decryptor user guide (v 1.1)

Miroslav Hruska

December 29, 2017

Contents

1 Introduction

decryptor is used for identification of variant peptides in shotgun proteomics data. As opposed to current proteogenomics approach, it might be used without matching sequencing data and therefore enables more universal use for almost all shotgun proteomics data. The data should be, however, measured with great depth as variants are rare among reference peptides (one variant peptide should be expected on average for around five hundred reference peptides).

2 Browsing results

This section introduces navigation, applicable to already available results. For clarity, it is done over guest account, such that each step can be performed even without registration. For actual submission of results see the next section.

2.1 Login

2.1.1 Root page

decryptor's main page informs about its use, i.e., identification of variant peptides, their genomic origin and report of metadata associated with identified variants. Note that there is no need for registration to explore this functionality. We'll continue by using "log-in" link.

decryptor (1.1) analyses data from tandem mass spectrometry of human proteome for presence of point alterations. Subsequently, decryptor deduces DNA/mRNA alterations whenever possible.

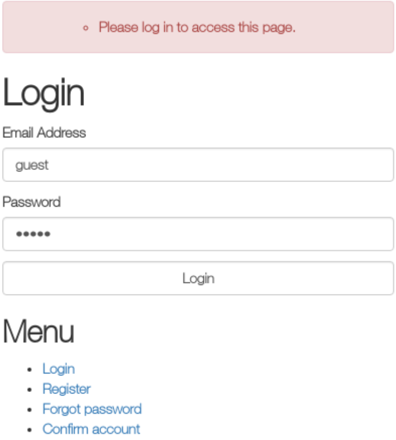
See example.
(email: *guest*, password: *guest*).

To use decryptor, you need to log-in. If not registered yet, sign-up.

Materials | Release notes | Changelog | Acknowledgement | Contact us

2.1.2 Login page

We'll login by providing guest account credentials.



- Please log in to access this page.

Login

Email Address: guest

Password: •••••

2.2 Navigation

2.2.1 Experiment view

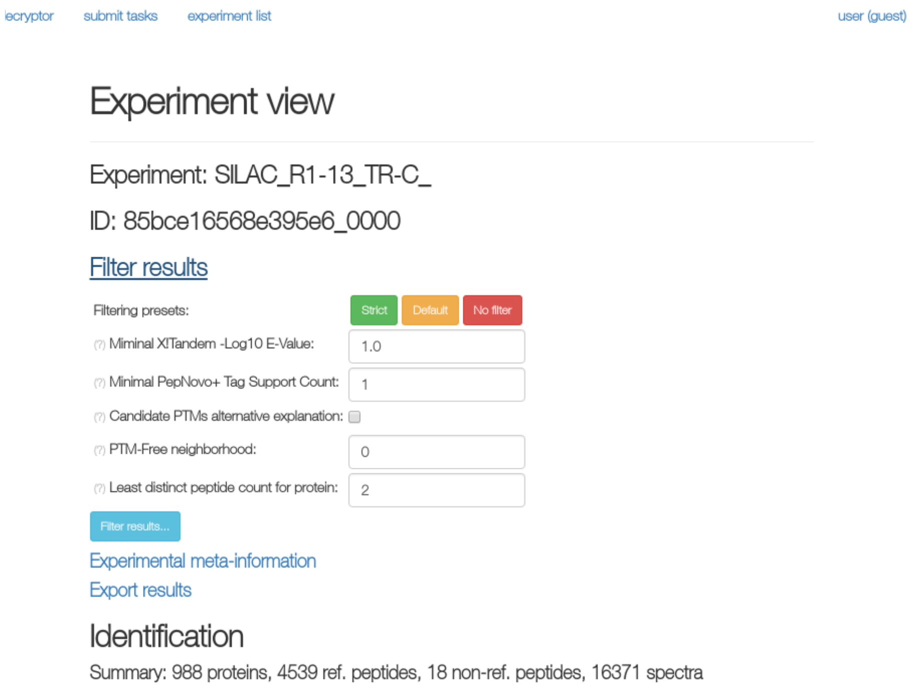
Overall results and their filtering The login lands on experiment view, which corresponds to overall view over evaluated sample (in this case, example). In the following figure, the “Filter results” was expanded. The filters give users additional control over selection of results and can be used in pre-defined way. The standard “Default” filter is one which should preserve most of the true result there, however lacking specificity. The “Strict filter”, on the other hand, should give results which are confident. To see explanation of particular fields, hover mouse over the “(?)” symbol, or continue reading following paragraphs.

