

Resolution and Quantification of Protein Isoforms in Complex Genomes

Pittcon 2009

D. Brent Weatherly
University of Georgia
BioInquire, LLC

The Problem

- The genome databases of complex organisms contain many families of related genes
 - Functional redundancy
 - Products of evolution/hybridization
 - Immune evasion (presumed for parasitic organisms)
- Label-free (relative) quantitation depends on accurate intensity measurements of individual proteins
 - Spectral counting
- How to calculate accurate intensity measurements using peptides shared by multiple proteins?

Case Study: Proteomics of *Trypanosoma cruzi*

- About *T. cruzi*
 - Protozoan parasite
 - Causative agent of Chagas Disease
 - 16-18 million infected, 500,000 deaths annually
 - No vaccine
 - 4 distinct life-cycle stages
- The *T. cruzi* genome
 - Reference strain was hybrid of 2 distinct parental strains
 - > 22K annotated genes
 - Alleles of heterozygous genes present (multiple sequences for 1 locus)
 - Multiple copies of members of gene families present (multiple loci)
- *T. cruzi* Proteomics
 - Which genes are functional in which life-cycle stages?

Goal: Create high-confidence, high-coverage quantitative proteome of *T. cruzi* across the 4 life-cycle stages

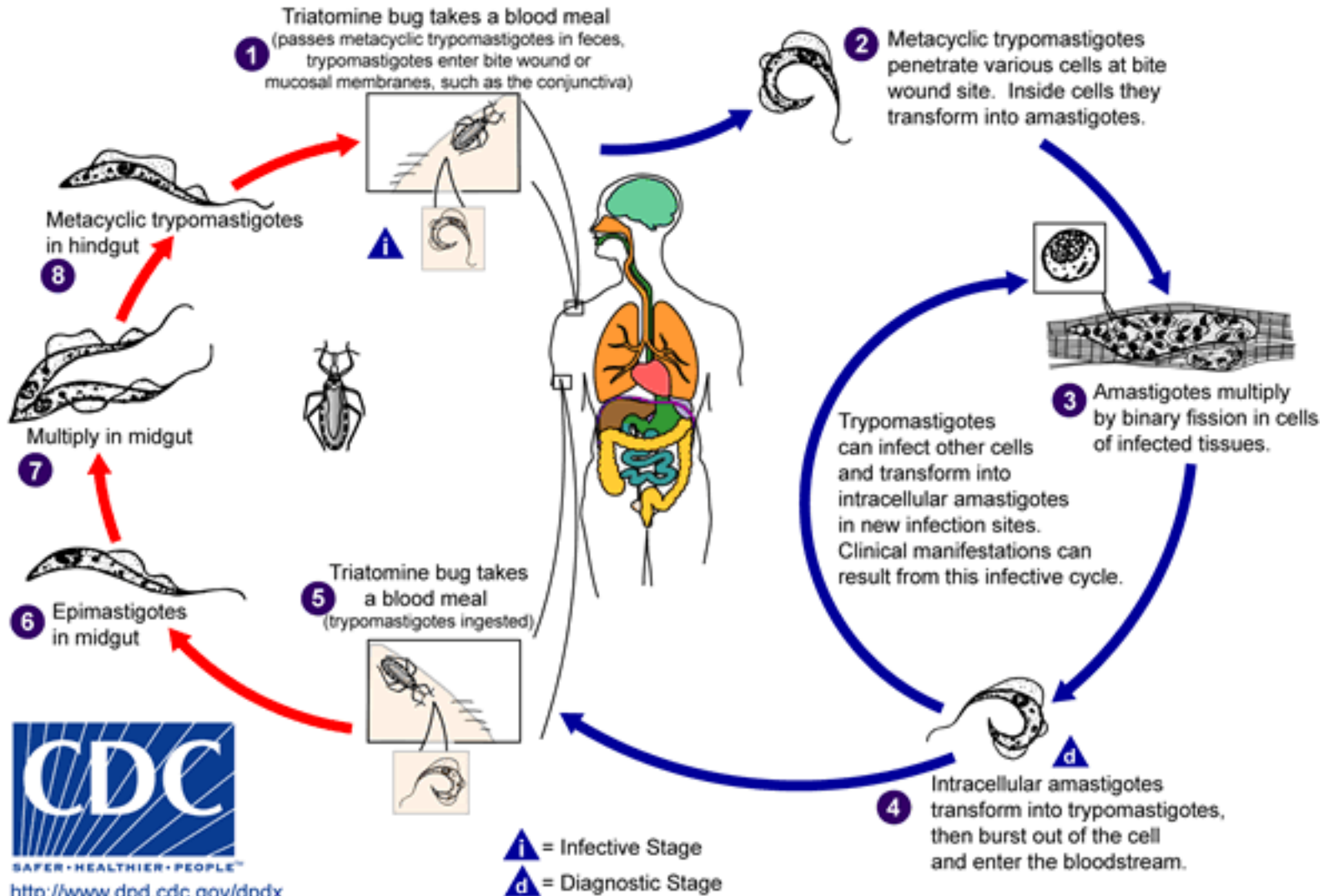
Problem: How can we differentiate alleles and/or gene family members?

T. cruzi – Cycle of Development

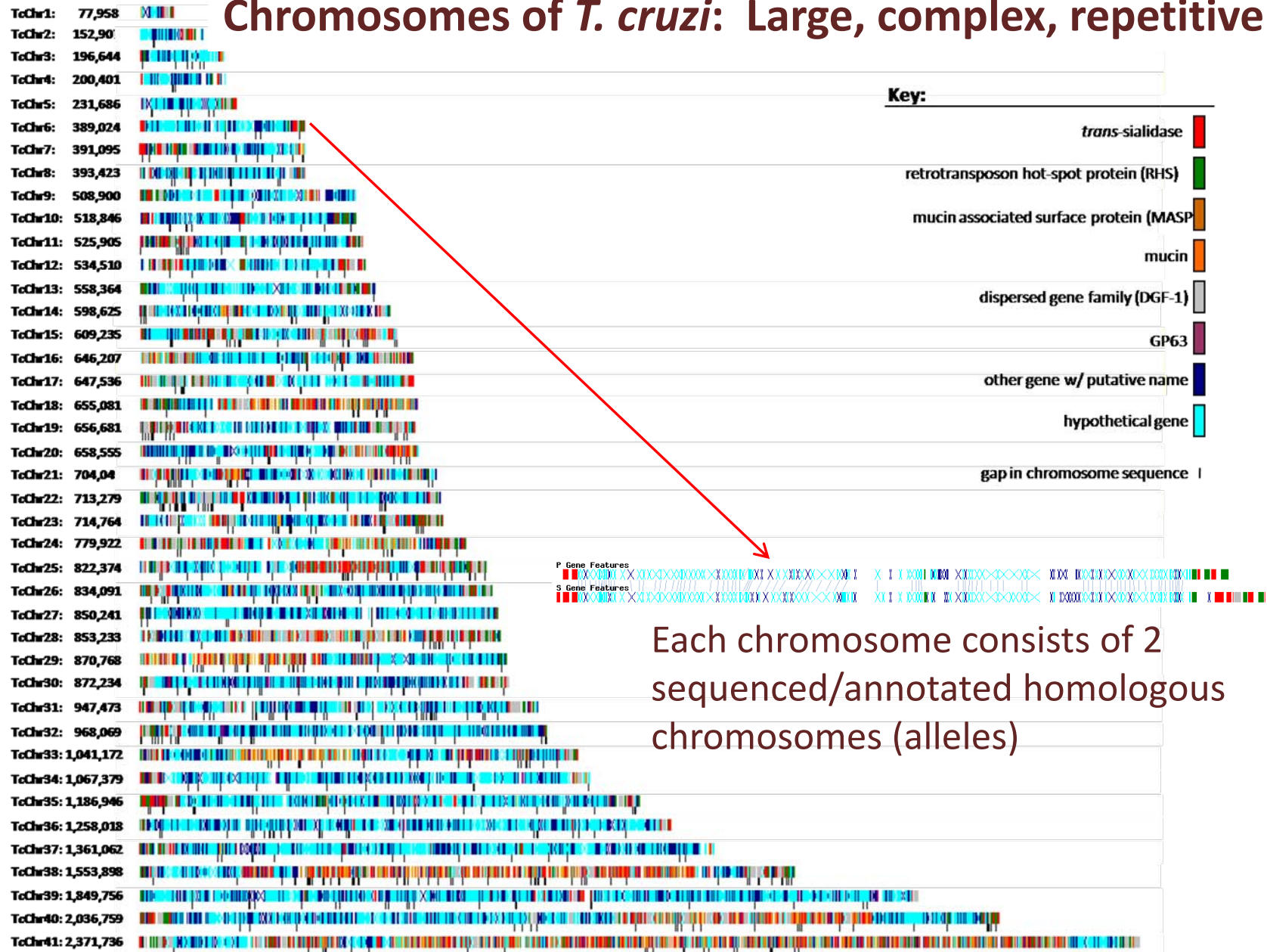


Triatomine Bug Stages

Human Stages

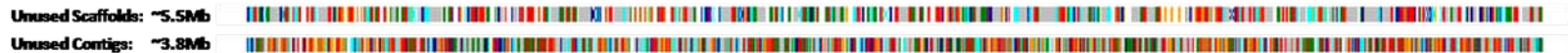


Chromosomes of *T. cruzi*: Large, complex, repetitive



Each chromosome consists of 2 sequenced/annotated homologous chromosomes (alleles)

