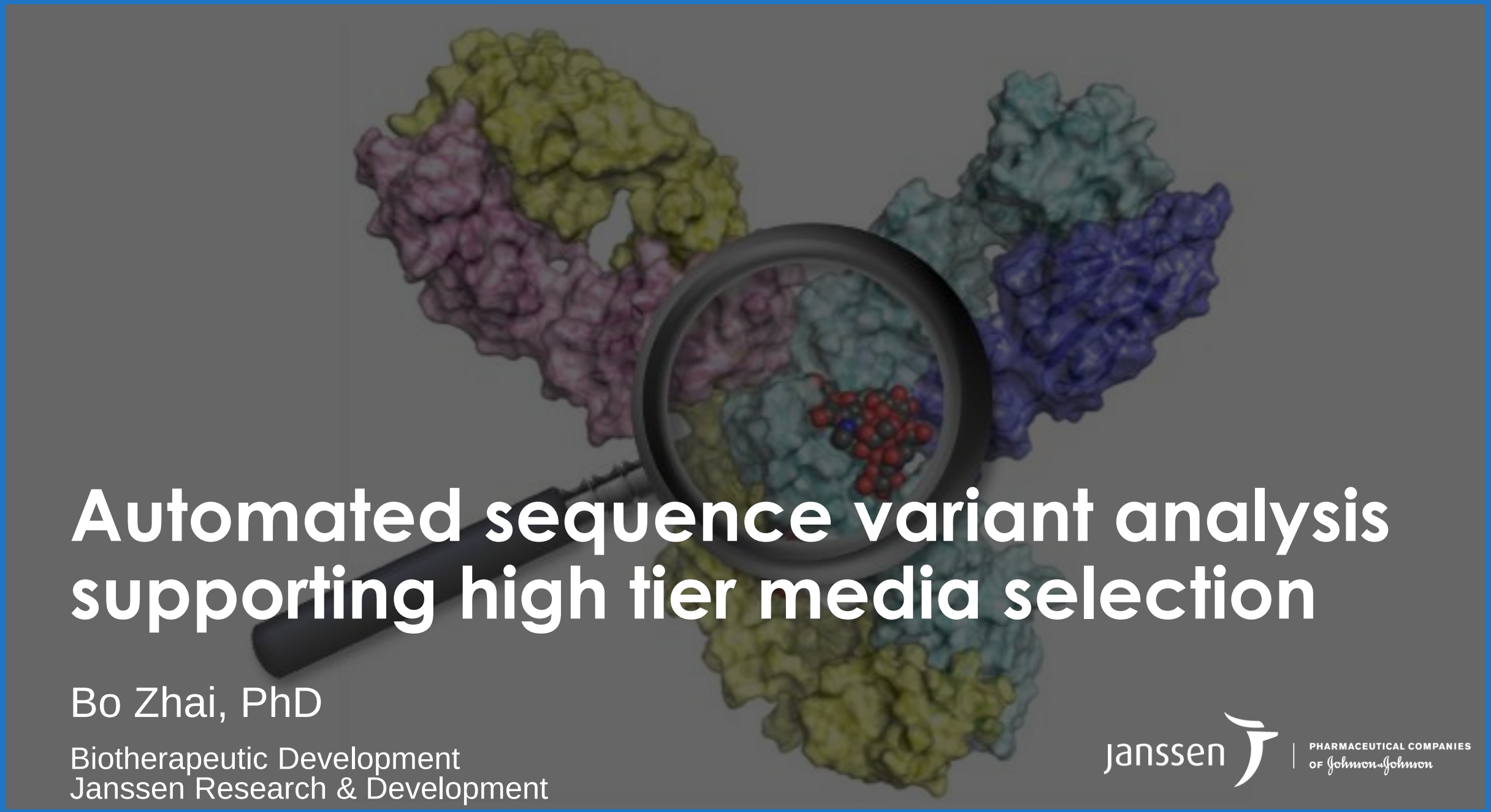


Automated sequence variant analysis supporting high tier media selection

Bo Zhai, PhD

Biotherapeutic Development
Janssen Research & Development

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Agenda

1

Introduction

Cell & Developability Sciences

Cell line selection process

2

Automated SVA

Minimal manual intervention

Less manual validation

3

Spent Media Analysis

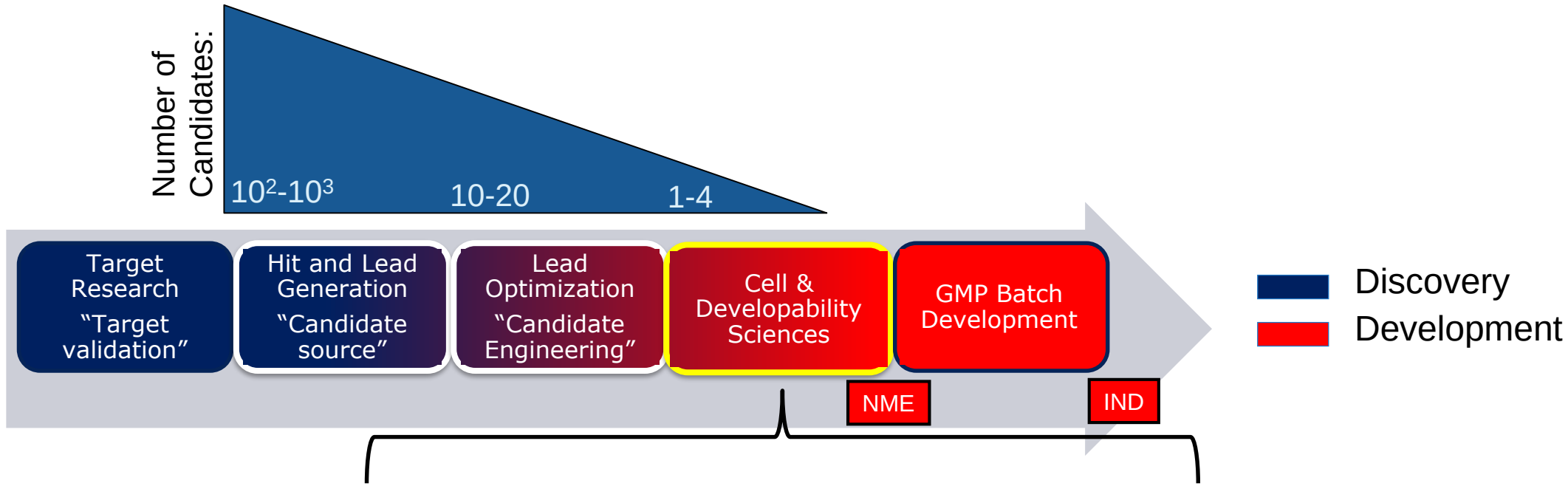
Verification

Cell line selection support

Introduction

1

Large Molecule Early Development

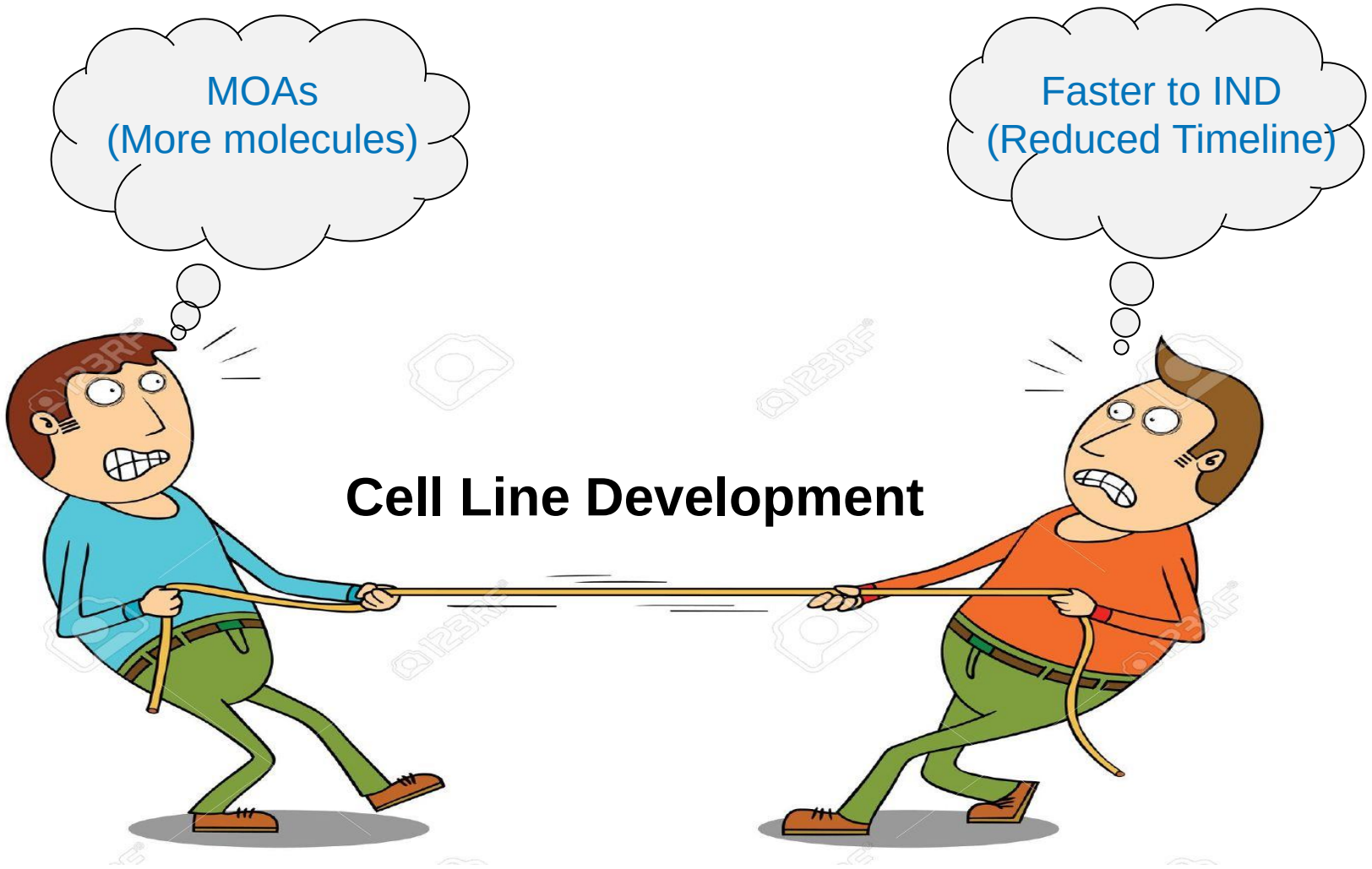


C e l l & D e v e l o p a b i l i t y S c i e n c e s

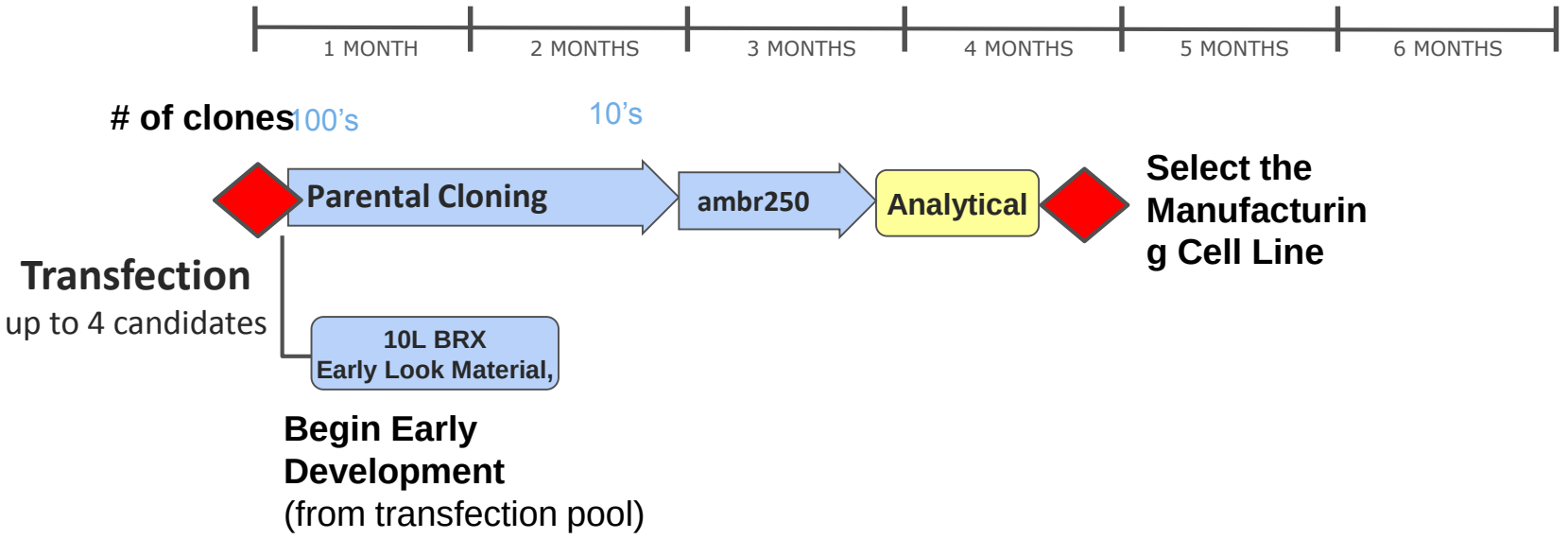
Cell Line Development | Early Formulation | pre-NME Analytical Support

