Standardization of Flow Cytometry Assay Data

Introduction

Flow Cytometry is used by specialty labs that support immunology, pathology or molecular biology investigations. This technology is key to Cell Phenotyping. These specialty labs may be contracted by sponsors conducting research or development of therapeutics or vaccines, or by health care providers who are treating patients.

When these assays are conducted for the purpose of submissions of study data to the regulatory agencies such as FDA, PMDA or EMA, very stringent conditions must be met for ensuring quality standards, repeatability of results from the technology used, documentation by the labs about their processes and the assays conducted. In these cases, only biomarkers recognized by the regulatory agencies as having a body published results about their value as a repeatable biological maker are acceptable. Further, the standardization in the form of their lineage must be accepted and published by the FDA or by CDISC. The labs conducting such assays must be CLIA certified, and they are subject to periodic audits, and they must have documented procedures and calibrated lab systems that meet stringent standards.

Since development of biologics and immunology-based therapies and vaccines are burgeoning, exploratory research is giving rise to many new or recent biomarkers and non-CLIA labs are used for these assays. In these cases, there is value in standardizing the assays using the reportables by the labs, although maybe not to FDA acceptable marker lineages, but at least to standardize them to various global registries that are cataloguing these newly discovered markers .

Submittable markers on patient's cell phenotyping assays need to be standardized to CDISC format. These are published periodically and standardization to these published marker lineages is necessary to include the patient data into the submission package. Since the reported markers and their lineages can vary considerably from lab to lab and based on the gating strategies used by the specialty labs, standardization begins with developing a transformation or interpretation of the reportables and their conventions as used by the lab to a global standard that may be a combination of global registries with published markers and scientifically recognized terminologies as well as CDISC published lineages. Finally, those markers that can be standardized to the CDISC marker form can be completed for submission preparation.

The development of source reportables to the global standard is an essential process that may call for some interaction with the labs, or the biomarker leads who contract the labs. Once this is completed, it is possible to re-use these transformations for the actual patient assay data.

Challenges in Standardization

Standardizing flow cytometry data to the Cell Phenotyping domain of SDTM can be challenging because flow cytometry data sourced from the lab may need to be interpreted before standardization can begin. This requires subject matter knowledge and the ability to look up research publications and global registries. The follow-on process of standardizing the markers to CDISC or regulatory agency acceptable form is complex and requires knowledge of CDISC standards and SDTM as well as the state of the frequently published and re-released lists of lineage standards.

The most common challenges faced are

- Interpreting the lab/vendor-provided source reportables or operational definition of test to global or CDISC terminologies.
- Identifying and representing key variables, including marker strings, sub-lineages, and cell states.
- Understanding and implementing various gating strategies used by the labs.
- Determining additional metadata based on raw event data (e.g., absolute cell counts) or derived populations (e.g., percentages, median fluorescence intensities).
- Adhering to SDTM rules and guidelines for the representation of different variables.
- Collaborating with laboratory data providers & scientists to identify the appropriate variables to include in the dataset.

SDTM Guidelines & Rules for Data representation

In the CP domain, interpreting flow cytometry data requires the variables **CPTEST** and **CPTESTCD** to describe the cell population. **CPMRKSTR** is an expected variable that provides the full marker string information which are used to identify and characterize different cell populations (CD3: A marker for T cells, CD19: A marker for B cells). CPSBMRKS (Cell Phenotype Subcellular Markers) captures markers that are found within the cell, such as intracellular proteins or other subcellular components (Granzyme B: An intracellular marker for cytotoxic T cells and NK cells)The variables **CPCELSTA ()**, **CPCSMRKS**, are used to further characterize the cell or subtype cell based on the functional state markers (Ki-67: A marker for cell proliferation).

The combination of **CPTEST**, **CPSBMRKS**, **CPCELSTA**, and **CPCSMRKS** is used to uniquely identify a test/measurement.

These standardized marker strings help ensure consistency and clarity in cell phenotyping data, facilitating better comparison and interpretation across different studies and regulatory submissions.

Furthermore, the process of data standardization should adhere to the guidelines and rules for accurate data representation.

