

Single Day Event

Clinical Trials in the Era of Precision Medicine: Innovative Approaches and Best Practices

Mohit Sehgal, PhD
Practice Head - Cell, Gene & Advanced
Therapies, Hexaware Technologies

Dr. Sapna Chandran
Lead - Subject Matter Expert -
Clinical Solutions & Biomarkers,
PointCross Life Sciences



Abstract

Precision Medicine and Data Quality Analysis: Innovative Approaches and Best Practices

With an ever-growing need for developing therapeutic technologies for more than 7,000 rare diseases and continuous efforts in making existing therapies better, precision medicine is revolutionizing healthcare by tailoring treatments to cohorts/individual patients based on genetic, environmental and lifestyle factors. The success of this approach heavily relies on the quality and analysis of the data. High-quality data ensures accurate diagnosis, effective treatment plans and improved patient outcomes. Innovative approaches in data quality involve advanced data collection methods, integration of diverse data sources, and robust validation techniques. Analysis techniques leverage AI/ML to uncover patterns and insights from vast datasets. Best practices include establishing standardized protocols for data management, ensuring data privacy and security, and fostering interdisciplinary collaboration among healthcare professionals, data scientists and researchers.

This presentation will underscore the critical role of data quality and analysis in advancing precision medicine, highlighting innovative strategies and best practices that drive this transformative field.



About me

- **PhD in Cell & Molecular Biology**
Drexel University
- **Postdoctoral Research**
Drexel University College of Medicine
The Wistar Institute
- **Senior Research Scientist**
Cell & Gene Therapy Division, Intas Pharmaceuticals
- **Technology Solutions for Life Sciences & Healthcare**
Cognizant
Hexaware Technologies

Research Areas

- **Immunology**
- **Oncology**
- **Infectious Diseases**
- **Cell & Gene Therapy**
- **Virology**