Drive candidate selection and development of the manufacturing cell line

More Molecules and Reduced Timeline

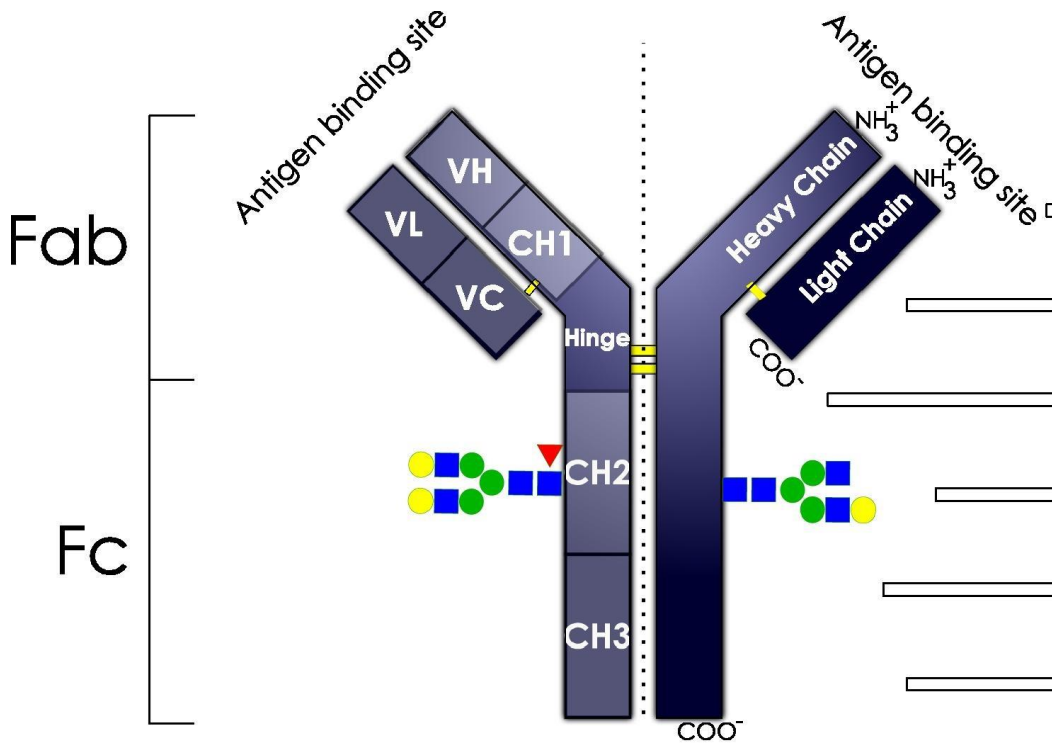


New Cell Line / New Process



- Eliminated subcloning reduced timeline by 1.5 months (VIPS Technology)
- Site directed integration. Titers are more predictable, screen fewer clones.
- 'Early Look Material' produced from transfection pool (not clonal): Start Development Sooner.

In Depth Characterization and Key Analytical Readouts



Proper chain recombination

N-terminal heterogeneity, Signal Peptide
Formation of Pyroglutamate

Modifications of amino acids
Deamidation, Isomerization, Glycation, Oxidation

Fragmentation
Cleavage in Hinge Region, Asp-Pro

Disulfide bonds
Free thiols, disulfide shuffeling, thioether

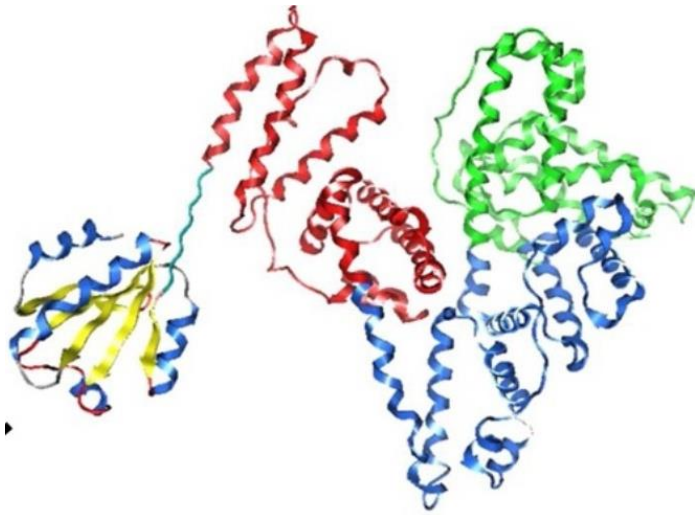
N-glycosylation & O-glycosylation
Fucosylation, Glalactosylation, Sialylation, Branching

