cdisc®

2026 Europe Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION

MILAN, ITALY | MAIN CONFERENCE: 20-21 MAY | TRAININGS & WORKSHOPS: 18, 19, & 22 MAY

CDISC GF Domain Implementation in Hematology: Standardized Cytogenetic Mapping for Leukemia and Lymphoma

Jessica Chioda, Senior Statistical Programmer, AstraZeneca (Barcelona)

Speakers

I started my career with a degree in Statistics and spent five years at a CRO, sharpening my skills as a statistical programmer on clinical trials. For the past two years, I've been at AstraZeneca, where I focus on Early Oncology Programming.

Jessica Chioda
Senior Statistical Programmer
AstraZeneca (Barcelona)

Agenda

1. Hematology & Cytogenetics: Clinical context
2. Cytogenetic data in the GF domain
3. Why standards are critical in hematology
4. AstraZeneca standards framework: RDS and SDTM
5. Challenges and Best Practices
6. Lessons Learned

Hematology & Cytogenetics: Clinical Context

Hematologic malignancies affect the blood, bone marrow, and lymphatic system (e.g., leukemia, lymphoma, multiple myeloma)

Unlike most solid tumors, these diseases are defined and stratified by cytogenetic abnormalities

Cytogenetic findings directly inform diagnosis and disease classification, risk stratification and prognosis, treatment selection, trial eligibility, and key endpoints

A single abnormality may:

- Define a disease subtype
- Drive therapeutic choice
- Predict response or resistance

Negative findings are often as clinically meaningful as positive ones

As a result, cytogenetics is:

- Routinely collected
- Longitudinal
- Interpreted in a disease-specific way

cdisc®

Cytogenetics: Concepts, Methods and Challenges

What Is Cytogenetics?

- Study of chromosome and gene-level alterations that drive cancer biology
- Common abnormality types:
  - Translocations (e.g., t(8;14), t(14;18))
  - Deletions (e.g., del(7q), 5q deletion)
  - Copy number amplifications (e.g., 9p24.1 amplification)
- Findings may be simple or highly complex, involving:
  - Multiple abnormalities
  - Clonal populations
  - Mosaicism

Key Interpretive Challenges

- **Karyotype**: Chromosome-level overview; often includes complex or multi-clone results
- **FISH** (Fluorescence In Situ Hybridization): Targeted assessment of specific genes or chromosomal regions
- **Array / NGS-based methods**: Detect copy number and regional changes at higher resolution

How Cytogenetic Data Are Generated

- Same laboratory techniques, **but different clinical meaning** by indication
- **Chromosomal regions** (e.g., 9p24.1) may include multiple relevant genes (PD-L1, PD-L2, JAK2) without a single defining target
- **Historically:** Results captured inconsistently, heavy reliance on free text, limited cross-study comparability

cdisc®

Genomics Findings (GF): Overview

The GF domain is designed to represent **genetic and genomic** findings in a structured, analysis-ready way

It supports results across:

- Genes, chromosomal regions, and copy number changes
- Different laboratory methods (e.g., karyotype, FISH, array, NGS)

GF focuses on:

- What abnormality was assessed
- What the result was (positive, negative, unknown)
- Where it applies (gene or chromosomal region)

The domain is method-agnostic:

- The same biological finding can be represented consistently, regardless of how it was measured

Source: CDISC SDTM IG 3.4 – GF domain

Complications in Using the CDISC GF Domain

- Implementation is not straightforward.
- Cytogenetic data represent chromosomal aberrations, which are structural events rather than simple gene-level findings and may involve multiple chromosomes, regions or clonal populations.
- **Same abnormality can be assessed using different methods**, such as karyotyping, FISH, or array-based techniques.
- Many hematology findings are best interpreted at the **chromosomal region level**, rather than being tied to a single gene. These region-level results don't always align neatly with gene-centric examples found in existing guidance.
- There is a grey area between GF and MI for chromosome-level findings which can lead to inconsistent implementations across studies or vendors.

Source: CDISC SDTM IG 3.4 – GF domain

Why Standardization Is Complex in Hematology

Complex Disease Biology	Heterogeneous subtypes and clonal evolution Measurable residual disease (MRD) Changing response criteria over time Longitudinal biomarker shifts require precise, consistent structures
Multi-Source Data Integration	Local vs central labs Flow cytometry, cytogenetics, NGS Pathology and imaging (PET/CT, MRI) Different formats, timing, and levels of detail
Consistent Endpoints and Biomarkers	Common definitions for response (CR/PR) Survival endpoints (PFS, OS) Cytogenetic and molecular markers Enables reproducible analyses and cross-study comparison
Regulatory and Cross-Study Efficiency	Standardized data accelerates review Enables pooling across studies Supports comparison to external and historical data

cdisc®

AstraZeneca Standards Framework: RDS and SDTM

Raw Data Standards (RDS)	• Therapy-area-specific • Aligned with CDASH where possible
SDTM	• Regulatory submission standard • Mapped from RDS using AZ business rules
Dedicated hematology cytogenetics modules:	• CYTGEN – Cytogenetics for leukemia • CYTOLYM – Cytogenetics for lymphoma

RDS → SDTM mapping uses sponsor business rules aligned with CDISC principles

cdisc®

Mapping in Leukemia and Lymphoma

Disease-specific separation:

