Molding and Maximizing the LabKey Platform for Clinical Translational Research

Anthony Corbett
University of Rochester Medical Center

2019 LabKey User Conference

Outline

- LabKey at URMC

- Philosophy

- Strategy
  - Clinical Data Capture
  - Specimen Management
  - Experimental Data
  - Reporting and Visualization

- Future Work

LabKey at URMC

Supported Projects

Multi-Center grants

Industry funded research projects

Biobanking

Departmental/Division wide

Reagent databases

LabKey at URMC

Research Areas

Infectious diseases

Autoimmune diseases

Lung development

Cancer

Clinical Interventions in Autism

Public Health (smoking/vaping)

Maternal and Children’s Health

Overview

Bio-Lab-Informatics-System (BLIS)

- Over 90 observational/registry studies, since 2010
- ~16,000 participants
- ~450,000 specimens (40 different primary specimen types)
- ‘omics and assay data integration

Biorepository for Investigation of Neonatal Diseases of the Lung (BRINDL)

- ~72,000 samples
- ~200 donors
- Imaging and ‘omics integration

Philosophy

- Improve Research Team Collaboration & Productivity
  - Participate as member of research teams
  - User, processes and workflow focused
- Enable “Real-time” Data
  - Visualizations, Analytics, Exploration, Sharing
- Enhanced Data Stewardship & Security (FISMA)
- Innovate Incorporating Best Tools and Practices

Philosophy

> “Organizations which design systems ... are constrained to produce designs which are copies of the communication structures of these organizations.”
>
> \- **Conway’s Law**

Thus, the structure of a system will reflect the social boundaries of the organization(s) that produced it, across which communication is more difficult.