Artificial Assemblies of Unassignable Contigs



A Proteome of *T. cruzi*: Unresolved Groups of Proteins

Protein Group	Is Top?	Sequence Id	Sequence Name	Protein Length (AA)	Protein Weight (kDa)	Total Group Probability	Total Protein Probability
53	No	Tc00.1047053511261.4	P-type H ⁺ -ATPase, putative [6086.t00017]	257	27.695	1	0
54	Yes	Tc00.1047053508297.7	trans-sialidase, putative [5567.t00008]	447	49.5	1	1
55	Yes	Tc00.1047053506577.80	trans-sialidase, putative [7005.t00008]	906	99.046	0.865	0.886
55	No	Tc00.1047053503769.20	trans-sialidase (pseudogene), putative [4743.t00002]	674	74.709	0.865	0
55	No	Tc00.1047053508355.80	trans-sialidase, putative [7695.t00008]	930	101.816	0.865	0
55	No	Tc00.1047053506173.70	trans-sialidase, putative [6839.t00007]	884	94.266	0.865	0
55	No	Tc00.1047053506667.50	trans-sialidase (pseudogene), putative [7047.t00005]	612	65.813	0.865	0
55	No	Tc00.1047053508313.14	trans-sialidase, putative [7681.t00016]	565	60.862	0.865	0
55	No	Tc00.1047053508761.40	trans-sialidase, putative [7861.t00004]	926	96.442	0.865	0
55	No	Tc00.1047053510715.90	trans-sialidase, putative [8472.t00009]	896	95.338	0.865	0
55	No	Tc00.1047053511173.160	trans-sialidase (pseudogene), putative [8613.t00016]	653	70.773	0.865	0
56	Yes	Tc00.1047053506725.20	hypothetical protein, conserved [7069.t00002]	588	66.413	1	1
56	No	Tc00.1047053507089.170	hypothetical protein, conserved [7204.t00017]	588	66.549	1	0
57	Yes	Tc00.1047053504423.30	hypothetical protein, conserved [6287.t00003]	822	93.646	1	1
57	No	Tc00.1047053507569.10	hypothetical protein, conserved [7388.t00001]	819	93.017	1	0.768
58	Yes	Tc00.1047053503689.4	rab1 small GTP-binding protein, putative [4703.t00005]	220	24.564	1	0.89
58	No	Tc00.1047053509571.79	rab1 small GTP-binding protein, putative [8135.t00008]	220	24.564	1	0.89
59	Yes	Tc00.1047053508045.70	hypothetical protein, conserved [7579.t00007]	458	49.976	1	1
60	Yes	Tc00.1047053503635.59	40S ribosomal protein S3a, putative [4676.t00009]	262	29.878	1	0.686
60	No	Tc00.1047053503635.68	40S ribosomal protein S3a, putative [4676.t00010]	262	29.878	1	0.686
60	No	Tc00.1047053511001.18	40S ribosomal protein S3a, putative [8564.t00025]	262	29.818	1	0.686
60	No	Tc00.1047053511001.9	40S ribosomal protein S3a, putative [8564.t00026]	262	29.818	1	0.686
60	No	Tc00.1047053510999.39	40S ribosomal protein S3a, putative [6029.t00005]	247	28.252	1	0
61	Yes	Tc00.1047053470827.20	trans-sialidase, putative [25155.t00002]	423	46.138	1	1
61	No	Tc00.1047053509875.120	trans-sialidase, putative [8229.t00012]	724	78.51	1	0
62	Yes	Tc00.1047053509601.70	vacuolar proton translocating ATPase subunit A, putative [8148.t00007]	774	87.852	1	0.952
62	No	Tc00.1047053509767.70	vacuolar proton translocating ATPase subunit A, putative [8196.t00007]	774	87.852	1	0.952
63	Yes	Tc00.1047053507757.21	hypothetical protein, conserved [7459.t00008]	378	43.072	0.998	0
63	No	Tc00.1047053506379.10	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [5362.t00001]	378	42.981	0.998	0
63	No	Tc00.1047053508929.20	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [7931.t00002]	377	42.554	0.998	0
63	No	Tc00.1047053424889.10	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [1577.t00001]	382	43.251	0.998	0
63	No	Tc00.1047053503615.10	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [4666.t00001]	378	42.837	0.998	0
63	No	Tc00.1047053509011.10	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase (pseudogene), putative [7963.t00001]	296	33.587	0.998	0
63	No	Tc00.1047053504557.10	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [6330.t00001]	382	42.983	0.998	0
63	No	Tc00.1047053509263.10	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [8053.t00001]	378	43.063	0.998	0
63	No	Tc00.1047053503949.40	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [4833.t00004]	378	42.701	0.998	0
63	No	Tc00.1047053508367.20	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [7701.t00002]	378	42.678	0.998	0
63	No	Tc00.1047053505971.10	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [6758.t00001]	378	42.898	0.998	0
63	No	Tc00.1047053504557.20	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [6330.t00002]	378	42.629	0.998	0
63	No	Tc00.1047053508547.9	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [7773.t00020]	356	39.927	0.998	0
63	No	Tc00.1047053508931.30	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [7932.t00003]	378	42.551	0.998	0
63	No	Tc00.1047053508975.30	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative [7946.t00003]	378	42.802	0.998	0

What is a Protein Group?

- Groups of proteins that share a set of peptides
 - Nesvizhskii et al 2003: **“A statistical model for identifying proteins by tandem mass spectrometry”**
 - Protein Prophet: probabilities of peptides matching to multiple proteins are penalized
 - Occam’s Razor assumption: Protein that could be responsible for all peptides is most likely identification
 - Choose 1 protein to represent the expression of all peptides

Protein Group	Is Top?	Sequence Id	Sequence Name	Protein Length (AA)	Protein Weight (kDa)	Total Group Probability	Total Protein Probability	Total Score	Total Peptides
1	Yes	Tc00.1047053509495.30	trans-sialidase, putative [8114.t00003]	941	99.422	1	1	1,062.07	19
1	No	Tc00.1047053506515.29	trans-sialidase, putative [6980.t00003]	273	29.728	1	0	211.8	3
1	No	Tc00.1047053508913.10	trans-sialidase (pseudogene), putative [7926.t00001]	505	55.297	1	0	211.8	3
1	No	Tc00.1047053510055.20	trans-sialidase, putative [8288.t00002]	203	22.106	1	0	211.8	3
1	No	Tc00.1047053508913.25	trans-sialidase, putative [7926.t00005]	204	22.217	1	0	161.62	2
1	No	Tc00.1047053508501.320	trans-sialidase, putative [7753.t00032]	172	19.261	1	0	130.73	3

1 MGKTVVGASRMFWMFMFVPLLLALCPSEPAHALAPESSRVELFKRQSSKVPFEKDGKVT 60
61 RVVHSFRLPALVNVVDGVMVAIADARYETSDNLSLIDTVAKYSVDDGETWETQIAIKNSRA 120
121 SSVSRVVDPTVIVKGNKLYVLVGSYNSSRSYWTSHGDARDWDILLAVGEVTKSTAGGKTT 180
181 ASIKWGPSVSLKFFPAEMEGMHTNQFLGGAGVAIVASNGNLVYPVQVTNKKKQVFSKIF 240
241 YSEDEGKTWKFGEGRSDFGCSEPVALEWEGKLIINTRVDWTRRLVYESSDMGNTWVEAVG 300
301 TLSRVWGPSPKSDQPGSQSSFTAVTIEGMRVMLFTHPLNFKGRWLRDRLLNLWLDNQRIY 360
361 NVGQVSIKDENSAYSSVLYKDDKLYCLHEINTNEVYSLV FARLVGELRIKSVLQSWKNW 420
421 DSHLSSICTPADPAASSSERGCGPAVTTVGLVGFSLGNASQNVWEDAYRCVNASTANAER 480
481 VPNGLKFAGVGGGALWPVVSQQGQNR YRFANHAFTLVASVTIHEAPRDATPLL GASLDSS 540
541 GGGKLLGLSYDEKHQWQPIYGS TPVTPGSWETGKR YHVLTMSNKIGSVYIDGEPLEGS 600
601 GQTVVPDGRTPDI SHFYVGGYKRSDMPTI SHVTVMNVLLYMRQLNAEEIRTLFLSQDLIG 660
661 TEAHMDSSSDSSAHGTPSIPVDS SAHGTPSTPVDSSAHATPSTPVDSSAHSTPSTPVDSS 720
721 AHGTPSTPVDSSAHGTPSTTTVDSSAHGTPSTPVDSSAHGTPSTTTVDSSAHATPSTPVDSS 780
781 AHSTPSTPVDSSAHGTPSTPVDSSAHGTPSTPVDSSAHGTPSTPVDSSAHGTPSTPVDSS 840
841 AHGTPSTPVDSSAHGKPSVVDSSAHGTPSTPVDSSAHSTPSTPADSSAHSTPSTPAGNG 900
901 ANGTVLILPDGAALSAPSGGGLLLCACALLLHVFFTAVFF* 960

Occam's Razor representative:
Tc00.1047053509495.30

**Protein Group Probability vs
Protein Probability**

Expression evidence for these
proteins is non-existent

1 MGRTVVGASRMFWMTCFVPLLLALCPNEPAHALAPGSSRVELFKRQNSTVPPFEENGEVVRQ 60
61 RVVHSFRLPALVNVVDGVMVAIADARYDTSNDNSLIDTVVKYSVDDGETWETQIAIKNSRA 120
121 SSVSRVVDPTVIVKGNKIYVLVGGYNSSRSYWTSHGDASDWDILLAVGEVKKCTVDDKTA 180
181 ASIKWGPSVSLKEFFPAEMEGMQTNQFLGGAGVAIVASNGNLVYPVQVTNKKKQVFSKIF 240
241 YSEDEGKTWKFGEGRSAPGCSSEPVALEWEGKLIINTRVDWTRRLVYESGGMGNTWVEAVG 300
301 TLSRVWGPSASPANMPGSSFTAVTIEGMRVMLFTHPLNFKGWMLRDRLLQLWLDNQRIY 360
361 NVGQLS IGDENSAYSSVLYKDDKLYCLHEINSNEVYSLV FARLVGELRLKSVVQSWKKW 420
421 DSHLSSNCTPADSAASSSERGCGPAVTTVGLVGFSLSHSANKTKWKDAYRCVDASTANAER 480
481 VSNGLKFAGVGGGALWPVVSQQGQN* 540

1 MGRTVVGASRMFWMTHFVPLLLALRPNEPAHALAPGSSRVELFKRQNSTVPPFEDKGVKVRQ 60
61 RVVHSFRLPALVNVVDGVMVAIADARYDTSNDNSLIDTVVKYSVDDGETWETQIAIKNSRV 120
121 SSASRVVDPTVIVKGNKIYVLVGSYNSSRSYWTSHGDASDWDILLAVGEVKKCTVDDKTA 180
181 ASIKWGPSVSLNEFFPAEMEGMQANQFLGGAGVAIVASNGNLVYPVQVTNKKKQVFSKIF 240
241 YSEDEGKTWKFGEGRSAPGCSSEPVALS GREAH* 300

