The Omics Standards Landscape and the Connection with CDISC

Presented by Christine Connolly, Head of Standards Projects, CDISC

Meet the Speaker

Christine Connolly

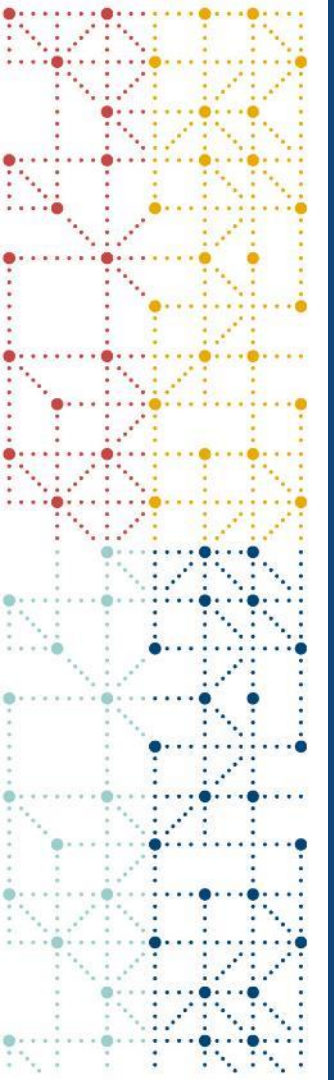
Title: Head of Standards Projects

Organization: CDISC

Ms. Connolly is the Head of Standards Projects for CDISC and advocates for data standardization given its potential to expedite approaches to improve health outcomes. Christine oversees the CDISC Genomics Subteam. She has led initiatives, developed, and implemented data standards for over fifteen years and has twenty-five years of experience working in global clinical trials in both academic and pharmaceutical settings.

Agenda

1. Omics and CDISC Standards
2. CDISC and Standards in Clinical Research
3. Reference Guide to CDISC Resources



Omics and CDISC Standards

Omics

The fields of research that use large scale sets of bioinformatics data to identify, describe and quantify the entire set of molecules and molecular processes that contribute to the form and function of cells, tissues and organisms.

Bioinformatics

The science of using computers, databases, and math to organize and analyze large amounts of biological, medical, and health information.

NCI Thesaurus (NCIt)
https://evsexplore.semantics.cancer.gov/evsexplore/welcome

Omics Improves Patient Outcomes

Comprehensive Biological Insights

Through holistic characterization and quantification of entire sets of biological molecules and the investigation of how they translate into the structure, function, and dynamics of an organism or group of organisms, support…

