

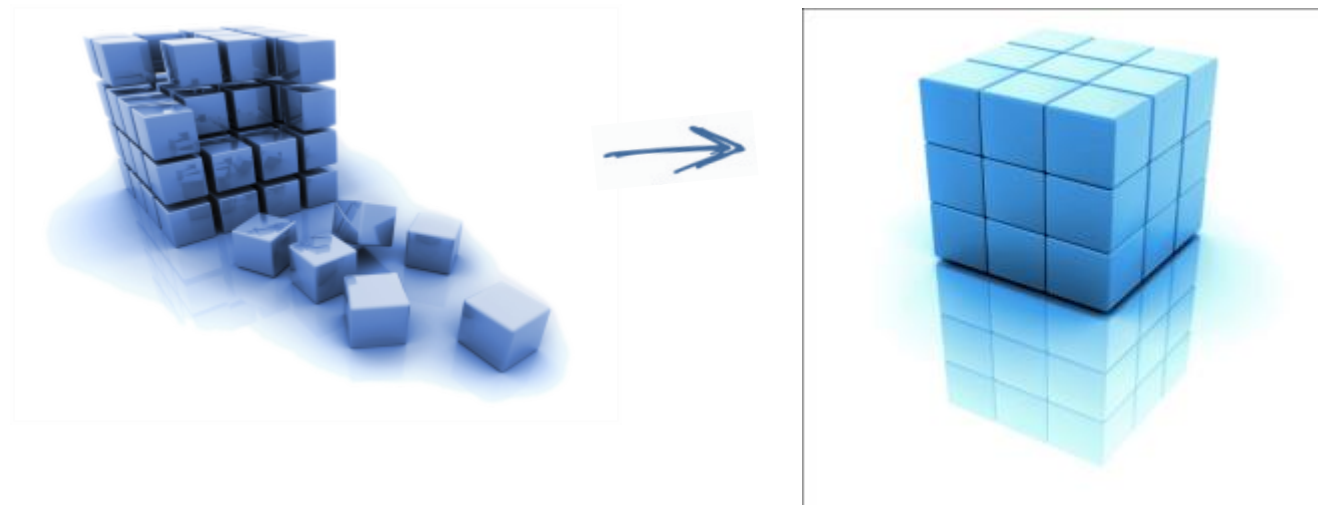
Importance of Semantic Integration of Disparate Clinical Sources Prior to Data Visualization

PhUSE SDE, Raritan, NJ – Clinical Data Visualization

Suresh Madhavan

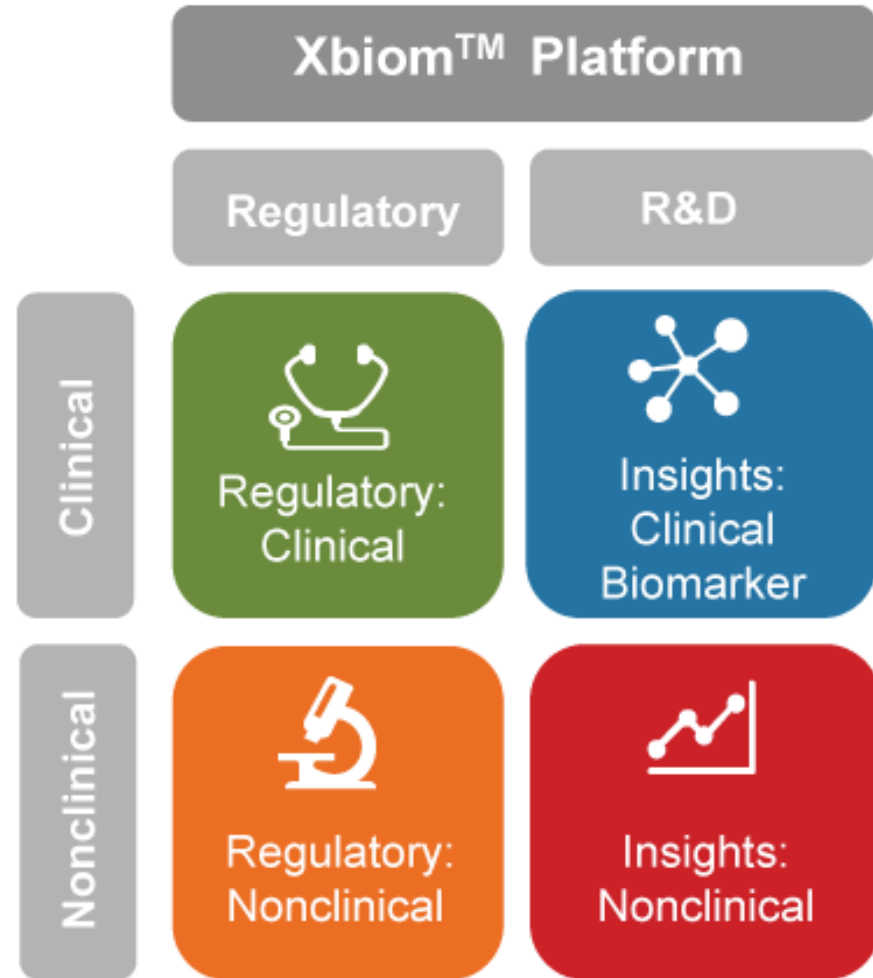
Jerry Zaborowski

June 8, 2017



PointCross Life Sciences

- Provides Regulatory Review solutions for Clinical and Nonclinical Data
- Provides Clinical and Nonclinical Research with search, analytics and insights from nonclinical, clinical, and genomic biomarker data
- Xbiom Suite – built on a big-data technology stack:
 - ✓ Spark HDFS, Cassandra
 - ✓ R for Analytics
 - ✓ Lucene/Solr distributed search
 - ✓ ExtJS for GUI
 - ✓ D3 for embedded data visualization
 - ✓ OpenData (Apache Olingo) Data Service Layer for data access to external applications, data visualization tools (SpotFire, Tableau, Clikview, Micro Strategy)