PointCross Offering in Cell Phenotyping Data Standardization

PointCross has Xbiom, a solutions platform that covers various aspects of clinical and nonclinical study data management from curation, to standardization, analysis, and bio-statistical analysis. A key component within Xbiom is its metadata repository (MDR) used to ensure that that study data model transformation to exchange standards, such as, SDTM are automated and managed using curated references for standards and terminology. The Cell Phenotyping markers are part of this capability.

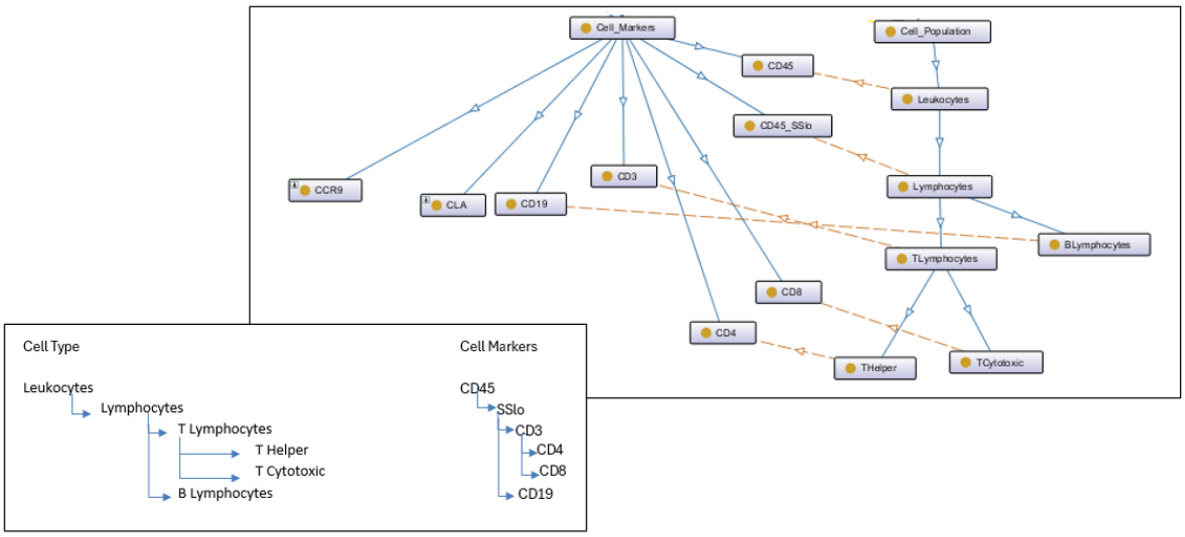
Xbiom Software as a Service (SaaS): Clients with large numbers of studies and patient data can subscribe to Xbiom under a SaaS agreement and training is provided for their biomarker teams to use the standardization modelling or actual standardization of patient data for submissions.

Xbiom Data as a Service (DaaS): Clients may also subscribe to the Data as a Service provided by PointCross where the Cell Phenotyping assays are standardized for patient assays or a transformation model to standardize reportable from a lab for an assay type is performed as a service.

Xbiom MDR

Xbiom MDR along with Xbiom Smart Transformation modules provides all the necessary tools, biomarker registries and CDISC CT and IG registries to standardize the marker strings from flow cytometry assays executed by different vendors and by multiple test panels.

Fig 1: Hierarchical structure of Cell Populations in MDR



The cell lineage of each population in Xbiom MDR, along with the markers that distinguish these cells, is defined and organized in a hierarchical structure. This framework represents the parent cell population, subpopulations, and functional states for each cell type.

Standardization Process Supported by Xbiom, or used by PointCross DaaS

The vendor-provided definitions imported into Xbiom undergo a comprehensive validation process. This includes: (i) an independent check of the reportable/source marker string against lexicons derived from known or previously standardized marker lists, (ii) verification of the syntactic accuracy and proper arrangement of markers and symbols, and (iii) validation of semantic accuracy to ensure the marker text conveys a meaningful biomarker or expression order.

Marker strings that successfully pass these validation steps are then advanced to the next phase, where they are reviewed for standardization recommendations to global registries or global Controlled Terminology (CT). Workflows are provided to allow interactions and queries to the laboratory, or the biomarker leads on source data that cannot be validated without further clarification.

