Bioinformatics and Database Support NIDA Neuroproteomics Center

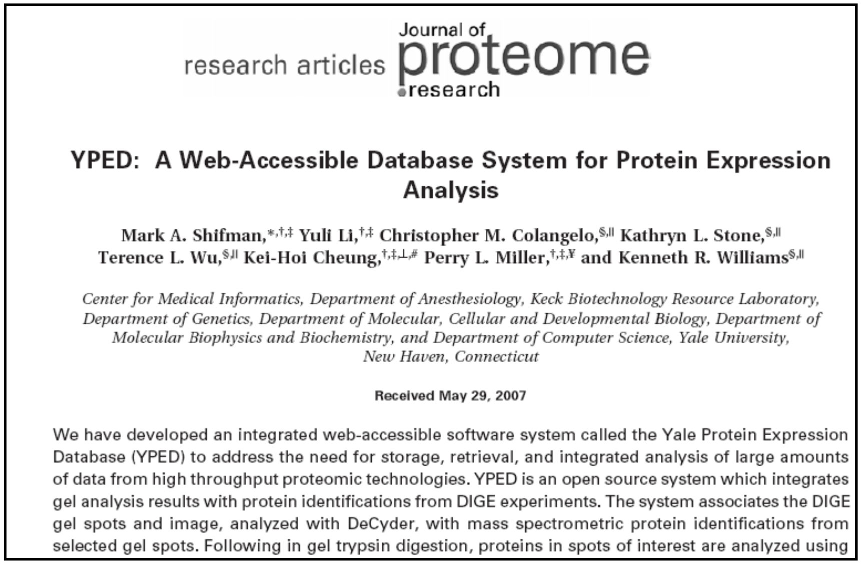
2008 Update

Kei Cheung, PhD

Mark Shifman, MD, PhD

Center for Medical Informatics

Yale Protein Expression Database (YPED)



research articles Journal of proteome research

YPED: A Web-Accessible Database System for Protein Expression Analysis

Mark A. Shifman,[*,†,‡] **Yuli Li,**[†,‡] **Christopher M. Colangelo,**[§,‖] **Kathryn L. Stone,**[§,‖]
Terence L. Wu,[§,‖] **Kei-Hoi Cheung,**[†,‡,⊥,#] **Perry L. Miller,**[†,‡,¥] **and Kenneth R. Williams**[§,‖]

Center for Medical Informatics, Department of Anesthesiology, Keck Biotechnology Resource Laboratory, Department of Genetics, Department of Molecular, Cellular and Developmental Biology, Department of Molecular Biophysics and Biochemistry, and Department of Computer Science, Yale University, New Haven, Connecticut

Received May 29, 2007

We have developed an integrated web-accessible software system called the Yale Protein Expression Database (YPED) to address the need for storage, retrieval, and integrated analysis of large amounts of data from high throughput proteomic technologies. YPED is an open source system which integrates gel analysis results with protein identifications from DIGE experiments. The system associates the DIGE gel spots and image, analyzed with DeCyder, with mass spectrometric protein identifications from selected gel spots. Following in gel trypsin digestion, proteins in spots of interest are analyzed using

- J Proteome Res. 2007 Oct;6(10):4019-24
- Web-accessible database (open source)
- Sample description and result reporting
- LC-MS/MS, MudPIT, 2D Gel, DIGE, and iTRAQ
- Java using Struts framework, Tomcat, 4TB file server, Oracle

Analytical Tool Infrastructure

- DeCyder and ABI GPS for DIGE DIA & BVA
- ProteinPilot for iTRAQ including both 4plex & 8plex
- Mascot Server running on our Linux cluster for LC-MS and SILAC
- X!!Tandem for PTM
- MultiQuant biomarker quantitation

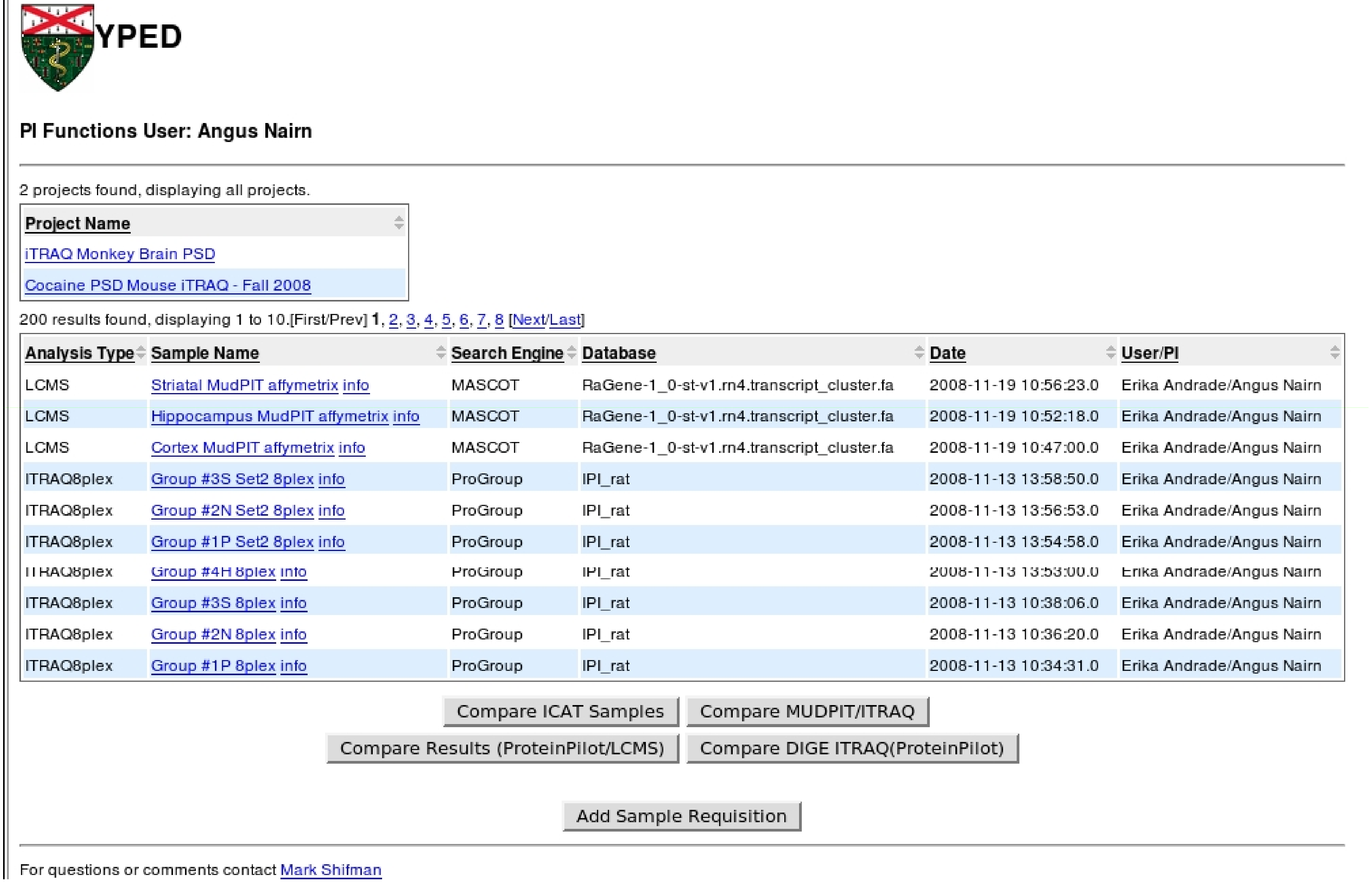
NIDA Usage (As of November, 2008)

