occurs in more than half of all CLL cases (H. Döhner,1999; H. Döhner, 2000), is usually linked with a slower-growing disease and a better prognosis, while defects in chromosomes 11 or 17 often indicate a poorer prognosis. Trisomy 12 observed in 10-20% of CLL cases (A Puiggros, 2014) doesn't seem to have much of an effect on prognosis.

Example	GFTESTCD	GFTEST	GFTSTDTL	GFCHROM	GFORRES
del(11q)	DEL	Deletion	OVERALL STATUS	11q	NEGATIVE
del(13q)	DEL	Deletion	OVERALL STATUS	13q	POSITIVE
del(17p)	DEL	Deletion	OVERALL STATUS	17p	UNKNOWN
trisomy 12	DUP	Duplication	OVERALL STATUS	12	POSITIVE
t(11; 14)	TRANS	Translocation	OVERALL STATUS	11;14	UNKNOWN

CONCLUSION

This paper provides a broad overview of mapping chromosome aberrations to the GF domain and aims to contribute towards categorizing chromosome abnormalities and incorporating them into the domain. It is important to note that due to the diverse nature of chromosome aberrations and ongoing cytogenetic discoveries, it is not possible to cover all scenarios encountered in real-world data in this discussion. In order to effectively integrate chromosome aberration data into disease identification, categorization, and even prognosis estimation, there is still a considerable amount of work needed to properly map it in the GF domain. However, we remain optimistic that this new domain will greatly benefit clinical trials throughout this ongoing journey.

REFERENCES

Nowell, P., & Hungerford, D. 1960. “A minute chromosome in human chronic granulocytic leukemia”. Science 132, 1497

Lobo, I. 2008. Chromosome abnormalities and cancer cytogenetics. Nature Education 1(1):68

Mitelman, F. 2005. “Cancer cytogenetics update 2005”. Atlas of Genetics and Cytogenetics in Oncology and Haematology

Xuejie Zhou, Liming Xie, Qiuan Hu, Shanshan Ma. Time to Get in the Genomics Findings (GF) Domain. PharmaSUG China 2022, SR145.

Klein U, Dalla-Favera R. 2010“New insights into the pathogenesis of chronic lymphocytic leukemia. Semin Cancer Biol. 20:377–383.

Kiefer, Y., Schulte, C., Tiemann, M., & Bullerdiek, J. (2012). Chronic lymphocytic leukemia-associated chromosomal abnormalities and miRNA deregulation. The application of clinical genetics, 5, 21–28. https://doi.org/10.2147/TACG.S18669

Döhner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Kröber, A., Bullinger, L., ... & Lichter, P. (2000). Genomic aberrations and survival in chronic lymphocytic leukemia. New England Journal of Medicine, 343(26), 1910-1916.

Döhner, H., Stilgenbauer, S., Döhner, K., Bentz, M., & Lichter, P. (1999). Chromosome aberrations in B-cell chronic lymphocytic leukemia: reassessment based on molecular cytogenetic analysis. Journal of Molecular Medicine, 77, 266-281.

Ho, A. K., Hill, S., Preobrazhensky, S. N., Miller, M. E., Chen, Z., & Bahler, D. W. (2009). Small B-cell neoplasms with typical mantle cell lymphoma immunophenotypes often include chronic lymphocytic leukemias. American journal of clinical pathology, 131(1), 27-32.

Puiggros, A., Blanco, G., & Espinet, B. (2014). Genetic abnormalities in chronic lymphocytic leukemia: where we are and where we go. BioMed research international, 2014.

Schmidt, T.M., Fonseca, R. & Usmani, S.Z. (2021). Chromosome 1q21 abnormalities in multiple myeloma. Blood Cancer J. 11, 83.

National Center for Biotechnology Information (US). Genes and Disease [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 1998-. Chromosome Map. Available from: https://www.ncbi.nlm.nih.gov/books/NBK22266/

ACKNOWLEDGEMENTS

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.