- Leukemia → CYTGEN; Lymphoma → CYTOLYM
- **One cytogenetic abnormality = one GF record**

Preserve:

- Abnormality type (deletion, duplication, translocation)
- Chromosomal location
- Overall result status (positive / negative / unknown)

Structured capture at raw data level

- Test performed
- Method (e.g. FISH, CBA)
- Abnormality assessed

SDTM GF mapping

- GFTEST / GFTESTCD → abnormality type
- GFCHROM → chromosome or region
- GFORRES / GFSTRESC → overall status

Explicit handling of negative and unknown findings

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

Challenge 1: Domain Ambiguity (GF vs MI)

- **Cytogenetic tests** (e.g., FISH, IHC, regional abnormalities) do not fit neatly into one SDTM domain

- **Literature and guidance show overlapping interpretations**
  - Chromosome-level findings → **MI**
  - Genomic/nucleic-acid–based tests → **GF**

- **Decision**
  - Adopt a GF-centric implementation for cytogenetic findings
  - Treat cytogenetics as genomic context when results support molecular interpretation

cdisc®

Best Practice: Domain Strategy

- Define **why data belong** in GF or MI before mapping
- Prioritize **consistency across studies** and downstream use
- **Document domain** assumptions clearly for:
  - Reviewers
  - Regulators
  - Future implementers

Recommendation: Choose one primary domain and justify exceptions only when needed

Challenge 2: Chromosome vs Gene Representation

Some tests target:

- Specific genes (e.g., MYC, BCL2)
- Chromosomal regions (e.g., 9p24.1, 7q deletion)

Ambiguity:

- Should region-based findings be forced into gene symbols?
- How granular should GFCHROM be?

Example:

- 9p24.1 is a chromosomal region
- Contains multiple biologically relevant genes (e.g., PD-L1, PD-L2, JAK2)
- Amplification impacts expression without a single “target gene”

Incorrect approach: assigning an arbitrary gene symbol

Adopted approach: treat as chromosomal region–level finding

cdisc®

Best Practice: Chromosome & Symbol Mapping

Use:

- GFCHROM for chromosomal identifiers
- GFSYM for gene or region symbols
- GFSYMTYP to distinguish:
  - GENE
  - REGION
- Leverage authoritative sources:
  - HUGO nomenclature
  - Peer-reviewed CDISC/PharmaSUG guidance

Outcome: Improved interpretability without over-specification

Challenge 3: Handling “NOT DONE”

Data collection systems often require an entry even when a test is not performed

- **“NOT DONE”** is:
  - Not a result
  - Not an observation

- Risk of polluting analysis datasets, if ‘NOT DONE’ is treated as a ‘result’

- This distinction becomes especially important when negative findings are clinically meaningful

Best practice: Representing ‘NOT DONE’

- **Adopted convention**
  - Populate GFSTAT = “NOT DONE”
  - Leave GFORRES / GFSTRESC empty
- **Applied consistently when**
  - Entire panel not performed
  - Gene-level test not performed
  - Individual assessment not performed

Outcome: aligns with SDTM intent, results ≠ execution status

Challenge 4: GFTEST / GFTESTCD Consistency

- **Initial mappings showed:**
  - Legacy test names
  - Inconsistent terminology
  - Increased risk of Pinnacle 21 findings

- "**Other**" findings and amplification results lacked standardization

Best Practice: Align Early with CDISC CT Intent

Standardize:

- GFTEST
- GFTESTCD
- GFTSTDTL
- Examples:
  - Use Copy Number Amplification instead of generic “Amplification”
  - Use Other Aberrations consistently
  - Normalize labels and technical formatting

Outcome: Improves cross-study consistency and reviewer confidence

Challenge 5: Free-Text & Complex Findings

- "**Other, specify**" fields may contain:
  - multiple abnormalities
  - narrative descriptions
  - mixed biological concepts

- Mapping into core result variables risks semantic dilution

Best Practice: Preserve Meaning via SUPPQUAL

- Map free-text and heterogeneous findings to SUPPGF
- Apply sub-setting rules:
  - Do not create SUPP records when value is missing
- Avoid over-normalizing narrative content into result variables

Outcome: Preserves information without compromising SDTM structure

Key Lessons Learned

- Cytogenetic data require intent-driven **SDTM design**
- **Biological meaning** should drive variable choice, not convenience
- Execution status ≠ result value
- Early alignment with **CDISC CT** prevents downstream rework
- **SUPPQUAL** is a strength, not a workaround
- Collaboration is key between biometrics, clinical and lab partners

What You Can Apply in Your Projects

- Start with biological and analytical intent
- Document domain decisions clearly
- Treat “NOT DONE” explicitly and consistently
- Use **GFCHROM / GFSYMTYP** to manage complexity
- Prefer clarity and traceability over forced standardization

Thank You!

Questions?

References

Shanshan Ma, Liming Xie, Xuejie Zhou, Qiuan Hu, BeiGene, Inc., (2023)
Challenges and Solutions in Chromosome Aberration Data Mapping to Genomics Findings (GF) Domain for Oncology Studies
https://www.cdisc.org/sites/default/files/2023-11/2023_us_interchange_paper_gf.pdf

CDISC SDTM Implementation Guide v3.4, CDISC, (2021)