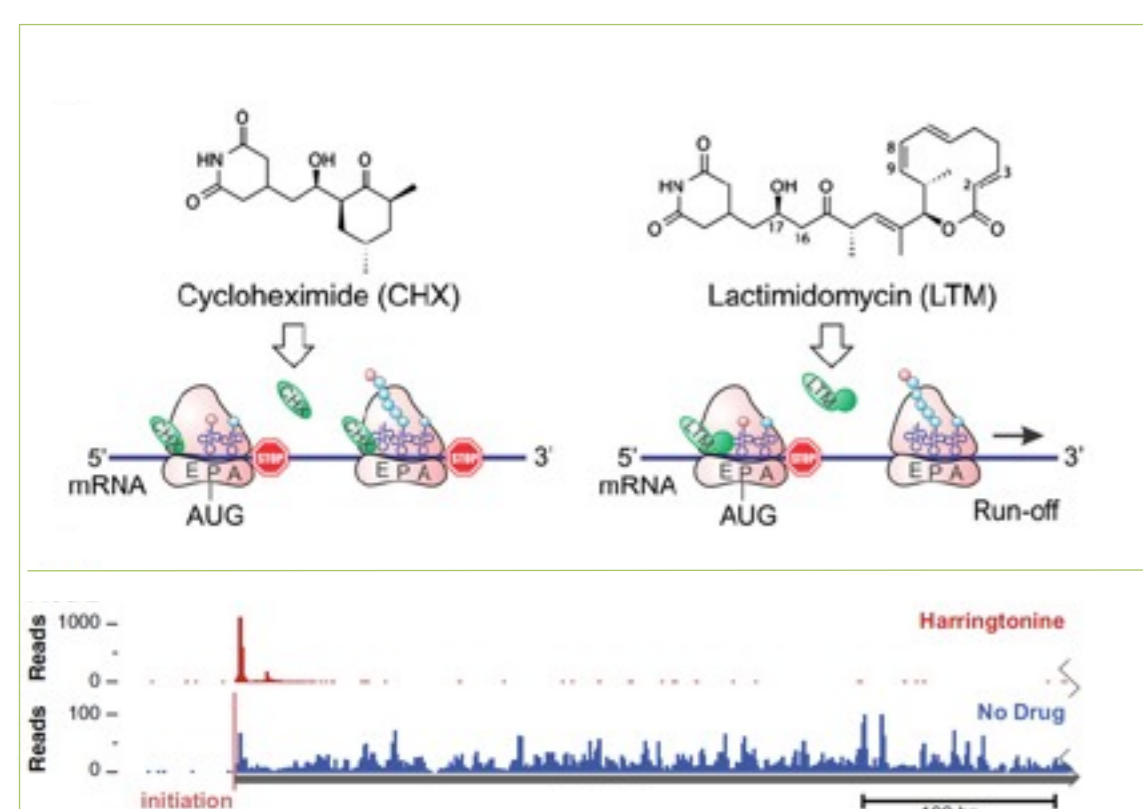