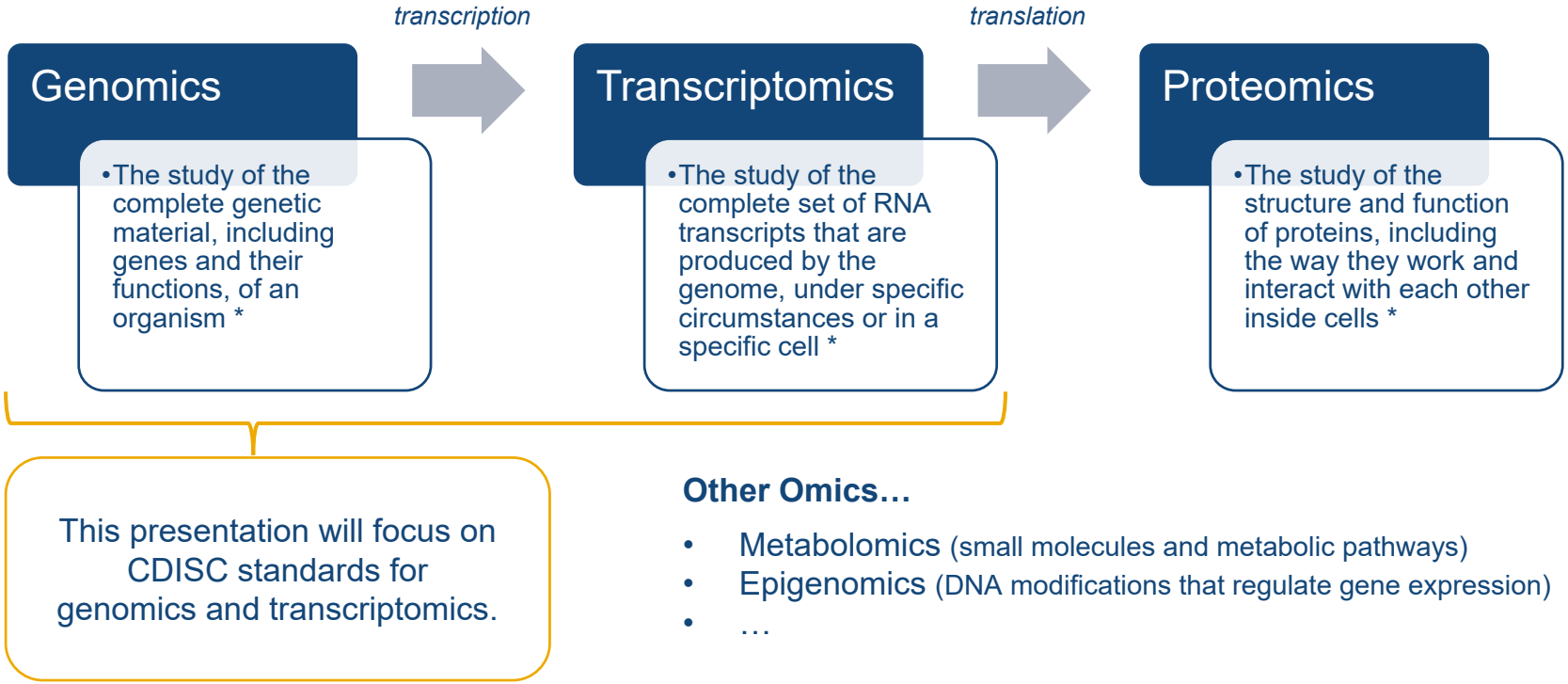
Personalized Medicine

Tailoring treatments to patients based on their unique characteristics

Improved Biomarkers

For disease susceptibility, diagnosis, and prognosis, and treatment safety, efficacy, and risks