C-terminal heterogeneity
Lysine variants, Amidation

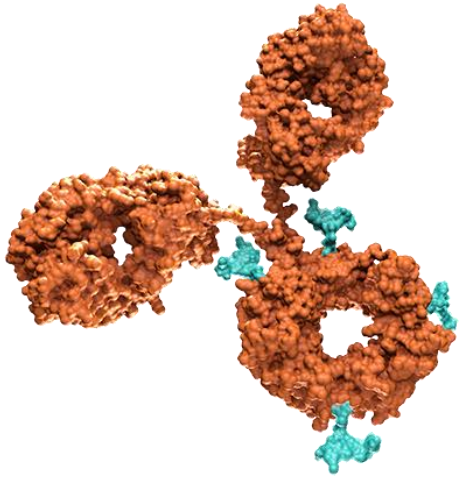
Sequence Variance

Off-Platform Programs Present New Challenges

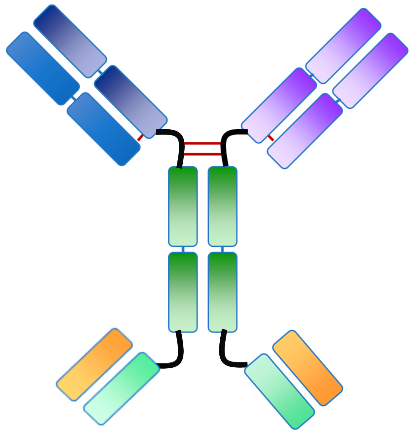
>50% of the Early Pipeline are not mAb



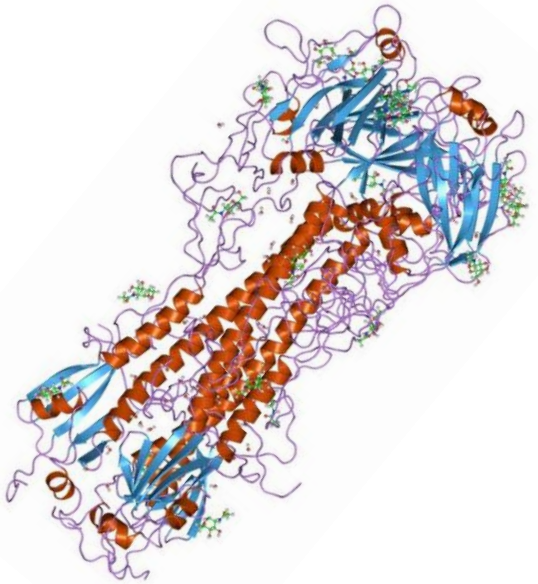
Alternative Scaffolds



ADC

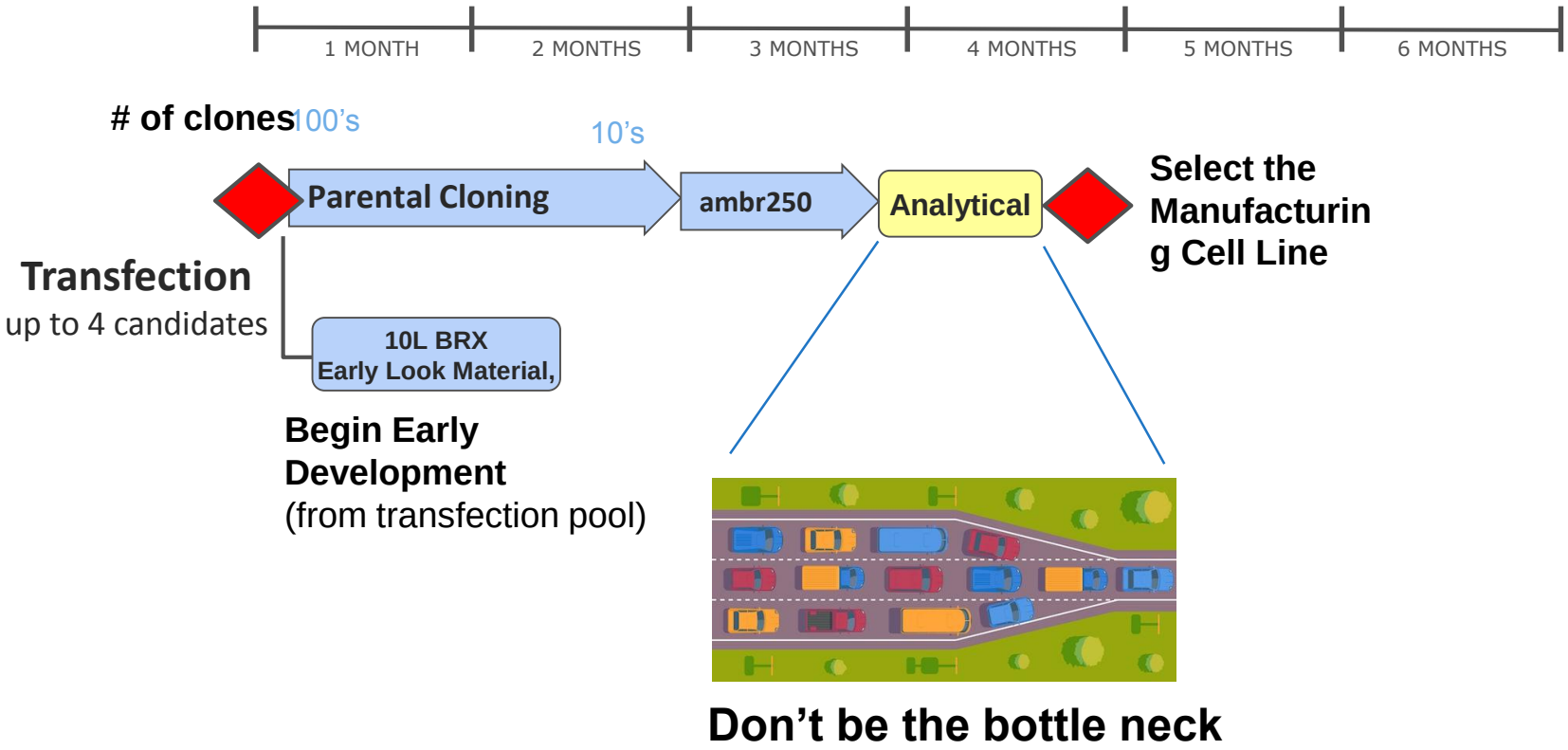


Multi-Specifics



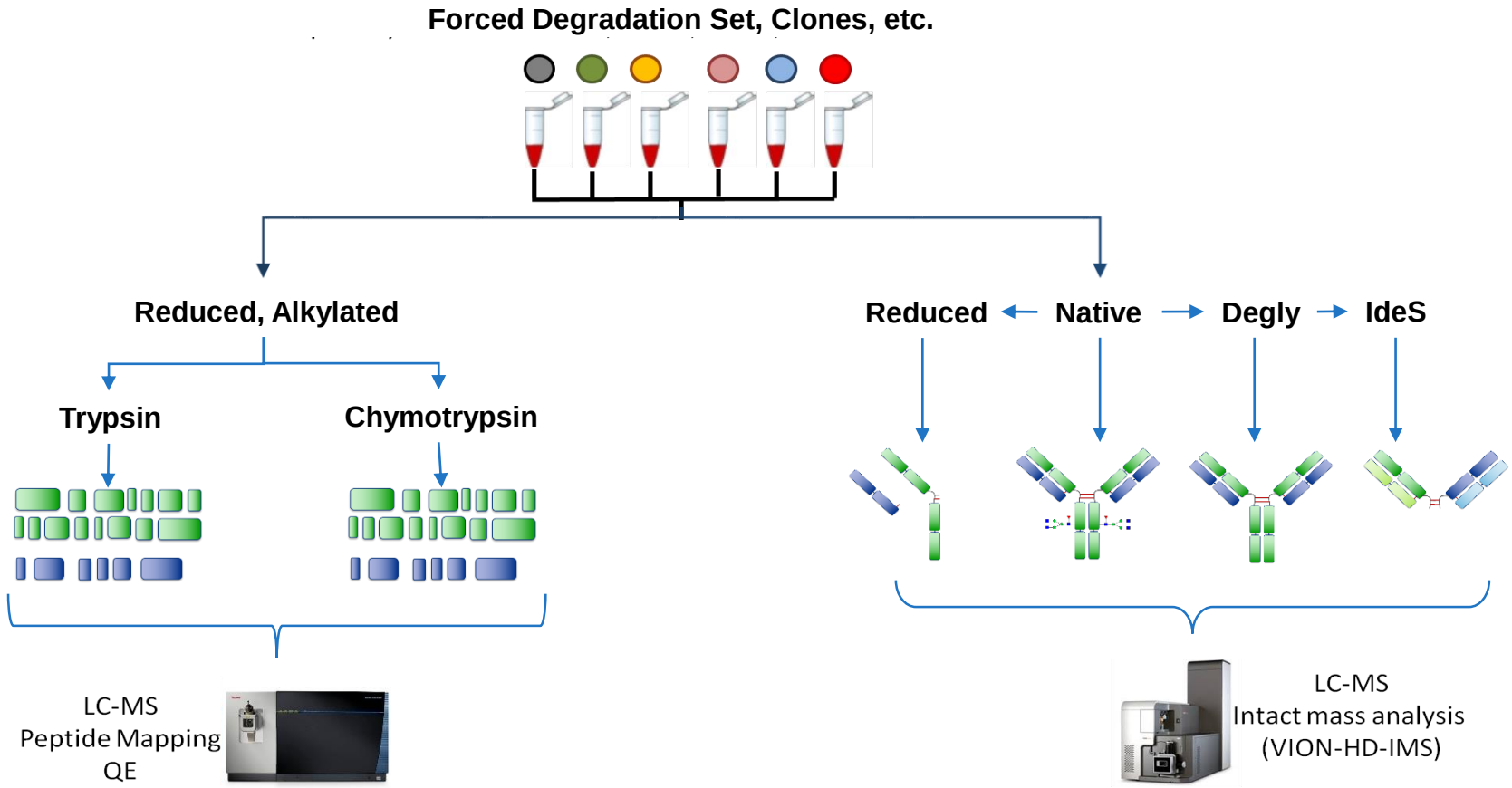
Vaccines

Don’t be the bottle neck



- Comprehensive characterization, high product quality in the lead cell line
- Analytical also has to shorten timelines and increase bandwidth

Bottom Up and Intact & Subunit Mass Approach



Peptide Mapping Analysis

Intact & Subunit Mass Analysis Pipeline

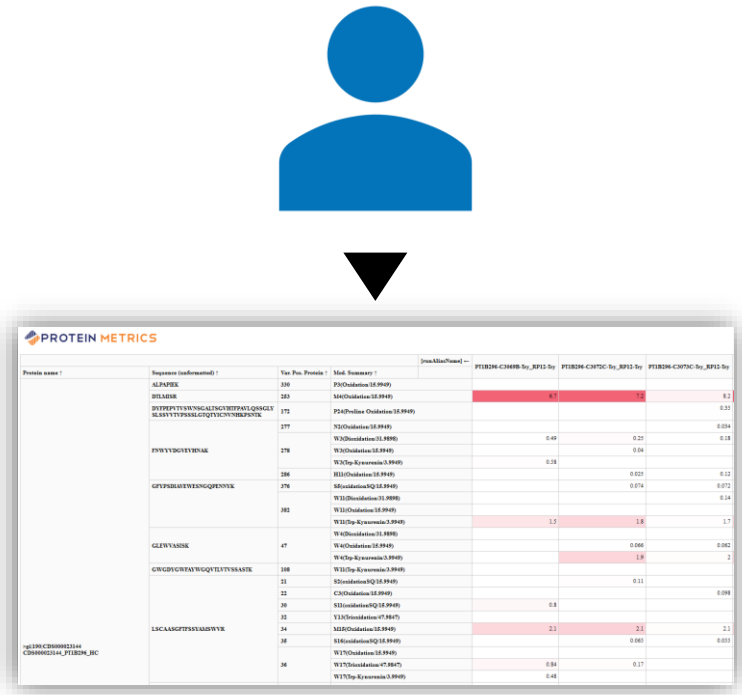
Peptide Mapping Auto Workflow with Byos™

PTMs

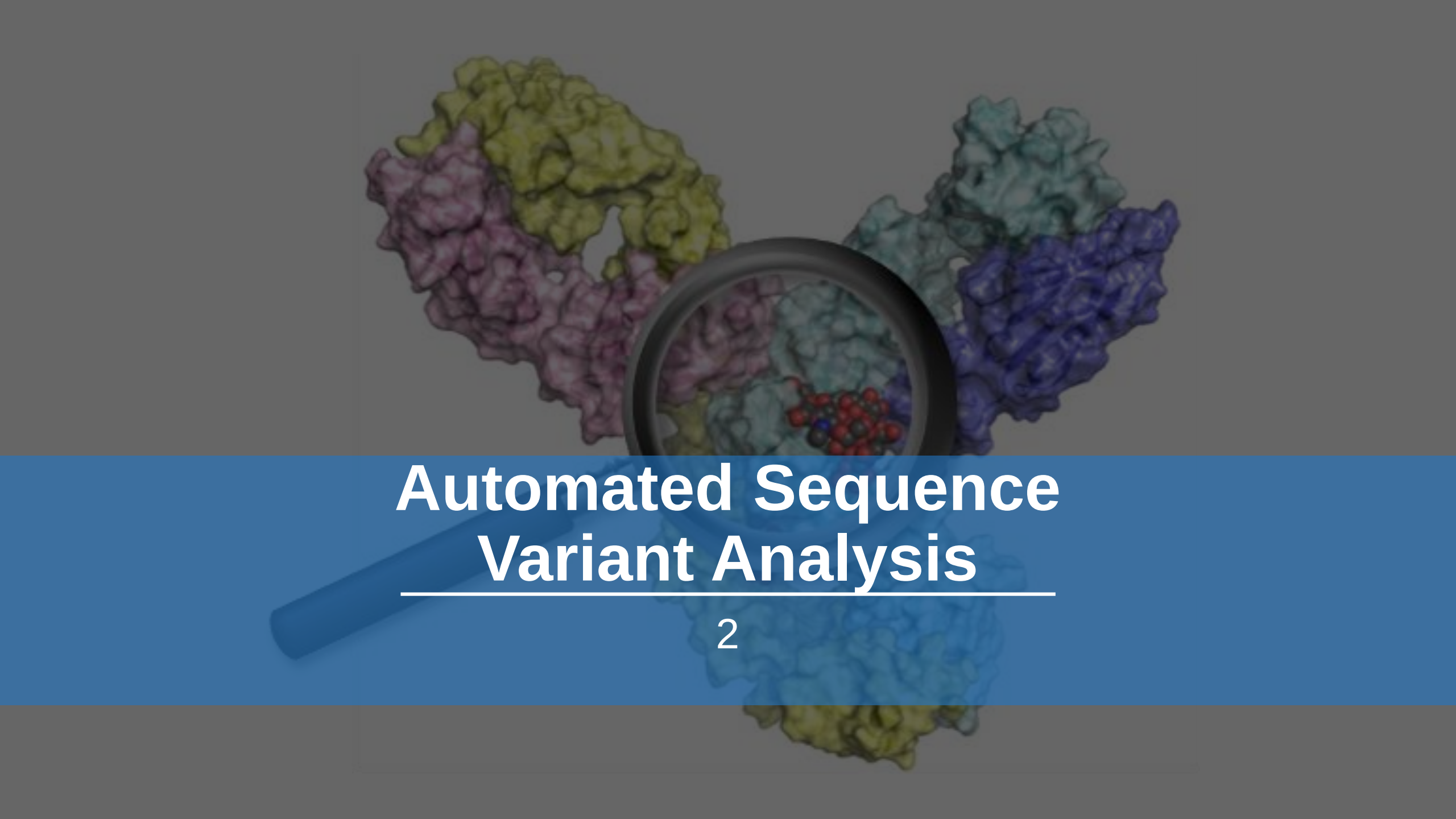
MS Files

Byos™

Protein Metrics Inc.