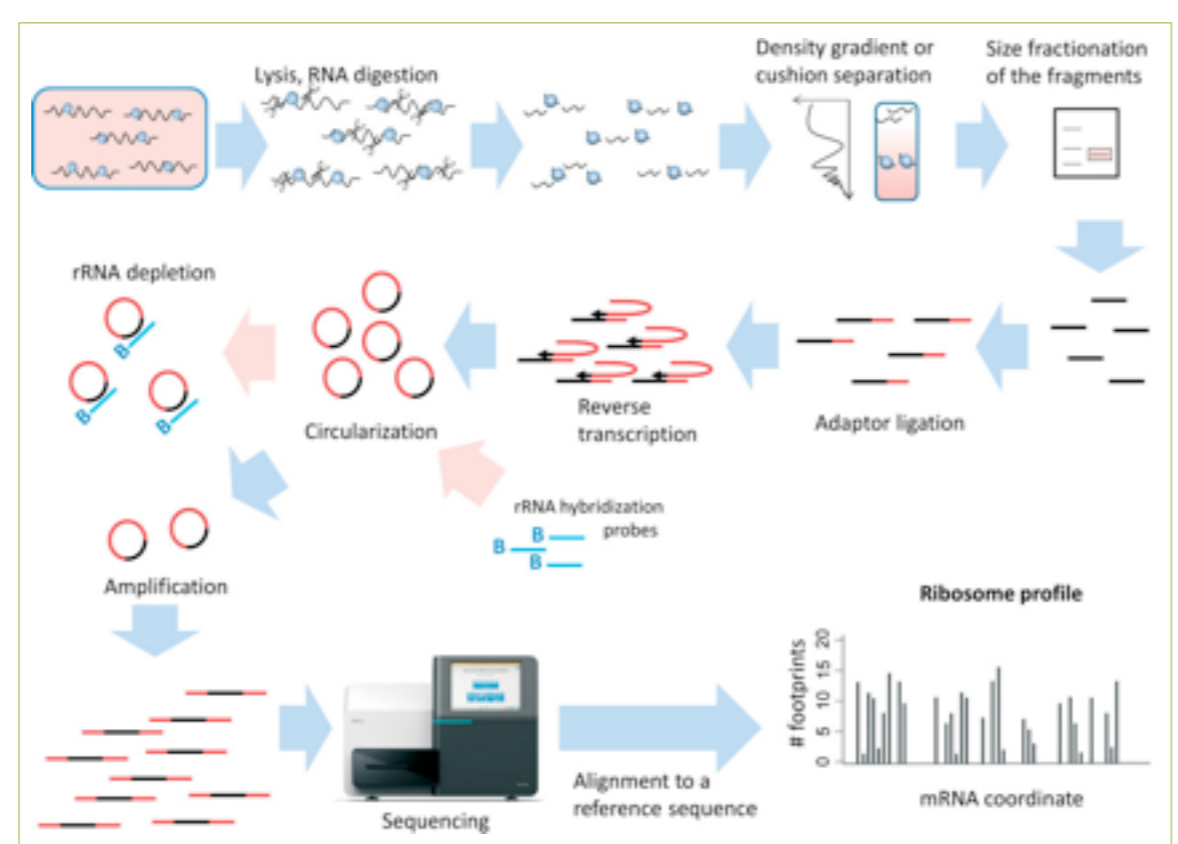


