Ghent, 21 December 2016

Dear Dr. Eisenacher,

We are here submitting on behalf of the authors the format proBAM to the PSI document process. The format represents systematically the output of proteogenomics analyses, by mapping peptide identification data retrieved from mass spectrometry (MS)-based experiments to the genome.

The original SAM format (Sequence Alignment/Map Format, http://samtools.github.io/hts-specs/SAMv1.pdf) is widely used in the genomics community for mapping individual reads to a reference genome. It is a TAB-delimited text, human-viewable format, corresponding to the binary BAM format. The format consists of a header and an alignment section. Each alignment line has 11 mandatory fields for essential alignment information (e.g. mapping position).

The proBAM format builds upon the structure of the original SAM/BAM format by extending it with other mandatory fields to accommodate unique features on MS-based proteomics peptide-spectrum matches (PSMs) information. The main difference to the other proteogenomic PSI-drafted format probed is that proBAM holds more information including the full alignment to the underlying genomic sequence and extra peptide/PSM annotation (normal, variant, indel …). Furthermore, a complete MS study result dataset can be exported, including both decoy information and low scoring and lower rank PSMs. The latter enables to use the proBAM format to reanalyse the data using different thresholds or perform a gene level inference procedure. proBed is recommended for annotation track building. ProBAM is also meant for that purpose but moreover is recommended for further downstream analyses at the gene level.

The documents I am attaching here are:

- proBAM draft version 1.0.0 specification document in pdf and Word formats.

In addition, four different example proBAM files (and all the related files) are linked from http://www.psidev.info/probAM. Links to proBAM tools and implementation can also be found on the site.

Best regards,

Dr. Gerben Menschaert and Dr. Xiaojing Wang
Ghent University Baylor College of Medicine
Ghent, Belgium Houston, Texas, US
on behalf of all authors (Dr. Eric W. Deutch, Dr. Andrew A. Jones, Dr. Juan Antonio Vizcaino)