Protein Group	Is Top?	Sequence Id	Sequence Name	Protein Length (AA)	Protein Weight (kDa)	Total Group Probability	Total Protein Probability	Total Score	Total Peptides
1	Yes	Tc00.1047053509495.30	trans-sialidase, putative [8114.t00003]	941	99.422	1	1	1,062.07	19
1	No	Tc00.1047053506515.29	trans-sialidase, putative [6980.t00003]	273	29.728	1	0	211.8	3
1	No	Tc00.1047053510055.20	trans-sialidase, putative [8288.t00002]	203	22.106	1	0	211.8	3
1	No	Tc00.1047053508913.25	trans-sialidase, putative [7926.t00005]	204	22.217	1	0	161.62	2
1	No	Tc00.1047053508501.320	trans-sialidase, putative [7753.t00032]	172	19.261	1	0	130.73	3
2	Yes	Tc00.1047053508913.10	trans-sialidase (pseudogene), putative [7926.t00001]	505	55.297	0.991	0.991	252.28	4

```

1   MGKTIVVGASRMFWLMPFVPLLLALCPSEPAHALAPESSRVELFKRQSSKVPFEKDGKVT 60
61  RVVHSFRLPALVNVVDGVMVAIADARYETSDNLSLIDTVAKYSVDDGETWETQIAIKNSRA 120
121 SSVSRVVDPTVIVKGNKLYVLVGSYNSSRSYWTSHGDARDWDILLAVGEVTKSTAGGKTT 180
181 ASIKWGPSVSLKKEFFPAEMEGMHTNQFLGGAGVAIVASNGNLVYPVQVTNKKKQVFSKIF 240
241 YSEDEGKTKWFKGGRSDFGCSEPVALEWEGKLIINTRVDWTRRLVYESSDMGHTWVEAVG 300
301 TLSRVWGPSPKSDQPGSQSSFTAVTIEGMRVMLFTHPLNFKGRWLRDRLNLWLTDNQRIY 360
361 NVGQVSIKDENSAYS SVLYKDDKLYCLHEINTNEVYSLV FARLVGELRIIKSVLQSWKNW 420
421 DSHLSSICTPADPAASSSERGCGPAVTTVGLVGLSGNASQNVWEDAYRCVNASTANAER 480
481 VPNGLKFAGVGGGALWVPSQQGQNR YRFANHAFTLVASVTIHEAPRDATPLL GASLDSS 540
541 GGGKLLGLSYDEKHQWQP IYGSTPVTPTGSWETGKRYHVVL TMSNKIGSVYIDGPELEGS 600
601 GQTVVPDGRTPDI SHFYVGGYKRSDMPTI SHVTVNNVLL YNRQLNAE IRTLFLSODLIG 660
661 TEAHMDDSSDSSAHGTPSIPVDSSAHGTPSTPVDSSAHATPSTPVDSSAHSTPSTPVDSS 720
721 AHGTPSTPVDSSAHGTPSTTVVDSSAHGTPSTPVDSSAHGTPSTTVVDSSAHATPSTPVDSS 780
781 AHSTPSTPVDSSAHGTPSTPVDSSAHGTPSTPVDSSAHGTPSTPVDSSAHGTPSTPVDSS 840
841 AHGTPSTPVDSSAHGKPSIPVDSSAHGTPSTPVDSSAHSTPSTPADSSAHSTPSTPAGNG 900
901 ANGTVLILPDGAALSAPSGGGLLLCACALLLHVFFTAFFF*

```

Identification of novel peptide to Tc00.1047053508913.10 resolves its expression from the group Id'd: LQLWLTDNQR

Limiting factor for Occam's Razor assumption is the degree of proteomic sampling

Tc00.1047053506515.29 still unresolved.

Note: Proteins have different lengths and weights

```

1   MGRTIVVGASRMFWLTCFVPLLLALCPNEPAHALAPGSSRVELFKRQNSTVPPFEENGEVRQ 60
61  RVVHSFRLPALVNVVDGVMVAIADARYDTSNDNSLIDTVVKYSVDDGETWETQIAIKNSRA 120
121 SSVSRVVDPTVIVKGNKIYVLVGGYNSRSYWTSHGDASDWDILLAVGEVKKCTVDDKTA 180
181 ASIKWGPSVSLKEFFPAEMEGMQTNQFLGGAGVAIVASNGNLVYPVQVTNKKKQVFSKIF 240
241 YSEDEGKTKWFKGGRSAFGCSEPVALEWEGKLIINTRVDWTRRLVYESCGMHTWVEAVG 300
301 TLSRVWGPSASNPMPGSSQSSFTAVTIEGMRVMLFTHPLNFKGWWLRDRLQLWLTDNQR IY 360
361 NVGQLS IGDENSAYS SVLYKDDKLYCLHEINSNEVYSLV FARLVGELRLIKSVYQSWKKW 420
421 DSHLSSNCTPADSAASSSERGCGPAVTTVGLVGLSHSANKTKWKDAYRCVDASTANAER 480
481 VSNGLKFAGVGGGALWVPSQQGQN*

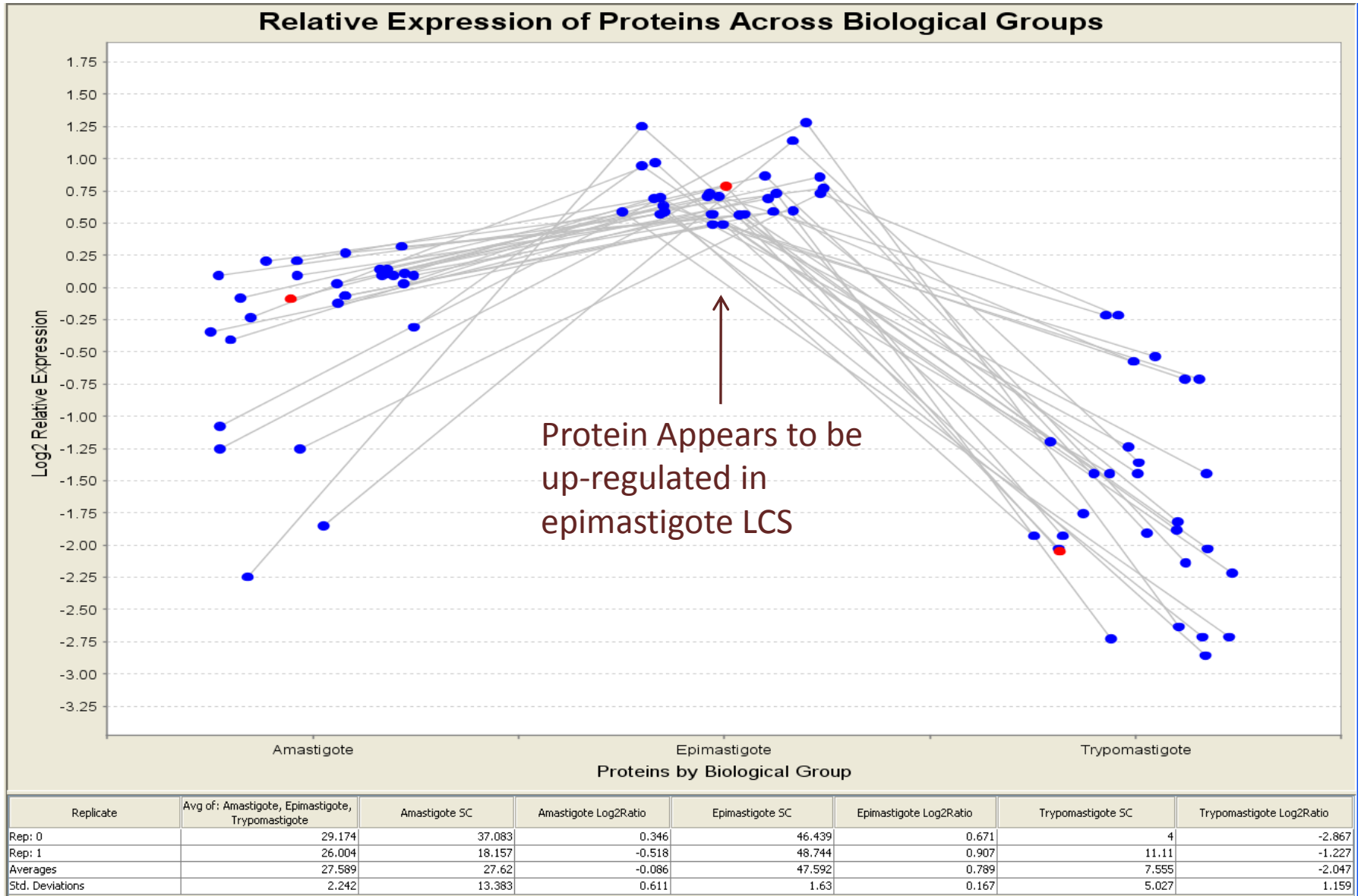
```

```

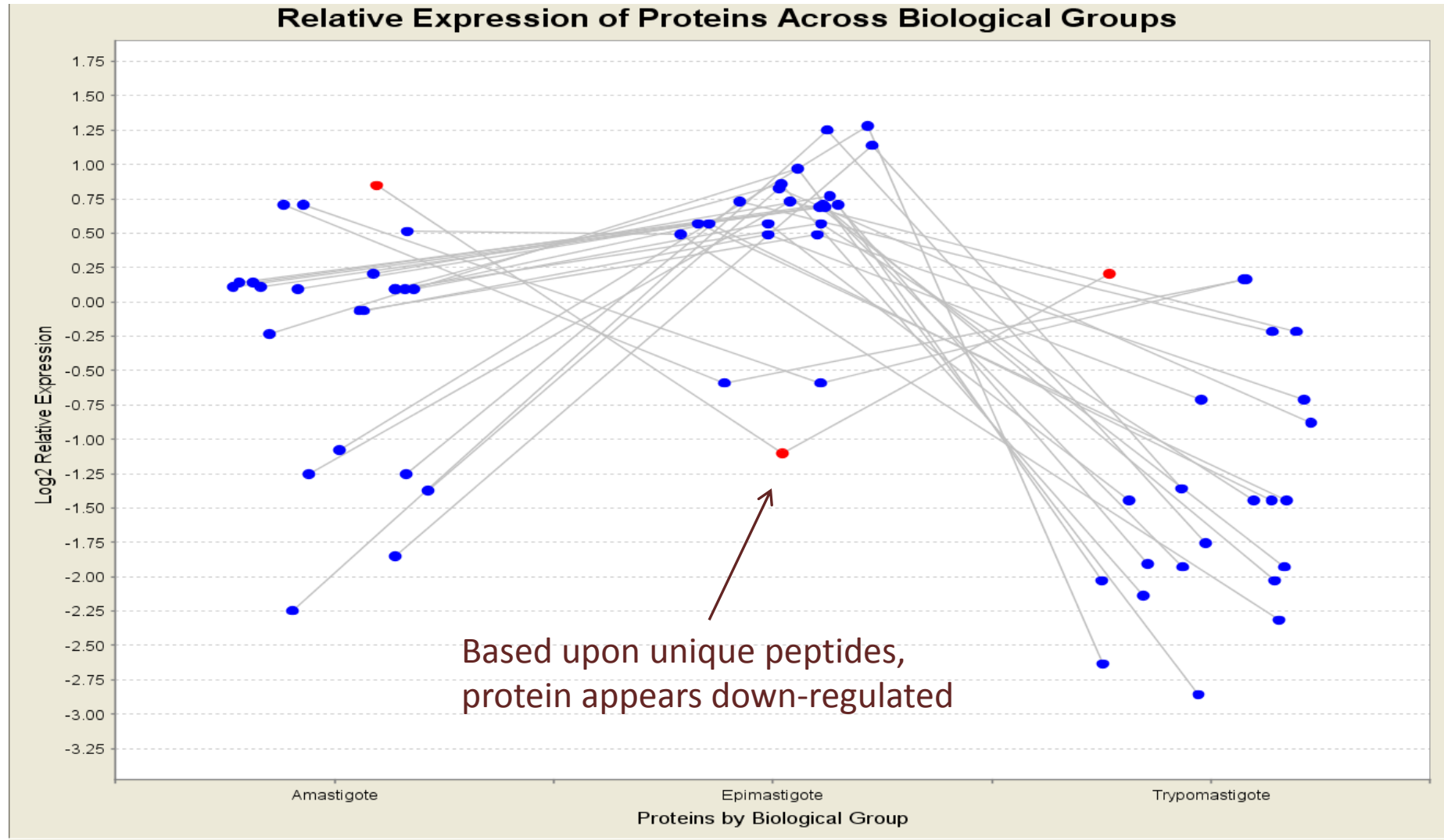
1   MGRTIVVGASRMFWLTHFVPLLLALRPNEPAHALAPGSSRVELFKRQNSTVPPFEDKGVKVRQ 60
61  RVVHSFRLPALVNVVDGVMVAIADARYDTSNDNSLIDTVVKYSVDDGETWETQIAIKNSRV 120
121 SSASRVVDPTVIVKGNKIYVLVGSYNSSRSYWTSHGDASDWDILLAVGEVKKCTVDDKTA 180
181 ASIKWGPSVSLNEFFPAEMEGMQANQFLGGAGVAIVASNGNLVYPVQVTNKKKQVFSKIF 240
241 YSEDEGKTKWFKGGRSAFGCSEPVALS GREAH*