Agenda

- ① Traditional Medicine vs. Precision Medicine

- ② Clinical Trial Design in Precision Medicine

- ③ Challenges in Genomics and Precision Medicine

- ④ Clinical Data Management: Innovative Approaches and Best Practices

- ⑤ Summary, Q & A



Traditional Medicine has been One Size Fits All

1



Travel to a physician
for a health check

2



The office looks the way
it did **40 years ago**

3



A large aliquot of blood
is drawn, **which hurts!**

4



Very **few measurements** are
made, many questionable

5

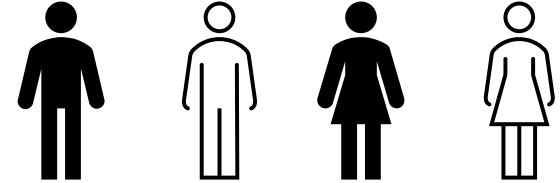


Treat you based on
population averages

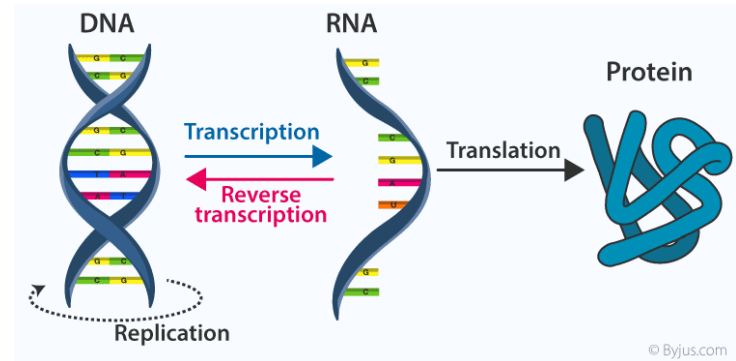
Understanding Human Genome



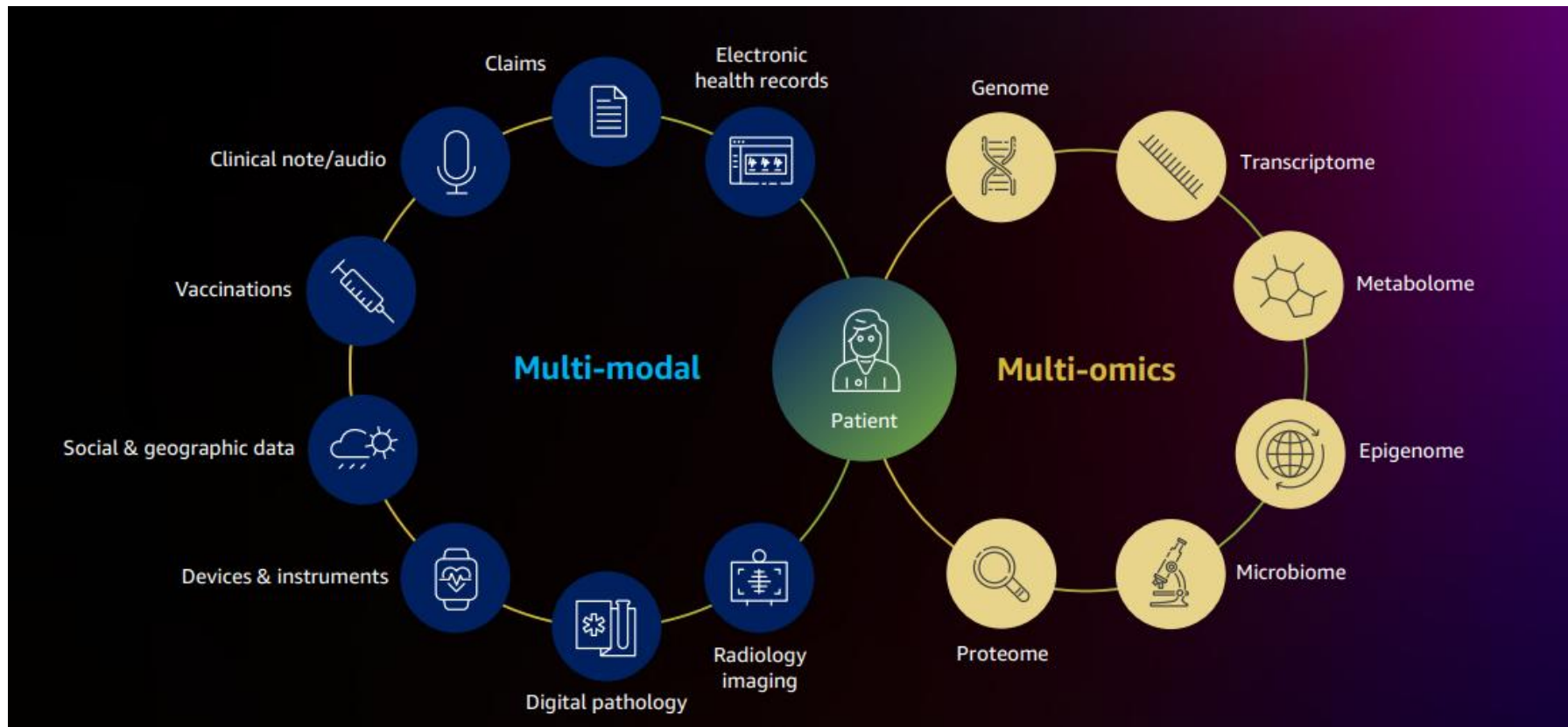
Human Genome is 3 Billion Base Pairs of DNA
in a single cell



Humans are 99.9% genetically identical,
which means that out of 3 billion base pairs,
0.1% i.e., 3 million pairs are unique

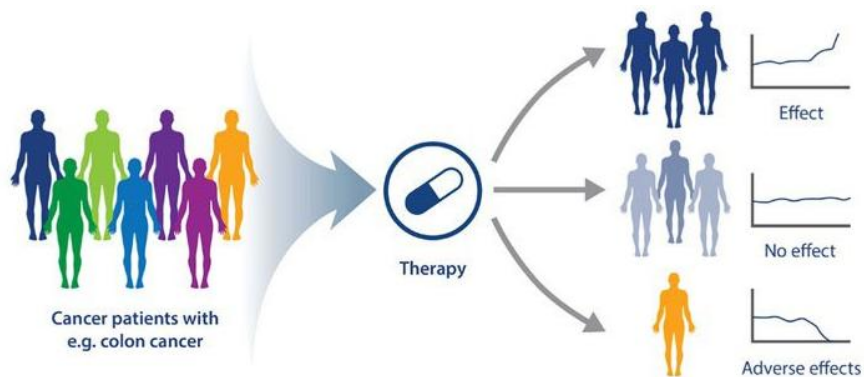


Creating a more holistic view of the patient

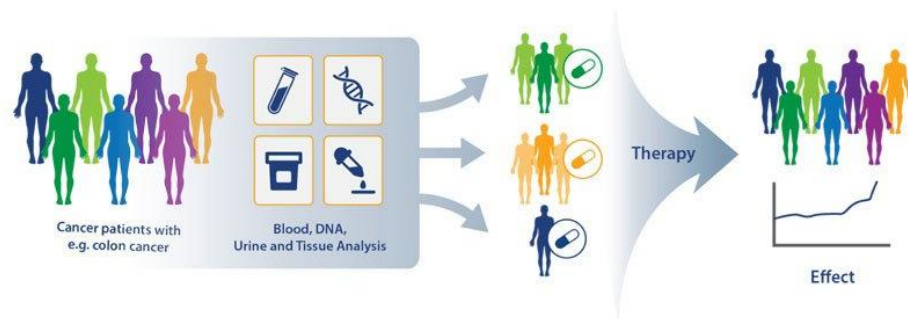


Medicine should not be “One Size Fits All”

