Genedata

Lunch & Learn Seminar

Streamlining Agile MS Workflows with Genedata

Aaron J. Cazy
Senior Scientist
Eli Lilly & Company, Bloomington, IN, USA

Lilly

CASSS
SHARING SCIENCE SOLUTIONS

danaher

Aaron J. Cazy

Senior Scientist

Biotechnology Discovery Research

Eli Lilly & Company, Bloomington, IN, USA

Midwestern man with a passion for science and technology. Aaron has been doing mass spectrometry for a little over a decade, with a focus on semi-automatic data analysis. Self-taught Python programmer who has used Knime, Raspberry Pi, and Genedata to automate workstream processes at QuarryBio, Labcorp, Merck & Lilly.
Outside of work, I stream as a competitive gamer as well as DJ/produce music.

A MEDICINE COMPANY

STREAMLINING AGILE MS WORKFLOWS WITH GENEDATA

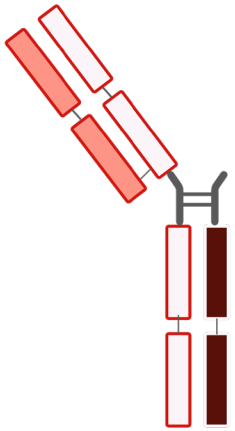
09.24.25

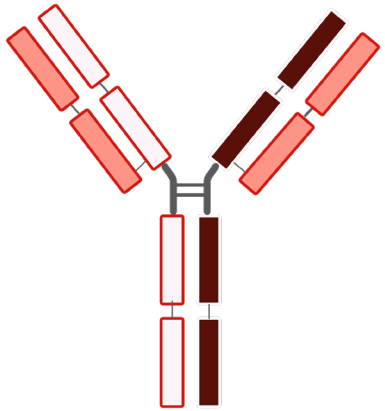
AARON CAZY

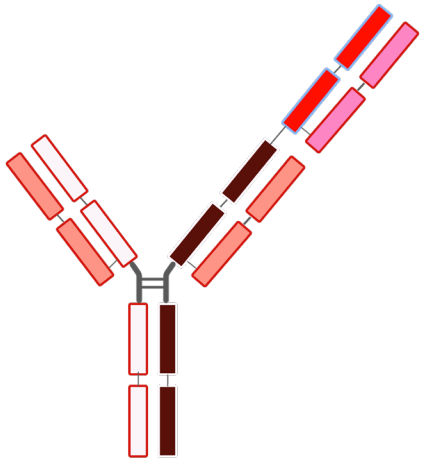
CASSS MS 25

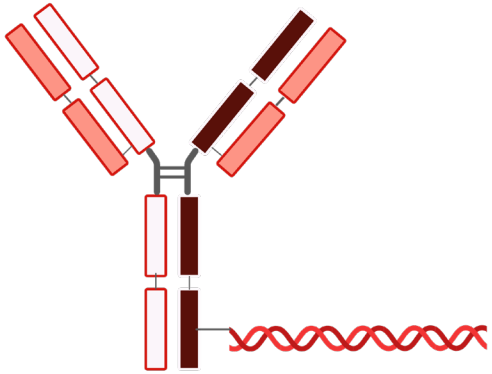
Lilly

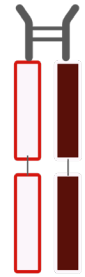
Complex Pipeline

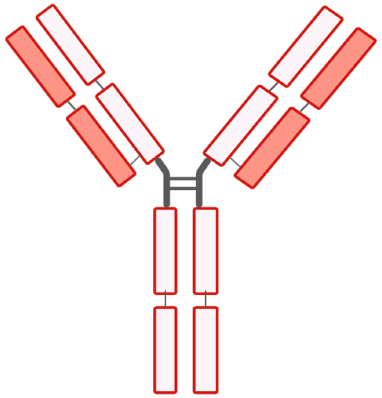


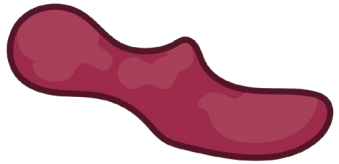


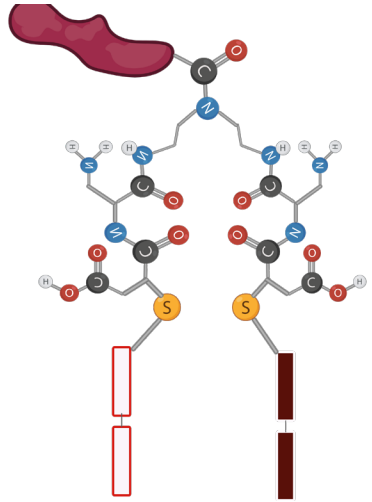












In This Presentation

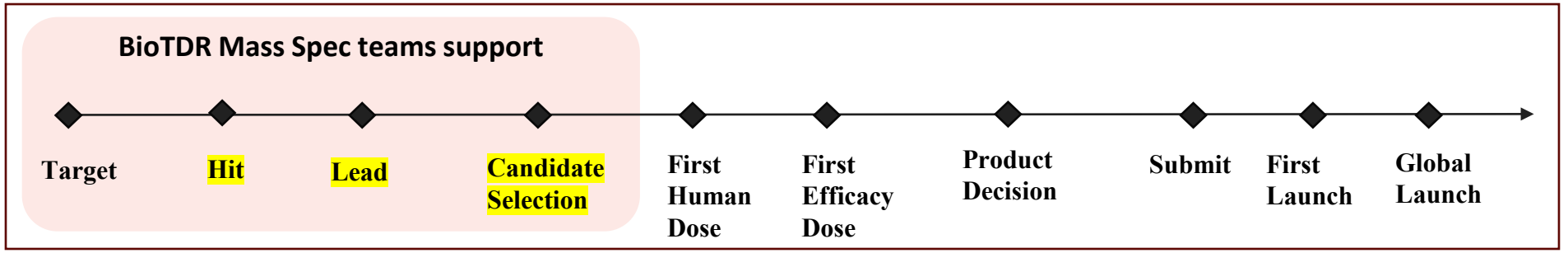
- Part 1

BioTDR mass spec team’s data digital transformation strategy and how we maximize our values and impacts of this effort

- Part 2

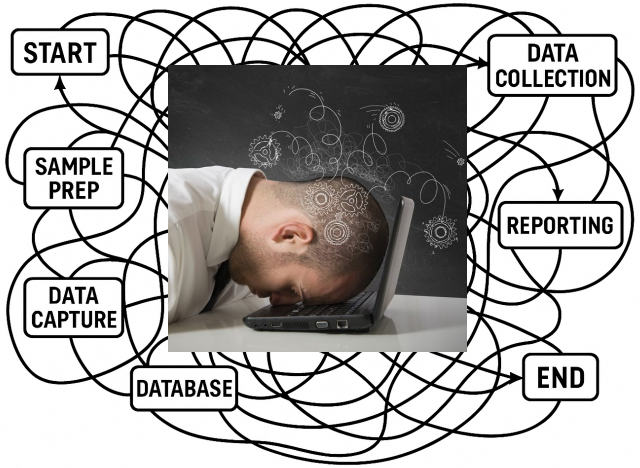
Assay specific case studies

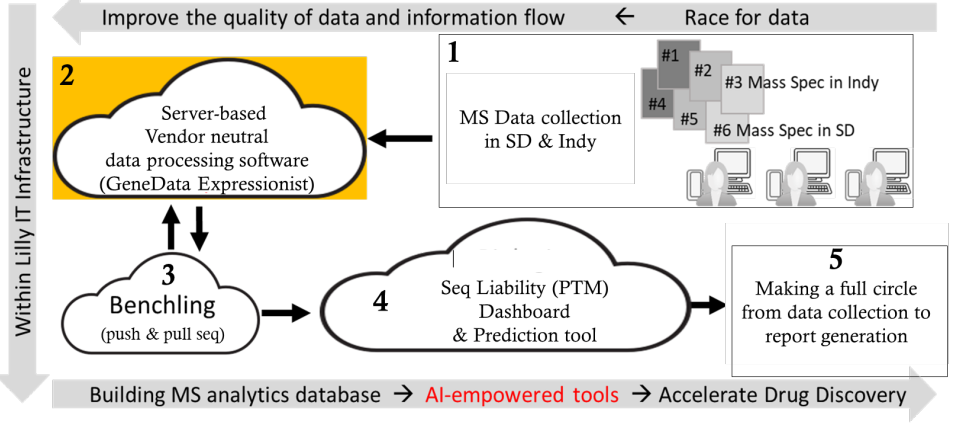
Our Goal in BioTDR is Accelerating Drug Discovery



We handle complex MS tasks ➡ What is our roadblock? ➡ Create a strategy roadmap

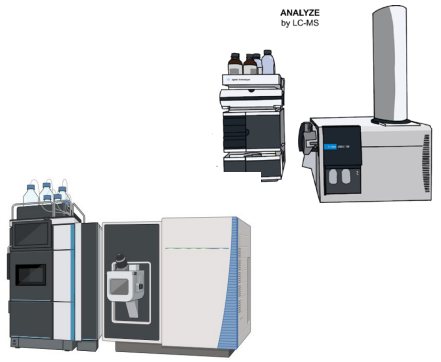
- Various types of discovery molecules
- High throughput platform methods for screening purpose
- At same time, low throughput methods for deeper characterization toward candidate selection (CS)
- Developability and stability assessment and impurity characterization shifted to earlier stage discovery to deliver CS faster





Joomi Ahn

What We Build and How We Connect



Benchling

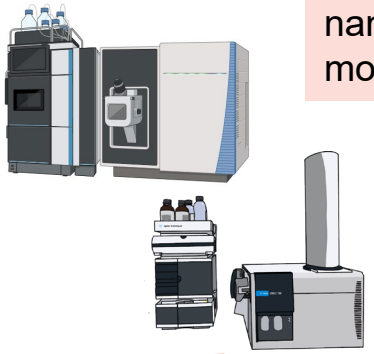
Genedata

Standardized &
harmonized

What We Build and How We Connect

6. **Raw file transfer tools** developed using file watcher auto-push to servers and/or Jarvis Phase1

1. **Structured data** established btw two sites (e.g. file name, method name, folder structure, metadata, molecule registration)



Benchling

2. **Customized Benchling plug-Ins** to automatically generate json sequence files with CDR and chain information
3. Tool developed to automatically generate metadata files containing json names, raw file names, and file path in structured Genedata Expressionist (GDE) server folders

Genedata

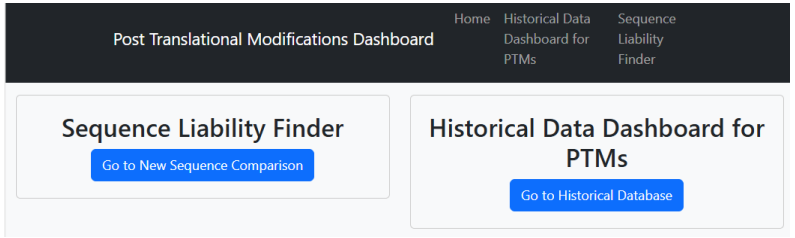
4. **Dev, test, production** servers established and validated

Plug-in Customization

5. **Report customization** is very powerful advancement. Report Plug-Ins with flexibility is key.