```

Why are Protein Groups a Problem?



Remove Common Peptides: Quantitation Affected



Replicate	Avg of: Amastigote, Epimastigote, Trypomastigote	Amastigote SC	Amastigote Log2Ratio	Epimastigote SC	Epimastigote Log2Ratio	Trypomastigote SC	Trypomastigote Log2Ratio
Rep: 0	3.151	8.374	1.41	1.08	-1.545	0	0
Rep: 1	3.509	4.272	0.284	2.216	-0.663	4.04	0.203
Averages	3.33	6.323	0.847	1.648	-1.104	2.02	0.102
Std. Deviations	0.253	2.9	0.796	0.803	0.623	2.857	0.144

Protein Group	Is Top?	Sequence Id	Sequence Name	Protein Length (AA)	Protein Weight (kDa)	Total Group Probability	Total Protein Probability	Total Score	Total Peptides
1	Yes	Tc00.1047053509495.30	trans-sialidase, putative [8114.t00003]	941	99.422	1	1	1,062.07	19
1	No	Tc00.1047053506515.29	trans-sialidase, putative [6980.t00003]	273	29.728	1	0	211.8	3
1	No	Tc00.1047053510055.20	trans-sialidase, putative [8288.t00002]	203	22.106	1	0	211.8	3
1	No	Tc00.1047053508913.25	trans-sialidase, putative [7926.t00005]	204	22.217	1	0	161.62	2
1	No	Tc00.1047053508501.320	trans-sialidase, putative [7753.t00032]	172	19.261	1	0	130.73	3
2	Yes	Tc00.1047053508913.10	trans-sialidase (pseudogene), putative [7926.t00001]	505	55.297	0.991	0.991	252.28	4

```

1  MGKTVVVGASRMFWMFVPLLLALCPSEPAHALAPESSRVELFKRQSSKVPFEKDGKVT 60
61  RVVHSFRLPALVNVVDGMVAIADARYETSDNSLIDTVAKYSVDDGETWETQIAIKNSRA 120
121  SSVSRVVDPTVIVKGNKIYVLVGSYNSRSYWTSHGDARDWDILLAVGEVTKTAGGKTT 180
181  ASIKWGPSVSLKFFPAEMEGMHTNQFLGGAGVAIVASNGMLVYPVQVTKKKQVFSKIF 240
241  YSEDEGKTWKFGEGRSDFGCSEPVVALEWEGKLIINTRVDWTRRLVYESSDMGNTWVEAVG 300
301  TLSRVWGPSFKSDQPGSQSSFTAVTIEGMRVMLFTHPLNFKGRWLRDRLLNLWTDNQRIY 360
361  NVGQVSIQDENSAYS SVLYKDDKLYCLHEINTNEVYSLV FARLVGELRLIKSVLQSWKNW 420
421  DSHLSSICTPADPAASSSERGCGPAVTTVGLVGF LSGNASQNVWEDAYRCVMASTANAER 480
481  VFNGLKFAVGGGALWVPSVQQGNQRYRPFANHAF TLVASVTIHEAPRDATPLL GASLDS 540
541  GKKLLGLSYDEKHQWQPIYGSTPVP TPGSWETGKRYHVVL TMSNKIGSVYIDGPELGS 600
601  GQTVVVDGRTPDISHFYVGGYKRSMP TI SHVTVNNVLLYNRQLNAEIRITFLS QDLIG 660
661  TEAHMSSSDSSAHGTPSIPVDSSAHGTPSPV DSSAHATPSTPVDSSAHSTPSTPVDSS 720
721  AHGTPSTPVDSSAHGTPSTPVDSSAHGTPSTPVDSSAHGTPSTPVDSSAHATPSTPVDSS 780
781  AHSTPSTPVDSSAHGTPSTPVDSSAHGTPSTPVDSSAHGTPSTPVDSSAHGTPSTPVDSS 840
841  AHGTPSTPVDSSAHGKPSIPVDSSAHGTPSTPVDSSAHSTPSTPADSSAHSTPSTPAGNG 900
901  ANGTVLILPDGAALSAFSGGGLLACACALLHVF FTAVFF* 960

```

```

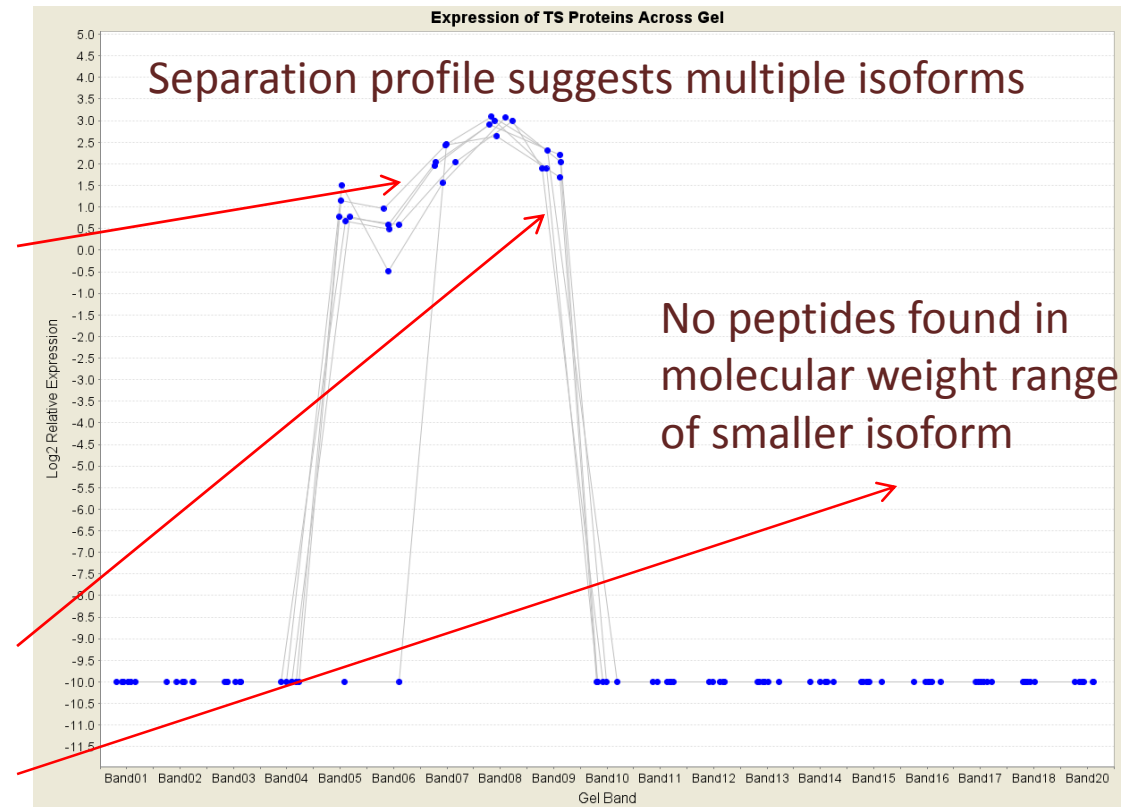
1  MGRTIVVGASRMFMLTCFVPLLLALCPNEPAHALAPGSSRVELFKRQNSTVFPFENGVEVRQ 60
61  RVVHSFRLPALVNVVDGMVAIADAR YDTSNDNSLIDTVVKYSVDDGETWETQIAIKNSRA 120
121  SSVSRVVDPTVIVKGNKIYVLVGSYNSRSYWTSHGDASDWDILLAVGEVKKCTVDDKTA 180
181  ASIKWGPSVSLKEFFPAEMEGMHTNQFLGGAGVAIVASNGMLVYPVQVTKKKQVFSKIF 240
241  YSEDEGKTWKFGEGRSAF GCSEPVVALEWEGKLIINTRVDWTRRLVYESGGMGNTWVEAVG 300
301  TLSRVWGPSASNMFGSQSSFTAVTIEGMRVMLFTHPLNFKGMWLRDRLLQLWLDNQRIY 360
361  NVGQLSIQDENSAYS SVLYKDDKLYCLHEINSNEVYSLV FARLVGELRLIKSVVQSWKKNW 420
421  DSHLSSNCTPADSAASSSERGCGPAVTTVGLVGF LSHSANKTKMKDAYRCVDASTANAER 480
481  VSNGLKFAVGGGALWVPSVQQGN* 540

```

```

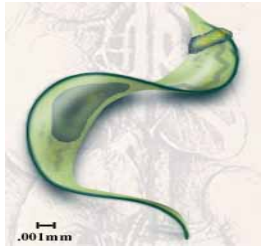
1  MGRTIVVGASRMFMLTHFVPLLLALRPNEPAHALAPGSSRVELFKRQNSTVFPFEDKGVKVRQ 60
61  RVVHSFRLPALVNVVDGMVAIADAR YDTSNDNSLIDTVVKYSVDDGETWETQIAIKNSRV 120
121  SSVSRVVDPTVIVKGNKIYVLVGSYNSRSYWTSHGDASDWDILLAVGEVKKCTVDDKTA 180
181  ASIKWGPSVSLNEFFPAEMEGMQANQFLGGAGVAIVASNGMLVYPVQVTKKKQVFSKIF 240
241  YSEDEGKTWKFGEGRSAF GCSEPVVALEWEGKLIINTRVDWTRRLVYESGGMGNTWVEAVG 300

```

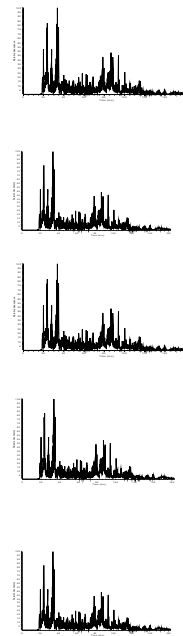
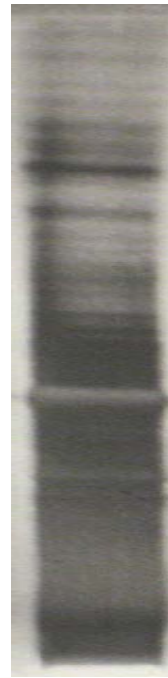


Gel-C Approach: Resolve Expression by Peptide Identifications to Proteins by Molecular Weight?

