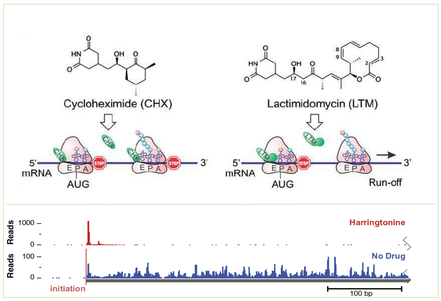
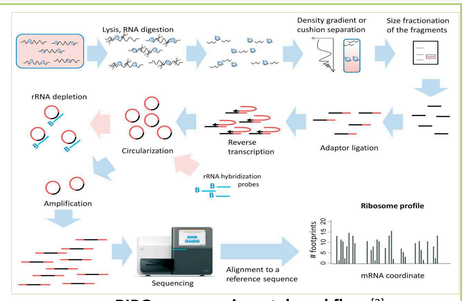
DEEP PROTEOME COVERAGE THROUGH RIBOSOME PROFILING AND MS INTEGRATION.

Elvis Ndah[a,b,c]; Jeroen Crappé[a]; Alexander Koch[a]; Gerben Menschaert[a]; Sandra Steyaert[a]; Petra Van Damme[b,c]

INTRODUCTION

Recent studies point to an underestimation of the complexity of the (eukaryotic) proteome [1]. With evidence pointing to the existence non-annotated translation products from alternative translation initiation resulting in the creation of N-terminal extended or truncated proteoforms, internal out-of-frame translation 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.

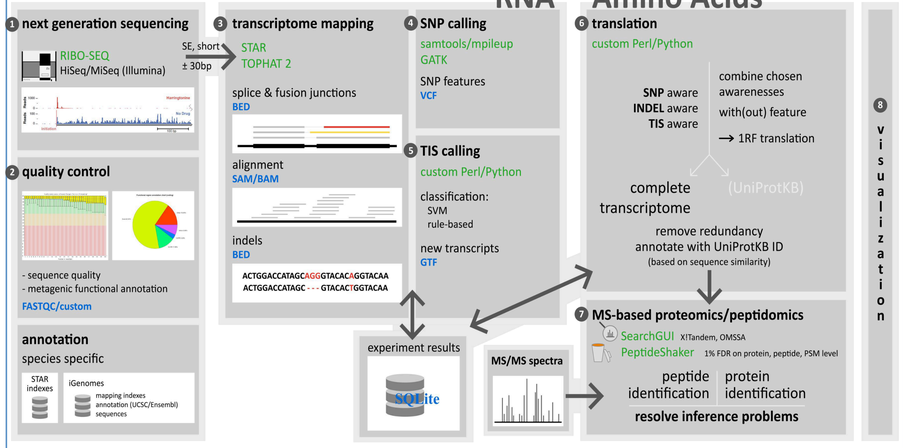




RIBO-seq experimental workflow [3]

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 [4] and/or TopHat2 [5] local aligners.

4. Single nucleotide polymorphism **(SNP) calling** based on samTools-mpileup.

5. Translation initiation site **(TIS) calling** based on trained Support Vector Machine (SVM) [2] or rule-based algorithm [1].

6. **Translation product assembly**, taken into account the SNP, TIS, INDEL awareness. Construct complete proteome, optionally combine with Canonical.

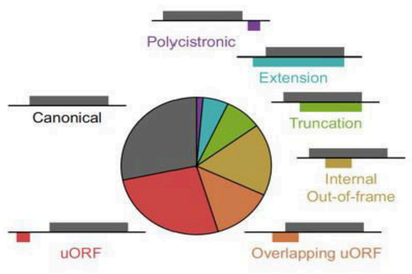
7. **MS-based proteomics/peptidomics** using SearchGui [6] and PeptideShaker [7] tools.

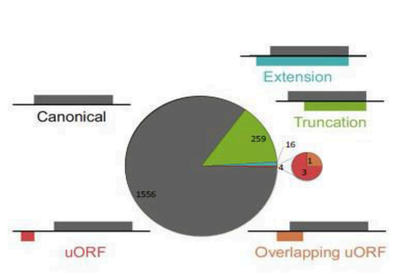
8. **Genome-centric visualization** of all generated information tracks (Ensembl, UCSC, IGV genome browsers).