During the standardization recommendation phase, each marker in the string is classified as a parent, child, or functional state and standardized accordingly.

The process involves identifying

1. Parent population – Leuk, B cell, T cell
2. Sub population – Helper , Cytotoxic, Monocytic , Granulocytic
3. Location /Origin – Germinal , Marginal , Peripheral
4. Structure – Alpha – beta , Gamma delta…

5. Functional stage – Mature , Immature , Functional
6. Sub lineages – Additional marker , expression
7. Cell state – Activated , Apoptotic , Senescent ,Proliferating, Viable…

Example: The string below will be analysed and broken down to identify the population.

"CD45+CD3+CD19-CD4+CD8-CD197+CD45RA-CD278+Ki67+7AAD-"

CD45 (Leukocyte)

CD3+CD19- (T lymphocyte)

CD4+CD8- (Helper)

CD197+CD45RA- (Central Memory)

CD278+ (Activated)

Ki67+ (Proliferating)

7AAD- (Viable)

The standardization process constructs variables according to the guidelines needed for CP dataset generation. Based on the identified population and sub-population, the appropriate CDISC term is mapped to CPTEST (Name of Measurement, Test or Examination). CPMRKSTR (Marker String) is created by tracing the cell lineage from parent to child populations, considering the type of measurement (absolute, relative, or fluorescence intensity). Sub-population markers are mapped to CPSBMRKS (Sub lineage Marker String), while markers indicating the functional state of the cell population are assigned to the variables CPCELSTA (Cell State) and CPCSMRKS (Cell State Marker String), respectively.

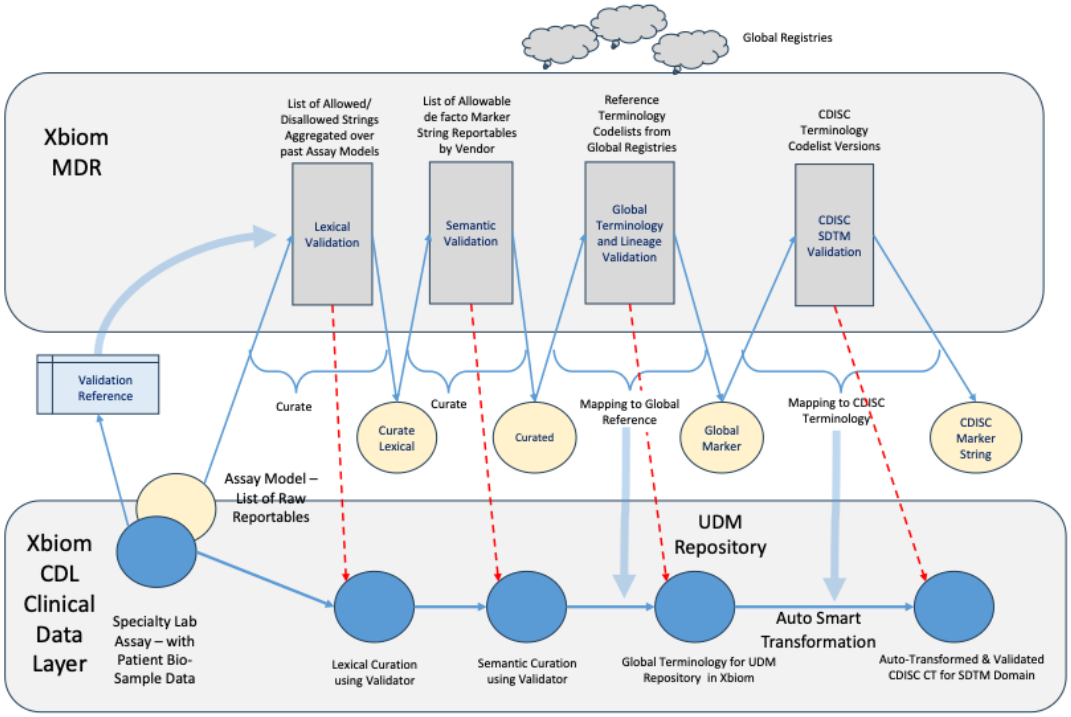
Gating strategy information CPGATE (Gate) and CPGATDEF (Gate Definition) are also an important information to be available in the datasets, which are available part of the assay validation reports which are panel specific, Xbiom facilitates extraction of these data from the reports if available else predicts the possible gating strategy as per the marker string which can be confirmed by the expert teams.