Quantitative, GelC Proteomics Approach

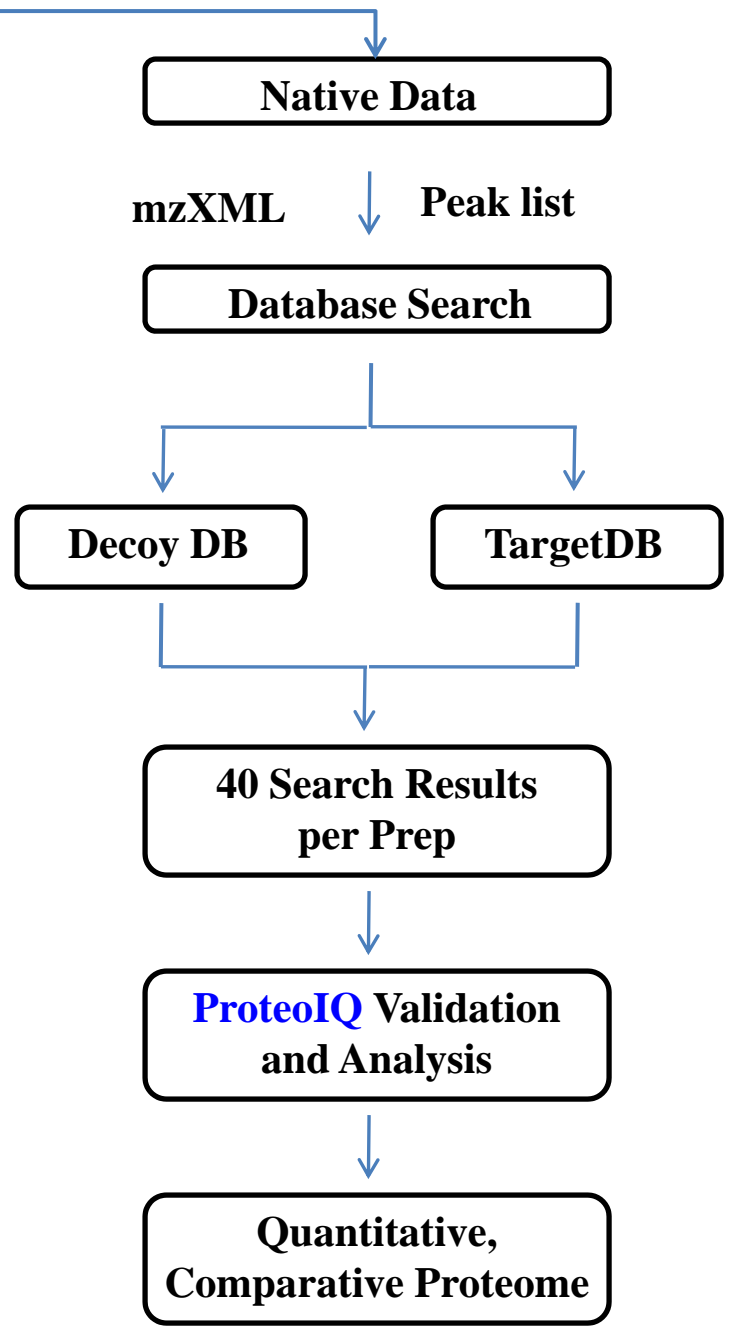


Preparation
X 3

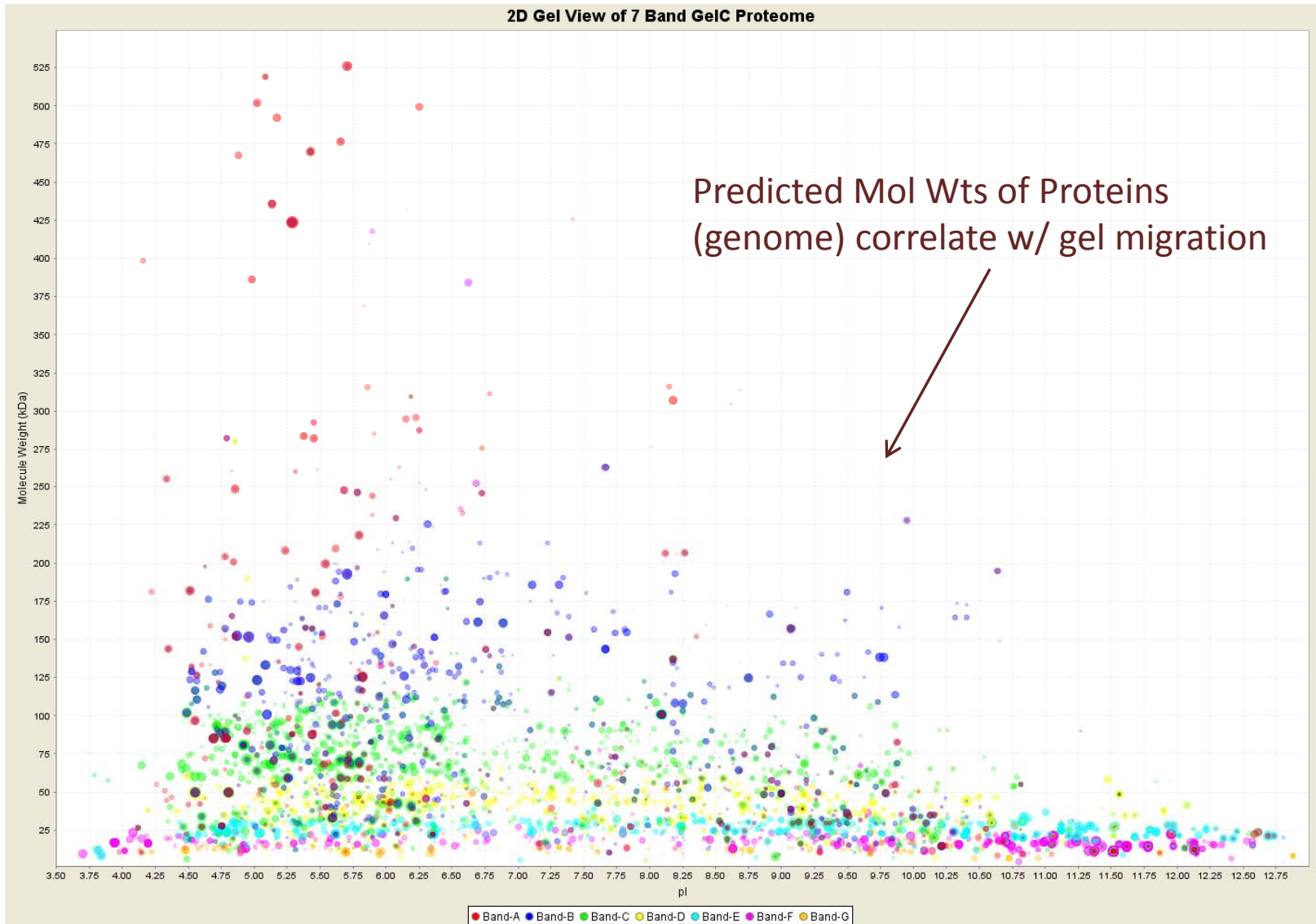


**20 Gel Bands
Per Replicate**

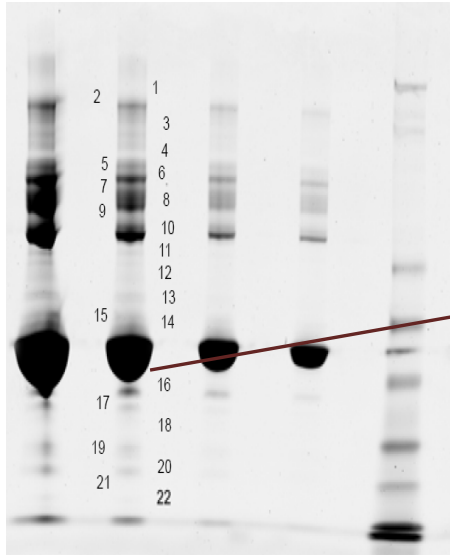
**20 LC-MS/MS
Per Replicate**



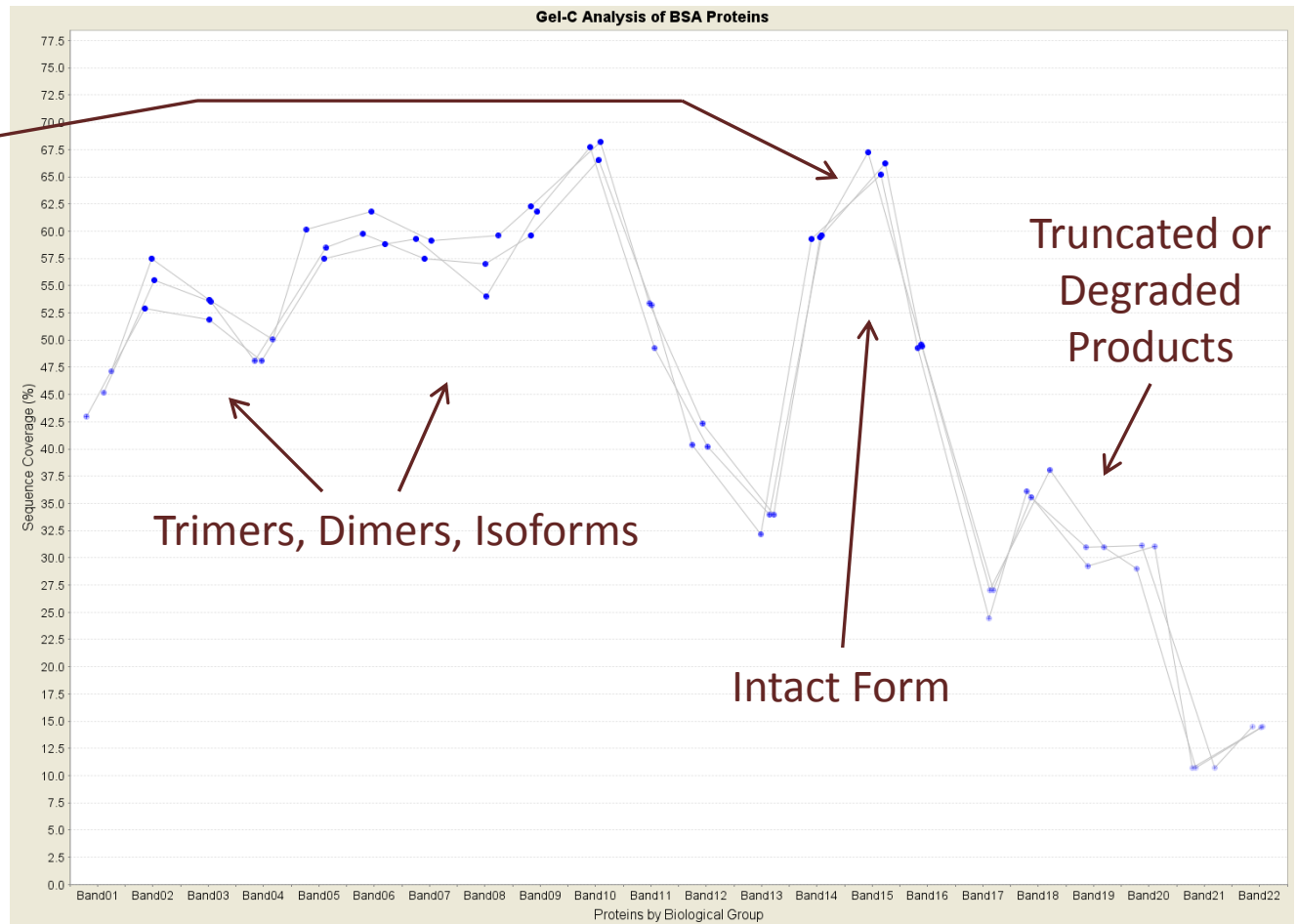
2D Gel View to Assess GelC Proteome Quality



BSA: Method Validation

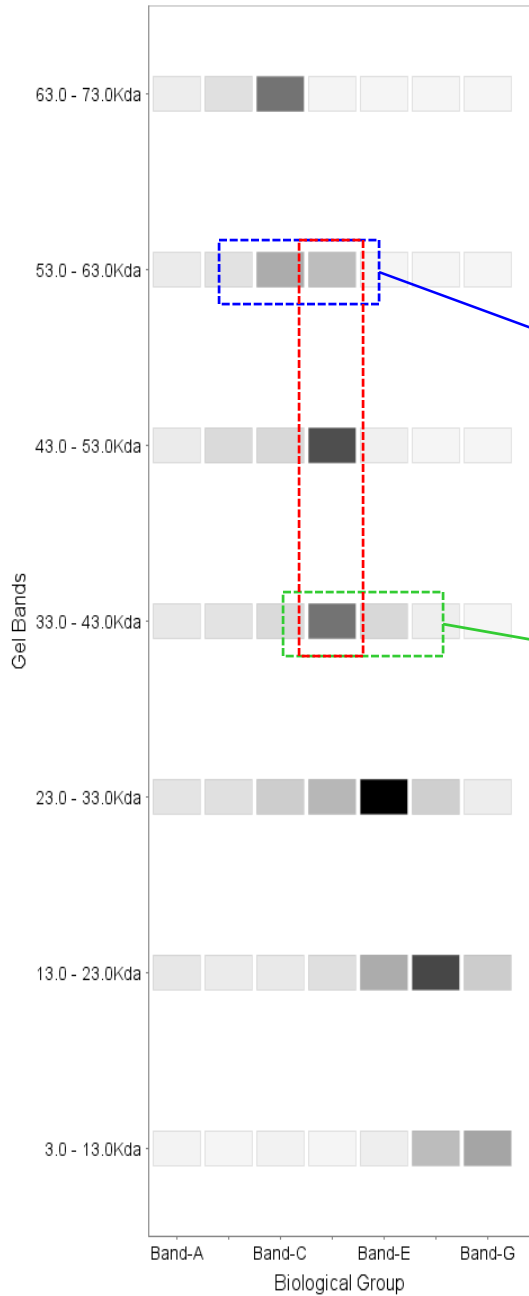


Protein Group	Is Top?	Sequence Id	Sequence Name	Protein Length (AA)	Protein Weight (kDa)	Total Group Probability	Total Protein Probability