X!Tandem E-Value The filter refers to minimal statistical significance of spectral match (XTandem’s HyperScore), which is necessary to report such match. Although the value does not directly relate to probability of spectral match, at value of 0.1 peptides, if evaluated on reference peptides, such number corresponds roughly to 90% of correct peptides.

PepNovo+ Tag Support Count In identification, effort is made to read the peptide sequence directly from the spectrum. The sequencing is performed in form of short subsequences, so-called peptide tags (here, set to length of three). These tags are then evaluated for correspondence with the sequencing identified from database search in X!Tandem.

PTM-Free neighborhood The value corresponds to the number of amino acid residues neighboring the variant amino acid to contain no post-translational modifications of mass corresponding to mass of substitution.

Least peptide count Minimal number of peptides per protein, for the protein to be reported. Due proteolysis, all peptides from protein are in the sample, or none; therefore it is unlikely that some peptide will be identified without corresponding reference peptides.



Experiment view

Experiment: SILAC_R1-13_TR-C_

ID: 85bce16568e395e6_0000

Filter results

Filtering presets:

(?) Miminal XITandem -Log10 E-Value: 1.0

(?) Minimal PepNovo+ Tag Support Count: 1

(?) Candidate PTMs alternative explanation: ☐

(?) PTM-Free neighborhood: 0

(?) Least distinct peptide count for protein: 2

Experimental meta-information

Export results

Identification

Summary: 988 proteins, 4539 ref. peptides, 18 non-ref. peptides, 16371 spectra

Experimental meta-information The experiment also contains meta-information which was filled in during the submission of the task to help organize the searches. See the next section.

Experimental meta-information

Parameter	Value
Fragmentation	CID
Protease	trypsin
Fragment tolerance	0.5 Da
Variable modifications	Oxidation (M)
Experiment info	SILAC_R1-13_TR-C_
Ip	127.0.0.1
Mail	hruska.miro@gmail.com
Fixed modifications	Carbamidomethyl (C)
Precursor tolerance	10 ppm