Build new MS tools & database & AI/ML

7. **New MS tools in Biologica**
- Historical Stability posttranslational modification liability database built
- Sequence liability finder implemented
- Deamidation prediction model developed (AI/ML)

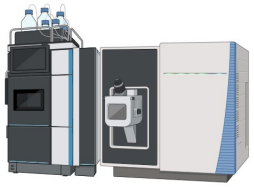


8. **Automated report generation tool** development via Research Data & SciQuery (stability report)

Joomi Ahn

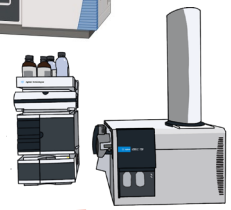
What Improvement We Achieved in GDE Full Circle

5. Raw files automatically pushed to servers and ready for data processing



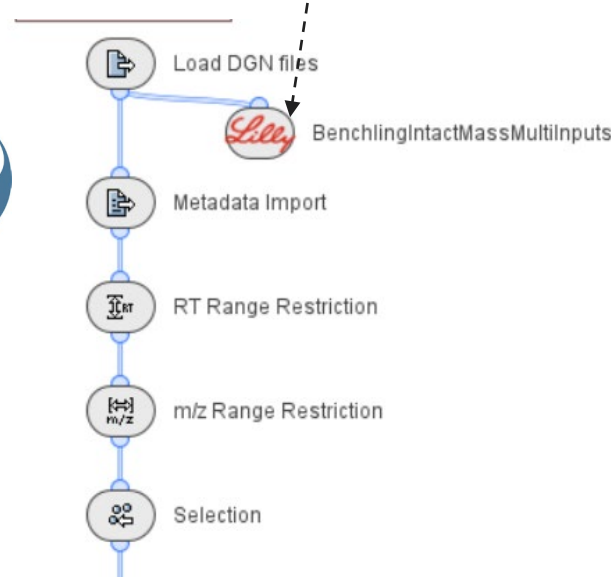
1. Reduced data error caused from minimal copy/paste of files

4. Lab automation and sample prep method simplification using a liquid handler



Benchling

2. Benchling and GDE are fully connected to push & pull metadata. We built Benchling custom activities as part of GDE workflows



Build new MS tools & database & AI/ML

Genedata

Plug-in Customization

3. Built foundation of transformed data:

- A living database of CDR sequence liabilities with post-translation modification
- Using this database, deamidation prediction model were developed (see page 10)
- Searchable stability knowledge is available with projects, molecule ID, and sequence motif, liabilities, etc

Joomi Ahn

Genedata Case Studies

Bispecific Mispairing

Improving the speed of data review

ARC Analysis

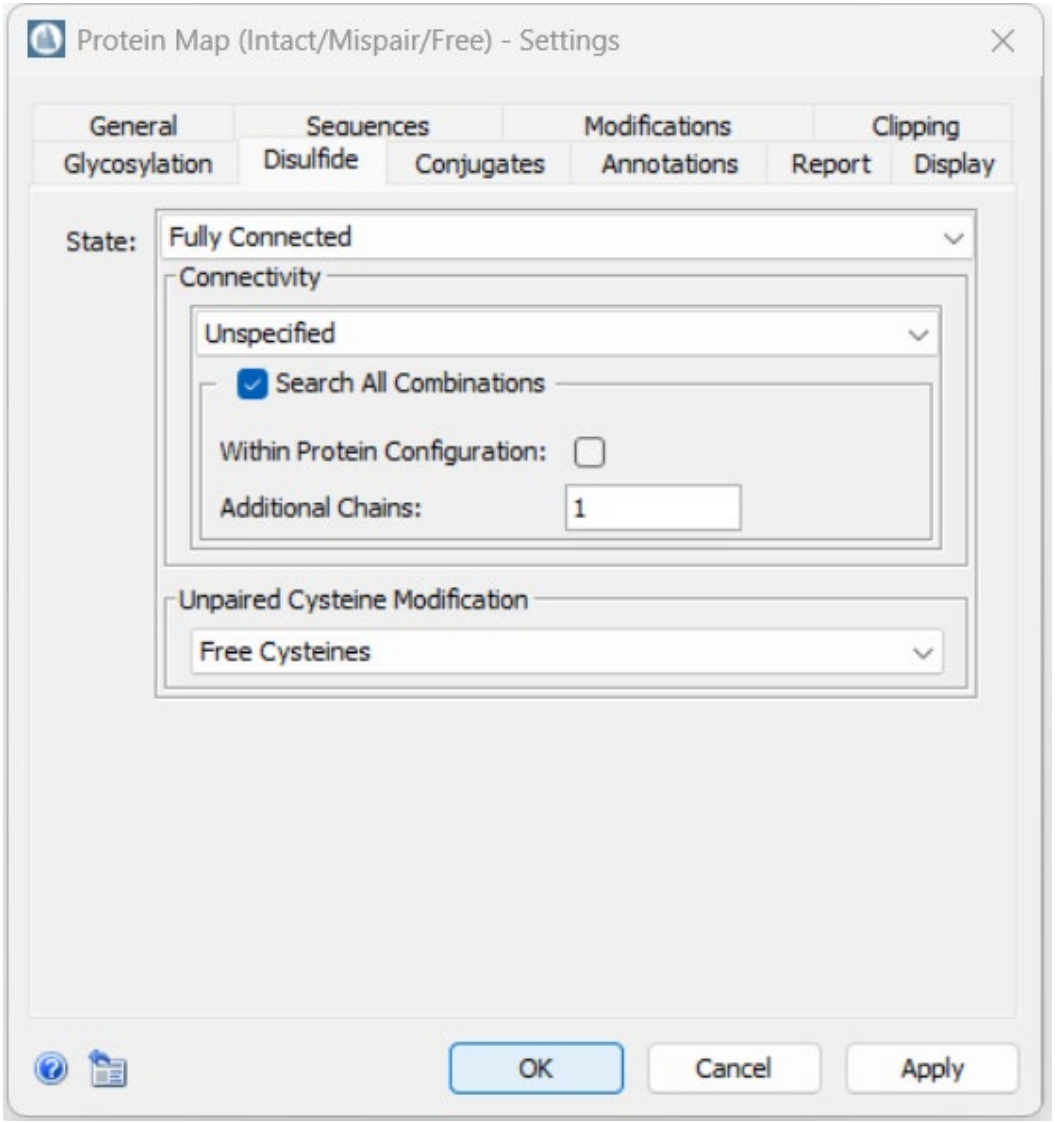
Integrating with Benchling to grab siRNA information and translating it into a Genedata input

