



Application of **PECAN** for confident peptide detection directly from data-independent acquisition (DIA) MS/MS data

Ying Sonia Ting

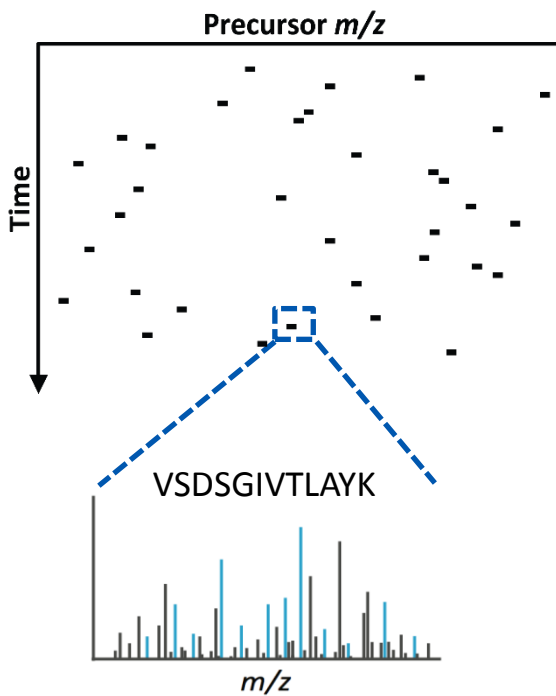
MacCoss Laboratory

Skyline User Group Meeting at ASMS 2015

What is Data-Independent Acquisition (DIA)?

DIA Combines the Merits of DDA and Targeted Acquisition at the Cost of Precursor Selectivity

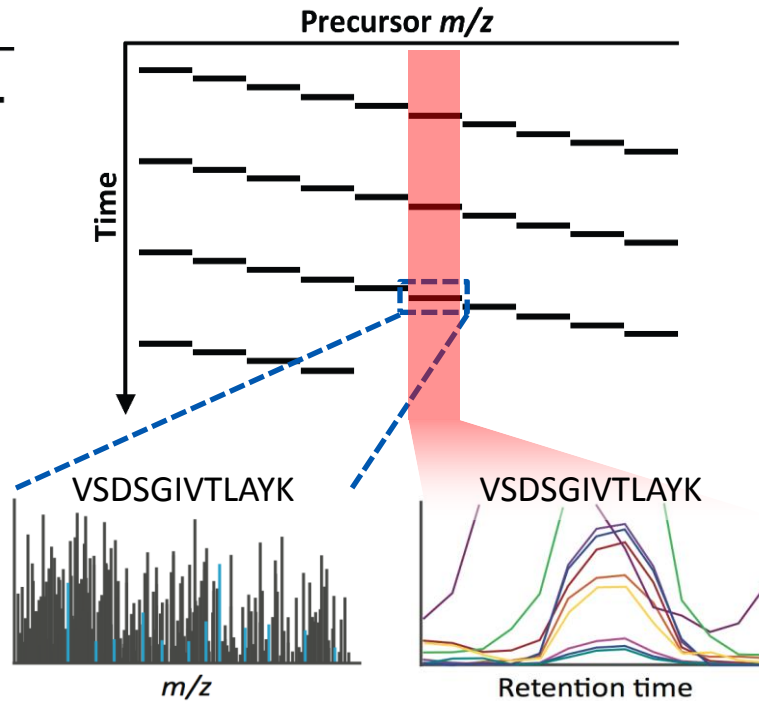
Data Dependent Acquisition (DDA)



Database Searching

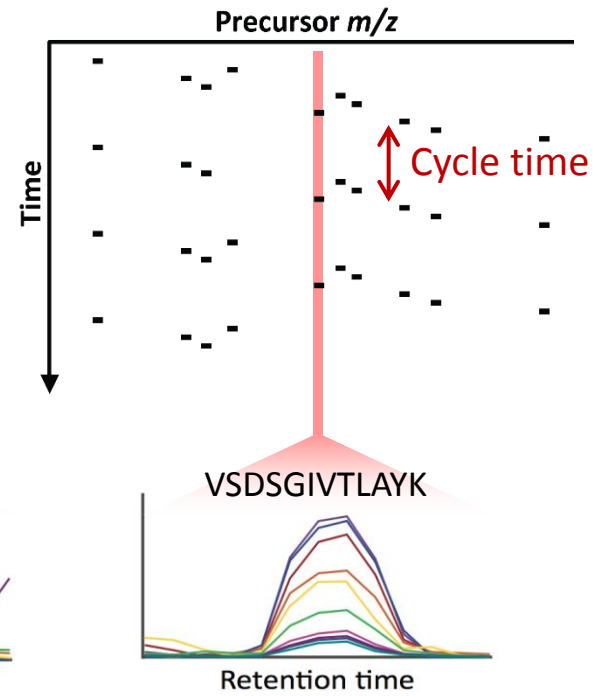
Peptide Identification

Data Independent Acquisition (DIA)



Retention time

Parallel Reaction Monitoring (PRM)



Retention time

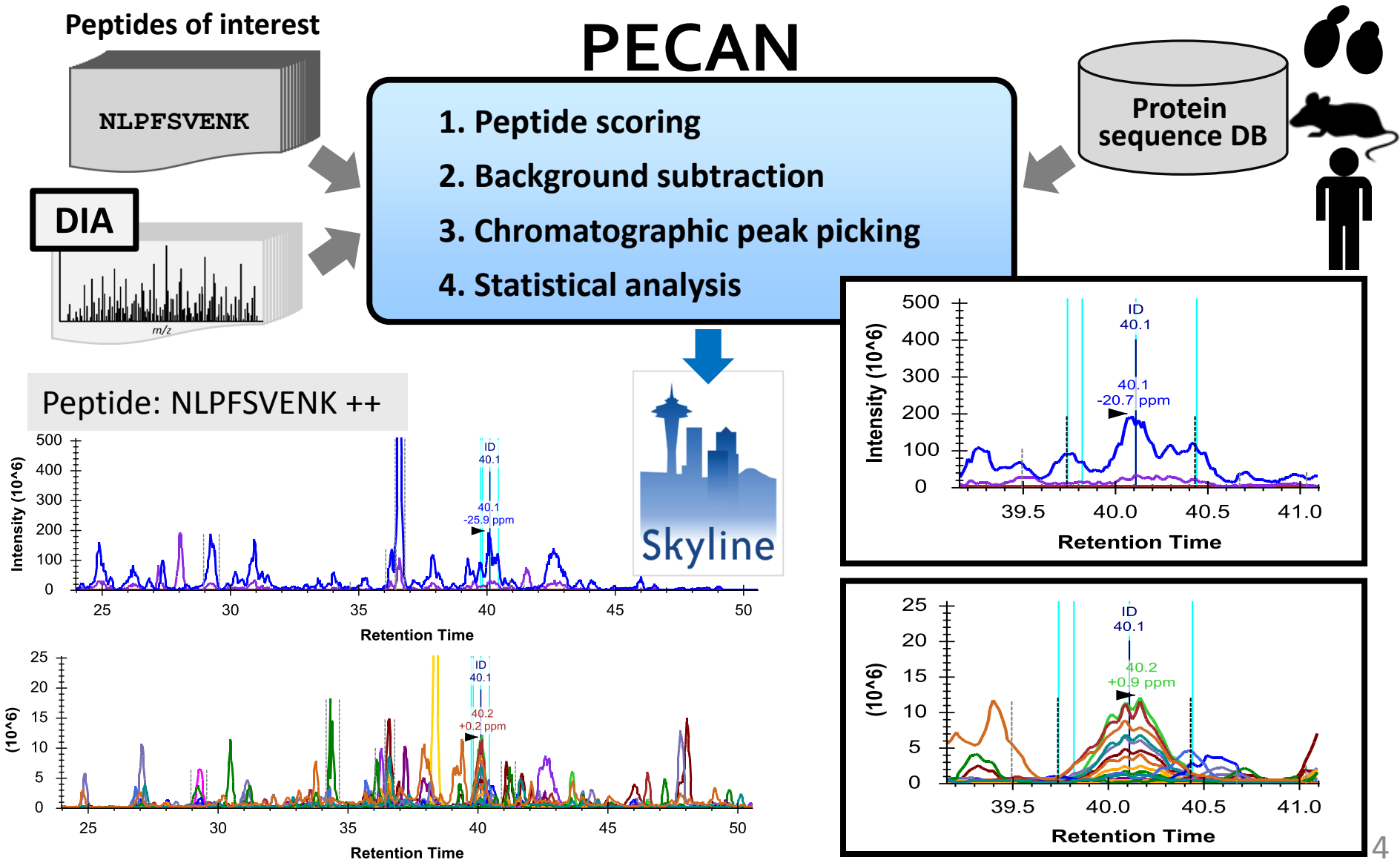
Targeted Analysis

Peptide Quantification

Can we confidently detect peptides *directly*
from DIA data?