Quantitation



Automated Sequence Variant Analysis

2

Sequence Variant and Causes

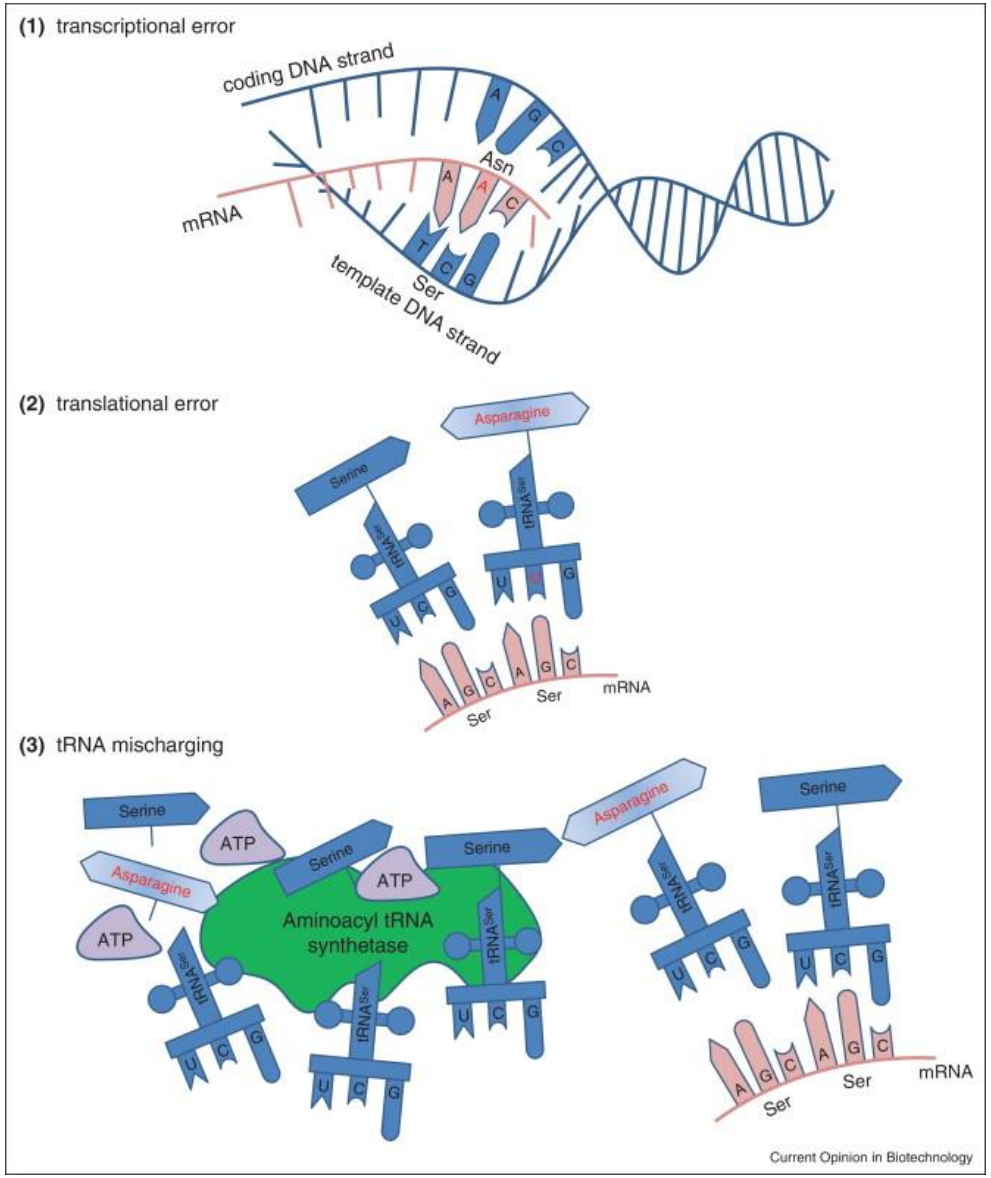
Unintended change in the amino acids sequences that could potentially contribute to efficacy, product safety and immunogenicity.

DNA mutations

- Replication error
- DNA damage

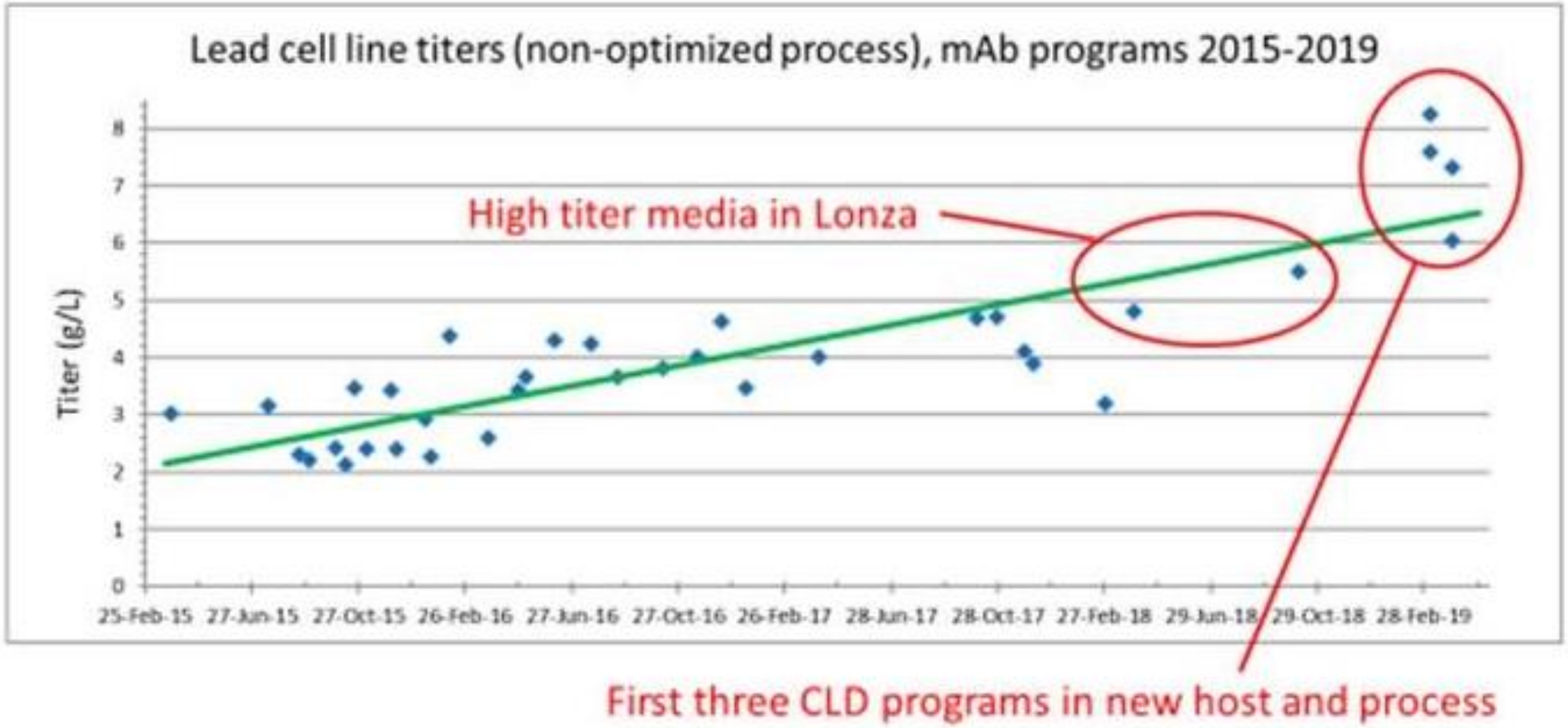
mRNA mistranslation

- Codon misreading (most common): incorrect tRNA is recruited during mRNA translation
- tRNA mischarging: incorrect amino acid is attached to the correct tRNA



1) Feeney, Lauren. et. al. (2013).. Biotechnology and Bioengineering. 110 (4)
2) Harris, Robert., Kilby, Peter. (2014) Current Opinion in Biotechnology. 30

Proprietary High Titer Media Improves Productivity



Sequence Variants Identified

Peptide	Position	Mod. Names	Clone 1	Clone 2	Clone 3	Clone 4	Clone 5	Clone 6	Clone 7	Clone 8
[illegible]	139/140	Cys->Tyr/3.0327	0.11	0.07	0.07	0.12	0.04	0.10	0.14	0.06
[illegible]	217	Cys->Tyr/3.0327							0.10	
[illegible]	302	Cys->Tyr/3.0327							0.22	
[illegible]	314	Cys->Tyr/3.0327	0.09	0.05	0.04	0.08	0.03	0.08	0.10	0.04
[illegible]	327	Cys->Tyr/3.0327		0.04	0.04	0.06		0.05	0.12	
[illegible]	328	Cys->Tyr/3.0327		0.04						

Peptide	Position	Mod. Names	Clone A	Clone B	Clone C	Clone D	Clone E	Clone F	Clone G	Clone H
[illegible]	119	Pro->Ala/-26.0157	0.26							
[illegible]	204	Pro->Ala/-26.0157	0.24				0.05			
[illegible]	330	Pro->Ala/-26.0157	0.35		0.10			0.18		
[illegible]	332	Pro->Ala/-26.0157	0.27					0.13		
[illegible]	30	Asn->Ser/-27.0109					0.04			0.39
[illegible]	287	Asn->Ser/-27.0109								0.52
[illegible]	362	Asn->Ser/-27.0109								0.50

Cys	TGC	**Tyr**	TAT
	TGT		TAC

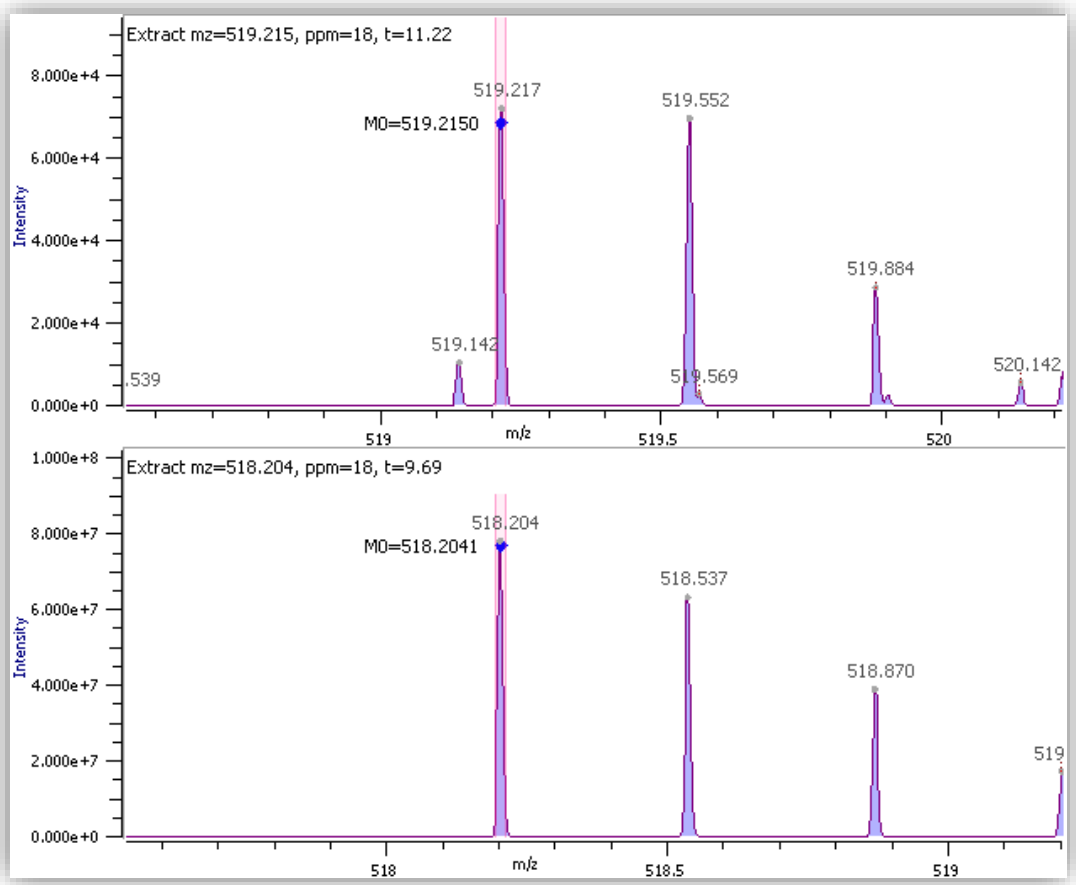
Pro	CCT	**Ala**	GCT
	CCA		GCA
	CCG		GCG
	CCC		GCC

