DEEP PROTEOME COVERAGE THROUGH RIBOSOME PROFILING AND MS INTEGRATION.
Elvis Ndah\textsuperscript{a,b,c}; Jeroen Crappé\textsuperscript{a}; Alexander Koch\textsuperscript{a}; Gerben Menschaert\textsuperscript{a}; Sandra Steyaert\textsuperscript{a}; Petra Van Damme\textsuperscript{b,c}

INTRODUCTION

Recent studies point to an underestimation of the complexity of the (eukaryotic) proteome [1]. With evidence pointing to the existence of non-annotated translation products from alternative translation initiation resulting in the creation of N-terminal extended or truncated proteforms, internal out-of-frame sequence variations products amongst others [1,2]. Typically, such alternative translation products cannot be identified from canonical protein databases, hence the need for a tailored search space.

The recently developed ribosome profiling (RIBO-seq) strategy provides a means to monitor protein synthesis at single to sub-codon level [1,3]. RIBO-seq derived data gives a more representative expression state and accounts for non-canonical information. Without taking this information into account, MS-based proteomic studies may fail to detect novel protein forms. We developed a pipeline [2] that takes advantage of RIBO-seq information to generate an optimized search space.

MATERIALS & METHODS (PROTEOFORMER Pipeline[1])

WORKFLOW OVERVIEW

1. Ribosome profiling (RIBO-seq) performed on HiSeq, illumina sequencing platform
2. Quality control based on existing tools as FastQC and custom metagenic functional assessment
3. Transcriptome mapping based on STAR \cite{5} and/or TopHat2 \cite{6} local aligners.
4. Single nucleotide polymorphism (SNP) calling based on samtools-snp.
5. Translation initiation site (TIS) calling based on trained Support Vector Machine (SVM) \cite{6} or rule-based algorithm \cite{6}.
6. Translation product assembly, taken into account the SNP TIS, INDEL awareness. Construct complete proteome, optionally combine with Canonical.
7. MS-based proteomics/peptidomics using SearchChick \cite{7} and PeptideShaker \cite{7} tools.
8. Genome-centric visualization of all generated information tracks (Ensembl, UCSC, IGV genome browser).

RESULTS

MS based proteomic and RIBO-seq experiments on two Mouse and Arabidopsis thaliana:

Mouse:

Arabidopsis Thaliana

Examples:

\begin{itemize}
\item LEFT: Depiction of the mouse HDGF gene’s 5’-extension predicted by RIBO-seq and identified using N-terminal proteomics experiment.
\item RIGHT: uORF with TIS called by RIBO-seq and a peptide [AGHKVAHATLKGPSVVKE] at the TIS identified by N-terminomics for gene AT3G62400 of Arabidopsis Thaliana TAIR10 genome.
\end{itemize}

REFERENCES


CONCLUSION & FUTURE WORK

Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events \cite{1}.

Future work will mainly focus on:
\begin{itemize}
\item Investigating read density and coverage to identify genes undergoing translation.
\item Implementing SVM in TIS calling algorithm
\end{itemize}

AFFILIATIONS

\textsuperscript{a} Laboratory of Bioinformatics and Computational Genomics, Department of Mathematical Modelling, Statistics and Bioinformatics, Faculty of Bioscience Engineering, Ghent University
\textsuperscript{b} Department of Medical Protein Research, Flemish Institute for Biotechnology, B-9000 Ghent
\textsuperscript{c} Department of Biochemistry, Ghent University, B-9000 Ghent

CONTACT: elvis.ndah@ugent.be, Gerben.Menschaert@Ugent.be