Omics in Clinical Research

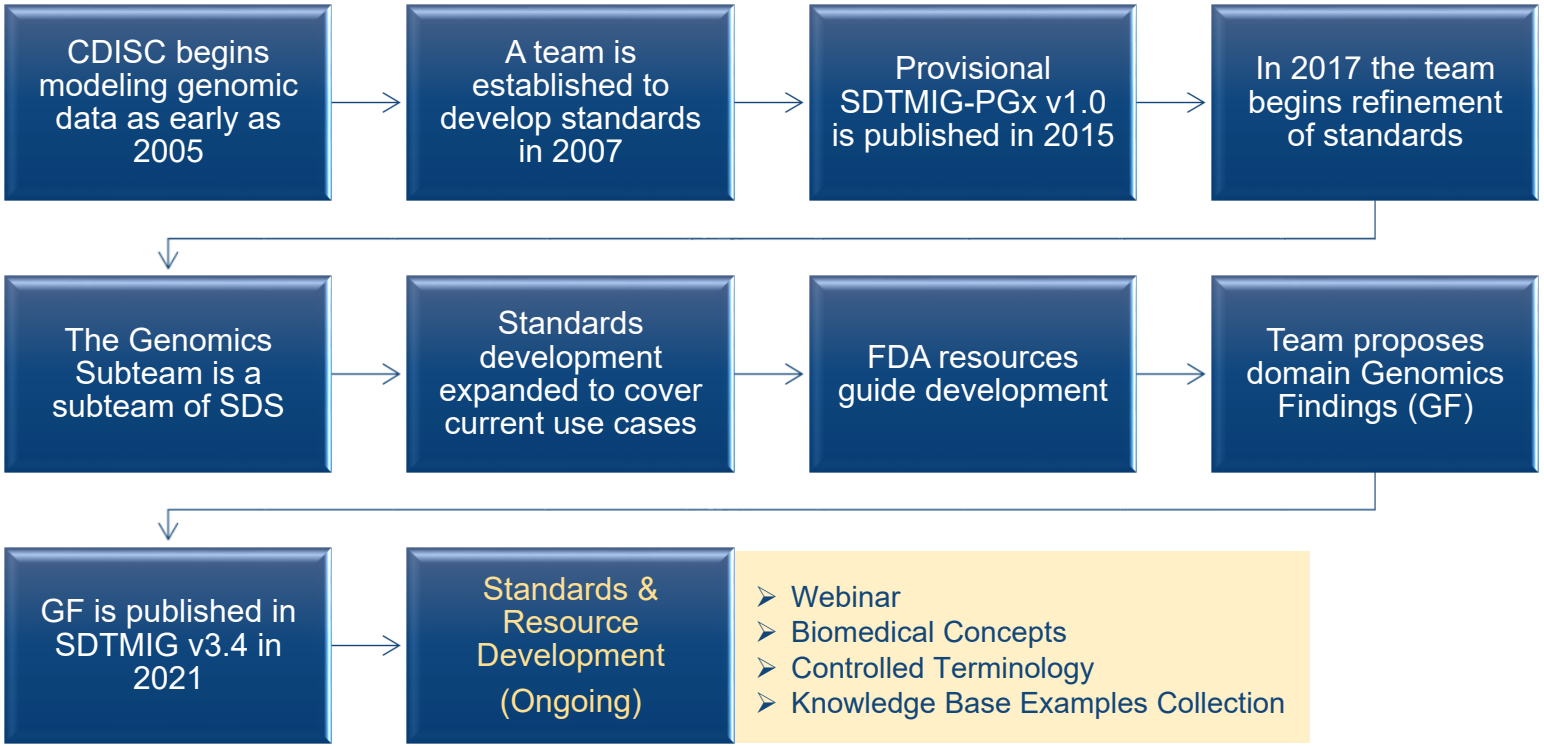


Other Omics…

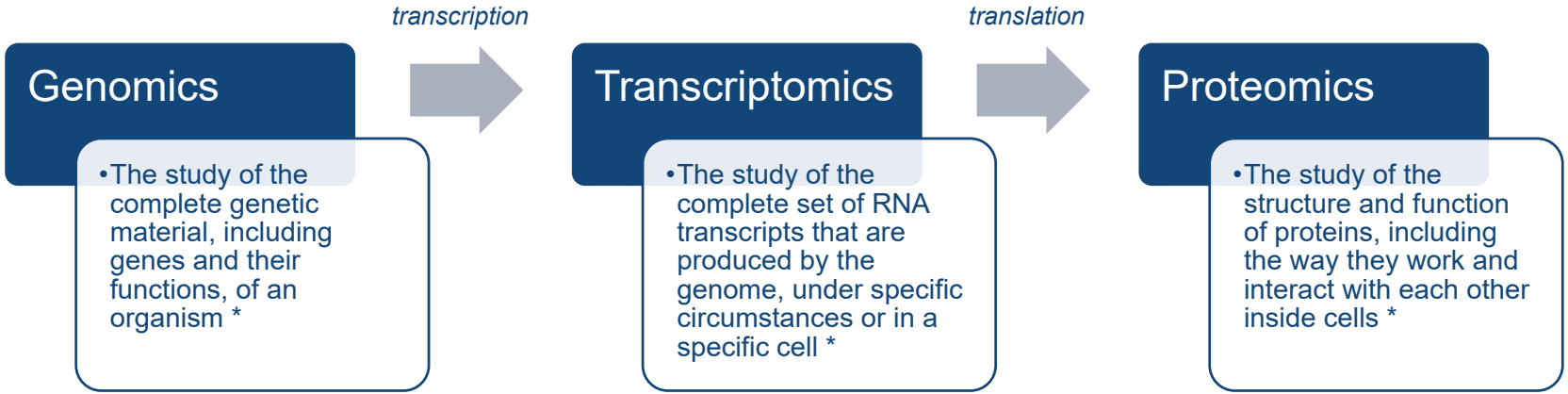
- Metabolomics (small molecules and metabolic pathways)
- Epigenomics (DNA modifications that regulate gene expression)
- …

* NCIt: https://evsexplore.semantics.cancer.gov/evsexplore/welcome

CDISC and Standards for Genomics



Genomics Findings (GF)



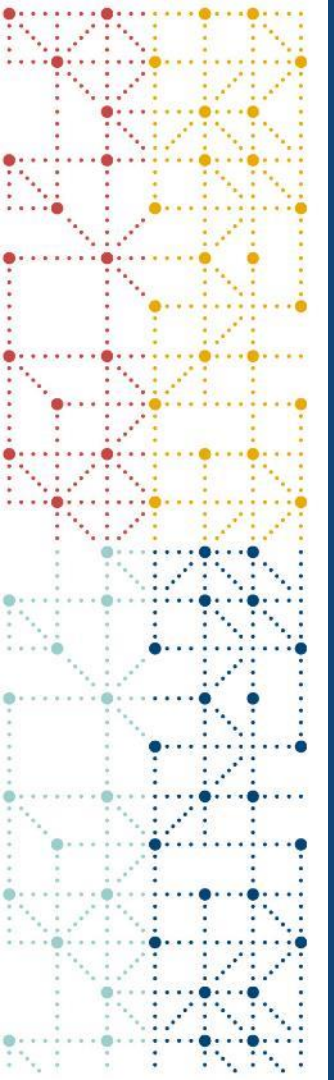
Scope

- Findings related to the structure, function, evolution, mapping, and editing of subject and non-host organism genomic material of interest
- Includes assessments for genetic variation and transcription, and summary measures derived from these assessments
- Characteristics assessed from nucleic acids and may include subsequent inferences and/or predictions about related proteins/amino acids

Out of Scope

- Direct assessments of proteins (e.g., amino acids)

