Structured Data Capture of Biotherapeutic Design & Analytical Results to Enable the Development of Predictive Models

2017-10-05

J. Alex Taylor

2017 LabKey User Conference

Mission: Design and apply innovative technologies to dramatically expand global access to biotherapeutics

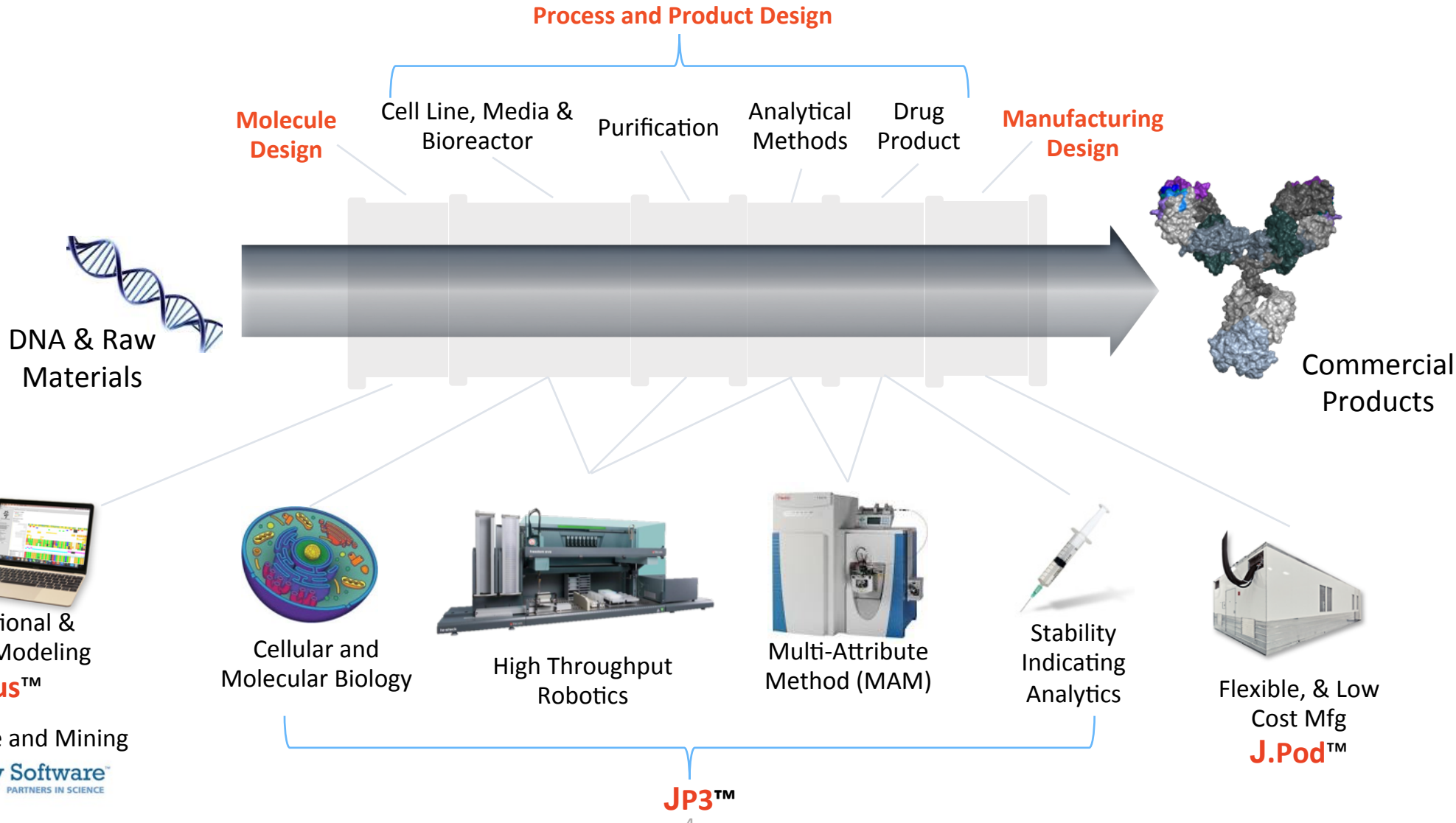
Outline

- Brief intro to Just
- Why we chose LabKey to play a critical role in meeting our goals
- Abacus (focus on Ab Classification & LabKey integration)
- Selection of LabKey-integrated solutions to automating data capture
  - Bioreactor sample generation
  - DNA sequencing assay
  - Cell line development automation
- JStruct
- Using data to create predictive models

Just was formed in 2014 to expand global access to important biotherapeutics

- Over 80% of the planet cannot afford most of the breakthrough therapeutics of modern biotechnology
- Just was formed to reduce the cost of biologics by at least 10x, making these therapeutics more accessible to a global population
- This will help expand markets for current therapies and lay a foundation for new therapeutic approaches

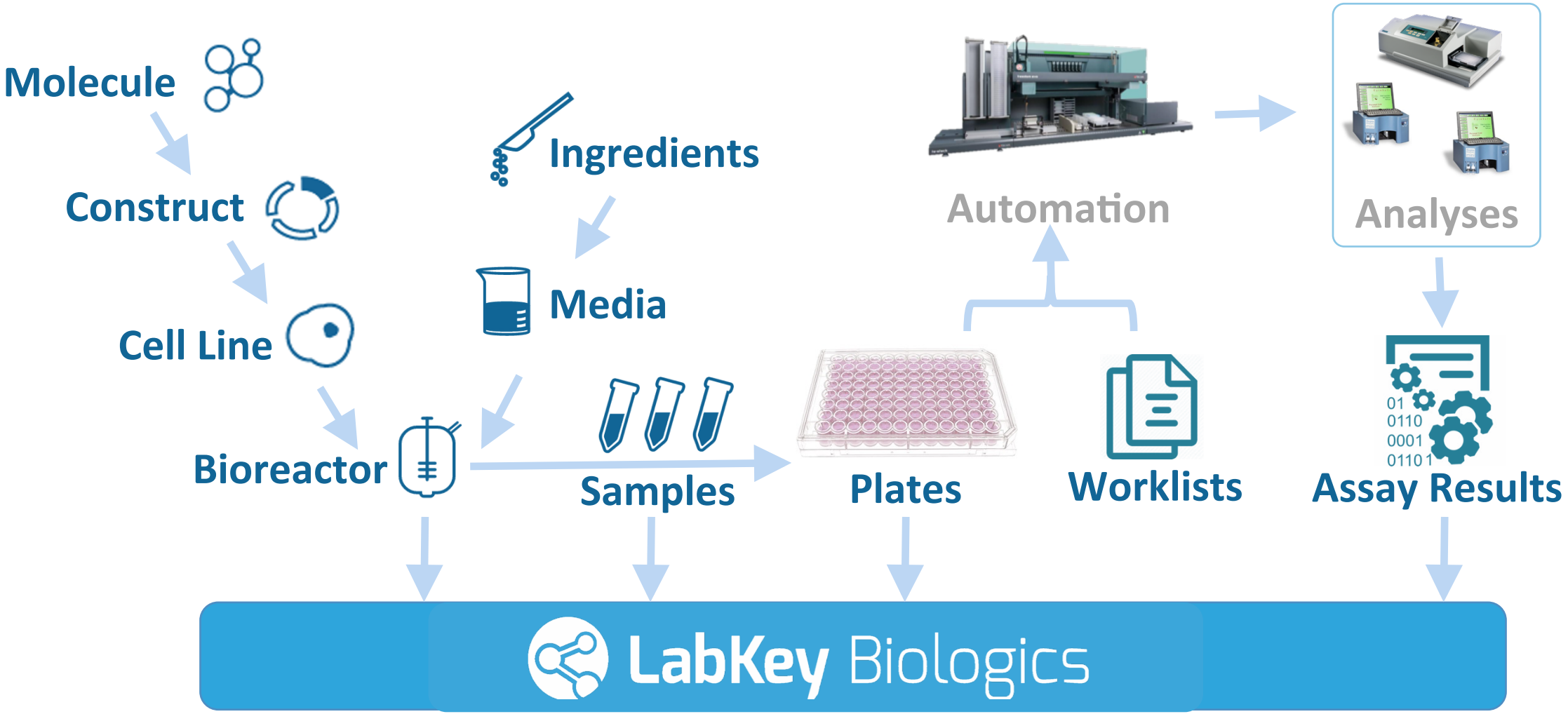
J.Design will accelerate development, improve product throughput and reduce the over cost of biologics