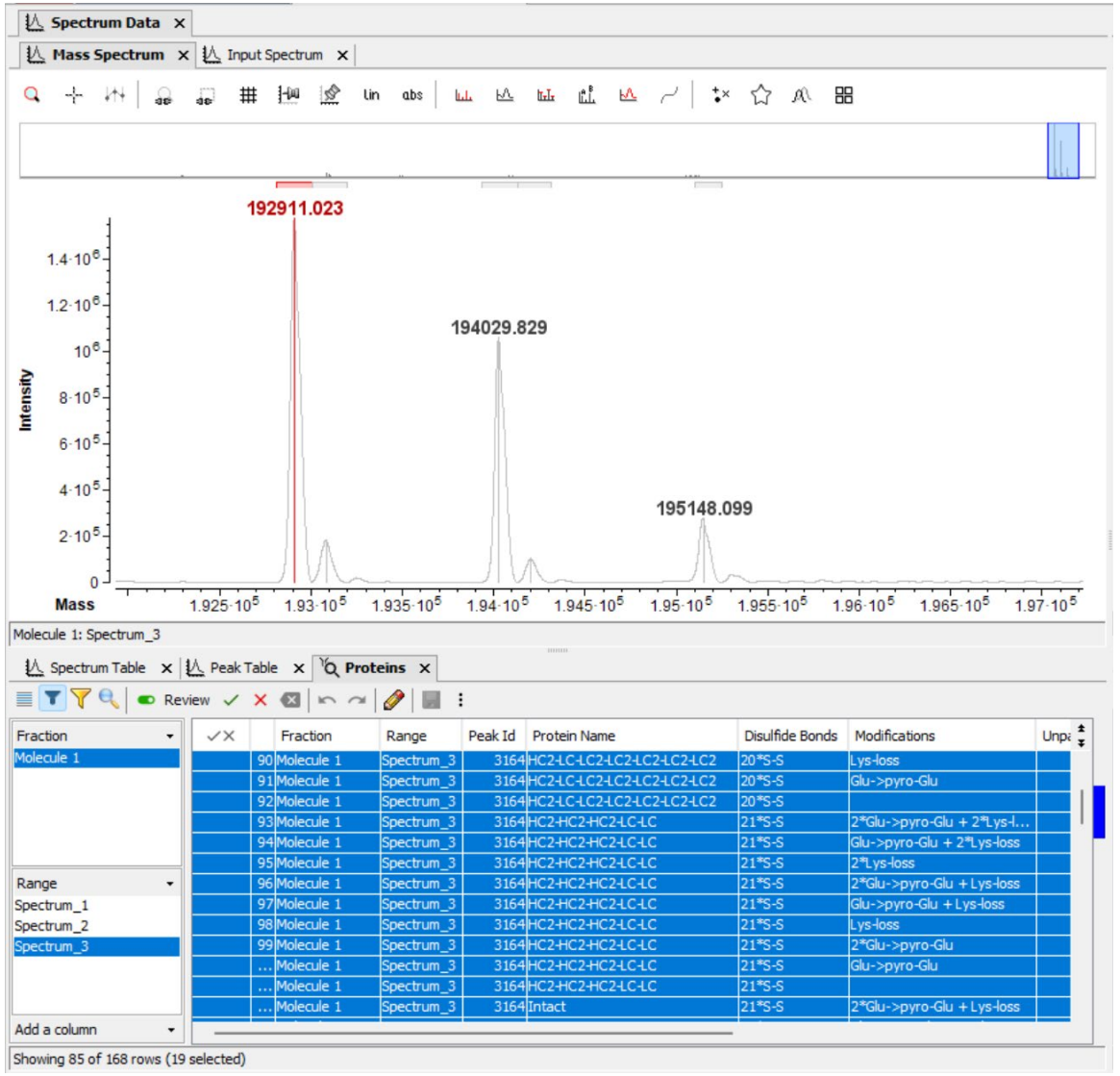
Peptide Mapping

Streamlining the workflow and accelerating the review

Case 1: Bispecific Mispairing – Quick Mispairing ID

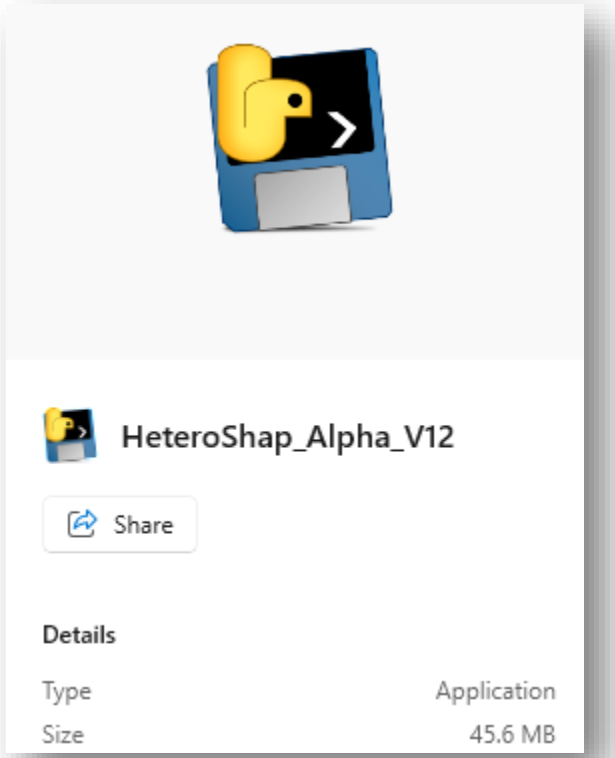
Untargeted mispairing algorithm requires manual review