* NCIt: https://evsexplore.semantics.cancer.gov/evsexplore/welcome

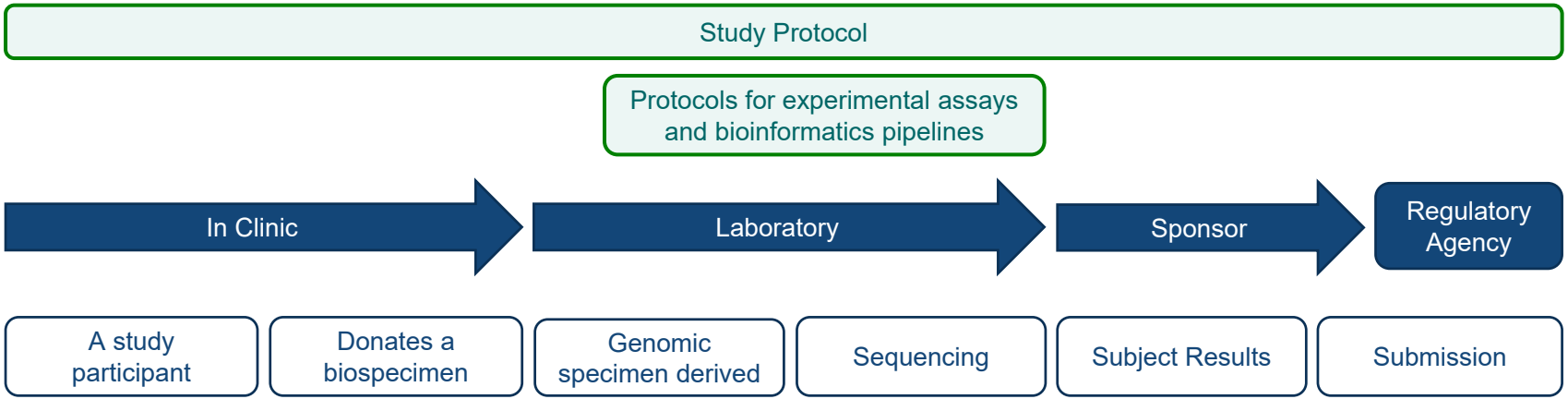


CDISC and Standards in Clinical Research

A Connected Standards Landscape

CDISC standards for genomics are part of a holistic standards landscape that supports clinical research.

Standards Across the Study Data Life Cycle



CDISC CDASH, SDTM, Controlled Terminology

- Demography (DM)
- Exposure (EC, EX)
- Biospecimen Collection & Handling (BE, BS, RELSPEC)

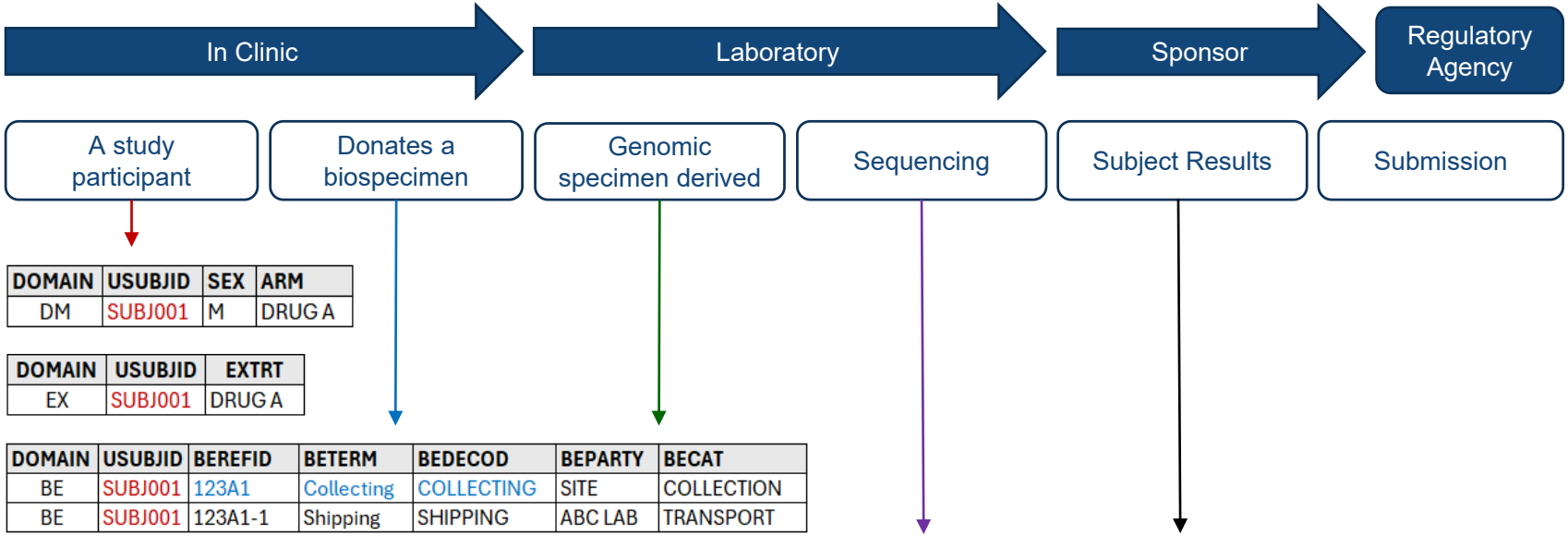
IEEE BioCompute

- Next Generation Sequencing (NGS) Bioinformatic Workflows

CDISC SDTM, ADaM, Controlled Terminology

- Biospecimen Collection & Handling (BE, BS, RELSPEC)
- Genomics Findings (GF)

