Proteomics as a tool for biomarker discovery

Katarina Davalieva

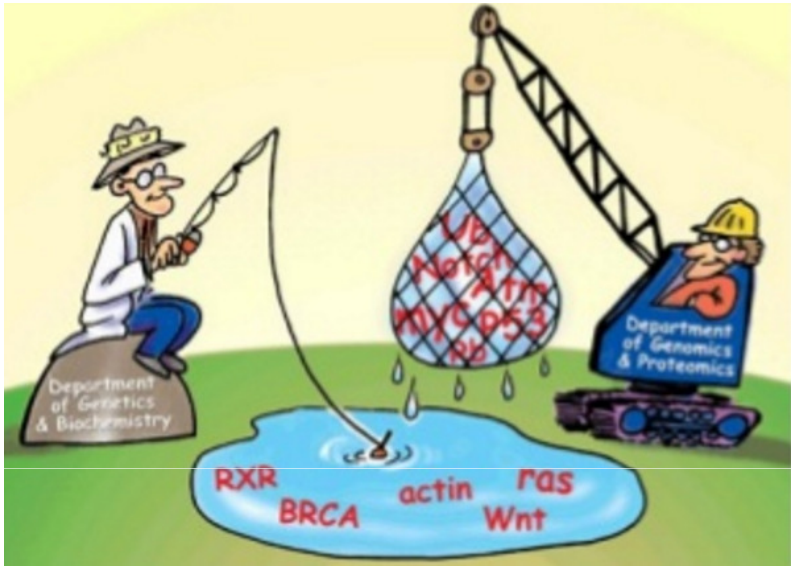
Research Centre for Genetic Engineering and Biotechnology “Georgi D Efremov”, Macedonian Academy of Sciences and Arts

What is PROTEOME and PROTEOMICS?

"PROTEOME“ =

PROTEin complement to a gen**OME**
(Marc Wilkins in 1994)

The full set of proteins comprising the structural, metabolic, and regulatory machinery of a cell, tissue, or organism



PROTEOMICS (1997)

Proteomics is a large-scale comprehensive study of a specific proteome, including information on protein abundances, their variations and modifications, along with their interacting partners and networks, in order to understand cellular processes.

Why PROTEOMICS?
Unique views vs GENETICS/GENOMICS

➢While genes are the "recipes" of the cell, **proteins are ultimately the functional players** that drive both normal and disease physiology

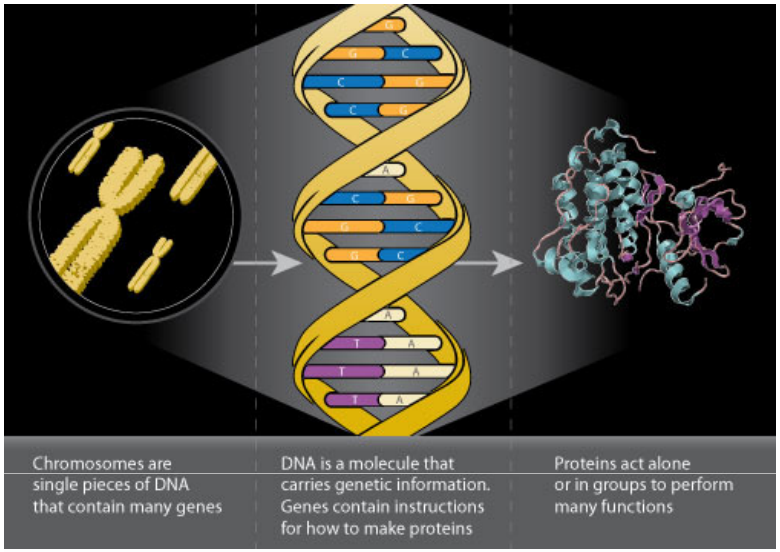
➢**Protein abundance and function is modulated at several levels** from transcription to post-translation and very little of this can be predicted from a simple analysis of nucleic acids alone

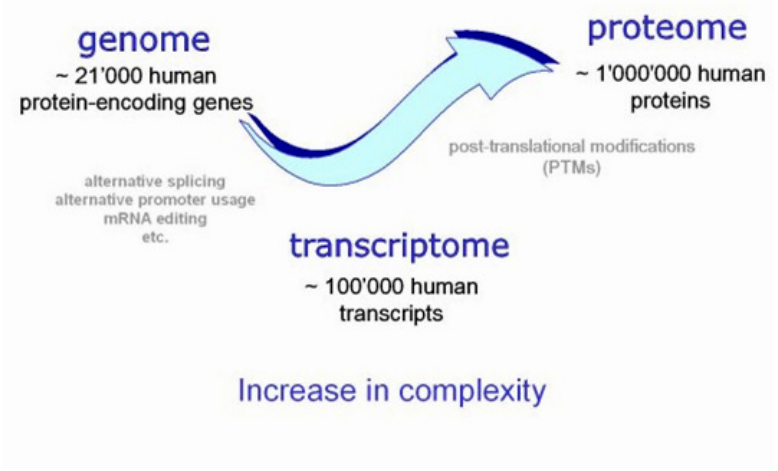
➢**Many proteins form complexes** with other proteins or RNA molecules, and only function in the presence of these other molecules

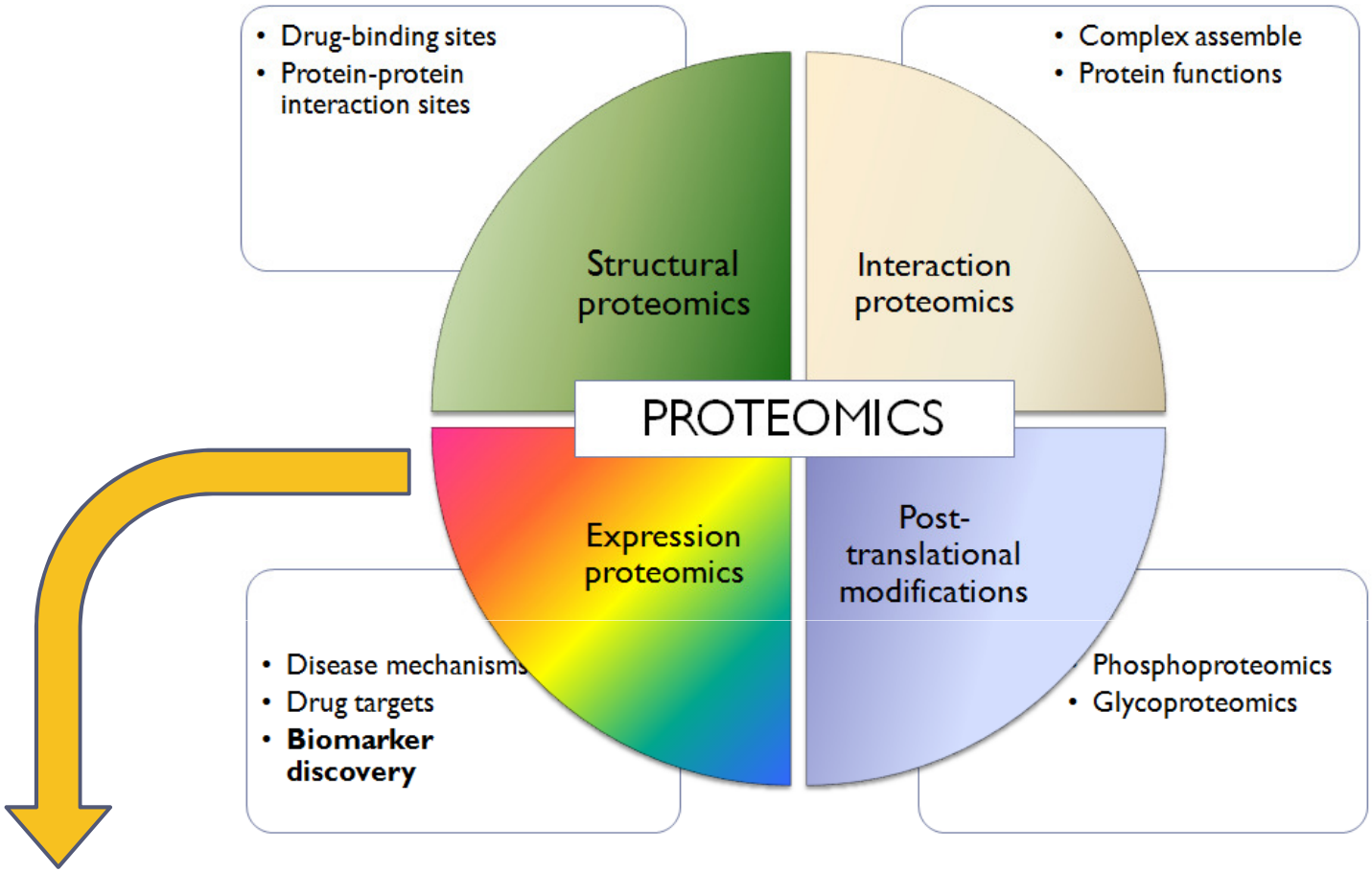
➢**One gene can encode more than one protein** through alternative splicing or alternative post-translational modifications

➢**Post-translational modifications** (phosphorilation, glycosylation, acetylation) that profoundly affect protein activities and function

➢**Proteins concentration poorly correlate with mRNA abundance**, transcribed from DNA





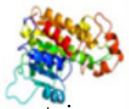


Clinical proteomics is a sub-discipline of proteomics that involves the application of proteomic technologies on clinical specimens to identify unique bio-signatures or biomarkers responsible for the diagnosis, prognosis and therapeutic prediction of such disease.

Biomarkers are biological molecules found in blood, other body fluids, or tissues that are a sign of a normal or abnormal process, or of a condition or disease. They may also be used to see how well the body responds to a treatment for a disease or condition.