About LabKey Biologics

- A collaboration between Just and LabKey, with input gathered from an advisory council of Pharma & Biotech representatives
- Goal: Build upon the foundation of LabKey Server to develop an application to support large molecule development
- Core components:
  - Entity Registration
  - Assay Data Integration
  - Querying & Analysis
  - Media Registration
  - Workflow Management
- Project kicked-off Fall 2015 | Product launched Spring 2017

LabKey Biologics Registers Entities, Tracks Samples, and Captures Assay Data



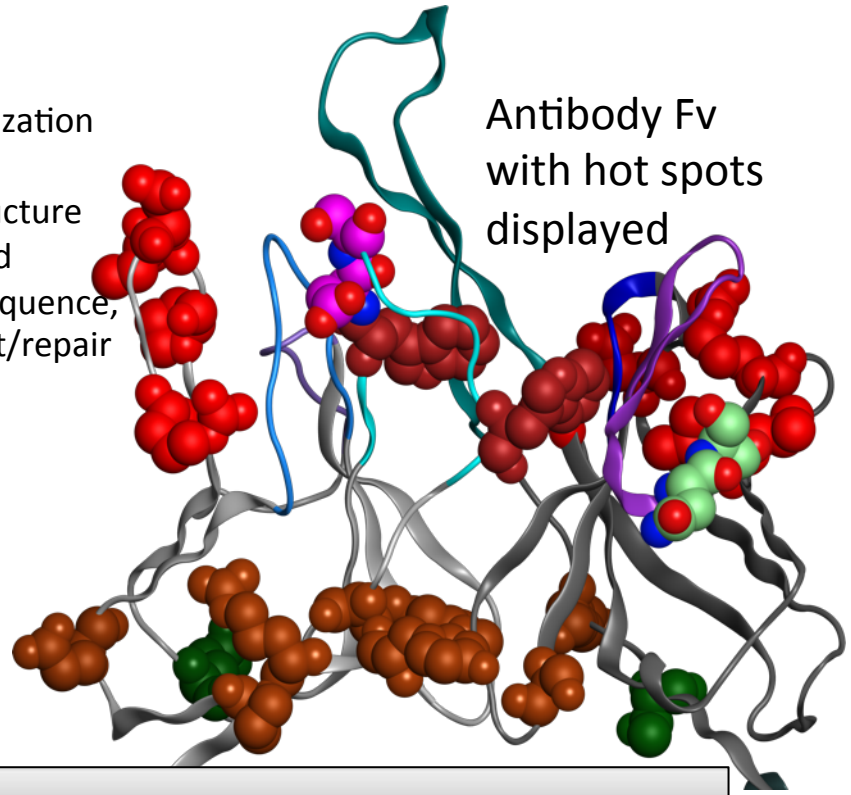
Abacus™ is a system for the analysis and engineering of antibodies and antibody-like molecules

Hot Spot Analysis
- Covariance violations (stability) (Gunasekaran 2004)
- Isomerization/Deamidation
- Much more…
- Issues communicated downstream

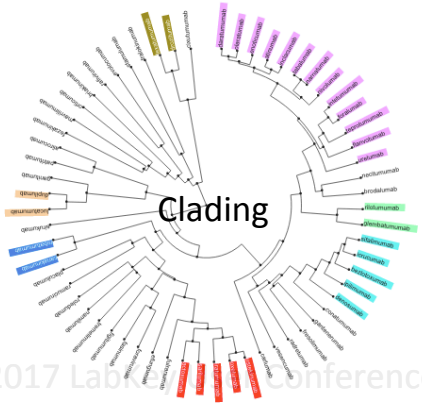
Variant Designs
- Utilize structure analysis and calculations
- Variants tracked across analyses
- Combinatorial variant sequence production
- Push to registration system to enable associated data capture
- Full analysis report generation

Sequence Analysis
- Structure-based numbering system (based on Honegger 2001)
- Full alignments with annotations
- Positional frequency analysis across specified reference antibodies
- Clading for diversity assessment and sibling engineering
- Germline alignments within or across subclasses
- Structure and stability-based CDR grafting / humanization
- Physical properties calculations

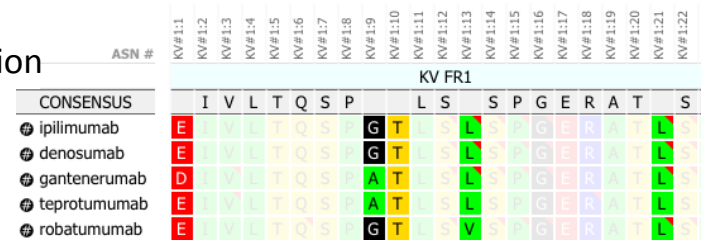
Machine Learning
- Antibody characterization data utilized
- Wide variety of structure attributes calculated
- Paired structure, sequence, data used to predict/repair characterization