39 NIDA users

(20 NIDA PI’s)

Analysis Type	Number of NIDA Samples
2D Gel	20
DIGE	246
ITRAQ4plex	76
ITRAQ8plex	26
LC-MS	144
MudPIT	10
SILAC	21

Different Types of Analysis Results Stored in YPED



PI Functions User: Angus Nairn

2 projects found, displaying all projects.

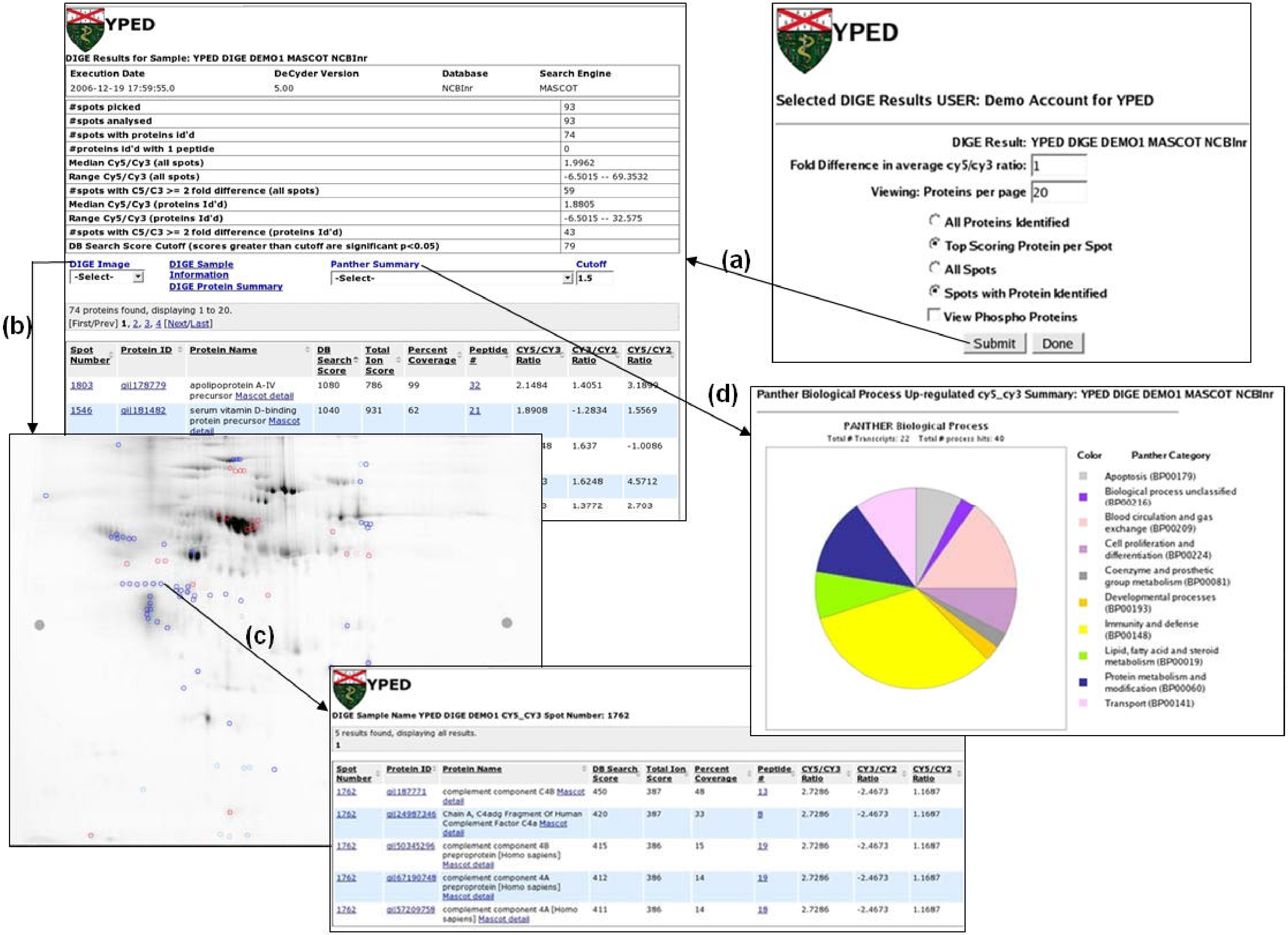
Project Name
iTRAQ Monkey Brain PSD
Cocaine PSD Mouse iTRAQ - Fall 2008

