

# decryptor

system for reliable identification of point mutations in peptides

Miroslav Hruška, Jiří Voller, Laskhman Varanasi, Petr Džubák, Marián Hajdúch





Identification

Problems

decryptor

Results

### Identification

Problems

decryptor

Results



 $MS^1$  spectrum



# $MS^2$ spectrum



### Identification

### Problems

decryptor

Results

- $\mathbb{P}$  set of peptides
- $\mathbb{M}$  set of measurements

### Informal identification task

Given measurement  $m\in\mathbb{M},$  obtain peptide  $p\in\mathbb{P}$  which produced the measurement.

### Identification approaches

- database
  - peptide database
  - spectral database
- de novo
  - peptide tags
  - peptide sequencing

## Theoretical spectrum



### Identification

### Problems

decryptor

Results

- there are usually many peptides agreeing with the measurement to the same degree
- this prevents straightfoward identification of mutant peptides as there are usually other interpretations of spectrum<sup>1</sup>
- these problems follow from insufficiencies of predictive models and/or insufficiencies of agreement evaluation models

<sup>&</sup>lt;sup>1</sup>For example, peptides with PTMs.

## I/L-peptide pair spectral difference



### Insufficiencies in spectra similarity models



Identification

Problems

decryptor

Results

- system for reliable identification of alterations in standard bottom-up protemics data
- deployed on computational cluster
- publicly available: http://decryptor.imtm.cz

### Architecture





Interface



### Submission



(?) MS/MS files [.mzML, .mzXML, .mgf; Spectra convertor] (max: 4 GB)



### Experiment view — overview

### Experiment view

### Experiment: SILAC\_R2-05\_TR-B\_

ID: 9cd623b520780c97\_0000

Filter results Experimental meta-information Export results

### Identification

Summary: 299 proteins, 1473 ref. peptides, 6 non-ref. peptides, 2231 spectra

#### Proteins with sequence alterations



### Experiment view — mutations

#### Proteins with sequence alterations

¢ Protein	Spectra / Peptides / Unique peptides	¢ Alterations	Detrimental <sup>¢</sup> effect	DNA/mRNA alteration source	Disease * relevance
plastin 3 (PLS3)	7/7/7	326: N>S	0.75	COSMIC v.68 — large_intestine	Osteoporosis (OSTEOP)
heat shock 70kDa protein 8 (HSPA8)	6/3/3	19: V>L	0.61	COSMIC v.68 —kidney ICGC 15.1— KIRC-US	
glucan (1,4-alpha-), branching enzyme 1 (GBE1)	5/4/4	333: I>V	0.38		Adult polyglucosan body disease (APBD)
glucosidase, alpha; acid (GAA)	6/6/6	222: R>H	0.35		Glycogen storage disease 2 (GSD2)
dipeptidyl-peptidase 3 (DPP3)	13/9/9	677: R>H	0.34		
TNF receptor-	15/10/10	97: R>G	0.15		

### Experiment view — reference

#### **Reference proteins**

Protein	\$ Unique peptides	¢
heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa) (HSPA5)	111 / 10 / 10	
heat shock 70kDa protein 8 (HSPA8)	41/8/8	
heat shock 70kDa protein 9 (mortalin) (HSPA9)	23 / 12 / 12	
heterogeneous nuclear ribonucleoprotein M (HNRNPM)	19/11/11	
hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit (HADHA)	19/12/12	
moesin (MSN)	16/7/7	
protein disulfide isomerase family A, member 4 (PDIA4)	14/8/8	
phosphofructokinase, liver (PFKL)	13/10/10	
glycyl-tRNA synthetase (GARS)	12/7/7	
ribophorin I (RPN1)	11/10/10	
DEAD (Asp-Glu-Ala-Asp) box helicase 3, X-linked (DDX3X)	10/6/6	
NOP58 ribonucleoprotein (NOP58)	10/6/6	

Spectra / Pentides /

### Protein view — overview

### Protein view

<< Experiment view

#### I. Protein/Gene information

Protein: plastin 3

Protein ID: ENSP00000445339

Coverage: 17.09 %, isoform-specific peptides: 1

Synonyms: BMND18, T-plastin

External: MIM:300131, HGNC:HGNC:9091, Ensembl:ENSG00000102024, HPRD:02133, Vega:OTTHUMG00000022237, Entrez:5358