# Mass spectrometry and ribosome profiling, a perfect combination towards a more comprehensive identification strategy of true *in vivo* protein forms.

Jeroen Crappé<sup>\*a</sup>; Alexander Koch<sup>\*a</sup>; Sandra Steyaert<sup>a</sup>; Wim Van Crielinge<sup>a</sup>; Petra Van Damme<sup>b,c</sup>; Gerben Menschaert<sup>a</sup>

## INTRODUCTION

An increasing number of studies involve integrative analysis of gene and protein expression data, taking advantage of new technologies such as next-generation transcriptome sequencing (RNA-Seq) and highly sensitive mass spectrometry (MS). Recently, a strategy, termed **ribosome profiling**, based on deep sequencing of ribosome-protected mRNA fragments, indirectly **monitoring protein synthesis**, has been described. When used in combination with **initiation-specific translation inhibitors**, it enables the identification of (alternative) translation initiations.

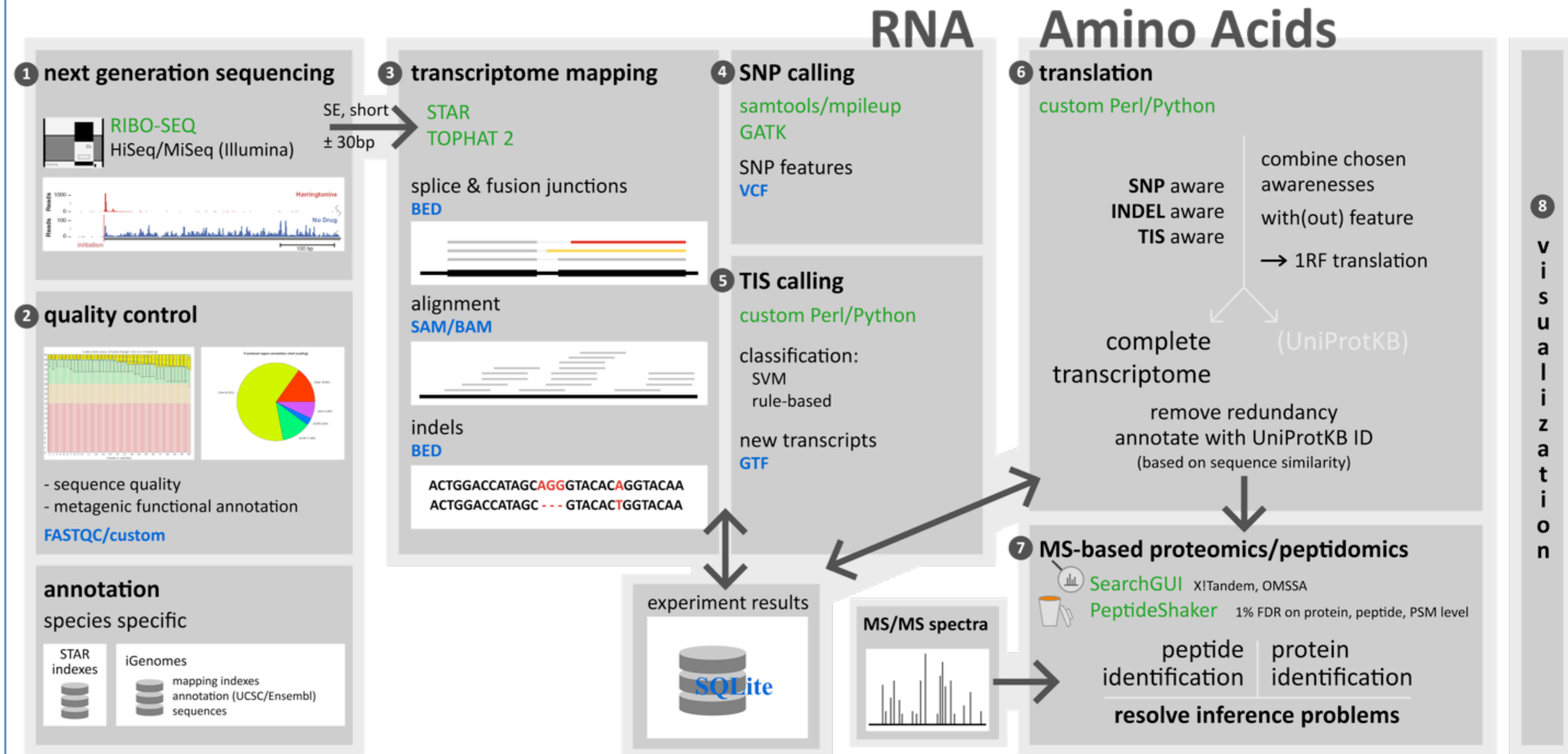


In contrast to routinely employed protein databases in proteomics searches, RIBO-seq derived data gives a **more representative expression state** and accounts for sequence variation information (**single nucleotide polymorphism, insertions, deletions and RNA-splice variants**) and **alternative translation initiation** leading to **N-terminal extended and/or truncated protein forms**. Furthermore, RIBO-seq reveals translation start at **near-cognate start sites**. Without taking this information into account, MS-based proteomic studies may fail to detect novel, important protein forms.

## GOALS

- ✓ Compile a **sample-specific protein search database** based on ribosome profiling sequencing data.
- ✓ Introduce **new translation products** in the MS search space: N-terminal extensions/truncations, trans-lated uORFs, near-cognate start sites.
- ✓ Bridging two omics worlds: **transcriptomics & MS-based proteomics** by means of **RIBO-seq**.

## MATERIALS & METHODS



### WORKFLOW OVERVIEW

- Ribosome profiling** of ribosome captured mRNA fragments performed on HiSeq, Illumina sequencing platform (**RIBO-seq**).
- Quality control** based on existing tools as FastQC and custom metagenic functional assessment.
- Transcriptome mapping** based on STAR [4] and/or TopHat2 [5] local aligners.