1	Yes	UniRef100_B0JYQ0	ALB protein n=1 Tax=Bos taurus RepID=B0JYQ0_BOVIN	607	69.23	1	1
2	Yes	UniRef100_P02769	Serum albumin n=1 Tax=Bos taurus RepID=ALBU_BOVIN	607	69.23	1	1
3	Yes	UniRef100_LUPI000179EC85	Serum albumin precursor (Allergen Bos d 6) (BSA). n=1 Tax=Bos taurus RepID=LUPI000179EC85	609	69.45	1	1



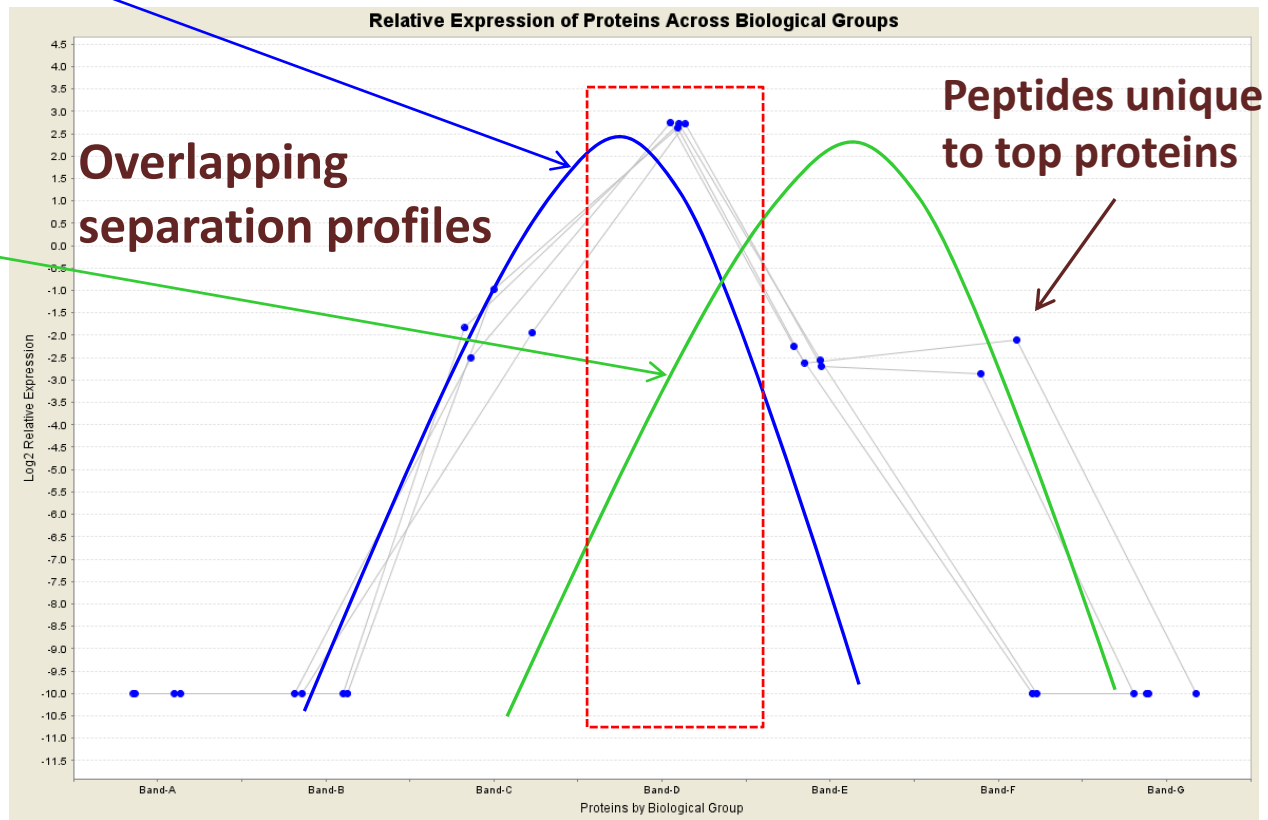
GelC Approach is sensitive enough to resolve isoforms

1D Gel View of Protein Expression



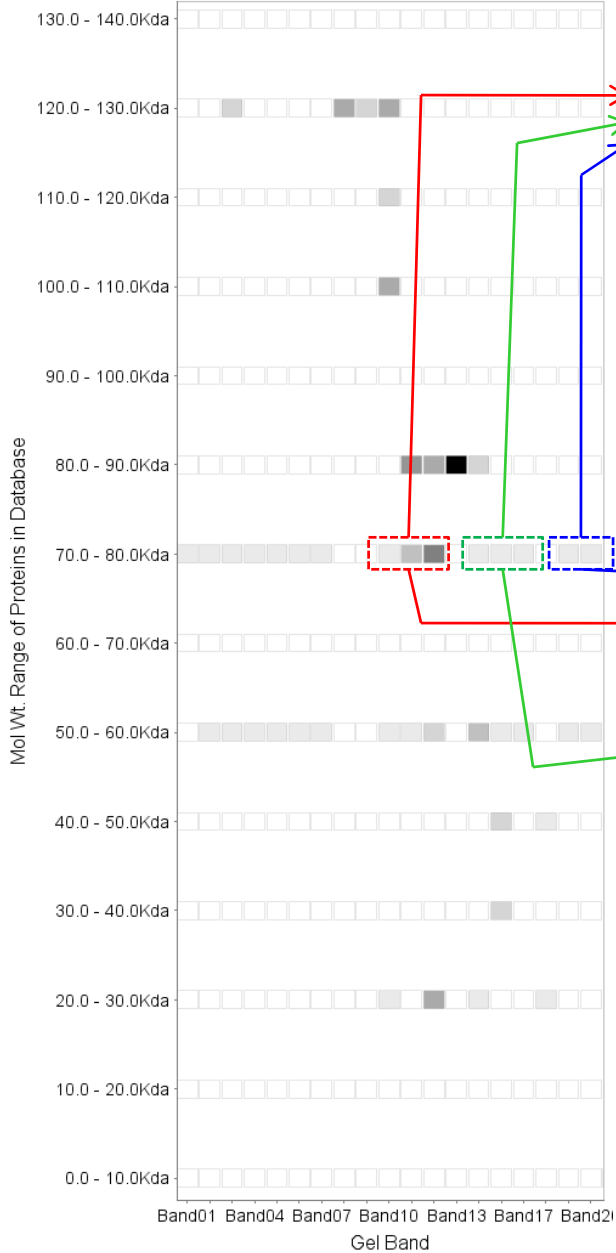
Resolution of 2 similar pairs (alleles) of hypothetical genes