Initial Report on Sequence Variants

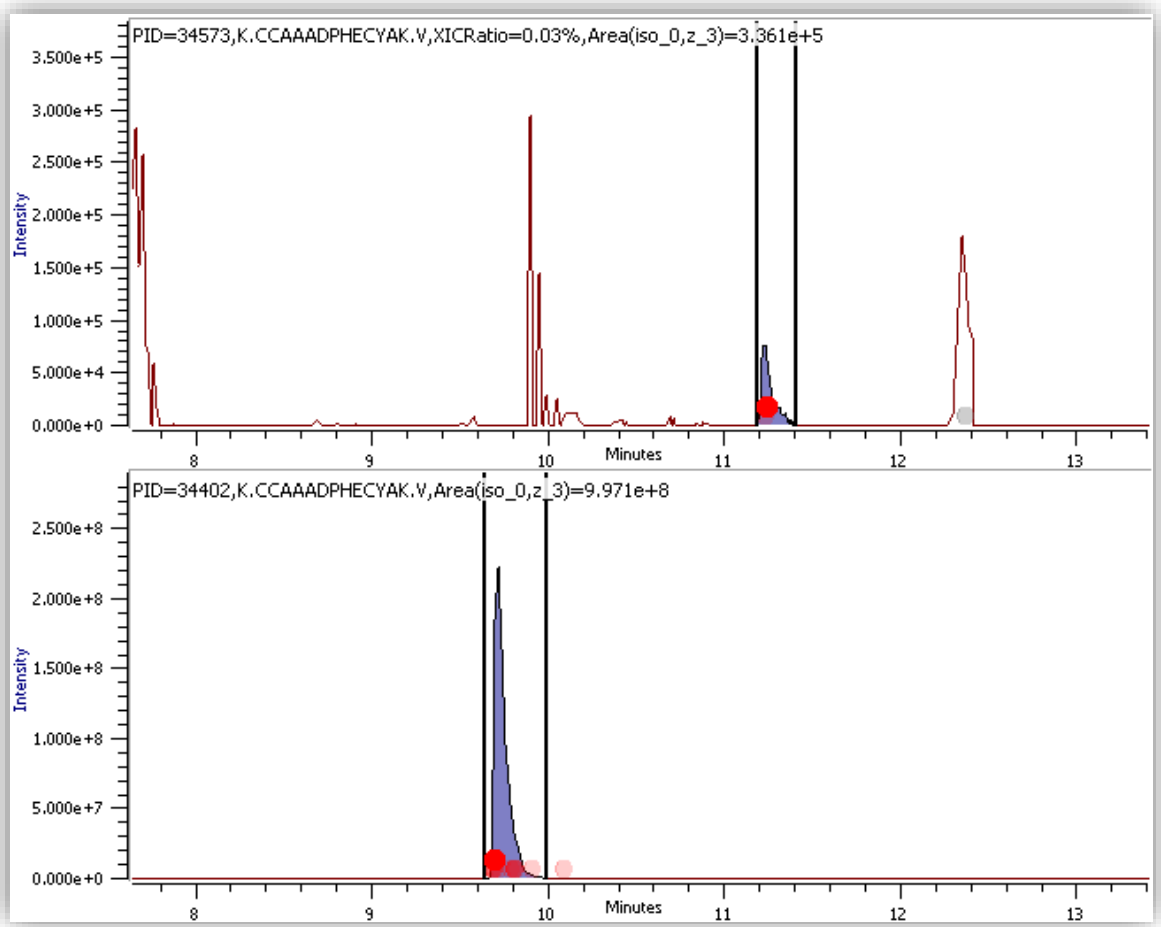
PROTEIN METRICS

				Digest name →			
				MS Id ←	9	10	11
				[runAliasName] ←	GGDW53-331714-RCN-10-Try_RP11-Try (%)	GGDW53-331714-RCN-5-Try_RP11-Try (%)	GGDW53-331715-RCN-13-Try_RP11-Try (%)
Protein name ↑	Sequence (unformatted) ↑	Var. Pos. Protein ↑	Mod. Names ↑				
	[redacted]	224	Ala->Pro/26.0157			0.03	0.02
	[redacted]	217	Cys->Tyr/3.0327		0.06	0.07	0.07
	[redacted]	222	Asp->Asn/-0.9840		0.34	0.31	0.33
	[redacted]	704	Ala->Pro/26.0157		0.05	0.03	0.03
	[redacted]	616	Cys->Tyr/3.0327		0.02		
	[redacted]	70	Ala->Pro/26.0157		0.03	0.02	0.02
	[redacted]	87	Asp->Asn/-0.9840				
	[redacted]	90	Lys->Arg/28.0062				0.01
	[redacted]	680	Ala->Ser/15.9949				
	[redacted]	262	Ala->Ser/15.9949				0.09
	[redacted]	409	Cys->Tyr/3.0327		0.03	0.02	0.02
	[redacted]	410	Cys->Tyr/3.0327		0.03	0.01	0.02
	[redacted]	418	Cys->Tyr/3.0327		0.06	0.05	0.04
	[redacted]	526	Cys->Tyr/3.0327		0.08	0.04	0.04
	[redacted]	530	Leu->Pro/-16.0313		18.14	16.90	
	[redacted]	530	Leu->Pro/-16.0313				
	[redacted]	540	Leu->Pro/-16.0313				
	[redacted]	376	Leu->His/23.9748		1.70	1.72	1.74
	[redacted]	380	Leu->His/23.9748				
	[redacted]	565	Leu->His/23.9748			1.55	
	[redacted]	568	Lys->Arg/28.0062			0.17	
	[redacted]	594	Lys->Arg/28.0062				
	[redacted]	595	Ala->Glu/58.0055		0.70	2.28	1.56