200 results found, displaying 1 to 10.[First/Prev] 1, 2, 3, 4, 5, 6, 7, 8 [Next/Last]

Analysis Type	Sample Name	Search Engine	Database	Date	User/PI
LCMS	Striatal MudPIT affymetrix info	MASCOT	RaGene-1_0-st-v1.rn4.transcript_cluster.fa	2008-11-19 10:56:23.0	Erika Andrade/Angus Nairn
LCMS	Hippocampus MudPIT affymetrix info	MASCOT	RaGene-1_0-st-v1.rn4.transcript_cluster.fa	2008-11-19 10:52:18.0	Erika Andrade/Angus Nairn
LCMS	Cortex MudPIT affymetrix info	MASCOT	RaGene-1_0-st-v1.rn4.transcript_cluster.fa	2008-11-19 10:47:00.0	Erika Andrade/Angus Nairn
ITRAQ8plex	Group #3S Set2 8plex info	ProGroup	IPI_rat	2008-11-13 13:58:50.0	Erika Andrade/Angus Nairn
ITRAQ8plex	Group #2N Set2 8plex info	ProGroup	IPI_rat	2008-11-13 13:56:53.0	Erika Andrade/Angus Nairn
ITRAQ8plex	Group #1P Set2 8plex info	ProGroup	IPI_rat	2008-11-13 13:54:58.0	Erika Andrade/Angus Nairn
ITRAQ8plex	Group #4H 8plex info	ProGroup	IPI_rat	2008-11-13 13:53:00.0	Erika Andrade/Angus Nairn
ITRAQ8plex	Group #3S 8plex info	ProGroup	IPI_rat	2008-11-13 10:38:06.0	Erika Andrade/Angus Nairn
ITRAQ8plex	Group #2N 8plex info	ProGroup	IPI_rat	2008-11-13 10:36:20.0	Erika Andrade/Angus Nairn
ITRAQ8plex	Group #1P 8plex info	ProGroup	IPI_rat	2008-11-13 10:34:31.0	Erika Andrade/Angus Nairn

For questions or comments contact Mark Shifman

DIGE Results



DIGE Results for Sample: YPED DIGE DEMO1 MASCOT NCBInr

Execution Date	DeCyder Version	Database	Search Engine
2006-12-19 17:59:55.0	5.00	NCBInr	MASCOT

#spots picked	93
#spots analysed	93
#spots with proteins id'd	74
#proteins id'd with 1 peptide	0
Median Cy5/Cy3 (all spots)	1.9962
Range Cy5/Cy3 (all spots)	-6.5015 -- 69.3532
#spots with C5/C3 >= 2 fold difference (all spots)	59
Median Cy5/Cy3 (proteins Id'd)	1.8805
Range Cy5/Cy3 (proteins Id'd)	-6.5015 -- 32.575
#spots with C5/C3 >= 2 fold difference (proteins Id'd)	43
DB Search Score Cutoff (scores greater than cutoff are significant p<0.05)	79

Cutoff: 1.5

74 proteins found, displaying 1 to 20.
[First/Prev] 1, 2, 3, 4 [Next/Last]

Spot Number	Protein ID	Protein Name	DB Search Score	Total Ion Score	Percent Coverage	Peptide #	CY5/CY3 Ratio	CY3/CY2 Ratio	CY5/CY2 Ratio
1803	gi|178779	apolipoprotein A-IV precursor Mascot detail	1080	786	99	32	2.1484	1.4051	3.1890
1546	gi|181482	serum vitamin D-binding protein precursor Mascot detail	1040	931	62	21	1.8908	-1.2834	1.5569
							48	1.637	-1.0086
							3	1.6248	4.5712
							3	1.3772	2.703

Selected DIGE Results USER: Demo Account for YPED

DIGE Result: YPED DIGE DEMO1 MASCOT NCBInr
Fold Difference in average cy5/cy3 ratio: 1
Viewing: Proteins per page 20

- ○ All Proteins Identified
- ● Top Scoring Protein per Spot
- ○ All Spots
- ● Spots with Protein Identified
- ☐ View Phospho Proteins

Panther Biological Process Up-regulated cy5_cy3 Summary: YPED DIGE DEMO1 MASCOT NCBInr

PANTHER Biological Process
Total # Transcripts: 22 Total # process hits: 40

Color	Panther Category
■	Apoptosis (BP00179)
■	Biological process unclassified (BP00216)
■	Blood circulation and gas exchange (BP00209)
■	Cell proliferation and differentiation (BP00224)
■	Coenzyme and prosthetic group metabolism (BP00081)
■	Developmental processes (BP00193)
■	Immunity and defense (BP00148)
■	Lipid, fatty acid and steroid metabolism (BP00019)
■	Protein metabolism and modification (BP00060)
■	Transport (BP00141)