Data Visualization needs are different for Regulatory Analysis and Research

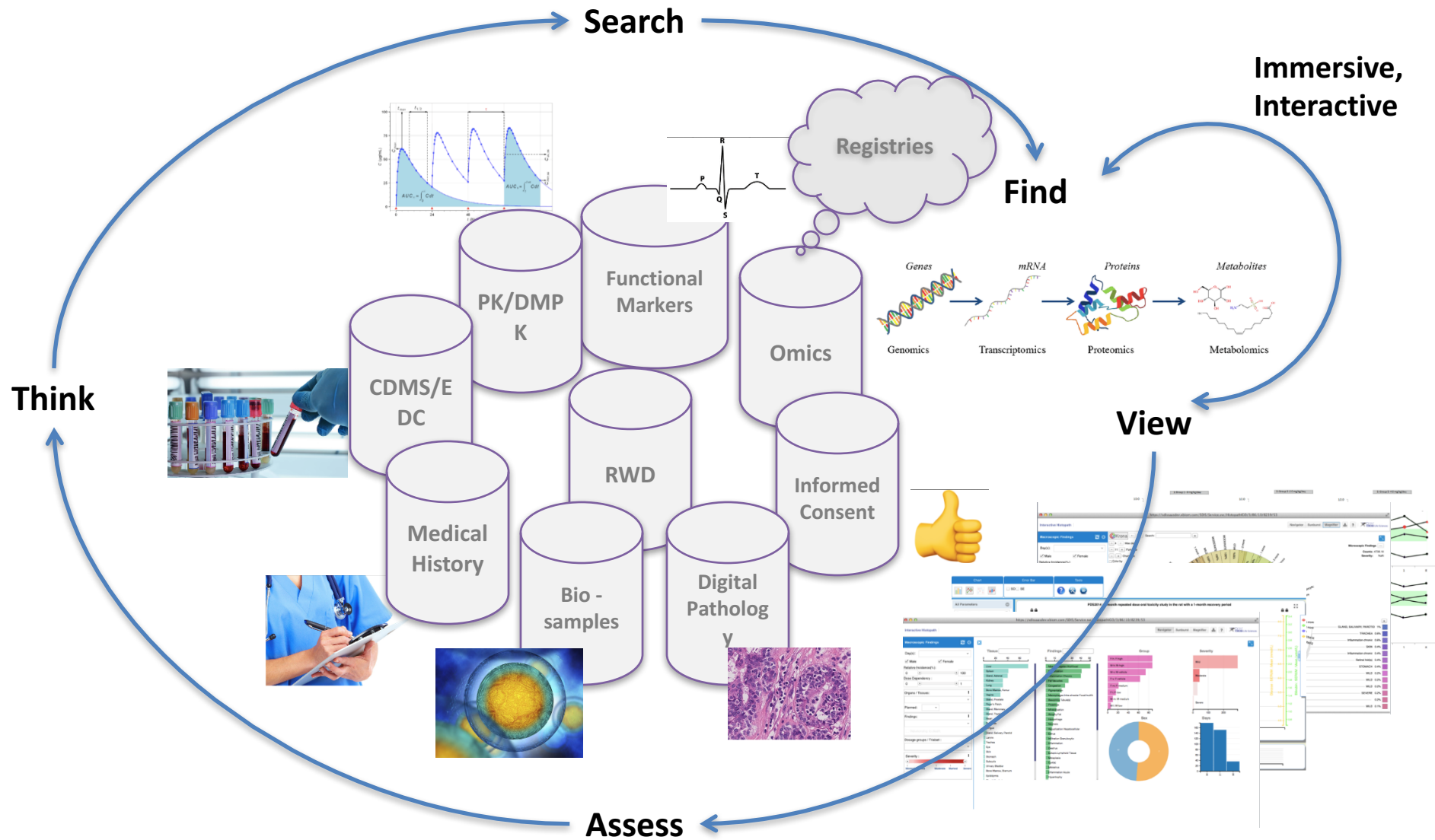
Regulatory

- Statistical Analysis Plans drive the regulatory analysis of a clinical trial study – clinical scientists are supported by programmers, bio-statisticians and data managers
- Data visualization through tabulations and figures is about presenting the results of SAP analysis – not much free form navigation through data
- Visualization is for data within a clinical study, and rarely across multiple studies

Research

- Researchers “consume” data through visualization aids to understand and analyze information – lone users with collaboration
- More likely to view data for meta-analysis across multiple studies or data from patients from multiple studies
- Navigation through the data space is driven by the researcher’s thinking process
- Interactive, immersive experience is key – think-search-find-view-assess-think...

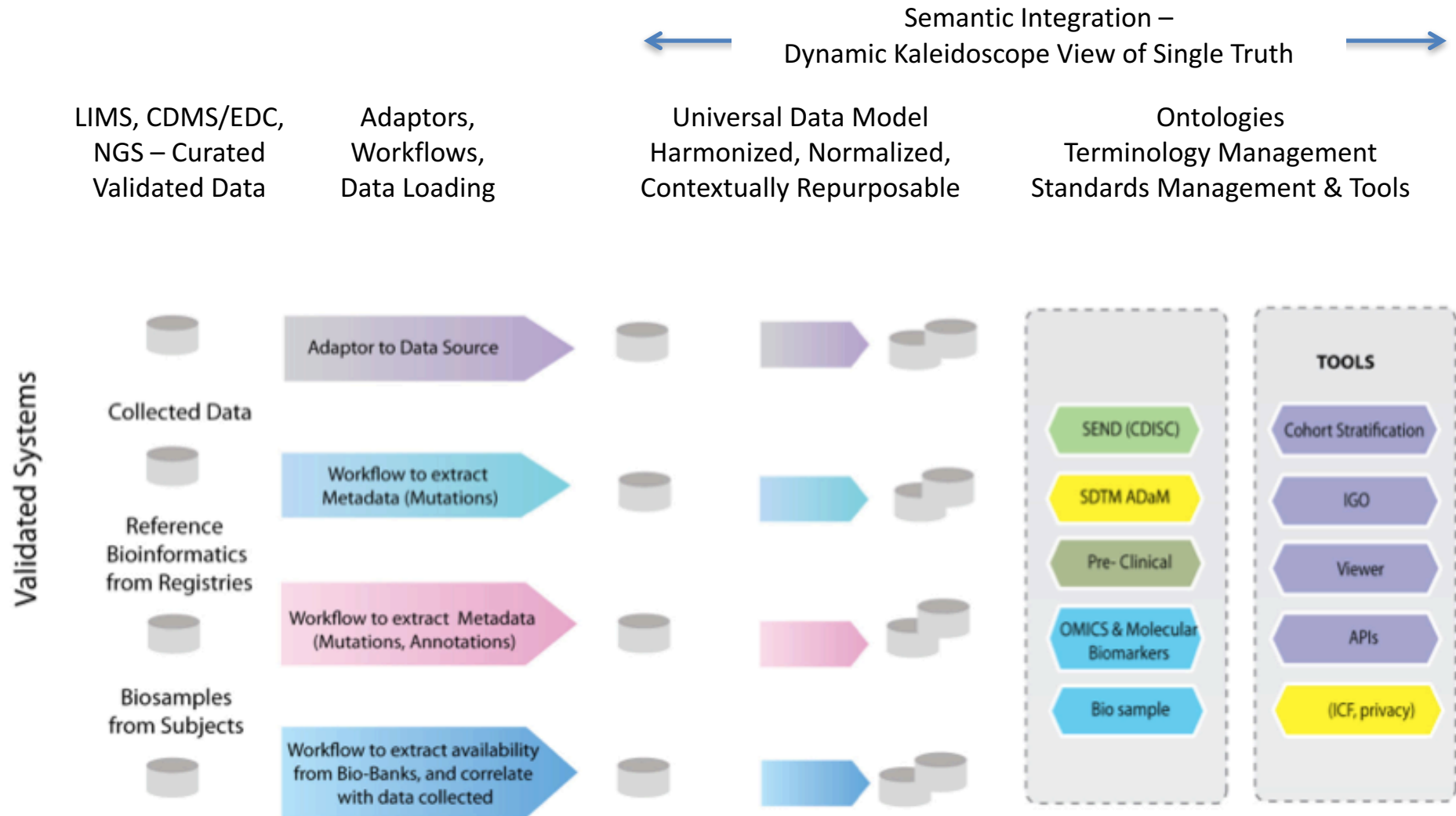
Data Visualization and the Research Cycle



Typical Challenges In Preparing Data for Visualization

- Clinical data is collected and stored in heterogeneous data models within EDC/CDMS
 - ✓ With study specific trial designs
 - ✓ With trial or site specific terminology – even if CDISC CTs are used
 - ✓ With disparate choice of units and technologies
- Updates and changes in Controlled Terminologies (e.g. CDISC) means that different studies can be coded with different standards – without an ontology engine dynamic harmonization is difficult
- Terminology is not standardized in many areas – histopathology, notations for mutations
- Curated meta-data from raw data such as from NGS or digital pathology interpretation may not meet the researcher's needs – data may need to be re-analyzed, bio-samples may need to be re-sequenced

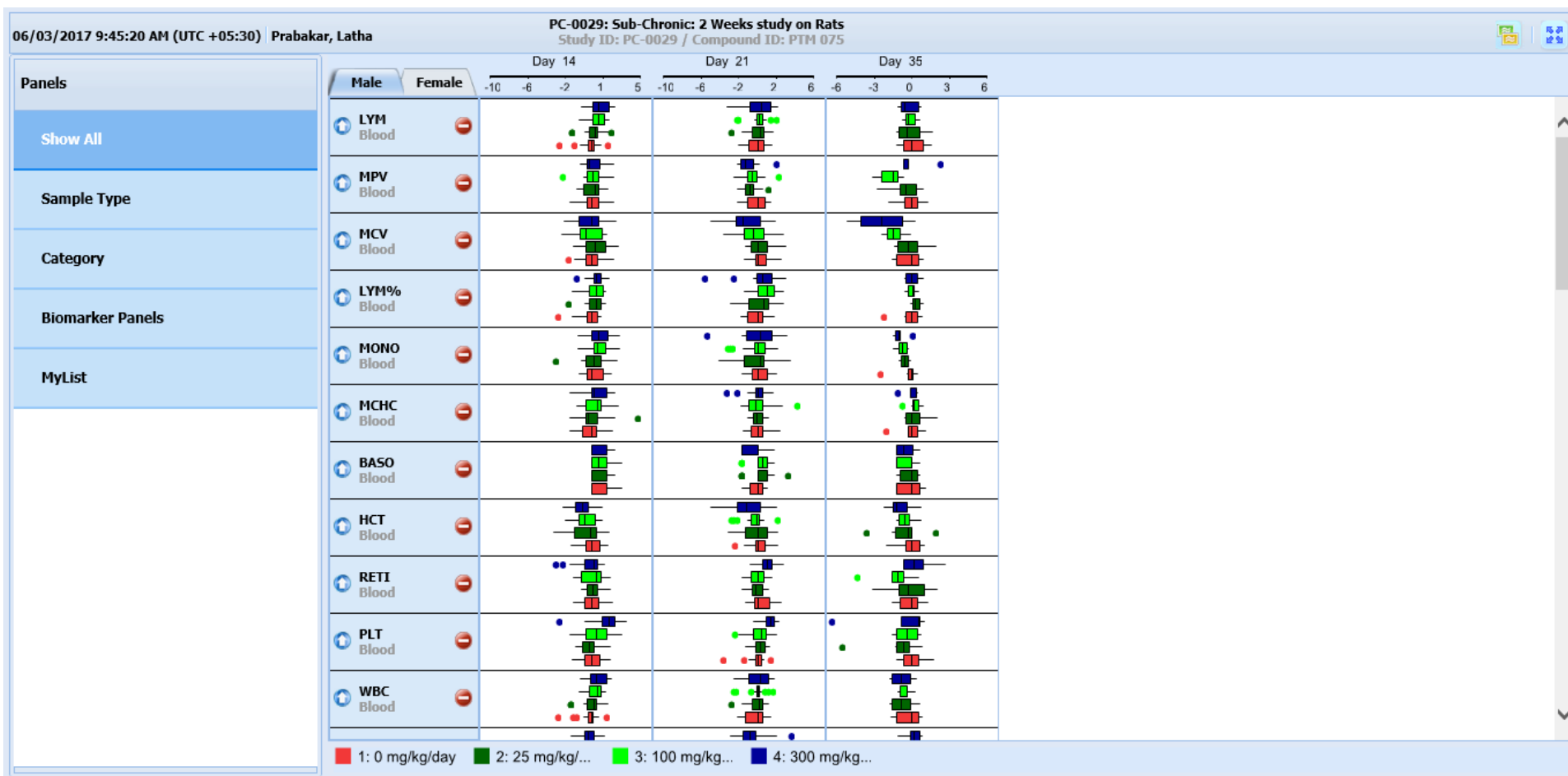
Making Clinical Data Ready for Rich Visualization