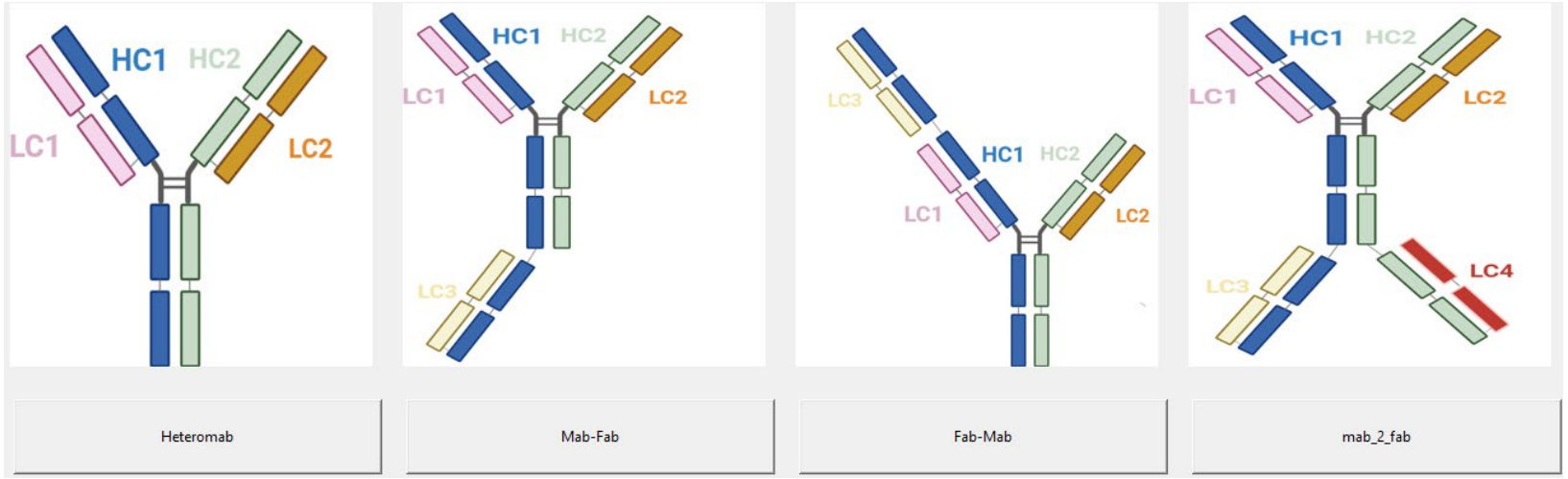


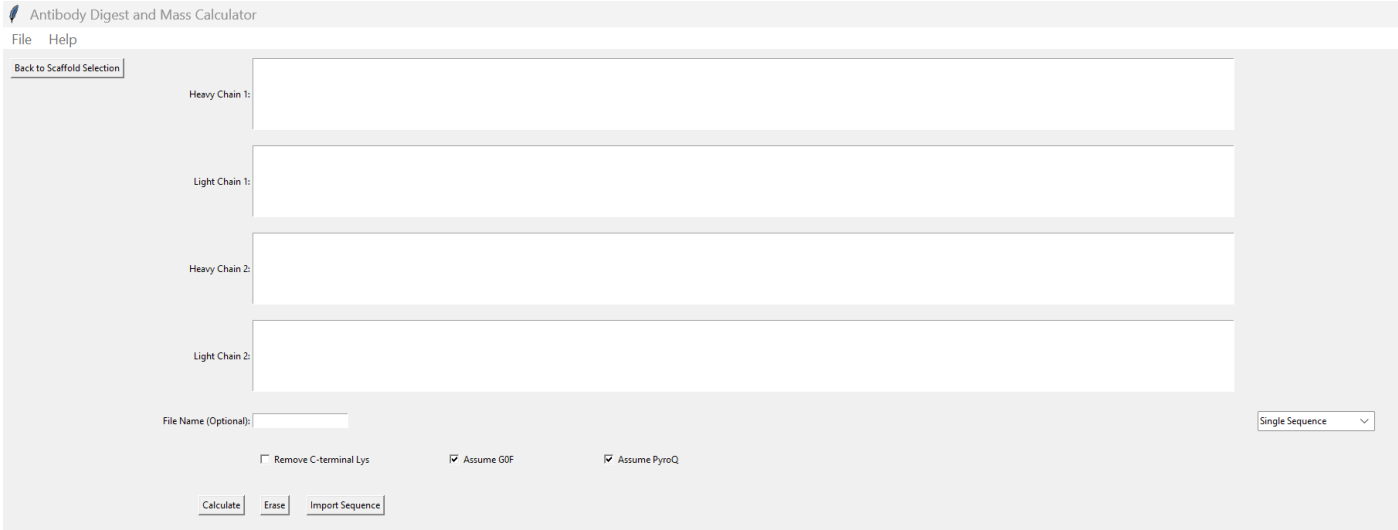


Case 1: Bispecific Mispairing - Custom Calculation Script

Standalone python application creates mass list for targeted annotation







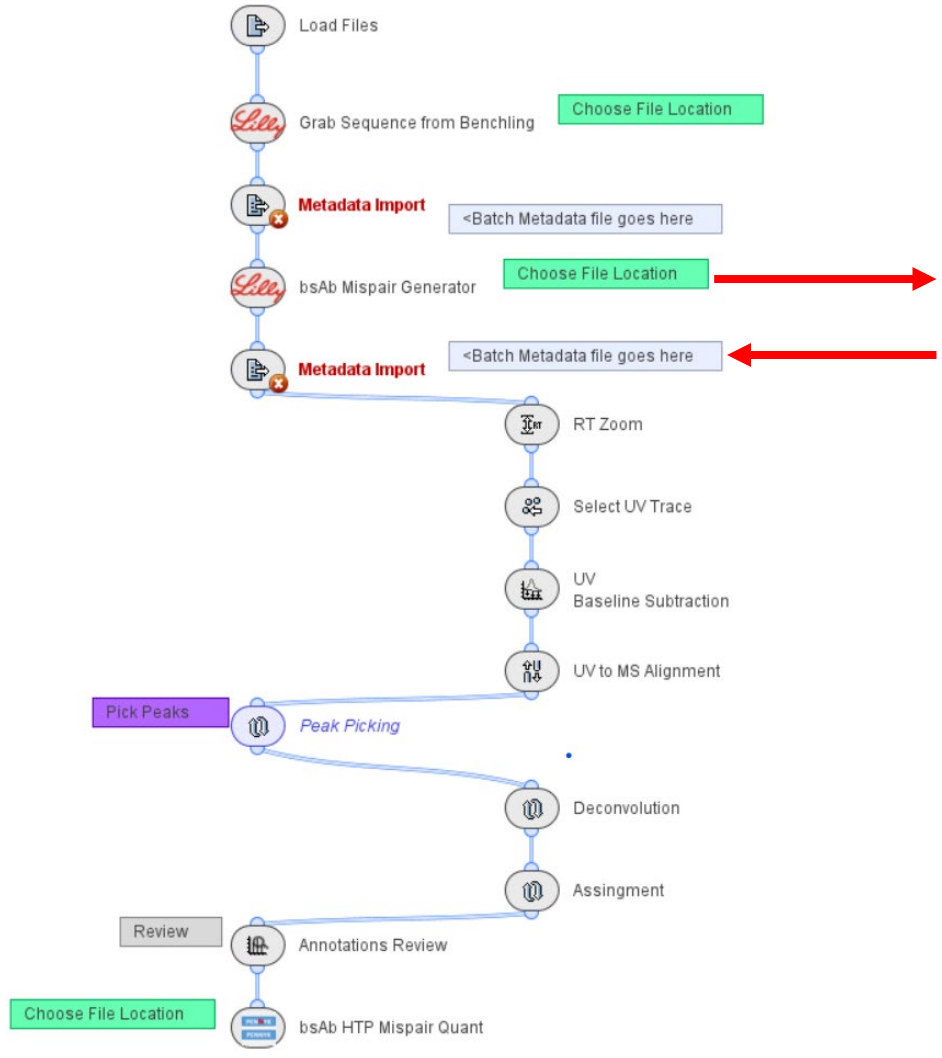
Standalone python calculator

Single/batch JSON import and process

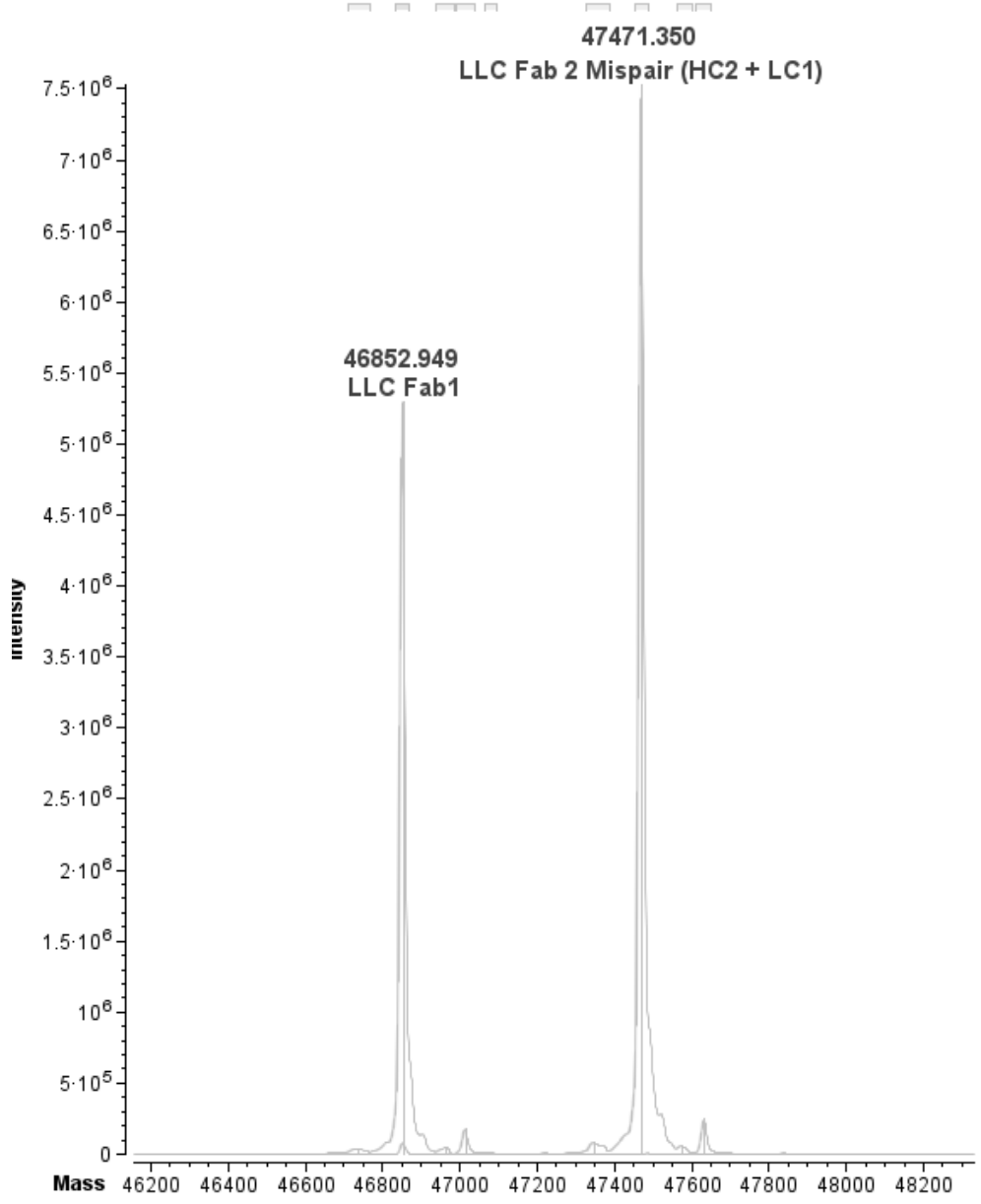
Metadata file creation (custom for Genedata input)

Case 1: Bispecific Mispairing – Integration

Final version will remove need to swap software and fully integrate the python calculator



	A	B
1	Identifier	Avg. Mass
2	Correct Pair	146623.4417
3	Mis Pair1	146141.1183
4	Mis Pair2	146690.6449
5	Mis Pair3	146556.2385
6	Mis Pair4	147105.7651
7	Mis Pair5	146208.3215
8	Mis Pair6	146073.9152
9	Mis Pair7	146623.4417
10	Mis Pair8	147172.9683
11	Mis Pair9	147038.5619
12	LLC Fab1	47054.86495
13	LLC Fab 2	46358.39435
14	LLC FC	53246.21305
15	LLC Fab 1 Mispair (HC1 + LC2)	46572.54155
16	LLC Fab 2 Mispair (HC2 + LC1)	46358.39435
17	LLC FC Mispair 1 (HC1 + HC1)	52964.86265
18	LLC FC Mispair 2 (HC2 + HC2)	53527.56345
19	Elas Fab1	47155.96885
20	Elas Fab 2	46459.49825
21	Elas FC	53044.00525
22	Elas Fab 1 Mispair (HC1 + LC2)	46673.64545
23	Elas Fab 2 Mispair (HC2 + LC1)	46459.49825
24	Elas FC Mispair 1 (HC1 + HC1)	52762.65485
25	Elas FC Mispair 2 (HC2 + HC2)	53325.35565
26	Ides Fab'2	146623.4417
27	IdeS Fab'2 Mispair 1 (HC1 + HC1)	146556.2385
28	IdeS Fab'2 Mispair 2 (HC2 + HC2)	146208.3215

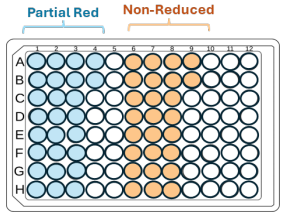


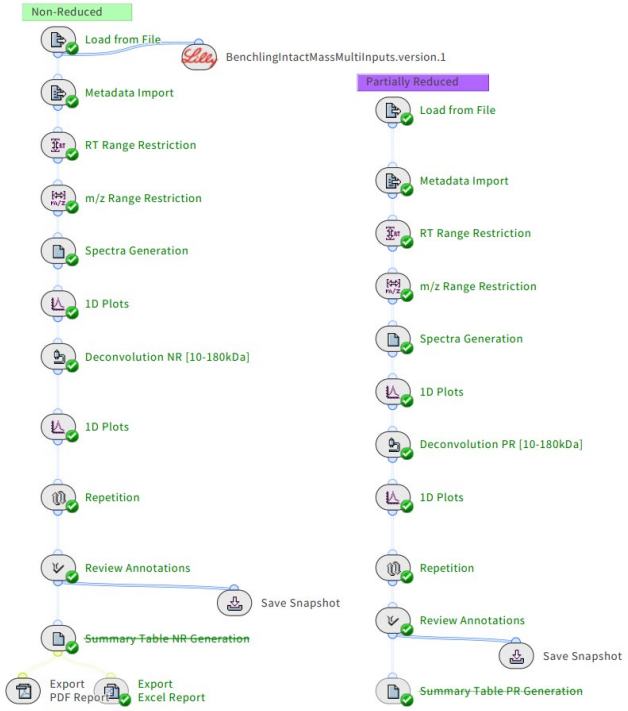
Case 1: Bispecific Mispairing – High-Throughput

Results reduced from weeks of analysis time to less than one workday

HT Standard MW ID

Batch size: 26 mAb
Goal: MW/ID confirmation
Non-reduced &
Partially reduced

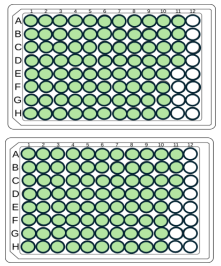


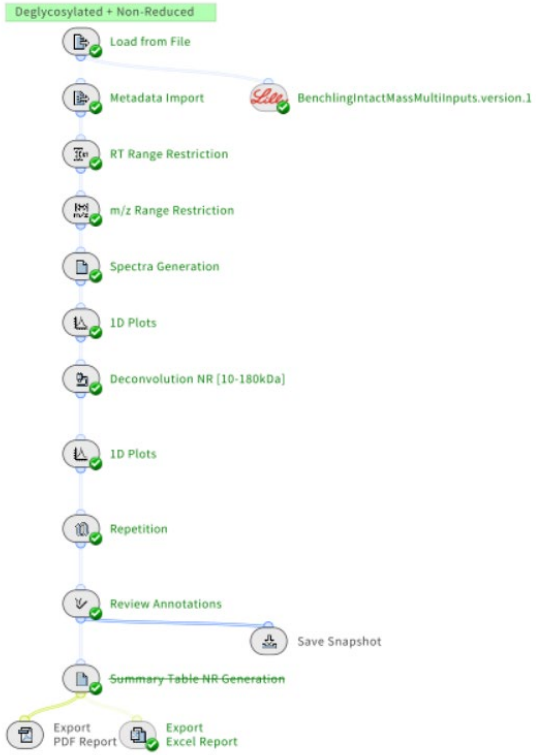


RESULT: 26/26 MATCHING
~ 1 day turnaround

HT BsAb MW ID

Batch size: 168 samples
Goal: Confirm AA sequence by MW
Condition: Deglycosylated NR
Hamilton + PNGase F Automation