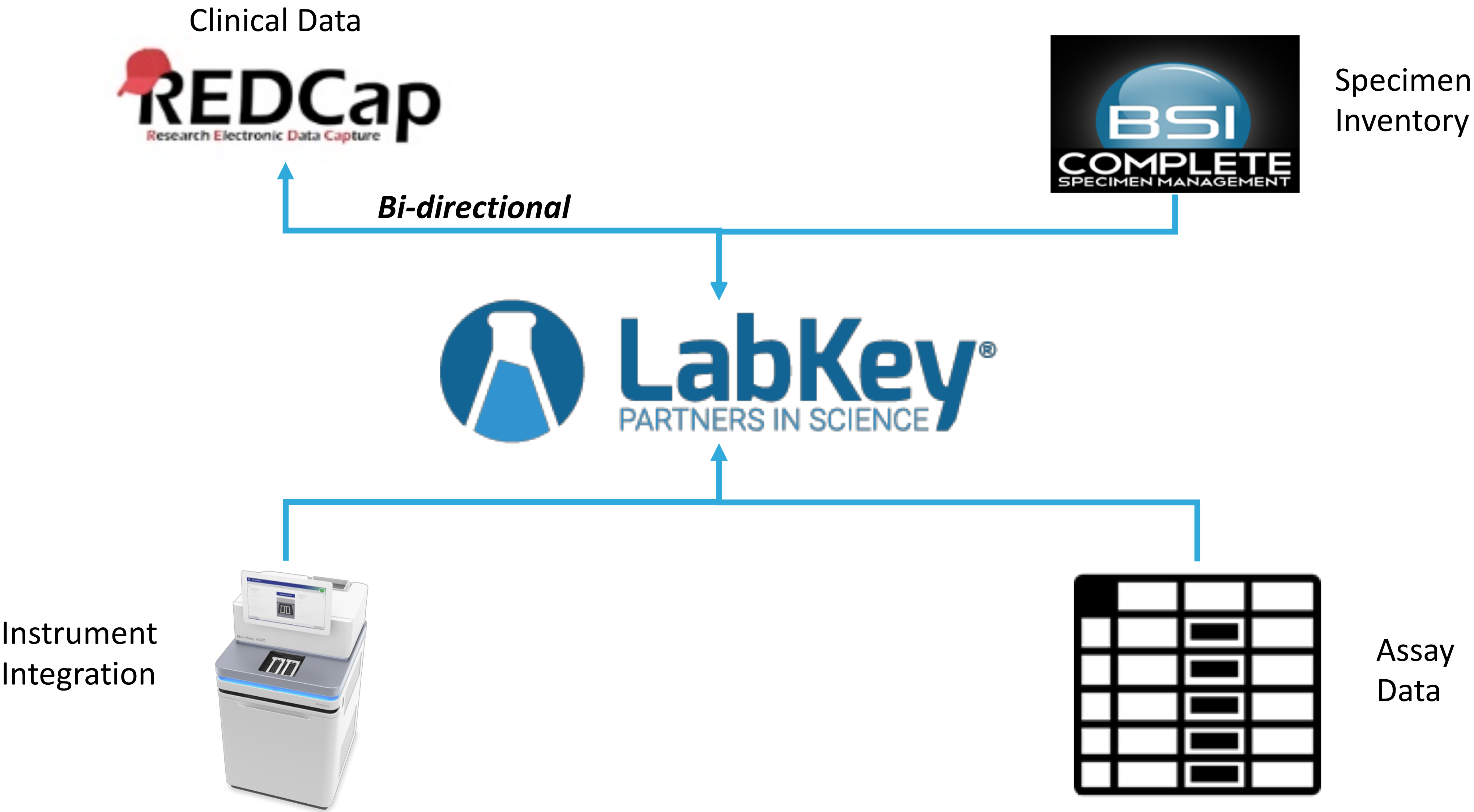
Philosophy

Factors driving LabKey adoption at URMC

Single Lab -> Collaborative Science

Single Purpose -> Discovery and Reuse

Strategy



Clinical Data

Bi-directional

Specimen Inventory

Instrument Integration

Assay Data

Clinical Data

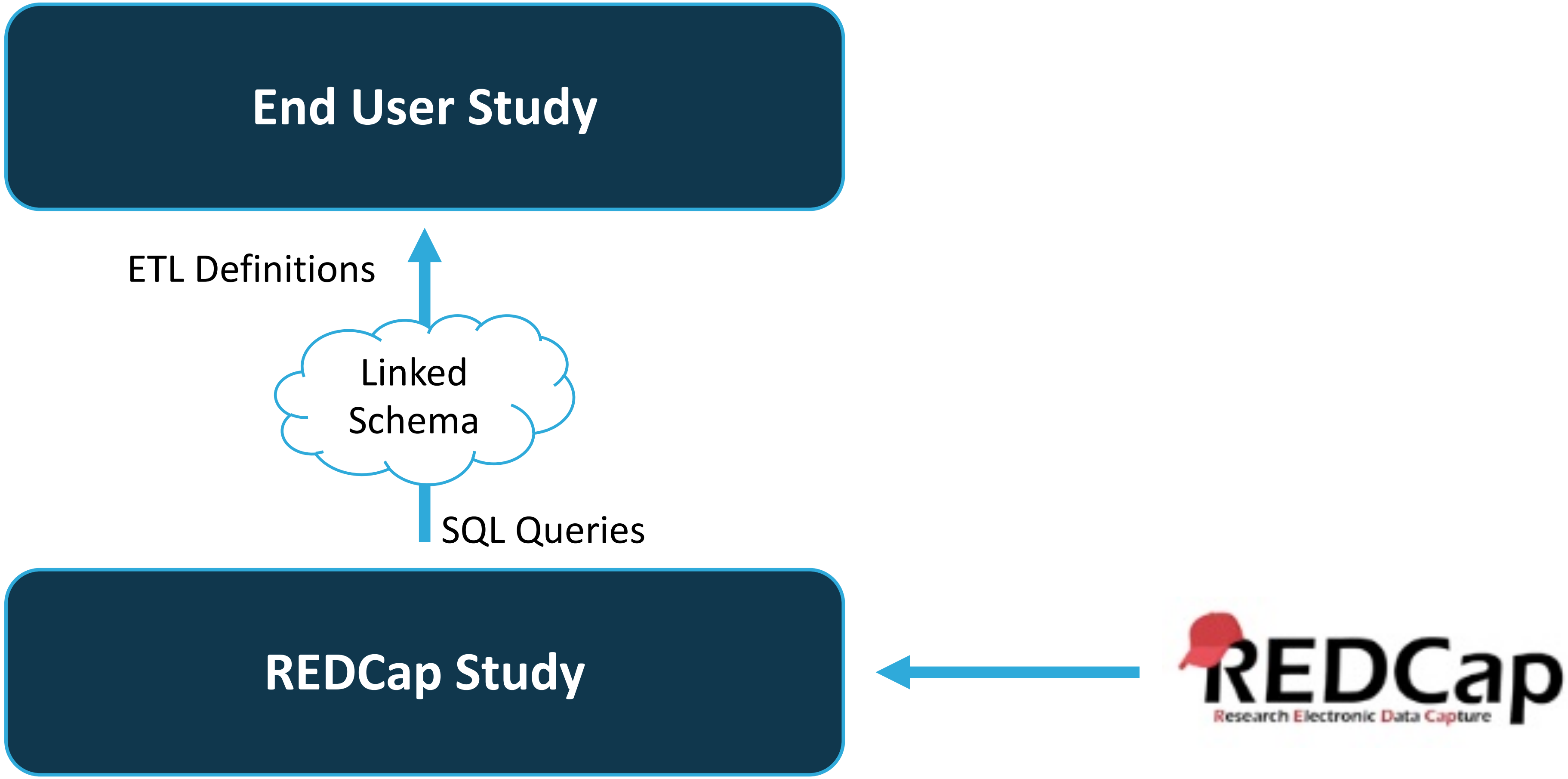
Clinical Data Capture

1. Using the redcap module to import clinical data in CDISC ODM format

2. Use the data integration module to write SQL and ETL processes
   - ✓ Restructure and subset the data
   - ✓ Calculated fields
   - ✓ QC metrics

Clinical Data Capture

REDCap



End User Study

ETL Definitions

Linked Schema

SQL Queries

REDCap Study

Custom REDCap Interactions

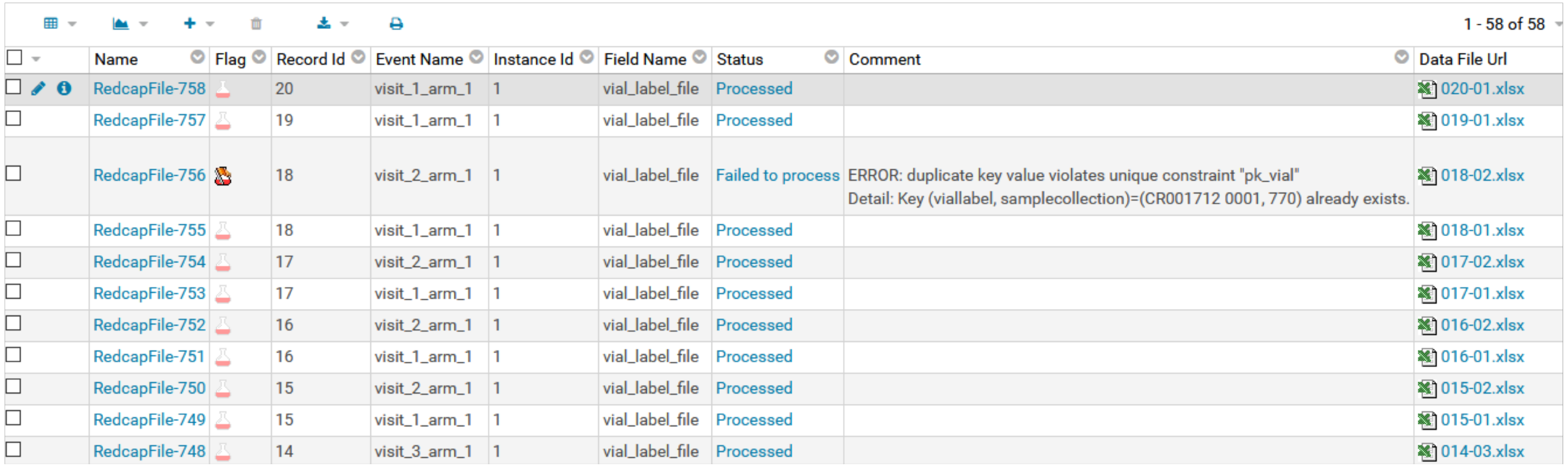
URMC developed module

Redcap API Service

- ✓ Import and Export Records
- ✓ Download files
- ✓ Run RedCAP Reports
- ✓ Download Events, Arms, Event-Form Mappings.