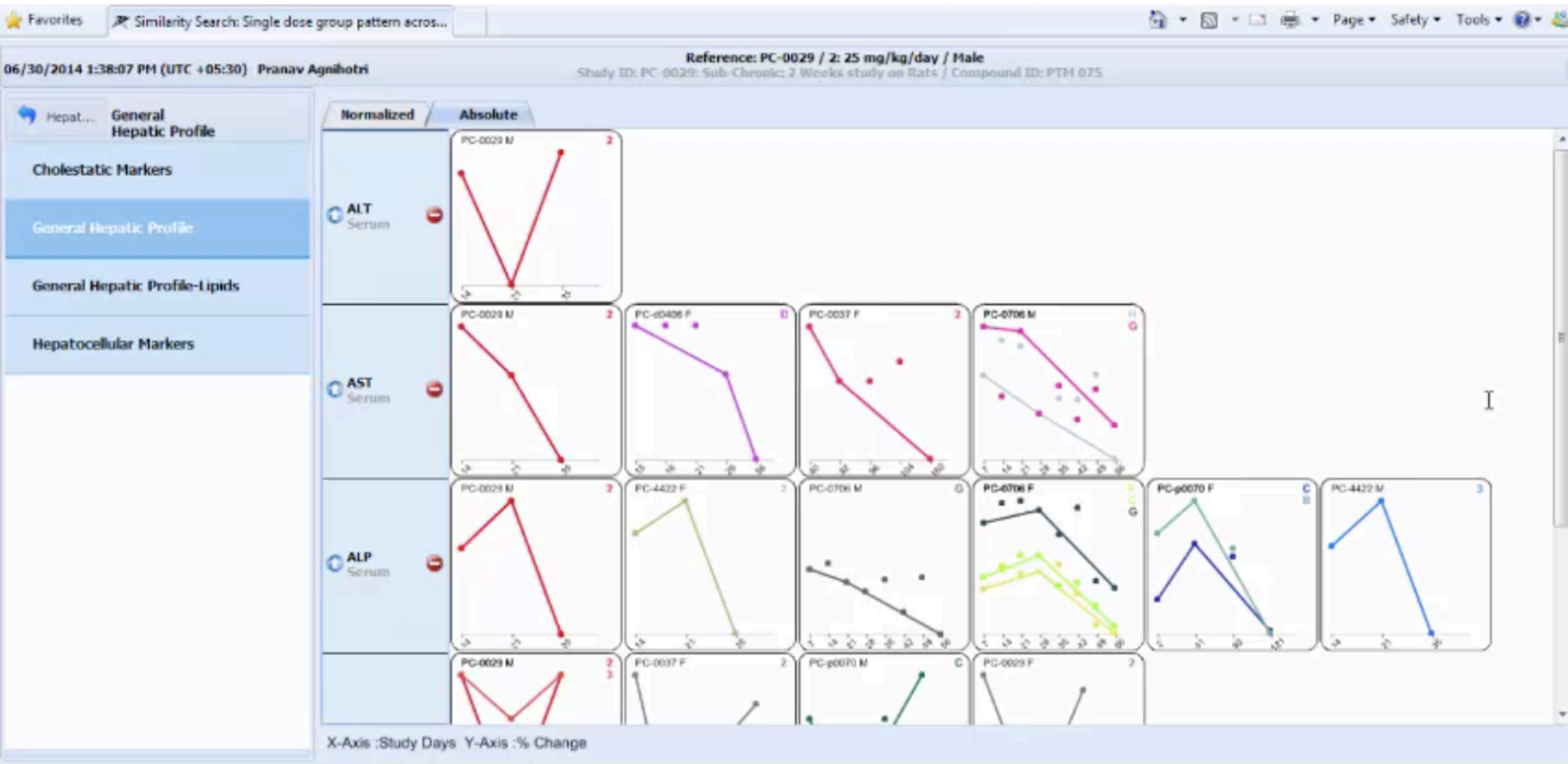
Key Steps in Preparing Data for Immersive Visualization

- Disparate techniques for disparate sources –
 - ✓ Curation – in search for “single truth”
 - ✓ Mapping – from source format to a consumable form
 - ✓ Ontology – to relate information to other related information
 - ✓ Semantic harmonization – synonyms make data searchable
 - ✓ Normalization – of data models and units for searchable data
 - ✓ Repurpose single truth data into multiple forms - kaleidoscope
 - ✓ Provenance – source or origin of data and time stamps
- Metadata maturation - tools, techniques and processes
- Integration – “big data” indexing, optimization
- Delivering data service layers to internal and external data visualization

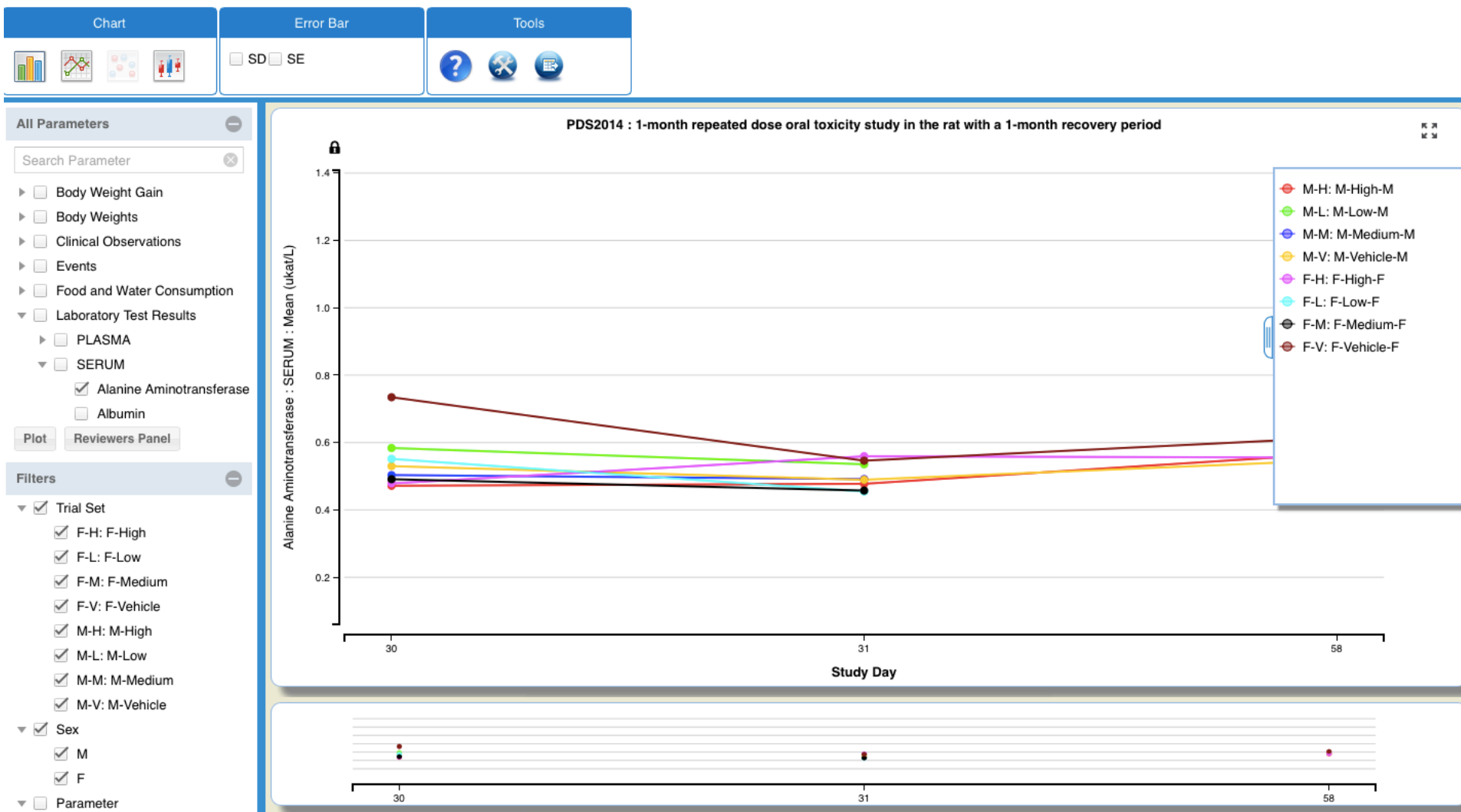
Z- Transforms for dose dependent response across multiple parameters and visit-days