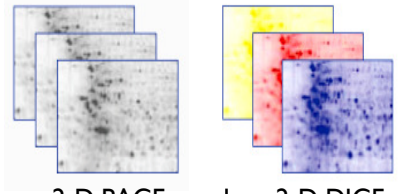
Principle of proteomic technologies

2-D gel electrophesis/MS

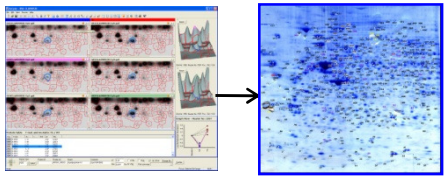


proteins

Label or label-free gel separation

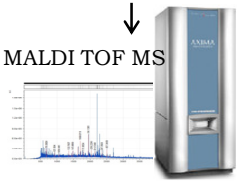


Quantification of protein expression



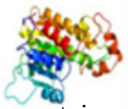
Selection of POI

MALDI TOF MS



Identification of proteins by MS

Shotgun LC-MS/MS



proteins

digestion

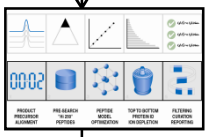
peptides

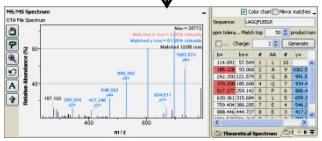
Label or label-free
Multidimensional separation

HPLC

MS

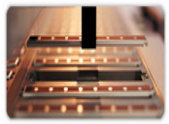
MS software





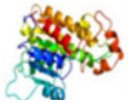
Protein identification and quantification

MS profiling

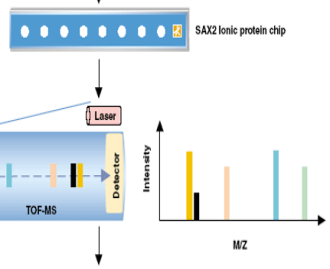


SELDI chip

MALDI plate



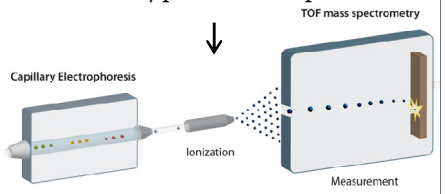
proteins



Biomarker(s) identification
based on pattern differences

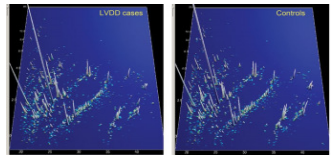
CE-MS

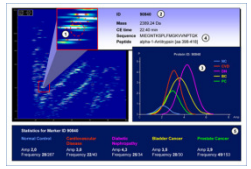
Urine/plasma sample



Capilary electrophoresis MS

Data evaluation



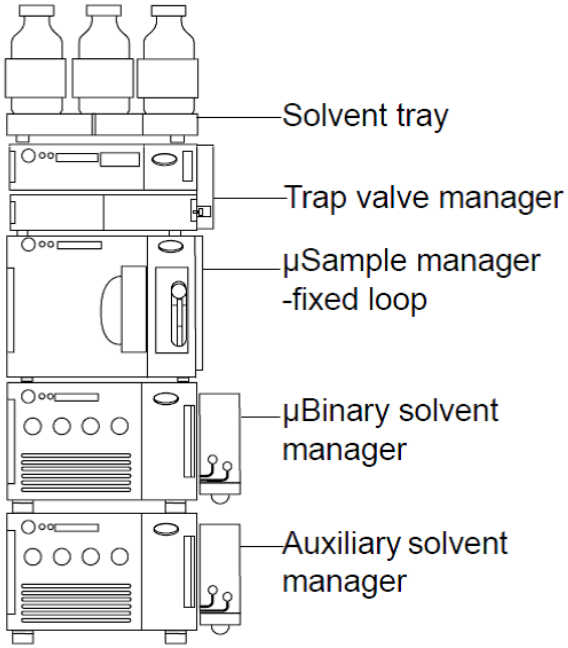


Biomarker panel identification

ACQUITY UPLC M-Class/ Synapt G2-Si
System introduction

Ultra Performance Liquid Chromatography (UPLC) + hybrid quadrupole/orthogonal acceleration, time-of-flight (oa-TOF) mass spectrometer

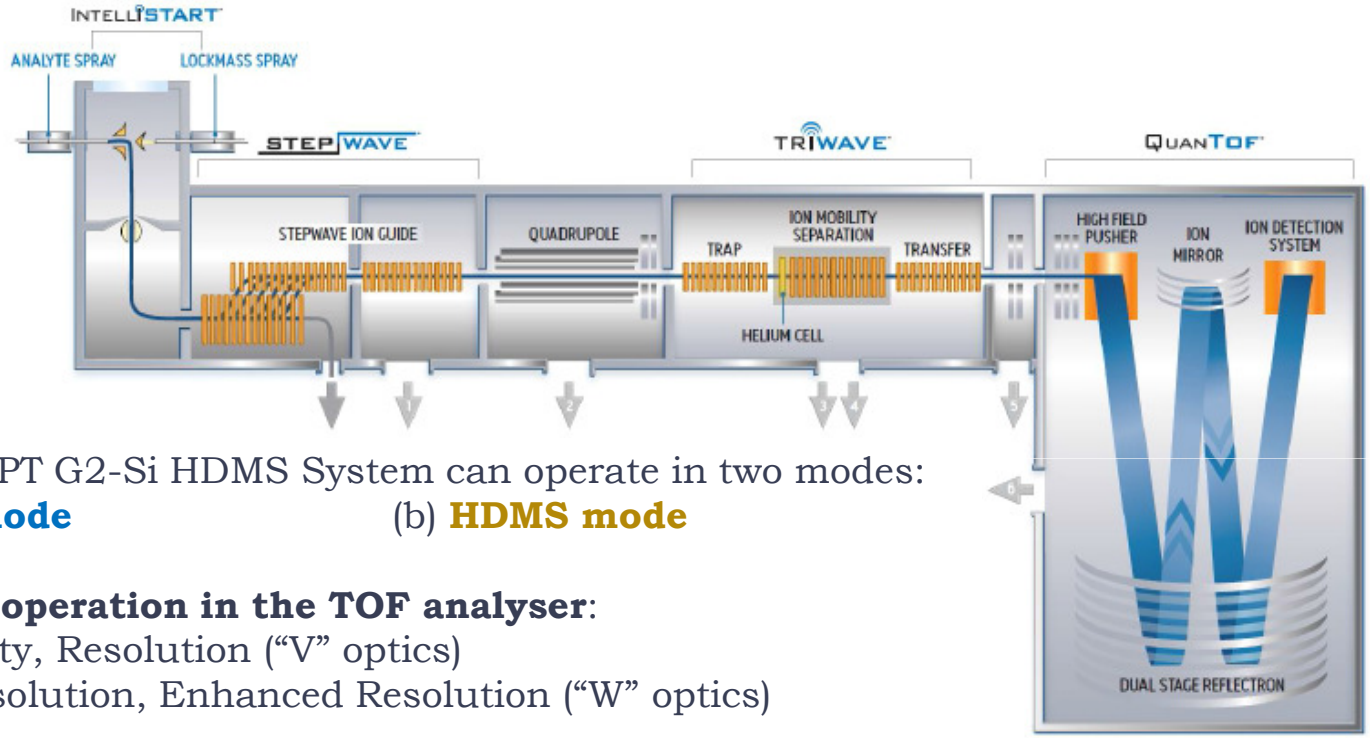
- Flow rates from 200 nL/min to 100 µL/min
- Max operation pressure 1034 bar (15,000 psi)



ACQUITY UPLC M-Class system

Synapt G2-Si

Synapt G2-Si overview



The SYNAPT G2-Si HDMS System can operate in two modes:
(a) **TOF mode** (b) **HDMS mode**

Modes of operation in the TOF analyser:
- Sensitivity, Resolution (“V” optics)
- High Resolution, Enhanced Resolution (“W” optics)

TOF mass range:	Sensitivity & Resolution	20-100000 m/z
	High Resolution	20- 32000 m/z
Mass resolution:	50,000 at 1000 m/z	
	35,000 at 3500 m/z	
LLOQ:	25 fg	
Mass accuracy:	< 1ppm	
Dynamic Range:	4 orders of magnitude	

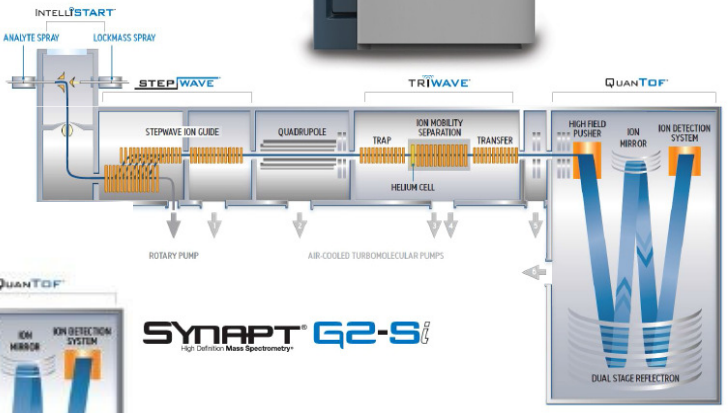
Improved sensitivity (x30)
Improved LOQ
(x10)

Improved resolution

2006

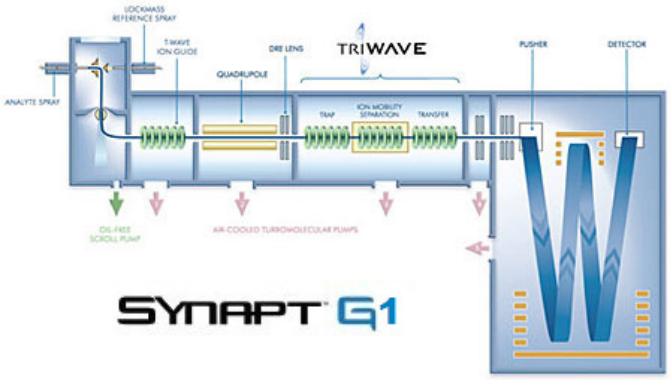
2009

2011



SYNAPT G2-S

SYNAPT G2



SYNAPT G1

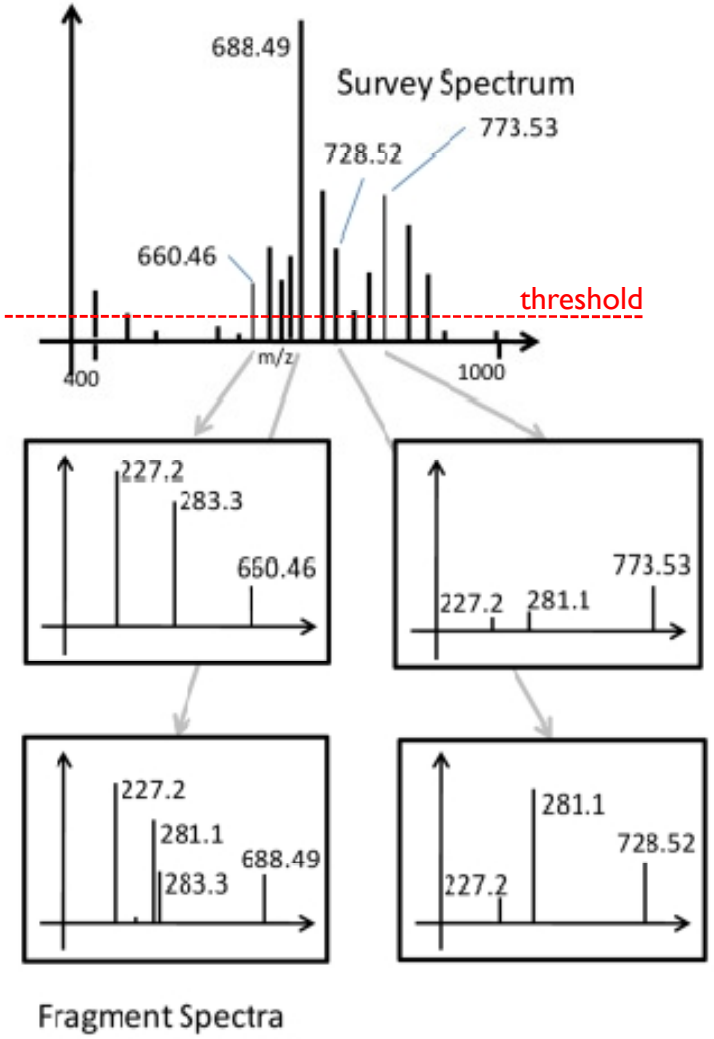
Why SYNAPT G2-Si?