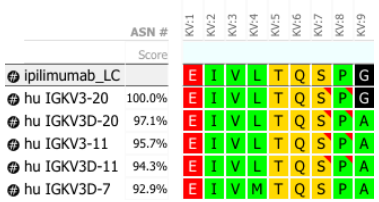
Antibody Fv with hot spots displayed



Clading



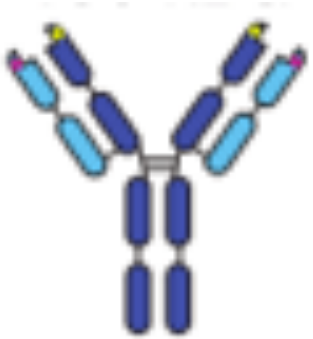
Structure-based alignments

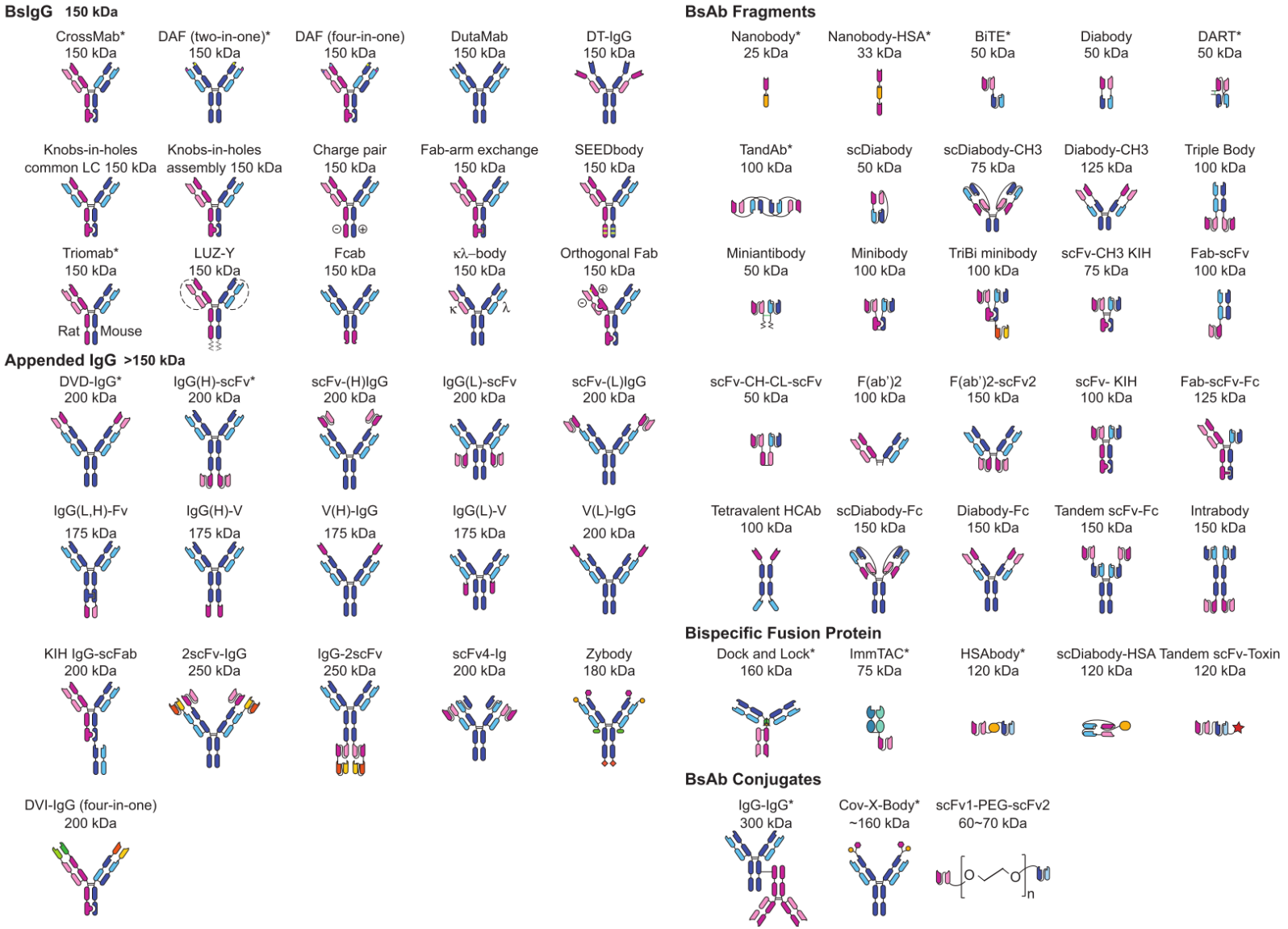


Germline

Hot spot detection

Ab domains can be modified and rearranged like building blocks





98

C. Spiess et al. / Molecular Immunology 67 (2015) 95–106

Just.

Antibody Classification Steps

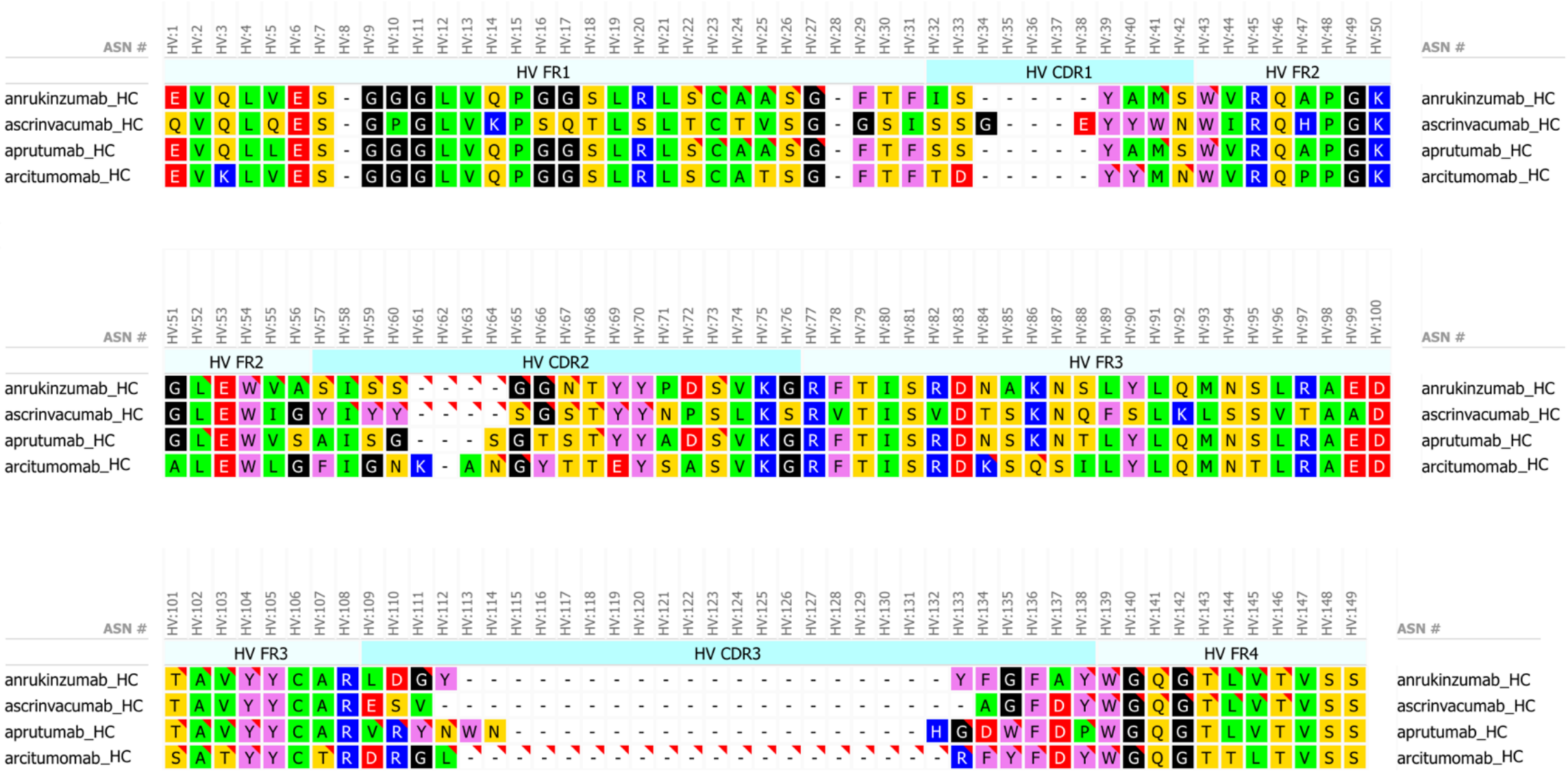
1. Initial detection performed with region (domain) PSSMs
2. Second pass matches to ASN-aligned germline sequences*
3. Region alignment tuning (i.e. CDR gapping)
4. Resolve region overlaps if present
5. Attempt to assign unrecognized regions to unstructured regions (hinge) and sequence parts
6. Assignment of a chain format based on the region pattern

**Available germlines include human, mouse, rat, rabbit and 6 other species with partial coverage.*
Users can provide their own:

- *Germline data as XML*
- *Custom sequence parts for classification*
- *Chain and structure formats.*

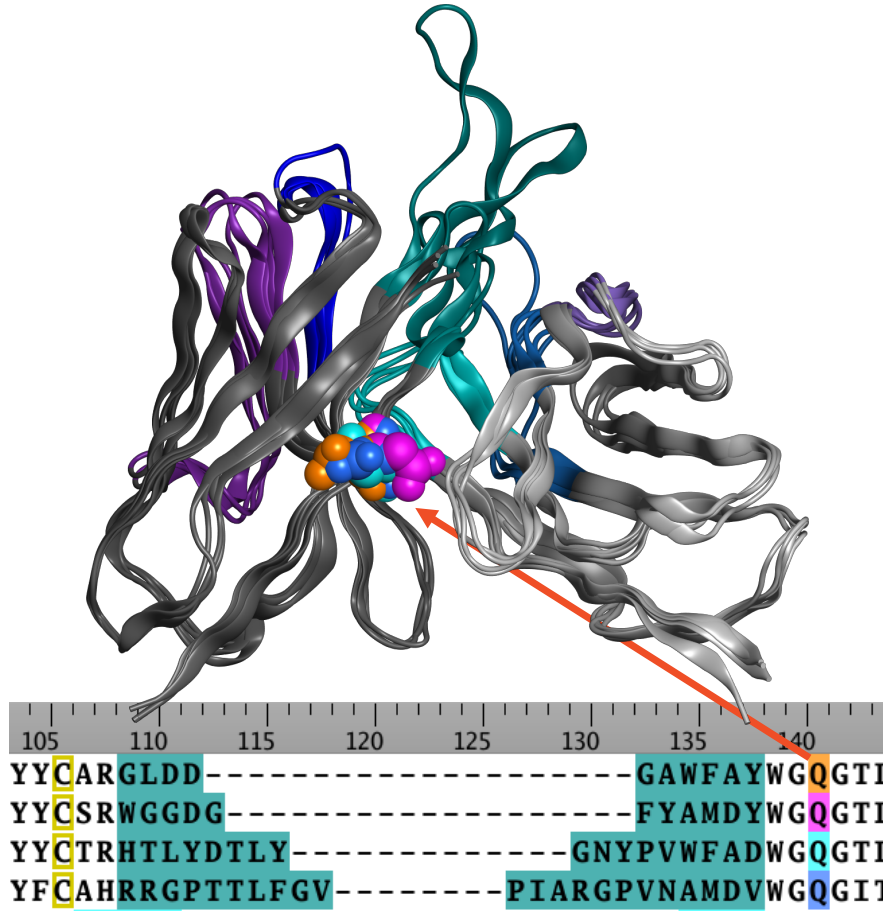
Result is a structurally-aligned sequence with region and feature annotation