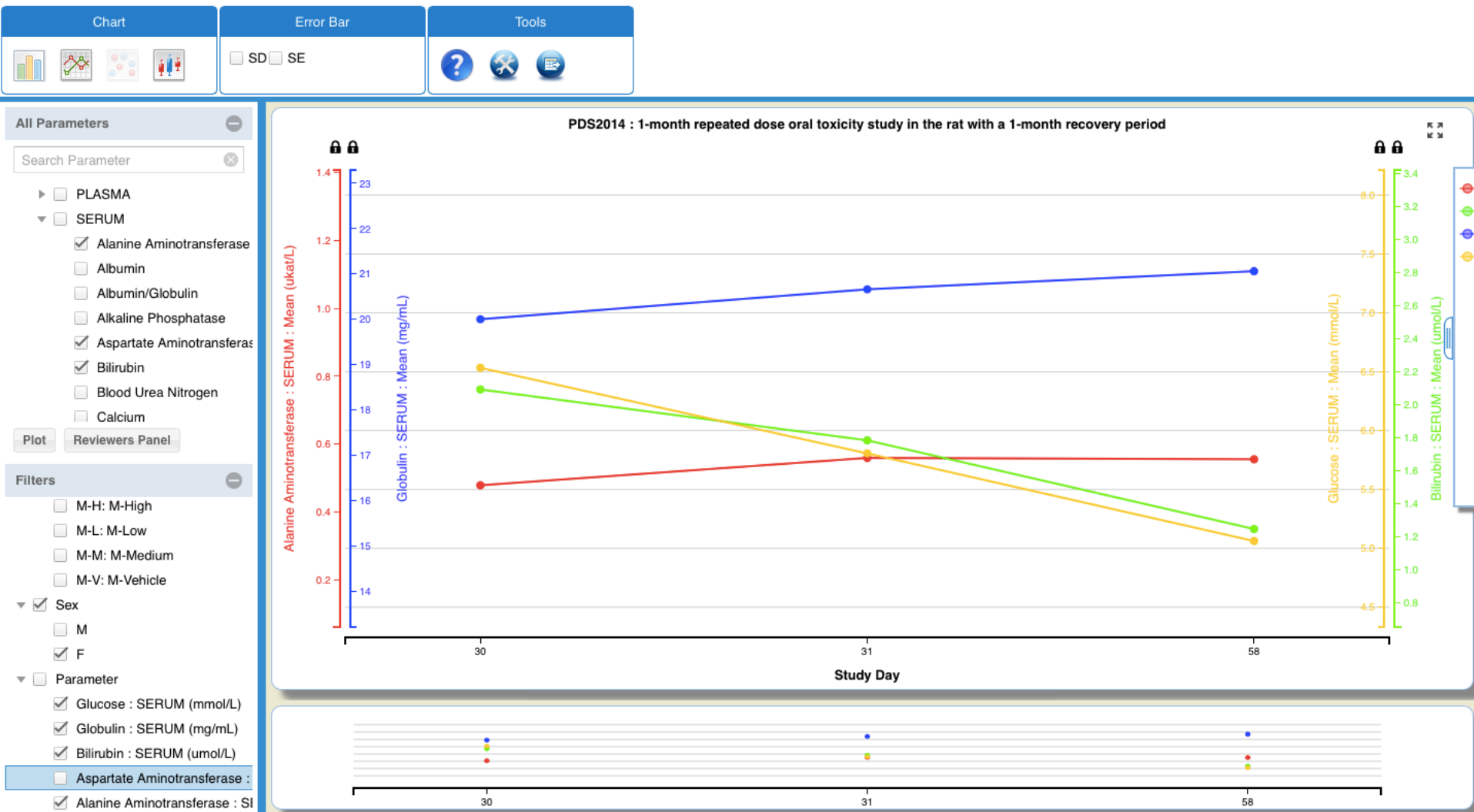
Find studies with similar patterns in a biomarker signal



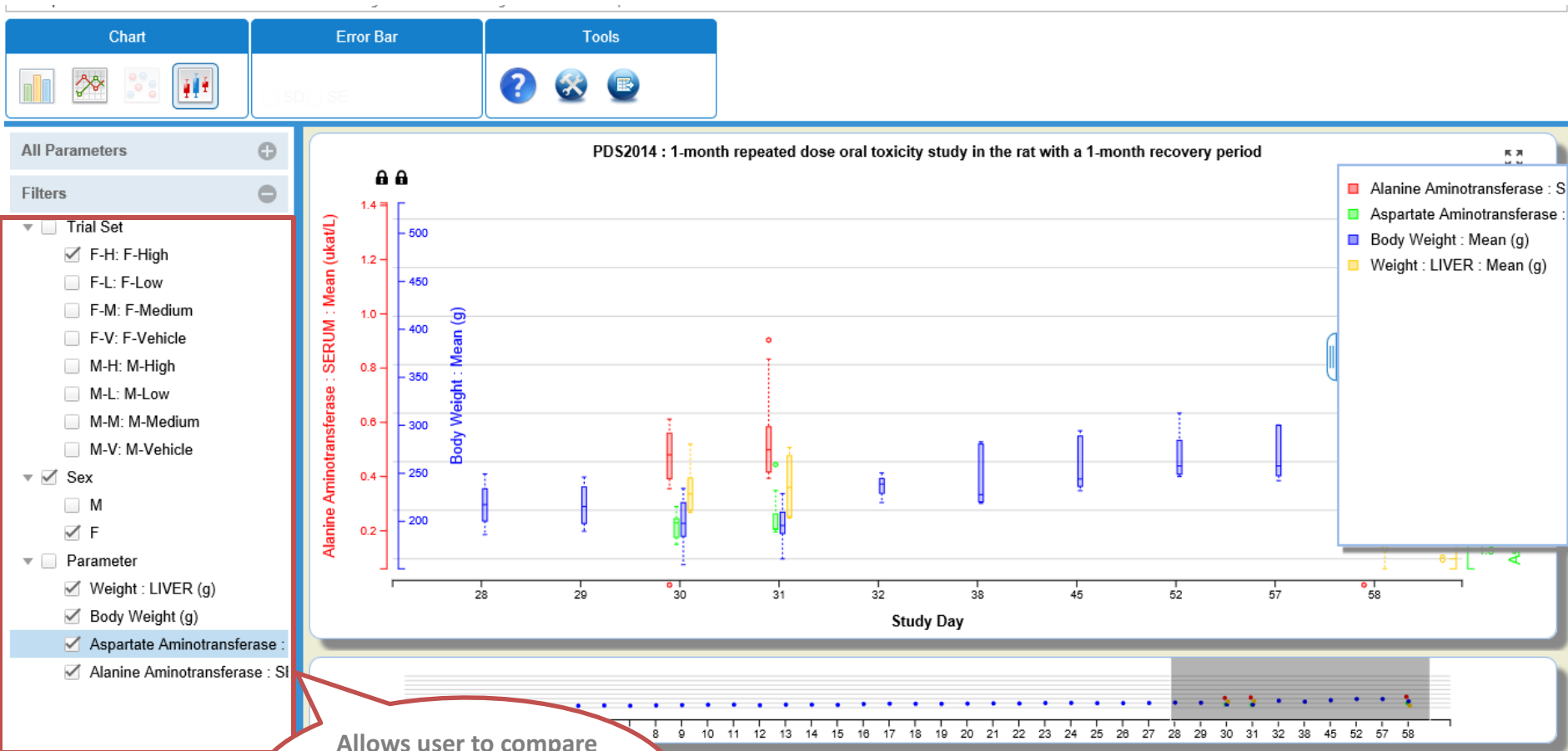
Immersive Interactive Graphics Objects



Dynamic selection and Multi-Parameter Cross Plots



IGO- Cross Domain Visualization



Stochastic Simulation Model Input

Smoker Profile: Michael Distributed_copy [Private]

Correlation matrix is a valid positive definite

Smoker Details Preset Limits Distribution Correlation

Preview Scatter Plot

BW * Body Weight (BW) in kg.

