



# Particle And Proteins: Pathway to Impact

JONATHAN HAYS

QUEEN MARY UNIVERSITY OF LONDON

# Searching for a Project : Autumn 2013

- ▶ Looking for a new project to complement physics work at the LHC
- ▶ Something that could lead to “impact”
- ▶ Things I've worked on...

- ✗ Instrumentation
- ✓ Data Handling
- ✓ Data Acquisition
- ✓ Algorithm development
- ✓ Machine Learning
- ✓ Statistical Analysis

Networking was key!  
Attended many events, met hundreds of people... no leads :(

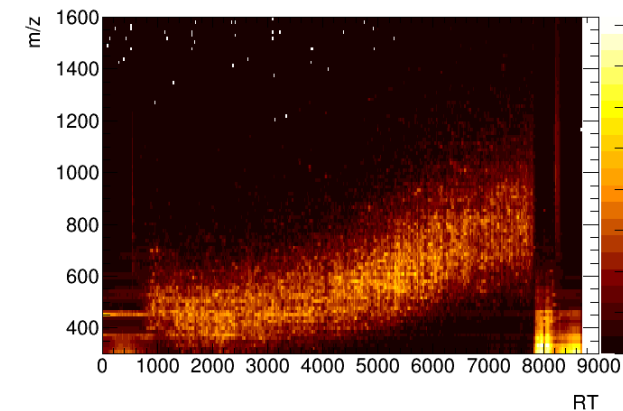
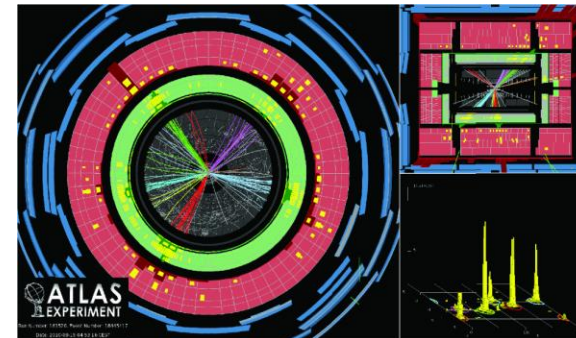
Met someone over coffee in Senior Common Room – invited to give seminar to Bioinformaticians – 3 leads

Met someone in a faculty teaching meeting – 1 lead

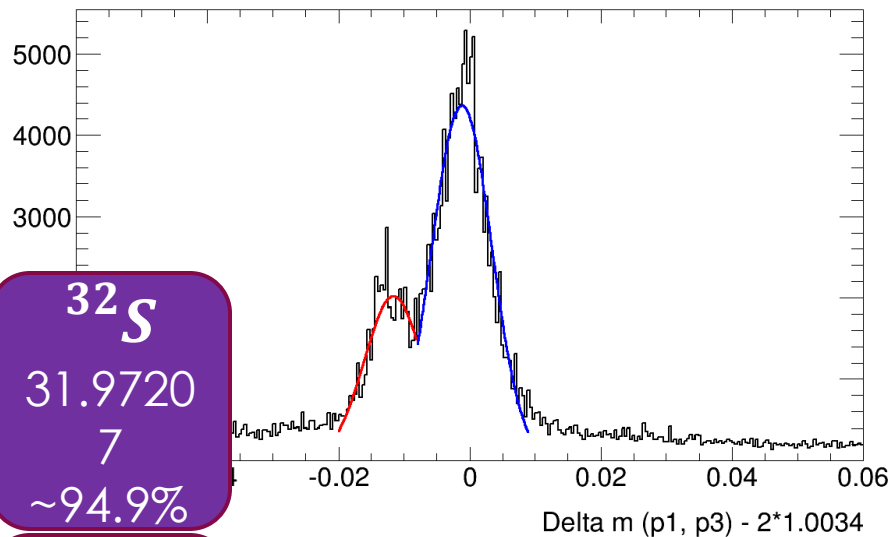
- ▶ Looking around at previous “impactful” projects – all based on instrumentation

# Particle to Proteins : Summer 2014

- ▶ Summer 2014: Summer UG intern
- ▶ Autumn 2014: Impact Accelerator Bid successful
  - ▶ £10k to examine distributed computing solutions for proteomics
- ▶ Winter 2014: Impact Accelerator Bid successful
  - ▶ £10k to examine new algorithms for analyzing proteomic data



# Particle to Proteins : 2014/2015



**$^{12}\text{C}$**   
12  
(def)  
~98.9%

**$^{32}\text{S}$**   
31.9720  
7  
~94.9%

**$^{13}\text{C}$**   
13.0033  
5  
~1.08%

**$^{34}\text{S}$**   
33.9678  
6  
4.3%

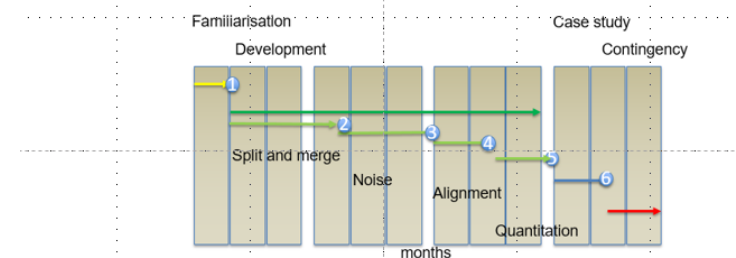
- ▶ UG Projects student 2014/15 working on algorithms
- ▶ Working with Post-doc
- ▶ 2015 – Impact Accelerator Pump-Prime bid successful £20k
- ▶ Demonstrated ability to extract information from the data – FAST!
- ▶ Building on this - applied for QMUL Life Science Institute PhD studentship with partners across two faculties and three schools