Traditional Medicine One Size Fits All



Precision Medicine Personalized





Precision Oncology



American Association
for Cancer Research®

myAACR

Search

DONATE TODAY

CANCER RESEARCHERS / OTHER
HEALTH CARE PROFESSIONALS

PATIENTS, CAREGIVERS,
AND ADVOCATES

GET INVOLVED

WAYS TO GIVE

ABOUT THE AACR

NEWSROOM

AACR ANNUAL
IMPACT REPORT

LEADERSHIP AND
GOVERNANCE

AACR FOUNDATION

HISTORY OF THE
AACR

BLOG

CONTACT US

Home > About the AACR > Newsroom > News Release > Nearly Half of Oncology Drugs Approved Since 1998 Are Precision Therapies

IN THIS SECTION

NEWS RELEASES

MEDIA ADVISORIES

MEDIA CONTACTS

MEDIA RESOURCES

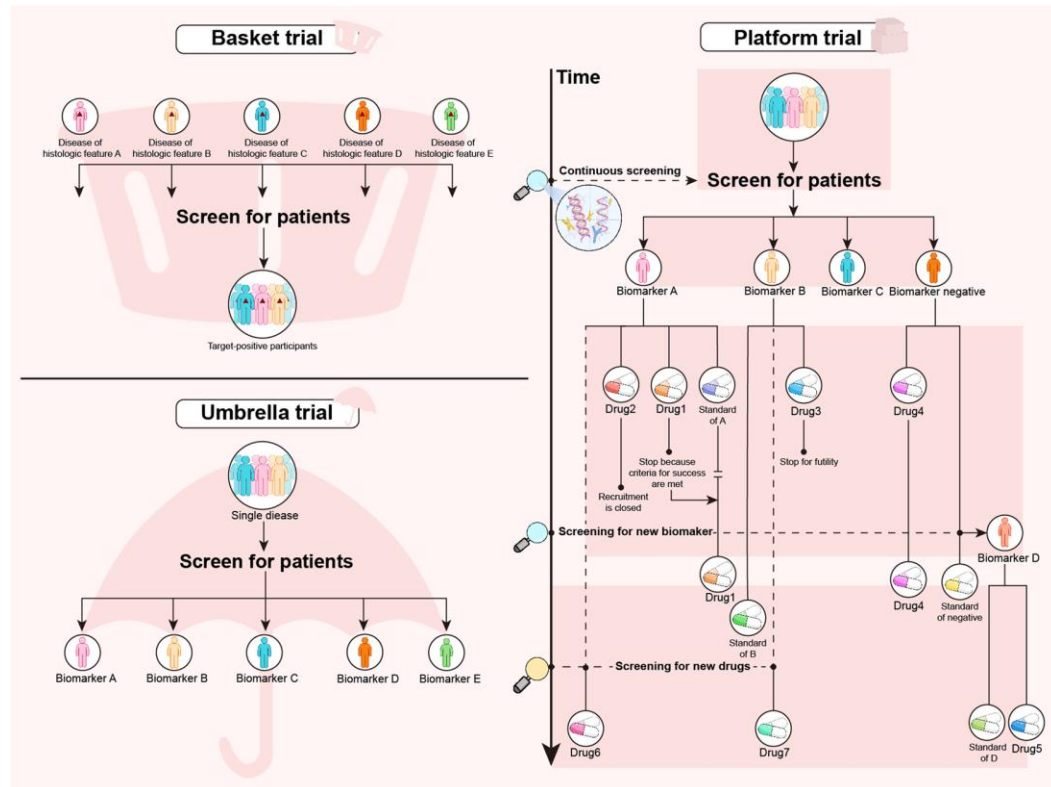


Nearly Half of Oncology Drugs Approved Since 1998 Are Precision Therapies

October 18, 2023

**CANCER PATIENTS ELIGIBLE FOR PRECISION ONCOLOGY THERAPIES
NEARLY DOUBLED BETWEEN 2017 AND 2022**

Clinical Trial Design in Precision Medicine





Challenges in Biomarker-guided Clinical Trials



Patient
recruitment
and selection



Data
Challenges



Biomarker
Assay
validation



Ethical
considerations



Regulatory
aspects

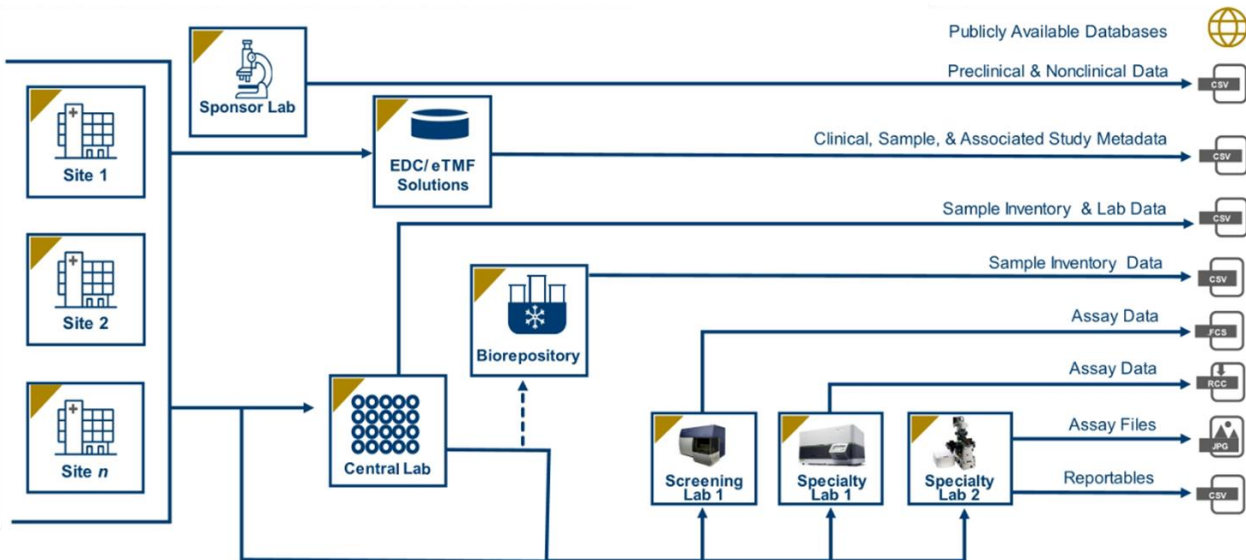


Cost
considerations



Disconnected Data, Samples, Technology