Custom ETL task definitions

- ✓ Export Records Task
- ✓ Download Files into DataClass and process in LabKey



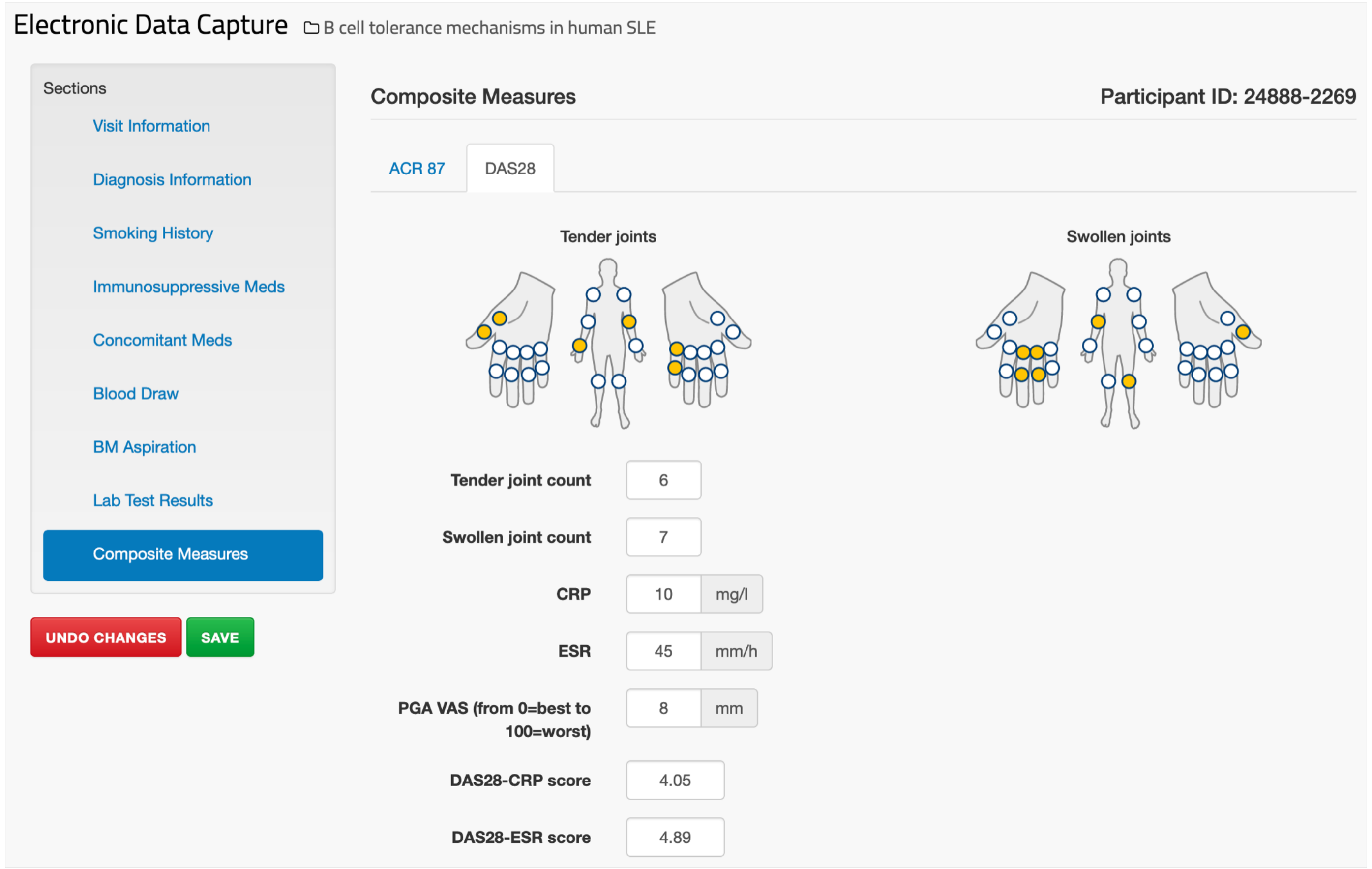
1 - 58 of 58

☐	Name	Flag	Record Id	Event Name	Instance Id	Field Name	Status	Comment	Data File Url
☐	RedcapFile-758		20	visit_1_arm_1	1	vial_label_file	Processed		020-01.xlsx
☐	RedcapFile-757		19	visit_1_arm_1	1	vial_label_file	Processed		019-01.xlsx
☐	RedcapFile-756		18	visit_2_arm_1	1	vial_label_file	Failed to process	ERROR: duplicate key value violates unique constraint "pk_vial" Detail: Key (viallabel, samplecollection)=(CR001712 0001, 770) already exists.	018-02.xlsx
☐	RedcapFile-755		18	visit_1_arm_1	1	vial_label_file	Processed		018-01.xlsx
☐	RedcapFile-754		17	visit_2_arm_1	1	vial_label_file	Processed		017-02.xlsx
☐	RedcapFile-753		17	visit_1_arm_1	1	vial_label_file	Processed		017-01.xlsx
☐	RedcapFile-752		16	visit_2_arm_1	1	vial_label_file	Processed		016-02.xlsx
☐	RedcapFile-751		16	visit_1_arm_1	1	vial_label_file	Processed		016-01.xlsx
☐	RedcapFile-750		15	visit_2_arm_1	1	vial_label_file	Processed		015-02.xlsx
☐	RedcapFile-749		15	visit_1_arm_1	1	vial_label_file	Processed		015-01.xlsx
☐	RedcapFile-748		14	visit_3_arm_1	1	vial_label_file	Processed		014-03.xlsx

Clinical Data Capture

Custom Forms

URMC developed EDC framework built on LabKey JavaScript API



Electronic Data Capture 🗀 B cell tolerance mechanisms in human SLE

Sections
- Visit Information
- Diagnosis Information
- Smoking History
- Immunosuppressive Meds
- Concomitant Meds
- Blood Draw
- BM Aspiration
- Lab Test Results
- Composite Measures

Composite Measures

Participant ID: 24888-2269

Tender joint count	6	
Swollen joint count	7	
CRP	10	mg/l
ESR	45	mm/h
PGA VAS (from 0=best to 100=worst)	8	mm
DAS28-CRP score	4.05	
DAS28-ESR score	4.89	

Specimen Inventory

Specimen Management

Many Sources

Excel spreadsheets

LDMS

BSI

Specimen Management

- Specimen accession and processing
- Freezer and location management
- Integration with BarTender
    - Barcoding
    - label printing