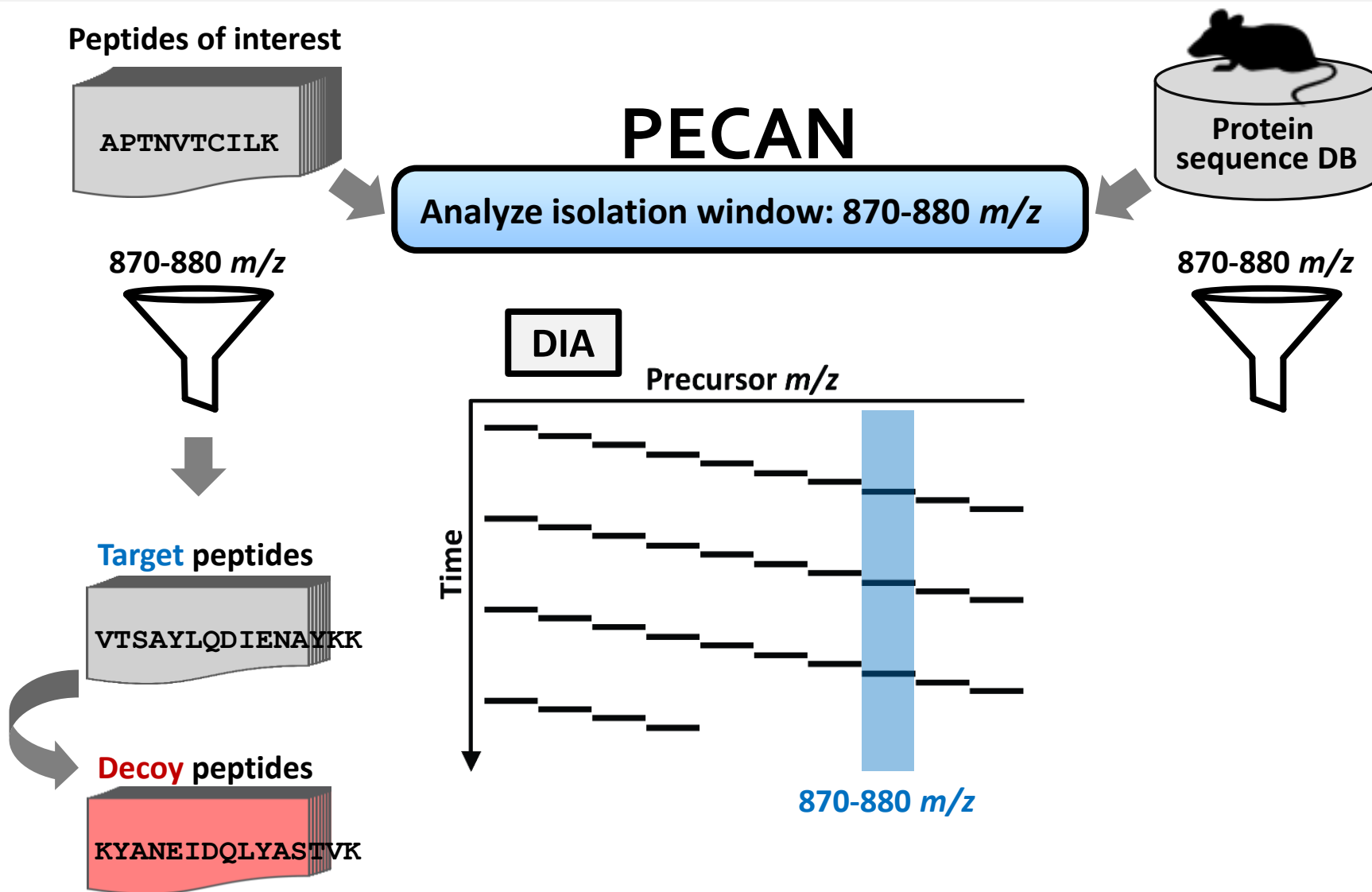
PECAN: Sequence-Based Peptide Detection Tool for DIA