- Powerful data independent & data dependent solutions
- Superior separation with high-efficiency T-Wave IMS
- Wide range of experimental options

Data dependent (DDA) acquisition

DDA principles
The most abundant ionized species are selected for subsequent isolation and tandem mass analysis

DDA limitations:
(1) restricted dynamic range of the proteome due to threshold minimum requirement for ion selection
(2) non-reproducible nature of peptides detected in replicate analyses of the same sample



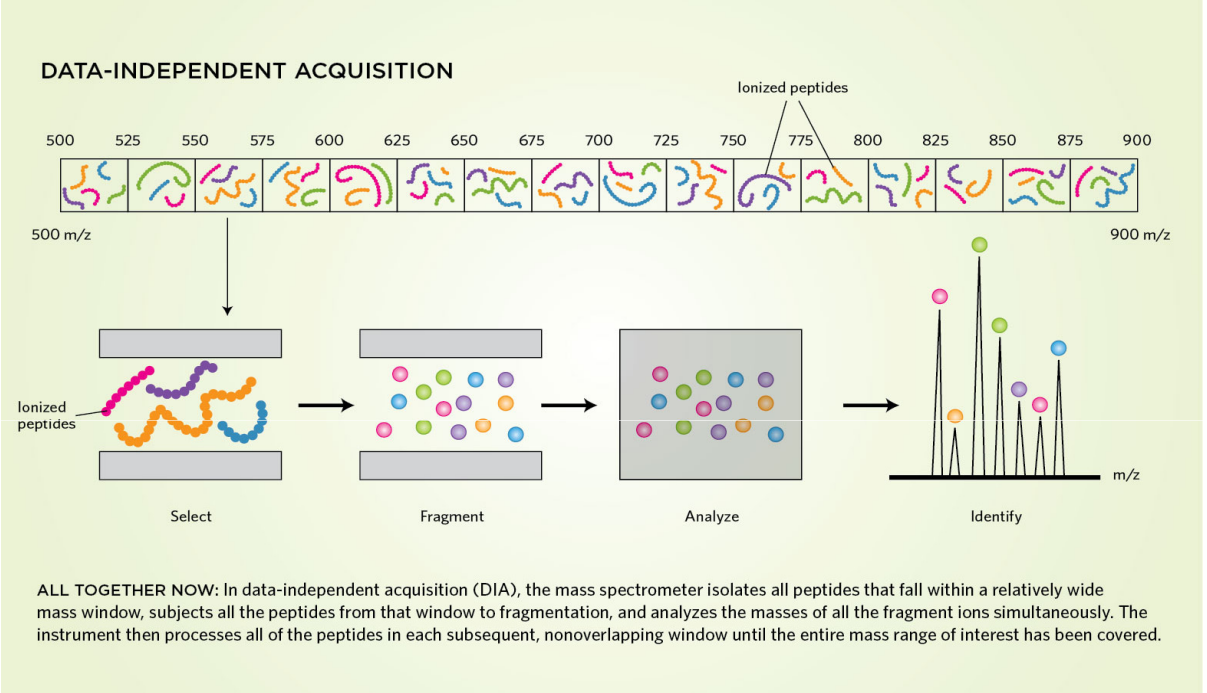
Data independent (DIA) acquisition

DIA principles
Do not use real-time ion selection based on precursor ion scans to obtain searchable spectra for peptide identification.

Predetermined m/z ranges are interrogated either:

(1) by fragmenting all ions entering the MS at a single point in chromatographic time

(2) by dividing the full m/z range into fixed smaller m/z isolation windows that are each independently and consecutively analyzed

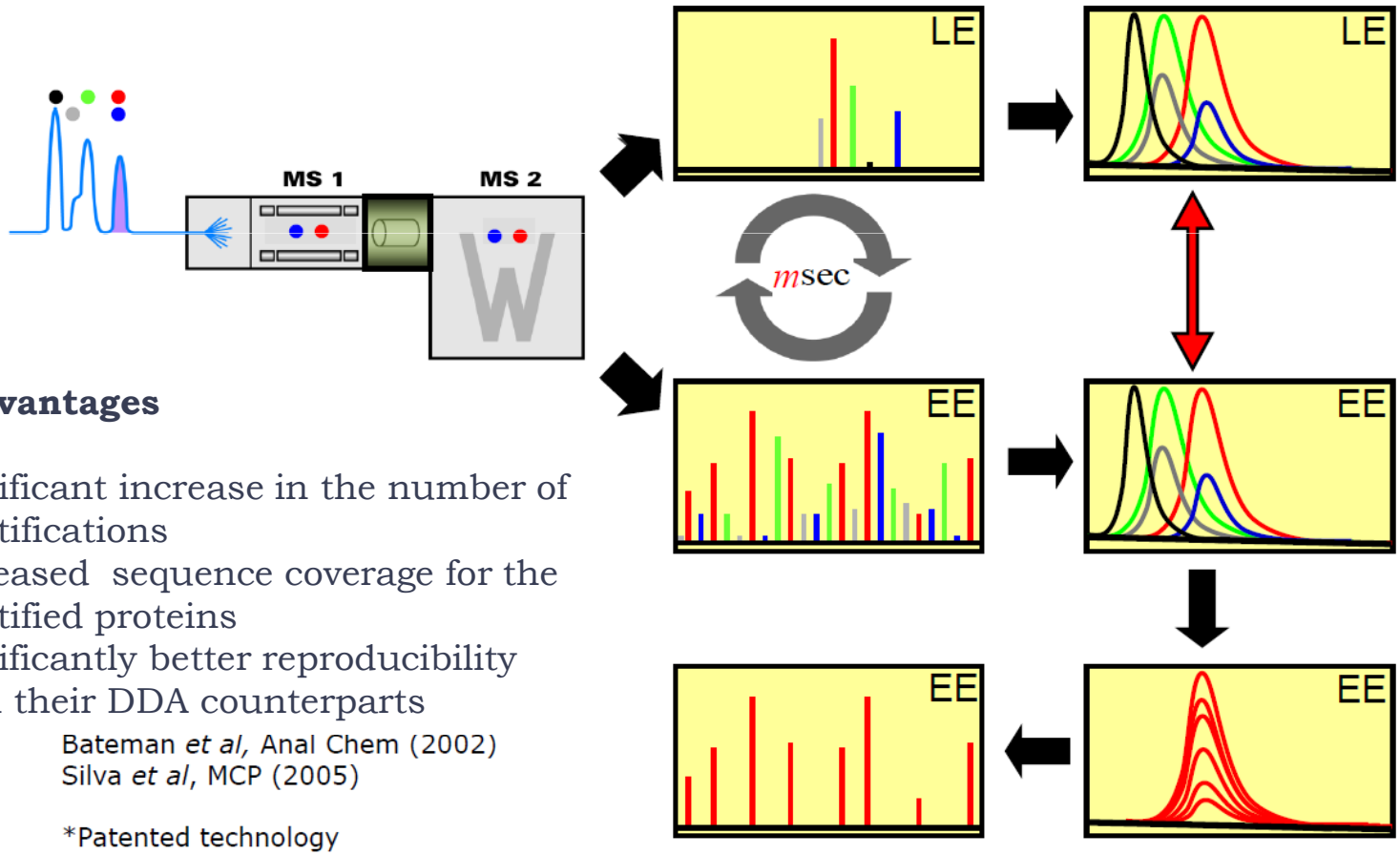


DIA advantages:
Wider detectable dynamic range of the proteome and increased accuracy of protein quantification

Waters Data Independent Acquisition (DIA) = MS^E

MS^E principles

Utilizes an LC/MS-TOF and cycled the collision cell energy voltage between a low energy to capture MS information, and a high energy to fragment all incoming precursor ions (MSMS). All precursor and fragment ion information could be captured at multiple points during the elution of a chromatographic peak