Specimen Management

BSI

One way: BSI -> LabKey

Pulled Nightly with

Custom Availability Rules

Want two way: LabKey <-> BSI

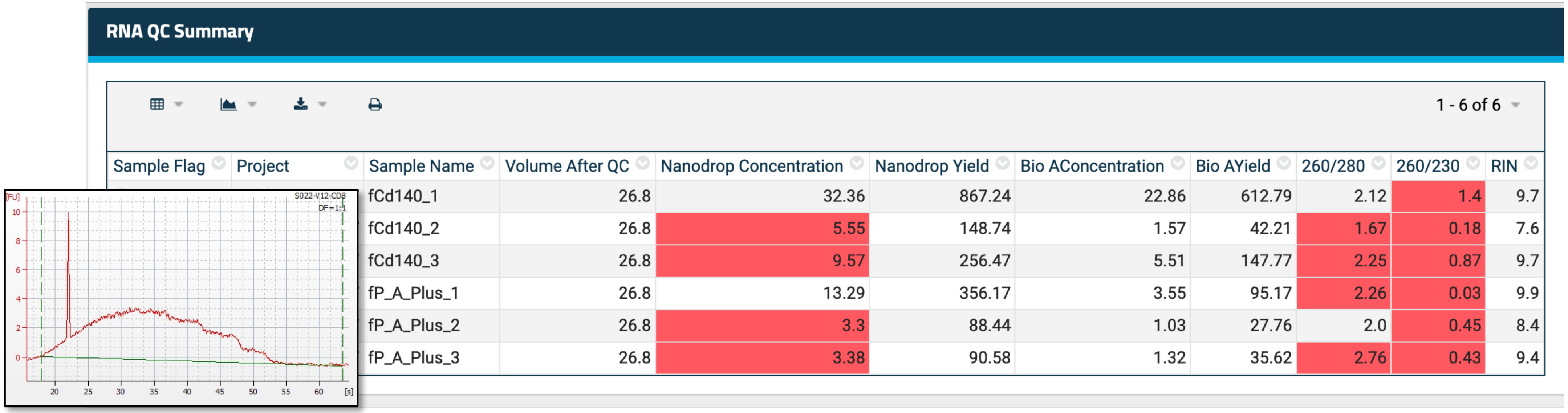
Specimen Requests -> Requisitions

Instrument and Assay Data

Experimental Data

Assay Results

BioAnalyzer and Nanodrop QC



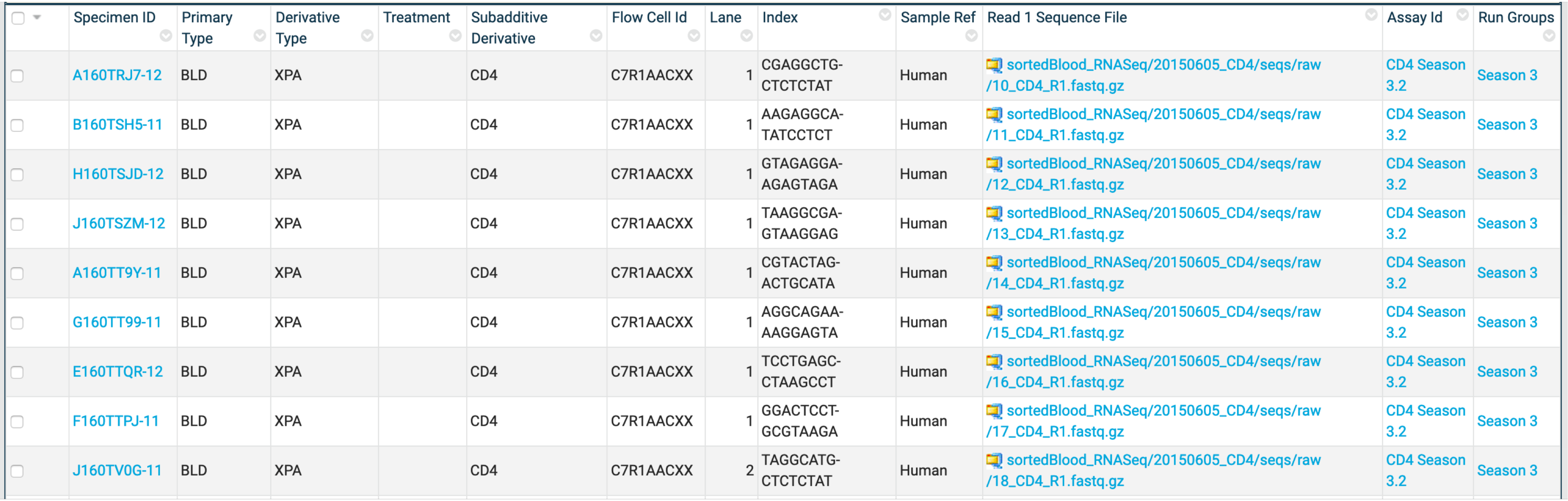
RNA QC Summary

Sample Flag	Project	Sample Name	Volume After QC	Nanodrop Concentration	Nanodrop Yield	Bio AConcentration	Bio AYield	260/280	260/230	RIN
		fCd140_1	26.8	32.36	867.24	22.86	612.79	2.12	1.4	9.7
		fCd140_2	26.8	5.55	148.74	1.57	42.21	1.67	0.18	7.6
		fCd140_3	26.8	9.57	256.47	5.51	147.77	2.25	0.87	9.7
		fP_A_Plus_1	26.8	13.29	356.17	3.55	95.17	2.26	0.03	9.9
		fP_A_Plus_2	26.8	3.3	88.44	1.03	27.76	2.0	0.45	8.4
		fP_A_Plus_3	26.8	3.38	90.58	1.32	35.62	2.76	0.43	9.4