DIGE Sample Name YPED DIGE DEMO1 CY5_CY3 Spot Number: 1762

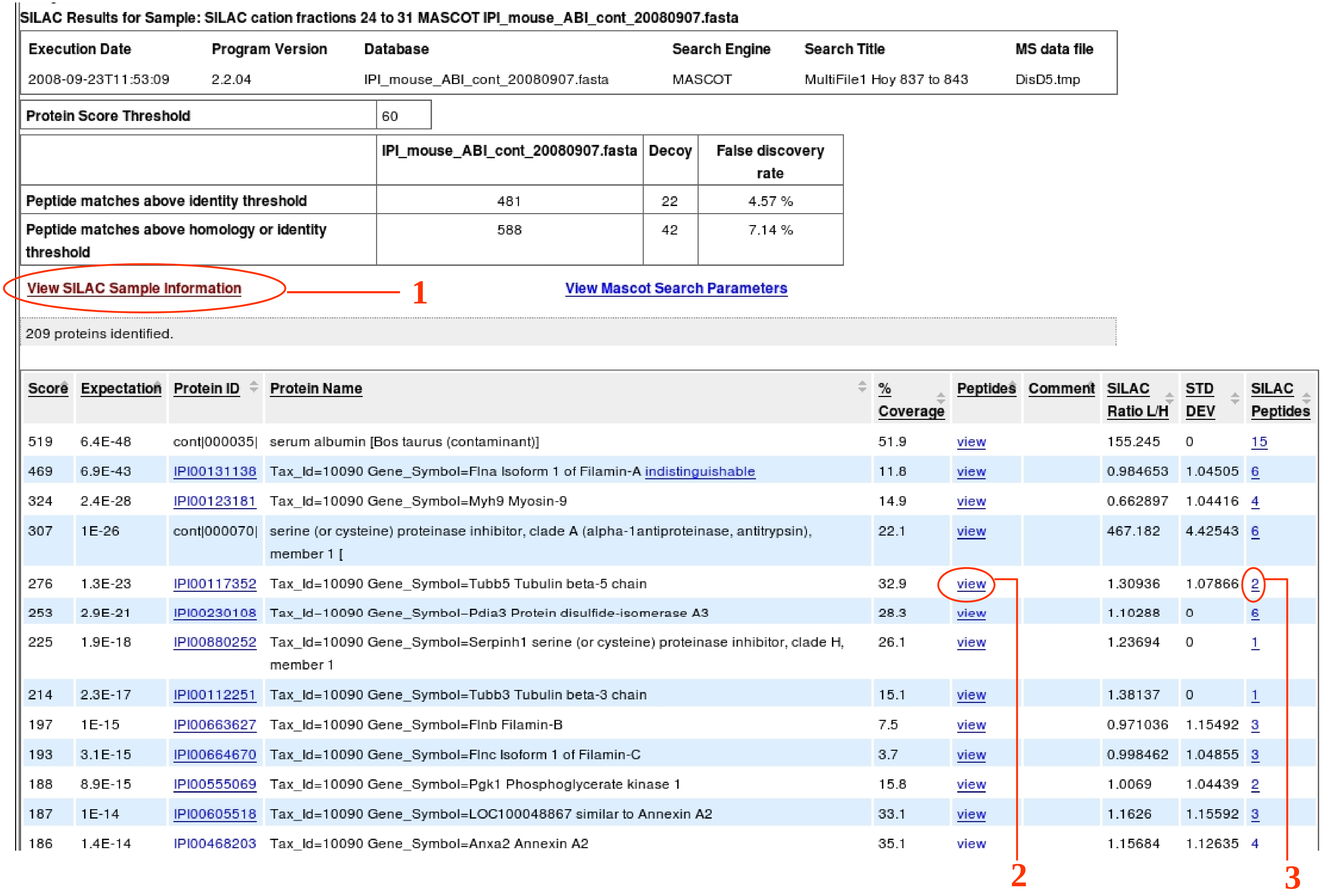
5 results found, displaying all results.
1

Spot Number	Protein ID	Protein Name	DB Search Score	Total Ion Score	Percent Coverage	Peptide #	CY5/CY3 Ratio	CY3/CY2 Ratio	CY5/CY2 Ratio
1762	gi|187771	complement component C4B Mascot detail	450	387	48	13	2.7286	-2.4673	1.1687
1762	gi|24987346	Chain A, C4adg Fragment Of Human Complement Factor C4a Mascot detail	420	387	33	8	2.7286	-2.4673	1.1687
1762	gi|50345296	complement component 4B preproprotein [Homo sapiens] Mascot detail	415	386	15	19	2.7286	-2.4673	1.1687
1762	gi|67190748	complement component 4A preproprotein [Homo sapiens] Mascot detail	412	386	14	19	2.7286	-2.4673	1.1687
1762	gi|57209759	complement component 4A [Homo sapiens] Mascot detail	411	386	14	18	2.7286	-2.4673	1.1687

Yale Protein Expression Database (Progress Since Last Site Visit)

- Upgraded Mascot server
- Refined LC-MS Mascot results
- SILAC with Mascot/Distiller
- DIGE BVA
- Synthetic peptide database
- Targeted proteomics database
- Began work on digital repository

SILAC Results



SILAC Results for Sample: SILAC cation fractions 24 to 31 MASCOT IPI_mouse_ABI_cont_20080907.fasta

Execution Date	Program Version	Database	Search Engine	Search Title	MS data file
2008-09-23T11:53:09	2.2.04	IPI_mouse_ABI_cont_20080907.fasta	MASCOT	MultiFile1 Hoy 837 to 843	DisD5.tmp

Protein Score Threshold	60

	IPI_mouse_ABI_cont_20080907.fasta	Decoy	False discovery rate
Peptide matches above identity threshold	481	22	4.57 %
Peptide matches above homology or identity threshold	588	42	7.14 %

