PharmaSUG China 2023 - Paper DS-152

Exploration on Cell Phenotype Findings (CP) Domain to Report Cell Marker Expression Data for Oncology Studies

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

ABSTRACT

Cell Phenotype Findings (CP) Domain has been introduced in SDTM IG v3.4 as a new specimen-based findings domain to capture cell phenotyping component based on cell expression markers and other indicators in disseminated tissue specimens and cell suspensions. It is usually used to identify or quantitate a specific population of cells. As a new domain, to explore the scope and appropriate use is very meaningful. In this paper, we would like to share our experience on implementing CP domain, introduce the related new SDTM variables and its definitions and difficulties in the application process, and explore the flexible use to report cell marker expression data for oncology studies.

BIOMARKER AND THE SIGNIFICANCE IN CLINICAL DEVELOPMENT

The FDA National Institutes of Health Biomarkers Definitions Working Group defines a biomarker as “A characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions”. A biomarker can be a gene, molecule, enzyme, specific cell or a protein. In the field of oncology, biomarkers are usually produced by tumor cells or immune cells in the body and reflect the change of biological matter. Biomarkers can be subdivided into susceptibility/risk biomarker, diagnostic biomarker, monitoring biomarker, prognostic biomarker, predictive biomarker, pharmacodynamic/response biomarker and safety biomarker (FDA/NIH BEST (FDA-NIH Biomarker Working Group, 2021).

Biomarkers can play an important role at the stages in drug development process from pre-clinical development to post-market safety surveillance. During the pre-clinical process, biomarkers can be used as a prognostic indicator or surrogate endpoints to help with the target selection and assessment of the toxicity of the drug. During clinical trials and post-market studies, biomarkers are used to assess the risk-benefit profile of the study medication. Besides, the new biomarker can be used to explore the effects of disease on molecular pathways and to find out the mechanism of actions of the drugs (Kiran Cherukuri, 2016; Prakash Dev, 2020).

Table 1. Biomarker Classification (By Methods)

Biomarker Type	Subtypes	Example	Detection Method
Genetic/Genomic	Gene mutation	EGFR mutations in NSCLC	PCR, FISH, Sequencing, DNA chip, etc.
	Gene copy number	HER2 amplification in Breast Cancer	

	Translocations	BCL-6 translocation in FL	
Protein	Tissue/cell protein	HER2 IHC status and staining intensity in Breast Cancer	Immunohistochemistry (IHC), Multiplex IHC, Western Blot, immunoassay, LC- MS, etc.
	Blood/fluids protein	Various	
Cells & Tissues	Cells	Circulating tumor cells (CTC) numbers in Breast Cancer	Flow cytometry, CTC, PET, CT, etc.
	Others	Various	

WHAT IS FLOW CYTOMETRY

Flow Cytometry is an analytical technique used to differentiate and discriminate between multiple different physical or chemical properties of cells by passing cells single file through a laser. The light from the laser can scatter differently based on the physical and chemical properties of the cell. The light scatter is recorded by sensors and render into graphical data by a computer. Filtering the cells one at a time allows for researchers to differentiate between many different cells within a heterogenous cellular population (Cheeky scientist, 2016). Besides, cells can be fluorescently tagged, which can create unique light scatter when passing through the laser corresponding to the fluorescent tag used. Using this technique, researchers can employ fluorescently conjugated antibodies that binds to specific cellular target receptors to gain insight into receptor activity for drug development.

WHY WE USE CP DOMIAN?

How to report marker expression data is very important in clinical research. At 26May2015, CDISC released SDTMIG-PGx provisional version 1.0, which was intended to provide guidance on the implementation of the SDTM for biomarker-related data by Pharmacogenomics/Genetics Biomarker (PB) domain and Subject Biomarker (SB) domain. PB domain was independent of data about study subjects, and it was used to relate genetic variations to the inference about those genetic variations (i.e., a medical statement). The SB domain provided the linkage between genetic findings (e.g., genetic mutation, viral genetics) for a study subject recorded in the PF domain and clinical statements defined within the PB domain. Both PB domain and SB domain were only used for genetic and genomic data, and they employed a special structure and did not belong to any of the CDISC three observational classes of findings, events or interventions. Besides, SDTMIG-PGx standed on its own like SDTMIG for Medical Devices (SDTMIG-MD), did not have the biomarker-related domains being incorporated into either SDTMIG or SENDING. SDTMIG-PGx v1.0 was deprecated and much of its content had been subsumed into SDTMIG v3.4. For example, the provisional PF domain had been deprecated and superseded by the GF domain; the provisional PG, PB, SB domains had been deprecated.

