decryptor user guide (v 1.1.3)

Miroslav Hruska

December 29, 2017

Contents

1 Introduction

decryptor is a computational system used for detection of variant peptides in *standard* shotgun proteomics data. **decryptor** can be used without sequencing data, as opposed to the more prevalent proteogenomics approach. The detection of variant peptides is thus not guided. In this respect, it enables universal applicability for most shotgun proteomics data. On the other hand, the data are required to be measured with great depth. This is because rare peptides are more common in deeper measurements. In particular, it is expected to detect more than double variant peptides if double of spectra was measured.

1.1 Applications

Unaided detection of variant peptides has variety of potential applications. For instance, variant peptides resulting from somatic variant can be used as tumour-specific biomarkers for monitoring of progression of disease. Germline variants can be used for establishment of origin of sample.

1.2 Word of caution

The detection of variant peptides is a complicated problem and **decryptor** might give false positives. The most reliably detected variant peptides are polymorphic peptides of population frequency higher than 1%. Therefore extra care should be taken in evaluation of identified somatic variants.

2 Browsing results

This section introduces navigation, applicable to already available results. Herein, it is exemplified over guest account, such that each step can be performed 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.

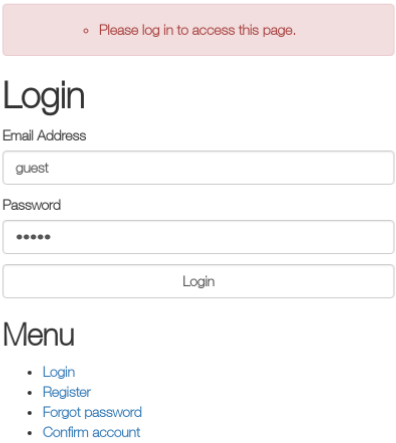
See example.
(email: *guest*, password: *guest*).

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.



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 but can be also used in predefined way (Default/Strict). 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.

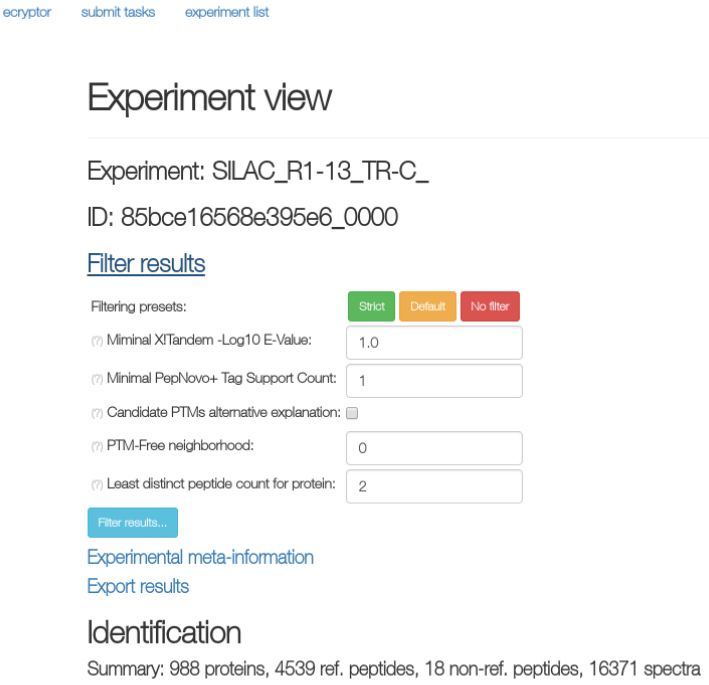
X!Tandem E-Value The filter refers to minimal statistical significance of spectral match (XTandem's HyperScore) to report spectral match. Note however, that the *value does not directly relate to probability that interpretation is correct*. However, as a guideline, at significance level of 0.1, one would expect around 90% of correct results on variant peptides of population frequency higher than 1%. Note however, that this does not extend intuitively for variant peptides of lower population frequency or somatic variants.

PepNovo+ Tag Support Count For identification, **decryptor** is tries to read the peptide sequence directly from the spectrum. The sequencing is performed in form of short subsequences, so-called peptide tags (here, set of length of three). These tags are evaluated for correspondence with the peptide sequencing as matched using 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. Thus for instance, if there is a candidate modification of N→D, but there exist *Deamidation* of N, such variant is not reported. Similarly, often A→S happens, but its mass is similar to that of *Oxidation*. Therefore, if M is nearby, it is more likely the *Oxidation* of M than the variant peptide.

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. Therefore, one would expect at least one other reference peptide for variant peptide identified.