without a prerequisite spectral/chromatogram library

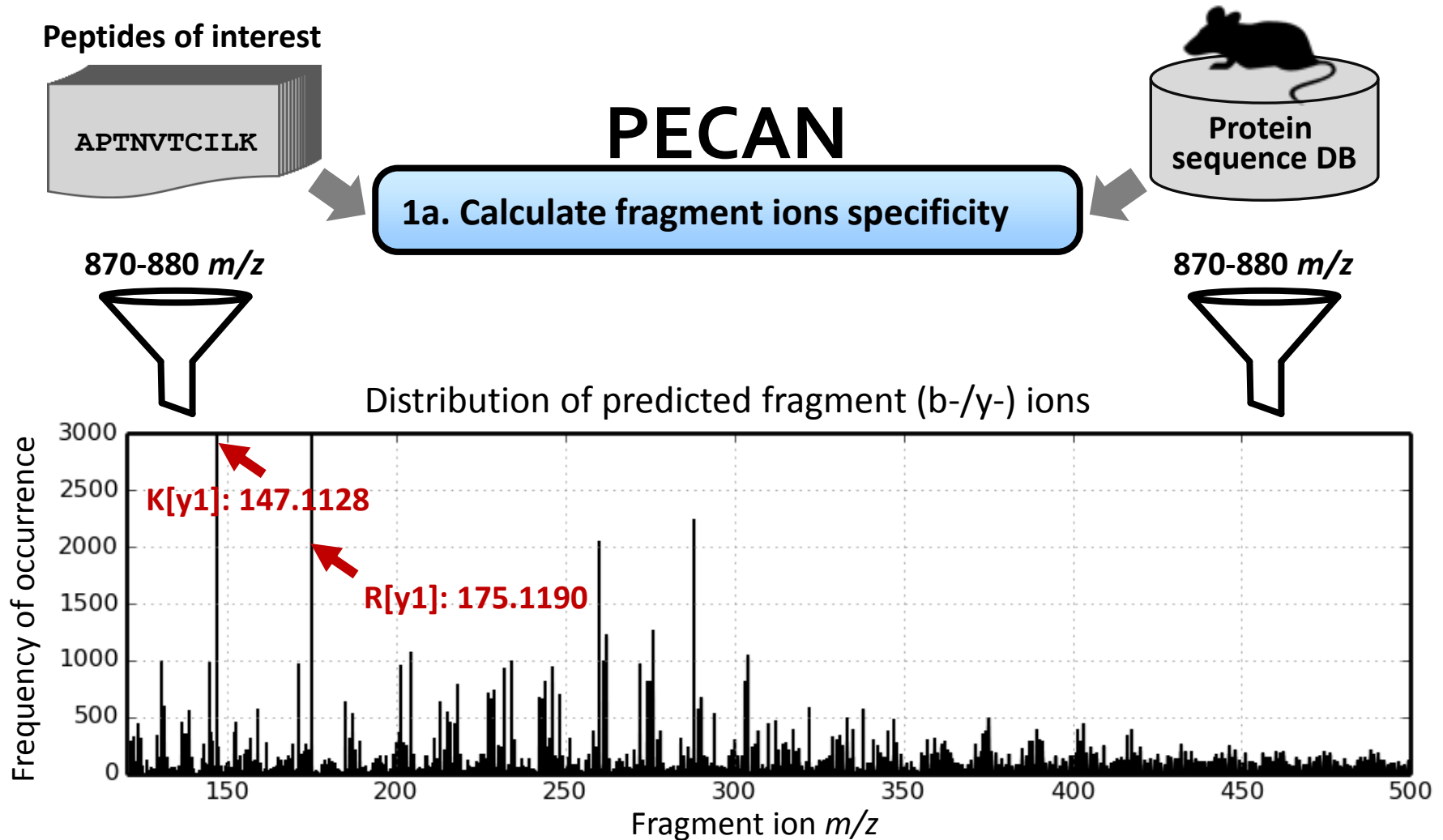


How does PECAN work?

PECAN First Analyzes each Isolation Window Independently



PECAN Step 1: Peptide Scoring

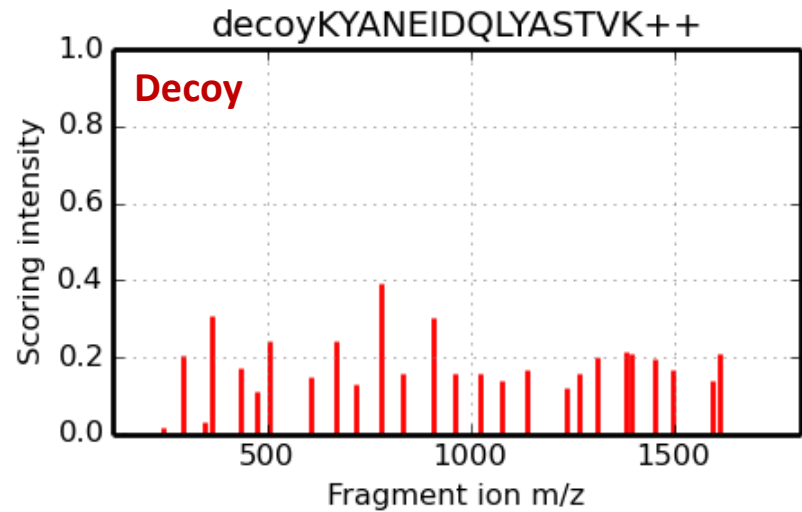
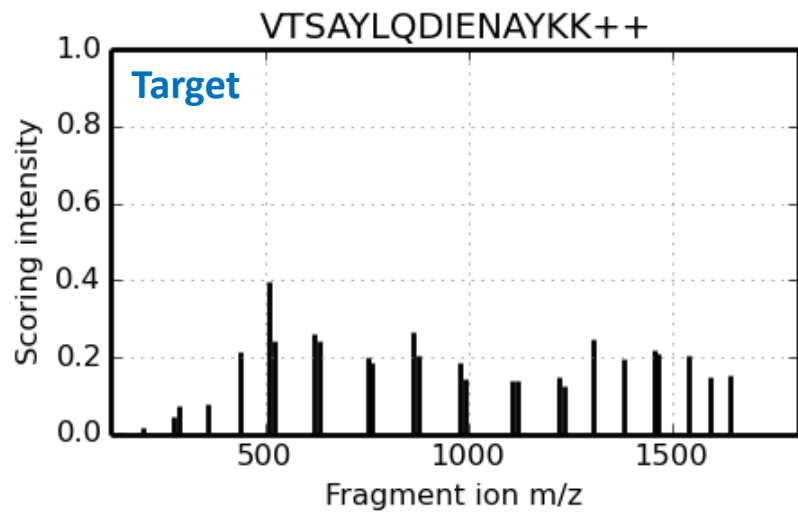


$$w_{mz} = \frac{100}{Occurr_{mz}}$$

PECAN Step 1: Peptide Scoring



PECAN generated spectra (scoring vectors)



$$w_{mz} = \frac{100}{Occurr_{mz}}$$

$$\vec{x}_{pep} = \frac{[w_{y1}, \dots, w_{yn}, w_{b2}, \dots, w_{bn}]}{\| [w_{y1}, \dots, w_{yn}, w_{b2}, \dots, w_{bn}] \|} \quad (\text{for all b-/y- ions})$$

PECAN Step 1: Peptide Scoring

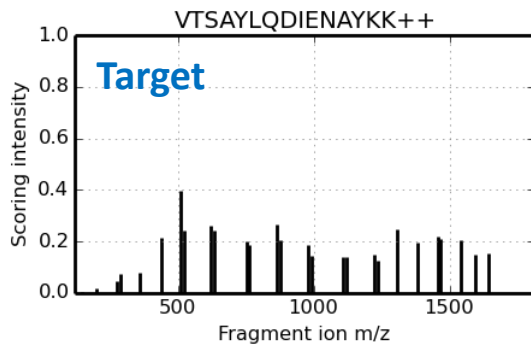
Peptides of interest

APTNVTCILK

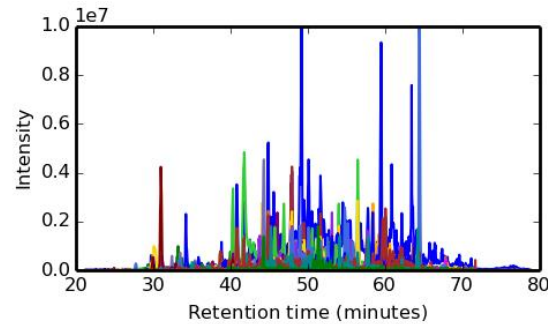
DIA

PECAN

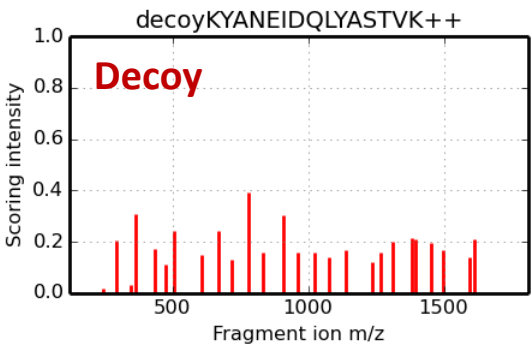
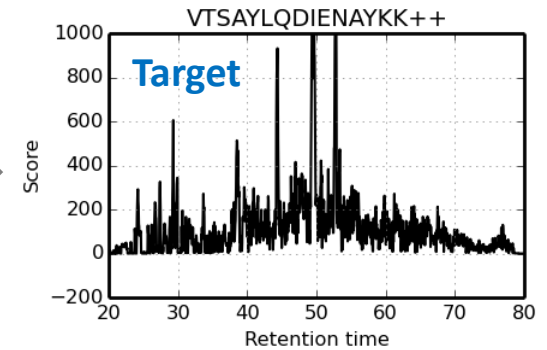
1c. Score with fragment ion XICs