RESULTS

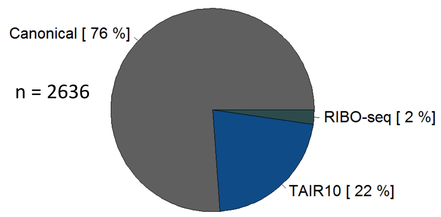
MS based proteomic and RIBO-seq experiments on two Mouse and Arabidopsis thaliana:

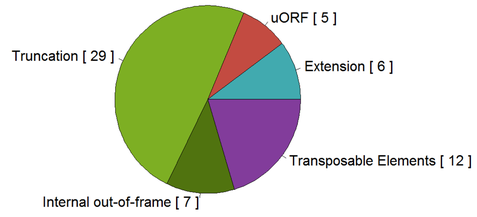
Mouse:



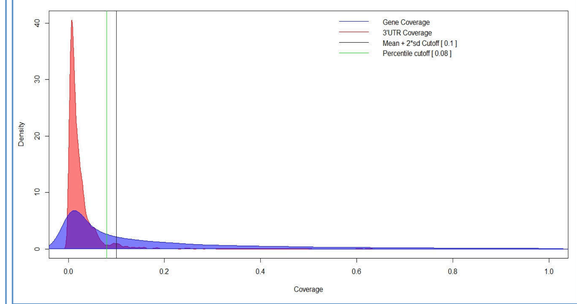


Arabidopsis Thaliana

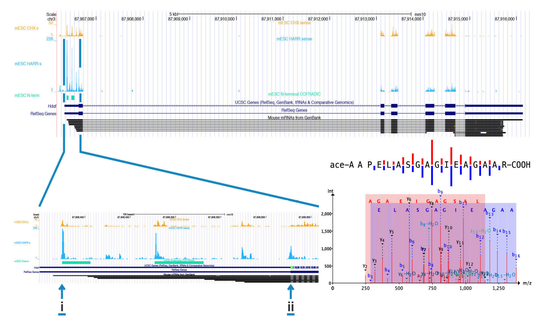


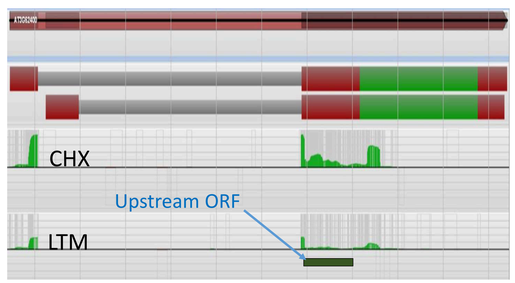


RPF coverage to determine transcripts being translated. Translation cutoff determined by 95th percentile of coverage values on untranslated regions



Examples:





LEFT: Depiction of the mouse HDGF gene's 5'-extension predicted by RIBO-seq and identified using N-terminal proteomics experiment.

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.

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** [1].

Future work will mainly focus on :

➔ Investigating read density and coverage to identify genes undergoing translation.

➔ Implementing SVM in TIS calling algorithm

REFERENCES

[1] Crappé, J., Ndah, E., Koch, A., Steyaert, S., Gawron, D., De Keulenaer, S., De Meester, E., De Meyer, T., Van Criekinge, W, Van Damme, P. and Menschaert, G. (**2014**) PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. Nucl. Acids Res. (11 March 2015) 43 (5).

[2] Lee, S.,et al. (2012) Global mapping of translation initiation sites in mammalian cells at single nucleotide resolution. Proc. Natl. Acad. Sci. U.S.A. 109, E2424–E2432.

[3] Ingolia, N et al. (**2011**) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147, 789–802.

[4] Michel, A. et al. (**2013**) Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. Wiley Interd. Rev. RNA. Epub ahead of print.

[5] Dobin A, et al. (**2013**) STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29 (1): 15-21.

[6] Menschaert, G, et al. (**2013**) Deep proteome coverage based on ribosome profiling aids MS-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. Mol Cell Prot.

AFFILIATIONS

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

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