Related Domains and Concepts



DOMAIN	USUBJID	SEX	ARM
DM	SUBJ001	M	DRUG A

DOMAIN	USUBJID	EXTRT
EX	SUBJ001	DRUG A

DOMAIN	USUBJID	BEREFID	BETERM	BEDECOD	BEPARTY	BECAT
BE	SUBJ001	123A1	Collecting	COLLECTING	SITE	COLLECTION
BE	SUBJ001	123A1-1	Shipping	SHIPPING	ABC LAB	TRANSPORT

BE	SUBJ001	123A1-1A0	Extracting	EXTRACTING	ABC LAB	EXTRACTION
BE	SUBJ001	123A1-1A0	Sequenced	SEQUENCING	ABC LAB	PREPARATION

DOMAIN	USUBJID	GFREFID	GFTESTCD	GFTEST	GFTSTDTL	GFORRES	GFORRESU	GFSYM	GFXFN	GFNAM	GFSPEC	GFMETHOD
GF	SUBJ001	123A1-1A0	TRNSCPTN	Transcription	NORMALIZED LEVEL	1.0523	/10^6	ACTB	AH749754.fastq.gz	ABC LAB	RNA	TARGETED TRANSCRIPTOME SEQUENCING
GF	SUBJ001	123A1-1A0	TRNSCPTN	Transcription	NORMALIZED LEVEL	1.9935	/MBP	ACTB	AH749754.fastq.gz	ABC LAB	RNA	TARGETED TRANSCRIPTOME SEQUENCING

Genomics-specific Concepts and External References

The GF domain includes concepts specific to genomics, some of which reference external databases and nomenclature.

Concepts Related to "Location"

Variable Name	Variable Label	Root Variable Definition	Examples
GFGENREF	Genome Reference	An identifier for the genome reference used to generate the reported result.	For example, Genome Reference Consortium Human Build 38 patch release 13 may be represented as "GRCh38.p13".
GFCHROM	Chromosome Identifier	The designation (name or number) of the chromosome or contig on which the variant or other feature appears.	"17"; "X"
GFSYM	Genomic Symbol	A published symbol for the portion of the genome serving as a locus for the experiment/test.	For human genetic data, standard nomenclature maintained in the HUGO Gene Nomenclature Committee (HGNC) database (www.genenames.org) is recommended.
GFSYMTYP	Genomic Symbol Type	A description of the type of genomic entity that is represented by the published symbol in --SYM.	"GENE WITH PROTEIN PRODUCT"
GFGENLOC	Genetic Location	Specifies the location within a sequence for the observed value in --ORRES.	"108175462"
GFGENSR	Genetic Sub-Region	The portion of the locus (--SYM) in which the variation was found.	"Exon 15", "Kinase domain"
GFSEQID	Sequence Identifier	A unique identifier for the sequence used as the reference to identify the genetic variation in the result.	"NM_001234", "ENSG00000182533", "ENST00000343849.2"

Genomics-specific Concepts and External References

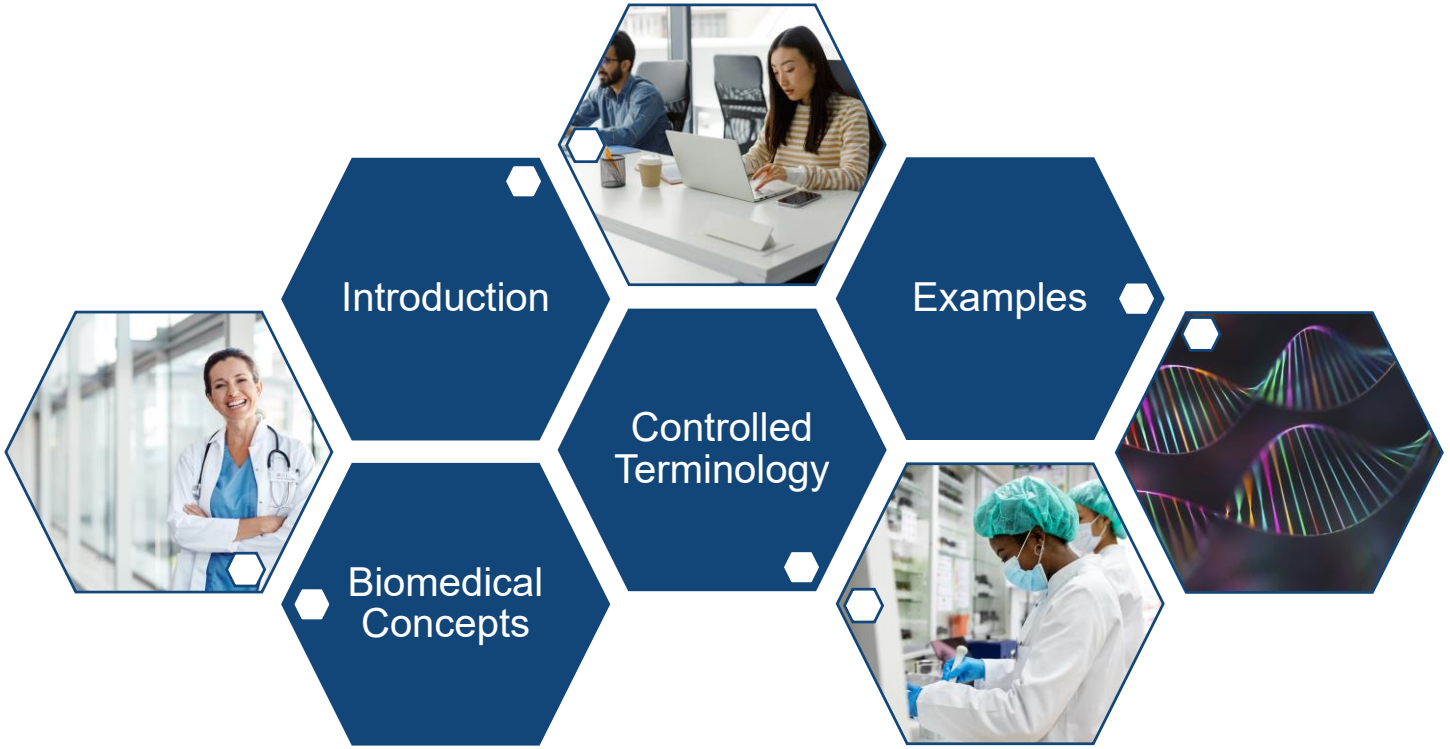
The GF domain includes concepts specific to genomics, some of which reference external databases and nomenclature.