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.

Filter results...

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

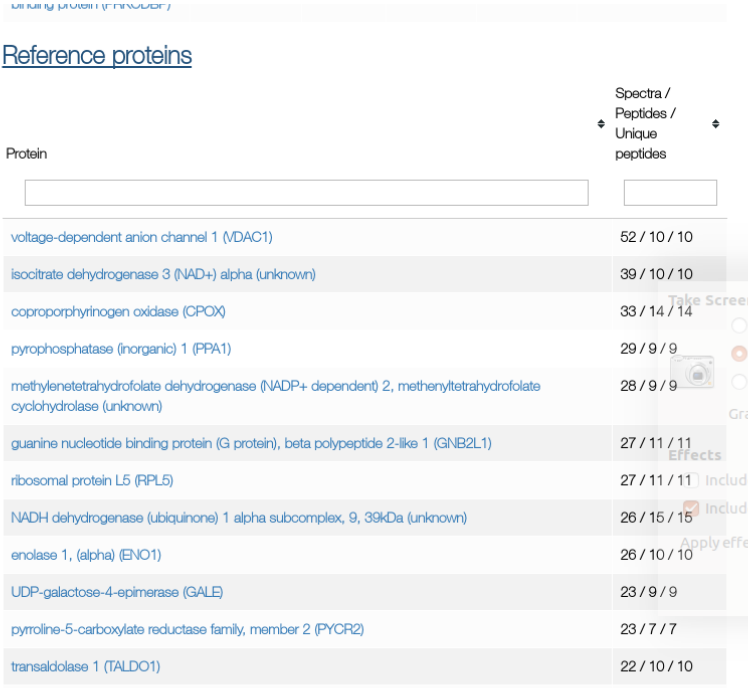
Export results

Proteins with alterations The most important view consists of proteins with claimed detected peptides. The view contains information about protein, the number of spectra (quantitative information) and peptides (number of distinct peptides per protein) identified. The detrimental effect over protein is aggregated value of predicted detrimental effect of individual variants (as calculated using dbNSFP, v 2.5). Disease relevance column contains information whether particular protein was linked to 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.