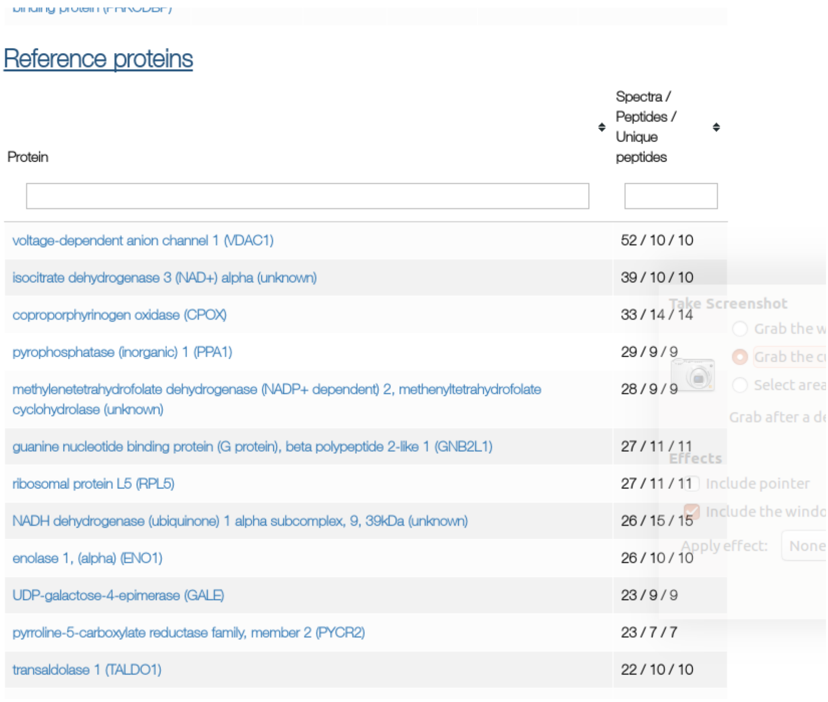
Export results

Proteins with alterations The most important view consists of proteins with claimed alerations. The view contains information about protein, the number of spectra (quantitative information) and peptides (number of distinct peptides per protein) identified. The detrimental effect is aggregated value of predicted detrimental effect of individual alterations; more details will be shown in Protein View. Disease relevance column contains information whether particular protein was linked in disease or in cancer (e.g., being an oncogene).

Proteins with sequence alterations

Protein	Spectra / Peptides / Unique peptides	Alterations	Detrimental effect	DNA/mRNA alteration source	Disease relevance
lactate dehydrogenase B (LDHB)	106 / 13 / 13	251: A>S 252: I>L	1.39		
malate dehydrogenase 2, NAD (mitochondrial) (MDH2)	51 / 11 / 11	235: V>I	0.67	COSMIC v.68 — endometrium ICGC 15.1—UCEC-US	
RNA terminal phosphate cyclase-like 1 (RCL1)	4 / 3 / 3	106: V>I	0.62	COSMIC v.68 — large_intestine	
eukaryotic translation elongation factor 1 alpha 1 (EEF1A1)	6 / 4 / 4	231: L>V	0.60	COSMIC v.68—liver ICGC 15.1—LINC-JP	
phosphoglycerate kinase 1 (unknown)	10 / 7 / 7	86: S>T	0.54	COSMIC v.68—kidney	
family with sequence similarity 192, member A (FAM192A)	5 / 3 / 3	47: V>I	0.49	ICGC 15.1—SKCM-US	

Reference proteins The rest of the view contains information about identified reference proteins.