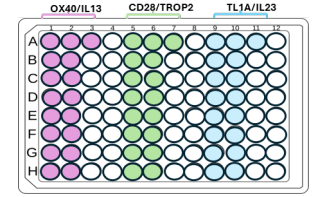




RESULT: 168/168 MATCHING
~ 1 day turnaround

HT BsAb Assembly Check + LC Mispair

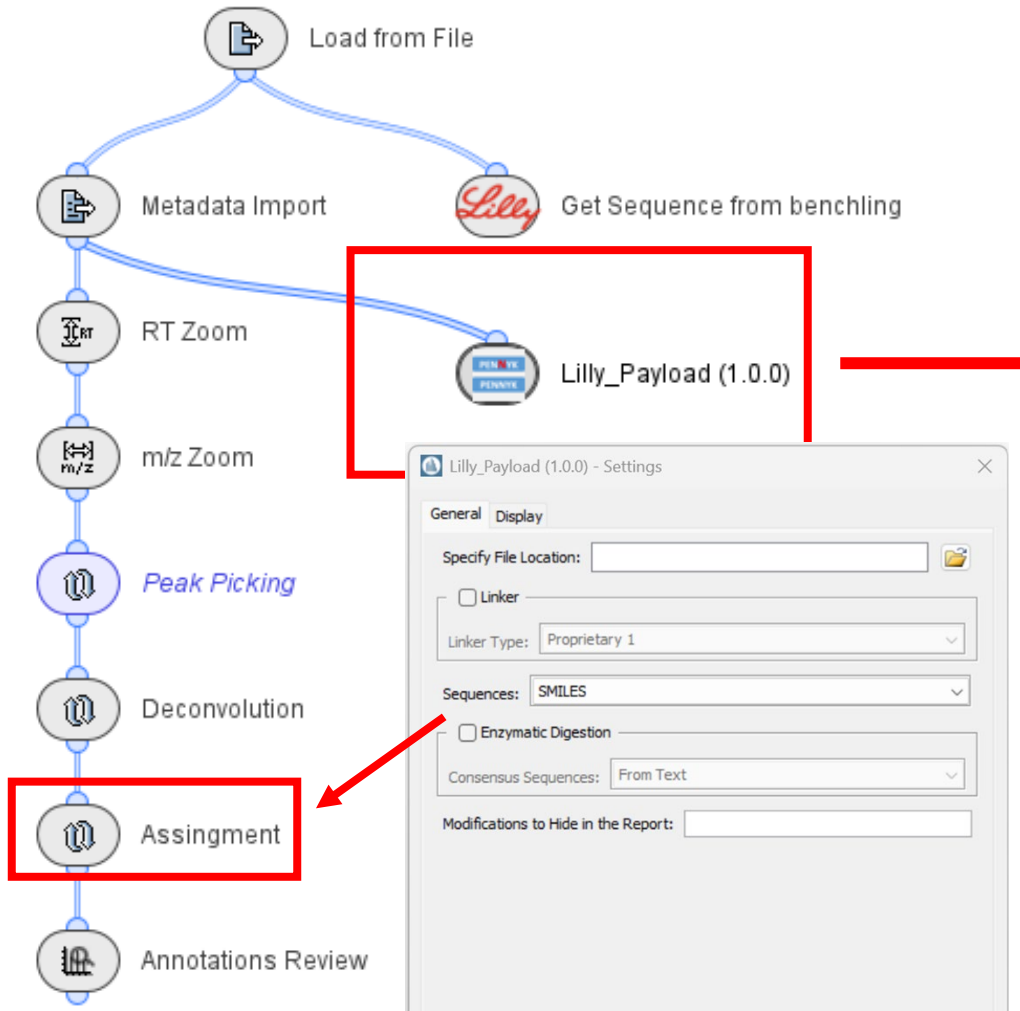
Batch size: 51 BsAbs (3 different projects with 17 different designs)
Goal: ID and LC mispairing
Conditions: Deglycosylated;
GingisKHAN (above-hinge digest)

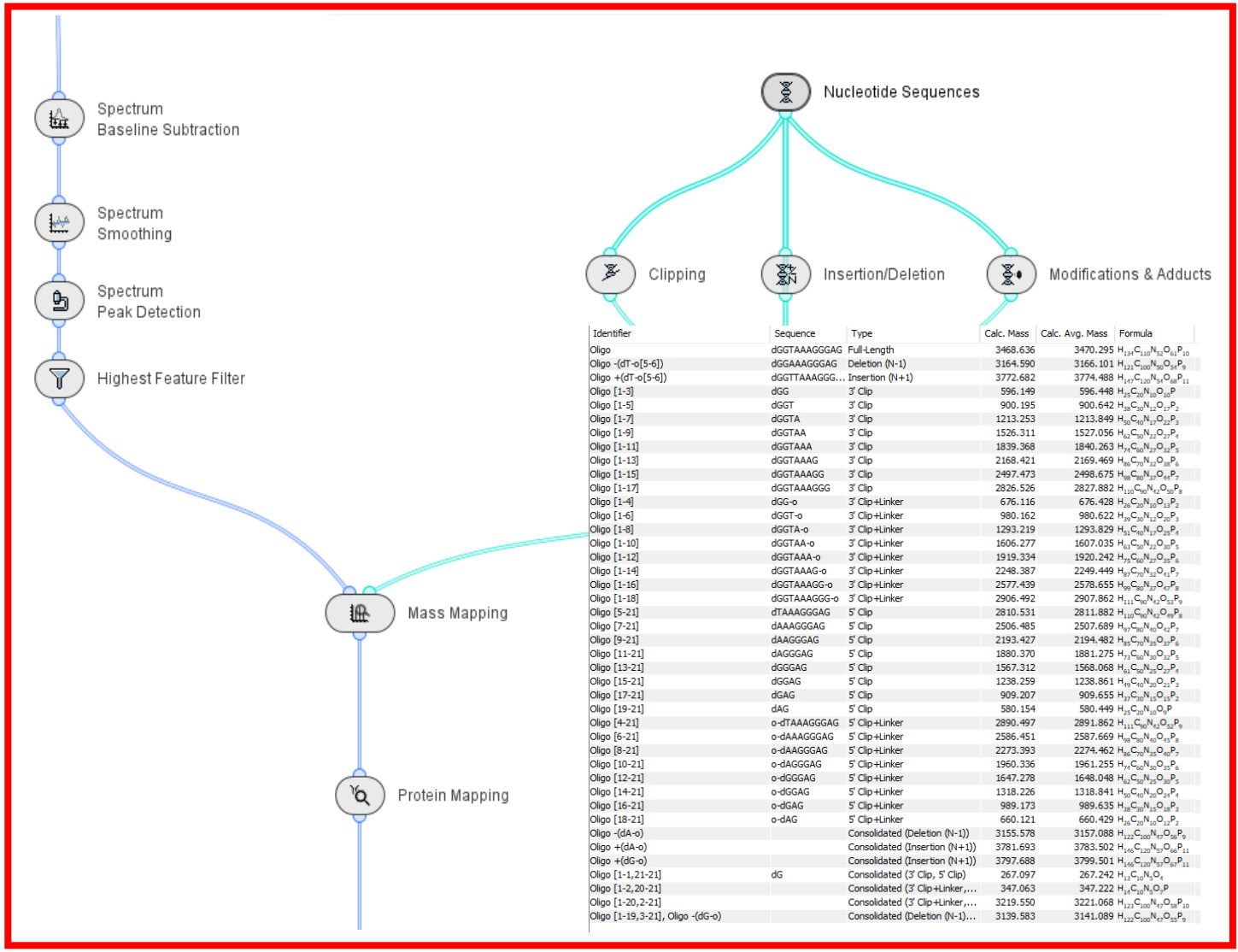


Result: 51/51 MATCHING
49/51 samples contain varying levels of mis-assembly.
~ 1 day turnaround