Distribution Type: LOGNORMAL

☒ Statistical Parameter gMean: 4.47

☐ Range (30 - 200) Estimated 5th Percentile: 50.91

Estimated 95th Percentile: 151.5

CpD * Consumption Per Day (CpD) in cig/day.

Distribution Type: CHISQUARED

☒ Statistical Parameter Degrees of Freedom: 17

☐ Range (0 - 100) Estimated 5th Percentile: 8.57

Estimated 95th Percentile: 27.73

EF * Exposure Frequency (EF) in day/yr.

Distribution Type: EXPONENTIAL

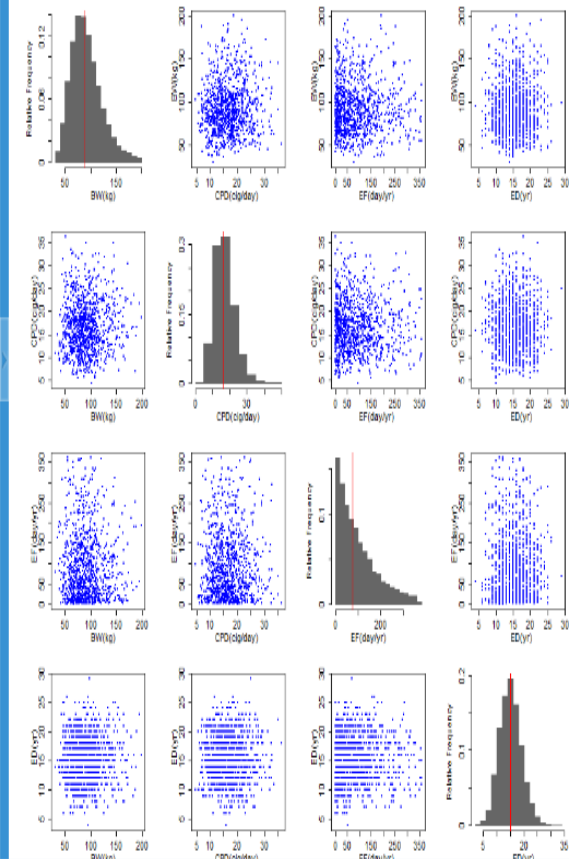
☒ Statistical Parameter Rate: 0.009

☐ Range (0 - 366) Estimated 5th Percentile: 5.62

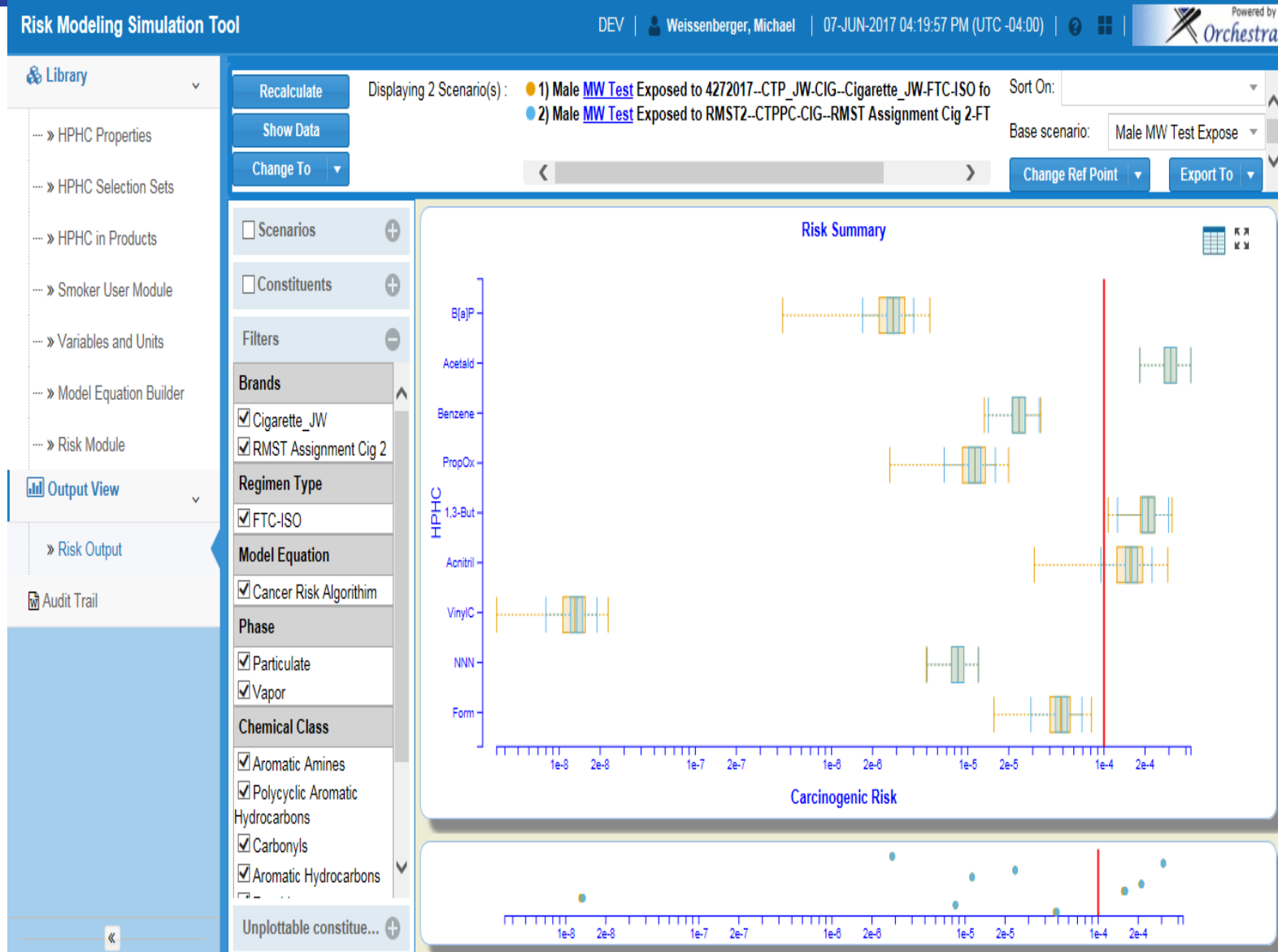
Estimated 95th Percentile: 332.56

ED * Exposure Duration (ED) in yr.