Data streams siloed across decentralized locations and within sponsor organizations



Manual, time-consuming data processing



Inability to map, transform, integrate, query data from various vendors/file formats



Hard to analyze and visualize data in one place



Difficult to generate shareable reports to key stakeholders



About me – Dr Sapna

- **Lead Subject Matter Expert**
Clinical Solutions, PointCross Life Sciences
- **Scientist**
Indian Council of Medical Research, ICMR-NCDIR
- **Regulatory Affairs**
- **PostGraduate**
Oral Pathology , Microbiology & Forensic Medicine

Research Areas

- **Oncology**
- **Immunotherapy**
- **Infectious Diseases**
- **Virology**
- **Forensic Medicine**

Challenges of Biomarker Data

Data Transfer Spec

Variation on DTS by vendor
Vendors ranging from a university lab to large assay platform
Assay Methodology Variability

Inconsistent definition & Classification

Inconsistent naming conventions across labs and assay types
Time-consuming manual standardization

Control Terms

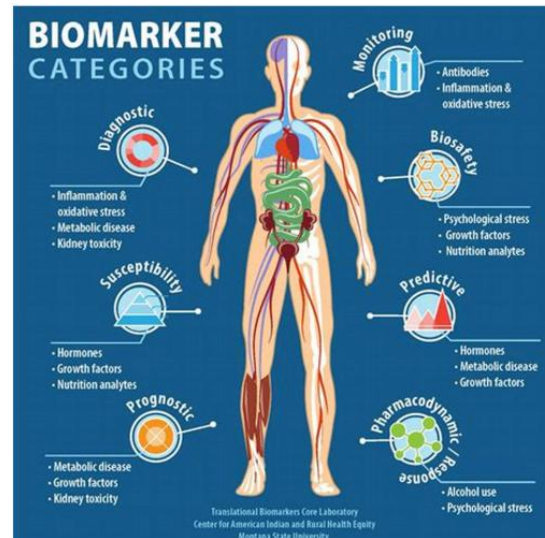
Limited Guidance /database from CDISC

Interdisciplinary knowledge across multiple roles such as scientists, data managers, programmers