Concepts Related to Results

Variable Name	Variable Label	Root Variable Definition	Examples
GFINHERT	Inheritability	Identifies whether the variation can be passed to the next generation.	"GERMLINE VARIATION", "MITOCHONDRIAL VARIATION", "SOMATIC VARIATION"
GFPVRID	Published Variant Identifier	A unique identifier for the variation that has been publicly characterized in an external database.	"rs2231142", "COSM41596"

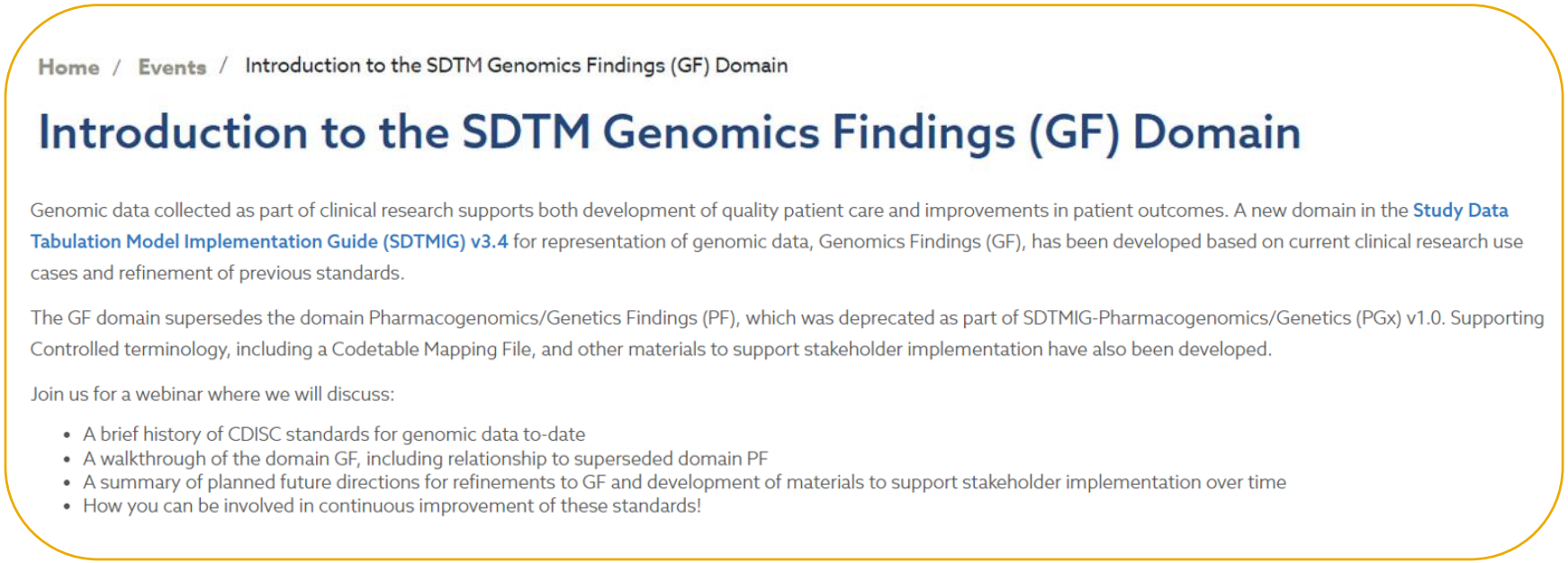
Reference Guide to CDISC Resources

CDISC Community Resources



Introduction

A public webinar introducing the GF domain is available in the CDISC website.