Experimental Data

Metadata Capture

Illumina Sequencing Run and Results

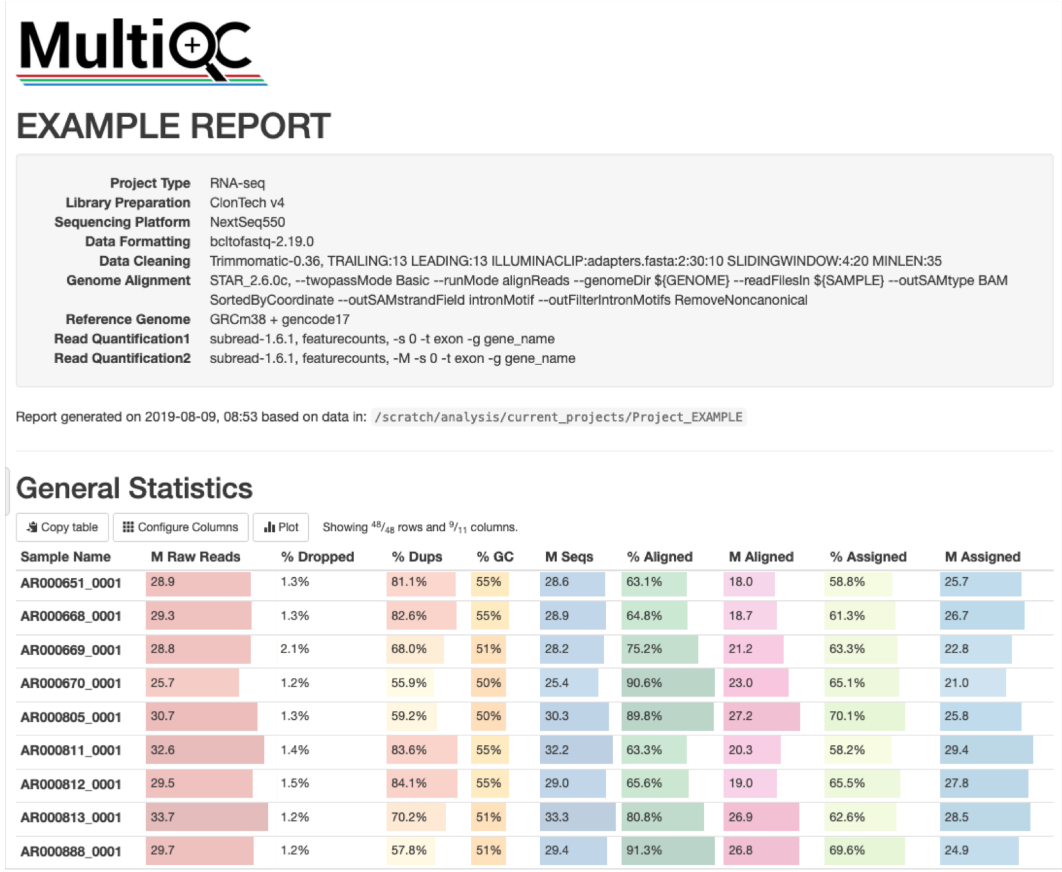


☐	Specimen ID	Primary Type	Derivative Type	Treatment	Subadditive Derivative	Flow Cell Id	Lane	Index	Sample Ref	Read 1 Sequence File	Assay Id	Run Groups
☐	A160TRJ7-12	BLD	XPA		CD4	C7R1AACXX	1	CGAGGCTG-CTCTCTAT	Human	sortedBlood_RNASeq/20150605_CD4/seqs/raw/10_CD4_R1.fastq.gz	CD4 Season 3.2	Season 3
☐	B160TSH5-11	BLD	XPA		CD4	C7R1AACXX	1	AAGAGGCA-TATCCTCT	Human	sortedBlood_RNASeq/20150605_CD4/seqs/raw/11_CD4_R1.fastq.gz	CD4 Season 3.2	Season 3
☐	H160TSJD-12	BLD	XPA		CD4	C7R1AACXX	1	GTAGAGGA-AGAGTAGA	Human	sortedBlood_RNASeq/20150605_CD4/seqs/raw/12_CD4_R1.fastq.gz	CD4 Season 3.2	Season 3
☐	J160TSZM-12	BLD	XPA		CD4	C7R1AACXX	1	TAAGGCGA-GTAAGGAG	Human	sortedBlood_RNASeq/20150605_CD4/seqs/raw/13_CD4_R1.fastq.gz	CD4 Season 3.2	Season 3
☐	A160TT9Y-11	BLD	XPA		CD4	C7R1AACXX	1	CGTACTAG-ACTGCATA	Human	sortedBlood_RNASeq/20150605_CD4/seqs/raw/14_CD4_R1.fastq.gz	CD4 Season 3.2	Season 3
☐	G160TT99-11	BLD	XPA		CD4	C7R1AACXX	1	AGGCAGAA-AAGGAGTA	Human	sortedBlood_RNASeq/20150605_CD4/seqs/raw/15_CD4_R1.fastq.gz	CD4 Season 3.2	Season 3
☐	E160TTQR-12	BLD	XPA		CD4	C7R1AACXX	1	TCCTGAGC-CTAAGCCT	Human	sortedBlood_RNASeq/20150605_CD4/seqs/raw/16_CD4_R1.fastq.gz	CD4 Season 3.2	Season 3
☐	F160TTPJ-11	BLD	XPA		CD4	C7R1AACXX	1	GGACTCCT-GCGTAAGA	Human	sortedBlood_RNASeq/20150605_CD4/seqs/raw/17_CD4_R1.fastq.gz	CD4 Season 3.2	Season 3
☐	J160TV0G-11	BLD	XPA		CD4	C7R1AACXX	2	TAGGCATG-CTCTCTAT	Human	sortedBlood_RNASeq/20150605_CD4/seqs/raw/18_CD4_R1.fastq.gz	CD4 Season 3.2	Season 3

Experimental Data

Analysis Reports

File share for QC and Analysis Reports



EXAMPLE REPORT

Project Type	RNA-seq
Library Preparation	ClonTech v4
Sequencing Platform	NextSeq550
Data Formatting	bcltofastq-2.19.0
Data Cleaning	Trimmomatic-0.36, TRAILING:13 LEADING:13 ILLUMINACLIP:adapters.fasta:2:30:10 SLIDINGWINDOW:4:20 MINLEN:35
Genome Alignment	STAR_2.6.0c, --twopassMode Basic --runMode alignReads --genomeDir ${GENOME} --readFilesIn ${SAMPLE} --outSAMtype BAM SortedByCoordinate --outSAMstrandField intronMotif --outFilterIntronMotifs RemoveNoncanonical
Reference Genome	GRCm38 + gencode17
Read Quantification1	subread-1.6.1, featurecounts, -s 0 -t exon -g gene_name
Read Quantification2	subread-1.6.1, featurecounts, -M -s 0 -t exon -g gene_name

Report generated on 2019-08-09, 08:53 based on data in: `/scratch/analysis/current_projects/Project_EXAMPLE`

General Statistics

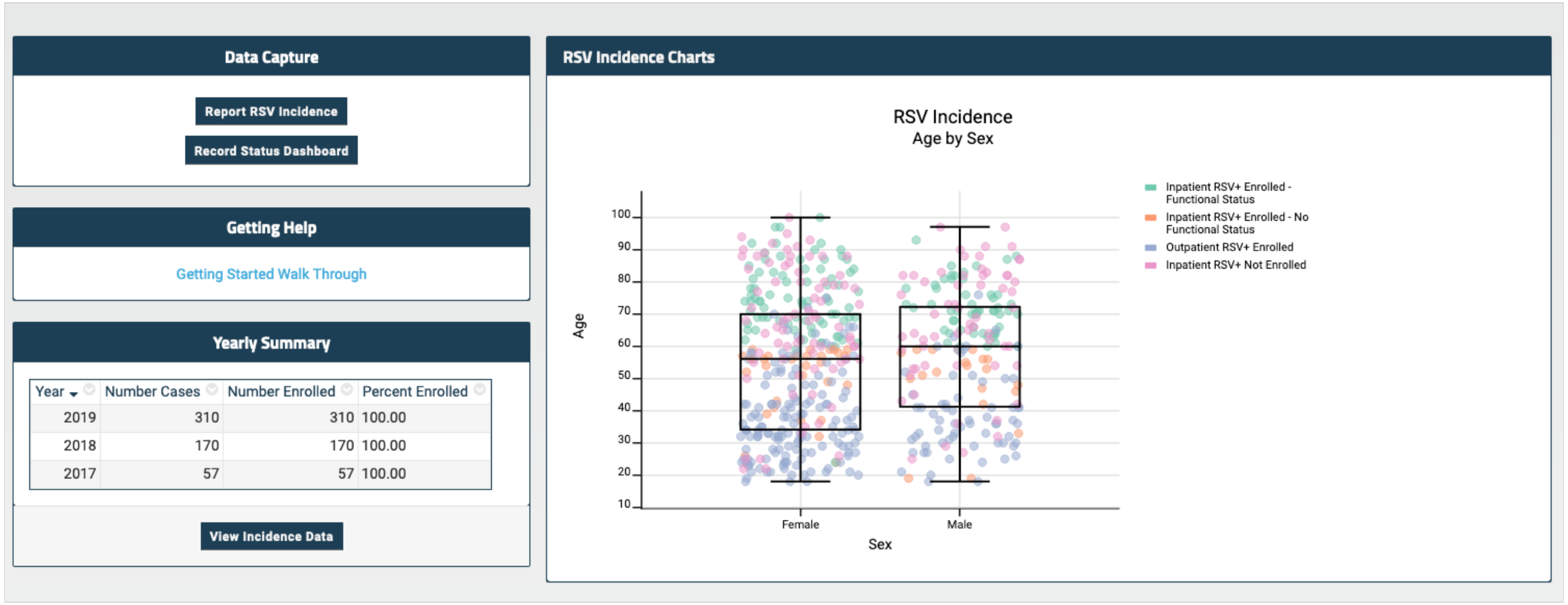
Showing $^{48}/_{48}$ rows and $^{9}/_{11}$ columns.