UniProt summary

Gene RefSeq

NCBI GeneRif

Gene Ontology

**II. Protein identification** 

Sequence coverage

### Protein view — coverage

Gene RefSeq

NCBI GeneRif

Gene Ontology

#### **II. Protein identification**

#### Sequence coverage

М					Y	K	۷	R	E	I			К		М					R	Ν	К			К	I	S	F	D	Е	F	۷	Y	I	F	Q	Е	۷	к
S			I	А	K	Т		R	К	A	I	Ν	R	К	Е		I		А				Т		Е							Q			Y				
К	Y	А		۷	Ν	W	I	Ν	Κ	A		Е	Ν					R	Н	V	I		М	Ν		Ν	Т					K	А	V				I	V
		K	М	I	Ν	L	S	۷	Ρ	D	Т	I	D	Е	R	А	I	Ν	К	Κ	K		Т			I	I			Ν		Ν		A		Ν		А	
A	I			Н	V	V	Ν			A				R	A		К		Н		V					W		I	I	Κ	I				A		I		
	R	Ν		А		А	А			R				Т		Е			М	Κ					Е				R	W	A	Ν	F	н	L	Е	Ν	S	G
W	Q	к	I	Ν	Ν			A		I	Κ			К	A	Y		Н			Ν	Q	I	A		K			К					R				Ν	М
			Ν		Т				К	R	А	Е		М				А		Κ				R			۷	Т		А		V	V			Ν	Р	К	L
Ν	L	А	F	۷	Α	N	L	F	Ν	K	Y	Ρ	А	L	т	к	Ρ	Е	Ν	Q	D	I	D	W	т	L	L	Е	G	Е	т	R			R	Т		R	Ν
W	Μ	Ν				V	Ν		Н	V	N			Y	A					٨									D		ĸ			V		W		К	V
Ν	K																																						
						K			Α	Ν	М	K	K		E	N		N	Y	A	V				K			A	K				v						
	N			Y		K T			A L	N A	M	K V	K W		E	N M	C	N R	Y Y	A T	V	EN	L V	G	K E	H	P L	A G	K D	F	s Q	L K	V A	G N	I D	G D	G	Q I	D V
L N	N W	D V	P G N	Y N R		K T L			A L A	N A G	M L K	K V S	K W T	L Q S	E L I	N M Q	C R S	N R F	Y Y K	A T D	V L K	E N T	L V I	GLS	× K E S	H D S	P L	A G A	K D V	F G V	S Q D	L K L	V A I	G N D	I D A	G D I	G I Q	Q I P	D V G
L N C	N W I	P D V N	P G N Y	Y N R D		K T L V	L S K		A L A G	N A G N	M L K L	K V S T	K W T E	L Q S D	E L I D	N M Q K	C R S H	N R F N	Y Y K N	A T D A	V L K	E N T Y	L V I A	GLSV	V K E S S	H D S M	P L L A	A G A R	K D V R	F G V I	S Q D G	L K L	V A I R	G N D V	I D A Y	G D I A	G I Q L	Q I P P	D V G E

#### Peptides with sequence alteration

Peptide / Spectral		Detrimental	Affected
count	modifications	offect	Domaine

Alternative mass

### Protein view — annotation

#### UniProt summary

Gene RefSeq

#### NCBI GeneRif

Text	\$	Pubmed id +	Timestamp \$
Loss of PLS3 is associated with spinal muscular atrophy.		24271012	2014-11-22 10:18
PLS3 expression and SMA phenotype: a commentary on correlation of PLS3 expression with disease severity in children with spinal muscular atrophy.		24284364	2014-07-26 10:33
$\ensuremath{PLS3}$ gene may have an age- and gender-specific role in the clinical severity of SMA in children afflicted with this condition.		24172247	2014-07-19 11:50
Plastin 3 (PLS3) appears to be important in human bone health, on the basis of pathogenic variants in PLS3 in five families with X-linked osteoporosis and osteoporotic fractures that we report here.		24088043	2013-10-26 10:59
study identified a common gene variant in PLS3 as an independent prognostic marker in female patients with stage II and III colon cancer		23549633	2013-09-28 11:20
PLS3 over-expression led to a stabilization of axons which, in turn, results in a significant delay of axon pruning, counteracting poor axonal connectivity in spina muscular atrophy neuromuscular junctions.	1	23263861	2013-08-31 10:14
Overexpression of PLS3 is associated with epithelial-mesenchymal transition and	d	23378342	2013-06-29

### Protein view — detrimental effect

### Protein view

<< Experiment view

#### I. Protein/Gene information

Protein: plastin 3

Protein ID: ENSP00000445339

Coverage: 17.09 %, isoform-specific peptides: 1

Syno <sup>2</sup> Exter	Details				2024. HPRD:0
Vega	Predictor	Effect	Predictor	Effect	
Unil	VEST3	24.2%	MutationAssessor	81.8%	
Ger	phastCons46way primate	47.3%	CADD raw	84.7%	
NCE	phyloP46way primate	48.1%	LRT converted	85.7%	
Ger	phyloP46way placental	55.9%	SIFT converted	87.9%	
OCI	SiPhy 29way logOdds	63.0%	GERP++ RS	89.5%	
II. P	MutationTaster converted	70.8%	FATHMM	95.7%	
Sog	phastCops100way vortebrata	71 406	phyloB100way vortabrata	06 906	

### Peptide view — details

## Peptide view

<< Protein view

#### Peptide sequence: LNLAFVASLFNK

#### Peptide-spectrum matches

sequence	modifications	charge	MZ	RT	XTandem -Log <sub>10</sub> E- Value	Alteration	Alternative intepretation of alteration
LNLAFVASLFNK		2	668.885193	4284.3406	6.318759	N>S (-27.0109 Da)	

Identification

Problems

decryptor

Results

## Filtering of results



## Identification of SNPs



### Identification based on RNA-Seq sequencing reads



- decryptor is publicly available system for identification of alterations in **standard** bottom-up proteomics data
- the system provides additional insights into sample of interest without additional requirements
- the results exhibit expected behaviour when compared to RNA sequencing of the corresponding sample

This work was supported by ELIXIR CZ research infrastructure project (MEYS Grant No: LM2015047) including access to computing and storage facilities.



Thank you for your attention.