Ab Classification Results



ASN (Antibody Structural Numbering)

- Builds upon AHo variable region numbering (Honegger 2001)
- Unique assignment for every residue in the sequence
- Constant region Ig domains built initially from SCOP structure-based alignments
- Allows for insertions when necessary
- Allows for flexible arrangement of regions and multiple regions of the same type



HV#2:141.1

Region prefix

Region instance #

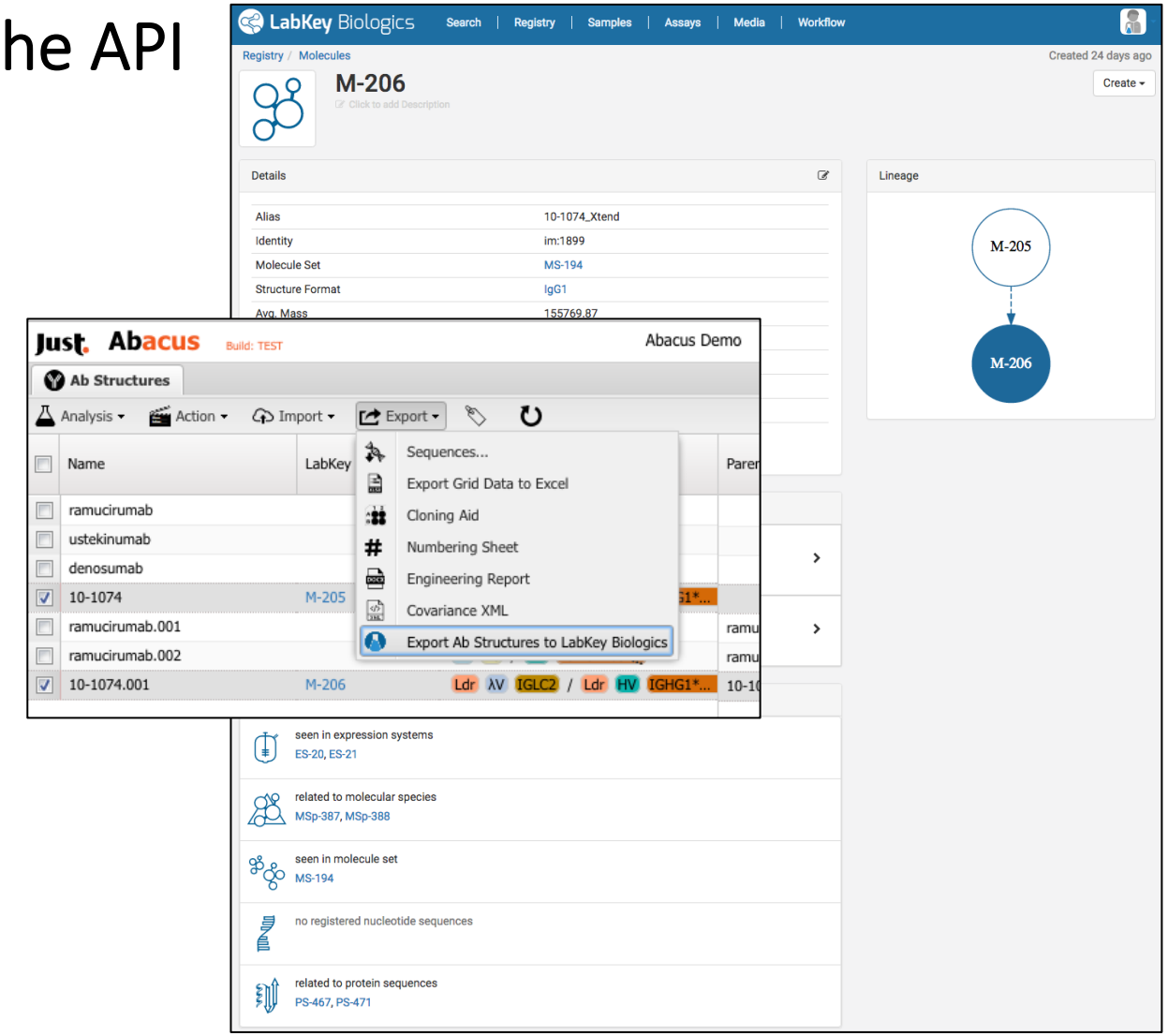
Aligned position

Aligned sub-position

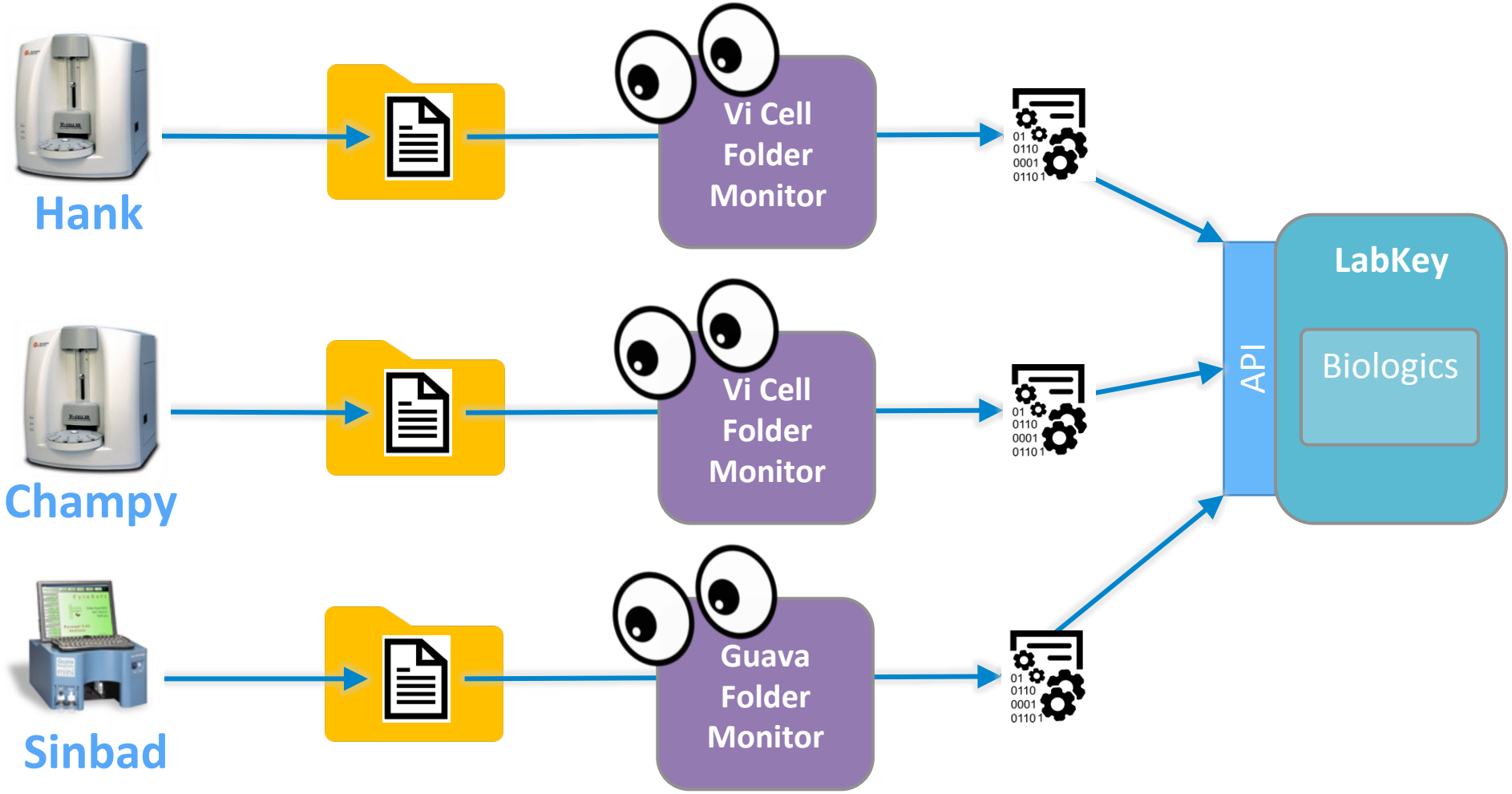
Structure-based alignment allows for the comparison of positions across multiple proteins