Protein Group	Is Top?	Sequence Id	Sequence Name	Protein Length (AA)	Protein Weight (kDa)	Total Group Probability	Total Protein Probability	Total Score	Total Peptides
1	Yes	Tc00.1047053506859.180	hypothetical protein, conserved [7120.t00018]	372	42.77	1	1	726.87	14
1	No	Tc00.1047053511815.180	hypothetical protein, conserved [6207.t00018]	372	42.778	1	0.973	660.41	13
1	No	Tc00.1047053511815.170	hypothetical protein, conserved [6207.t00017]	537	60.517	1	0.939	611.76	12
1	No	Tc00.1047053506859.170	hypothetical protein, conserved [7120.t00017]	537	60.409	1	0.892	531.34	11



Resolution of 3 ATPase proteins

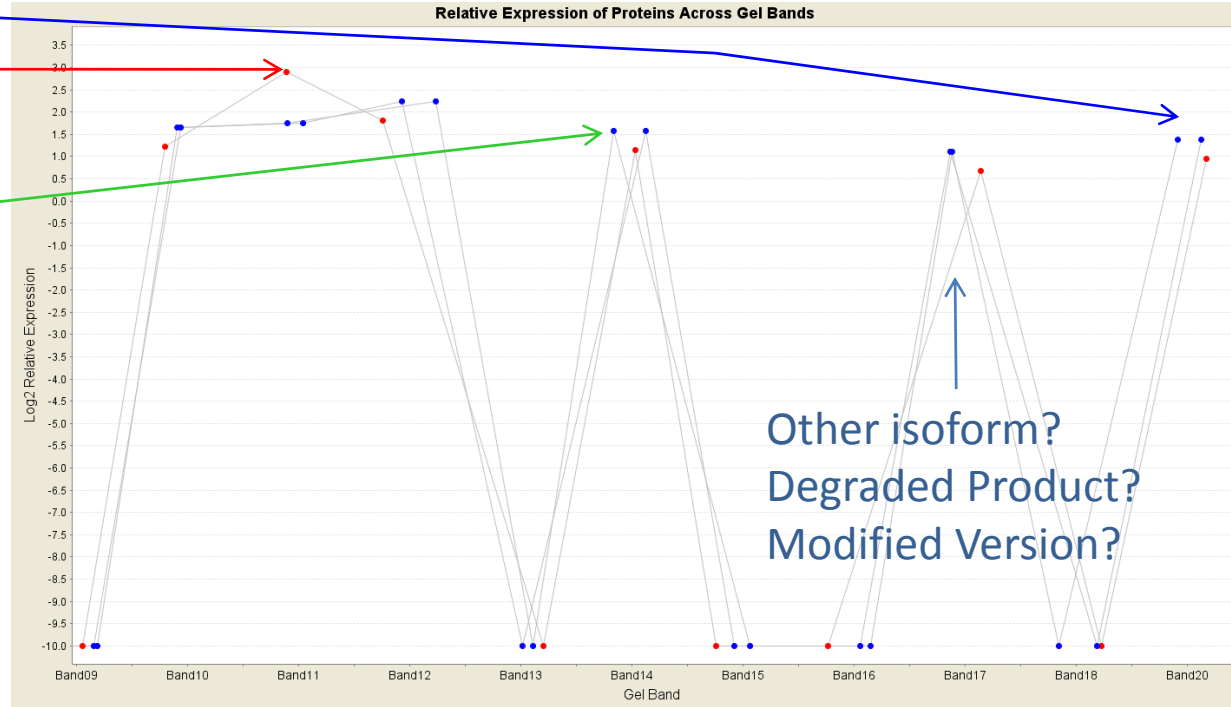
1D Gel View - ATPase



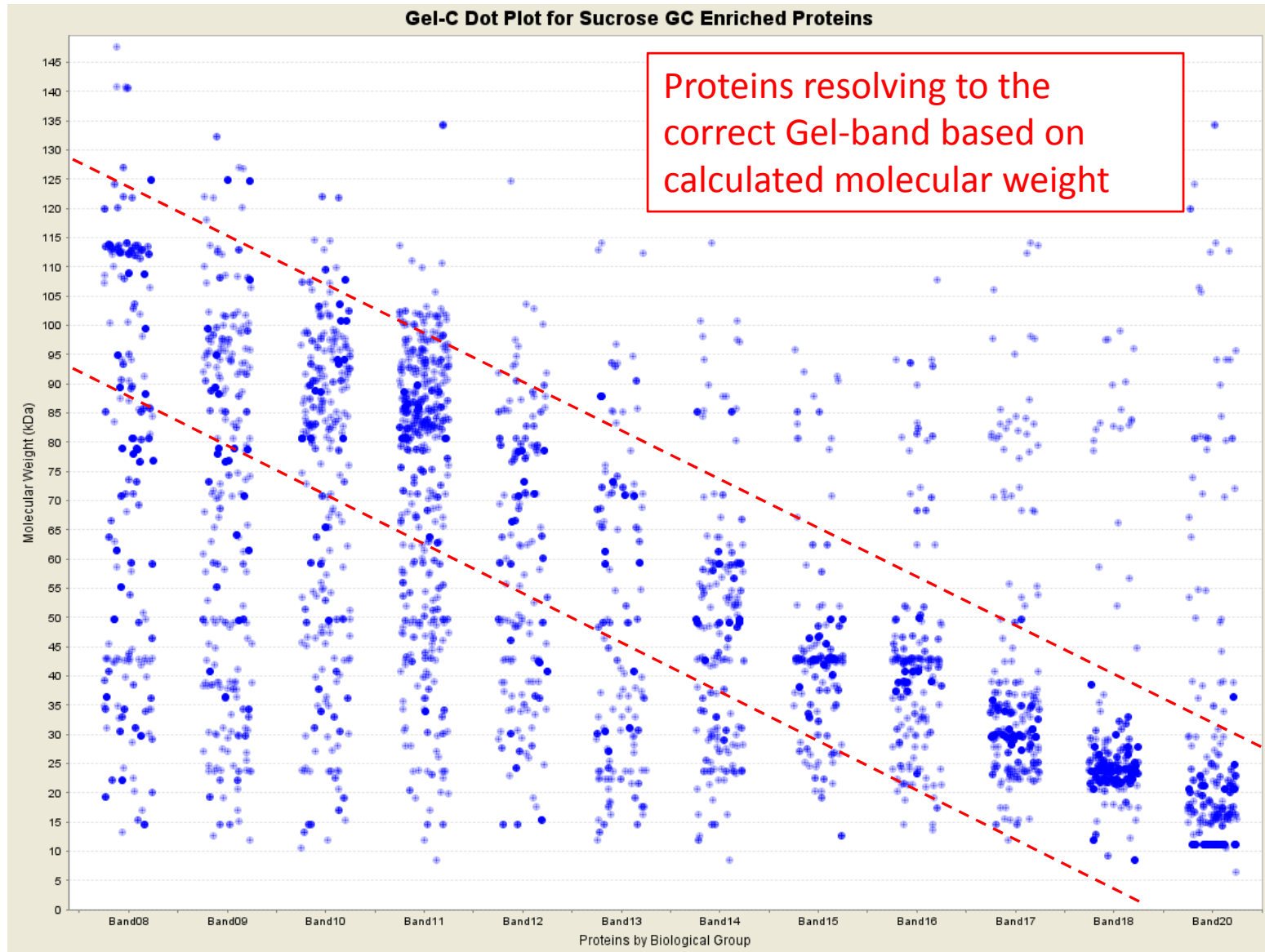
Protein Group	Is Top?	Sequence Id	Sequence Name	Protein Length (AA)	Protein Weight (kDa)	Total Group Probability	Total Protein Probability	Total Score
1	Yes	Tc00.1047053505763.19	P-type H ⁺ -ATPase, putative [6697.t00002]	646	70.686	1	1	226.36
1	No	Tc00.1047053506333.10	plasma-membrane proton-efflux P-type ATPase, putative [6901.t00001]	485	53.419	1	0	130.01
1	No	Tc00.1047053509809.39	P-type H ⁺ -ATPase, putative [5788.t00004]	260	28.007	1	0	130.01

```

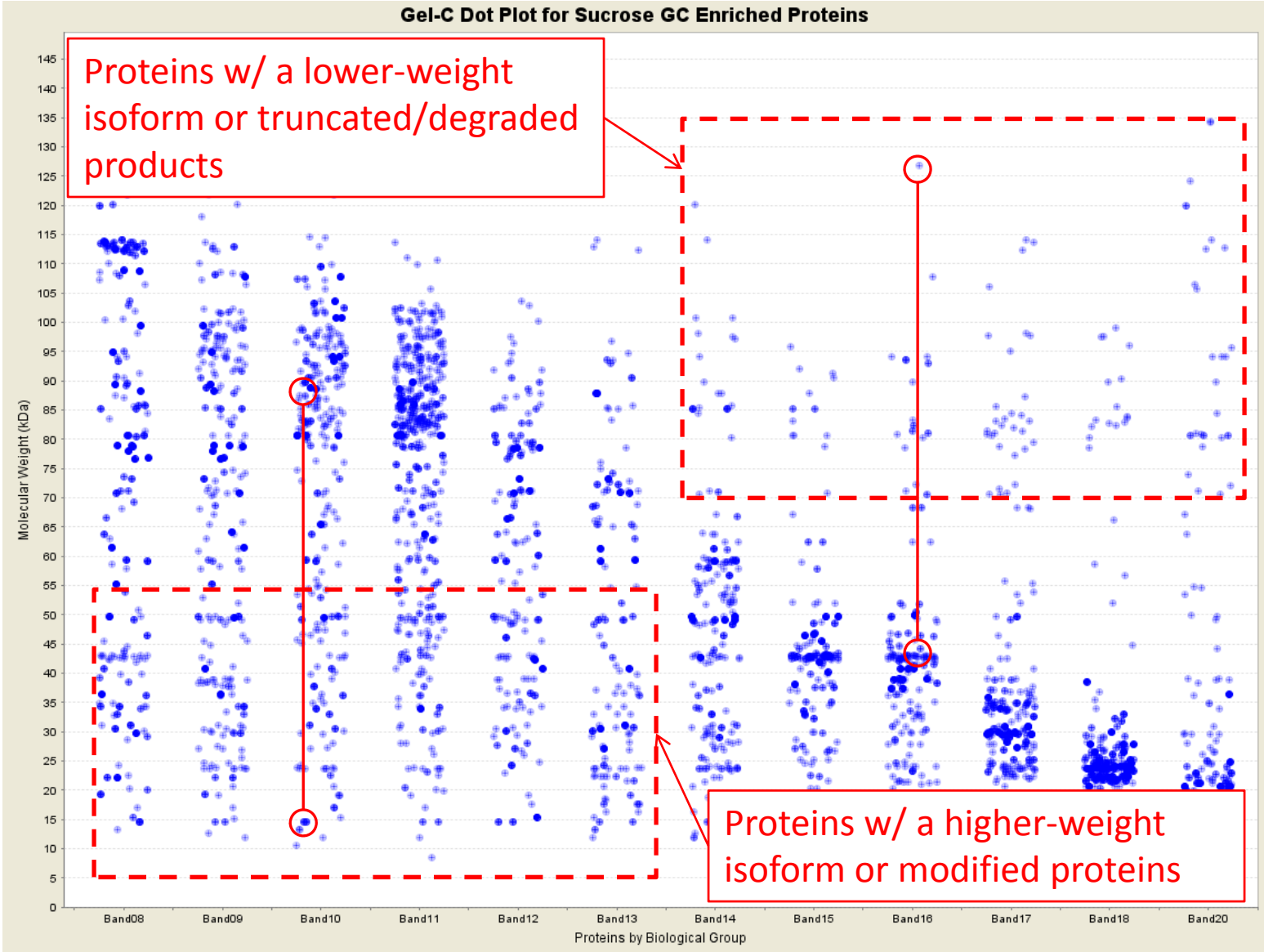
1  MLPPSKGLTTAEAEELLAKYGRNELPEKKTSPSMLIFVRNLWGMPFALWVAIIIEFALEN 60
61  WPDGAILLVIQLANATIGWYETIKAGDAVAALKNSLKPVATVHRDGAWQQLDAALLVPGD 120
121  LVKLAGSVAVPADCSINEGVIDVDEAALTGESLPTMGTDHMPKMGSNVVRGEVDGTVQY 180
181  TGQNTFFGKTAVLLQSVESDLGNIHVLSRVVMVLTSPFSLCLICFIYLMKFYETFR 240
241  SLQFSVVVLVYSIPIALEIVVTTTLAGVSKLSRHKIIVTKLTAIEMMSGVNMLCSDKTG 300
301  TLTLNKMEIQDQCPTFEKGYDLRSVLAALAAKMRPPRDALDITMVLGAADLDECDNYT 360
361  QTEFVFPDPTTKRTAATLVDRKRTNEKFSVTKGAPHVILQLVYNQDEINDQVVEIIDS LAA 420
421  RGVRCLSVAKTDSQGRMHLGILTFLDPPRPDTKETIRRSKQYGVVKMITGDHVLIAKE 480
481  MCRMLNLDPNILTADKLPKVDVNDLPPDLGKEYGEMMLGVGGFAQVFPPEHKFMIVEALRQ 540
541  YGFTCAMTGDGVNDAPALKRADVGIADVQCATDARAADAADMVLTPGLSVVVEAMLVSRQV 600
601  FQCMLSFLTYRISATLQLVCFPIACPSLTPRNYGSDADPQFFH* 660
    
```