Gateways in Standardization Process

During the standardization process, validation phases for the source marker text or reportable marker may detect specific issues based on the predefined validation configurations. When such issues arise, the marker is flagged with a tailored prompt question to gather input from the expert team, facilitating well-informed decisions on how to proceed to the subsequent steps in the standardization workflow.

In the later stages of standardization, the system may present multiple closely matching recommendations or, in some cases, no recommendation at all. In such scenarios, users are expected to review the options, decide, or flag the issue for collaboration or consultation with other biomarker or standards leads.

These validation gateways ensure a streamlined and efficient process for standardizing the data.



Steps Employed in Recommendations & Standardization of Flow data

1. Vendor provided reportables/test are processed to exclude texts related to reagents/reads.

2. The markers are processed if there are any gaps in the vendor representation (e.g., 3+4+ to CD3+CD4+).

3. In the complete marker string, the markers related to Functional stage, Cell state, Location are looked at and segregated.

4. In the next step of processing the markers, each marker is looked at in the hierarchical definition of cell lineage maintained in Xbiom.

5. From the definition, all possible lineages for the marker are recommended.

6. Based on the identified lineage and the units provided, the cell type measured is decoded to be Parent or subpopulation or relative or expressions.

7. Next step is to construct the Marker string (complete lineage along with expression). The order of markers - Cell hierarchy from highest to lowest, followed by additional non-lineage-defining markers, and ending with cell state and viability markers.

8. Markers related to Cell state are translated (e.g., Activated, Proliferation, etc).

9. When the reportables are identified to be subpopulation, the respective markers are considered as Sub-lineage markers

10. Unit of measurement, Absolute, Relative, Events, Fluorescence intensity are key for recommendations

11. At the final step, the tool aligns the values as per the standard definition and recommends CPTEST, CPTESTCD, CPSBMRKS, CPCELSTA, CPCSMRKS, CPRESTYP, CPRESSCL.