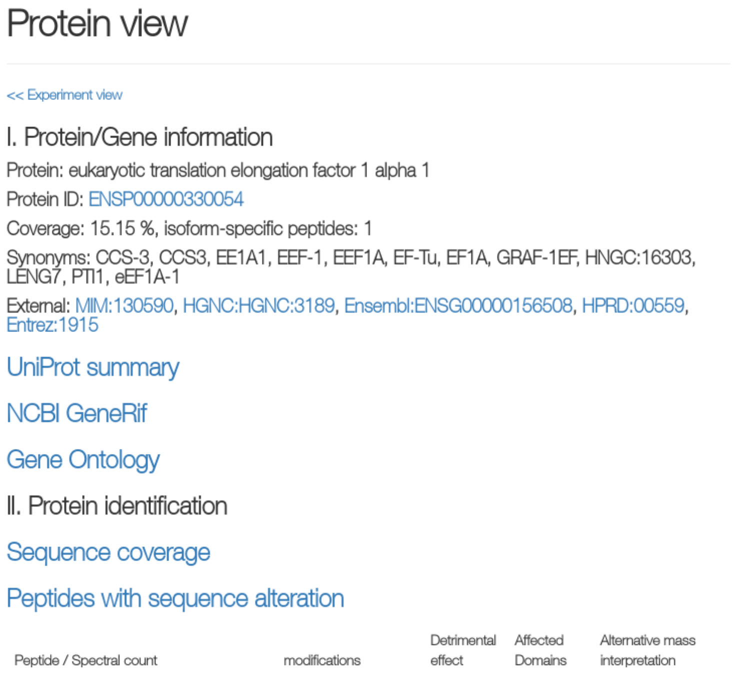


Reference proteins

Protein	Spectra / Peptides / Unique peptides
voltage-dependent anion channel 1 (VDAC1)	52 / 10 / 10
isocitrate dehydrogenase 3 (NAD+) alpha (unknown)	39 / 10 / 10
coproporphyrinogen oxidase (CPOX)	33 / 14 / 14
pyrophosphatase (inorganic) 1 (PPA1)	29 / 9 / 9
methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase (unknown)	28 / 9 / 9
guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1 (GNB2L1)	27 / 11 / 11
ribosomal protein L5 (RPL5)	27 / 11 / 11
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9, 39kDa (unknown)	26 / 15 / 15
enolase 1, (alpha) (ENO1)	26 / 10 / 10
UDP-galactose-4-epimerase (GALE)	23 / 9 / 9
pyrroline-5-carboxylate reductase family, member 2 (PYCR2)	23 / 7 / 7
transaldolase 1 (TALDO1)	22 / 10 / 10

2.2.2 Protein view

Overall view The protein view contains information relevant to particular protein and corresponding gene, external links and extracted information from several sources (UniProt, NCBI GeneRif, Gene Ontology).



Protein view

<< Experiment view

I. Protein/Gene information

Protein: eukaryotic translation elongation factor 1 alpha 1

Protein ID: ENSP00000330054

Coverage: 15.15 %, isoform-specific peptides: 1

Synonyms: CCS-3, CCS3, EE1A1, EEF-1, EEF1A, EF-Tu, EF1A, GRAF-1EF, HNGC:16303, LENG7, PTI1, eEF1A-1

External: MIM:130590, HGNC:HGNC:3189, Ensembl:ENSG00000156508, HPRD:00559, Entrez:1915

UniProt summary

NCBI GeneRif

Gene Ontology

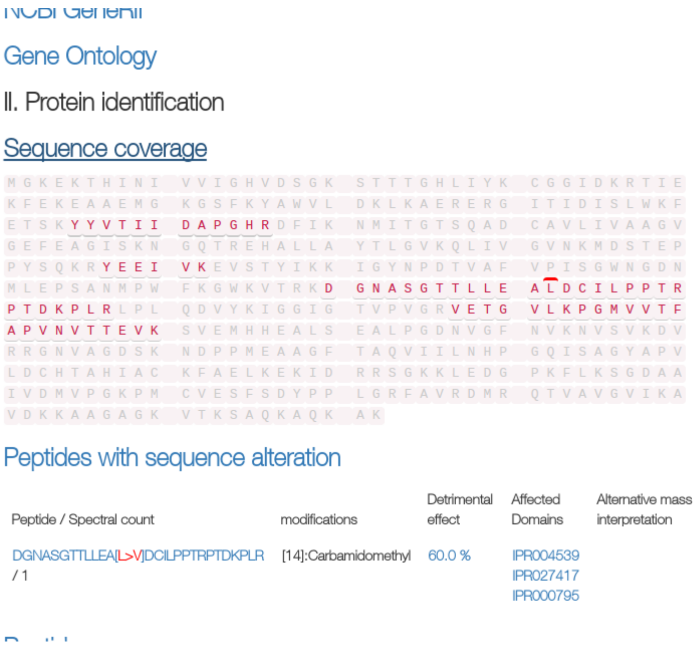
II. Protein identification

Sequence coverage

Peptides with sequence alteration

Peptide / Spectral count	modifications	Detrimental effect	Affected Domains	Alternative mass interpretation

Sequence coverage The sequence coverage shows the peptide and its identified subsequences, together with identified alterations.