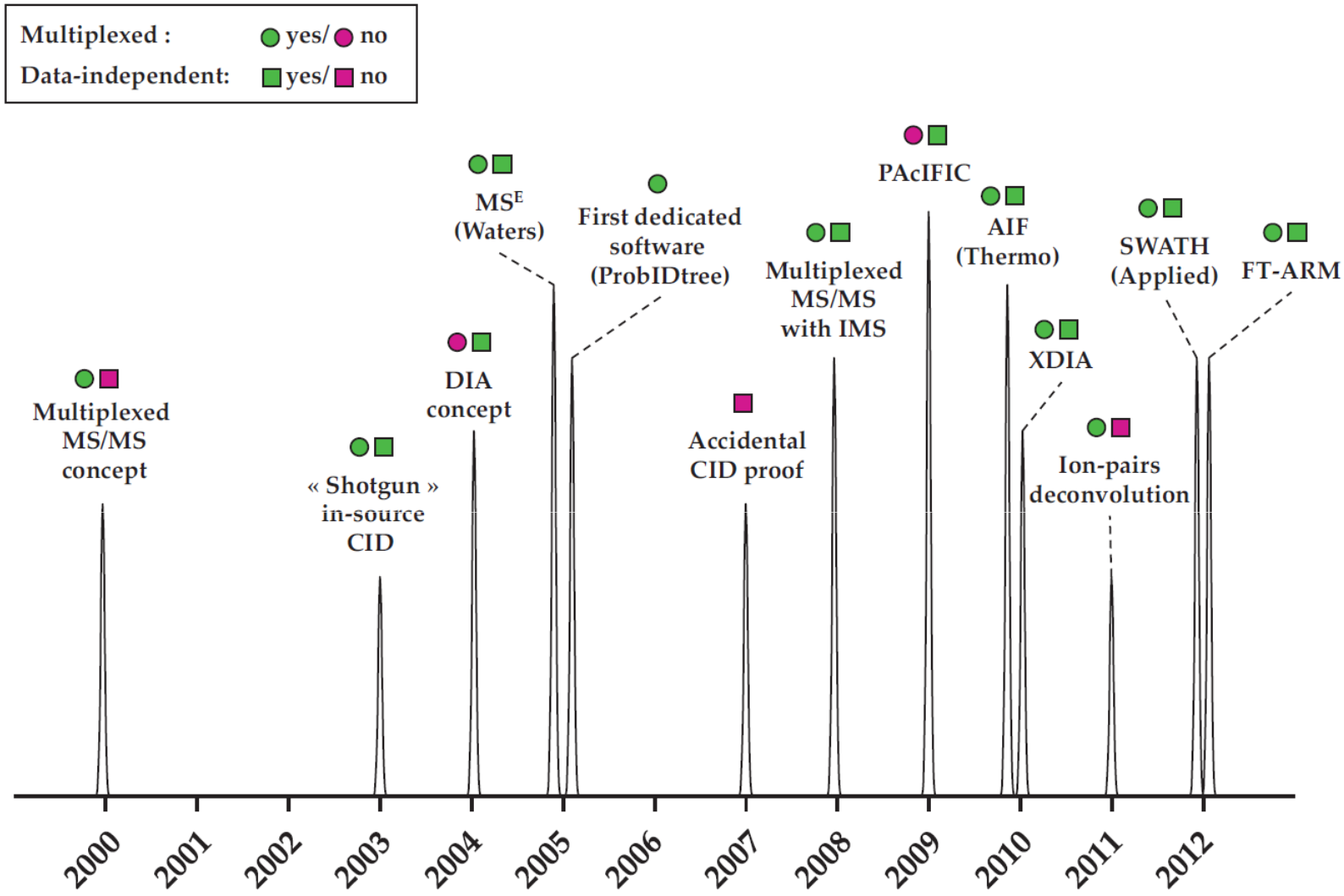


MS_E advantages

- Significant increase in the number of identifications
- Increased sequence coverage for the identified proteins
- Significantly better reproducibility than their DDA counterparts

Bateman *et al*, Anal Chem (2002)
Silva *et al*, MCP (2005)

*Patented technology



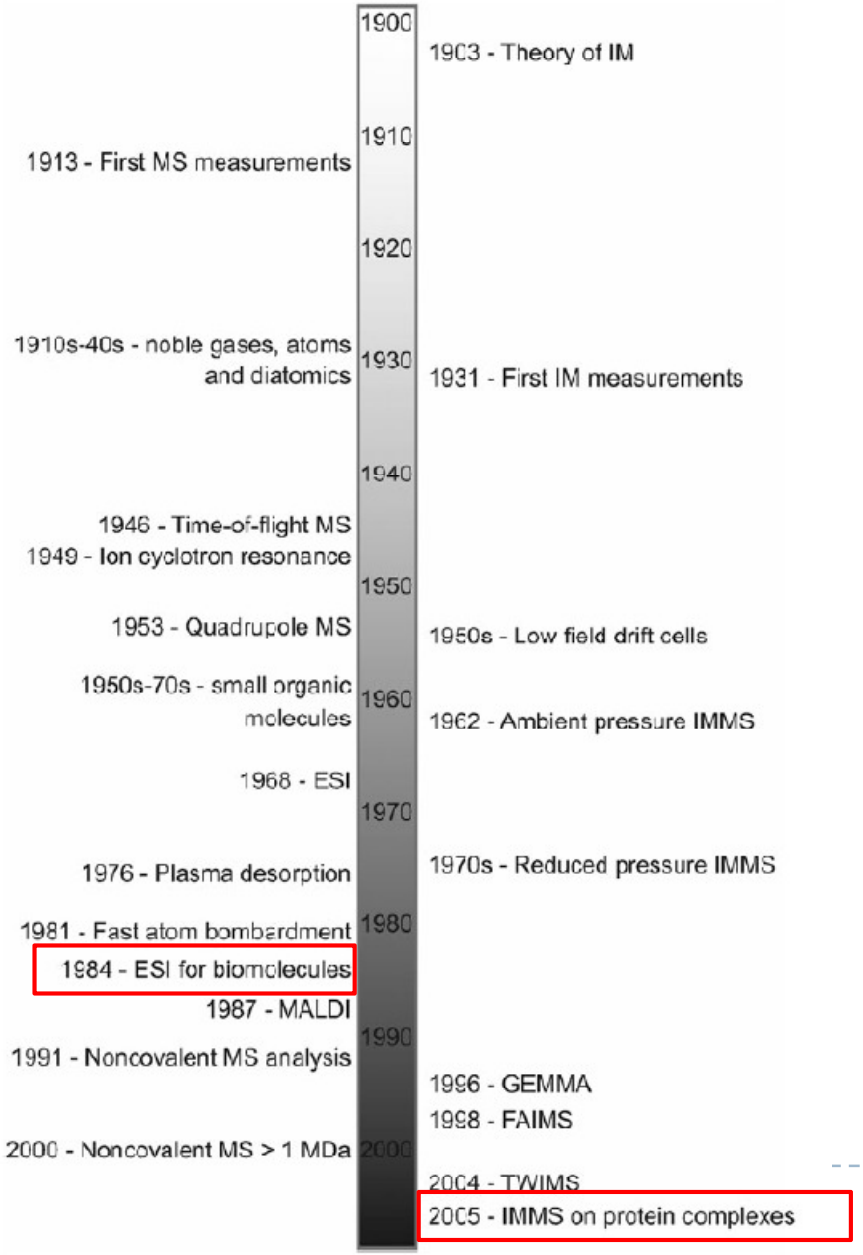
Timeline indicating some of the important milestones in the field of data independent tandem mass spectrometry for proteome profiling

Chapman et al., Mass Spectrometry Reviews 2014, 33, 452–470

Why SYNAPT G2-Si?

- Powerful data independent & data dependent solutions
- Superior separation with high-efficiency T-Wave IMS
- Wide range of experimental options

Development of IM and MS



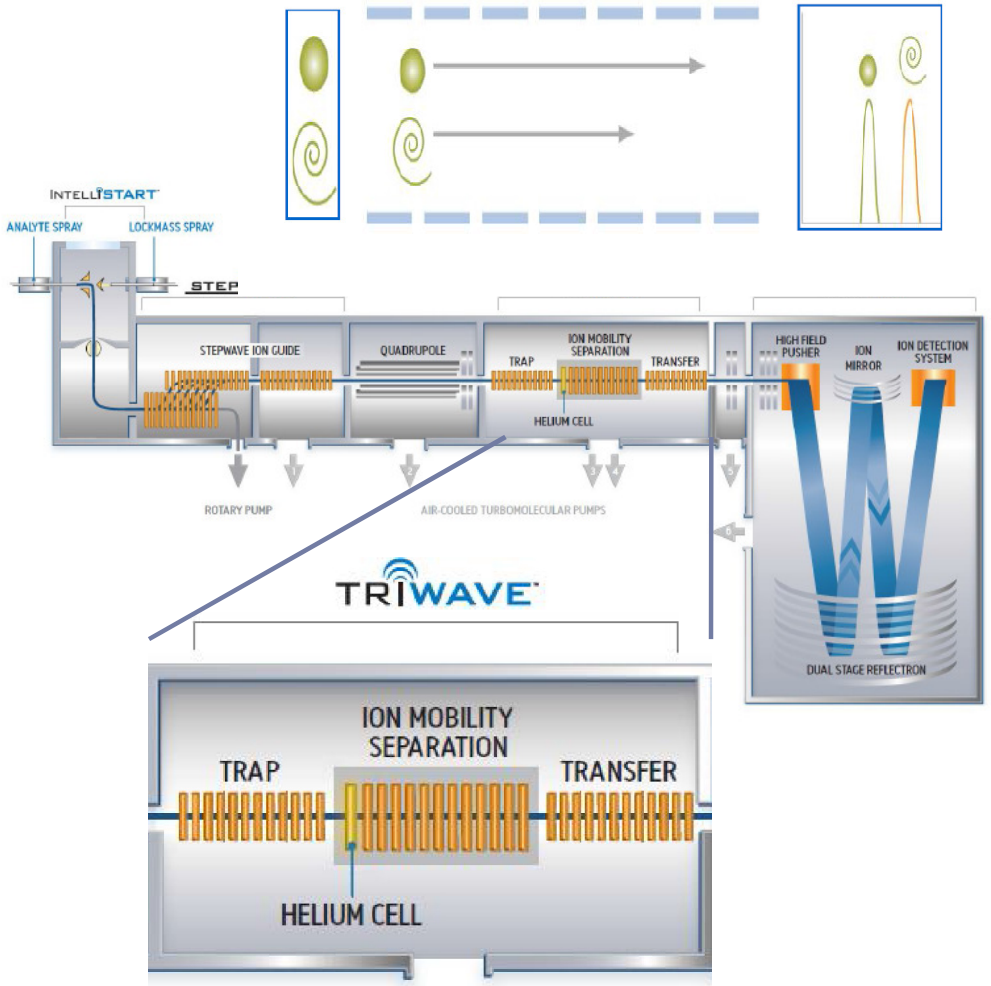
Thomson, J. J.; Rutherford, G. P., "Conduction of Electricity Through Gases", Dover, NY, 1928.

Thomson, J. J.; "Rays of Positive Electricity", Green and Co., London, England, 1933.