Case 2: ARC Analysis

Genedata can be connected directly to LIMS system for direct metadata transfer

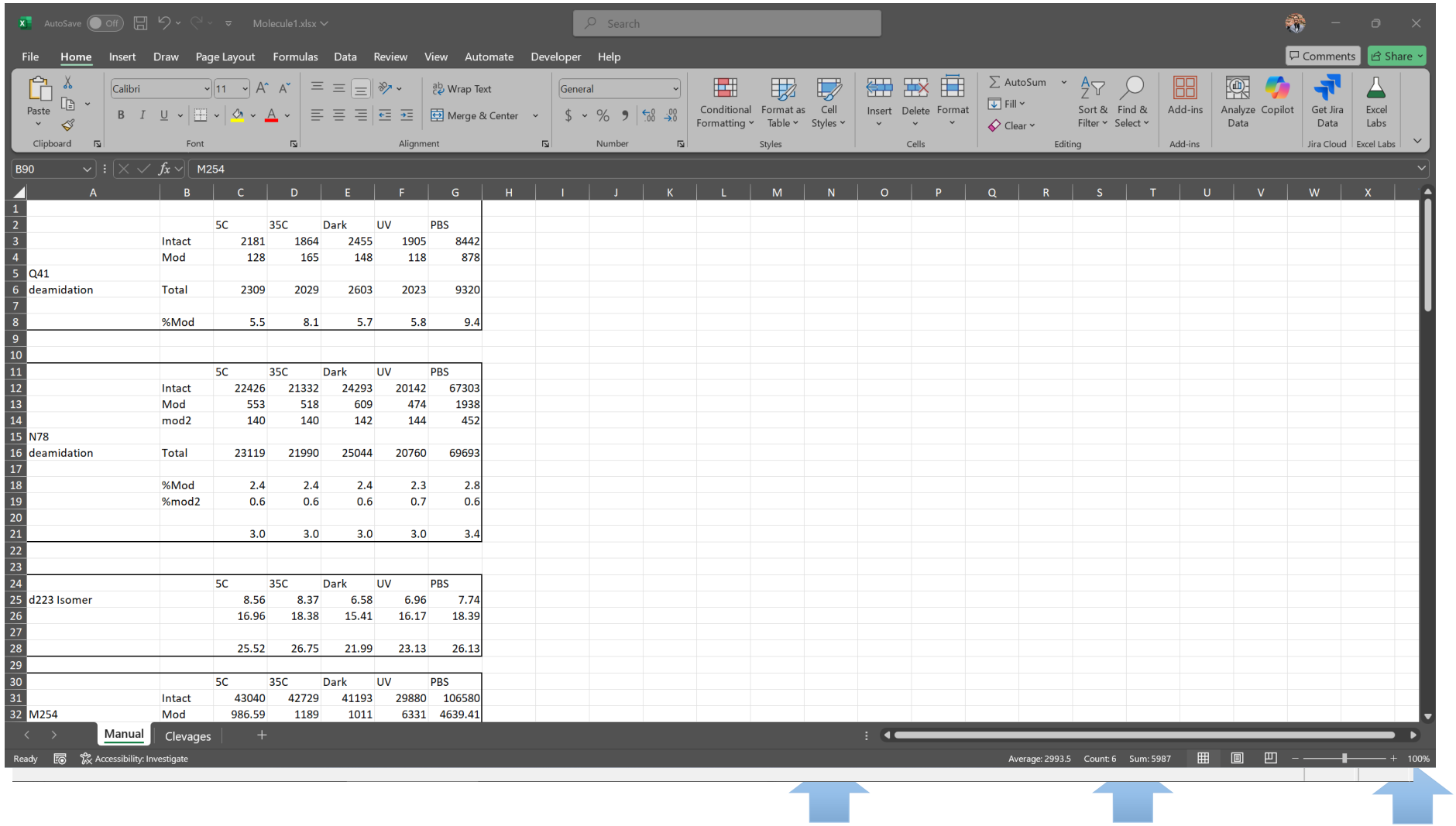




Identifier	Sequence	Type	Calc. Mass	Calc. Avg. Mass	Formula
Oligo	dGGTAAAGGGAG	Full-Length	3468.636	3470.295	$H_{114}C_{110}N_{32}O_{61}P_{10}$
Oligo -(dT-o[5-6])	dGGAAAGGGAG	Deletion (N-1)	3164.590	3166.101	$H_{121}C_{100}N_{36}O_{54}P_9$
Oligo +(dT-o[5-6])	dGGTTAAAGGG...	Insertion (N+1)	3772.682	3774.488	$H_{147}C_{120}N_{34}O_{68}P_{11}$
Oligo [1-3]	dGG	3' Clip	596.149	596.448	$H_{25}C_{20}N_{10}O_{10}P$
Oligo [1-5]	dGGT	3' Clip	900.195	900.642	$H_{38}C_{30}N_{12}O_{17}P_2$
Oligo [1-7]	dGGTA	3' Clip	1213.253	1213.849	$H_{50}C_{40}N_{17}O_{23}P_3$
Oligo [1-9]	dGGTAA	3' Clip	1526.311	1527.056	$H_{62}C_{50}N_{22}O_{29}P_4$
Oligo [1-11]	dGGTAAA	3' Clip	1839.368	1840.263	$H_{74}C_{60}N_{27}O_{35}P_5$
Oligo [1-13]	dGGTAAAG	3' Clip	2168.421	2169.469	$H_{86}C_{70}N_{32}O_{38}P_6$
Oligo [1-15]	dGGTAAAGG	3' Clip	2497.473	2498.675	$H_{98}C_{80}N_{37}O_{44}P_7$
Oligo [1-17]	dGGTAAAGGG	3' Clip	2826.526	2827.882	$H_{110}C_{90}N_{42}O_{50}P_8$
Oligo [1-4]	dGG-o	3' Clip+Linker	676.116	676.428	$H_{26}C_{20}N_{10}O_{11}P_2$
Oligo [1-6]	dGGT-o	3' Clip+Linker	980.162	980.622	$H_{39}C_{30}N_{12}O_{18}P_3$
Oligo [1-8]	dGGTA-o	3' Clip+Linker	1293.219	1293.829	$H_{51}C_{40}N_{17}O_{24}P_4$
Oligo [1-10]	dGGTAA-o	3' Clip+Linker	1606.277	1607.035	$H_{63}C_{50}N_{22}O_{30}P_5$
Oligo [1-12]	dGGTAAA-o	3' Clip+Linker	1919.334	1920.242	$H_{75}C_{60}N_{27}O_{36}P_6$
Oligo [1-14]	dGGTAAAG-o	3' Clip+Linker	2248.387	2249.449	$H_{87}C_{70}N_{32}O_{41}P_7$
Oligo [1-16]	dGGTAAAGG-o	3' Clip+Linker	2577.439	2578.655	$H_{99}C_{80}N_{37}O_{47}P_8$
Oligo [1-18]	dGGTAAAGGG-o	3' Clip+Linker	2906.492	2907.862	$H_{111}C_{90}N_{42}O_{57}P_9$
Oligo [5-21]	dTAAAGGGAG	5' Clip	2810.531	2811.882	$H_{110}C_{90}N_{32}O_{60}P_8$
Oligo [7-21]	dAAAGGGAG	5' Clip	2506.485	2507.689	$H_{97}C_{80}N_{42}O_{51}P_7$
Oligo [9-21]	dAAGGGAG	5' Clip	2193.427	2194.482	$H_{85}C_{70}N_{37}O_{42}P_6$
Oligo [11-21]	dAGGGAG	5' Clip	1880.370	1881.275	$H_{73}C_{60}N_{32}O_{33}P_5$
Oligo [13-21]	dGGGAG	5' Clip	1567.312	1568.068	$H_{61}C_{50}N_{22}O_{24}P_4$
Oligo [15-21]	dGGAG	5' Clip	1238.259	1238.861	$H_{49}C_{40}N_{20}O_{15}P_3$
Oligo [17-21]	dGAG	5' Clip	909.207	909.655	$H_{37}C_{30}N_{15}O_{6}P_2$
Oligo [19-21]	dAG	5' Clip	580.154	580.449	$H_{25}C_{20}N_{10}O_6P$
Oligo [4-21]	o-dTAAAGGGAG	5' Clip+Linker	2890.497	2891.862	$H_{111}C_{90}N_{34}O_{62}P_9$
Oligo [6-21]	o-dAAAGGGAG	5' Clip+Linker	2586.451	2587.669	$H_{98}C_{80}N_{45}O_{53}P_8$
Oligo [8-21]	o-dAAGGGAG	5' Clip+Linker	2273.393	2274.462	$H_{86}C_{70}N_{35}O_{44}P_7$
Oligo [10-21]	o-dAGGGAG	5' Clip+Linker	1960.336	1961.255	$H_{74}C_{60}N_{30}O_{35}P_6$
Oligo [12-21]	o-dGGGAG	5' Clip+Linker	1647.278	1648.048	$H_{62}C_{50}N_{25}O_{26}P_5$
Oligo [14-21]	o-dGGAG	5' Clip+Linker	1318.226	1318.841	$H_{50}C_{40}N_{23}O_{17}P_4$
Oligo [16-21]	o-dGAG	5' Clip+Linker	989.173	989.635	$H_{38}C_{30}N_{18}O_8P_3$
Oligo [18-21]	o-dAG	5' Clip+Linker	660.121	660.429	$H_{26}C_{20}N_{10}O_{11}P_2$
Oligo -(dA-o)		Consolidated (Deletion (N-1))	3155.578	3157.088	$H_{122}C_{100}N_{37}O_{56}P_9$
Oligo +(dA-o)		Consolidated (Insertion (N+1))	3781.693	3783.502	$H_{146}C_{120}N_{37}O_{66}P_{11}$
Oligo +(dG-o)		Consolidated (Insertion (N+1))	3797.688	3799.501	$H_{146}C_{120}N_{37}O_{67}P_{11}$
Oligo [1-1,21-21]	dG	Consolidated (3' Clip, 5' Clip)	267.097	267.242	$H_{13}C_{10}N_5O_4$
Oligo [1-2,20-21]		Consolidated (3' Clip+Linker,...	347.063	347.222	$H_{14}C_{10}N_5O_7P$
Oligo [1-20,2-21]		Consolidated (3' Clip+Linker,...	3219.550	3221.068	$H_{123}C_{100}N_{47}O_{58}P_{10}$
Oligo [1-19,3-21], Oligo -(dG-o)		Consolidated (Deletion (N-1)...	3139.583	3141.089	$H_{122}C_{100}N_{47}O_{55}P_9$

Case 3: Peptide Mapping

Traditional workflow requires manual candidate generation & verification

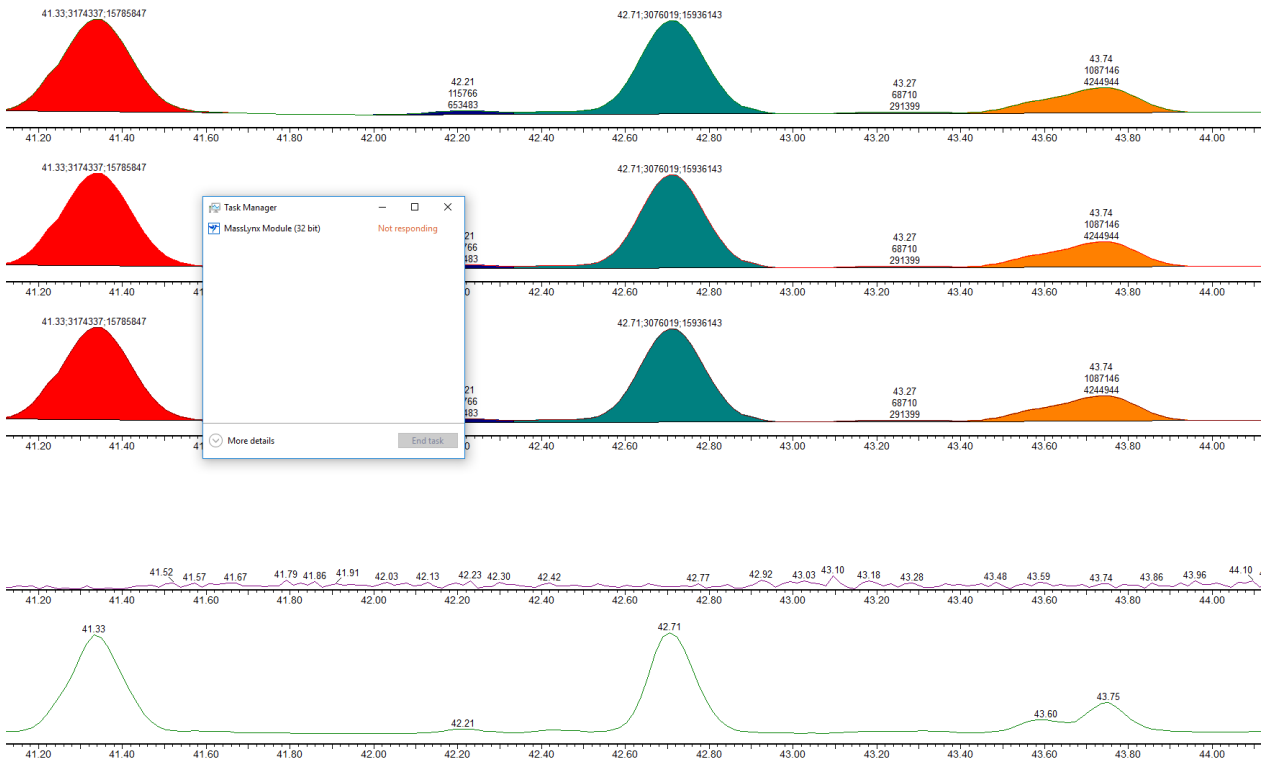


	A	B	C	D	E	F	G
1							
2			5C	35C	Dark	UV	PBS
3		Intact	2181	1864	2455	1905	8442
4		Mod	128	165	148	118	878
5	Q41						
6	deamidation	Total	2309	2029	2603	2023	9320
7							
8		%Mod	5.5	8.1	5.7	5.8	9.4
9							
10							
11			5C	35C	Dark	UV	PBS
12		Intact	22426	21332	24293	20142	67303
13		Mod	553	518	609	474	1938
14		mod2	140	140	142	144	452
15	N78						
16	deamidation	Total	23119	21990	25044	20760	69693
17							
18		%Mod	2.4	2.4	2.4	2.3	2.8
19		%mod2	0.6	0.6	0.6	0.7	0.6
20							
21			3.0	3.0	3.0	3.0	3.4
22							
23							
24			5C	35C	Dark	UV	PBS
25	d223 Isomer		8.56	8.37	6.58	6.96	7.74
26			16.96	18.38	15.41	16.17	18.39
27							
28			25.52	26.75	21.99	23.13	26.13
29							
30			5C	35C	Dark	UV	PBS
31		Intact	43040	42729	41193	29880	106580
32	M254	Mod	986.59	1189	1011	6331	4639.41