NCBI GeneRIF

Gene Ontology

II. Protein identification

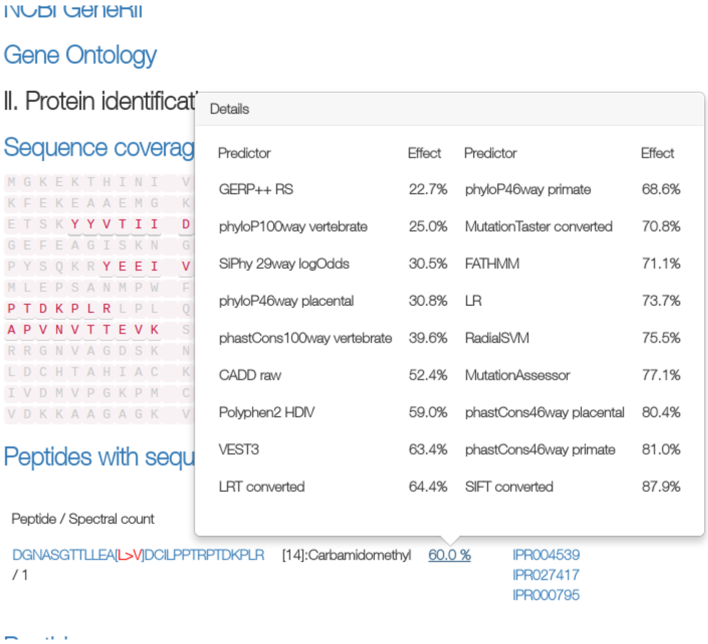
Sequence coverage

```
M G K E K T H I N I   V V I G H V D S G K   S T T T G H L I Y K   C G G I D K R T I E
K F E K E A A E M G   K G S F K Y A W V L   D K L K A E R E R G   I T I D I S L W K F
E T S K Y Y V T I I   D A P G H R D F I K   N M I T G T S Q A D   C A V L I V A A G V
G E F E A G I S K N   G Q T R E H A L L A   Y T L G V K Q L I V   G V N K M D S T E P
P Y S Q K R Y E E I   V K E V S T Y I K K   I G Y N P D T V A F   V P I S G W N G D N
M L E P S A N M P W   F K G W K V T R K D   G N A S G T T L L E   A L D C I L P P T R
P T D K P L R L P L   Q D V Y K I G G I G   T V P V G R V E T G   V L K P G M V V T F
A P V N V T T E V K   S V E M H H E A L S   E A L P G D N V G F   N V K N V S V K D V
R R G N V A G D S K   N D P P M E A A G F   T A Q V I I L N H P   G Q I S A G Y A P V
L D C H T A H I A C   K F A E L K E K I D   R R S G K K L E D G   P K F L K S G D A A
I V D M V P G K P M   C V E S F S D Y P P   L G R F A V R D M R   Q T V A V G V I K A
V D K K A A G A G K   V T K S A Q K A Q K   A K
```

Peptides with sequence alteration

Peptide / Spectral count	modifications	Detrimental effect	Affected Domains	Alternative mass interpretation
DGNASGTTLLEA[L>V]DCILPPTRPTDKPLR / 1	[14]:Carbamidomethyl	60.0 %	IPR004539 IPR027417 IPR000795	

Detrimental effect Clicking the number of detrimental effect at particular peptide expands the aggregated value into its consituent parts as predicted by dbNSFP.



NCBI GeneRIF

Gene Ontology

II. Protein identificat

Sequence coverag

```
M G K E K T H I N I   V
K F E K E A A E M G   K
E T S K Y Y V T I I   D
G E F E A G I S K N   G
P Y S Q K R Y E E I   V
M L E P S A N M P W   F
P T D K P L R L P L   Q
A P V N V T T E V K   S
R R G N V A G D S K   N
L D C H T A H I A C   K
I V D M V P G K P M   C
V D K K A A G A G K   V
```