Data standard evolution

Evolving CDISC standard in biomarker data standardization





Biomarker Validation & Standardization Process

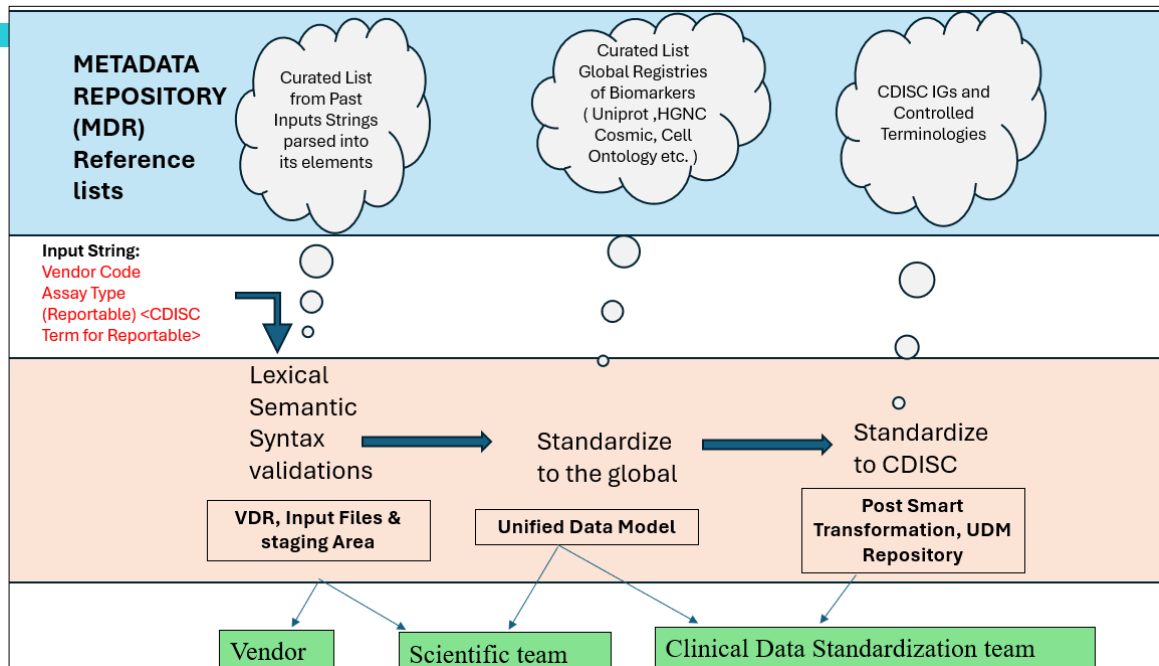
Automated Validation & Data Transformation

- To ensure reliability and reproducibility in clinical trials
- To confirm their predictive power and clinical relevance, thereby enhancing the credibility of trial outcomes.

MDR - “the ontology engine”.

- Intelligent Standards Management & Integrations

Scientific-Regulatory Bridge Framework



DOMAIN	ISSUE	EXAMPLE	PROMPT QUESTION	DECISION
GF	When Gene Symbol cannot be identified from HGNC	ZBA1	Gene Symbol not found in the HUGO database	Request advice from Scientific team



Example :Genomic Findings

- **Description**

To include assessments & results for genetic variation , transcription, gene expression & fusion

- **Dataset name**

gf.xpt, Genomic findings

- **Structure**

One record per finding per observation per specimen per subject

- **Key Variables**

Study ID, Subject ID , Method of Test or Examination , Specimen Material Type , Visit Number, Name of Genomic Measurement;
Genome Reference

Subject Identifier	Sample ID	Sample Category	Processed Specimen Type	Visit	Assay Method	Assay Lab/ Assay Name	Assay Lab Analyte Name	Analyte Result	DNA sequence change	Protien description	Variant Allele Frequency	Reference sequence
S12345	R001	Blood	RNA	D1	qRT-PCR	Transcript Lab/RNA Analysis	BRCA1	Detected	c.68_69del	BRCA1 - Breast Cancer	0.002	GRCh38
S12346	D002	Blood	DNA	D2	PCR	Genomics Lab/DNA Sequencing	TP53	Not Detected	c.215C>T	TP53 - Tumor Suppressor	0.01	GRCh38

↓
GFSPEC
(Specimen Material
Type)

↓
GFMETHOD
(Method of Test)

↓
GFSYM
(Genomic Symbol)

↓
GFORRES(Result
/Finding in Units)

↓
GFSEQID(Sequence
Identifier)



Standardizing to Global : HUGO Registry

Biomarker Standardization

Dashboard : CR Details **CR1234567** GF

Validate and recommend(Auto) Verify Prompt Questions and Flag Awaiting Response Review Close

Cha...

Rol...

Met...

Test...

Input...