Cys->Tyr Variant Manual Validation

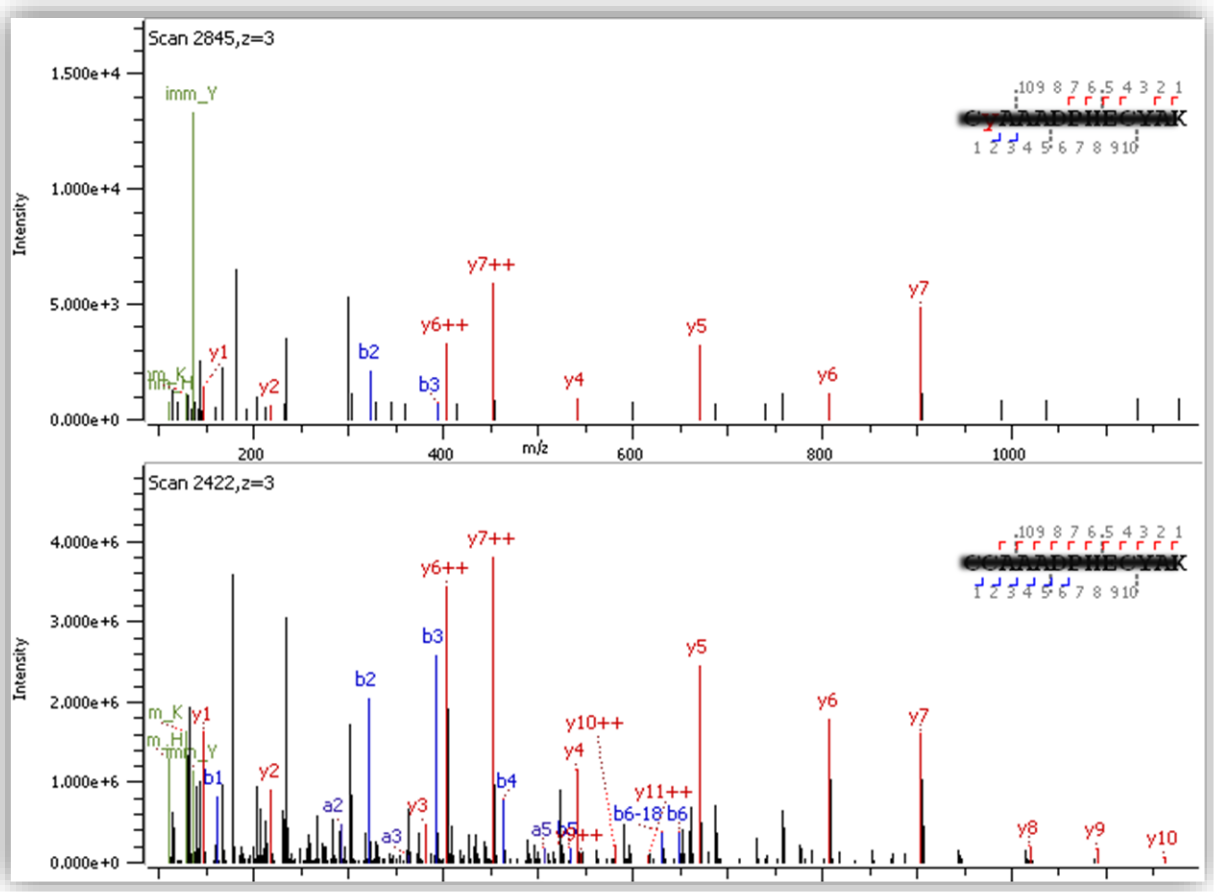


MS1

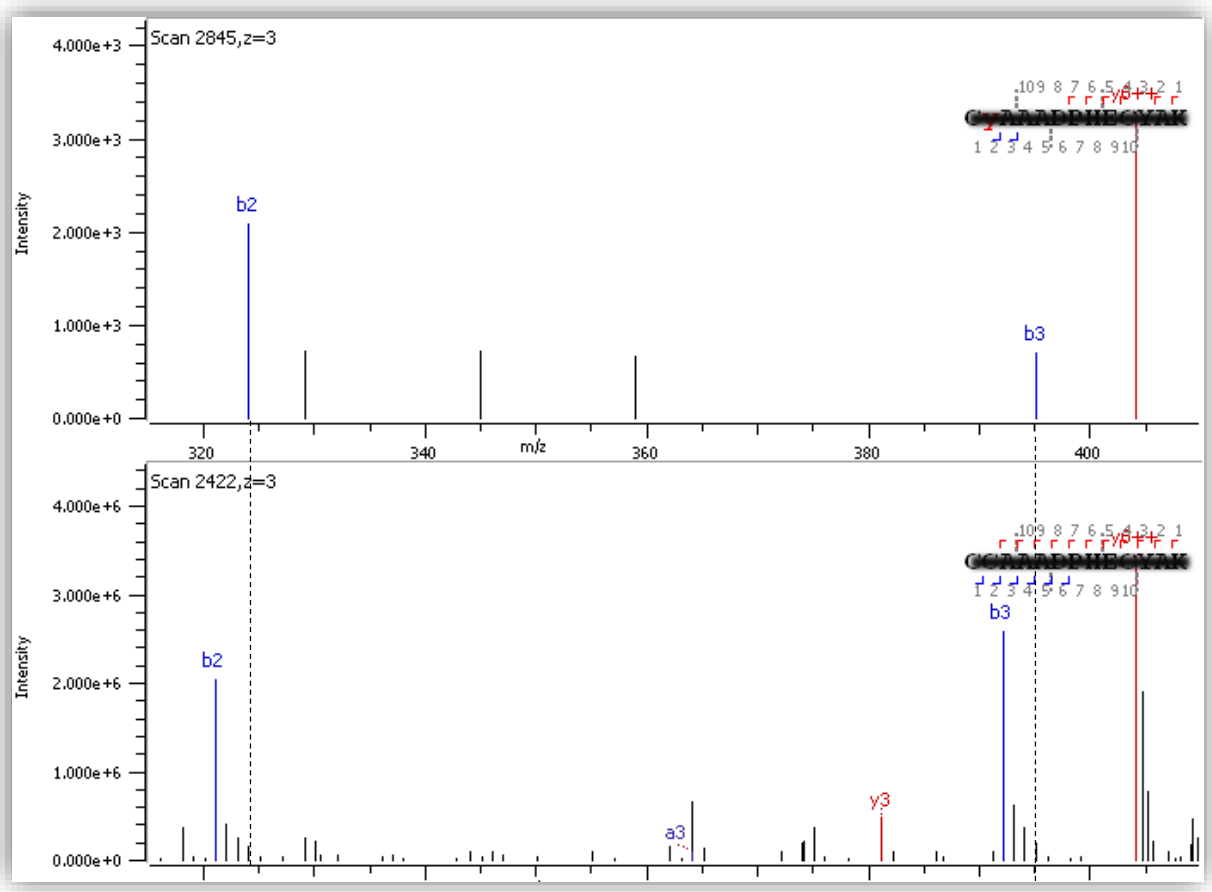


XIC

Cys->Tyr Variant Manual Validation



MSMS



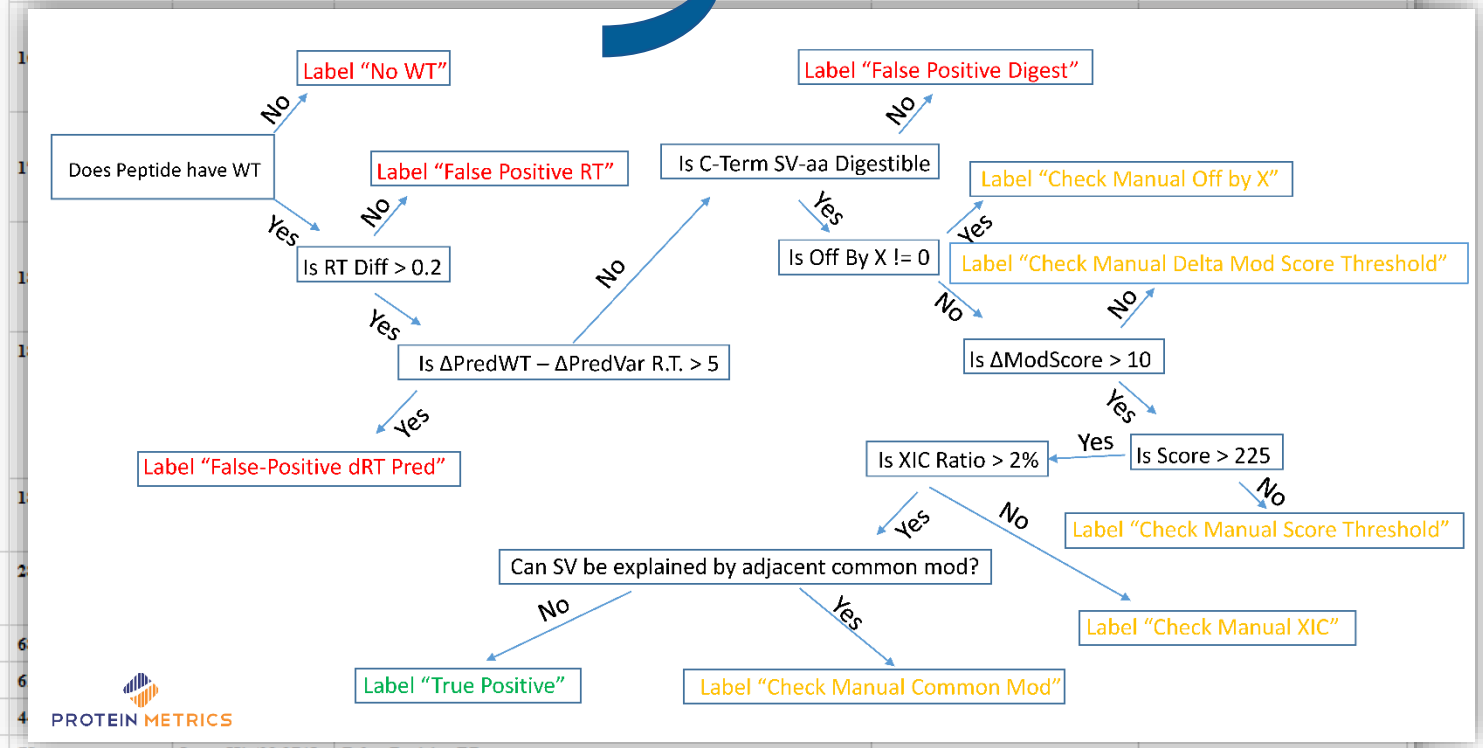
MSMS

Sequence Variant Annotated Based on Filter Logic

PROTEIN METRICS

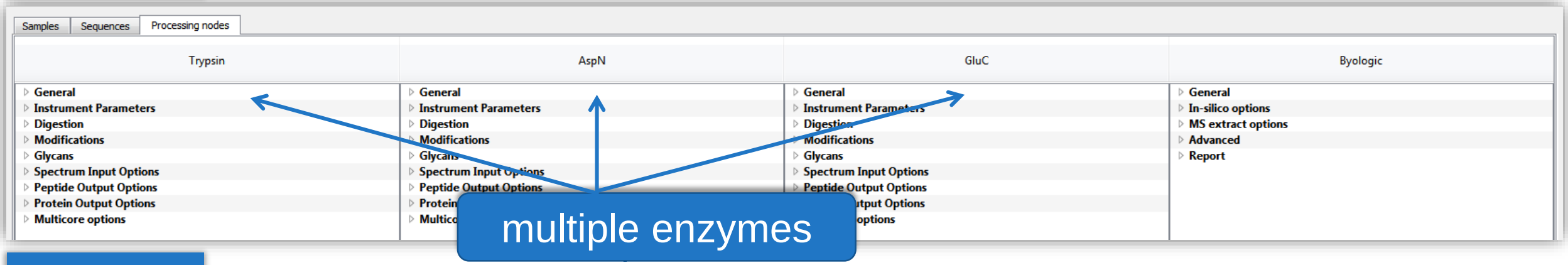
Protein name ↑	Sequence (unformatted) ↑	Var. Pos. Protein ↑	Mod. Names ↑	{SVA Filter} ↑	[runAliasName] ←	PT1B296-C3069B-Try_RP12-Try	PT1B296-C3072C-Try_RP12-Try
>gi|190|CDS000023144 CDS000023144_P[redacted]_HC	DNSKNTLYLQMNSLR	84	Ser->Phe/60.0364	False-Positive dRT Pred			
	[redacted]GLY	160		Check Manually Off by X			
			Asn->His/23.0160	Check Manually XIC			3
				False-Positive RT			
			Asn->Ile/-0.9589	Check Manually Off by X			
		163	Ala->Pro/26.0157	False-Positive RT			
	FTISR	6					
	[redacted]	6					
	[redacted]	4					
	[redacted]	78	Leu->His/23.9748	False-Positive RT			

- Monoisotopic peak
- Peptide score
- Retention time shift
- Explained by common mod
- K/R variance

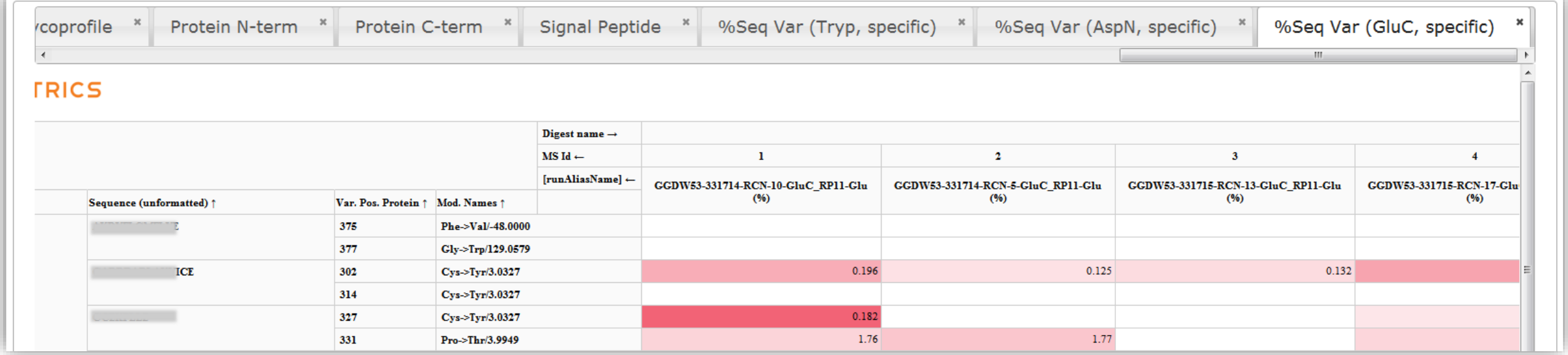


Cross validation with multiple enzymes

Processing



Reporting

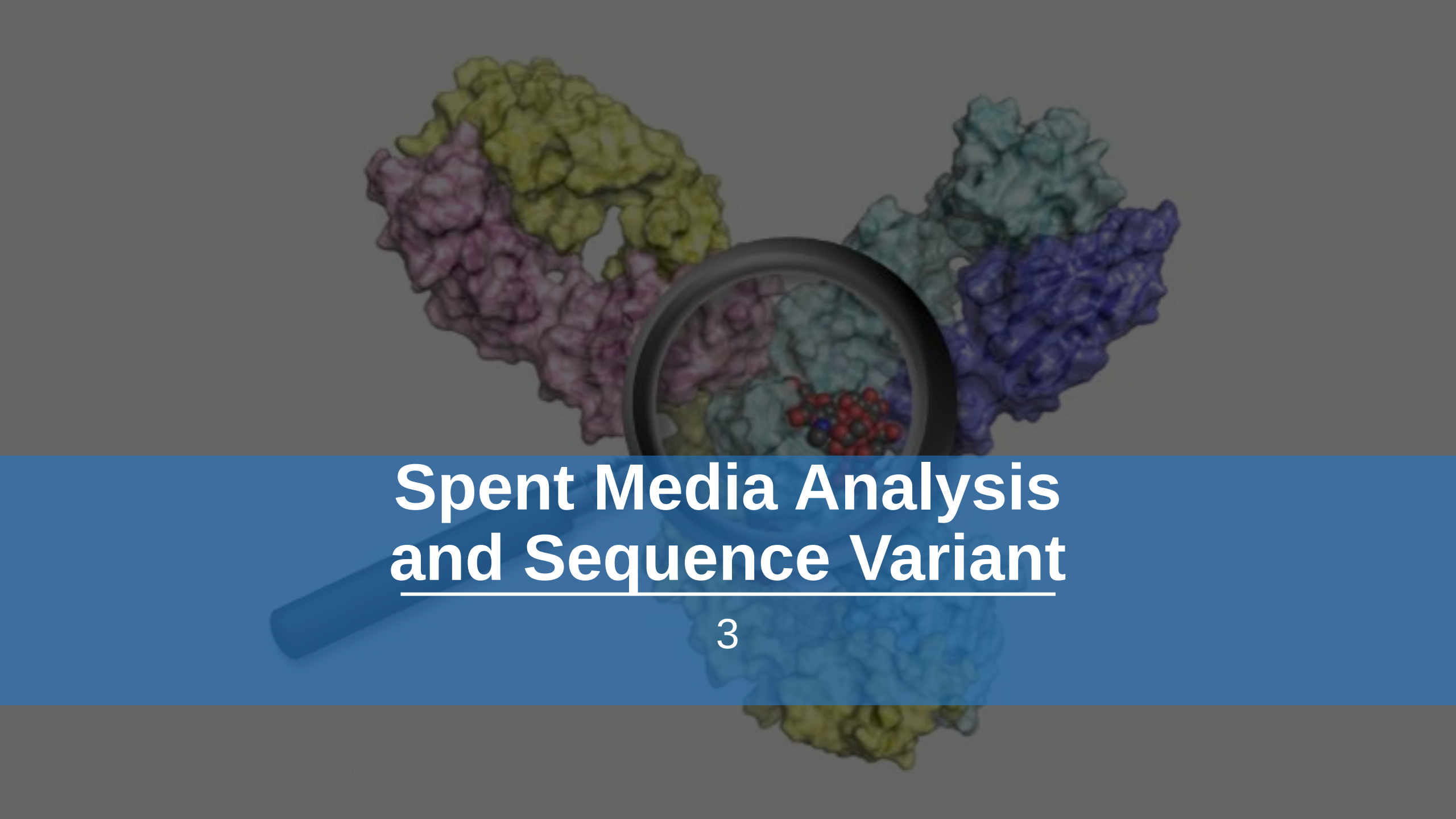


			Digest name →				
			MS Id ←	1	2	3	4
			[runAliasName] ←	GGDW53-331714-RCN-10-GluC_RP11-Glu (%)	GGDW53-331714-RCN-5-GluC_RP11-Glu (%)	GGDW53-331715-RCN-13-GluC_RP11-Glu (%)	GGDW53-331715-RCN-17-Glu (%)
	Sequence (unformatted) ↑	Var. Pos. Protein ↑	Mod. Names ↑				
	[redacted]	375	Phe->Val/-48.0000				
		377	Gly->Trp/129.0579				
	[redacted]ICE	302	Cys->Tyr/3.0327	0.196	0.125	0.132	
		314	Cys->Tyr/3.0327				
	[redacted]	327	Cys->Tyr/3.0327	0.182			
		331	Pro->Thr/3.9949	1.76	1.77		