3+4+8-197-45RA-152-Ki67+ ABS

↓*Text exclusions/ Decode the text*

3+4+8-197-45RA-152-Ki67+

↓*Filling gaps*

CD3+CD4+CD8-CD197-CD45RA-CD152-Ki67+

↓ *Identifies Cell State markers if any*

CD3+CD4+CD8-CD197-CD45RA-CD152-

↓ *Identifies Subpopulation markers*

CD3+CD4+CD8-CD197-CD45RA-

↓ *Lineage Predictions*

CD45+CD3+CD4+CD197-CD45RA-

CD45+CD3+CD4+CD127+CD197-CD45RA-

CD45+CD3+CD4+CD195+CD197-CD45RA-

CD45+CD3+CD4+CD29+CD197-CD45RA-

CD45+CD3+CD4+CD196+CD197-CD45RA-

↓ *Identification of Cell populations*

TLym Help Eff Mem ---> CD45+CD3+CD4+CD197-CD45RA-

TLym Help Conv Eff Mem ---> CD45+CD3+CD4+CD127+CD197-CD45RA-

TLym Help 1 Eff Mem ---> CD45+CD3+CD4+CD195+CD197-CD45RA-

TLym Help 2 Eff Mem ---> CD45+CD3+CD4+CD29+CD197-CD45RA-

TLym Help 17 Eff Mem ---> CD45+CD3+CD4+CD196+CD197-CD45RA-

↓ *Tool Recommendations**

Confidence Score	Marker String	Sub Marker String	Cell State Marker	Cell State	Test	Unit	Result Scale	Result Type
0.8	CD45+CD3+CD4+CD197-CD45RA-CD152-KI67+	CD152-	KI67+	Proliferating	TLym Help Eff Mem Sub	10^6/L	QUANTITATIVE	NUMBER CONCENTRATION
0.67	CD45+CD3+CD4+CD127+CD197-CD45RA-CD152-KI67+	CD152-	KI67+	Proliferating	TLym Help Conv Eff Mem Sub	10^6/L	QUANTITATIVE	NUMBER CONCENTRATION
0.67	CD45+CD3+CD4+CD195+CD197-CD45RA-CD152-KI67+	CD152-	KI67+	Proliferating	TLym Help 1 Eff Mem Sub	10^6/L	QUANTITATIVE	NUMBER CONCENTRATION
0.67	CD45+CD3+CD4+CD29+CD197-CD45RA-CD152-KI67+	CD152-	KI67+	Proliferating	TLym Help 2 Eff Mem Sub	10^6/L	QUANTITATIVE	NUMBER CONCENTRATION
0.67	CD45+CD3+CD4+CD196+CD197-CD45RA-CD152-KI67+	CD152-	KI67+	Proliferating	TLym Help 17 Eff Mem Sub	10^6/L	QUANTITATIVE	NUMBER CONCENTRATION

**Top 5 recommendations*

Input sent to Xbiom

Vendor Code	Reportable as defined by vendor	Unit
ABC	3+8+4- ABS	Cells/uL
ABC	CD3+CD4+ ABS	Cells/uL
ABC	CD3+CD4+CD223+CD279+ (%CD4)	%
ABC	3+4- 8+197+45RA+_CD152_BV421_MFI	MFI
ABC	Lin-CD14+HLA-DR-/low #Events	Events
ABC	3+4-8+197+45RA-152-Ki67+ ABS	Cells/uL
ABC	Lin-DR-lowCD11b+CD33+ ABS	Cells/µL
ABC	CD3-CD56brCD16-CD366+ ABS	Cells/uL
ABC	B cells (% TNC)	%
ABC	3+4-8+197-45RA-CD152+(%EMCD8)	%
ABC	Event flag (TumorBcells/Kappa+)	Events