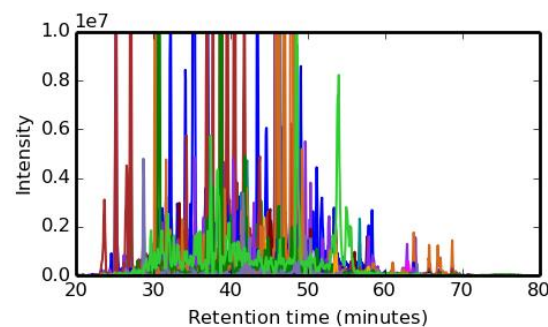
×



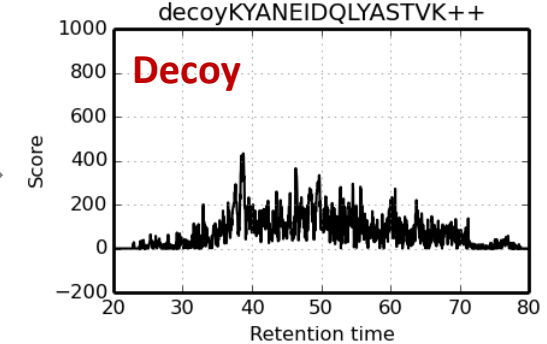
→



×



→



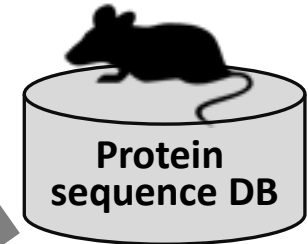
PECAN Step 2: Background Subtraction

Peptides of interest

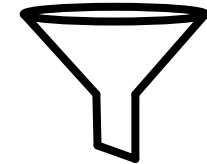


PECAN

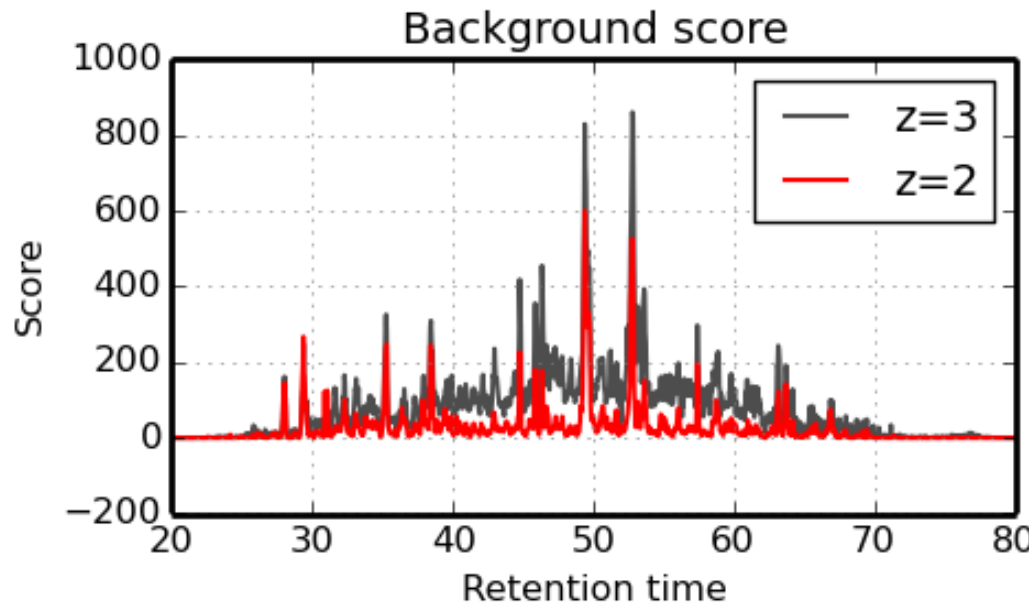
2a. Estimate background score



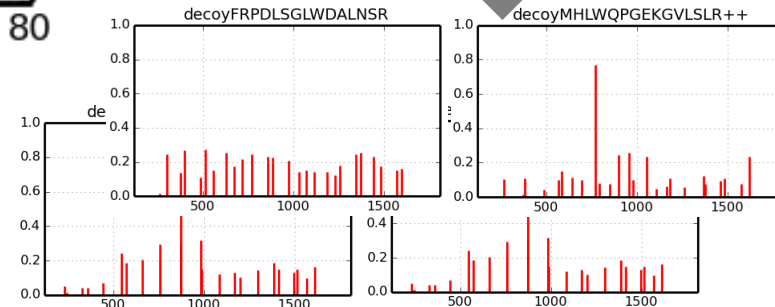
870-880 m/z



Thousands of randomly shuffled peptides



Scoring vectors



$$\vec{s}_{pep} = \vec{x}_{pep} \times \mathbf{XIC}_{pep}$$

$$\vec{b}_z = \frac{1}{m} \sum_{i=1}^m \vec{s}_{shuffled.pep} \quad z: \text{charge state}$$

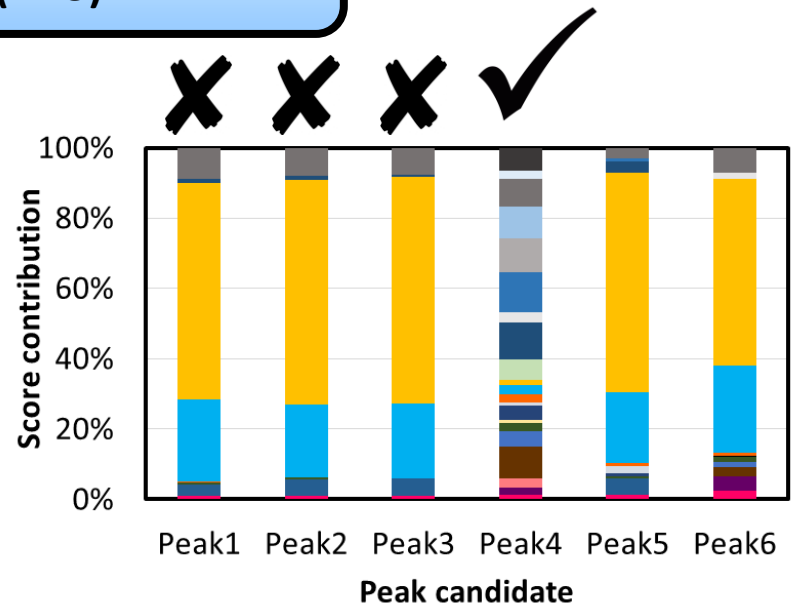
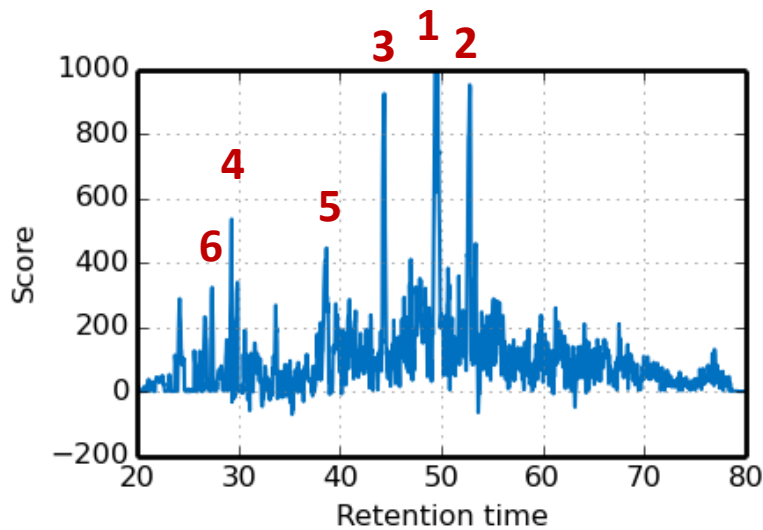
PECAN Step 3: Peak Picking

Peptides of interest



PECAN

3a. Pick the top N peaks ($N=3$)



$$\vec{s}_{pep} = \vec{x}_{pep} \times \mathbf{XIC}_{pep}$$

$$\vec{s}_{mz} = w_{mz} \times \overline{\vec{x}ic}_{mz} = w_{mz} \times [i_{mz,t_1}, i_{mz,t_2}, \dots, i_{mz,t_T}]$$