Molecules and their variants can be pushed to LabKey Biologics via the API

- Works in bulk
- Registers nucleotide and protein sequences as well as molecules
- Sequences registered once by identity
- Light and Heavy chains are associated
- Antibody features are auto-detected
- Parental lineage is maintained
- LKB then allows registration of constructs, expression systems, etc., with **associated data**

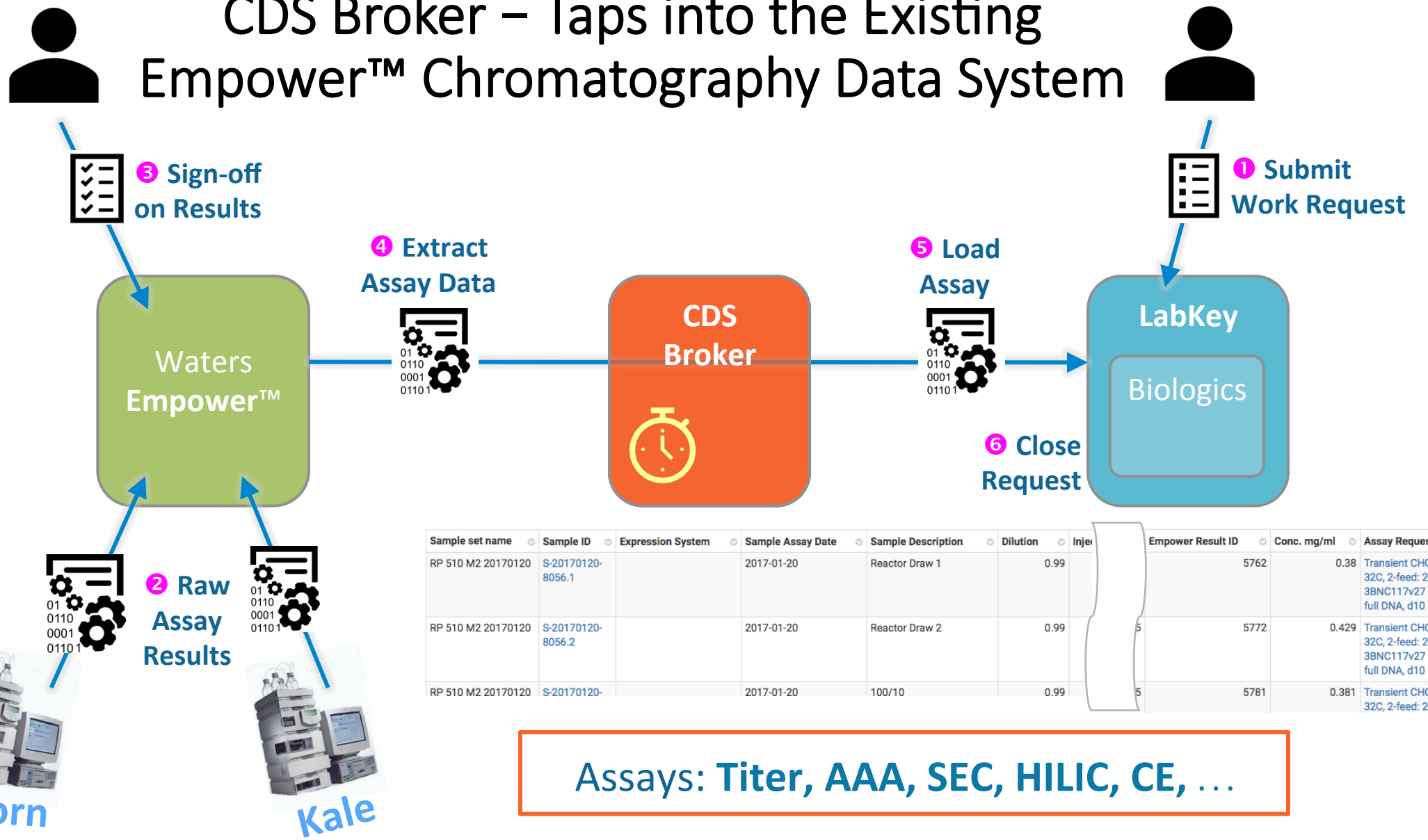


Simple Assay Data Import: Folder Monitor Daemons



Sample ID	Day	VCD (E6 cells/ml)	Total Cell Conc. (E6 cells/ml)	Viability (%)	Avg. Diameter (microns)	Run Date	Operator	Name
S-20170331-33		3.398	3.575	95	16.81	2017-03-31		Hank

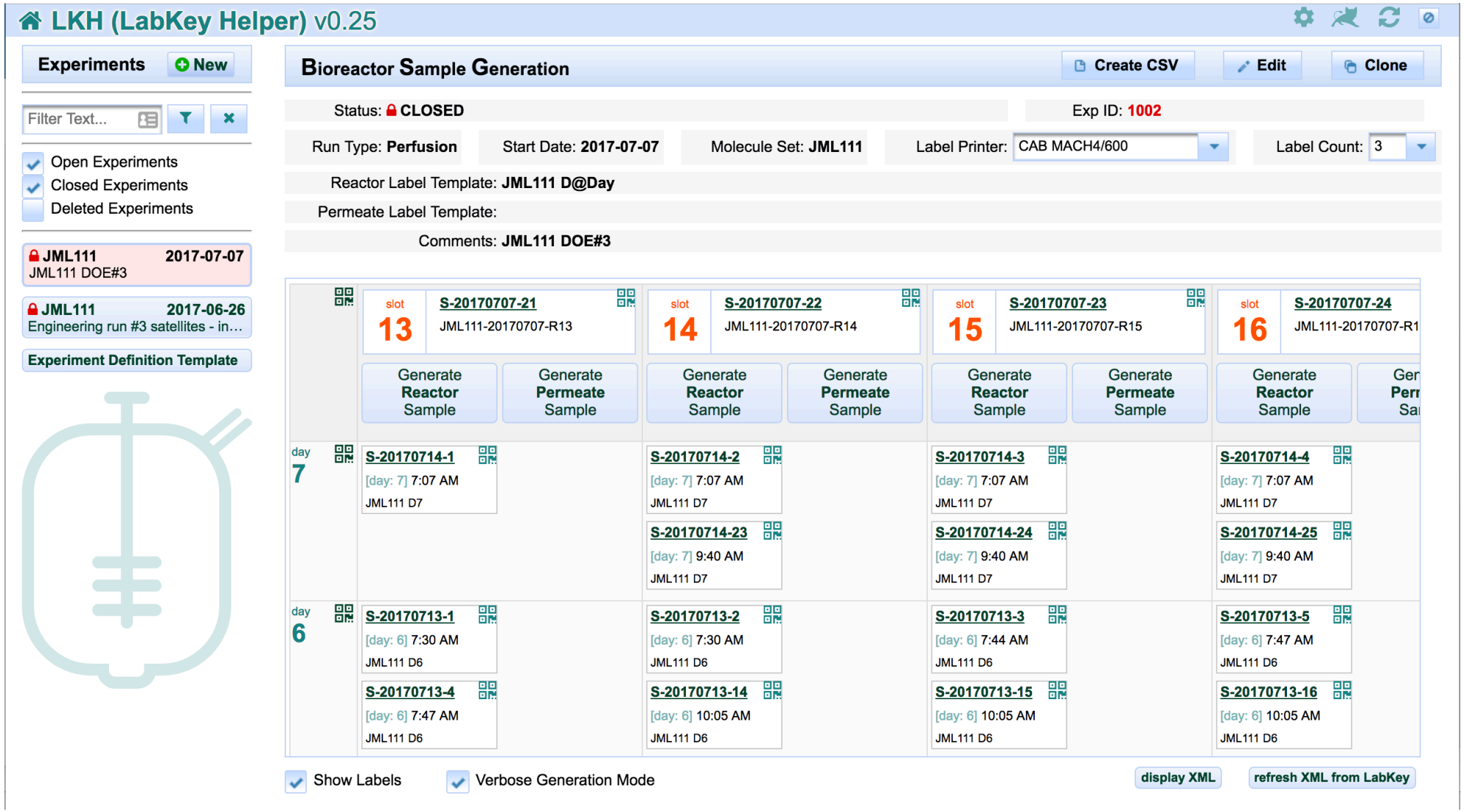
CDS Broker – Taps into the Existing Empower™ Chromatography Data System