Cell Phenotype Findings (CP) Domain was introduced in SDTM IG v3.4 as a new specimen-based findings domain to report marker data at 29Nov2021. CP domain includes but is not limited to genetic and genomic biomarker data and it also support tests associated with the characterization of cell phenotype, lineage, and function based on expression of specific markers in single cell or particle suspensions. It is a finding domain meets CDISC data structure and many new variables and concepts has been added to support the report of marker data.

DISTINGUISH CP DOMAIN VS. OTHER SPECIMEN-BASED FINDINGS DOMAINS

Individual domains (e.g., CP, GF, LB, MI, IS, MB, MS) for laboratory measurements or examinations performed on collected biological specimens (e.g., blood, urine, bone marrow aspirates, tumor tissue, disseminated tissue specimens) are classified into specimen-based findings domains section of SDTMIG v3.4. The distinction and relationship between CP domain and other specimen-based findings domains are summarized in Figure 1 (Xuejie Zhou, 2022).

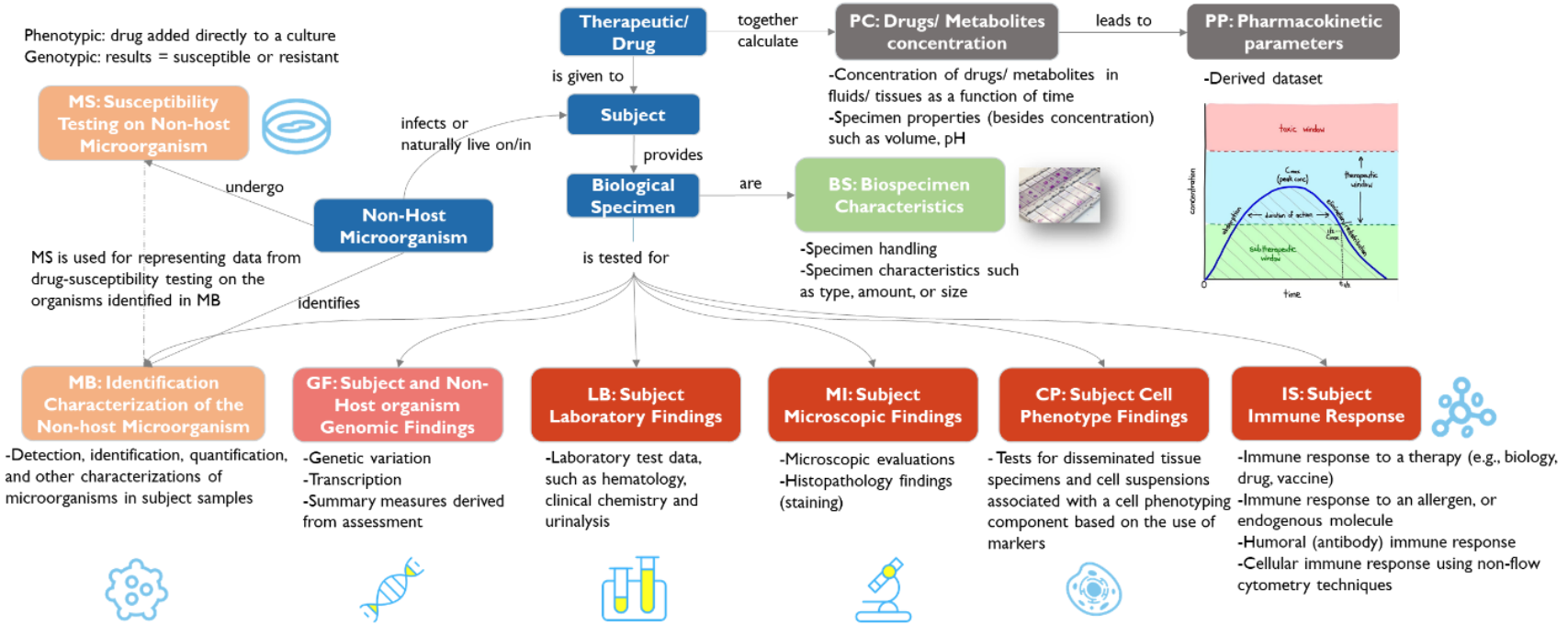

Figure 1. Specimen-based Findings Domains Summary

CP domain is aimed to support tests based on phenotyping markers in disseminated tissue specimens (e.g., blood and other body fluids, bone marrow aspirates) and cell suspensions measured by flow cytometry and is not currently applied for solid tissue specimens. It can be used in the following cases: cell populations or subpopulations identified or classified based on the expression of markers, the level of marker expression quantified, other cell properties based on characterization of expression marker(s) and/or substances that interact with the markers (e.g., target/receptor occupancy).

CP vs LB, GF, MI, IS

Other domains tested from biology specimen (LB domain, GF domain, MI domain and IS domain) represent findings about laboratory examination of fluid specimens, microscopic examination of tissue samples, nucleic acids or subsequent inferences and/ or predictions about related proteins/ amino acids, 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. GF domain focuses on genomic findings from subject or non-host organism, such as assessments and results for gene variation, chromosomes variation, gene rearrangement. Microscopic examination data in MI domain include histopathology findings and microscopic evaluations, such as the assessment of HER2 receptor protein amount using Immunohistochemistry (IHC) method with stain. 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.

Since tests in CP domain are associated with marker-based phenotyping, it cannot replace LB domain for routine lab hematology (e.g., blood cell differentials), and MI domain from microscopic assessment of cells, including those employing immunohistochemical (IHC) techniques.