Cross validation with multiple enzymes

Peptide (Trypsin)	Position	Mod. Names	Clone 1	Clone 2	Clone 3	Clone 4	Clone 5	Clone 6	Clone 7	Clone 8
[illegible]	139/140	Cys->Tyr/3.0327	0.11	0.07	0.07	0.12	0.04	0.1	0.14	0.06
[illegible]	217	Cys->Tyr/3.0327							0.1	
[illegible]	302	Cys->Tyr/3.0327							0.22	
[illegible]	314	Cys->Tyr/3.0327	0.09	0.05	0.04	0.08	0.03	0.08	0.1	0.04
[illegible]	327	Cys->Tyr/3.0327		0.04	0.04	0.06		0.05	0.12	
[illegible]	328	Cys->Tyr/3.0327		0.04						

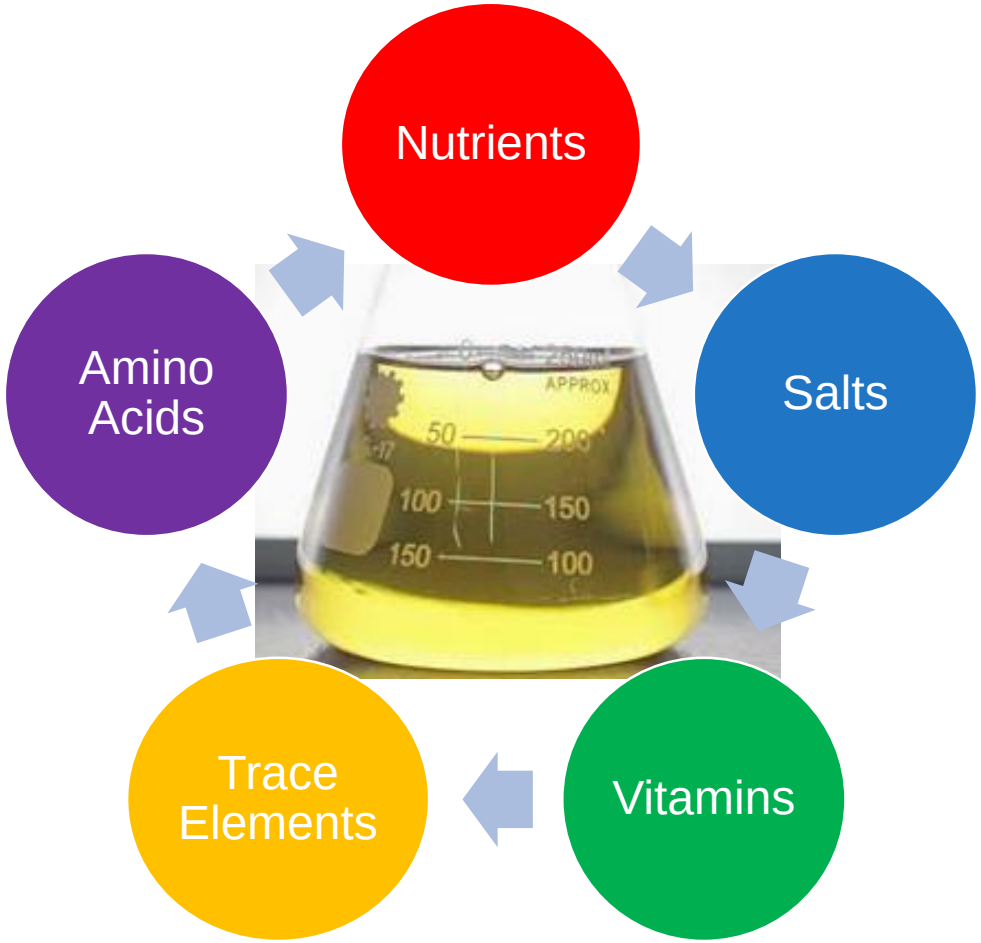
Peptide (GluC)	Position	Mod. Names	Clone 1	Clone 2	Clone 3	Clone 4	Clone 5	Clone 6	Clone 7	Clone 8
[illegible]	139/140	Cys->Tyr/3.0327	0.09	0.05	0.06	0.10	0.04	0.08	0.11	0.05
[illegible]	314	Cys->Tyr/3.0327	0.20	0.13	0.13	0.21	0.08	0.19	0.30	0.12
[illegible]	302	Cys->Tyr/3.0327							0.11	
[illegible]	327	Cys->Tyr/3.0327	0.19			0.06		0.07	0.09	0.04

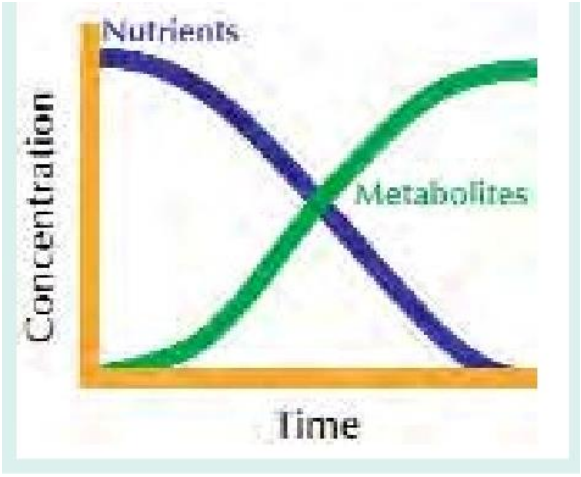


Spent Media Analysis and Sequence Variant

3

Introduction of Spent Media Analysis





- Amino acids and nutrients consumed
- Metabolites and by-products secreted

Purpose of Spent Media Analysis

- Optimize the growth conditions for clones
- Promote high cell viability
- Maximize productivity (titer)
- Ensure high quality product (Minimize clipping, PTM)
- Reduce/eliminate misincorporations (sequence variant)

Spent Media LC-MS Workflow (Scheduled MRM)

1290 Infinity II LC

UHPLC separation

AdvanceBio MS Spent Media HILIC Column

A wide range of compound classes

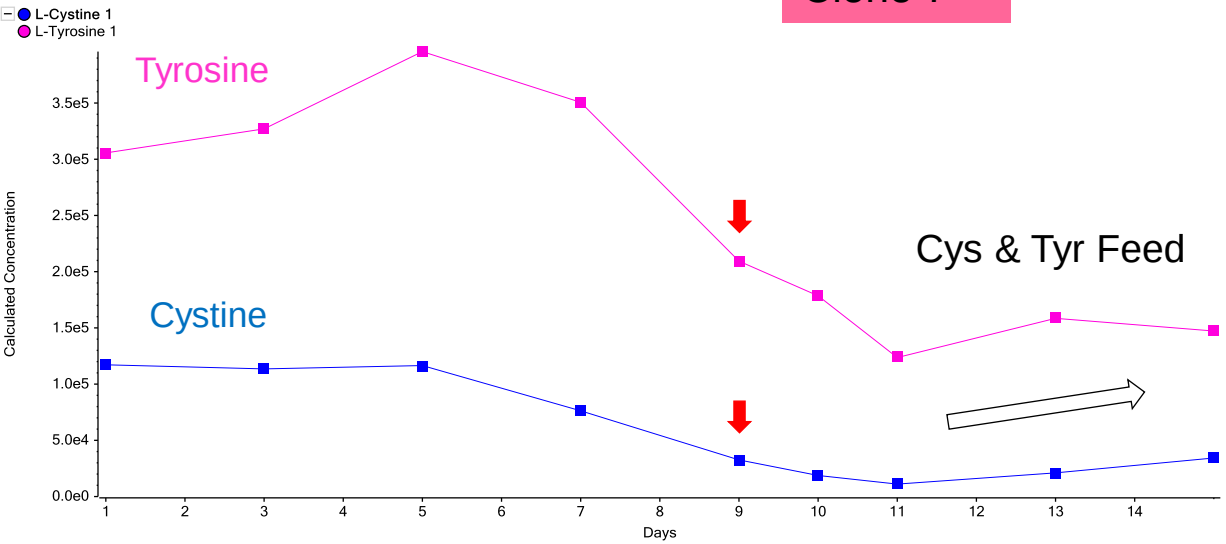
SCIEX QTRAP 6500+

Speed and Sensitivity

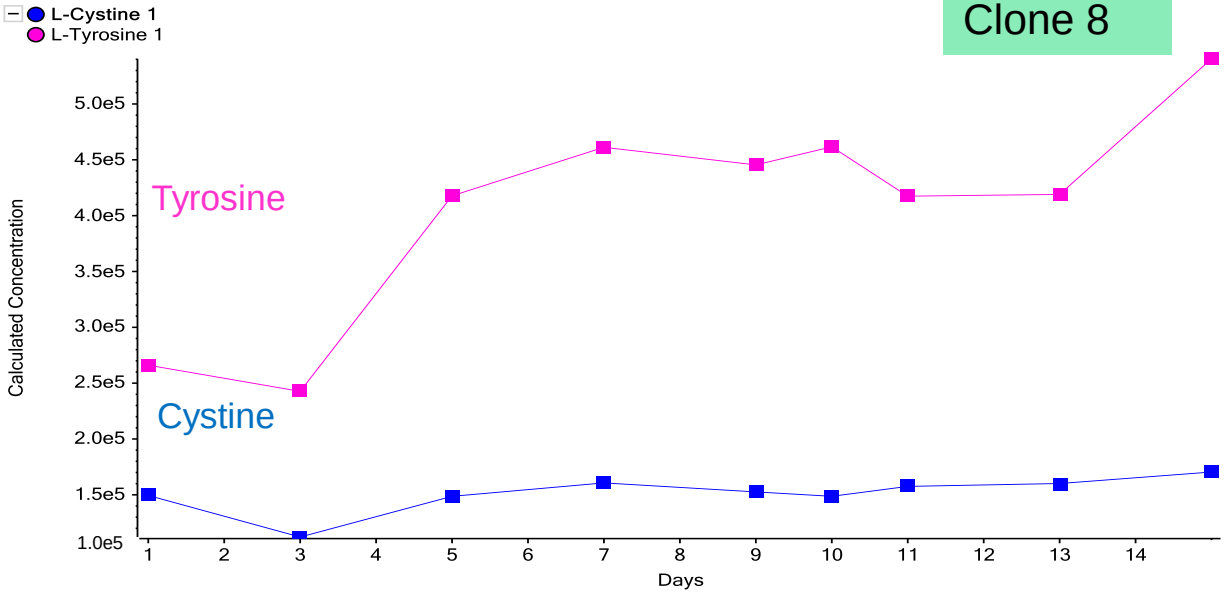
P-119-T: Andrew Mahan

Correlating the C->Y to Amino Acids Depletion

Clone 7



Clone 8



Trend plots show the timepoint the depletion started.