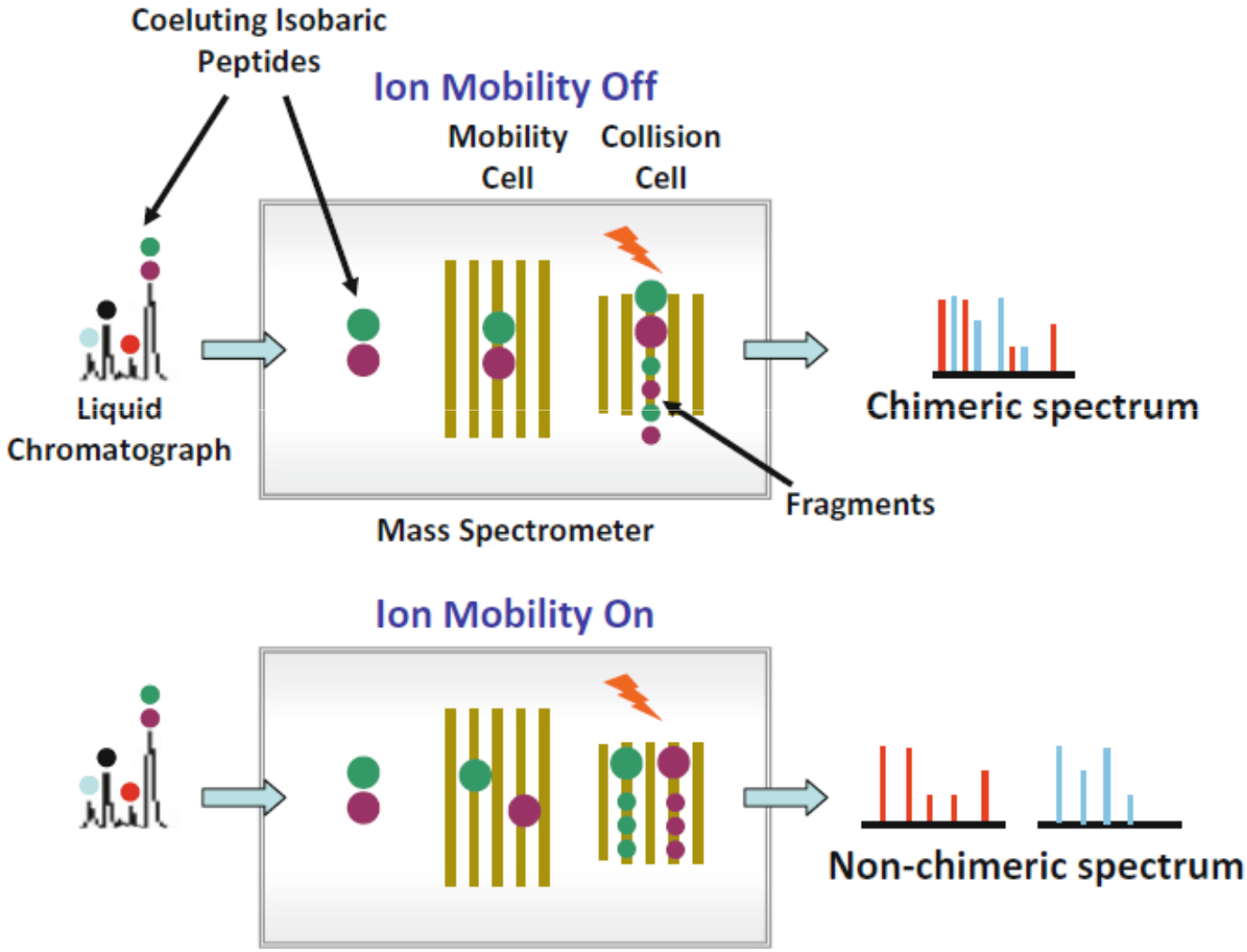
Uetrecht et al., Chem. Soc. Rev., 2010, 39, 1633–1655

Ion mobility separation (IMS)

- IMS offers additional dimension of ion separations based on size and shape
- IMS occurs in IMS ion guide in TRIWAVE region
- TRIWAVE region consists of 3 ion guides: trap, IMS and transfer
- Trap and transfer allow fragmentation of ions prior and/or after IMS

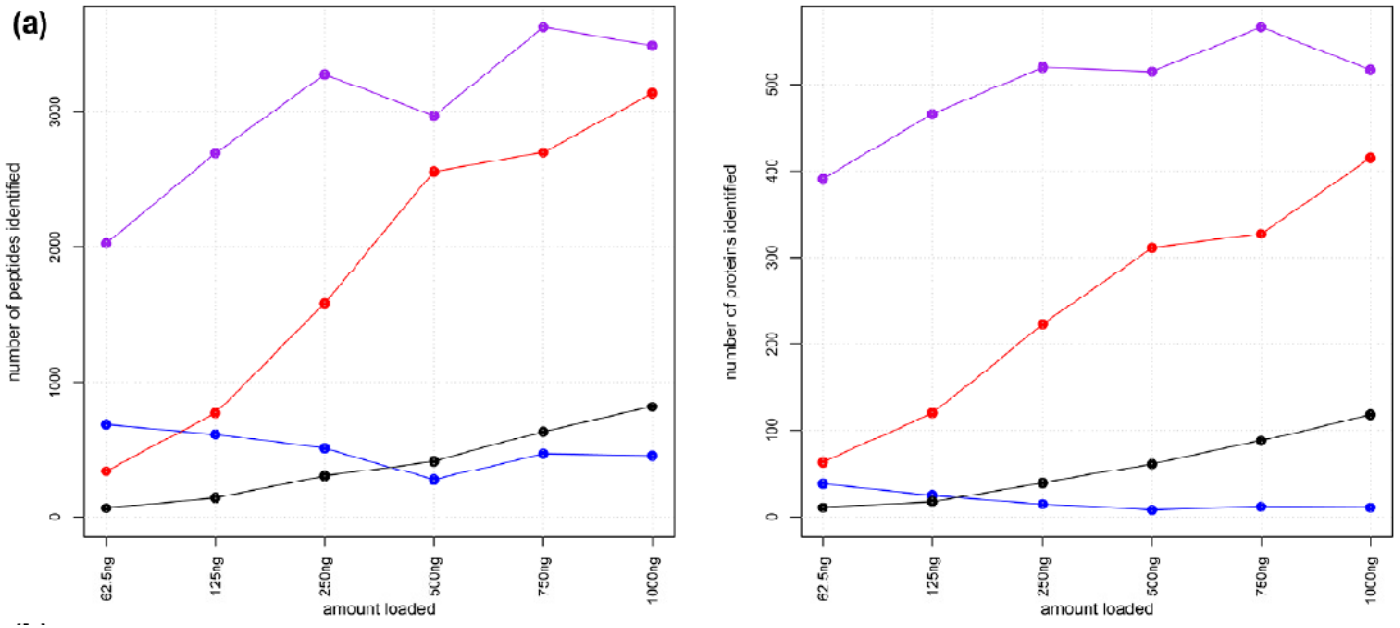


Advantages of ion mobility separation – Comparison between MSE and HDMSE



Effect of ion mobility separation in proteomics

- Up to 60% higher proteome coverage and higher confidence of protein and peptide identifications
- Limited linear range of quantitation, compromising quantitation accuracy of high intensity peptides
- More accurate and precise quantitation of lower intensity precursor ions

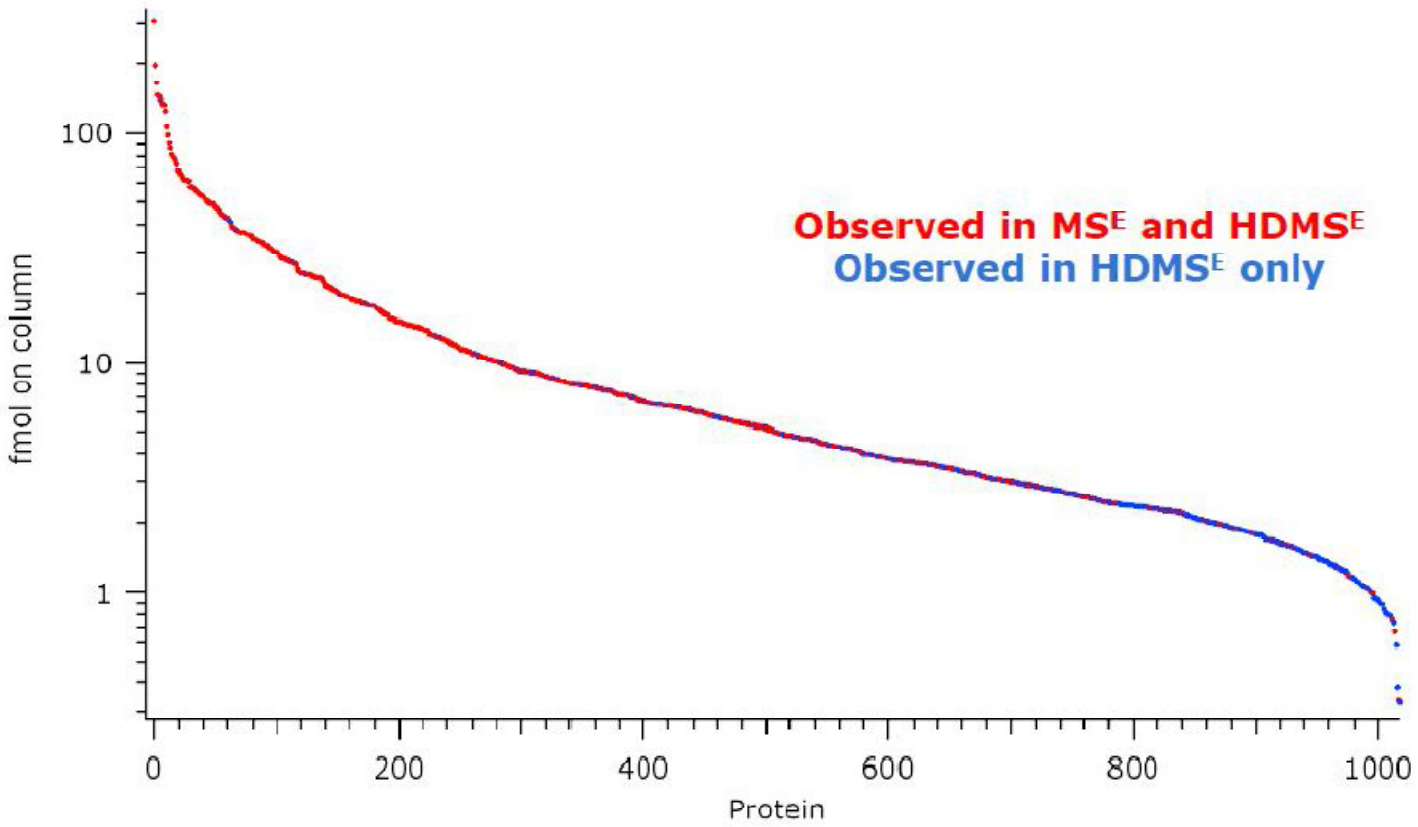


Number of protein and peptide identifications unique to MS^E (blue) and $HDMS^E$ (red) modes, common in both modes (purple), above saturation threshold in $HDMS^E$ (black)