CP 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 CP domain represents cell marker information, PC domain and PP domain focus on the study of the time course of drug absorption, distribution, metabolism, and excretion.

CP vs BS, 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. For example, if 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.

HOW TO INTERPRET CELL MARKER EXPRESSON DATA IN CP DOMAIN?

To meet the flexible needs to report cell marker expression data, which can range widely in complexity, many new SDTM variables have been created in SDTMIG v3.4 to interpret CP domain data. CPTEST and CPTESTCD are required variables and used to describe and classify the cell population or phenotype test name (e.g., Leukocytes, Monocytes, B-Lymphocytes). CPMRKSTR is expected variable and used to keep the full marker string information. When the marker information is unclear for same cases, CPTEST will be use alone and CPMRKSTR must still be included in the dataset as a null column. CPCELSTA, CPCSMRKS and CPSBMRKS are permissible variables and used to further subdivide the cell or subtype cell population into more detailed subpopulations based on more additional markers, and it can be not provided in a CP domain dataset when sponsor or vendor cannot provide the corresponding information. The combination of CPTEST, CPSBMRKS, CPCELSTA and CPCSMRKS is used to uniquely identify a test. In this section, we summarize the usage of following key variables in CP domain based on SDTMIG v3.4 and present our understanding through some examples:

Table 2. Some key Variables for Cell Phenotype Findings (CP) Domain

Variable Name	Variable Label	CDISC Codelist	CDISC Notes	Core
CPTESTCD	Test or Examination Short Name	(CPTESTCD)	Short name of the measurement, test, or examination described in CPTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in CPTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g., "1TEST" is not valid). CPTESTCD cannot contain characters other than letters, numbers, or underscores. Examples: "MONO", "MNS".	Req
CPTEST	Name of Measurement, Test or Examination	(CPTEST)	Long name for CPTESTCD. For cell phenotyping, the name (often abbreviated) of the cell population, as it is generally accepted by the scientific community, is populated (rather than a colloquial designation based on a primary marker, e.g., TLym Help rather than CD4). When the test is for a sublineage which can only be identified by specifying additional markers (i.e., has not been given a name) or which is further restricted to a subpopulation based on a particular cell state (e.g., activated, proliferating, apoptotic), the Sublineage Marker String (CPSBMRKS), Cell State (CPCELSTA), and Cell State Marker String (CPCSMRKS) variables are additionally populated and the value in CPTEST is suffixed with "Sub" to denote that it is a subset of the	Req