PECAN Step 3: Peak Picking

Peptides of interest



PECAN

3b. Generate features for each peak

Fragment ions

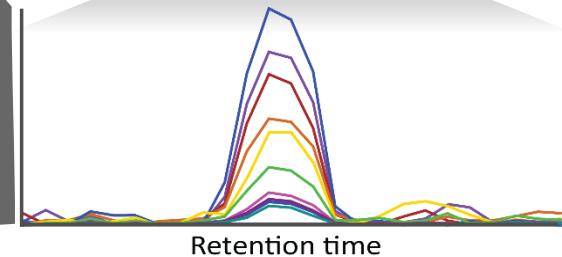
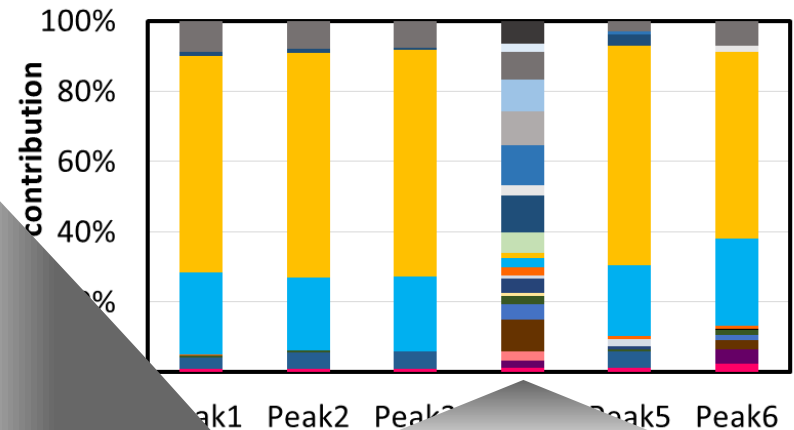
- raw score
- background subtracted score
- spectrum similarity
- number of contributing ions
- mass error mean
- mass error variance
- rank

Precursor ions

- idotp (isotope dot product)
- mass error mean
- mass error variance

Peptide sequence

- charge state
- peptide length
- SSR hydrophobicity index



PECAN Step 4: Statistical analysis

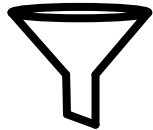
Peptides of interest



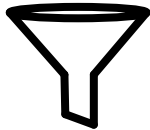
PECAN

4. Statistical analysis

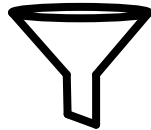
400-410m/z



410-420m/z

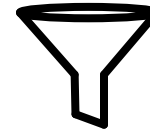


420-430m/z

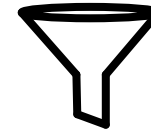


...

880-890m/z



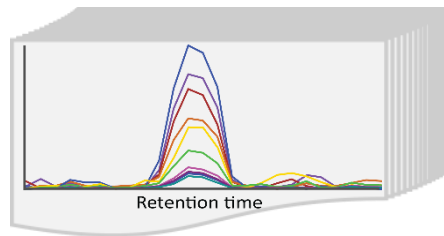
890-900m/z



Peptides of interest



Top N



Decoy peptides



Top N



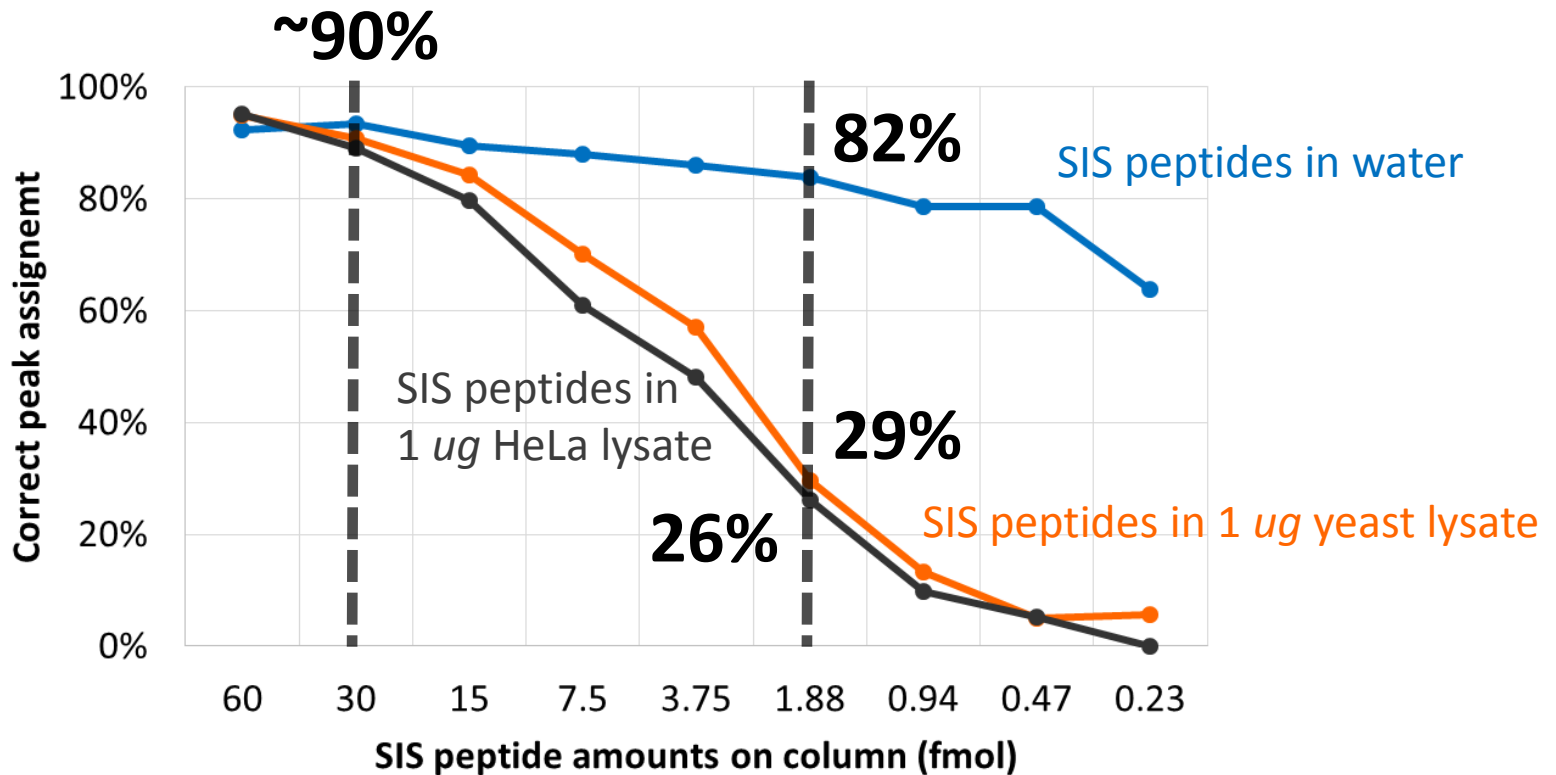
Percolator



p-value
q-value

Does PECAN work?