Home / Events / Introduction to the SDTM Genomics Findings (GF) Domain

Introduction to the SDTM Genomics Findings (GF) Domain

Genomic data collected as part of clinical research supports both development of quality patient care and improvements in patient outcomes. A new domain in the **Study Data Tabulation Model Implementation Guide (SDTMIG) v3.4** for representation of genomic data, Genomics Findings (GF), has been developed based on current clinical research use cases and refinement of previous standards.

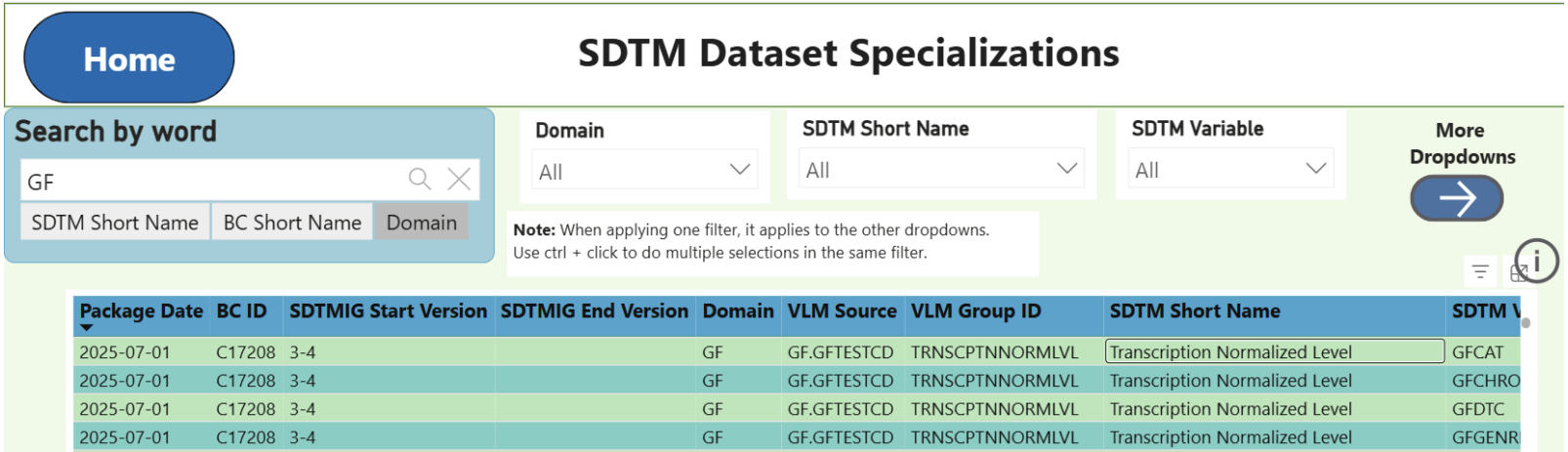
The GF domain supersedes the domain Pharmacogenomics/Genetics Findings (PF), which was deprecated as part of SDTMIG-Pharmacogenomics/Genetics (PGx) v1.0. Supporting Controlled terminology, including a Codetable Mapping File, and other materials to support stakeholder implementation have also been developed.

Join us for a webinar where we will discuss:

- A brief history of CDISC standards for genomic data to-date
- A walkthrough of the domain GF, including relationship to superseded domain PF
- A summary of planned future directions for refinements to GF and development of materials to support stakeholder implementation over time
- How you can be involved in continuous improvement of these standards!

Biomedical Concepts

Genomic concepts and GF dataset specializations are published in the CDISC Biomedical Concepts Browser.



Package Date	BC ID	SDTMIG Start Version	SDTMIG End Version	Domain	VLM Source	VLM Group ID	SDTM Short Name	SDTM V
2025-07-01	C17208	3-4		GF	GF.GFTESTCD	TRNSCPTNNORMLVL	Transcription Normalized Level	GFCAT
2025-07-01	C17208	3-4		GF	GF.GFTESTCD	TRNSCPTNNORMLVL	Transcription Normalized Level	GFCHRO
2025-07-01	C17208	3-4		GF	GF.GFTESTCD	TRNSCPTNNORMLVL	Transcription Normalized Level	GFDTC
2025-07-01	C17208	3-4		GF	GF.GFTESTCD	TRNSCPTNNORMLVL	Transcription Normalized Level	GFGENR

- Additionally, the CDASHIG and SDTMIG provide domain implementation guidance.

Controlled Terminology

SDTM Terminology

Codelist Name	CDISC Submission Value	CDISC Definition
Genomic Findings Test Name	Short Variation	An assessment of the variability in a short sequence of nucleotides (generally defined as 50 base pairs or less), when compared to a reference sequence.