GF								
# Issues								
Show All	Source Reportable Validation	Standardization	CDISC Completeness					
<input type="checkbox"/>	Sp...	Test Code *	Test *	Method *	Genomic Sym...	Genomic Symbol Type	Category	
<input type="checkbox"/>	Source		IL-17F	TLDA				
<input type="checkbox"/>	Target Term	TRNSCPTN	Transcription	REVERSE TRANSCRIPTASE PCR	IL17F	GENE WITH PROTEIN P...	Gene Expression	
<input type="checkbox"/>	Source		IL-1a	TLDA				
<input type="checkbox"/>	Target Term	TRNSCPTN	Transcription	REVERSE TRANSCRIPTASE PCR	IL1A	GENE WITH PROTEIN P...	Gene Expression	
<input type="checkbox"/>	Source		KLK5	TLDA				
<input type="checkbox"/>	Target Term	TRNSCPTN	Transcription	REVERSE TRANSCRIPTASE PCR	KLK5	GENE WITH PROTEIN P...	Gene Expression	
<input type="checkbox"/>	Source		MKG67	TLDA				
<input type="checkbox"/>	Target Term	TRNSCPTN	Transcription	REVERSE TRANSCRIPTASE PCR	MKG67	GENE WITH PROTEIN P...	Gene Expression	
<input type="checkbox"/>	Source		CCL17	TLDA				
<input type="checkbox"/>	Target Term	TRNSCPTN	Transcription	REVERSE TRANSCRIPTASE PCR	CCL17	GENE WITH PROTEIN P...	Gene Expression	
<input type="checkbox"/>	Source		CD209	TLDA				
<input type="checkbox"/>	Target Term	TRNSCPTN	Transcription	REVERSE TRANSCRIPTASE PCR	CD209	GENE WITH PROTEIN P...	Gene Expression	

Completed - Output With Errors Initiate Transformation

HGNC

Gene data Tools Downloads VGNC Contact us More

Search symbols, keywords or IDs

Request symbol

All download files including the archive files are now in a publicly accessible Google Storage Bucket. Downloads page links have been updated.

Symbol report for CCL17

Report HCOP homology predictions

HGNC data for CCL17

Approved symbol CCL17

Approved name C-C motif chemokine ligand 17

Locus type gene with protein product

HGNC ID HGNC:10615

Symbol status Approved

Previous symbols SCYA17

Previous names " small inducible cytokine subfamily A (Cys-Cys), member 17 "
" chemokine (C-C motif) ligand 17 "

Alias symbols TARC; ABCD-2

Chromosomal location 16q21

Gene groups Chemokine ligands

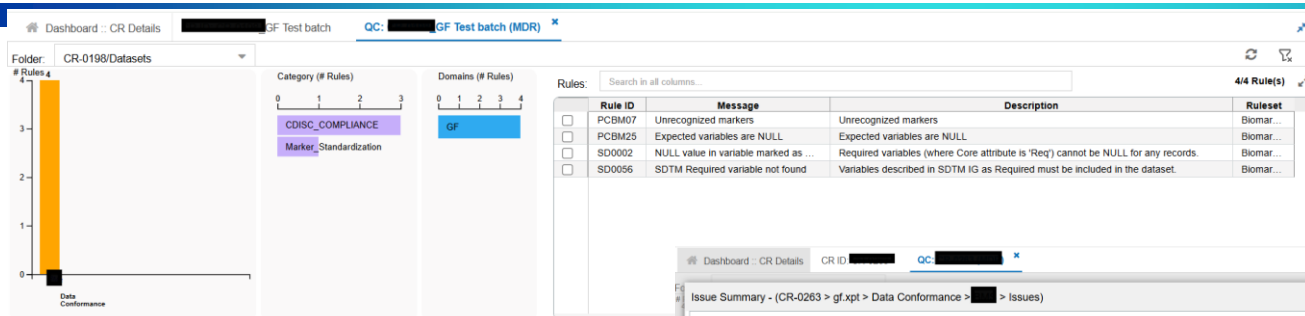
GFSYM = CCL17

GFSYMTYP = GENE WITH PRODUCT

Vendor provided data		Text Interpretation-based on CDISC Linkage			Text Interpretation- derived from HUGO database			
Vendor Reportable Name	Vendor Genomic Sym	Genomic Findings Test Name	Category	Method of Test or Examination	Approved Name	Genomic Symbol Type	Genetic Location	HGNC ID
PREDICTED CODING SEQUENCE CHANGE	BAG3	Single Nucleotide Variation	Genetic Variati	TARGETED GENOME SEQUENCING	BAG cochaperone 3	GENE WITH PROTEIN PRODUCT	121410891-121437331	939
PREDICTED AMINO ACID CHANGE	EMD	Single Nucleotide Variation	Genetic Variati	TARGETED GENOME SEQUENCING	emerin	GENE WITH PROTEIN PRODUCT	153607654-153609883	3331
Gene expression	STAT1	Transcription	Gene Expressio	REVERSE TRANSCRIPTASE PCR	Signal transducer and activator of transcription 1	GENE WITH PROTEIN PRODUCT	2q32.2	11362



Data Quality : Validation with eDV



(71 Open, 0 Closed)

1 file(s) loaded

File/Dataset	Domain	# Rows	Data Conform...	File Status
gf.xpt	GF	4	Issues	Active

Issue Summary - (CR-0263 > gf.xpt > Data Conformance > > Issues)