Shliaha et al., J. Proteome Res. 2013, 12, 2323−2339

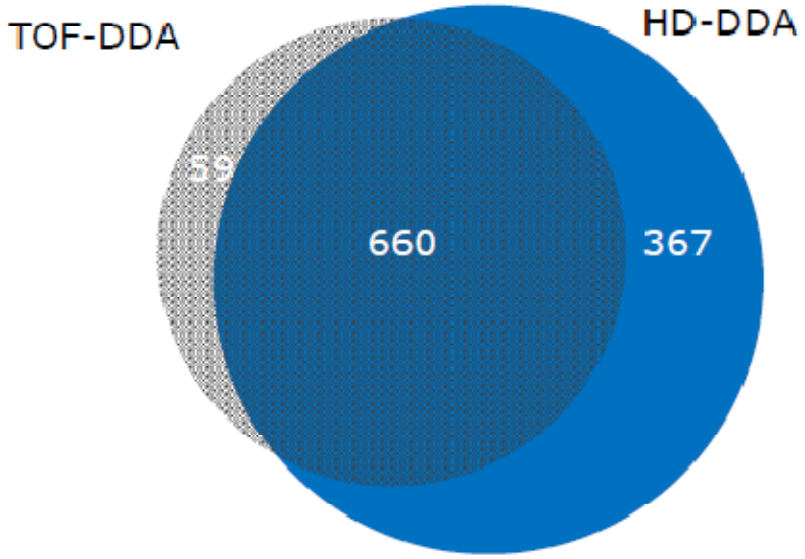
Quantification Proteins Shewanella Oneidensis

Waters
THE SCIENCE OF WHAT'S POSSIBLE.®



Applied to IMS-Data Dependent MS^2(HD-DDA)

Waters
THE SCIENCE OF WHAT'S POSSIBLE.®



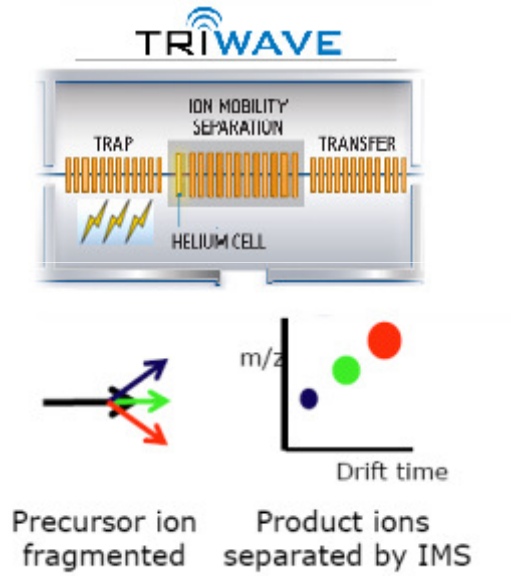
Venn intersection of protein identifications from triplicate injections of 500ng E.Coli searched with Mascot with and without wideband enhancement

Why SYNAPT G2-Si?

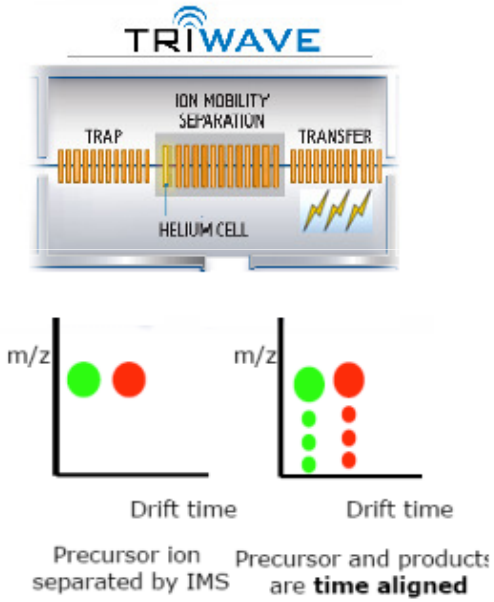
- Powerful data independent & data dependent solutions
- Superior separation with high-efficiency T-Wave IMS
- Wide range of experimental options

Fragmentation options in Synapt G2-Si

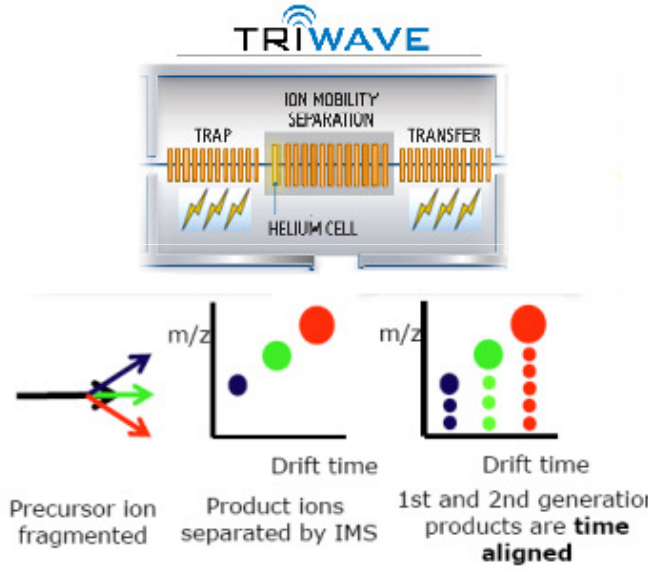
TRAP Fragmentation



Transfer Fragmentation



TAP Fragmentation

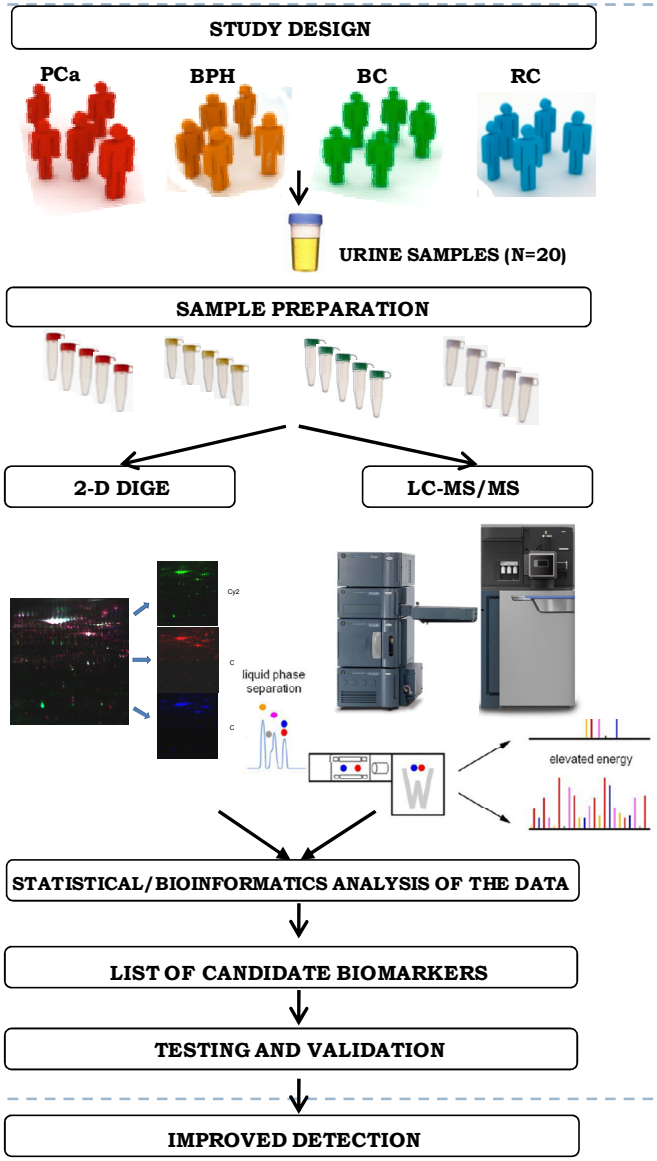


Data processing and analysis

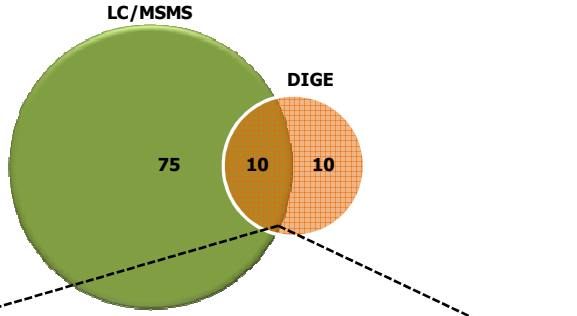
ProteinLynx Global SERVER (PLGS)

Progenesis QI
for proteomics

Non-invasive PCa biomarkers in urine – Comparative proteomics analysis of urine (4 groups)



Number of proteins with statistically significant difference in abundance (Anova ≤ 0.05) and fold change of ≥ 1.5



Protein name	Uniprot Accession No.	LC/MSMS PCa vs BPH	LC/MSMS PCa vs BC	DIGE PCa vs BPH	DIGE PCa vs BC
FGA	P02671	↓	↓	-	↑
FGG	P02679	↓	↓	-	↓
HBB	P68871	↓	↓	↑	-
PTGDS	P41222	↑	↑	-	↑
ITIH4	Q14624	↑	↑	↑	↑
CD59	P13987	-	↑	-	↑
IGKC	P01834	-	↑	-	↑
KNG1	P01042	-	↑	↑	↑
AMBP	P02760	-	↑	↑	↑
AZGP1	P25311	-	↑	-	↑