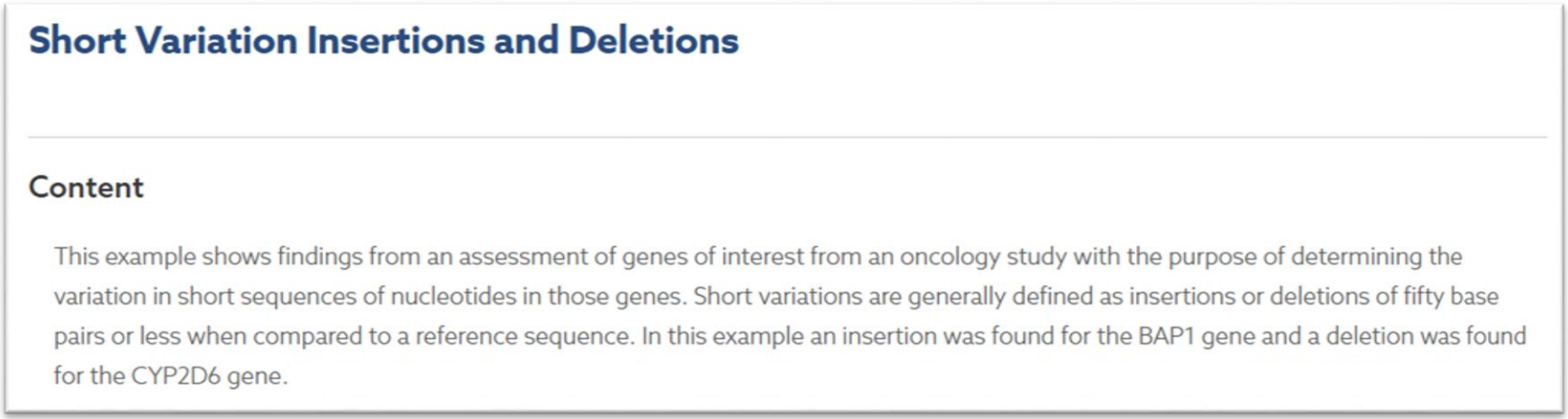
CDISC CONTROLLED TERMINOLOGY RULES:
Genomics Findings Domain (GF)

- Test Code/Name Codelists
- Test Detail Codelist
- Analytical Method Calculation Formula

GF Codetable Mapping

C-code (Concept Code)	Genomic Findings Test Code (GFTESTCD) (codelist code = C181178)	Genomic Findings Test Name (GFTEST) (codelist code = C181179)	C-code (Concept Code)	Genomic Findings Test Detail (GFTSDTL) (codelist code = C181180)	C-code (Concept Code)	Unit of Measure (UNIT) (codelist code = C71620)
C181334	SHRTVAR	Short Variation	C181343	PREDICTED AMINO ACID CHANGE		
C181334	SHRTVAR	Short Variation	C181344	PREDICTED CODING SEQUENCE CHANGE		
C181334	SHRTVAR	Short Variation	C181347	VARIANT READ DEPTH/READ DEPTH	C25613	%

Examples



Short Variation Insertions and Deletions

Content

This example shows findings from an assessment of genes of interest from an oncology study with the purpose of determining the variation in short sequences of nucleotides in those genes. Short variations are generally defined as insertions or deletions of fifty base pairs or less when compared to a reference sequence. In this example an insertion was found for the BAP1 gene and a deletion was found for the CYP2D6 gene.

SDTMIG
- Single Nucleotide and Copy Number Variation
- Single Nucleotide Variation
- Transcription
- Microsatellite Instability

** Preparing for Internal Review

Knowledge Base Examples Collection
- Variable Number of Tandem Repeats
- Short Variation Insertions and Deletions
- Tumor Mutation Burden
- Sequence Rearrangement Fusion
- ** Cytogenetic Abnormality
- ** Drug Absorption, Distribution, Metabolism, and Excretion (ADME) Genes
- ** Cytogenetic Assessment of Copy Number Amplification
- ** Human Leukocyte Antigens (HLA) Genotyping

CDISC Genomics Resource Reference

Resource	Description	Links
Introduction	Introduction to the SDTM Genomics Findings Domain	CDISC Public Webinar: • https://www.cdisc.org/events/webinar/introduction-sdtm-genomics-findings-gf-domain
Biomedical Concepts	Standards agnostic, defined concepts and GF dataset specializations	CDISC Biomedical Concepts Browser: • https://www.cdisc.org/cdisc-biomedical-concepts CDISC Implementations: CDASHIG v2.3, SDTMIG v3.4 • https://www.cdisc.org/standards/foundational/cdash • https://www.cdisc.org/standards/foundational/sdtmig
Controlled Terminology	• SDTM Terminology • GF Controlled Terminology Rules • GF Codetable Mapping	CDISC Controlled Terminology: • https://www.cdisc.org/standards/terminology/controlled-terminology
Examples	Informative, real-life examples of GF implementation	CDISC SDTMIG v3.4 GF Examples: • https://www.cdisc.org/standards/foundational/sdtmig CDISC Knowledge Base Examples Collection • https://www.cdisc.org/kb/examples

How can I participate?

We invite you to volunteer as a member of the Genomics Subteam.

Sign-up through the CDISC Website

- https://www.cdisc.org/volunteer/form

Reach out if you need help!

- Email: **Christine Connolly,** cconnolly@cdisc.org

Thank You!