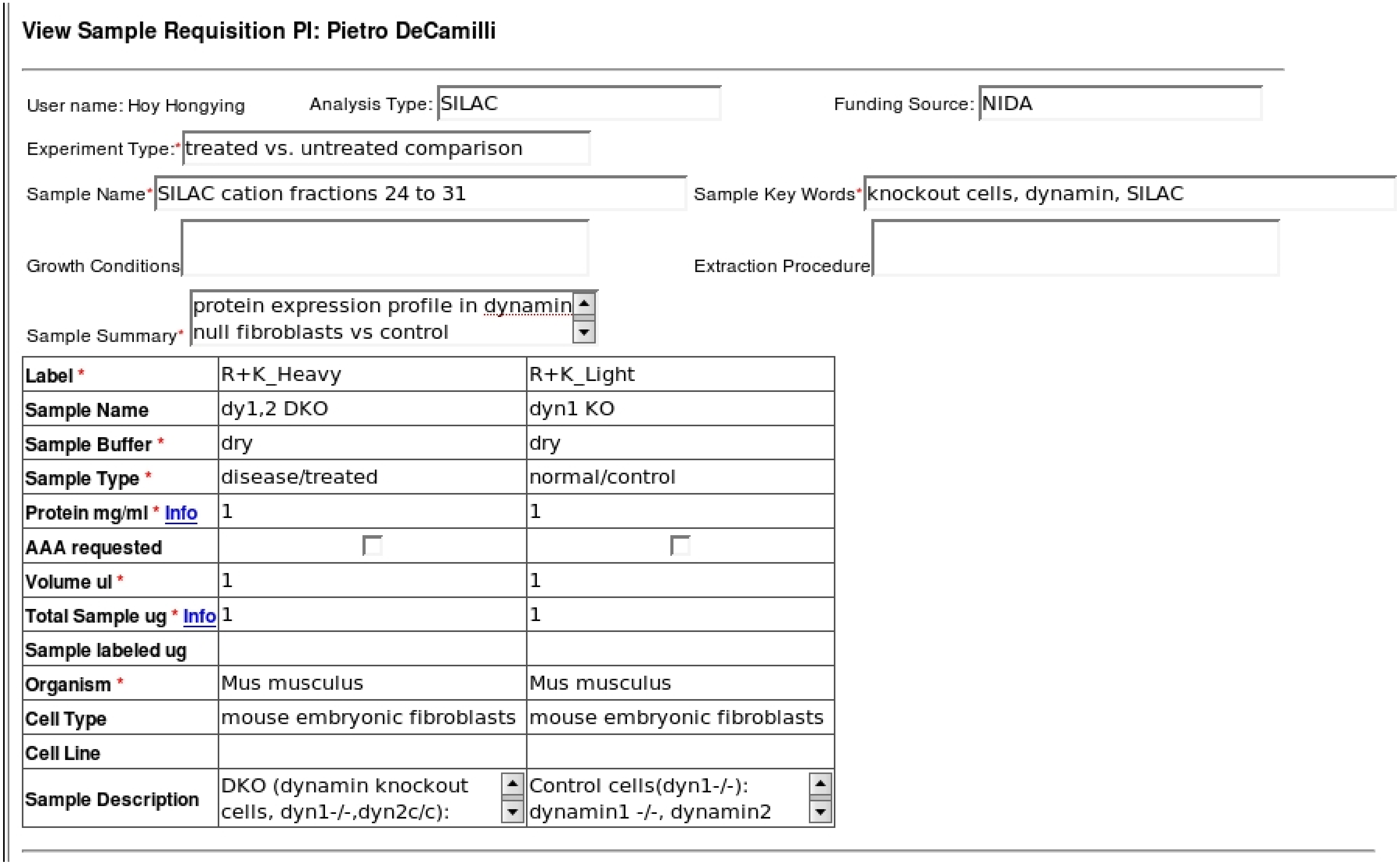
View SILAC Sample Information

View Mascot Search Parameters

209 proteins identified.

Score	Expectation	Protein ID	Protein Name	% Coverage	Peptides	Comment	SILAC Ratio L/H	STD DEV	SILAC Peptides
519	6.4E-48	cont|000035|	serum albumin [Bos taurus (contaminant)]	51.9	view		155.245	0	15
469	6.9E-43	IPI00131138	Tax_Id=10090 Gene_Symbol=Flna Isoform 1 of Filamin-A indistinguishable	11.8	view		0.984653	1.04505	6
324	2.4E-28	IPI00123181	Tax_Id=10090 Gene_Symbol=Myh9 Myosin-9	14.9	view		0.662897	1.04416	4
307	1E-26	cont|000070|	serine (or cysteine) proteinase inhibitor, clade A (alpha-1antiproteinase, antitrypsin), member 1 [	22.1	view		467.182	4.42543	6
276	1.3E-23	IPI00117352	Tax_Id=10090 Gene_Symbol=Tubb5 Tubulin beta-5 chain	32.9	view		1.30936	1.07866	2
253	2.9E-21	IPI00230108	Tax_Id=10090 Gene_Symbol=Pdia3 Protein disulfide-isomerase A3	28.3	view		1.10288	0	6
225	1.9E-18	IPI00880252	Tax_Id=10090 Gene_Symbol=Serpinh1 serine (or cysteine) proteinase inhibitor, clade H, member 1	26.1	view		1.23694	0	1
214	2.3E-17	IPI00112251	Tax_Id=10090 Gene_Symbol=Tubb3 Tubulin beta-3 chain	15.1	view		1.38137	0	1
197	1E-15	IPI00663627	Tax_Id=10090 Gene_Symbol=Flnb Filamin-B	7.5	view		0.971036	1.15492	3
193	3.1E-15	IPI00664670	Tax_Id=10090 Gene_Symbol=Flnc Isoform 1 of Filamin-C	3.7	view		0.998462	1.04855	3
188	8.9E-15	IPI00555069	Tax_Id=10090 Gene_Symbol=Pgk1 Phosphoglycerate kinase 1	15.8	view		1.0069	1.04439	2
187	1E-14	IPI00605518	Tax_Id=10090 Gene_Symbol=LOC100048867 similar to Annexin A2	33.1	view		1.1626	1.15592	3
186	1.4E-14	IPI00468203	Tax_Id=10090 Gene_Symbol=Anxa2 Annexin A2	35.1	view		1.15684	1.12635	4