			population identified in CPTEST (e.g., Monocytes Sub). The value in CPTEST cannot be longer than 40 characters.	
CPSBMRKS	Sublineage Marker String		Used to further subset the cell population identified in CPTEST based on the use of additional marker(s) that define a sublineage. The value in CPSBMRKS is used in combination with values in CPTEST and CPCELSTA to fully describe the cell population being measured. As such, it is an essential component of the full test name. For example, three unnamed sublineages of monocytes have been identified as: CCR2+CD16-, CCR2- CD16+, and CCR2+CD16+. Whereas the entire monocyte cell population can be defined as CD14+cells, the additional CCR2 and CD16 markers are used to differentiate one sublineage from another. As none of these sublineages have been given names, they are only known by the CCR2 and CD16 marker combinations. By associating the CPTEST value of "Monocytes Sub" with, for example, a value of "CCR2+CD16-" in CPSBMRKS, the full test is defined to be the CCR2+CD16- monocyte subpopulation.	Perm
CPCELSTA	Cell State	(CELSTATE)	A textual description of a subset of the cell population identified in CPTEST based on a particular functional and/or biological state (e.g., "ACTIVATED", "PROLIFERATING", "SENESCENT"). When populated, the values in CPCELSTA and CPSMRKS, in combination with the values in CPTEST and CPSBMRKS, fully describe the cell population being measured.	Perm
CPCSMRKS	Cell State Marker String		Identifies the marker(s) or indicator(s) used to define the cell state (i.e., the value in CPCELSTA). For example, when Ki67 expression is used to determine that a cell population is in a proliferating state (i.e., CPCELSTA value="PROLIFERATING"), the value "Ki67+" in CPCSMRKS indicates that positive expression of Ki67 was used to define the population as proliferating. Similarly, a value of "Ki67-" in CPCSMRKS would indicate that lack of expression of Ki67 defined the "NON-PROLIFERATING" cell state in CPCELSTA. The CPCSMRKS value is useful for quickly determining which marker(s) were used to classify (i.e., operationally define) a cell population based on a functional/biological state.	Perm

CPMRKSTR	Marker String		The text string identifying the full set of markers/indicators used by the laboratory to operationally define the complete test based on the combination of CPTEST, CPSBMRKS, and CPCELSTA. Because laboratories often use different markers/indicators to identify a cell population, the relationship between a named cell population in CPTEST (as combined with CPSBMRKS and CPCELSTA values) and the set of markers used to identify that population is many-to-one. To ensure nuances important for accurately interpreting the data are accounted for and which arise from the use of different sets of markers, it is necessary to operationally define the test in terms of the complete set of markers/indicators used to perform that test.	Exp
CPCAT	Category	(CPCAT)	Used to define a category of topic-variable values across subjects. Examples: "IMMUNOPHENOTYPING", "CELL FUNCTION", "TARGET ENGAGEMENT".	Perm

Note: CDISC Codelist is based on SDTM Terminology 2023-03-31.

Per SDTMIG v3.4 and CDISC rules for CP domain, here are the basic usage for the variables:

1. CPTEST, CPTESTCD and CPCAT

CPTESTCD and CPTEST are populated with the name of the cell population type being measured. Several of the new variables (i.e., CPSBMRKS, CPCELSTA, CPCSMRKS) are used to further subdivide or describe the population reported in CPTEST/CPTESTCD into more subpopulations based on 1 or more additional markers. This approach provides a more easily understood and standardized cell phenotype test name and enables development of controlled terminology for CPTEST and CPTESTCD. CPCAT represents the supported category of CPTEST/CPTESTCD with Immunophenotyping, cell function and target engagement (e.g., target/receptor occupancy).

For the values of CPTESTCD, CPTEST and CPCAT, it is suggested to check with CDISC codelist at first and select the appropriate values.

2. CPCSMRKS, CPSBMRKS and CPMRKSTR

CPCSMRKS is useful for determining which marker(s) are used to identify the cell state (CPCELSTA) based on a functional/biological state. CPSBMRKS is applied to further subset the cell population identified in CPTEST based on additional marker(s). The value in CPSBMRKS is used in combination with values in CPTEST and CPCELSTA to fully describe the cell population being measured. CPMRKSTR represents the full marker sting information and reflects the laboratory- specific definition of the test measurement.