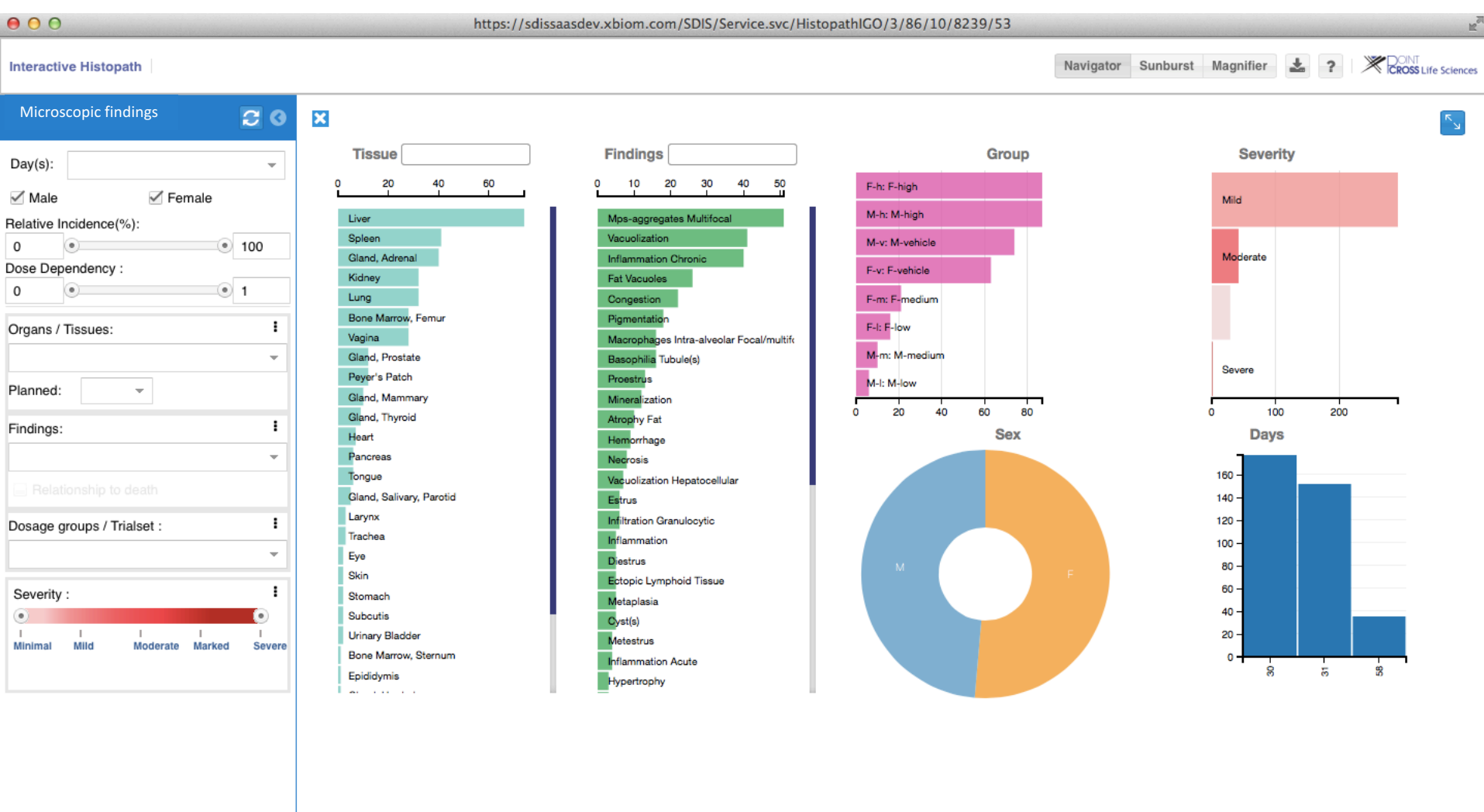
The vertical line in red color indicates median of data



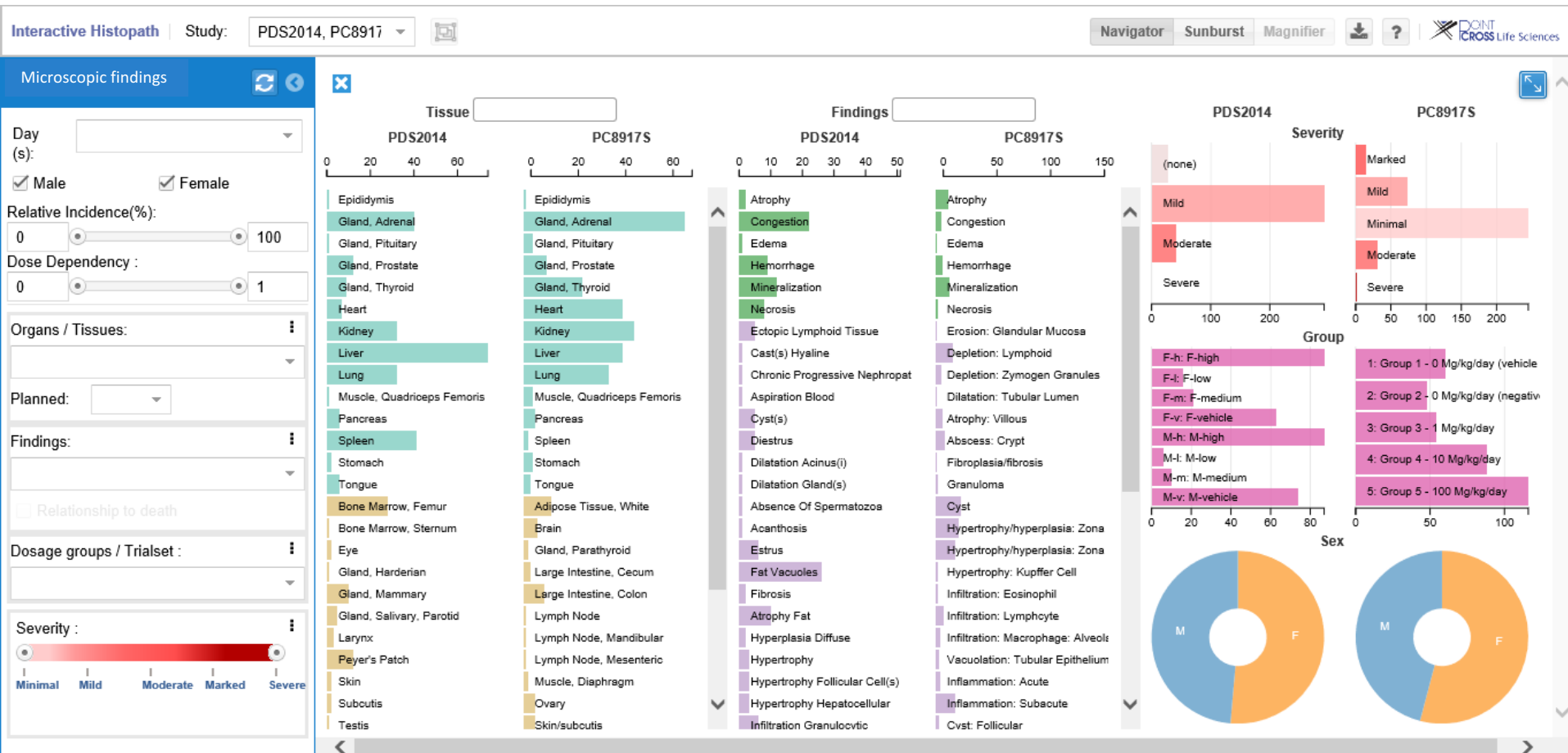
Accumulated Cancer Risk Output of Stochastic Simulation



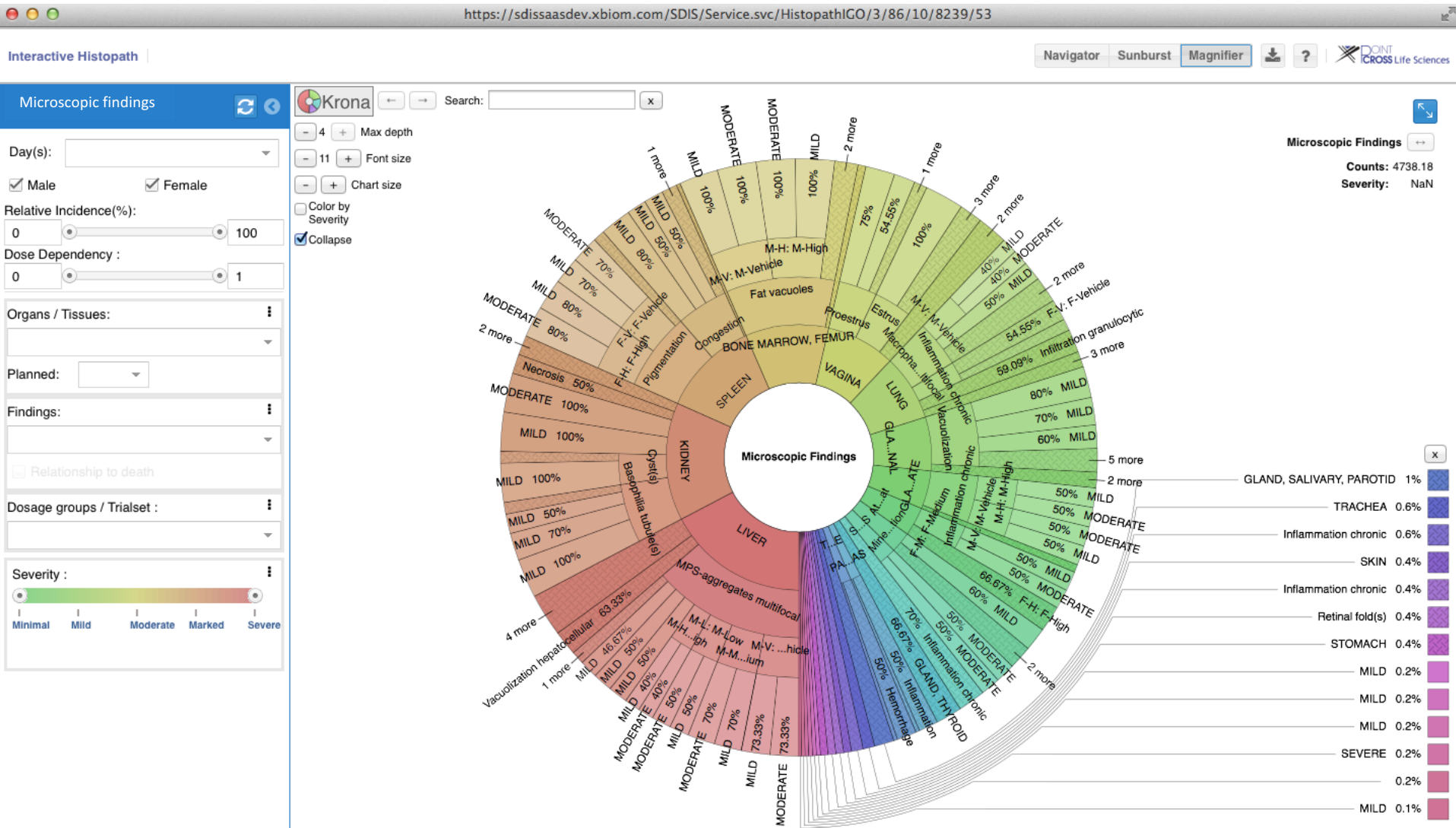
Interactive Histopathology Data (PhUSE CSS WG)