Sample Description (1)



View Sample Requisition PI: Pietro DeCamilli

User name: Hoy Hongying

Analysis Type: SILAC

Funding Source: NIDA

Experiment Type:* treated vs. untreated comparison

Sample Name* SILAC cation fractions 24 to 31

Sample Key Words* knockout cells, dynamin, SILAC

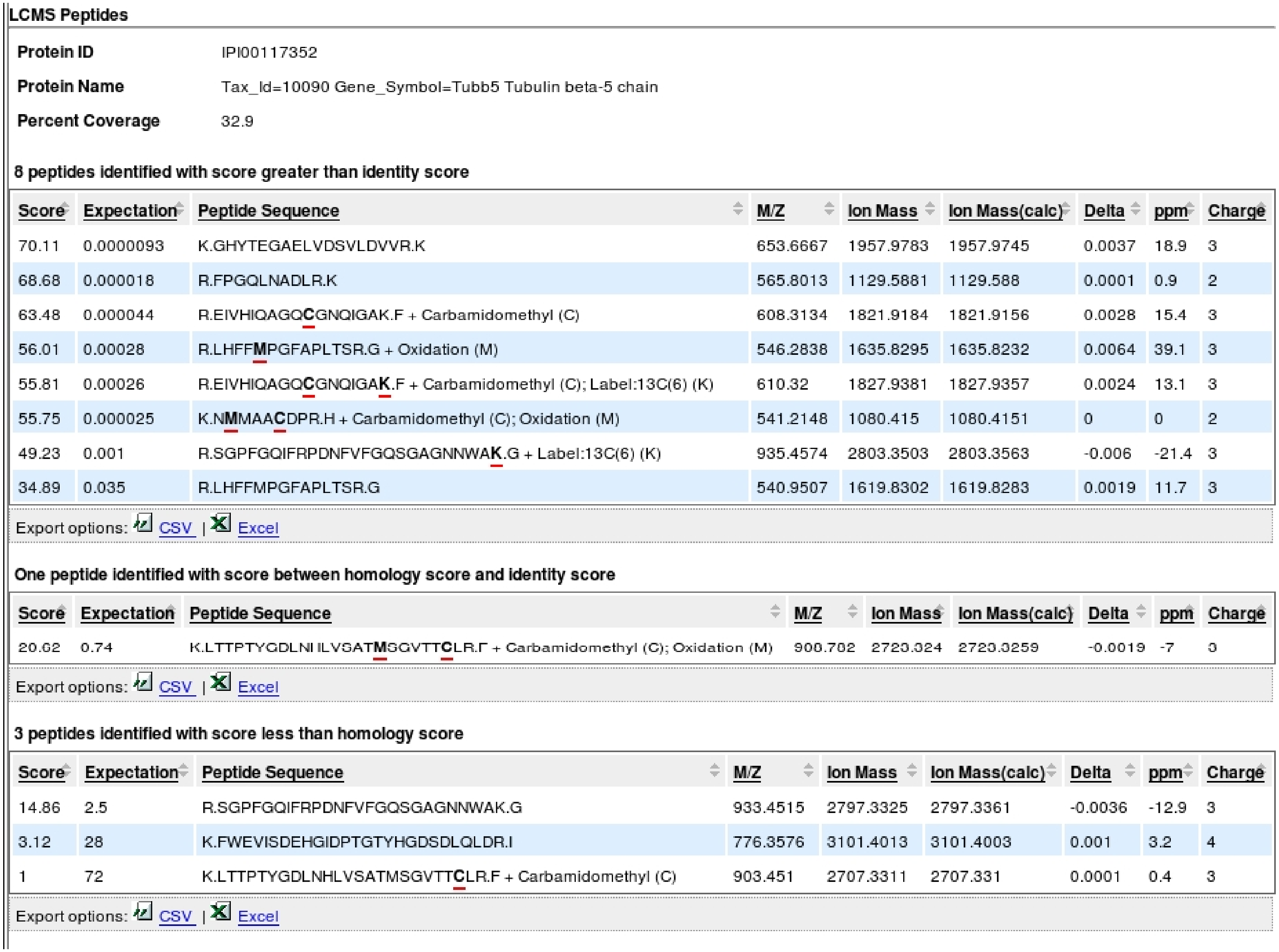
Growth Conditions

Extraction Procedure

Sample Summary* protein expression profile in dynamin null fibroblasts vs control

Label *	R+K_Heavy	R+K_Light
Sample Name	dy1,2 DKO	dyn1 KO
Sample Buffer *	dry	dry
Sample Type *	disease/treated	normal/control
Protein mg/ml * Info	1	1
AAA requested	☐	☐
Volume ul *	1	1
Total Sample ug * Info	1	1
Sample labeled ug		
Organism *	Mus musculus	Mus musculus
Cell Type	mouse embryonic fibroblasts	mouse embryonic fibroblasts
Cell Line		
Sample Description	DKO (dynamin knockout cells, dyn1-/-,dyn2c/c):	Control cells(dyn1-/-): dynamin1 -/-, dynamin2

LC-MS Peptides (2)



LCMS Peptides

Protein ID	IPI00117352
Protein Name	Tax_Id=10090 Gene_Symbol=Tubb5 Tubulin beta-5 chain
Percent Coverage	32.9