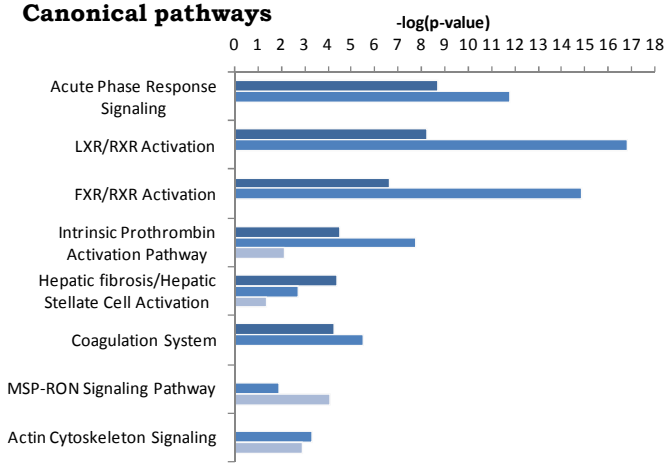
LC-MS/MS results
1882 peptides -> 226 proteins

85 proteins differentially expressed

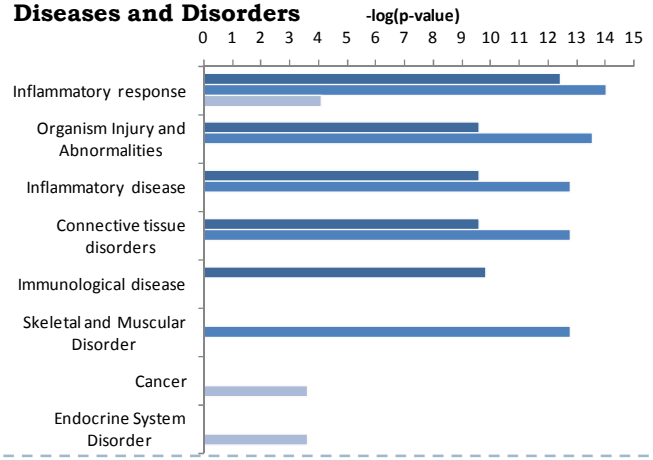
35 proteins differentially expressed in PCa in more than one group comparison

IPA analysis of proteomics data. Top significantly associated:

Canonical pathways



Diseases and Disorders



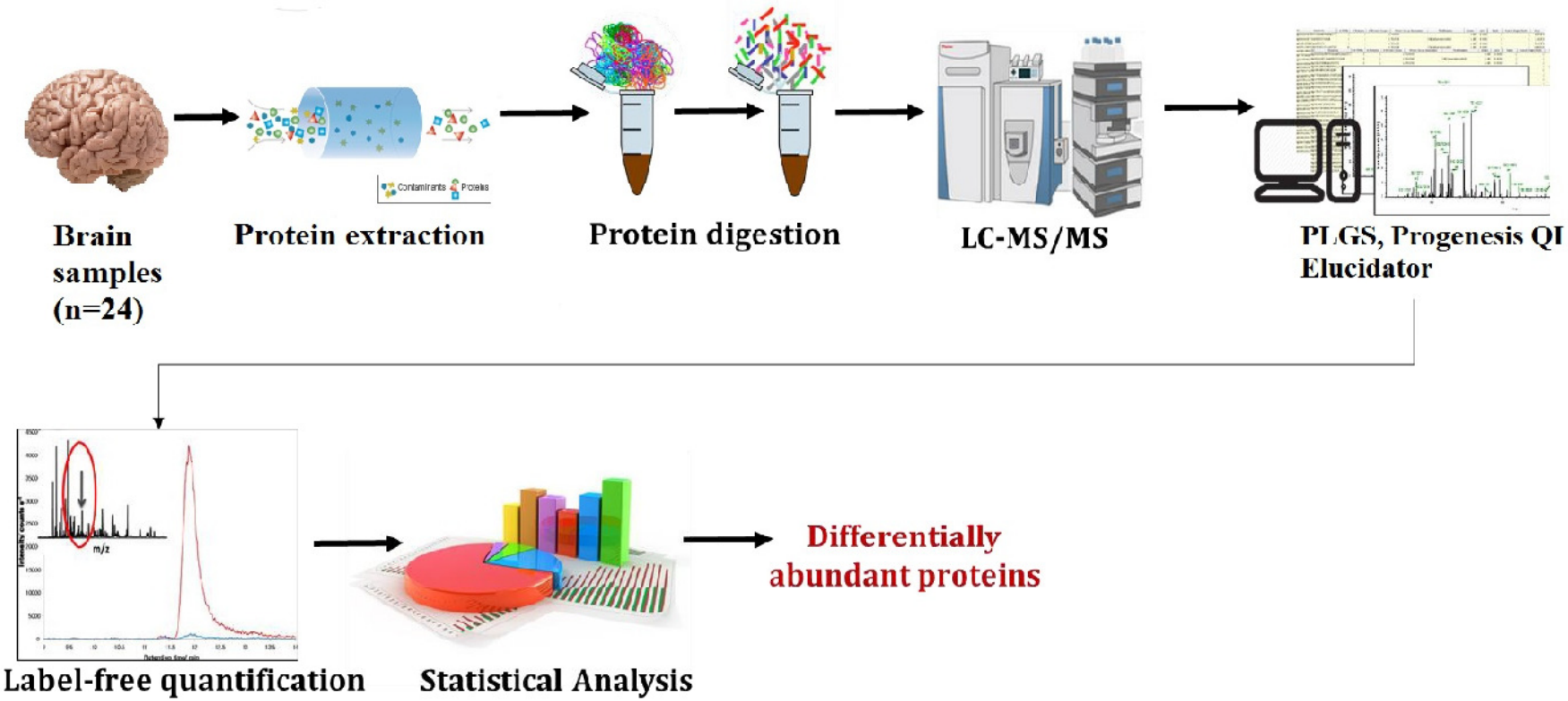
Uniprot Acc No.	Gene Name	Protein abundance in PCa in this	Normal tissue						Cancer		
			Prostate	Seminal vesicle	Epididymis	Testis	Kidney	Bladder	Prostate	Renal	Urothelial
Differential abundance in PCa vs BPH, PCa vs BC and PCa vs RC											
P08571	CD14	↑									
Differential abundance in PCa vs BPH and PCa vs BC											
P62736	ACTA2										
P02765	AHSG	↑									
P04745	AMY1A	↑									
P19961	AMY2B	↑									
P06733	ENO1										
P04083	ANXA1	↑									
P20160	AZU1										
P10909	CLU										
P12109	COL6A1										
P39060	COL18A1	↑									
P01024	C3	↑									
P24855	DNASE1	↑									
O94919	ENDOD1	↑									
P02671	FGA										
P02679	FGG										
P28799	GRN										
P68871	HBB										
P02790	HPX										
P01857	IGHG1		/	/	/	/	/	/	/	/	/
P05451	REG1A	↑									
P11117	ACP2		/	/	/	/	/	/	/	/	/
P08246	ELANE										
P10153	RNASE2	↑									
P10451	SPP1	↑									
P04746	AMY2A	↑									
P30086	PEBP1										
P20742	PZP										
P41222	PTGDS	↑									
P06702	S100A9										
Q8WZ75	ROBO4		/	/	/	/	/	/	/	/	/
Q12907	LMAN2	↑									
Q14624	ITIH4	↑									
Differential abundance in PCa vs BC and PCa vs RC											
P63267	ACTG2	↑									
P07858	CTSB										

Comparative proteomic analysis of white matter from human postmortem brain – comparison of 3 groups

Proteomes from ventral prefrontal white matter of post-mortem brains from 8 clinical triplets with

- SCH,
- major depressive disorder (MDD)
- non-psychiatric controls

	Schizophrenia	MDD	Controls
Comparative proteomics set			
Number of subjects	8	8	8
Gender	M	M	M
Age at death (years)	**42.5**±12.0	**42.6**±10.9	**42.1**±13.1
PMI (h)	**15.8**±8.5	**13.9**±4.5	**14.5**±6.1
Brain pH	**6.36**±0.22	**6.55**±0.19	**6.58**±0.12
Causes od death			
suicide	2	7	0
cardiac	0	0	4
accident	0	0	3
other	6	1	1

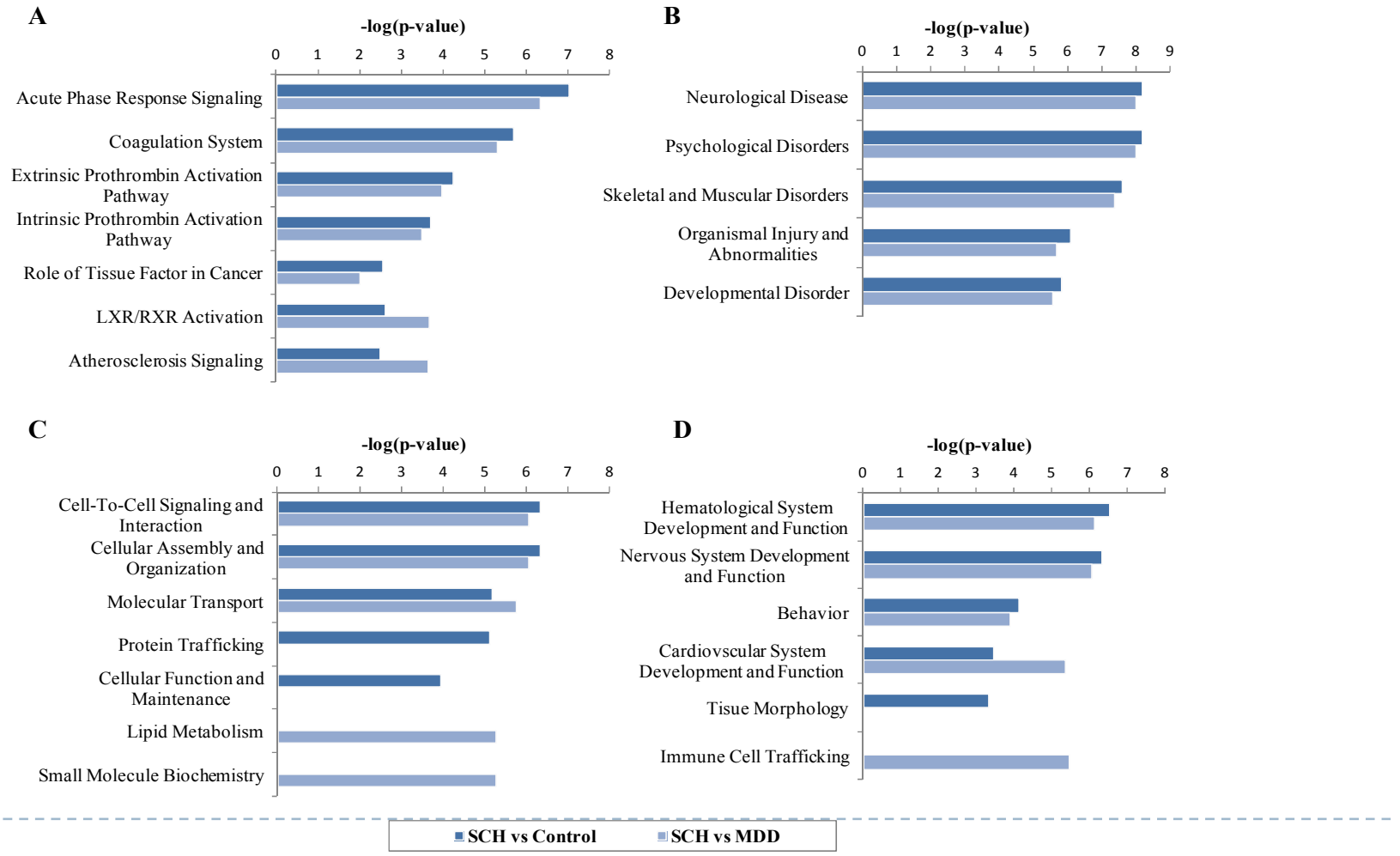


LC-MS/MS results
5249 peptides -> 1077 proteins

→ **745 proteins**
Identified based ≥ 2 peptides

→ **29 proteins** differentially expressed
Fold change ≥ 1.5 p≤ 0.01

In silico analysis of the proteins with differential abundance among the SCH, MDD and Control groups using IPA. The charts represent the top significantly associated:
A) Canonical pathways B) Diseases and Disorders C) Molecular and Cellular Functions D) Physiological System Development and Function