- Single nucleotide polymorphism (SNP) calling** based on samTools-mpileup and/or GATK.
- Translation initiation site (TIS) calling** based on trained Support Vector Machine (SVM) [2] or rule-based algorithm [1].
- Translation product assembly**, taken into account the SNP, TIS, INDEL awareness. Construct complete proteome, optionally combine with SwissProt.

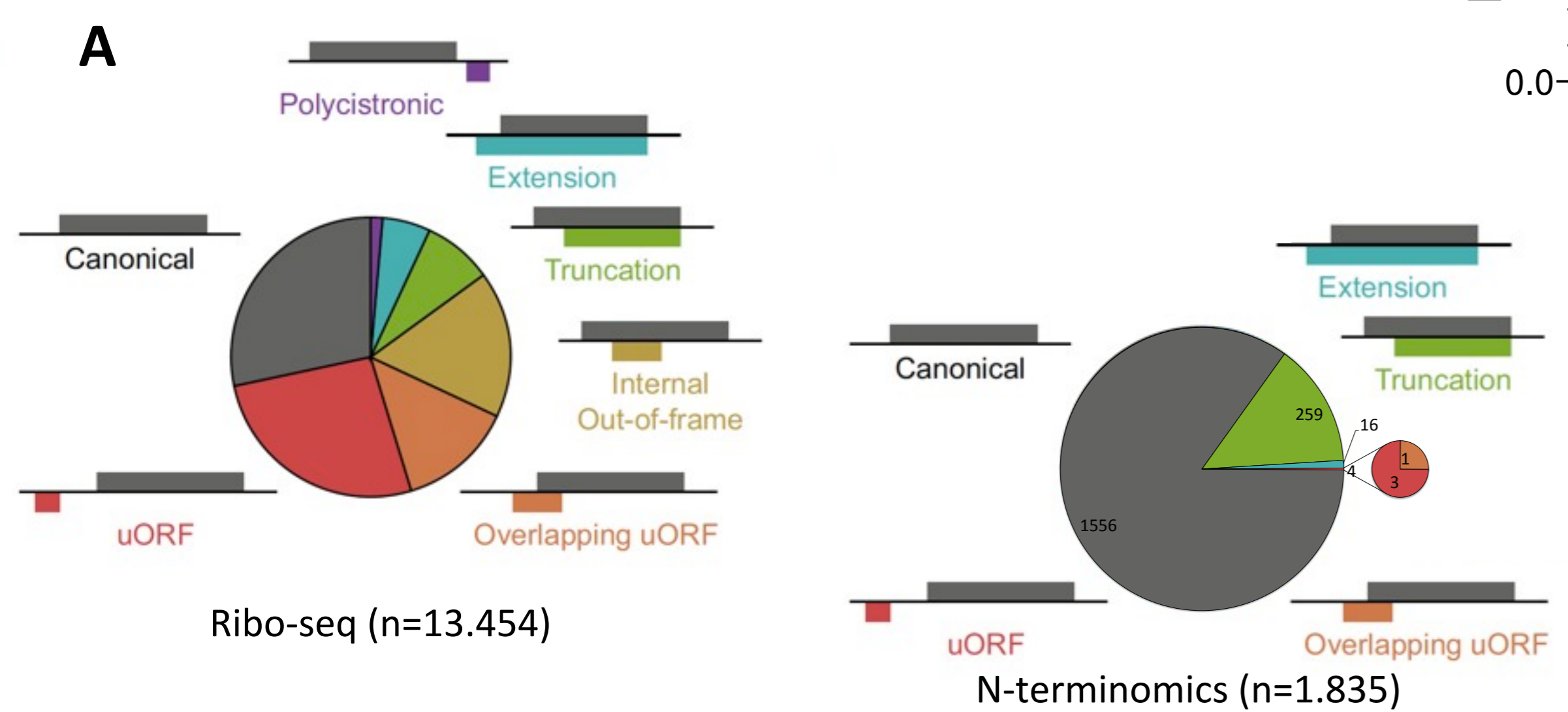
- MS-based proteomics/peptidomics** using SearchGui [6] and PeptideShaker [7] tools.
  - Genome-centric visualization** of all generated information tracks (Ensembl, UCSC, IGV genome browsers).
- ⇒ All results are stored in a **relational SQLite database** allowing further detailed analysis.