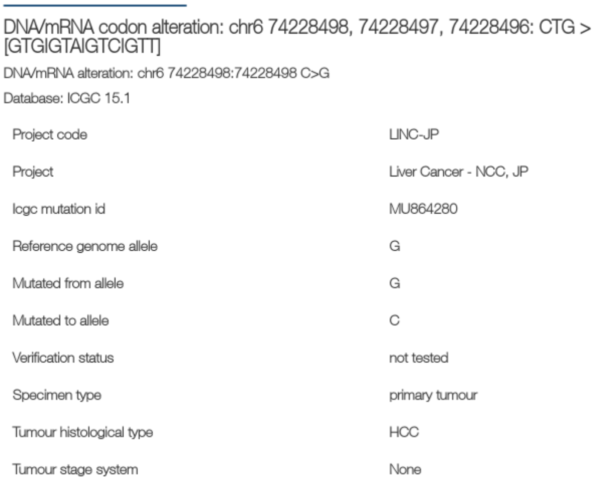
Peptides with sequ

Details

Predictor	Effect	Predictor	Effect
GERP++ RS	22.7%	phyloP46way primate	68.6%
phyloP100way vertebrate	25.0%	MutationTaster converted	70.8%
SiPhy 29way logOdds	30.5%	FATHMM	71.1%
phyloP46way placental	30.8%	LR	73.7%
phastCons100way vertebrate	39.6%	RadialSVM	75.5%
CADD raw	52.4%	MutationAssessor	77.1%
Polyphen2 HDIV	59.0%	phastCons46way placental	80.4%
VEST3	63.4%	phastCons46way primate	81.0%
LRT converted	64.4%	SIFT converted	87.9%

Peptide / Spectral count			
DGNASGTTLLEA[L>V]DCILPPTRPTDKPLR / 1	[14]:Carbamidomethyl	60.0 %	IPR004539 IPR027417 IPR000795

Alteration details For more detailed information, the alteration detail contains the source of this alteration, where it was observed and source-specific details.



DNA/mRNA codon alteration: chr6 74228498, 74228497, 74228496: CTG > [GTG|GTA|GTC|GTT]

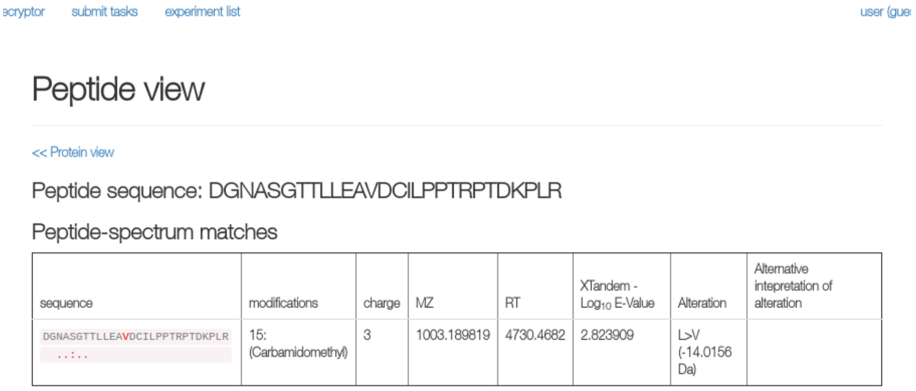
DNA/mRNA alteration: chr6 74228498:74228498 C>G

Database: ICGC 15.1

Project code	LINC-JP
Project	Liver Cancer - NCC, JP
Icgc mutation id	MU864280
Reference genome allele	G
Mutated from allele	G
Mutated to allele	C
Verification status	not tested
Specimen type	primary tumour
Tumour histological type	HCC
Tumour stage system	None

2.2.3 Peptide view

If there is a need, one could see the details of identification of particular peptide; which contains additional mass-spectrometric data such as charge, mass-to-charge ratio and retention time. The XTandem E-Value column contains log10 of statistical significance of spectral match. Column corresponding to alternative interpretation of alteration contains other candidate explanations of observed mass changes with respect to reference peptide. In this case, there are no other explanations known; note: modifications are drawn from UniMod.



Peptide view

Peptide sequence: DGNASGTTLLEAVDCILPPTRPTDKPLR

Peptide-spectrum matches

sequence	modifications	charge	MZ	RT	XTandem - Log$_{10}$ E-Value	Alteration	Alternative intepretation of alteration
DGNASGTTLLEAVDCILPPTRPTDKPLR ..:..	15: (Carbamidomethyl)	3	1003.189819	4730.4682	2.823909	L>V (-14.0156 Da)	

3 Task submission

For the ability to submit tasks, user needs to be registered by filling up the corresponding registration form.