3. CPCELSTA and CPCSMRKS

CPCELSTA is used in conjunction with CPCSMRKS. When CPCELSTA is populated, CPCSMRKS must also be populated. Conversely, when CPCSMRKS is populated, CPCELSTA must be populated.

It is recommended that CPCELSTA and CPCSMRKS generally not be used unless a selective viability stain is included in the test to differentiate the record for viable cells from record(s) for cells in a different vital state.

EXAMPLE OF CELL IMMUNOPHENOTYPING

In this example, CPTEST is populated with the name of the cell type being measured, not with the set of markers used to define the cell type. It is expected that CPMRKSTR is populated, and that it contains all markers used to define the test, including those that are also present in CPSBMRKS and/or CPCSMRKS. When 1 or more of the variables CPSBMRKS, CPCELSTA, or CPCSMRKS are populated, the test name (CPTEST) must be populated with the test name variant containing the "Sub" suffix to indicate that the finding/result pertains to a subpopulation of the cell type named in CPTEST.

Example	CPTESTCD	CPTEST	CPCELSTA	CPCSMRKS	CPMRKSTR	CPCAT	CPORRES	CPORRESU
(CD45+) Lymphocytes count	LYM	Lymphocytes			CD45+	IMMUNOPHENOTYPING	1000	10^6/L
Abnormal (CD19+) B-Lymphocytes Cell Count	BLYCE	B-Lymphocytes			CD45+CD19+	IMMUNOPHENOTYPING	100	10^6/L
Abnormal (CD19+) B-Lymphocytes/ CD45+ Lymphocytes	BLYCELY	B-Lymphocytes/ Lymphocytes			CD45+CD19+/ CD45+	IMMUNOPHENOTYPING	10	%
(CD45+) Lymphocytes count for viable cells	LYS	Lym Sub	VIABLE	7AAD-	CD45+7AAD-	IMMUNOPHENOTYPING	200	10^6/L
(CD45+) Lymphocytes sub-population count for non-viable cells	LYS	Lym Sub	NON-VIABLE	7AAD+	CD45+7AAD+	IMMUNOPHENOTYPING	800	10^6/L

EXAMPLE OF CELL FUNCTION

This example shows data from a CCR5 cytotoxic T helper assay. CPTSTCND is used to identify the test condition applied to the assay and CPCNDAGT is used to impose the condition identified in CPTSTCND. CPCELSTA and CPCSMRKS indicate that it is activated T helper and the

cell state is measured by pCCR5+ marker. CPGATE contains a full gate definition or gating strategy used to virtually isolate the cell population of interest based on a set of defined characteristics (e.g., presence, absence, or intensity of expression of various markers; physical size; internal complexity or granularity) and CPGATEDEF contains the gate text string definition or gating strategy.

CPTESTCD	CPTEST	CPCELSTA	CPCSMRKS	CPTSTCND	CPCNDAGT	CPMRKSTR	CPGATE	CPGATDEF	CPCAT	CPORRES
TLC	TLym Cytx	ACTIVATED	pCCR5+	WITHOUT STIMULATING AGENT		CD3+CD8+pCCR5+	TLym Cytx	CD3+CD8+	CELL FUNCTION	10
TLC	TLym Cytx	ACTIVATED	pCCR5+	WITH STIMULATING AGENT	MACROPHAGE INFLAMMATORY PROTEIN 1BETA	CD3+CD8+pCCR5+	TLym Cytx	CD3+CD8+	CELL FUNCTION	15

EXAMPLE OF TARGET ENGAGEMENT

Receptor Occupancy (RO) assays provide a means to measure the interaction of therapeutics with their cell surface targets, and it is commonly performed by flow cytometry using fresh whole blood specimens. There are three common methods for measuring receptor occupancy: drug free receptor assay, drug occupied/bound receptor assay, and total receptor assay. Free receptor assay measures the proportion of cell surface receptors not bound by the drug. This is accomplished by fluorescently tagging the drug itself, a competitive antibody, or the receptor ligand. Drug occupied receptor assay measures the proportion of drug occupied receptors by introducing a fluorescently tagged antidrug that does not compete with the drug binding receptor. This antidrug will bind to the drug itself after the drug has bound to the receptor, allowing us to determine how many receptors have been bound by the drug. Total receptor assays measures both free and bound receptors by using a fluorescently tagged non-competing antibody (Inna Vainshtein, 2016).