## RESULTS

Shotgun proteomics: 24 LC runs → 112.974 MS/MS spectra. Positional proteomics (N-term COFRADIC): 45 LC runs → 68.523 MS/MS spectra

Version 1 Custom DB [8] based on:

- SVM based TIS calling [2]
- Using reference genomic sequence
- Combi of RIBO-seq derived and SwissProt



(A) COFRADIC: Pie charts depicting the **identification of novel translation products**: N-terminal extensions/truncations, uORF translations, internal out-of-frame translations (left panel is RIBO-seq based, right panel is MS-based using custom DB). (B) SHOTGUN: Pie chart showing an **+2.5% gain in protein identification** using the combined RIBO-seq derived and SwissProt search database. (C) Weblogos depicting the sequence context (three bases upstream and four bases downstream) of the newly identified translation initiation sites, clearly pointing to **near-cognate translation initiation**.

## 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 :

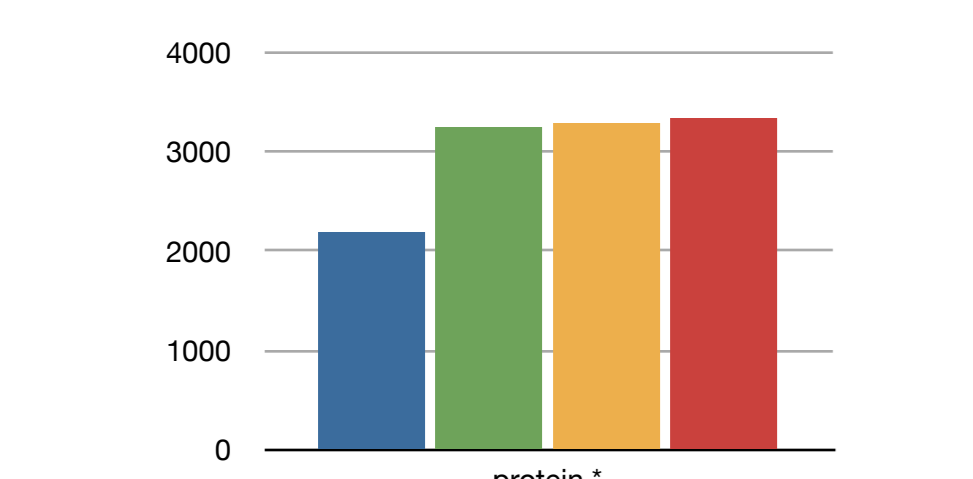
- Further investigation of the differential expression on translation level of UniProtKB-SwissProt and RIBO-seq derived translation products: technological and/or biological relevance? Detailed assessment of the difference between *in vivo* measurement of **protein synthesis** (RIBO-seq) and **protein presence** (MS-based proteomics).
- **Generalize the pipeline** to all types of next-generation sequencing RNA-seq data: (directional) A+/A- RNA-seq, CLIP-seq, exome-seq, ribo-seq
- **Quantitative correlation** of RIBO-seq and (non) labelled MS-based proteomics
- Incorporate Pipeline into **Galaxy-P** [9]