# Particles to Proteins : 2015/16

- ▶ Using results from Pump-prime activities applied for BBSRC funding – unsuccessful
- ▶ Work begins with PhD student to work on new software for protein analysis in Barts Cancer Institute
- ▶ Secured £50k from Queen Mary Innovations to produce simple prototype example with benchmarking to develop future partnerships
- ▶ Recruitment problems ☹️
  - ▶ Turns out hiring people for short-term contracts is tough

## Project Plan & Experimental design

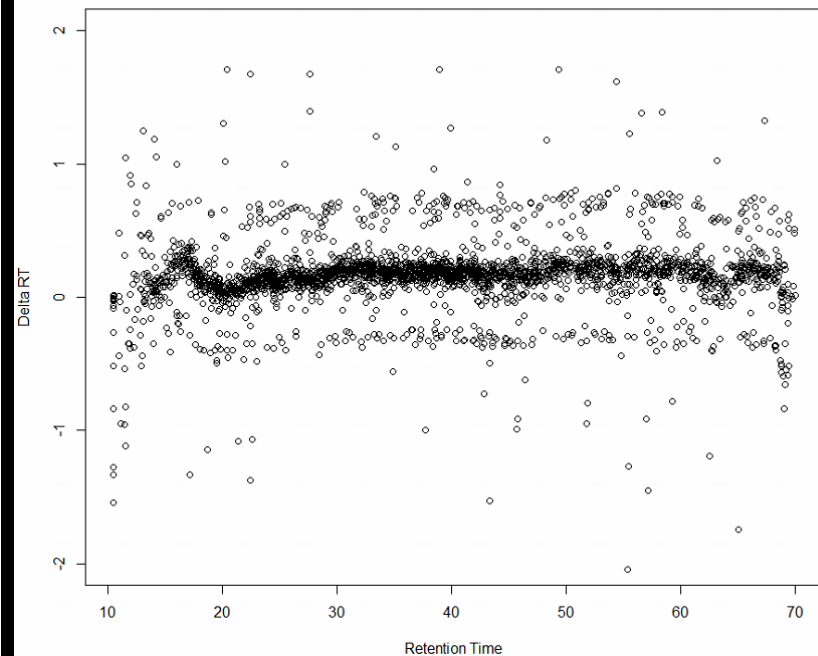
Time line and milestones



1. Up and running with existing code
2. Cluster split and merge
3. Noise evaluation
4. Cross-run alignment
5. Simple quantitation using peak integration and workflow integration
6. Case study – comparative study of performance in real workflow

Pipeline step	Intensity Cut				
	0	1000	2000	3000	4000
Prefiltering and reading data	11	11	12	10	10
Peak Finding	12	7	5	4	3
Clustering Peaks	89	43	17	8	5
2D Geometric Clustering	26	18	11	10	8
Writing outputs	58	40	28	20	16

# Particles to Proteins : 2016/17



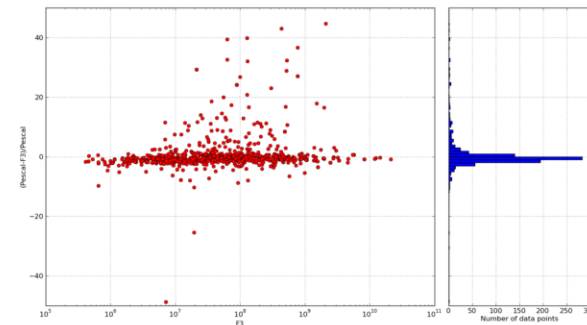
- ▶ Finally succeeded in recruiting a new post-doc
- ▶ Work continues, but now more closely with team at Barts
- ▶ Looking at more realistic analysis pipelines and trying to produce benchmarking scenarios

# Particles to Proteins : 2017/18

- ▶ Post-doc funding runs out
- ▶ PhD student almost complete with PhD work by end Summer 2018
- ▶ Failed application for CLASP funding ☹️
  - ▶ Too low TRL and poorly aligned with call
- ▶ But – PhD student now post-doc working on proteomics
- ▶ Have successful demonstration of working code
- ▶ Need to find commercial partners

## Quantitation Performance Comparison

PESCAL++ versus F3



New results – hot off the press – as part of our on-going case study  
**Very good agreement in quantitation.** Timing acceptable with some caveats.

Improving on this is the primary part of the core development work proposed

Stage	Time to complete [s]	Cumulative time [s]
Running PescalPP	159	159
Filtering data	15	174
Get RT slices	<1	174
Finding peaks	14	188
Finding clusters	13	201
Finding 2D clusters	24	225
Read in peptide IDs	<1	225
Matching IDs and 2D clusters	919	744
Quantitation	<1	744
Filling histograms	29	773

Data sample	No. peptides Pescal++	% total peptides	No. peptides F3Pescal++	% total peptides
1	12500	95.7	12059	92.3
2	12364	94.7	11886	91.0
3	12453	95.4	12049	92.3
4	12174	93.2	11704	89.6
5	12066	92.4	11572	88.6
6	12388	94.9	11993	91.8
7	10878	83.3	10213	78.2
8	10892	83.4	10252	78.5
9	11039	84.5	10344	79.2
10	11032	84.5	10320	79.0
11	10848	83.1	10200	78.3
12	10808	82.8	10114	77.5