LC-MS/MS results
5249 peptides -> 1077 proteins

→

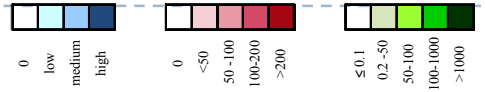
745 proteins
Identified based ≥ 2 peptides

→

29 proteins differentially expressed
Fold change ≥ 1.5 p≤ 0.01

Brain cells specificity of the proteins with differential abundance among the SCH, MDD and Control groups

Protein name	Gene name	Human brain*											Mouse brain**						
		Protein										mRNA***	mRNA						
		CC				H		LV		C			CC						
		N	E	G	Ne	N	G	N	G	P	GL	ML	A	N	OPC	NFO	MO	MG	E
Differential abundance in SCH vs Control, SCH vs MDD and MDD vs Control																			
Annexin A2	*ANXA2*																		
Differential abundance in SCH vs Control and SCH vs MDD																			
Alpha-1-acid glycoprotein 1	*ORM1*																		
Alpha-1-antichymotrypsin	*SERPINA3*																		
Alpha-1-antitrypsin	*SERPINA1*																		
Fibrinogen beta chain	*FGB*																		
Fibrinogen gamma chain	*FGG*																		
Mitochondrial import receptor subunit TOM70	*TOMM70A*																		
Synapsin-1	*SYN1*																		
Synapsin-2	*SYN2*	N/A																	
Synaptotagmin-1	*SYT1*																		
V-type proton ATPase subunit d 1	*ATP6V0D1*																		
Differential abundance in SCH vs Control and MDD vs Control																			
Tubulin alpha-1C chain	*TUBA1C*																		
Differential abundance in SCH vs Control																			
Heat shock protein beta-1	*HSPB1*																		
NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial	*NDUFS3*																		
Neuronal membrane glycoprotein M6-a	*GPM6A*																		
Transcriptional activator protein Pur-alpha	*PURA*	N/A																	
Differential abundance in SCH vs MDD																			
Apolipoprotein E	*APOE*																		
Excitatory amino acid transporter 2	*SLC1A2*																		
Hemoglobin subunit alpha	*HBA1*																		
Histone H2AX	*H2AFX*																		
Ig gamma-1 chain C region	*IGHG1*	N/A										N/A	N/A						
Major vault protein	*MVP*																		
Phospholysine phosphohistidine inorganic pyrophosphate phosphatase	*LHPP*																		
Plasma membrane calcium-transporting ATPase 1	*ATP2B1*																		
Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	*SDHA*																		
V-type proton ATPase subunit a	*ATP6V0A1*																		
Differential abundance in MDD vs Control																			
Claudin-11	*CLDN11*																		
Kinesin-like protein KIF2A	*KIF2A*																		
Perilipin-3	*PLIN3*																		



29 proteins

5 not detected in CNS

3 data NA

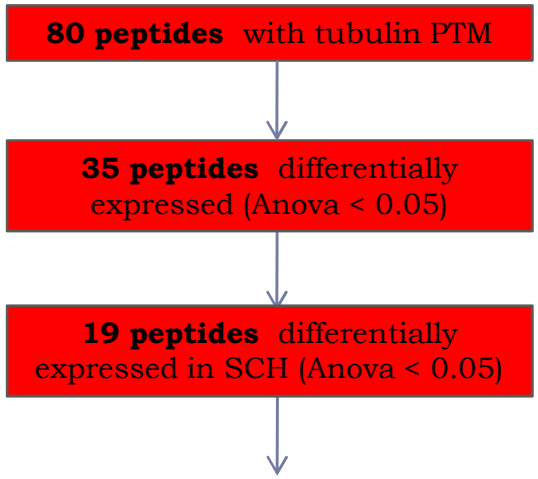
13 have ubiquitous cytoplasmic and/or nuclear expression

2 are plasma proteins with expression in cerebral cortex endothelial cells

6 proteins have distinct selective expression in the CNS

Tubulin PTM in brain white matter

Tubulin PTM	PLGS variable modification
Acetylation of Lysine	Acetyl K
Phosphorylation on Ser/Thr/Tyr	Phosporyl S/T/Y
O-Linked β-N-acetylglucosamine on serine or threonine residues	O-GlcNAc ST
Palmitoylation of Cysteine/Serine/Threonine	Palmitoyl CST
Constant variable modification*	
Oxidation of Methionin	Oxidation M
Constant fixed modification*	
Carbamidomethylation of Cysteine	Carbamidomethyl C



*=product of sample preparation and reduction and alkylation of Cysteine

Description	Sequence	Modifications	Max fold change	Highest mean condition	Lowest mean condition	Anova
Tubulin alpha-1B chain OS=Homo sapiens GN=TUBA1B PE=1 SV=1	SIQFVDWCPTGFK	[8] Carbamidomethyl C|[10] Phosphoryl STY	2.1	SCH	MDD	2.43E-02
Tubulin alpha-1C chain OS=Homo sapiens GN=TUBA1C PE=1 SV=1	AYHEQLTVAEITNACFEPANQMVK	[7] O-GlcNac ST|[12] O-GlcNac ST|[15] Carbamidomethyl C	17.8	Control	SCH	1.18E-04
Tubulin alpha-8 chain OS=Homo sapiens GN=TUBA8 PE=1 SV=1	LMYAK	[2] Oxidation M	1.5	SCH	Control	6.53E-03
Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV=2	ALTVPELTQQVFDAK	[8] O-GlcNac ST	1.3	Control	SCH	2.11E-02
Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV=2	ISVYYNEATGGKYVPR	[12] Acetyl K	1.4	SCH	MDD	1.55E-02
Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV=2	ALTVPELTQQVFDAK	[15] Acetyl K	1.2	Control	SCH	3.16E-04
Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV=2	GSQQYRALTVPELTQQVFDAK	[9] Phosphoryl STY|[14] Phosphoryl STY|[21] Acetyl K	1.1	MDD	SCH	1.39E-02
Tubulin beta-2A chain OS=Homo sapiens GN=TUBB2A PE=1 SV=1	GSQQYRALTVPELTQQMFDSK	[21] Acetyl K	1.7	MDD	SCH	3.11E-04
Tubulin beta-2A chain OS=Homo sapiens GN=TUBB2A PE=1 SV=1	EVDEQMLNVQNKNSSYFVEWIPNNVK	[14] O-GlcNac ST|[15] O-GlcNac ST	2.2	Control	SCH	1.17E-02
Tubulin beta-2A chain OS=Homo sapiens GN=TUBB2A PE=1 SV=1	ALTVPELTQQMFDSK	[14] Phosphoryl STY|[15] Acetyl K	1.5	MDD	SCH	5.07E-03
Tubulin beta-4A chain OS=Homo sapiens GN=TUBB4A PE=1 SV=2	EVDEQMLSVQSK	[12] Acetyl K	1.4	MDD	SCH	1.32E-03
Tubulin beta-4A chain OS=Homo sapiens GN=TUBB4A PE=1 SV=2	NMMAACDPR	[2] Oxidation M|[6] Carbamidomethyl C	8.1	SCH	MDD	2.63E-02
Tubulin beta-4A chain OS=Homo sapiens GN=TUBB4A PE=1 SV=2	NSSYFVEWIPNNVKTAVCDIPPR	[2] Phosphoryl STY|[18] Carbamidomethyl C	58.4	SCH	MDD	2.02E-02
Tubulin beta-6 chain OS=Homo sapiens GN=TUBB6 PE=1 SV=1	TLKLTTPTYGDLNHLVSATMSGVTTSLR	[5] Phosphoryl STY|[17] Phosphoryl STY|[20] Oxidation M	1.9	MDD	SCH	1.88E-02
Tubulin beta-6 chain OS=Homo sapiens GN=TUBB6 PE=1 SV=1	MREIVHIQAGQCGNQIGTK	[12] Carbamidomethyl C|[12] Palmitoyl CST	1.6	SCH	Control	1.28E-02
Tubulin beta-8 chain OS=Homo sapiens GN=TUBB8 PE=1 SV=2	ALTVAELTQQMFDAK	[15] Acetyl K	1.1	Control	SCH	4.03E-02
Tubulin beta-8 chain OS=Homo sapiens GN=TUBB8 PE=1 SV=2	ALTVAELTQQMFDAK	[3] Phosphoryl STY	1.4	MDD	SCH	1.76E-03
Tubulin beta-8 chain OS=Homo sapiens GN=TUBB8 PE=1 SV=2	LPTPTYGDLNHLVSATMSGVTTCLR	[14] Phosphoryl STY|[16] O-GlcNac ST|[16] Phosphoryl STY|[23] Carbamidomet	12.4	SCH	MDD	1.26E-02
Tubulin beta-8 chain-like protein LOC260334 OS=Homo sapiens PE=1 SV=1	MSATFIGNNAAIQELFTCVSEQFTAMFR	[18] Carbamidomethyl C	1.3	Control	SCH	2.56E-02

THANK YOU…….

DO YOU HAVE ANY QUESTİONS ?