This is a target engagement example. The interaction of cell marker CDxx cell-associated target protein expressed on cytotoxic T-lymphocytes (defined as CD45+CD3+CD8+CDxx+) and the therapeutic antibody (CDxx antibody) is assessed by the proportion of bound and total target/report occupancy. CPMRKSTR, CPGATE, and CPGATEDEF are used to identify the cell population on which the binding of interest is measured. CPBDAGNT identifies the CDxx antibody (therapeutic antibody) as the target-bound substance of interest. CPTSTCND is used to indicate the conducted condition of CDxx antibody. CPSPTSTD is the test description used by a sponsor.

CPTESTCD	CPTEST	CPTSTCND	CPBDAGNT	CPMRKSTR	CPGATE	CPGATDEF	CPSPTSTD	CPCAT	CPSCAT	CPORRES
TGB	Target Bound		ANTI-CDxx ANTIBODY	CDxx MdFI CD45+CD3+ CD8+CDxx+	TLym Cytx	FSC SSC|CD45+| CD3+CD8+	CDxx Bound Expression MdFI T-Lymphocytes Cytotoxic Rep 1	TARGET ENGAGEMENT	RECEPTOR OCCUPANCY	28

TGB	Target Bound	SATURATED CONDITION WITH BINDING AGENT	ANTI-CDxx ANTIBODY	CDxx MdFI CD45+CD3+ CD8+CDxx+	TLym Cytx	FSC SSC|CD45+| CD3+CD8+	CDxx Bound Expression MdFI T-Lymphocytes Cytotoxic Rep 1	TARGET ENGAGEMENT	RECEPTOR OCCUPANCY	160

CONCLUSION

As a quite newly introduced domain, CP domain is closely related to biomarkers, which are extensive, complex and unintelligible. To report cell marker data flexibility, several new SDTM variables have been created, but most of the new variables are permissible and don't have CDISC Controlled Terminology codelists. We hope our experience in implementing this domain can provide extra helpful information beyond SDTMIG v3.4. However, there are some limitations in this paper due to limited knowledge. With the more exploratory application and the ongoing evolvement of CP 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

FDA-NIH Biomarker Working Group. "BEST (Biomarkers, EndpointS, and other Tools) Resource". November 29, 2021. Available at https://www.ncbi.nlm.nih.gov/books/NBK326791/pdf/Bookshelf_NBK326791.pdf.

Kiran Cherukuri, Seattle Genetics, Inc., Bothell, WA. Transforming Biomarker Data into an SDTM based Dataset. PharmaSUG US 2016, DS15. Available at https://www.lexjansen.com/pharmasug/2016/DS/PharmaSUG-2016-DS15.pdf.

Prakash Dev, Covance, Durham, NC. Integration of Biomarkers Data in Clinical Trials. PharmaSUG US 2020, TT04. Available at https://www.lexjansen.com/phuse-us/2020/tt/TT04.pdf.

Inna Vainshtein, Amy K. Schneider. Multiplexing of Receptor Occupancy Measurements for Pharmacodynamic Biomarker Assessment of Biopharmaceuticals. Cytometry Part B (Clinical Cytometry) 90B:128–140 (2016). Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5057311/pdf/CYTO-90-128.pdf.

Xuejie Zhou, Liming Xie, Qiuan Hu, Shanshan Ma. Time to Get in the Genomics Findings (GF) Domain. PharmaSUG China 2022, SR145. Available at https://www.lexjansen.com/pharmasug-cn/2022/SR/Pharmasug-China-2022-SR145.pdf.

Cheekscientist blog "Measuring Receptor Occupancy with Flow Cytometry". 2016. Available at https://expert.cheekyscientist.com/measuring-receptor-occupancy-with-flow-cytometry.

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 v1.0*

CONTACT INFORMATION

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

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

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

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

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