LC-MS Spent Media Analysis and Clone Selection

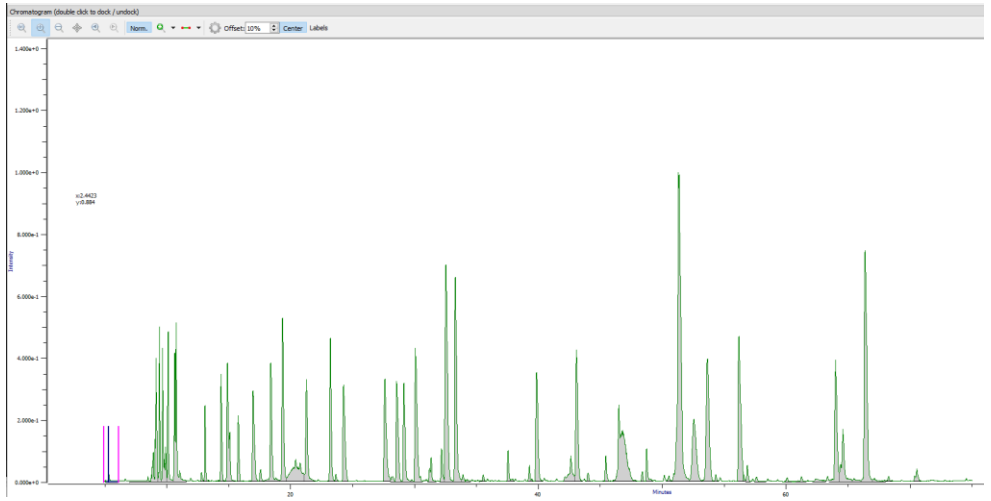
LC-MS based Spent Media Analysis

Allows for the quantitative measure of different classes of metabolites & amino acids in a 20 min LC-MS run.

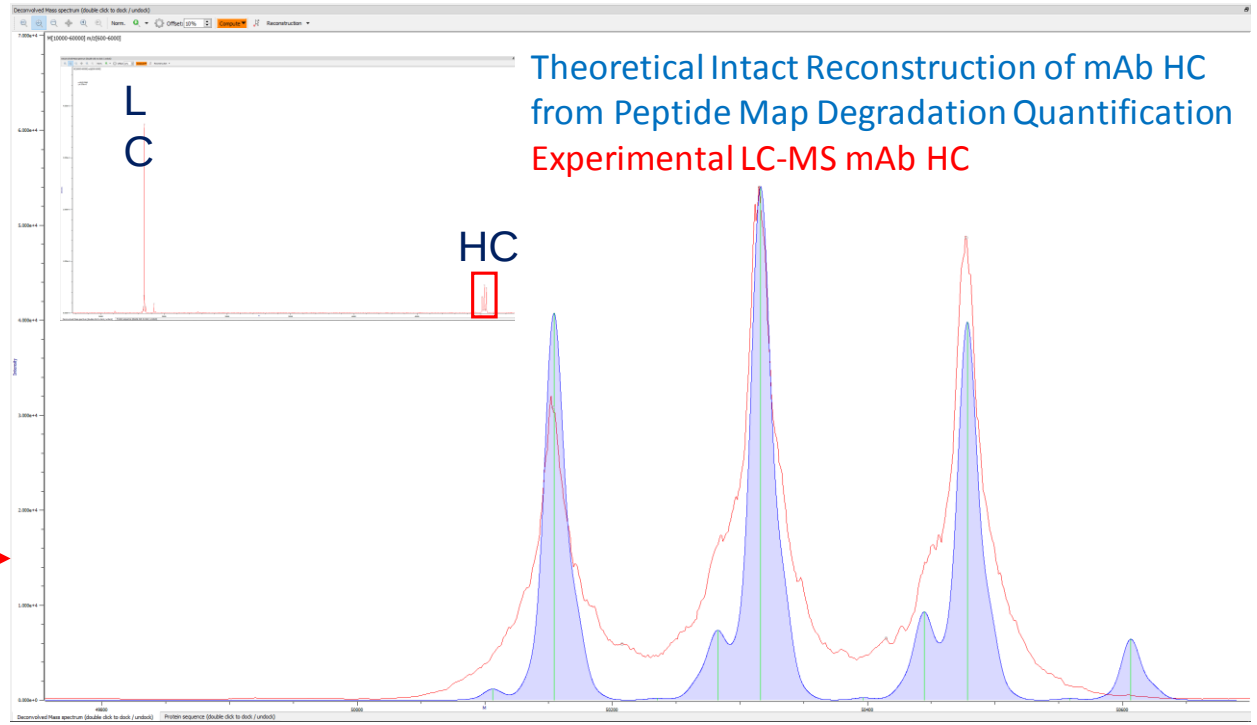
Implementation into Clone Selection

Verifying of misincorporation observed in peptide map analysis.
Monitoring amino acids and metabolites profiles of different clones.

Intact Mass Reconstruction with Peptide Mapping Data



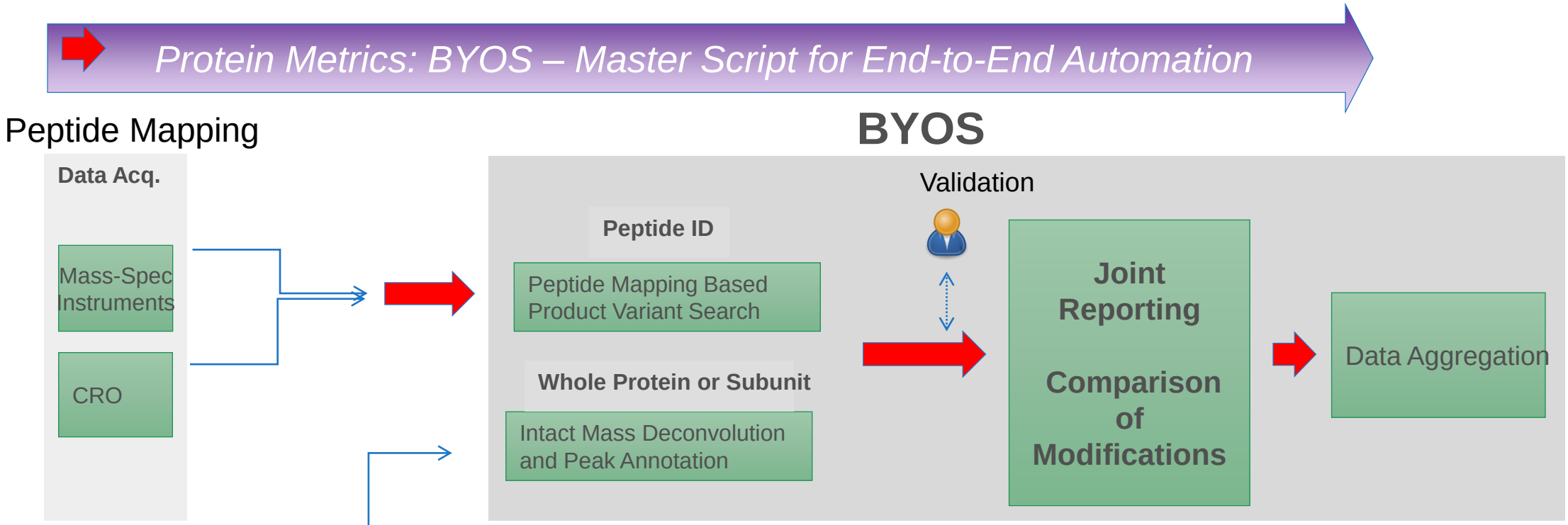
						MS Id ←	1
						MS Alias name ←	AD226_C4LB231-01_Trypsin_5ug_REL (%) ↓
Protein name	Sequence (unformatted)	Mod. Names	Mod. AAs	Var. Pos. Protein	Labels		
gi|270|CDS000028779 CDS000028779_C4LB231_HC	QVQLVQSGAEVK	Asn-Succinimide/-17.0265	Q	3			3.93
gi|271|CDS000024830 CDS000024830_C4LB231_LC	TVAAPSVFIFPPSDEQLK	Transpeptidation R/156.1011	NTerm	108			1.19
gi|270|CDS000028779 CDS000028779_C4LB231_HC	VVSVLTVLHQDWLNGKEYK	Deamidated/0.9840	N	316			0.447
gi|271|CDS000024830 CDS000024830_C4LB231_LC	SGTASVVCLLNNFYPR	Deamidated/0.9840	N	137			0.441
gi|270|CDS000028779 CDS000028779_C4LB231_HC	VVSVLTVLHQDWLNGK	Deamidated/0.9840	N	316			0.366
gi|271|CDS000024830 CDS000024830_C4LB231_LC	DIQMTQSPSSLSASVGDR	Over Alkylation/57.0215	M	4			0.27
gi|270|CDS000028779 CDS000028779_C4LB231_HC	GFYPSDIAVEWESNGQPENNYK	Deamidated/0.9840	N	385			0.267
gi|271|CDS000024830 CDS000024830_C4LB231_LC	TVAAPSVFIFPPSDEQLK	Pro->Arg/69.0483	P	113			0.25
gi|270|CDS000028779 CDS000028779_C4LB231_HC	DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGL YSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVD	Transpeptidation K/128.0950	CTerm	214			0.223
	DTLMISR	Oxidation/15.9949	M	253			0.22
	QAPGQGLEWMGWISPIFGNTNYAQK	Over Alkylation/57.0215	M	48			0.219
gi|271|CDS000024830 CDS000024830_C4LB231_LC	LLIYYANSLQSGVPSR	Pro->Arg/69.0483	P	59			0.215
gi|270|CDS000028779 CDS000028779_C4LB231_HC	HTCPPCPAPEAAGASSVFLFPPKPK	Ala->Gly/-14.0157	A	232			0.213
	DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGL VSLSSVVTVPSSSLGTQTYICNVNHKPSNTK	Tyr->Cys/-3.0327	Y	181			0.204
	QAPGQGLEWMGWISPIFGNTNYAQK	Gly->Asp/58.0055	G	49			0.19
	VVSVLTVLHQDWLNGK	Asn-Succinimide/-17.0265	Q	312			0.187
	VTITADESTSTAYMELSSLR	Over Alkylation/57.0215	M	81			0.177
	QVQLVQSGAEVKKPGSSVK	Gln-pyro-Glu/-17.0265	Q	1			0.158
		Asn-Succinimide/-17.0265	Q	3			0.158
gi|271|CDS000024830 CDS000024830_C4LB231_LC	VDNALQSGNSQESVTEQDSK	Leu->His/23.9748	L	154			0.152
	TVAAPSVFIFPPSDEQLK	Ser->Phe/60.0364	S	114			0.141
	DIQMTQSPSSLSASVGDR	Pro->Arg/69.0483	P	8			0.138
gi|270|CDS000028779 CDS000028779_C4LB231_HC	DTLMISR	Dethiomethyl/-48.0034	M	253			0.137
	WQQGNVFSCSVMHEALHNHYTQK	Oxidation/15.9949	M	429			0.134
		Over Alkylation/57.0215	C	426			0.13
gi|271|CDS000024830 CDS000024830_C4LB231_LC	ASQSISSYLNWYQQKPGK	Dehydrated/-18.0106	S	31			0.12



- Can help determine if modifications are stochastic or not.
- Quickly compare quantification between bottom up and intact data.
- Can elucidate underlying modifications that contribute to Intact LC-MS peak shape, particularly combinations of modifications.
- Can be used to quickly evaluate analysis parameters.

P-173-T: Andrew C Nichols

Integrating Both Workflows Into One Data Analysis Pipeline



1. Automated data sweep into search software – *Vendor Agnostic*
2. Extensive Product Variant Search
3. Reduce manual validation time – *Define Criteria to flag false-positives*
4. Automated export of results
5. Aggregate data with molecule information to build ‘In-house’ knowledge base

Summary

1. The need for fast sequence variant analysis in cell line development.
2. Automated analysis workflow reduced sequence variant analysis time.
3. LC-MS based spent media analysis provides complimentary information.

Team

Hirsh Nanda
Andy Mahan
Harsha Gunawardena
Eric Beil
Jeffrey Brelsford
Elsa Gorre

Eric Carlson
Jing Li
Yong J. Kil
Rose Lawler
Andrew Nichols
K.Ilker Sen