A comprehensive characterization of the pig islet proteome: PTMs, amino acid substitutions and novel isoforms

Ebbing de Jong[1], Bernhard Hering[2,3], Pratik Jagtap[4], John Chilton[4], Getiria Onsongo[4], Timothy Griffin[1]

[1] Department of Biochemistry, Molecular Biology and Biophysics, [2] Department of Surgery, [3] Schulze Diabetes Institute, [4] Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN

*contact: ebbingdejong@umn.edu

Overview

- Characterization of the porcine islet proteome is required for an understanding of differences in transplantable quality
- 4-plex iTRAQ labeling strategy comparing three replicates of two high- and two low-quality islet preparations, with prefractionation and Orbitrap Velos analysis
- 4527 proteins identified and 9797 modified peptides, including cases where peptide abundances differ from protein ratios
- Proteogenomics workflow identified 28 previously uncharacterized protein products at high stringency

Introduction

Pancreatic islet transplantation is an emerging treatment option for patients with type 1 diabetes. However, assays for selecting islet preparations suitable for transplantation remain to be optimized. We have undertaken a large-scale MS-based proteomics investigation of purified pig islets. This investigation includes iTRAQ labeling for comparative quantitative analysis between four samples (two high and two low quality) across three biological replicates. Back-end analyses include (i) an unbiased database search for PTMs and amino acid substitutions whose relative abundances can be compared with that of the protein group, and (ii) a proteogenomics search to better annotate expressed proteins and identify potential variants.

Challenges to this project are that (i) pigs are not a model organism often used in biomedical research and the proteome is poorly annotated, and (ii) this sort of bioinformatics analysis often requires knowledge of several software packages which do not always interface well together[1]. Here we show how the newly developed Galaxy-P suite of tools can allow a non-expert user to perform an in-depth analysis of a proteomic dataset.

Figure 1: High quality islets preparation (left) and low quality islets preparation (right). Note the irregular shape of islets in the low quality preparations.

Methods

Islets from individual pigs were isolated, purified graded for transplantable quality, lysed, and the proteins digested. Samples were labeled using the 4-plex iTRAQ reagents, across three biological replicates. Peptide fractionation was performed using high-pH reversed phase chromatography and analyzed by RP-LC-MSMS on an Orbitrap Velos using HCD. A recently developed framework, Galaxy-P, allowed multiple search functions to be performed in an integrated manner, including all steps described here. Database searching was performed using ProteinPilot including searches for amino acid substitutions and unbiased search for PTMs. iQuant was used to quantify proteins as well as peptides containing substitutions or modifications. For proteogenomics search, database generated by using AUGUSTUS[2] was used to identify novel gene products.

[1] S. Renuse *et al.*, Proteomics, 11(4): 620-630 (2011)
[2] http://gbi.agrsci.dk/pig/sscrofa10_2_annotation/
[3] R. Wu *et al.*, Mol. Cell Proteomics, 10: 1-12 (2011)
[4] G. Onsongo *et al.*, Proteomics, 10(19): 3533 – 3538 (2010)

Results

Database search results

Table 1: Search results excluding modifications.

	Replicate 1	Replicate 2	Replicate 3	Total
Spectra	112183	56022	79590	247795
Spectral IDs at 5% local FDR	59441	31111	45123	136759
Percent of spectra identified	53.0	55.5	56.7	55.2
Proteins identified at 1% global FDR	3827	3045	3443	4527

Modified peptides and amino acid substitutions

Levels of PTM-containing peptides must be compared to protein-level data for correct interpretation, similar to a process described by Wu *et al.*[3] A tool has been created in Galaxy-P, based on the iQuant software[4], which can extract both peptide and protein-level abundance data, allowing users to identify PTMs of interest which differ between samples.

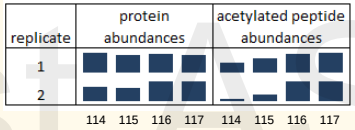

Figure 2: Peptide N-terminal acetylation was detected on a peptide which is significantly upregulated, compared to its parent protein, in the two low quality islet preparations 116 and 117 vs 114 and 115.

An overview of the peptide modifications are shown in Table 2.

Table 2: Summary of PTMs and amino acids substitutions search

	Replicate 1	Replicate 2	Replicate 3	Total
Total distinct peptides	36467	22652	33509	52821
Unmodified	28021	18842	24389	37764
Modified	5202	2907	6436	9797
Containing an amino acid substitution	3244	903	2684	5260

Proteogenomics

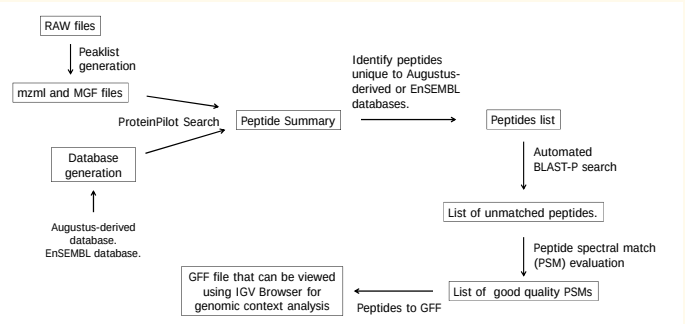

Figure 3: Workflow used to discover novel protein products.

For more details on proteogenomics, please see poster TP 248.

Table 3: Novel peptide sequences detected in the porcine islet proteome.

Genomic rearrangement	Number of peptides
On an intron	6
On a UTR	2
Novel exon junctions	2
Different frame	2
Pseudogene	2
Identified by Augustus as an exon but not by EnSEMBL	14
Total	*28*

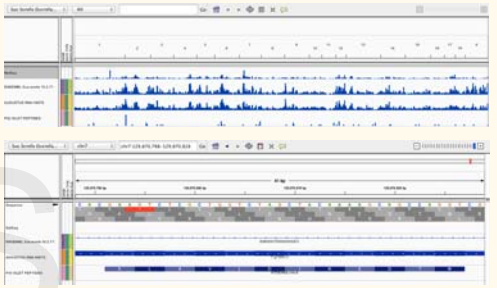

Figure 4: Screen captures of the Integrative Genomics Viewer (IGV) program, displaying previously unidentified pig islet peptides, (top) mapped onto the genome; and (bottom) zoomed in on a peptide expressed from a genomic region which was previously annotated as an intron on chromosome 7.

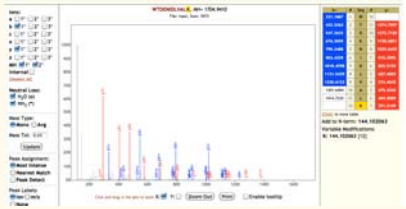

Figure 5: PSM evaluation tool showing high quality of the MS^2 spectrum assigned to the sequence in Figure 4. This peptide was identified by a total of 7 MS^2 spectra.

Conclusions

- Porcine islets produce a very rich dataset, allowing thorough testing of the Galaxy-P framework
- Large numbers of PTMs and amino acid substitutions can be quantitatively compared to overall protein ratios
- High mass accuracy and good spectral quality allow for the proteogenomic identification of novel peptides

Acknowledgements

- Funding from SDI
- NSF grant 1147079
- Jeff Ansite, SDI
- Dr. LeeAnn Higgins, MSP
- Todd Markowski, MSP
- James Johnson, MSI
- Center for Mass Spectrometry and Proteomics, U Minnesota