Case 3: Peptide Mapping - Liabilities Galore

Traditional workflows requires multiple programs and lots of copy/paste

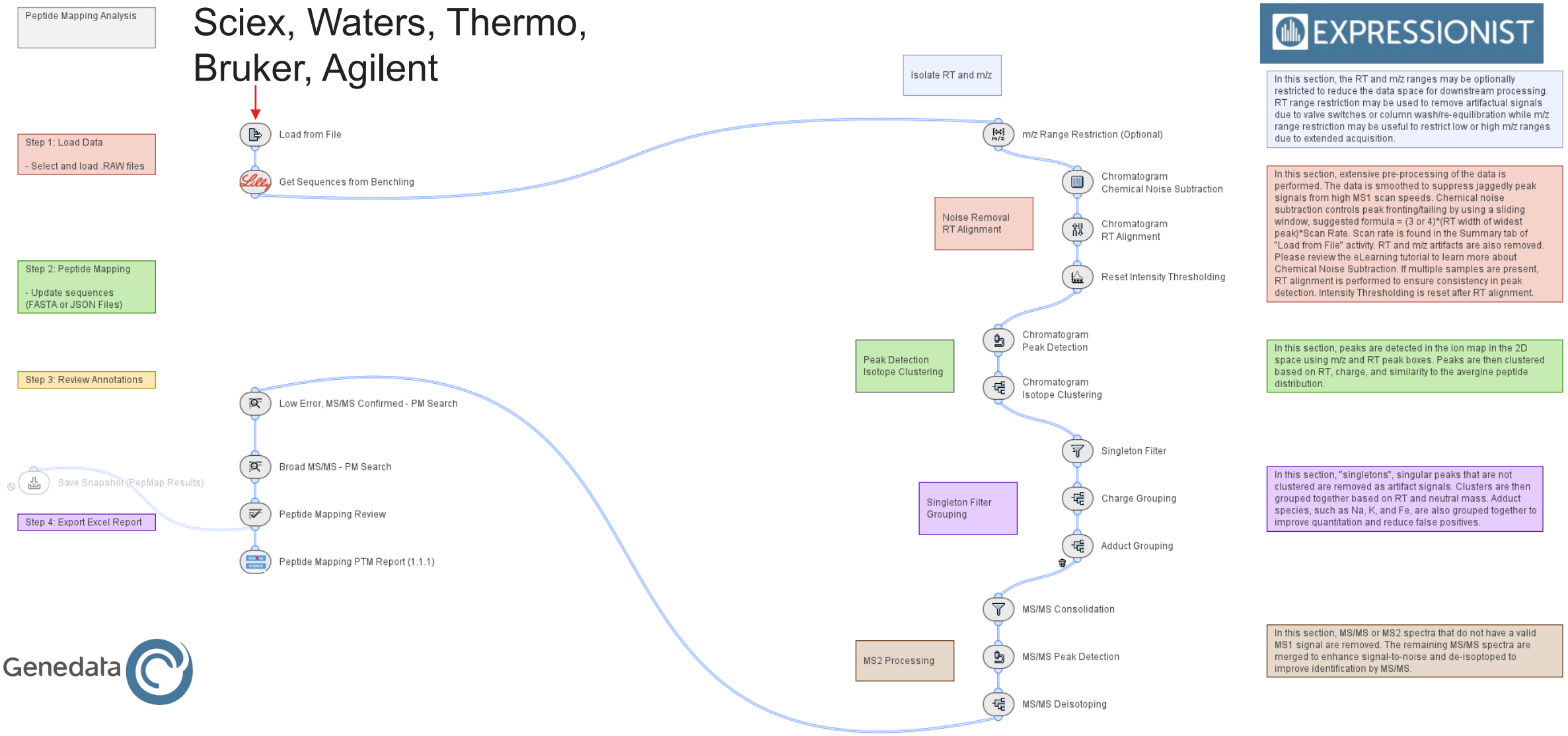
Waters → Biopharmalynx → Excel → Masslynx → Excel → PPT → in-house parser → benchling



1. Software swapping
2. Copy/paste
3. Spinning wheel of sadness
4. Alternate vendor
    1. Sciex
    2. Thermo
    3. Agilent
    4. Bruker
5. Training new hires

Case 3: Peptide Mapping - Genedata PM Workflow

All steps combined into one streamlined workflow regardless of instrument vendor

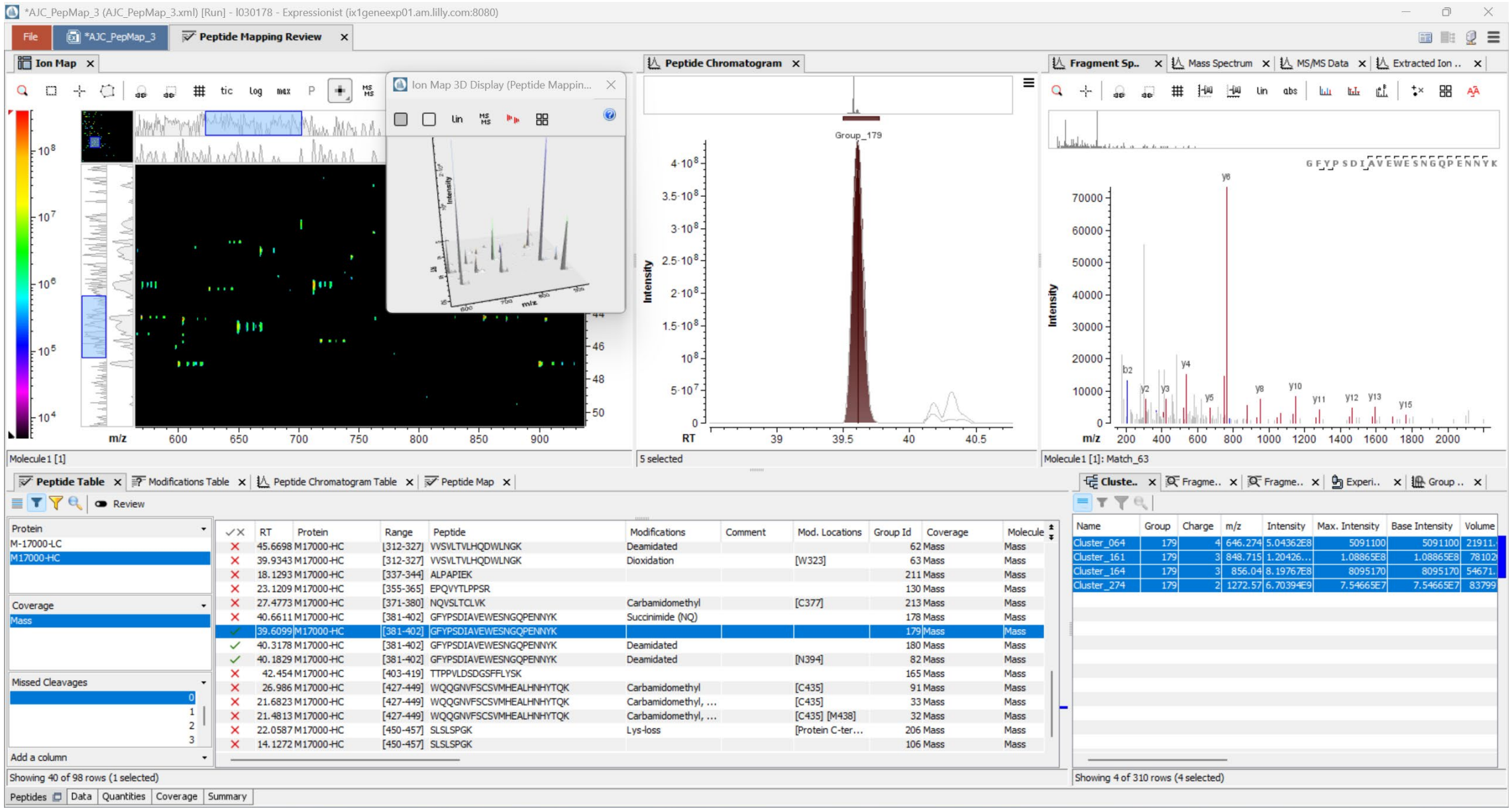


Genedata

Lilly

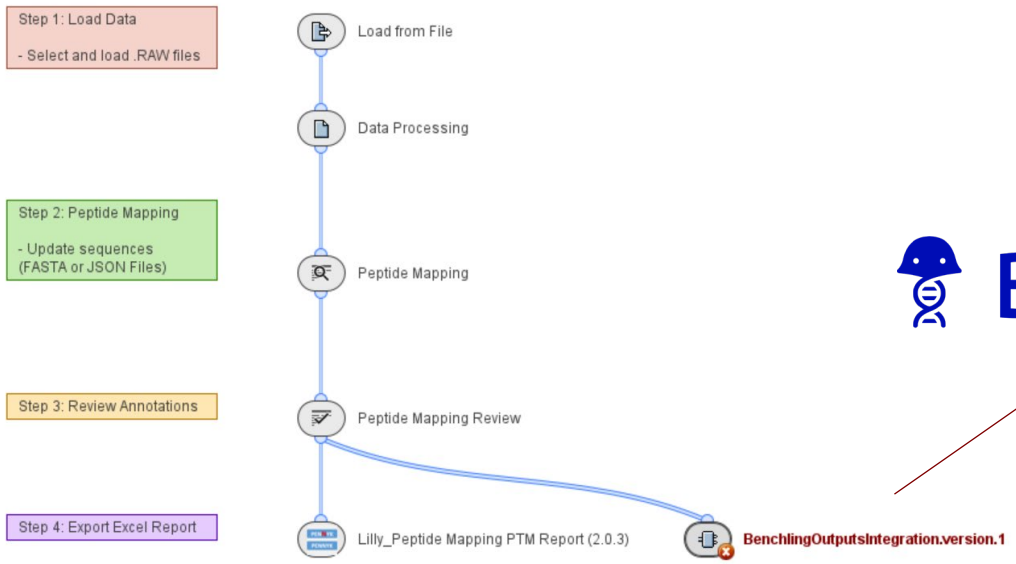
Case 3: Peptide Mapping - Fully Customizable UI