Sample set name	Sample ID	Expression System	Sample Assay Date	Sample Description	Dilution	Inje	Empower Result ID	Conc. mg/ml	Assay Request	Slop
RP 510 M2 20170120	S-20170120-8056.1		2017-01-20	Reactor Draw 1	0.99		5762	0.38	Transient CHO-S, 32C, 2-feed: 20mL 3BNC117v27 1:1, full DNA, d10	
RP 510 M2 20170120	S-20170120-8056.2		2017-01-20	Reactor Draw 2	0.99		5772	0.429	Transient CHO-S, 32C, 2-feed: 20mL 3BNC117v27 1:1, full DNA, d10	
RP 510 M2 20170120	S-20170120-		2017-01-20	100/10	0.99		5781	0.381	Transient CHO-S, 32C, 2-feed: 20mL	

Assays: **Titer, AAA, SEC, HILIC, CE, …**

Simplifying Bioreactor Sample Generation



DNA Sequencing Assay

LKH (LabKey Helper) v0.25

DNA Sequencing

[1] Input/Review → [2] Samples/Request

Sequence Input

Review Input

Nucleotide Sequences: @NucSeq

NS-1062
NS-1063
NS-1064
NS-1065
NS-1066
NS-1067
NS-1068

Clones per Sequence: 4 @Clone

Vector: V-15 @Vector

Fwd Sequencing Primer: CMV-Forward

Rev Sequencing Primer: 10-336528137

Sample Label Template: Dna seq. sample for @NucSeq, Clone: @Clone

Notes:

Review

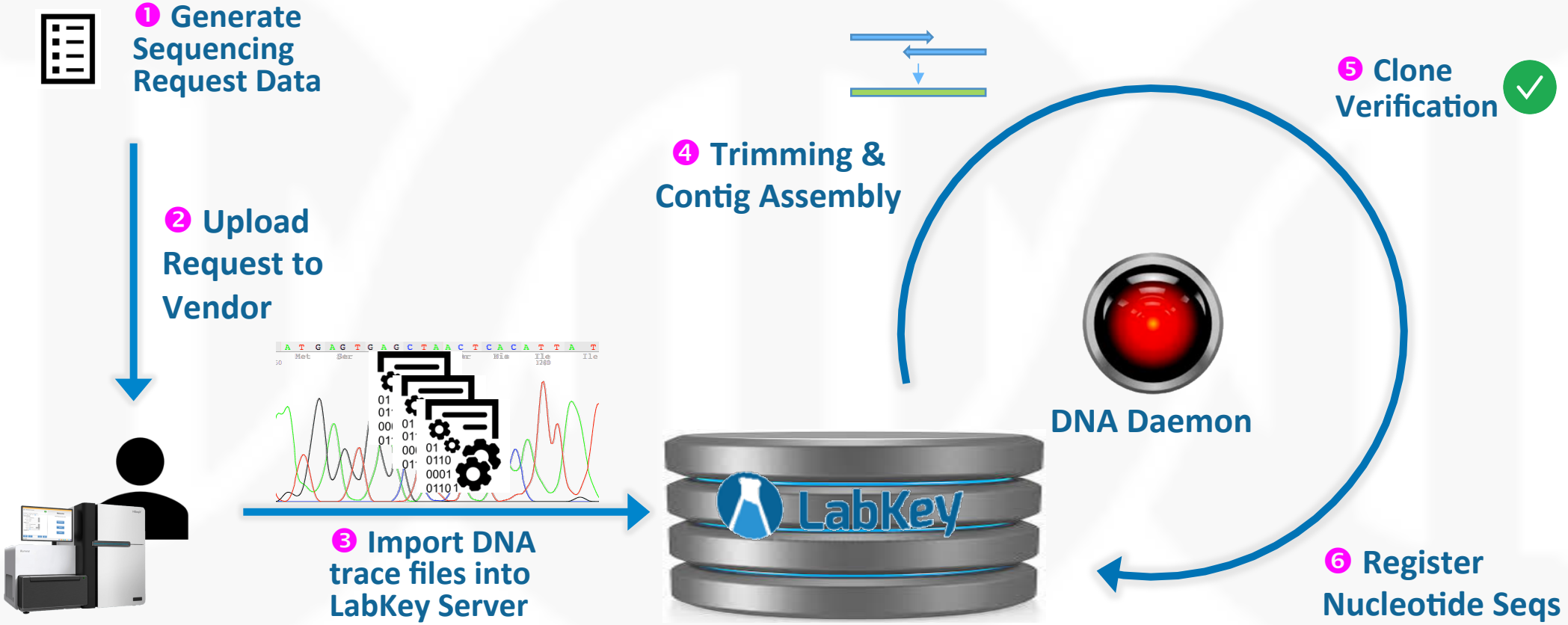
→ Create Samples

Protein Seq.	Ab Chain Format	Nucleotide Seq.	Nuc. Sequence Length	Genewiz Primer(s)*	Sample Label	Notes
PS-1493	IgG1 Heavy Chain	NS-1062	1395	CMV-Forward;10-336528137	Dna seq. sample for NS-1062, Clone: <clone_number>	
PS-1495	Kappa Light Chain	NS-1063	696	CMV-Forward	Dna seq. sample for NS-1063, Clone: <clone_number>	
PS-1496	IgG1 Heavy Chain	NS-1064	1404	CMV-Forward;10-336528137	Dna seq. sample for NS-1064, Clone: <clone_number>	
PS-1498	Kappa Light Chain	NS-1065	696	CMV-Forward	Dna seq. sample for NS-1065, Clone: <clone_number>	
PS-1499	IgG1 Heavy Chain	NS-1066	1407	CMV-Forward;10-336528137	Dna seq. sample for NS-1066, Clone: <clone_number>	
PS-1501	Kappa Light Chain	NS-1067	699	CMV-Forward	Dna seq. sample for NS-1067, Clone: <clone_number>	
PS-1502	IgG1 Heavy Chain	NS-1068	1404	CMV-Forward;10-336528137	Dna seq. sample for NS-1068, Clone: <clone_number>	

Click and drag rows to re-order

*Rev Primer used if Nuc. Seq. length excedes 900 base pairs

DNA Sequencing Assay



Liquid Handler Worklists for Cell Line Development Automation

LKH (LabKey Helper) v0.23 — Alex Taylor

CLD Automation — Reset

[1] Enter Passaging Data → [2] Review and Configure → **[3] Download Worklists** → [4] Plates Created!

Download Worklists

Derive new LabKey Sample IDs in destination plate(s)

← Review and Configure | Register Plates and Download Worklists

Protocol: **96-96**
Compact? **false**

Cell Transfer Tip: **DiTi 1000uL SBS LiHa** Min/Max Capacity: **25 µL / 940 µL**
Media Transfer Tip: **DiTi 1000uL SBS LiHa** Min/Max Capacity: **25 µL / 940 µL**

Worklist Display | Destination Plates/Samples

```
C;Passaging Protocol
C;Generated 2017.09.15 15:05 PM
C;**************************************************
C;Setting tips for the media transfer
B;
S;7
C;Media transfer
C;Place reagent at deck position Reagent-1
C;Place dest at deck position Dest-1
A;Reagent-1;;Trough 350ml;1;;133;Trough detect 1tip;;
D;Dest-1;;96 Well DeepWell;1;;133;Trough detect 1tip;;
W;
A;Reagent-1;;Trough 350ml;2;;93;Trough detect 1tip;;
D;Dest-1;;96 Well DeepWell;3;;93;Trough detect 1tip;;
```