Sample Name	M Raw Reads	% Dropped	% Dups	% GC	M Seqs	% Aligned	M Aligned	% Assigned	M Assigned
AR000651_0001	28.9	1.3%	81.1%	55%	28.6	63.1%	18.0	58.8%	25.7
AR000668_0001	29.3	1.3%	82.6%	55%	28.9	64.8%	18.7	61.3%	26.7
AR000669_0001	28.8	2.1%	68.0%	51%	28.2	75.2%	21.2	63.3%	22.8
AR000670_0001	25.7	1.2%	55.9%	50%	25.4	90.6%	23.0	65.1%	21.0
AR000805_0001	30.7	1.3%	59.2%	50%	30.3	89.8%	27.2	70.1%	25.8
AR000811_0001	32.6	1.4%	83.6%	55%	32.2	63.3%	20.3	58.2%	29.4
AR000812_0001	29.5	1.5%	84.1%	55%	29.0	65.6%	19.0	65.5%	27.8
AR000813_0001	33.7	1.2%	70.2%	51%	33.3	80.8%	26.9	62.6%	28.5
AR000888_0001	29.7	1.2%	57.8%	51%	29.4	91.3%	26.8	69.6%	24.9

Dashboard and Reports

Dashboards

Project Specific



Data Capture

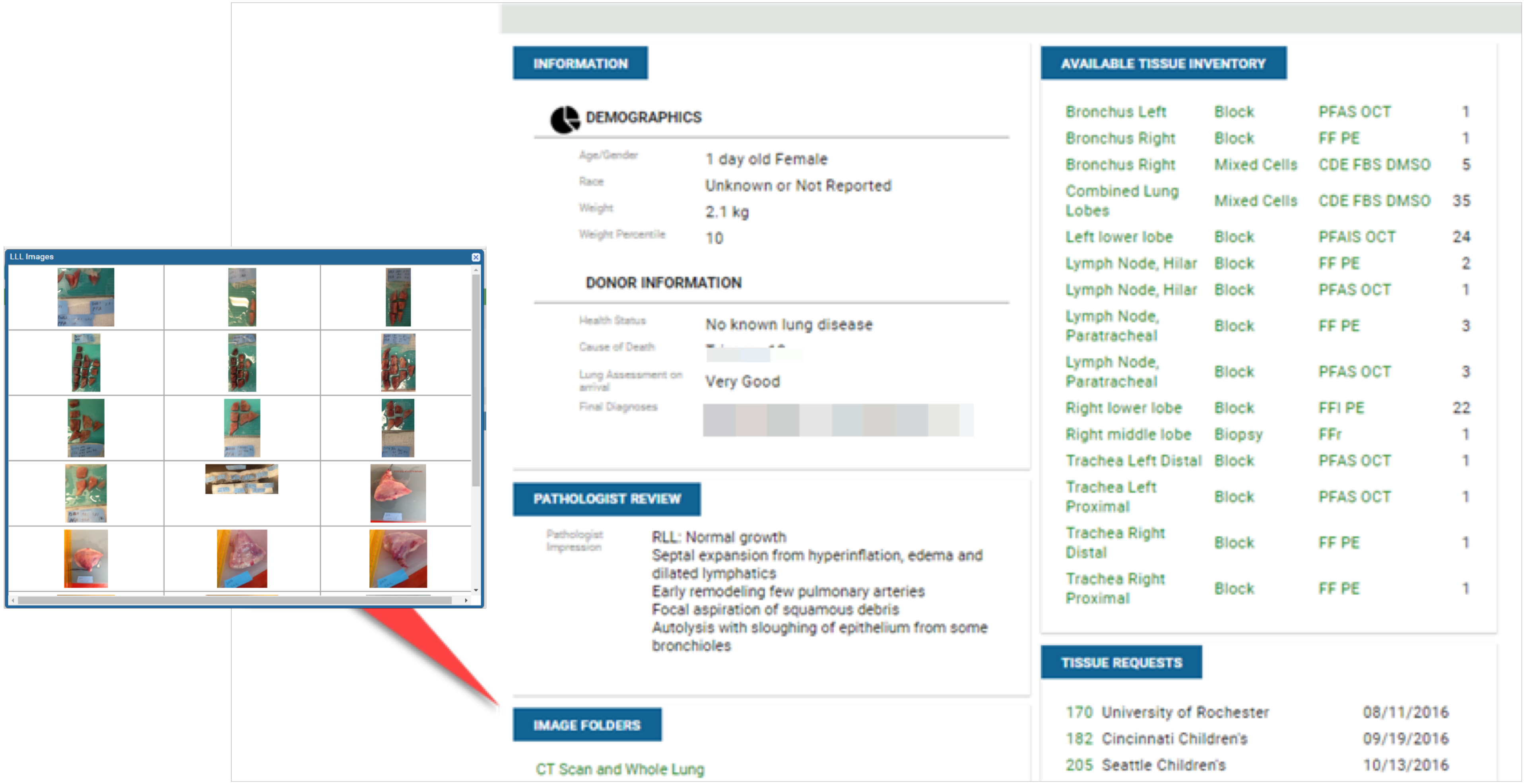
Getting Help

Yearly Summary

Year	Number Cases	Number Enrolled	Percent Enrolled
2019	310	310	100.00
2018	170	170	100.00
2017	57	57	100.00

RSV Incidence Charts

Dashboards



INFORMATION

DEMOGRAPHICS

Age/Gender	1 day old Female
Race	Unknown or Not Reported
Weight	2.1 kg
Weight Percentile	10