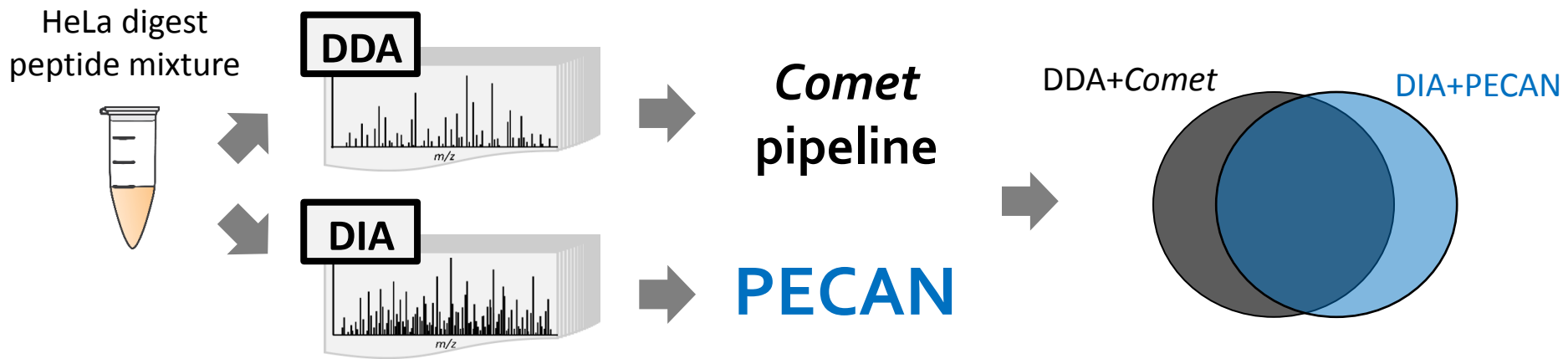
Chromatographic Peak Picking Accuracy of PECAN



SWATH Gold Standard dataset (OpenSWATH, Röst *et al.*, 2014)

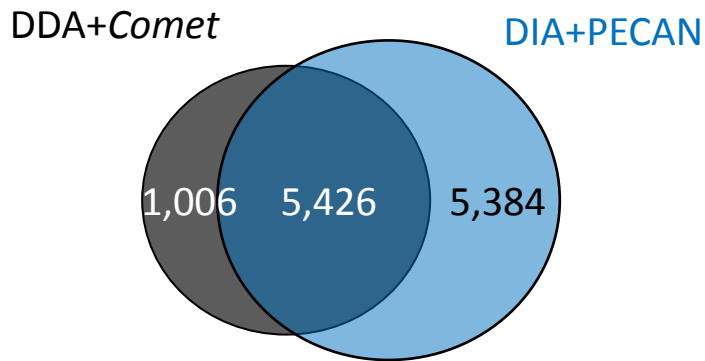
- **422 stable isotope-labeled standard (SIS)** peptides with 10 two-fold dilution steps
- SIS peptides spiked into water, yeast, or HeLa cell lysate
- DIA isolation window width: **25 m/z**

Peptide Detection Capability of PECAN

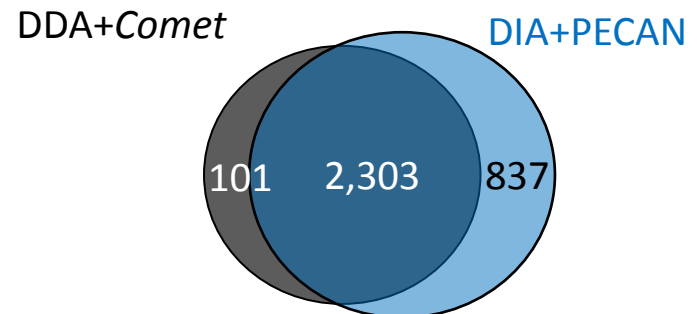


Query **8,207** GST-fusion-proteins from DNASU human cDNA plasmid library

Peptide comparison



Protein comparison



84 peptides from
PECAN only detection

SRM validation

82 peptides

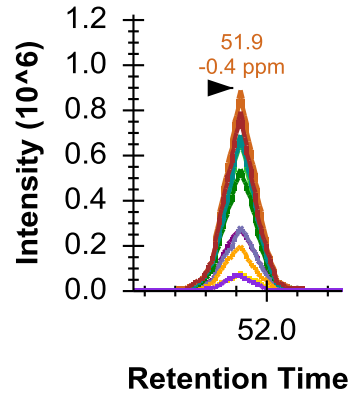
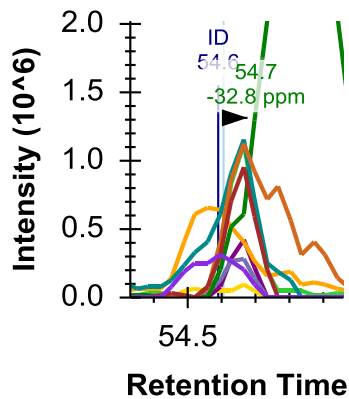
4xGPF covering 500-600,600-700,700-800,800-900 m/z range
Isolation window size: DDA 1.5 m/z -wide, DIA 5 m/z -wide

DIA+PECAN Specific Detection Validated by SRM from Analytical Standards

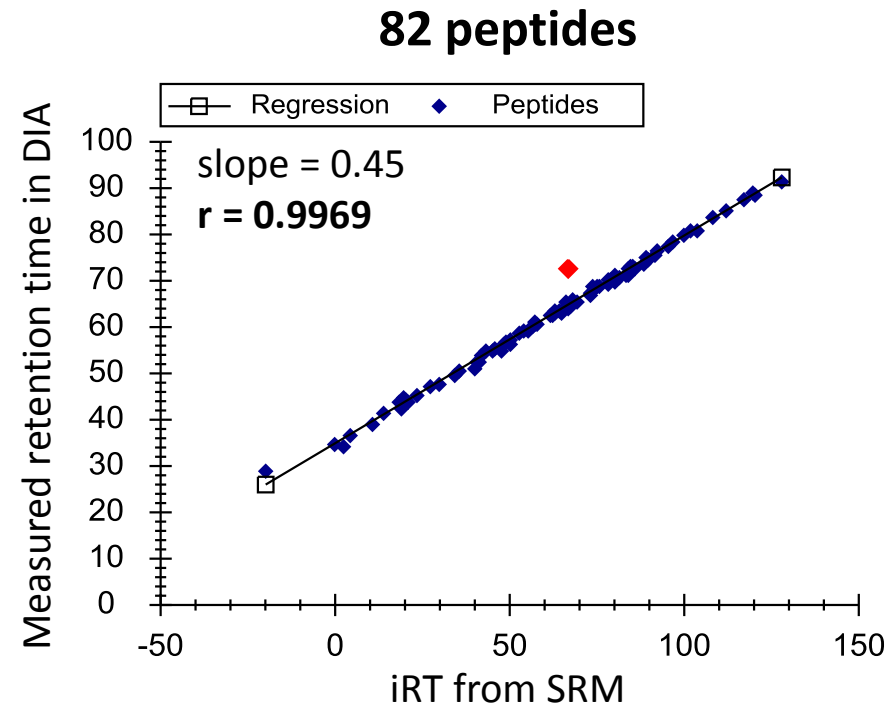
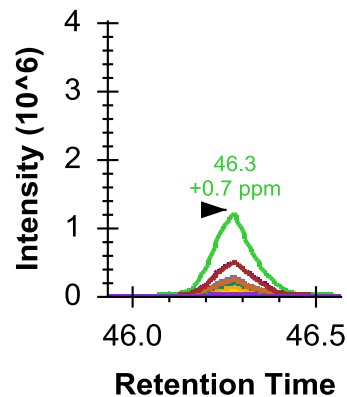
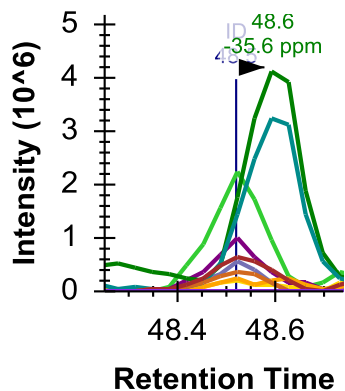
HeLa digest
5mz DIA

16 IVT proteins mixture
2mz PRM

K.LMVEESDLLGQLK.I [342, 354]



R.LIEEENMLAPSLK.Q [242, 254]



What can PECAN do?