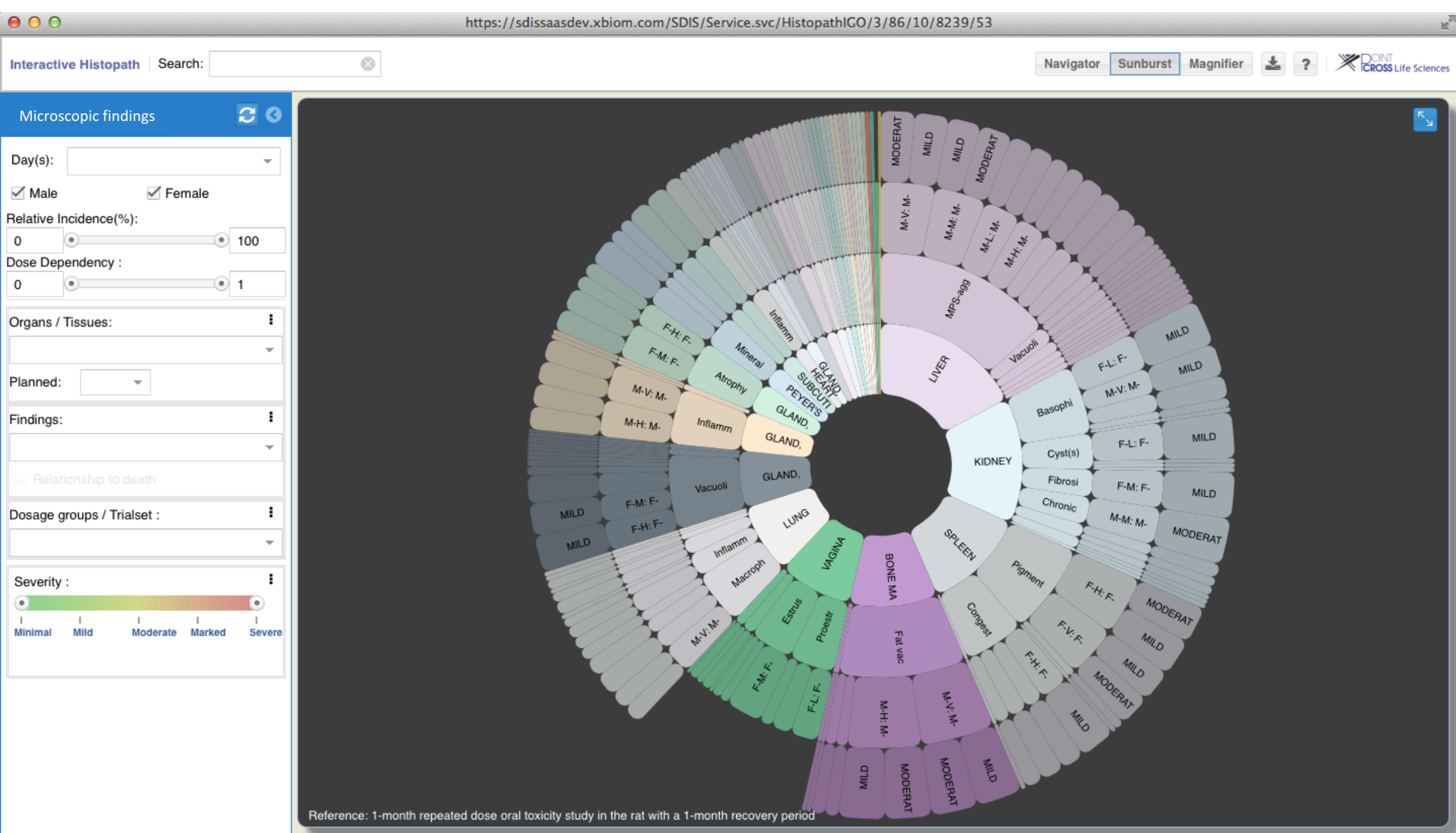
Histopathology Data-Cross study Visualization



Magnifier View for Histopathology data (PhUSE CSS WG)



Sunburst view for Histopathology (PhUSE CSS WG)



A Short Video

Study Data Integration and Search System

Development | Administrator, Orchestra | 06-JUN-2017 02:44:30 PM (UTC +05:30) |  Powered by Orchestra

Saved Search Criteria | Setup Search Criteria | Show Search Results | **ToxVision**

ToxVision

Study Definition/Findings

Enter Search Criteria

Group By: Parameter

Study Data Definitions

Study Activities View

Study/Trial Definition

- Trial Design Domains
- Special-Purpose Domains

Test Findings

- Body Weight Gain - SEND
- Body Weights - SEND
- Laboratory Test Results - SEND
- Macroscopic Findings - SEND
- Microscopic Findings - SEND
- Organ Measurements - SEND

Search Criteria: A

Study ID: PC-26SD-(Reviewer Copy)

Trial Summary

Study Identifier	PC-26SD
Study Title	PC26SD2: One-Month Oral Toxicity Study In Monkeys
Study Type	REPEAT DOSE TOXICITY
Investigational Therapy or Treatment	PC26SD2
Treatment Vehicle	Water
Route of Administration	ORAL GAVAGE
Doses	0 mg/kg/day 1 mg/kg/day 10 mg/kg/day 100 mg/kg/day
Dosing Duration	P28D
Time to Terminal Sacrifice	P29D
Study Start Date	2008-05-23
Experimental Start Date	2008-05-23
Experimental End Date	2008-07-06
Sponsor's Group Label	Group 1 - 0 mg/kg/day (Vehicle Control) Group 2 - 0 mg/kg/day (Saline control) Group 3 - 1 mg/kg/day Group 4 - 10 mg/kg/day Group 5 - 100 mg/kg/day
Species	MONKEY
Strain/Substrain	CYNOMOLGUS
Age Unit	YEARS
Age Text	2-3
Test Subject Supplier	CRB
Primary Treatment CAS Registry Number	NOT AVAILABLE
Test Facility Name	PC, MD
Test Facility Location	PointCross, Department of Research and Development
Test Facility Country	USA