Barcoding and scanning

S-20170906-105

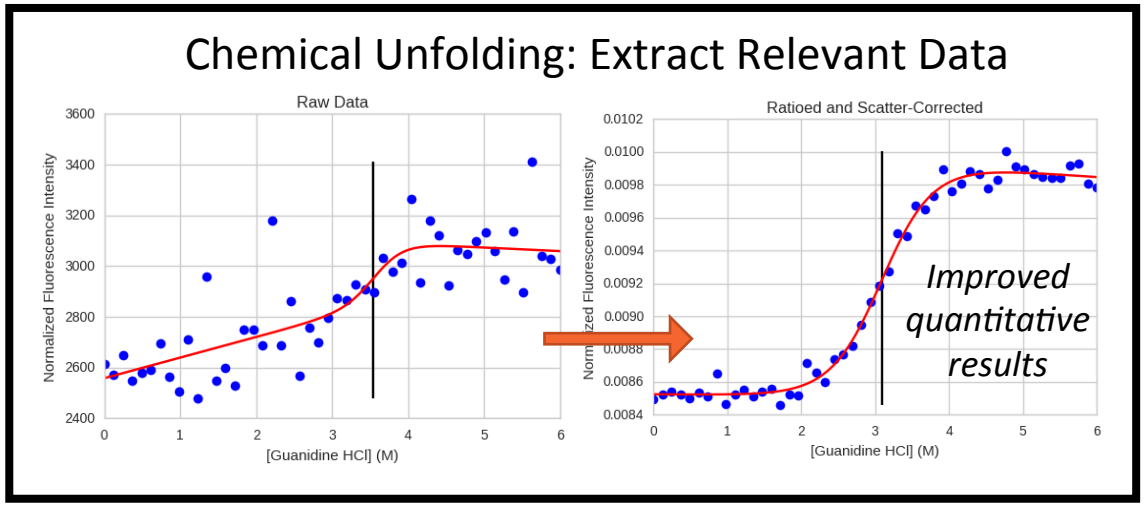
Engineering Run Lot #2

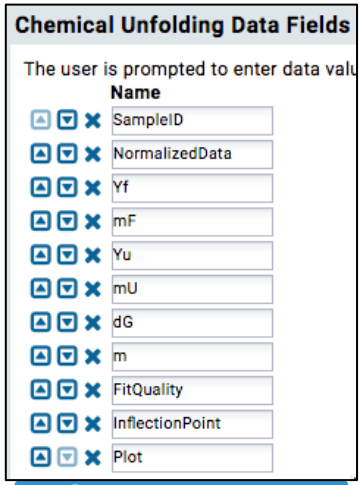
P-20170725-99

- Barcode everything!
- LabKey Inventory module can generically store plates as “boxes” and track sample well locations
- Currently using BarTender to construct label templates and as a generic printing webservice

Raw data interpretation can be subjective, so we are automating data analysis to improve results from characterization assay data

- Develop analysis methods that process raw assay results into values that are
  - Defined
  - Objective
  - Consistent
  - Reliable
  - Curated
  - Structured
  - Informative

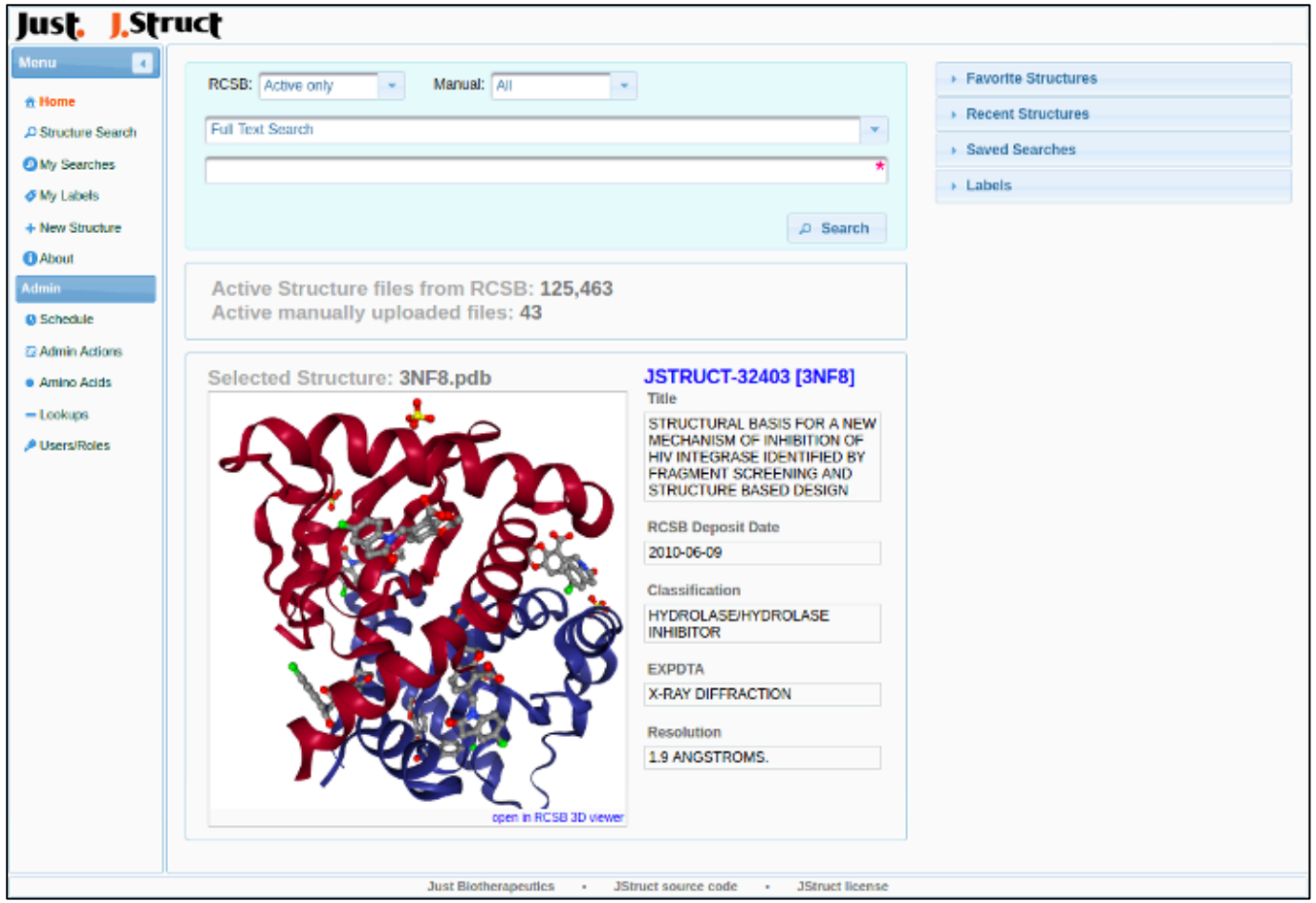




- Which leads to data that is more useful for machine learning
- Data analysis automation also improves analysis throughput and capacity

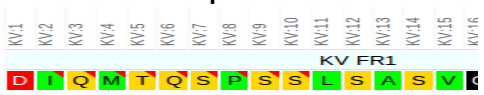
J.Struct - storage and retrieval of public and proprietary three-dimensional macromolecular structures

- **JStruct** is an open source stand-alone software tool designed and developed by Just Biotherapeutics for the storage and retrieval of public and proprietary three-dimensional macromolecular structures as well as information parsed from their content.
- **JStruct** provides a robust and secure means to store, access, search and utilize both public and proprietary structures within the same system. With JStruct it is possible to search, calculate and compare structures across a full set of files while keeping internal sequences from being publicly exposed.
- The **JStruct** open source code is available via Bitbucket and is licensed under the GNU General Public License.
- **JStruct** will facilitate the calculation/storage/retrieval of structure attributes that will be used for machine learning.
- To download or for more information: http://www.justbiotherapeutics.com/jstruct



Machine Learning in Antibody Engineering Derives from Sequence, Structure, Data

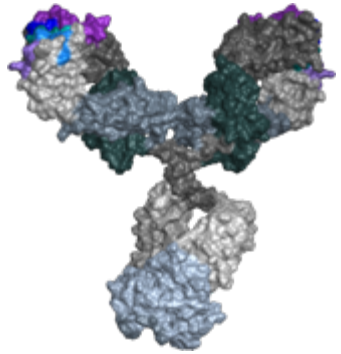
Sequence



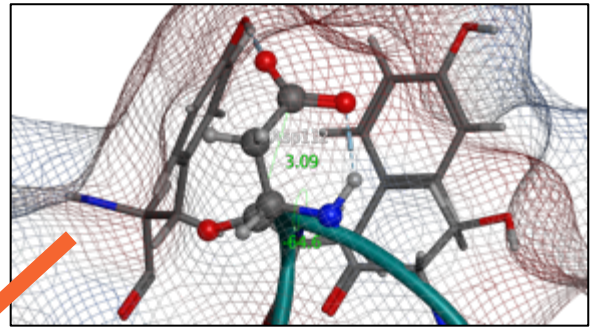
Protein Synthesis

Measured Physical Properties

Structural Model



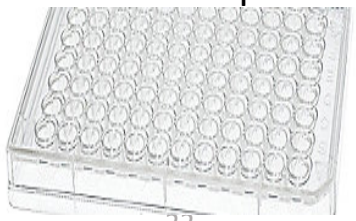
Structural Features



(100s – 100,000s of features per variant)

Machine Learning

Estimated Properties



Recommended Sequence Changes

Devise new variants and estimate the likely properties of these new variants.