8 peptides identified with score greater than identity score

Score	Expectation	Peptide Sequence	M/Z	Ion Mass	Ion Mass(calc)	Delta	ppm	Charge
70.11	0.0000093	K.GHYTEGAELVDSVLDVVR.K	653.6667	1957.9783	1957.9745	0.0037	18.9	3
68.68	0.000018	R.FPGQLNADLR.K	565.8013	1129.5881	1129.588	0.0001	0.9	2
63.48	0.000044	R.EIVHIQAGQ**C**GNQIGAK.F + Carbamidomethyl (C)	608.3134	1821.9184	1821.9156	0.0028	15.4	3
56.01	0.00028	R.LHFF**M**PGFAPLTSR.G + Oxidation (M)	546.2838	1635.8295	1635.8232	0.0064	39.1	3
55.81	0.00026	R.EIVHIQAGQ**C**GNQIGA**K**.F + Carbamidomethyl (C); Label:13C(6) (K)	610.32	1827.9381	1827.9357	0.0024	13.1	3
55.75	0.000025	K.N**M**MAA**C**DPR.H + Carbamidomethyl (C); Oxidation (M)	541.2148	1080.415	1080.4151	0	0	2
49.23	0.001	R.SGPFGQIFRPDNFVFGQSGAGNNWA**K**.G + Label:13C(6) (K)	935.4574	2803.3503	2803.3563	-0.006	-21.4	3
34.89	0.035	R.LHFFMPGFAPLTSR.G	540.9507	1619.8302	1619.8283	0.0019	11.7	3

Export options: CSV | Excel

One peptide identified with score between homology score and identity score

Score	Expectation	Peptide Sequence	M/Z	Ion Mass	Ion Mass(calc)	Delta	ppm	Charge
20.62	0.74	K.LTTPTYGDLNHLVSAT**M**SGVTT**C**LR.F + Carbamidomethyl (C); Oxidation (M)	908.782	2723.324	2723.3259	-0.0019	-7	3

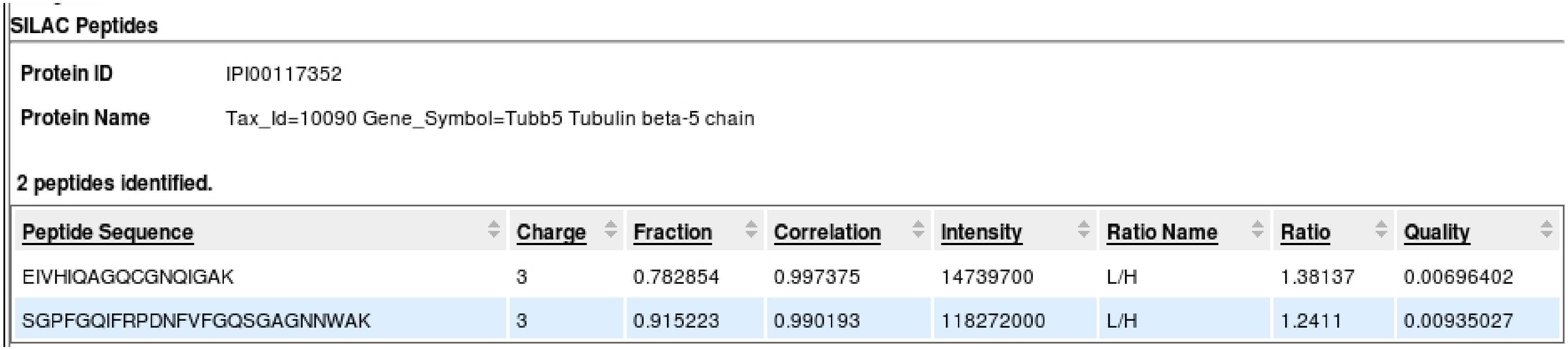
Export options: CSV | Excel

3 peptides identified with score less than homology score

Score	Expectation	Peptide Sequence	M/Z	Ion Mass	Ion Mass(calc)	Delta	ppm	Charge
14.86	2.5	R.SGPFGQIFRPDNFVFGQSGAGNNWAK.G	933.4515	2797.3325	2797.3361	-0.0036	-12.9	3
3.12	28	K.FWEVISDEHGIDPTGTYHGDSDLQLDR.I	776.3576	3101.4013	3101.4003	0.001	3.2	4
1	72	K.LTTPTYGDLNHLVSATMSGVTT**C**LR.F + Carbamidomethyl (C)	903.451	2707.3311	2707.331	0.0001	0.4	3

Export options: CSV | Excel

SILAC Peptides (3)