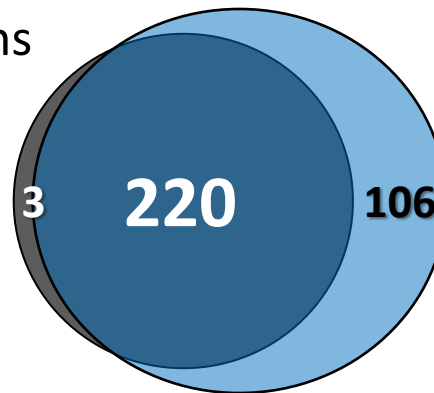
Application 1: What Fraction of the Mitochondrial Proteome Can be Detected?

Query 1,265 mouse mitochondrial proteins
from MitoP2 database
= 62 K peptide precursors



DDA
+ SEQUEST
223 mt proteins

DIA
+ PECAN
326 mt proteins



Proteins with at least one unique peptide(s)

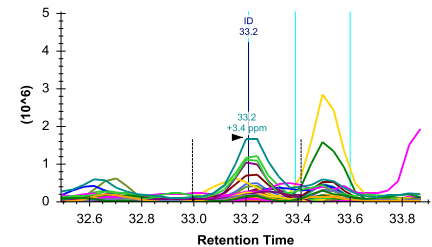
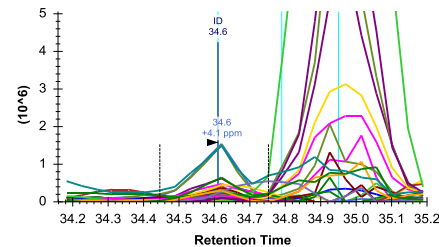
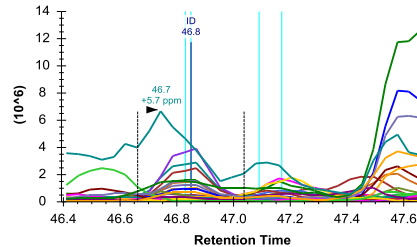
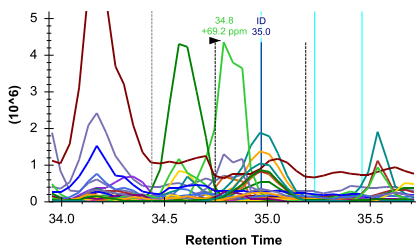
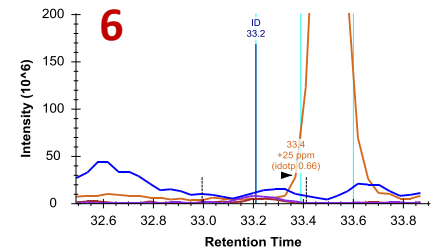
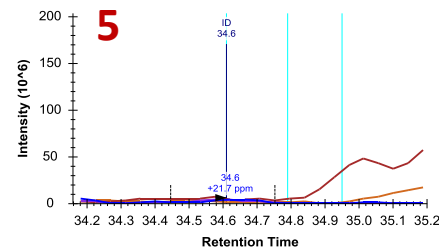
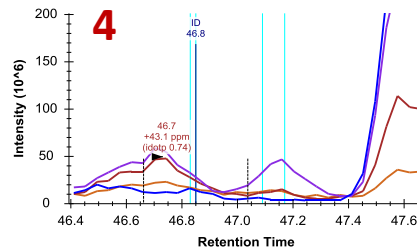
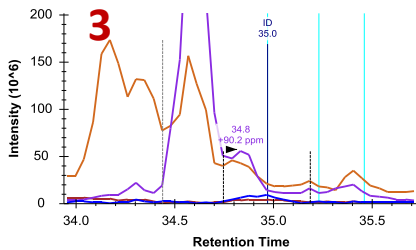
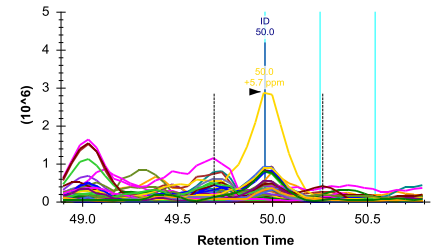
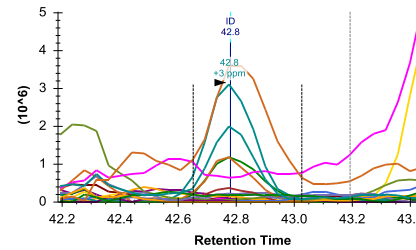
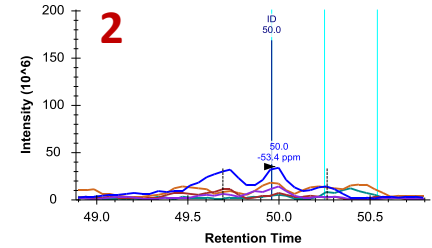
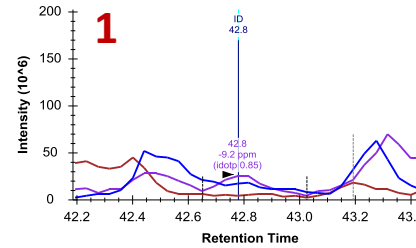
M. musculus heart proteome

- DDA covering 325-1450 m/z analyzed by SEQUEST (with Percolator)
- DIA (5 m/z -wide isolation windows) covering 400-900 m/z analyzed by PECAN (with Percolator)

Application 1: What Fraction of the Mitochondrial Proteome Can be Detected?

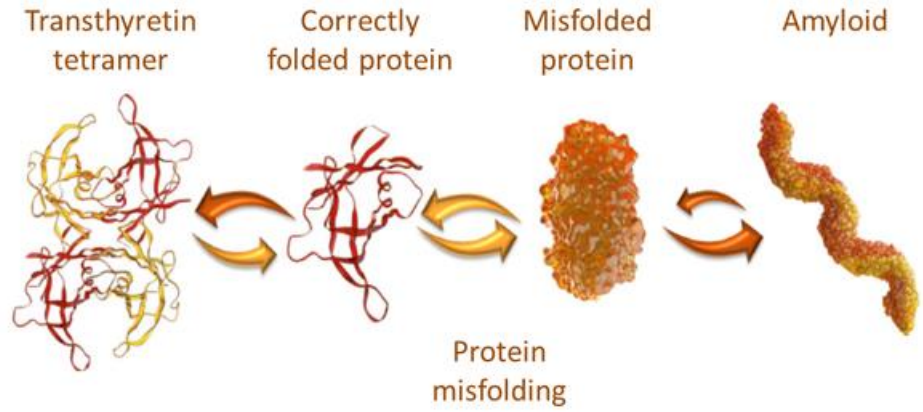
Q3U3J1: 2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial (*BCKDHA*)

10 20 30 40 50 60
 MAIVMSAAKI WRPSRGLRQA ALLLLGRSGV RGLARSHPSR QQQQQFPSLD DKPQFPGASA
 70 80 90 100 110 120
 EFVDKLEFIQ PNVISGIPIY RVMDRQGQII NPSEDPHLPQ EEVLKFYRSM TLLNTMDRIL
 130 140 150 160 170 180
 YESQRQGRIS FYMTNYGEEG THVGSAAALE RTDLVFGQYR EAGVLMYRDY PLELFMSQCY
 190 200 210 220 230 240
 GNVNDPGKGR QMPVHYGCKE RHFVTISSPL ATQIQAVGA AYAAKRANAN RIVICYFGEG
 250 260 270 280 290 400
 AASEGDAHAG FNFAATLECP IIFFCRNNGY AISTPTSEQY RGDGIAARGP GYGIMSIRVD
 310 320 330 340 350 360
GNDVFAVYNA TKEARRRAVA ENQPFLEIAM TYRIGHHSTS DDSSAYRSVD EVNYWDKODH
 370 380 390 400 410 420
PISRLRQYLL NQGWDEEQE KAWRKQSRKK VMEAFEQAER KLKPNPSSLF SDVYQEMPAQ
 430 440
 LRRQQESLAR HLOTYGEHYP LDHFDK