DONOR INFORMATION

Health Status	No known lung disease
Cause of Death	[redacted]
Lung Assessment on arrival	Very Good
Final Diagnoses	[redacted]

PATHOLOGIST REVIEW

Pathologist Impression

RLL: Normal growth
Septal expansion from hyperinflation, edema and dilated lymphatics
Early remodeling few pulmonary arteries
Focal aspiration of squamous debris
Autolysis with sloughing of epithelium from some bronchioles

IMAGE FOLDERS

CT Scan and Whole Lung

AVAILABLE TISSUE INVENTORY

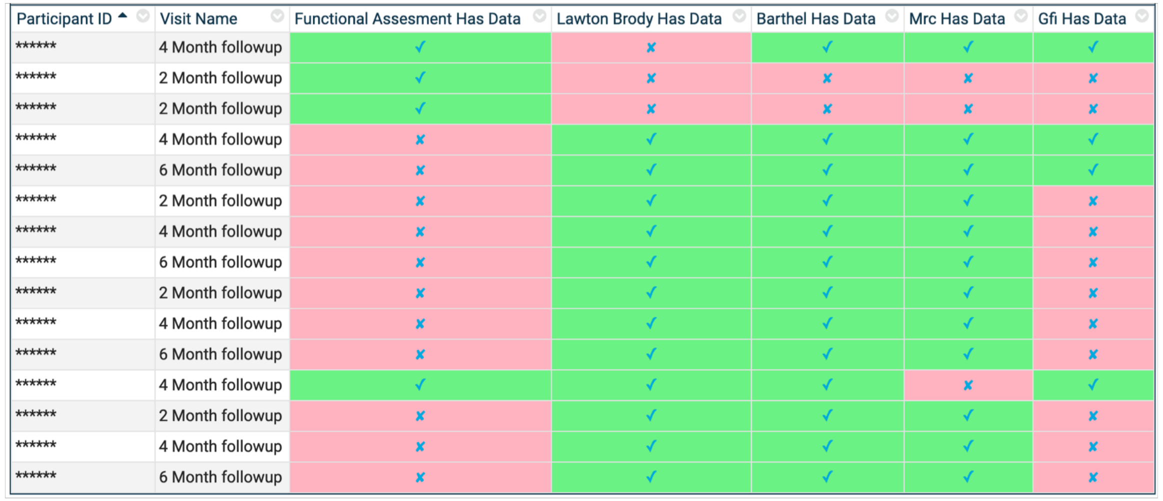
Bronchus Left	Block	PFAS OCT	1
Bronchus Right	Block	FF PE	1
Bronchus Right	Mixed Cells	CDE FBS DMSO	5
Combined Lung Lobes	Mixed Cells	CDE FBS DMSO	35
Left lower lobe	Block	PFAIS OCT	24
Lymph Node, Hilar	Block	FF PE	2
Lymph Node, Hilar	Block	PFAS OCT	1
Lymph Node, Paratracheal	Block	FF PE	3
Lymph Node, Paratracheal	Block	PFAS OCT	3
Right lower lobe	Block	FFI PE	22
Right middle lobe	Biopsy	FFr	1
Trachea Left Distal	Block	PFAS OCT	1
Trachea Left Proximal	Block	PFAS OCT	1
Trachea Right Distal	Block	FF PE	1
Trachea Right Proximal	Block	FF PE	1

TISSUE REQUESTS

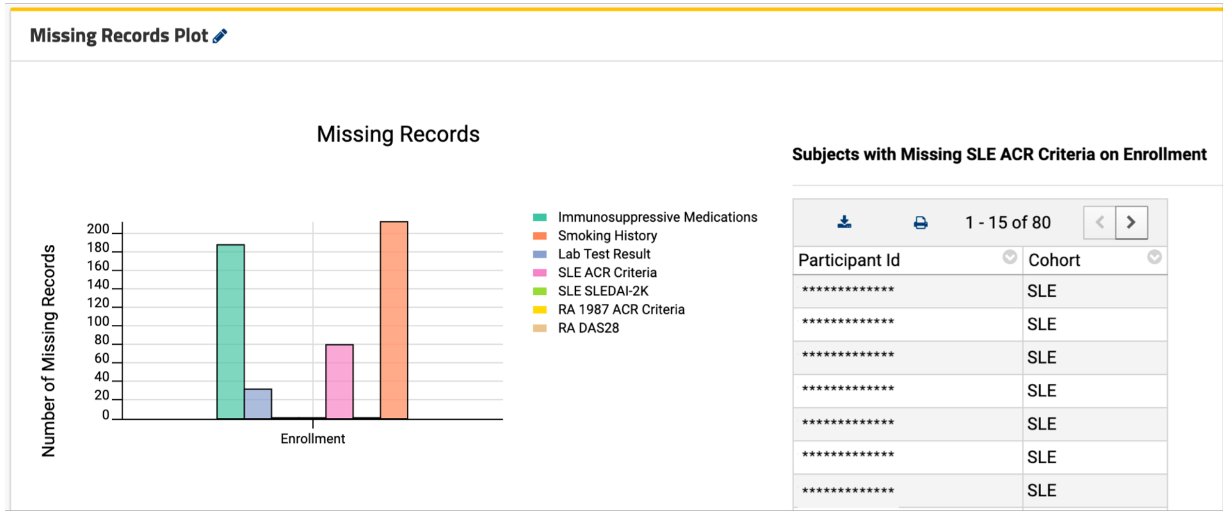
170	University of Rochester	08/11/2016
182	Cincinnati Children's	09/19/2016
205	Seattle Children's	10/13/2016

Interactive Reports

Quality Control



Participant ID ▲	Visit Name	Functional Assesment Has Data	Lawton Brody Has Data	Barthel Has Data	Mrc Has Data	Gfi Has Data
******	4 Month followup	✓	✗	✓	✓	✓
******	2 Month followup	✓	✗	✗	✗	✗
******	2 Month followup	✓	✗	✗	✗	✗
******	4 Month followup	✗	✓	✓	✓	✓
******	6 Month followup	✗	✓	✓	✓	✓
******	2 Month followup	✗	✓	✓	✓	✗
******	4 Month followup	✗	✓	✓	✓	✗
******	6 Month followup	✗	✓	✓	✓	✗
******	2 Month followup	✗	✓	✓	✓	✗
******	4 Month followup	✗	✓	✓	✓	✗
******	6 Month followup	✗	✓	✓	✓	✗
******	4 Month followup	✓	✓	✓	✗	✓
******	2 Month followup	✗	✓	✓	✓	✗
******	4 Month followup	✗	✓	✓	✓	✗
******	6 Month followup	✗	✓	✓	✓	✗