## REFERENCES

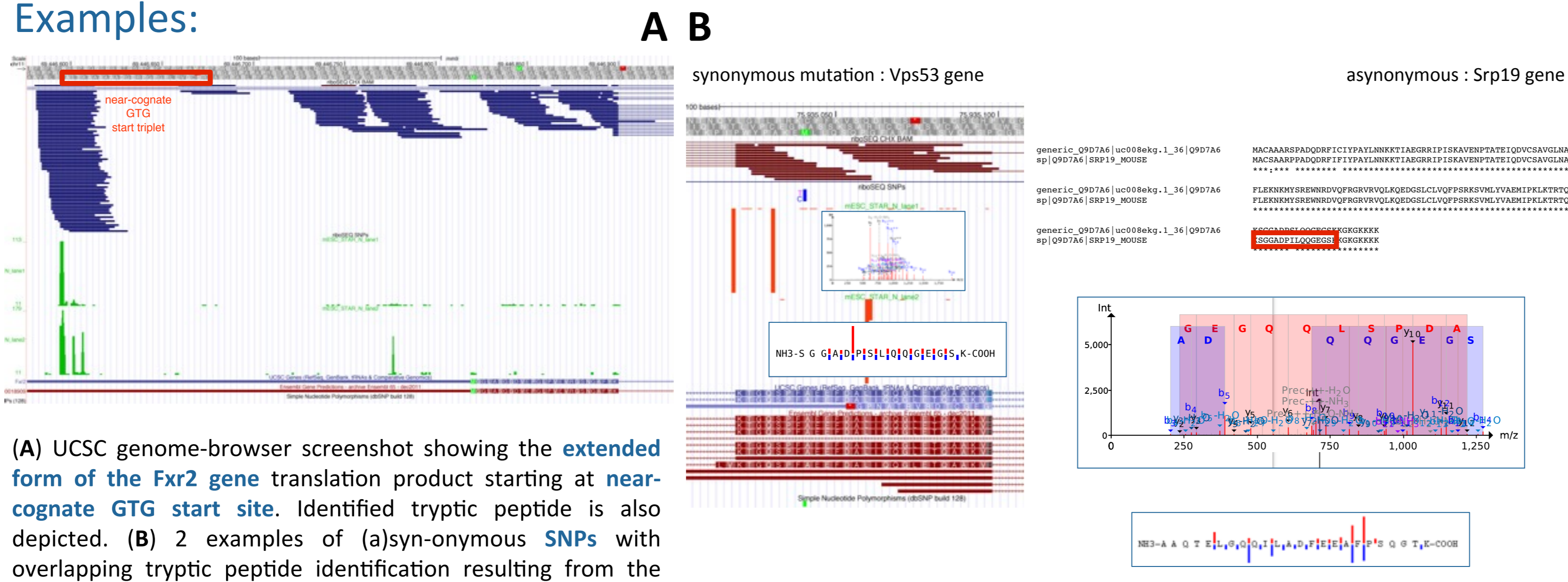
- Lee, S., Liu, B., Lee, S., Huang, S. X., Shen, B., and Qian, S. B. (2012) Global mapping of translation initiation sites in mammalian cells at single nucleotide resolution. Proc. Natl. Acad. Sci. U.S.A. 109, E2424–E2432.
- Ingolia, N. T., Lareau, L. F., and Weissman, J. S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147, 789–802.
- Michel, A. M., Baranov, P.V. (2013) Ribosome profiling: a Hi-Def monitor for protein synthesis at the genome-wide scale. Wiley Interdisc. Rev. RNA. Epub ahead of print.
- Dobin A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski C., Jha S., Batut P., Chaisson, M., Gingeras T. R. (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29 (1): 15–21.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S. L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 2013, 14:R36.
- Vaudel, M., Barsnes, H., Berven, F.S., Sickmann, A., Martens, L. (2011) SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. Mar;11(5):996–9.
- http://peptide-shaker.googlecode.com
- Menschaert, G., Van Crielinge, W., Notelaers, T., Koch, A., Crappé, J., Gevaert, K., Van Damme, P. (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. Epub ahead of print.
- http://getgalaxy.org

Version 2 Custom DB based on:

- rule-based TIS calling [1]
- Including SNP-INDEL information
- Only Ribo-Seq derived sequences



Examples:



## AFFILIATIONS

- <sup>a</sup> Laboratory of Bioinformatics and Computational Genomics, Department of Mathematical Modelling, Statistics and Bioinformatics, Faculty of Bioscience Engineering, Ghent University, B-9000, Ghent
- <sup>b</sup> Department of Medical Protein Research, Flemish Institute for Biotechnology, B-9000 Ghent
- <sup>c</sup> Department of Biochemistry, Ghent University, B-9000 Ghent
- \* Co-first authors

## DWNL



**CONTACT:** Gerben.Menschaert@UGent.be, Jeroen.Crappe@UGent.be