Issue Summary - (CR-0263 > gf.xpt > Data Conformance > > Issues)

Study ID	Domain	Publisher ID	Rule ID	Message	Description	Ruleset
GF	GF	CG0014, CG0208, 12, 257	SD0002	NULL value in variable marked as Required	Variables should be populated with terms from its CDISC controlled terminology codelist. New terms...	Biomar...
GF	GF	CG0014, CG0313, CG0554	SD0056	SDTM Required variable not found	Variables are NULL	Biomar...
GF	GF	FDAB017	CT2002	'GENE WITH PROTEIN PRODUCT;immunoglobulin gene; GENE WITH PROTEIN PRODUCT;C	variables (where Core attribute is 'Req') cannot be NULL for any records.	Biomar...
GF	GF	FDAB017	CT2002	'MRD 10^5' value not found in 'GFTSTDL' extensible codelist	described in SDTM IG as Required must be included in the dataset.	Biomar...
GF	GF	FDAB017	CT2002	'MRD' value not found in 'GFTSTCD' extensible codelist		
GF	GF	FDAB017	CT2002	'Measurable Residual Disease' value not found in 'GFTEST' extensible codelist		
GF	GF	FDAB017	CT2002	'Per Million Count Lower 95th Percentile' value not found in 'GFTSTDL' extensible codelist		
GF	GF	FDAB017	CT2002	'immunoglobulin gene' value not found in 'SYMTPGPF' extensible codelist		

Issue Drilldown - (> gf.xpt > GFSYMTYP > 'immunoglobulin gene' value not found in 'SYMTPGPF' extensible codelist)

MetaData

Row No.	Study ID	DOMAIN	GFRPID	GFTSTCD	GFTEST	GFTSTDL	GFCA	GFSYM	GFSYMTYP	GFNAM	GFSP	GFMETHOD
4	Study05	GF	CARDIONEXT	MRD	Measurable	MRD 10^5	B-CELL TR...	IGHD	immunoglob...	AGC	DNA	NEXT GEN...
6	Study05	GF	PCR	MRD	Measurable	MRD 10^5	B-CELL TR...	IGHD	immunoglob...	PCR		NEXT GEN...

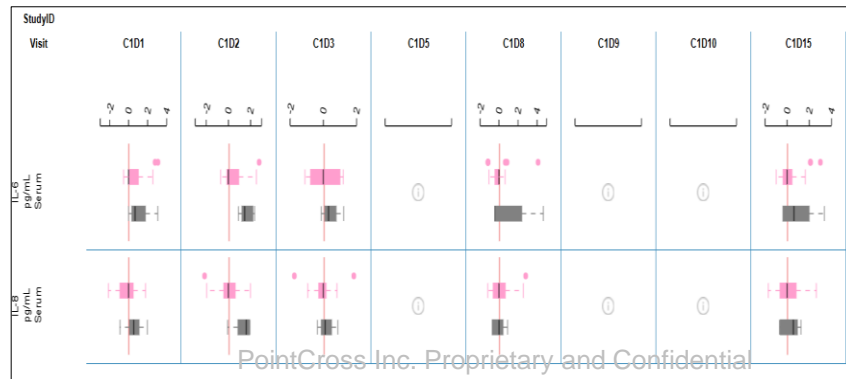
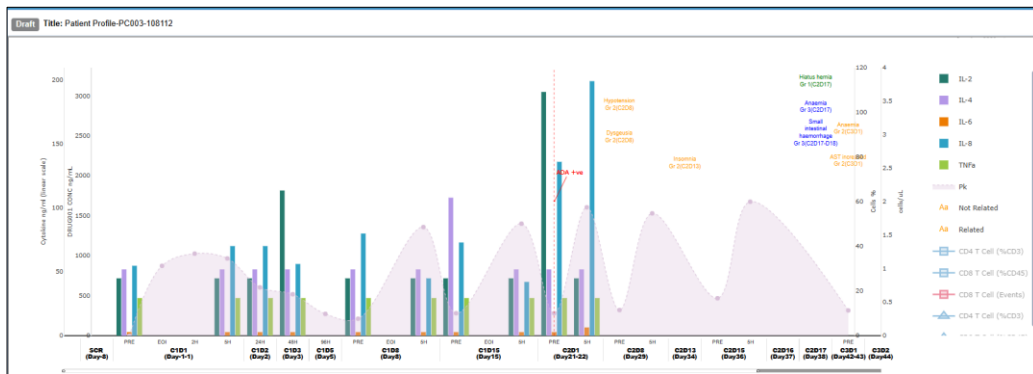
	Study ID	Domain	Publisher ID	Rule ID	
<input type="checkbox"/>		GF	CG0014, CG0208, 12, 257	SD0002	NULL value in varia
<input type="checkbox"/>		GF	CG0014, CG0313, CG0554	SD0056	SDTM Required va
<input type="checkbox"/>		GF	FDAB017	CT2002	'GENE WITH PROTEIN PRODUCT;immunoglobulin gene;GENE WITH PROTEIN PRODUCT;C
<input type="checkbox"/>		GF	FDAB017	CT2002	'MRD 10^5' value not found in 'GFTSTDL' extensible codelist
<input type="checkbox"/>		GF	FDAB017	CT2002	'MRD' value not found in 'GFTSTCD' extensible codelist
<input type="checkbox"/>		GF	FDAB017	CT2002	'Measurable Residual Disease' value not found in 'GFTEST' extensible codelist
<input type="checkbox"/>		GF	FDAB017	CT2002	'Per Million Count Lower 95th Percentile' value not found in 'GFTSTDL' extensible codelist
<input type="checkbox"/>		GF	FDAB017	CT2002	'immunoglobulin gene' value not found in 'SYMTPGPF' extensible codelist
<input type="checkbox"/>		GF		PCBM25	Expected variables are NULL
<input type="checkbox"/>		GF		PCBM25	Expected variables are NULL
<input type="checkbox"/>		GF		PCBM25	Expected variables are NULL
<input type="checkbox"/>		GF		PCBM25	Expected variables are NULL
<input type="checkbox"/>		GF		PCBM25	Expected variables are NULL