PointCross

19

 POINT CROSS

Putting Knowledge to Work

© 2017 PointCross. All Rights Reserved

Tabulations of selected signals of interest – for export or charting

Study Data Integration and Search System

SDIS SaaS Dev | Madhavan, Suresh | 30-MAY-2017 02:16:44 PM (UTC -07:00) |

Powered by
Orchestra

Saved Search Criteria

Setup Search Criteria

Show Search Results

ToxVision

Business Base: PCBLRMD

ToxVision

Study Definition/Findings

Enter Search Criteria

Group By: Parameter

Special-Purpose Domains

Interventions

Relationship Datasets

Test Findings

Body Weight Gain - SEND

Body Weights - SEND

Clinical Observations - SEND

Events

Food and Water Consumption - SEND

Laboratory Test Results - SEND

PLASMA - SEND

Activated Partial Thromboplastin Time

Albumin/Globulin

Fibrinogen

Globulin

Prothrombin Time

SERUM - SEND

Alanine Aminotransferase

Albumin

Albumin/Globulin

Alkaline Phosphatase

Aspartate Aminotransferase

Bilirubin

Blood Urea Nitrogen

Calcium

Chloride

Cholesterol

Creatinine

Free Fatty Acid

Gamma Glutamyl Transferase

Globulin

Glucose

Phosphate

Potassium

Search Criteria: A

SEND Implementation Guide Version

SEND Implementation Guide Version 3.0

SEND Controlled Terminology Version

SEND Terminology 2016-09-30

Study Category

TOX

Study Design

PARALLEL

Weight

F: 152.5 - 202.7 g

M: 272.2 - 305.2 g

Sponsor's Reference ID

PDS2014

Group Summary

Individual Results

Individual Details

Page 1 of 2

View Days as Columns

Menu

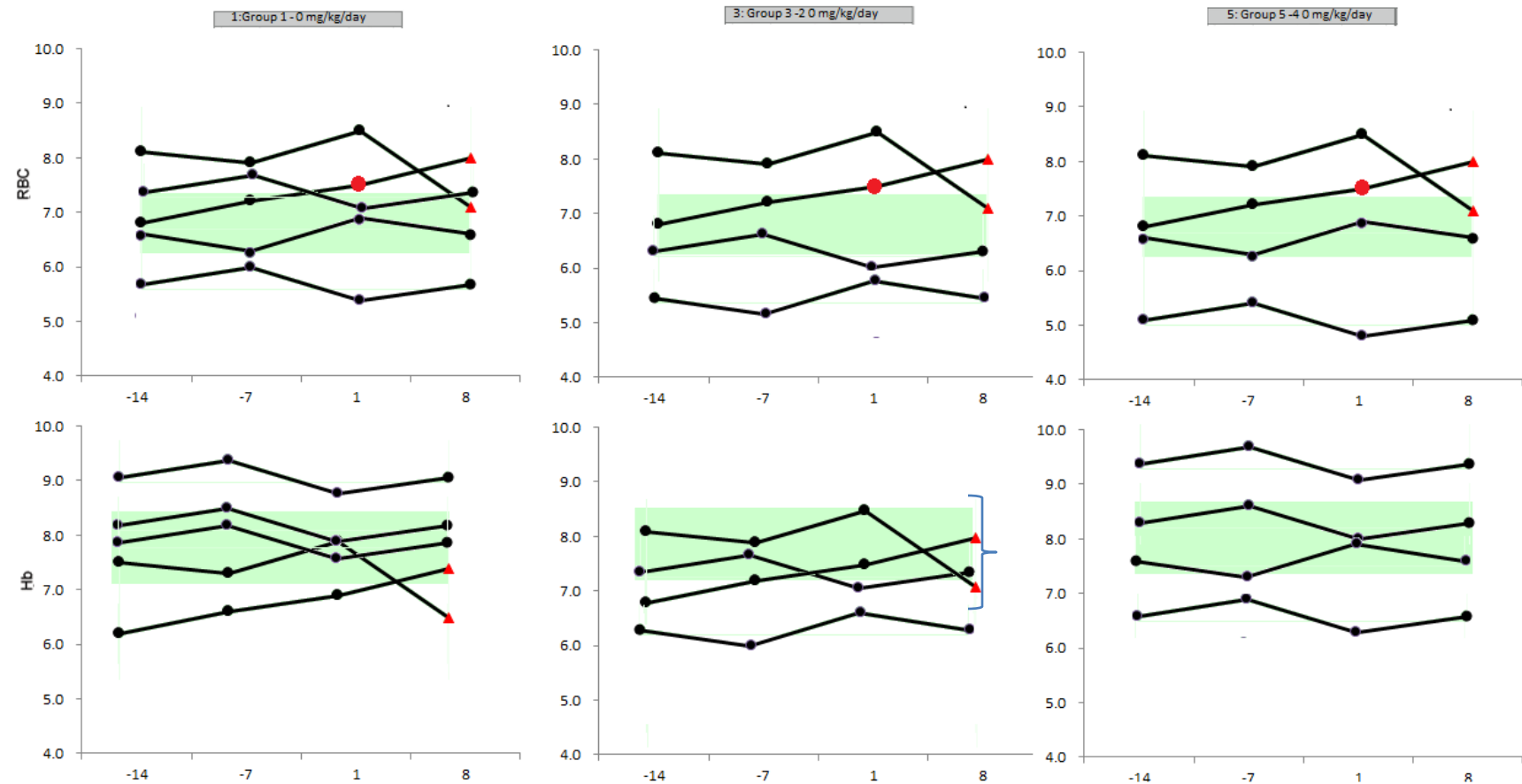
Day 30

Day 31

Day 58

	Group	Sex	Sample Type	Parameter	Unit	Mean	SD	N	% Change	Mean	SD	N	% Change	Mean	SD	N	% Change
☐	M-H: M-High	M	SERUM	Alanine Aminotransferase	ukat/L	0.4717	0.0556	10	-10.97	0.4774	0.08593	5	-2.452	0.5671	0.06615	5	3.593
☐	M-L: M-Low	M	SERUM	Alanine Aminotransferase	ukat/L	0.5837	0.09429	5	10.16	0.5354	0.05242	5	9.4				
☐	M-M: M-Medium	M	SERUM	Alanine Aminotransferase	ukat/L	0.5034	0.08354	5	-5.001	0.4914	0.08399	5	0.4087				
☐	M-V: M-Vehicle	M	SERUM	Alanine Aminotransferase	ukat/L	0.5299	0.06922	10	0.0	0.4894	0.02376	5	0.0	0.5474	0.08134	5	0.0
☐	F-H: F-High	F	SERUM	Alanine Aminotransferase	ukat/L	0.4784	0.09642	10		0.5591	0.2065	5		0.5551	0.1012	5	
☐	F-L: F-Low	F	SERUM	Alanine Aminotransferase	ukat/L	0.5514	0.1177	5		0.4544	0.03038	5					
☐	F-M: F-Medium	F	SERUM	Alanine Aminotransferase	ukat/L	0.491	0.1075	5		0.4574	0.08547	5					
☐	F-V: F-Vehicle	F	SERUM	Alanine Aminotransferase	ukat/L	0.7344	0.6748	10		0.5461	0.08612	5		0.6147	0.139	5	
☐	M-H: M-High	M	SERUM	Aspartate Aminotransferase	ukat/L	1.707	0.1662	10	-20.08	1.707	0.09717	5	-18.23	1.906	0.3541	5	-5.143
☐	M-L: M-Low	M	SERUM	Aspartate Aminotransferase	ukat/L	2.122	0.425	5	-0.6709	2.078	0.2718	5	-0.4789				
☐	M-M: M-Medium	M	SERUM	Aspartate Aminotransferase	ukat/L	1.975	0.1359	5	-7.551	1.711	0.1963	5	-18.05				
☐	M-V: M-Vehicle	M	SERUM	Aspartate Aminotransferase	ukat/L	2.136	0.2215	10	0.0	2.088	0.3523	5	0.0	2.009	0.1598	5	0.0
☐	F-H: F-High	F	SERUM	Aspartate Aminotransferase	ukat/L	1.732	0.1209	10		1.84	0.262	5		1.896	0.1615	5	
☐	F-L: F-Low	F	SERUM	Aspartate Aminotransferase	ukat/L	1.989	0.1252	5		1.966	0.06262	5					
☐	F-M: F-Medium	F	SERUM	Aspartate Aminotransferase	ukat/L	1.712	0.3789	5		1.847	0.3428	5					
☐	F-V: F-Vehicle	F	SERUM	Aspartate Aminotransferase	ukat/L	2.375	0.8362	10		2.111	0.2259	5		2.091	0.3652	5	
☐	M-H: M-High	M	SERUM	Bilirubin	umol/L	1.385	0.446	10	-13.18	0.9612	0.1913	5	-21.72	1.217	0.1823	5	-8.717
☐	M-L: M-Low	M	SERUM	Bilirubin	umol/L	1.388	0.228	5	-12.96	1.275	0.2684	5	3.899				
☐	M-M: M-Medium	M	SERUM	Bilirubin	umol/L	1.358	0.3637	5	-14.89	1.084	0.4143	5	-11.69				
☐	M-V: M-Vehicle	M	SERUM	Bilirubin	umol/L	1.595	0.3396	10	0.0	1.228	0.514	5	0.0	1.334	0.376	5	0.0

Plotting of Reference Ranges



Xbiom – Built on a Robust Technology Stack

Web Server



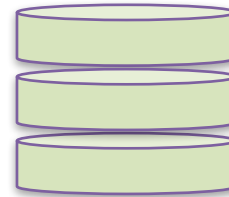
RDBMS



Big Data Processing

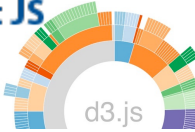
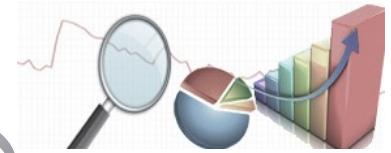


Normalized Data
Big Table



Cassandra

Data Analysis,
Interaction and Search



Orchestra™

Single Sign-on
Web Services
HTTPS

User Accounts & Roles
Object meta-data
Crawler configurations
And Mappers

Data Processing
and normalization,
Document Storage

Normalized data with
Precompiled Summaries
and statistics

Data visualization,
Search and interactivity