Standardized Output generated from Xbiom

CPTESTCD	CPTEST	CPMRKSTR	CPSBMRKS	CPCELSTA	CPCSMRKS	CPGATE	CPGATDEF	CPSPEC	CPCAT	CPORRESU	CPSTRESU	CPRESSCL	CPRESTYP	CPMETHOD
TLC	TLym Cytx	CD45+CD3+CD8+				TLym Cytx	FSC SSC|Viable|CD45+SSClo|CD3+CD56-|CD4-CD8+	BLOOD	IMMUNOPHENOTYPING	10^6/L	10^6/L	QUANTITATIVE	NUMBER CONCENTRATION	FLOW CYTOMETRY
TLYH	TLym Help	CD45+CD3+CD4+				TLym Help	FSC SSC|Viable|CD45+SSClo|CD3+CD56-|CD4+CD8-	BLOOD	IMMUNOPHENOTYPING	10^6/L	10^6/L	QUANTITATIVE	NUMBER CONCENTRATION	FLOW CYTOMETRY
TLHSP	TLym Help Sub/TLym Help	CD45+CD3+CD4+CD223+CD279+/CD45+CD3+CD4+		ACTIVATED; EXHAUSTED	CD223+;CD279+	TLym Help	FSC SSC|Viable|CD45+SSClo|CD3+CD56-|CD4+CD8-	BLOOD	IMMUNOPHENOTYPING	%	%	QUANTITATIVE	NUMBER FRACTION	FLOW CYTOMETRY
CD152X	CD152 Expression	CD152 MFI CD45+CD3+CD8+CD197+CD45RA+CD152+				TLym Cytx	FSC SSC|Viable|CD45+SSClo|CD3+CD56-|CD4-CD8+|CD197+|CD45RA+	BLOOD	IMMUNOPHENOTYPING	FIU	FIU	QUANTITATIVE	FLUORESCENCE INTENSITY	FLOW CYTOMETRY
MNS	Mono Sub	CD45+CD3-CD19-CD56-CD14+HLADR-Lo	HLADR-Lo			Monocytes, Viable	FSC SSC|LDB-|CD45+SSC|CD3-CD19-|HLADR+CD14lo+	BLOOD	IMMUNOPHENOTYPING	EVENTS	EVENTS	QUANTITATIVE	NUMBER	FLOW CYTOMETRY
TLCCMS	TLym Cytx Cen Mem Sub	CD45+CD3+CD8+CD197+CD45RA-CD152-KI67+	CD152-	PROLIFERATING	KI67+	TLym Help Cen Mem	FSC SSC|Viable|CD45+SSClo|CD3+CD56-|CD4+CD8-|CD197+CD45RA-	BLOOD	IMMUNOPHENOTYPING	10^6/L	10^6/L	QUANTITATIVE	NUMBER CONCENTRATION	FLOW CYTOMETRY
MDSCS	MDSC Sub	CD45+CD3-CD19-CD56-CD11B+CD33+HLADR-Lo	HLADR-Lo			Myeloid-Derived Suppressor Cells	FSC SSC|Viable|CD45+SSClo|CD3-CD56-CD19-|CD11b+|CD33+	BLOOD	IMMUNOPHENOTYPING	10^6/L	10^6/L	QUANTITATIVE	NUMBER CONCENTRATION	FLOW CYTOMETRY
NKS	NK Cells Sub	CD45+CD56+HiCD366+	CD366+			NK Cells	FSC SSC|Viable|CD45+SSClo|CD3-CD56+	BLOOD	IMMUNOPHENOTYPING	10^6/L	10^6/L	QUANTITATIVE	NUMBER CONCENTRATION	FLOW CYTOMETRY
BLYCELE	BLym/Leuk	CD45+CD19+/CD45+				B-Lymphocytes	FSC SSC|CD45+SSClo|CD19+	BLOOD	IMMUNOPHENOTYPING	%	%	QUANTITATIVE	NUMBER FRACTION	FLOW CYTOMETRY
TLCEMSP	TLym Cytx Eff Mem Sub/TLymCEM	CD45+CD3+CD8+CD197-CD45RA-CD152+/CD45+CD3+CD8+CD1		ACTIVATED	CD152+	TLym Cytx Eff Mem	FSC SSC|Viable|CD45+SSClo|CD3+CD56-|CD4-CD8+|CD197-CD45RA-	BLOOD	IMMUNOPHENOTYPING	%	%	QUANTITATIVE	NUMBER FRACTION	FLOW CYTOMETRY
BLYS	BLym Sub	CD45+CD19+KAPPA+	KAPPA+			B-Lymphocytes	FSC SSC|CD45+SSClo|CD19+	BLOOD	IMMUNOPHENOTYPING	EVENTS	EVENTS	QUANTITATIVE	NUMBER	FLOW CYTOMETRY

Data Conformance Rules extended for CP domain in Xbiom eDV

The eDataValidator (eDV) is a validation engine that applies conformance rules published by CDISC,FDA and PMDA as well as additional rules that PointCross developed for ensuring that the standardization is conformant to the published standards. Xbiom's eDV is equipped with enhanced rules to validate input strings from vendors, facilitating efficient standardization, as well as rules to validate outputs generated by Xbiom. 14 additional rules have been developed and included by PointCross. This ensures that the data is accurately represented in compliance with CDISC-published rules and guidelines.

Input string Validation Rules

1. Reportable seems to be invalid, as it does not contain any known cell marker/population.
2. Marker in the reportable has cell expression with indication as both positive and negative.
3. Markers of two different populations are part of the reportable.
4. Parent marker is placed after sub-population marker.
5. The tool has predicted multiple parent populations lineage for the input reportable.
6. Parent population not identified for the marker

Validation Rules for CP output

1. Variable value is missing when its paired variable value is available (Ex: GATE & GATEDEF)
2. CPTEST contains "* Expression" with Sub-lineage marker (SBMRKS)/Cell State marker populated
3. Sub-lineage marker (SBMRKS) contains markers of parent population
4. CPTEST is not tagged to a single marker expression, when the measurement type is of Fluorescence intensity
5. Either Sub-lineage marker (SBMRKS)/Cell State marker (CSMRKS) is expected to be populated for Sub populations
6. Result Type should be defined as per the ORRESU/STRESU
7. When unit is %. Test names should always be relative, and Markers defined for it (/)
8. When unit is describing the Fluorescence intensity (level of expression), Test names should always be Expression, and Markers defined should contain delimiting text is the abbreviation for the unit of measure used to report the level of expression of the quantified marker.