Application 2: Query Peptides from Disease Specific Alleles in Amyloidosis

(Han-Yin Yang)

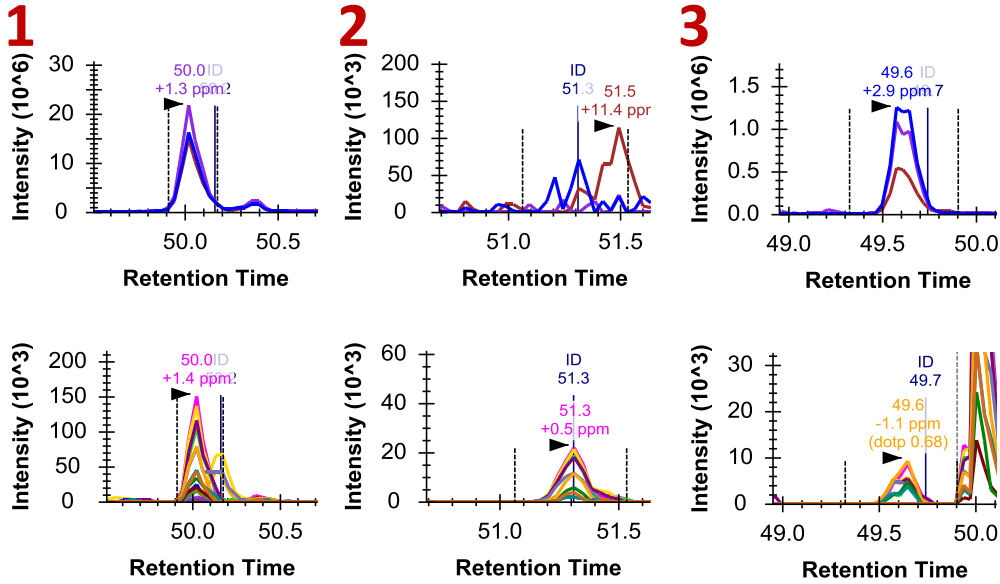


Transthyretin TTR

MASHRLLLLCLAGLVFVSEAGPTGTGESKCP
 LMVKVLD¹AVRGSPAINVAVHVFRKAADDTWE
 PFASGKTSESGELHGLTTEEEFVEGIYKVEI
 DTKSYWKALGISPFHEHAEVVFTANDSGPRR
 YTIAALLSPYSYSTTAVVTNPKE

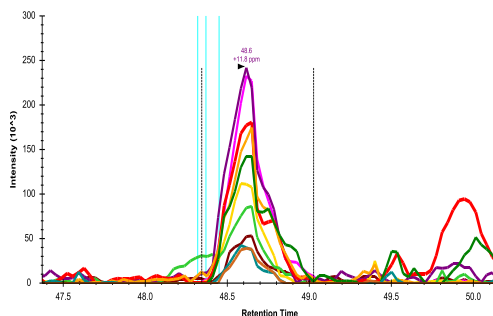
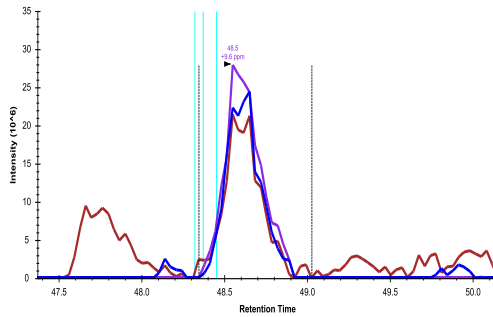
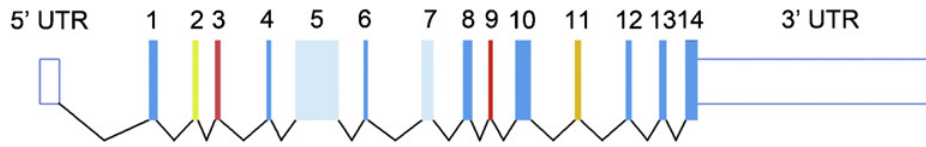
H → R TTR variance caused by an A>G

MASHRLLLLCLAGLVFVSEAGPTGTGESKCP
 LMVKVLD²AVRGSP³AINVAVHVFRKAADDTWE
 PFASGKTSESGELRGLTTEEEFVEGIYKVEI
 DTKSYWKALGISPFHEHAEVVFTANDSGPRR
 YTIAALLSPYSYSTTAVVTNPKE

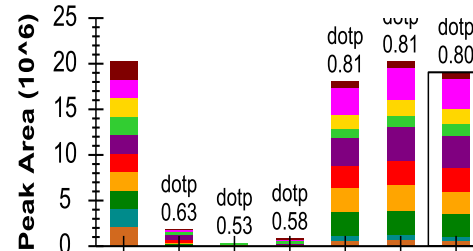
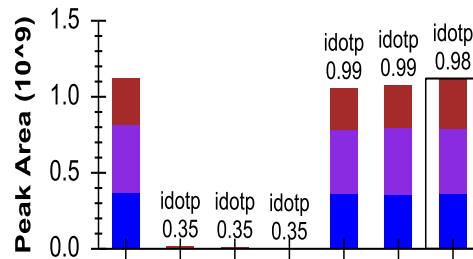


Application 3: Query Isoform Specific Peptides

Tau: Microtubule associated protein



Sample 1 Sample 2



Canonical Tau Sequence

10	20	30	40	50
MAEPRQEFEV	MEDHAGTYGL	GDRKDQGGYT	MHQDQEGDTD	AGLK
60	70	80	90	100
110	120	130	140	150
AEEAGIGD	TPSLEDEAAG	HVTQEPESGK	VVQ	
160	170	180	190	200
210	220	230	240	250
260	270	280	290	300
310	320	330	340	350
360	370	380	390	400
		ARMVS	KSKDGTGSDD	KKAKTSTRSS
410	420	430	440	450
AKTLKNRPLCL	SPKHPTPGSS	DPLIQPSSPA	VCPEPPSSPK	YVSSVTSRTG
460	470	480	490	500
SSGAKEMKLLK	GADGKTKIAT	PRGAAPPGQK	GQANATRIPA	KTPPAPKTPP
510	520	530	540	550
SSGEPKSGD	RSGYSSPGSP	GTPGSRSRTP	SLPTPTREP	KKVAVVRTPP
560	570	580	590	600
KSPSSAKSRL	QTAPVMPDL	KNVKSIGST	ENLKHQPGGG	KVQIINKKLD
610	620	630	640	650
LSNVQSKCGS	KDNIKHVPGG	GSVQIVYKPV	DLSKVTSKCG	SLGNIHHPKG
660	670	680	690	700
GGQVEVKSEK	LDFKDRVQSK	IGSLDNITHV	PGGGNKKIET	HKLTFRENAK
710	720	730	740	750
AKTDHGAEIV	YKSPVSGDT	SPRHLSNVSS	TGSIDMVDSP	QLATLADEV
ASLAKQGL				

Conclusions

About DIA

- Samples comprehensively
- Combines the merits of DDA and targeted acquisitions
- Often compromise precursor selectivity

About PECAN

- Confident peptide detection for DIA data
- No prerequisite spectral/chromatogram library
- Detects peptides with target-decoy FDR control
- Enables both discovery and targeted data analysis

Acknowledgements

University of Washington

Seattle, Washington

Michael MacCoss

Jarrett Egertson

James Bollinger

Han-Yin Yang

Brendan MacLean

Richard Johnson

Gennifer Merrihew

Peter Rabinovitch

William Noble

Jimmy Eng

Andrew Stergachis

Thermo Fisher Scientific, Inc.

San Jose, California

Romain Huguet

Michael Blank

Philip Remes

Derek Bailey

Graeme McAlister

Vlad Zabrouskov

Pacific Northwest National

Laboratory

Richland, Washington

Samuel Payne

Sangtae Kim