3.1 Submission

During the submission, the usual mass-spectrometric information is filled in; moreover, last configuration of modifications can be reloaded. The user is informed on completion of the evaluation, if e-mail is provided. Please note that although decryptor supports wide variety of modifications (obtained from UniMod), its use with data with unusual modifications was not tested and its performance is not guaranteed.

(?) MS/MS files [.mzML, .mzXML, .mgf; Spectra convertor] (max: 4 GB)
No file chosen

(?) Fixed modifications

(?) Variable modifications

(?) Filter modifications

(?) Modifications
15dB-biotin (C +626.39 Da)
2-succinyl (C +117.02 Da)
2HPG (R +282.05 Da)
3-deoxyglucosone (R +144.04 D
3sulfo (N-term +183.98 Da)
4-ONE (C +154.10 Da)
4-ONE (H +154.10 Da)

(?) Protease: trypsin

(?) Fragmentation: CID

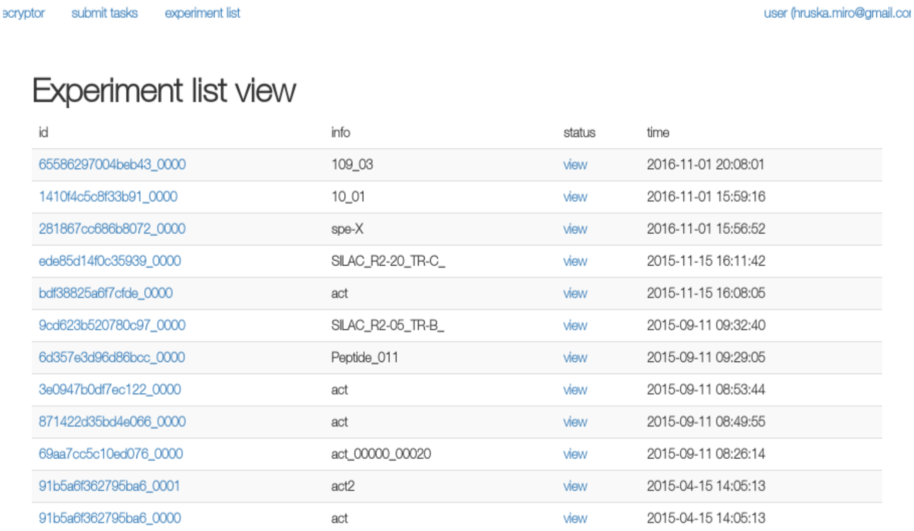
(?) Precursor tolerance: 10 ppm

(?) Fragment tolerance: 0.5 Da

(?) Mail: hruska.miro@gmail.com

3.2 Selection of results

The results can be then accessed through the experiment list.



Experiment list view

id	info	status	time
65586297004beb43_0000	109_03	view	2016-11-01 20:08:01
1410f4c5c8f33b91_0000	10_01	view	2016-11-01 15:59:16
281867cc686b8072_0000	spe-X	view	2016-11-01 15:56:52
ede85d14f0c35939_0000	SILAC_R2-20_TR-C_	view	2015-11-15 16:11:42
bdf38825a6f7cfde_0000	act	view	2015-11-15 16:08:05
9cd623b520780c97_0000	SILAC_R2-05_TR-B_	view	2015-09-11 09:32:40
6d357e3d96d86bcc_0000	Peptide_011	view	2015-09-11 09:29:05
3e0947b0df7ec122_0000	act	view	2015-09-11 08:53:44
871422d35bd4e066_0000	act	view	2015-09-11 08:49:55
69aa7cc5c10ed076_0000	act_00000_00020	view	2015-09-11 08:26:14
91b5a6f362795ba6_0001	act2	view	2015-04-15 14:05:13
91b5a6f362795ba6_0000	act	view	2015-04-15 14:05:13