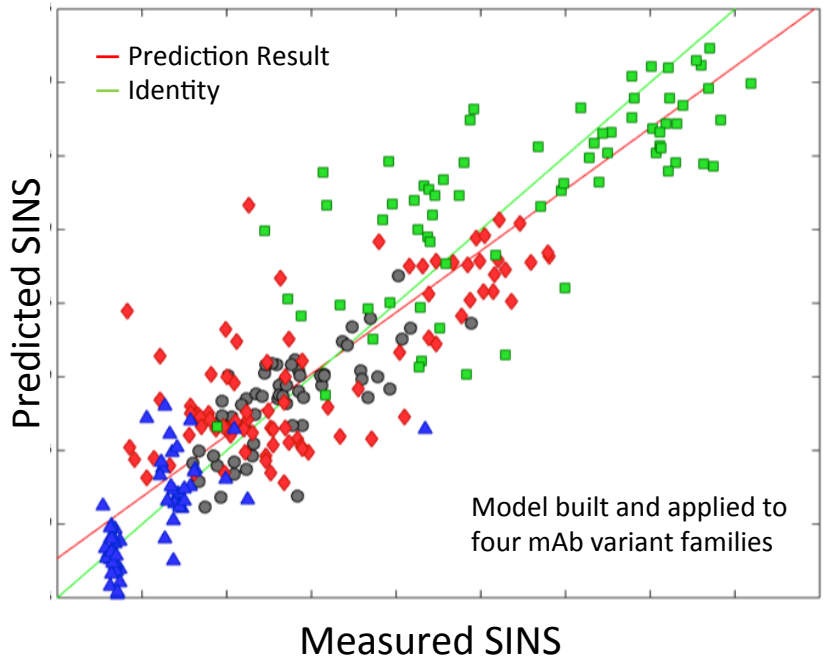
- Reduced cost and time
- Expanded experimental design
- Increased understanding of molecular properties

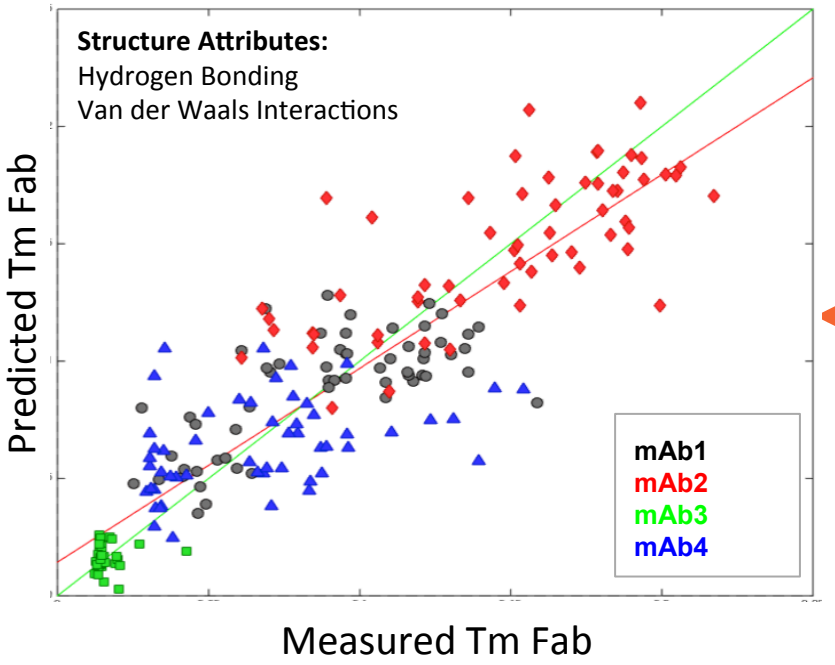
Just.

Machine Learning being applied to predict molecular properties across multiple antibodies and antibody variants

Targeted Properties:

- Stable Titer
- Molecular Stability (Tm, pH, Chemical Stability)
- Aggregation
- High Molecular Weight
- Molecular Interactions
- Viscosity





Summary

- Expertise → Platform → Data → Predictive Models → **Speed, Capacity, Cost**
- Capture all virtual entities and associated data
- Democratize all data
- Develop data analysis systems to be Objective, Consistent, Reliable, Informative
- Automate data capture and analysis
- **Build machine learned predictive models to select and engineer antibodies for optimal characteristics**

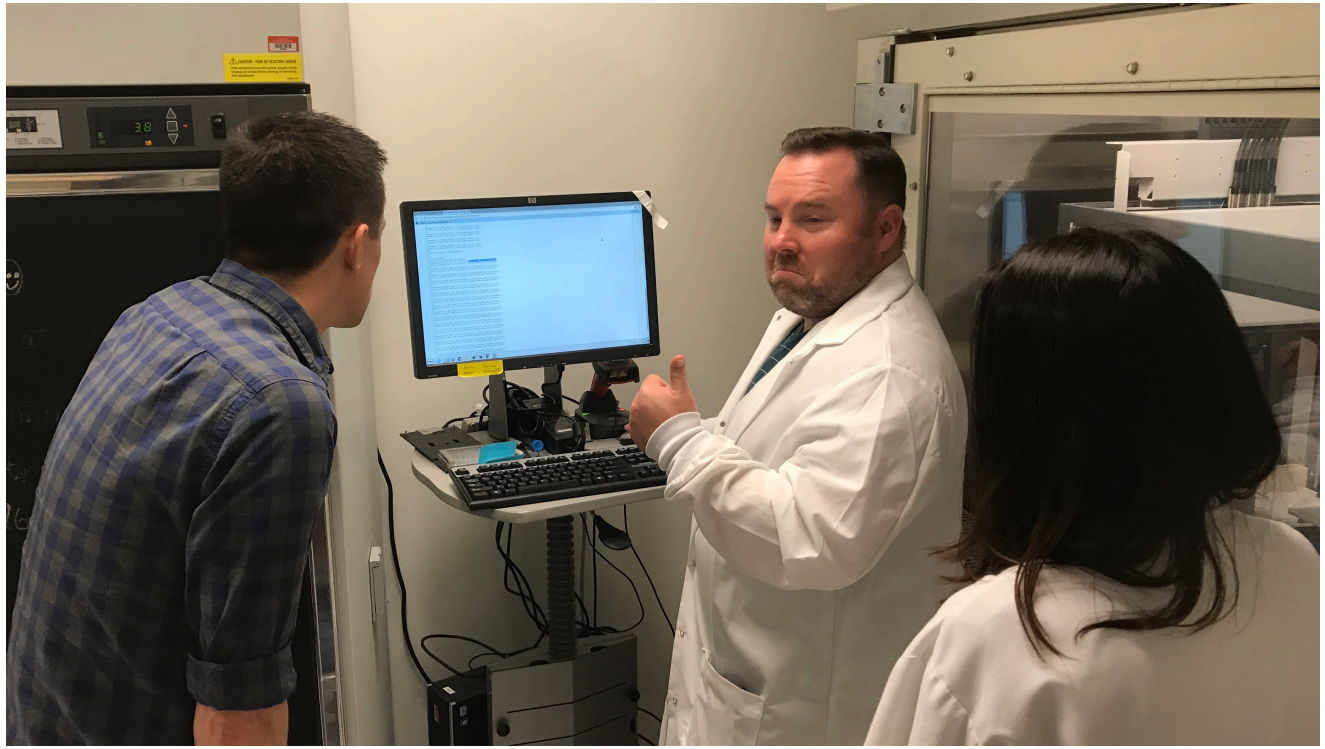
Aknowledgements

Just Biotherapeutics

Tileli Amimeur
Rutilio Clark
Alaina Floyd
Alison Gillespie
Jeff James
Bruce Kerwin
Randal Ketchem
Alison Leonard
Jeremy Shaver
Christine Siska
Pauline Smidt
Russell Williams

LabKey

Kevin Krouse
Ryan Luce



just

adjective

based on or behaving according to what is morally right and fair.
"a just and democratic society"

synonyms: fair, fair-minded, equitable, unbiased, unprejudiced, open-minded, nonpartisan

Just.