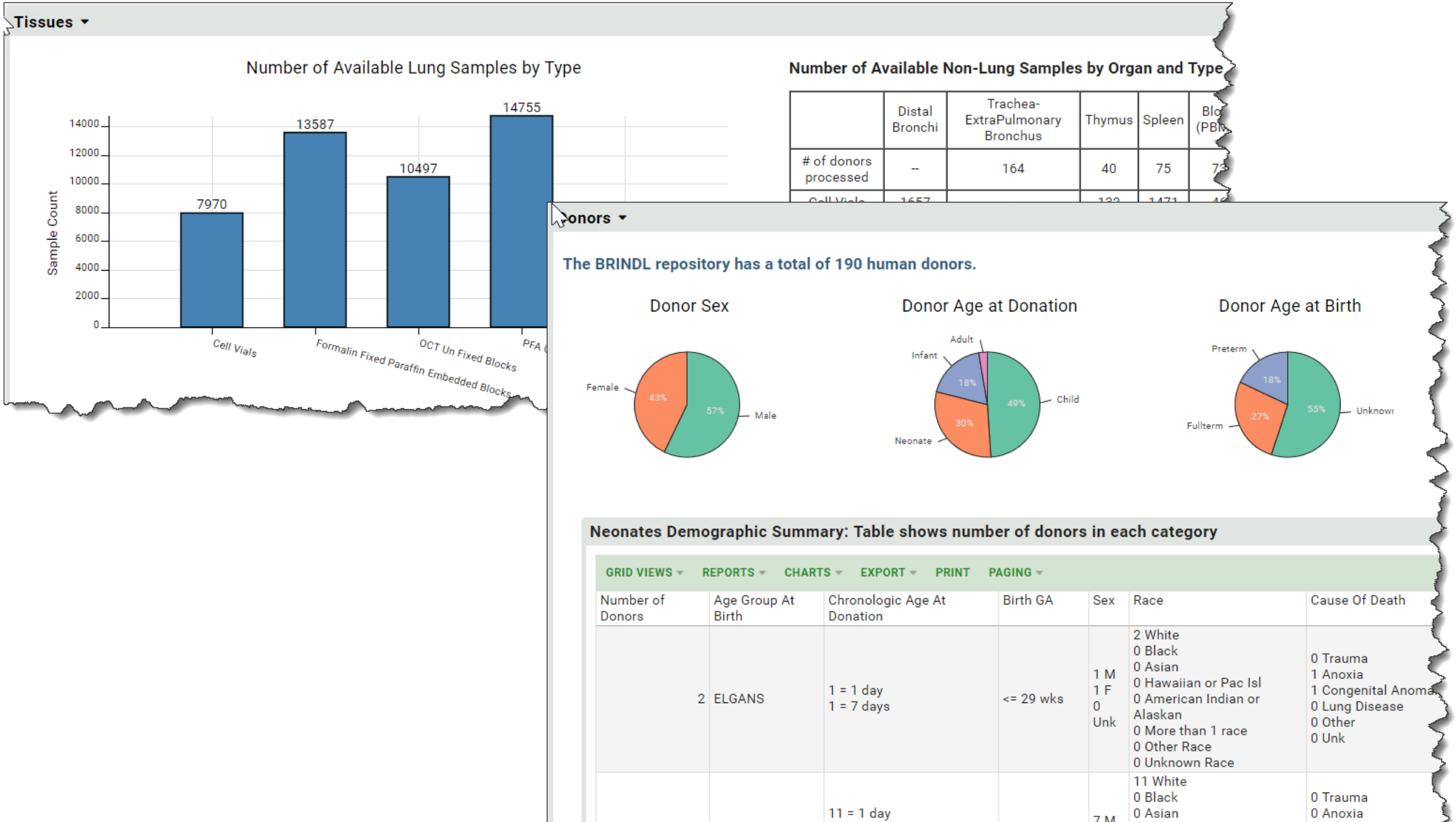


Subjects with Missing SLE ACR Criteria on Enrollment

Participant Id	Cohort
*************	SLE
*************	SLE
*************	SLE
*************	SLE
*************	SLE
*************	SLE
*************	SLE

Visualizations

Real-time Reporting



Tissues

Number of Available Lung Samples by Type

Number of Available Non-Lung Samples by Organ and Type

	Distal Bronchi	Trachea-ExtraPulmonary Bronchus	Thymus	Spleen	Blo (PBM
# of donors processed	--	164	40	75	7
Cell Vials	1657		132	1471	4

onors

The BRINDL repository has a total of 190 human donors.

Donor Sex

Donor Age at Donation

Donor Age at Birth

Neonates Demographic Summary: Table shows number of donors in each category

Number of Donors	Age Group At Birth	Chronologic Age At Donation	Birth GA	Sex	Race	Cause Of Death
2	ELGANS	1 = 1 day 1 = 7 days	<= 29 wks	1 M 1 F 0 Unk	2 White 0 Black 0 Asian 0 Hawaiian or Pac Isl 0 American Indian or Alaskan 0 More than 1 race 0 Other Race 0 Unknown Race	0 Trauma 1 Anoxia 1 Congenital Anoma 0 Lung Disease 0 Other 0 Unk
		11 = 1 day		7 M	11 White 0 Black 0 Asian	0 Trauma 0 Anoxia

Visualizations

Real-time Reporting

Clinical Adjudication Summary

Dataset: Final Adjudication Results (Form 10), All Visits

Transcriptional Profiling

Future Work

Integration with additional clinical data sources:

EMR and/or Clinical Data Warehouse

Better Integration with OMERO:

Image Viewing and Analysis

Managing Neuroimaging (fMRI) data:

Brain Imaging Data Structure (BIDS)

Integration with computational workflow engines:

Galaxy and CWL/WDL

Acknowledgements

RDIA Team

Jeanne Holden-Wiltse, MPH, MBA, Director
Anthony Corbett, MS, Technical Lead
Sanjukta Bandyopadhyay, MS
Jennifer Dutra
Jill Payne
Alicia Tyrell, MBA
Andrew Watts, MA
Jeff Williams
Zidian Xie, PhD
Mike Conklin
Kim Kaukeinen

LabKey Corporation
CTSI Informatics
Center for Integrated Research Computing

Major Funding

National Institutes of Health

BIM

CFAR
CENTERS FOR AIDS RESEARCH

LungMAP
Molecular Atlas of Lung Development Program

CROFT
WNY CENTER FOR RESEARCH ON FLAVORED TOBACCO

UNIVERSITY of ROCHESTER MEDICAL CENTER
RPRC
NIAID Respiratory Pathogens Research Center

UNIVERSITY of ROCHESTER MEDICAL CENTER
NYICE
New York Influenza Center of Excellence

MELIORA
UNIVERSITY of ROCHESTER MEDICAL CENTER
CTSI
CLINICAL & TRANSLATIONAL SCIENCE INSTITUTE