Genomic Finding : Mapping to CDISC Standards

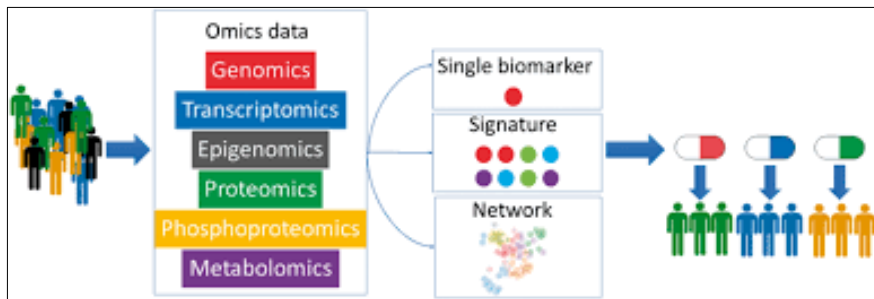
Subject Identifier	Sample ID	Sample Category	Genetic Material Sample	Visit	Assay Method	Assay Lab/ Assay Name	Assay Lab Analyte Name	Analyte Result	DNA sequence change	Protein description	Variant Allele Frequency	Reference sequence
S12345	R001	SKIN TISSUE	RNA	D1	TaqMan Low-Density Array (TLDA)	Transcript Lab/RNA Analysis	MMP12	CT	c.442C>T (p.Arg148C)	Matrix Metalloproteinase-12 (MMP-12)	0.002	NM_002425.4 (for MMP12)
S12346	D002	BLOOD	DNA	D2	Targeted Genome Sequencing	Genomics Lab/DNA Sequencing	ALPK3	Detected	c.1300G>A (p.Val434A)	Alkaline Phosphatase 3 (ALPP)	0.01	NM_003770.3 (for ALPK3)

Domain Abbreviation	Ref.ID	Sequence Number	Group ID	Specimen Material	Method of Test or Examination	Category for Genomic Finding	Short Name of Genomic Measurement	Name of Genomic Measurement	Measurement Test or Examination Detail	Genomic Symbol	Genomic Symbol Type	Results or Findings in original Units	Sequence Identifier	Visit
Domain	GFREFID	GFSEQ	GFGRPID	GFSPC	GFMETHOD	GFCAT	GFTESTCD	GFTEST	GFTSTDTL	GFSYM	GFSYMTYP	GFORRES	GFSEQID	VISIT
GF	R001	1	1	RNA	REVERSE TRANSCRIPTASE PCR	Gene Expression	TRNSCPTN	Transcription	GENETIC TRANSCRIPTION INDICATOR	MMP12	GENE WITH PROTEIN PRODUCT	Ct	NM_002425.4	D1
GF	D002	2	2	DNA	TARGETED GENOME SEQUENCING	Genetic Variation	SNV	Single Nucleotide Variation	PREDICTED CODING SEQUENCE CHANGE	ALPK3	GENE WITH PROTEIN PRODUCT		NM_003770.3	D2





Harmonizing the Complexity



Centralized, Version Controlled Standards Repository (MDR)

Intelligent Classification System/Agile CT Management

AI/ML based data mapping/transformations

Automated Validation/Quality Control

Unified Repository for Patient data

Visualization tools

Thank You!



Mohit Sehgal, PhD
mohits12@hexaware.com



Dr. Sapna Chandran
sapna@pointcross.com