Every piece of information needed to review potential IDs can be displayed on one screen

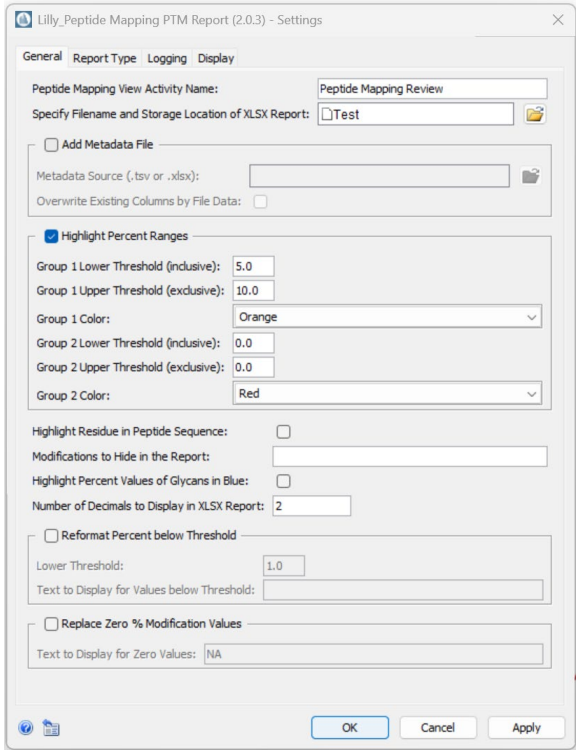


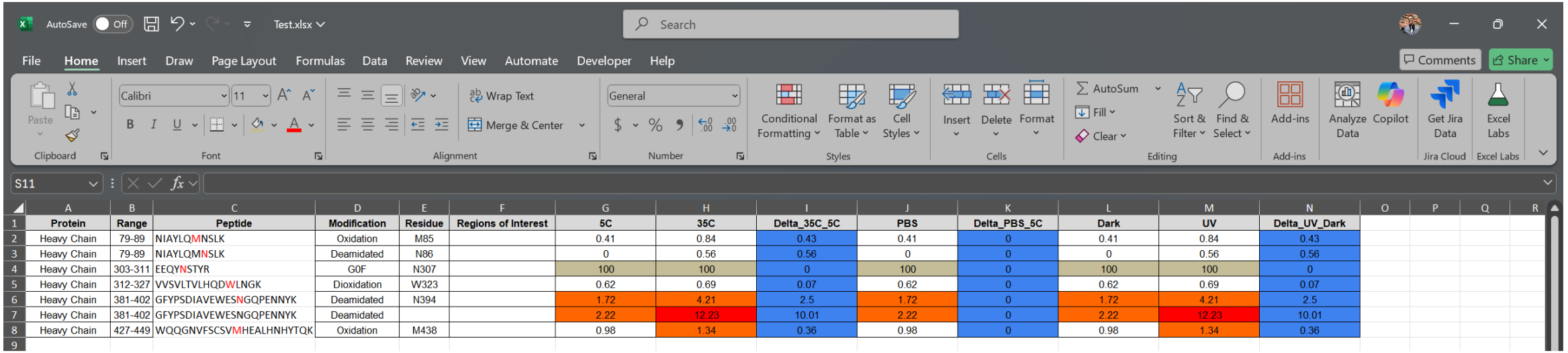
Case 3: Peptide Mapping - Reporting

Accepted results automatically sent to distributable report as well as LIMS



Benchling



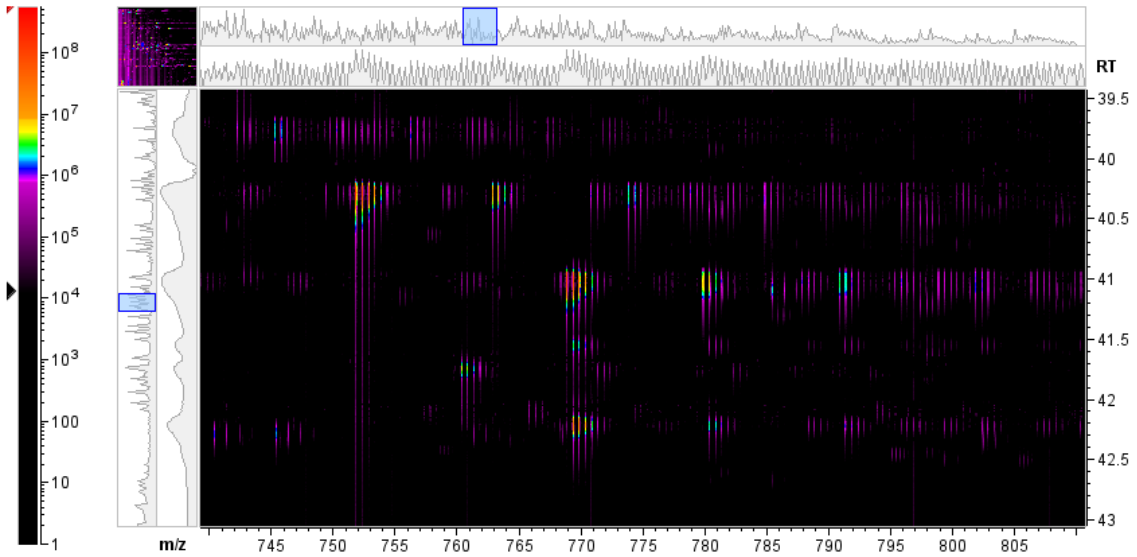


	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Protein	Range	Peptide	Modification	Residue	Regions of Interest	5C	35C	Delta_35C_5C	PBS	Delta_PBS_5C	Dark	UV	Delta_UV_Dark
2	Heavy Chain	79-89	NIAYLQMNSLK	Oxidation	M85		0.41	0.84	0.43	0.41	0	0.41	0.84	0.43
3	Heavy Chain	79-89	NIAYLQMNSLK	Deamidated	N86		0	0.56	0.56	0	0	0	0.56	0.56
4	Heavy Chain	303-311	EEQYNSTYR	G0F	N307		100	100	0	100	0	100	100	0
5	Heavy Chain	312-327	VVSVLTVLHQDWLNGK	Dioxidation	W323		0.62	0.69	0.07	0.62	0	0.62	0.69	0.07
6	Heavy Chain	381-402	GFYPSDIAVEWESNGQPENNYK	Deamidated	N394		1.72	4.21	2.5	1.72	0	1.72	4.21	2.5
7	Heavy Chain	381-402	GFYPSDIAVEWESNGQPENNYK	Deamidated			2.22	12.23	10.01	2.22	0	2.22	12.23	10.01
8	Heavy Chain	427-449	WQQGNVFSCSVMHEALHNHYTQK	Oxidation	M438		0.98	1.34	0.36	0.98	0	0.98	1.34	0.36
9														

Genedata Feature Highlights

These attributes are what empower scientist when using Genedata

Genedata



Vendor Neutral: Sciex, Bruker, Waters, Thermo, mZmL, Agilent, Shimadzu

Modular: Activity-based flexible workflows

Extensible: Functions can be modded with plugins from user

Interoperable: API is made to easily work with other software

Values and Impact

Completely Changed Ways of Our Working

Faster turnaround time

Reducing turn-around time (days to hours)

Enabling HT data processing
- HT Stability
- Bispecific QC
- HT Intact QC (a scale of 100s-1000s per batch)

Harmonized MS data analysis in both LBC and LO MS labs

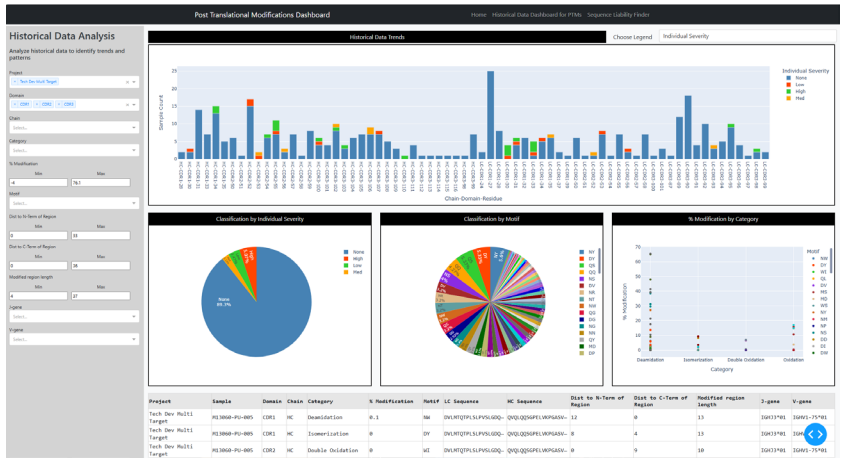
Data Foundation

From Unstructured Raw/Processed Data
- Manual transfer from source to report
- Manual authoring of descriptive narrative

To Structured Data
- Automated import of source to report
- Human and machine-readable output
- Data-driving raw to processed

Easier to access MS information

- Enabling AI/ML prediction model development for PTM
- Customized PTM Dashboard of repurposing historical data
- A living database in Benchling



Streamlined workflows are simply easy and efficient!

Joomi Ahn

Thank you

Q&A Session

Want to Digitalize & Automate Biotherapeutics R&D?

LEARN HOW

Genedata.com

CASSS™
SHARING SCIENCE SOLUTIONS