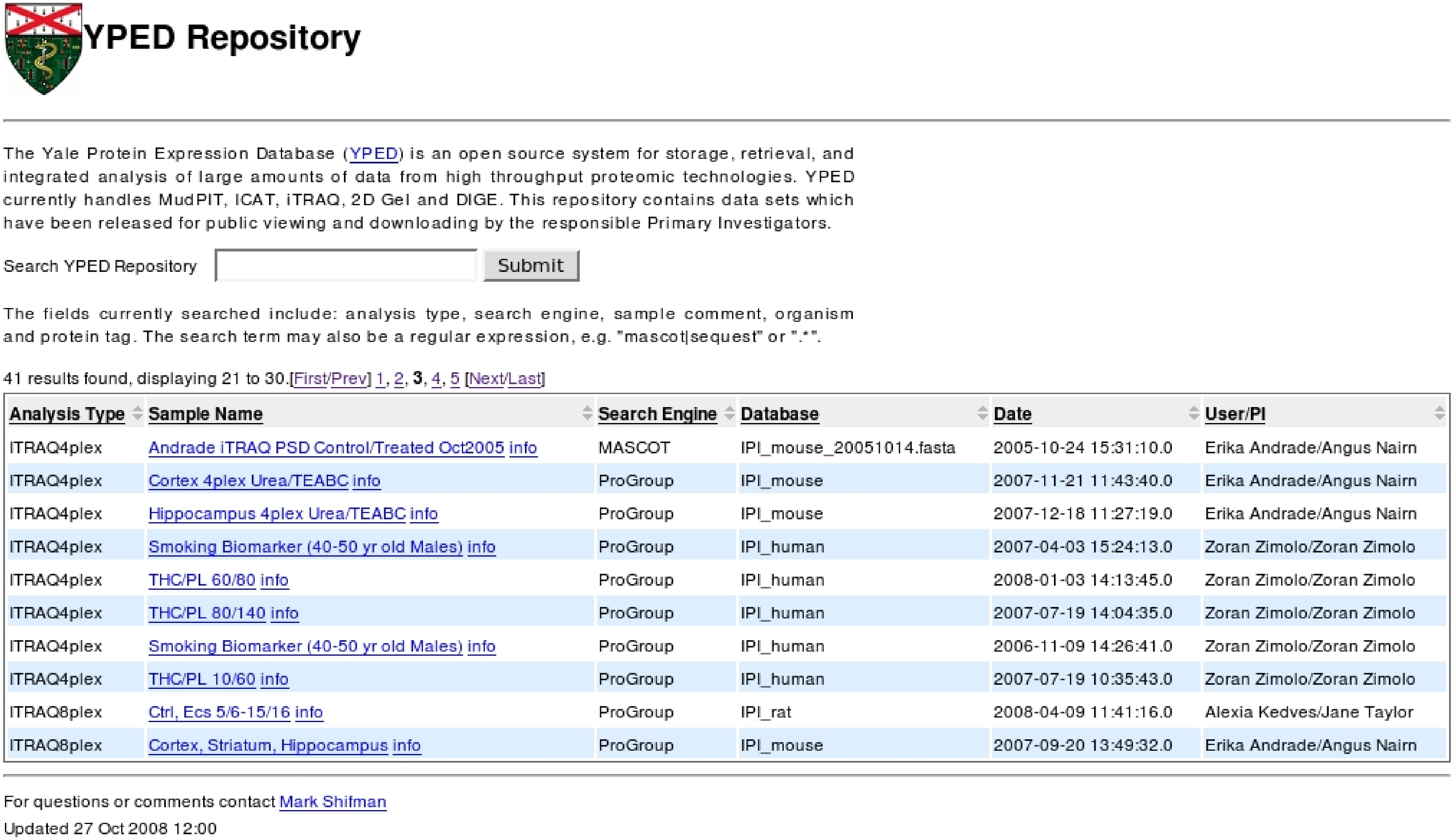
SILAC Peptides

Protein ID	IPI00117352
Protein Name	Tax_Id=10090 Gene_Symbol=Tubb5 Tubulin beta-5 chain

2 peptides identified.

Peptide Sequence	Charge	Fraction	Correlation	Intensity	Ratio Name	Ratio	Quality
EIVHIQAGQCGNQIGAK	3	0.782854	0.997375	14739700	L/H	1.38137	0.00696402
SGPFGQIFRPDNFVFGQSGAGNNWAK	3	0.915223	0.990193	118272000	L/H	1.2411	0.00935027

Prototype Repository for Public Data Access



YPED Repository

The Yale Protein Expression Database (YPED) is an open source system for storage, retrieval, and integrated analysis of large amounts of data from high throughput proteomic technologies. YPED currently handles MudPIT, ICAT, iTRAQ, 2D Gel and DIGE. This repository contains data sets which have been released for public viewing and downloading by the responsible Primary Investigators.

Search YPED Repository

The fields currently searched include: analysis type, search engine, sample comment, organism and protein tag. The search term may also be a regular expression, e.g. "mascot|sequest" or ".*".

41 results found, displaying 21 to 30.[First/Prev] 1, 2, **3**, 4, 5 [Next/Last]

Analysis Type	Sample Name	Search Engine	Database	Date	User/PI
ITRAQ4plex	Andrade iTRAQ PSD Control/Treated Oct2005 info	MASCOT	IPI_mouse_20051014.fasta	2005-10-24 15:31:10.0	Erika Andrade/Angus Nairn
ITRAQ4plex	Cortex 4plex Urea/TEABC info	ProGroup	IPI_mouse	2007-11-21 11:43:40.0	Erika Andrade/Angus Nairn
ITRAQ4plex	Hippocampus 4plex Urea/TEABC info	ProGroup	IPI_mouse	2007-12-18 11:27:19.0	Erika Andrade/Angus Nairn
ITRAQ4plex	Smoking Biomarker (40-50 yr old Males) info	ProGroup	IPI_human	2007-04-03 15:24:13.0	Zoran Zimolo/Zoran Zimolo
ITRAQ4plex	THC/PL 60/80 info	ProGroup	IPI_human	2008-01-03 14:13:45.0	Zoran Zimolo/Zoran Zimolo
ITRAQ4plex	THC/PL 80/140 info	ProGroup	IPI_human	2007-07-19 14:04:35.0	Zoran Zimolo/Zoran Zimolo
ITRAQ4plex	Smoking Biomarker (40-50 yr old Males) info	ProGroup	IPI_human	2006-11-09 14:26:41.0	Zoran Zimolo/Zoran Zimolo
ITRAQ4plex	THC/PL 10/60 info	ProGroup	IPI_human	2007-07-19 10:35:43.0	Zoran Zimolo/Zoran Zimolo
ITRAQ8plex	Ctrl, Ecs 5/6-15/16 info	ProGroup	IPI_rat	2008-04-09 11:41:16.0	Alexia Kedves/Jane Taylor
ITRAQ8plex	Cortex, Striatum, Hippocampus info	ProGroup	IPI_mouse	2007-09-20 13:49:32.0	Erika Andrade/Angus Nairn

For questions or comments contact Mark Shifman

Updated 27 Oct 2008 12:00

Yale Protein Expression Database (Future directions)

- Enhance targeted proteomics support including:
  - Synthetic peptide database
  - Web and database module for viewing biomarker quantification
- Build a prototype for the digital repository for preservation in collaboration with Yale Library

The End