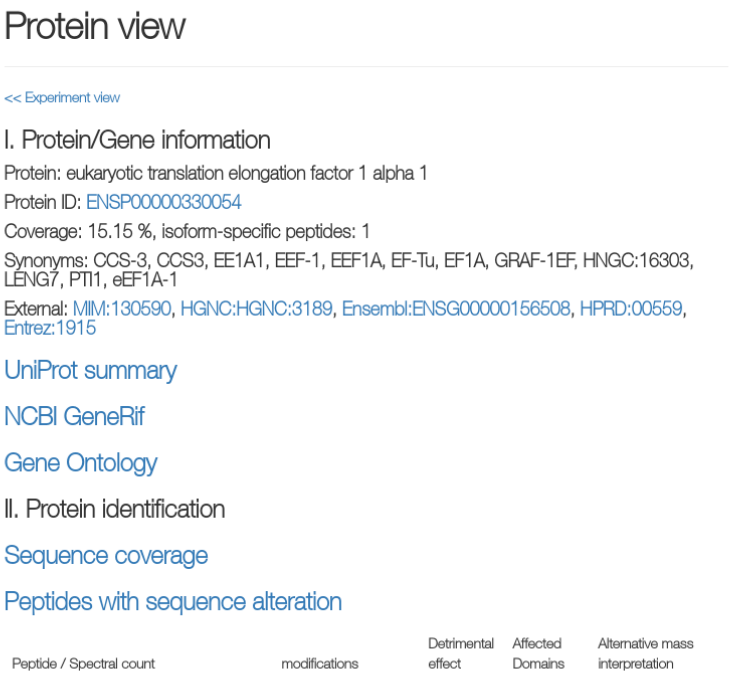


Reference proteins

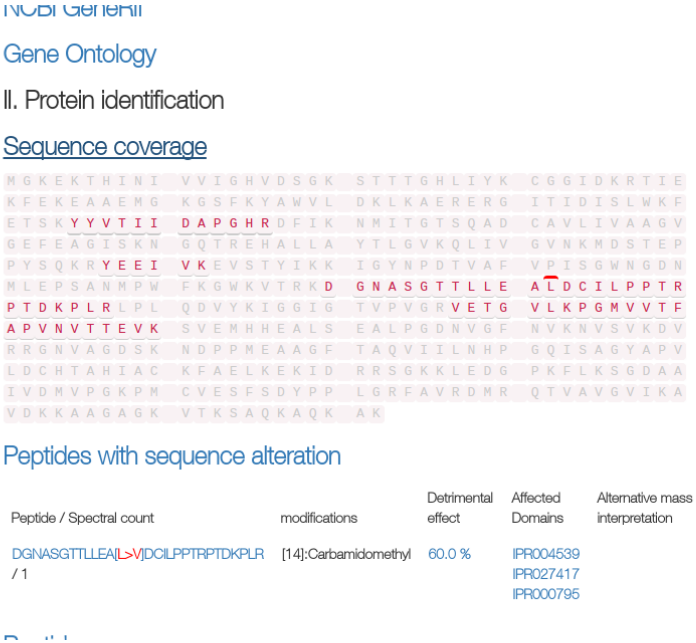
Protein	Spectra / Peptides / Unique peptides
voltage-dependent anion channel 1 (VDAC1)	52 / 10 / 10
isocitrate dehydrogenase 3 (NAD+) alpha (unknown)	39 / 10 / 10
coproporphyrinogen oxidase (CPOX)	33 / 14 / 14
pyrophosphatase (inorganic) 1 (PPA1)	29 / 9 / 9
methylenetetrahydrofolate dehydrogenase (NADP+ 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

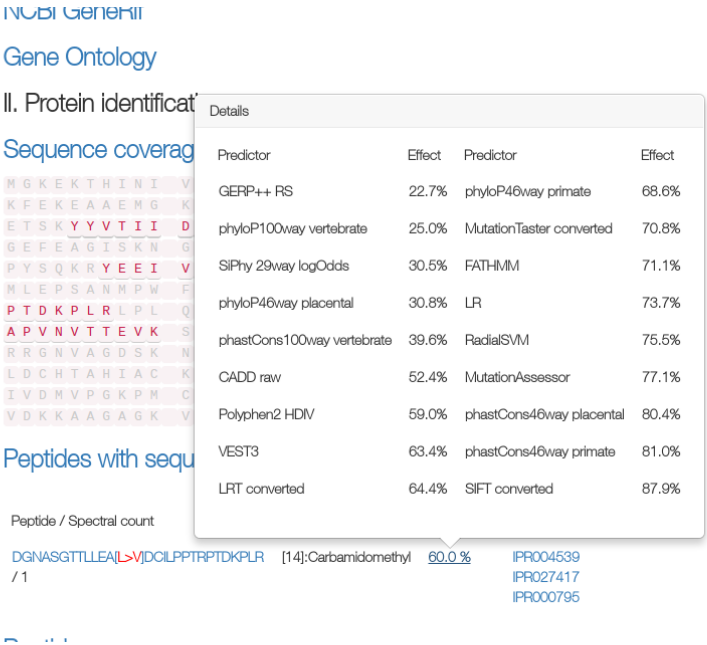
Overall view The protein view contains information relevant to particular protein and corresponding gene, external links and extracted information from several sources (UniProt, NCBI GeneRif, Gene Ontology).



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



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



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 > [GTGIGTAIGTCIGTT]

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

One could also see the details of identification of particular peptide from spectral match. This view contains additional mass-spectrometric data such as charge, *mz* and retention time. The E-Value column contains log10 of statistical significance of spectral match (X!Tandem, HyperScore). Further one could see alternative explanations of observed mass changes with respect to reference peptide. In this case, there are no other explanations known. The modifications are drawn from UniMod.

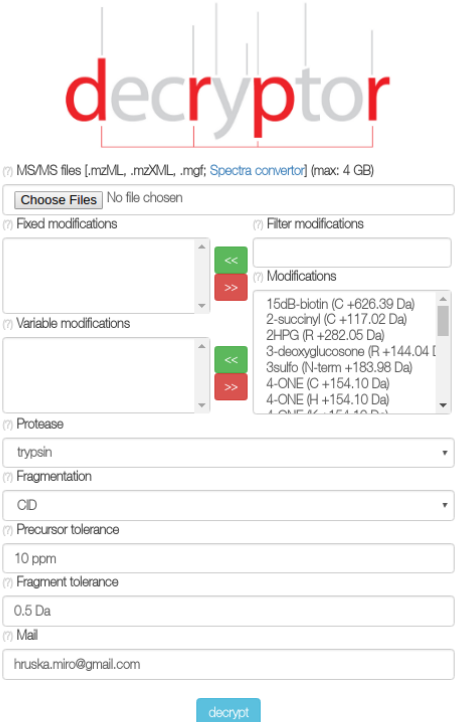
3 Task submission

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.



3.2 Selection of results

The results can be then accessed through the experiment list.