Expanded 1D Gel to ID Proteins of Interest



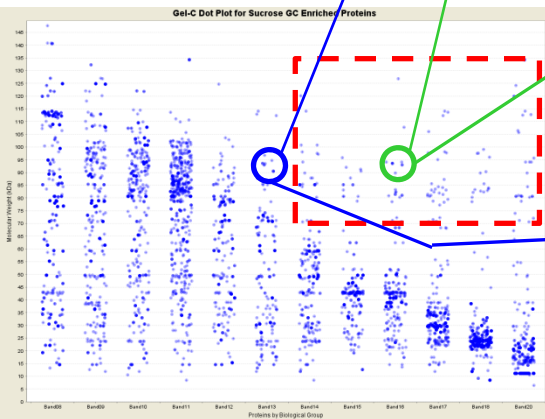
Expanded 1D Gel to ID Proteins of Interest



Unresolved hypotheticals. Degraded product? Or sequencing error?

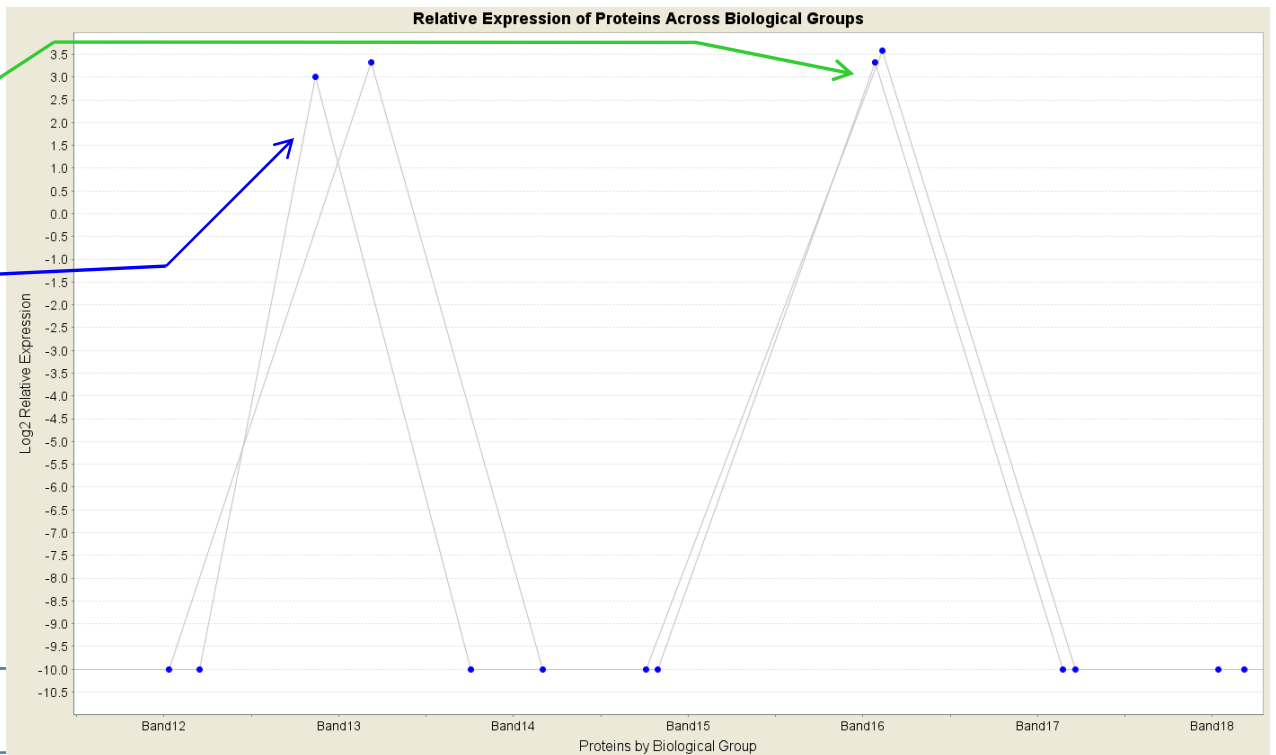
Protein Group	Is Top?	Sequence Id	Sequence Name	Protein Length (AA)	Protein Weight (kDa)	Total Group Probability	Total Protein Probability
1	Yes	Tc00.1047053504423.30	hypothetical protein, conserved [6287.t00003]	822	93.646	1	1
1	No	Tc00.1047053507569.10	hypothetical protein, conserved [7388.t00001]	819	93.017	1	0.768

1	MAAAKTELAMELNDVVCFERLCADREWRNLTGKVI	60	1	MAAAKTELAMELNDVVCFERLCADREWRNLTGKVI	60
61	VDEKIQEMATDTGKLAVWEEKLELDHQLQQA	120	61	VDEKIQEMATDTGKLAVWEEKLELDHQLQQA	120
121	RETMEALDEARKAMTRVFPFRQMRDVRAPGR	180	121	RETMEALDEARKAMTRVFPFRQMRDVRAPGR	180
181	MRQQNFTERLMHLDCTVTFPLARRKKIVAEL	240	181	MRQQNFTERLMHLDCTVTFPLARRKKIVAEL	240
241	SAAHATQIEIAMRDMLLAQVACSEAREAEHV	300	241	SAAHATQIEIAMRDMLLAQVACSEAREAEHV	300
301	KEKILEKCLAICGDTPLAALLIKDAEASKD	360	301	KEKILEKCLAICGDTPLAALLIKDAEASKD	360
361	EEQLDEENALFSLTPEEVMQLRDAARAANE	420	361	EEQLDEENALFSLTPEEVMQLRDAARAANE	420
421	EELRHKENFTEDDEEEMRQLDAALTDVVR	480	421	EELRHKENFTEDDEEEMRQLDAALTDVVR	480
481	VYSELNLTFSGDKWGVLLANTEYLMVETAL	540	481	VYSELNLTFSGDKWGVLLANTEYLMVETAL	540
541	TVKHSRDVTAEQLQALANEAEFLGLRGLY	600	541	TVKHSRDVTAEQLQALANEAEFLGLRGLY	600
601	NFAGAGLPEKLEDFADDSGEGETDENYRR	660	601	NFAGAGLPEKLEDFADDSGEGETDENYRR	660
661	FEALFEGKRTQLRAAMYPEEEEEAAEVA	720	661	FEALFEGKRTQLRAAMYPEEEEEAAEVA	720
721	TAVEAEDVHAVHTDHDVAADHFEHAEL	780	721	TAVEAEDVHAVHTDHDVAADHFEHAEL	780
781	ADAVGSAGFLEPKETLEAPKDGQRDPDS	840	781	SEHVDSVGSAGFLEPKETLEAPKDGQRDP	840



Band 13: ~60-90Kda

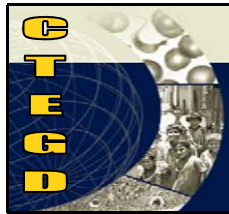
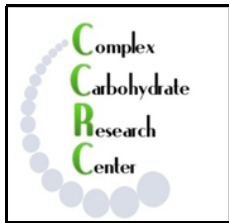
Band 16: ~30-50Kda



Conclusions

- Protein isoforms can be resolved by assigning peptides uniquely to proteins by molecular weight and gel band identification
 - Protein coverage must be high in order to have high-confidence de-convolution
 - Truncated/degraded products and/or modified proteins complicate the problem
- Label-free quantitation calculations will improve with the additional intensity information
 - i.e. spectral counts of resolved peptides

Acknowledgments



Dr. James Atwood, COO

Ms. Lin Lin, Software Engineer

Dr. Rick Tarleton (CTEGD)

Dr. Ron Orlando, CCRC

Dr. Todd Minning, CTEGD

For more information about ProteoIQ: www.bioinquire.com