

Onco-proteogenomics: a novel approach to identify cancer-specific mutations combining proteomics and transcriptome deep sequencing

Mohamed Helmy^{1,2*}, Naoyuki Sugiyama¹, Masaru Tomita¹, Yasushi Ishihama¹

From Beyond the Genome: The true gene count, human evolution and disease genomics
Boston, MA, USA. 11-13 October 2010

Background

The accumulation of somatic mutation is a common property in all cancer genomes. These mutations include several patterns of mutagenesis such as small insertions, chromosomal rearrangement and nucleotide substitutions. Consequently, the mutated genomes produce mutant transcriptome and, therefore, mutant proteins that give the cancer cell its oncogenic properties [1]. For such mutated proteins, however, mass spectrometry-based identification by shotgun proteomics is generally difficult because the identification is dependent on databases containing normal proteins or hybrid database with normal and mutated proteins. Here, we present 'onco-proteogenomics, a novel proteogenomics approach to identify the cancer-related peptides (phospho- and non-phospho peptides) and proteins.

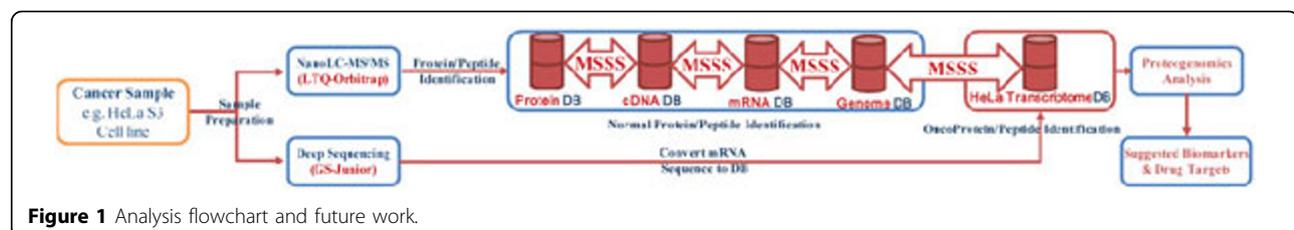
Methods

We analyzed 15 MS/MS runs of HeLa S3 cells, as a test sample, by shotgun proteomics and phosphoproteomics. The obtained data was analyzed by an extended version of MSSS (MS Spectra Sequential Subtraction), the proteogenomic approach that we used before in the identi-

fication of novel genomic features in Rice plant [2]. In our onco- proteogenomic approach, we used four databases containing normal sequences (Human protein, cDNA, mRNA and genome databases) for Mascot peptide identification and removed all the MS/MS spectra that corresponds to all identified peptides. The reminder MS/MS spectra were searched against one cancer-driven database obtained through deep sequencing of HeLa S3 cells to identify cancer-specific peptides.

Results

The four databases that contain normal sequences were used sequentially to identify all potential peptide sequences and phosphorylation sites that can be generated from the normal genome. This includes the potential protein sequences, junction-peptides and exon-skipping peptides (protein and cDNA databases), exonic peptides (mRNA database) and extragenic peptides (genome database). Following each Mascot search, we removed all the MS/MS spectra corresponding to the identified peptide sequences and created new files containing the reminder MS/MS spectra. Next, we constructed HeLa S3 transcriptome



¹Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan
Full list of author information is available at the end of the article

database with data obtained from deep sequencing of HeLa S3 cells (obtained from NCBI UniGene Database). The constructed database contains over 60,000 entries. For the remaining unidentified MS/MS spectra, we performed Mascot search against this transcriptome database. Consequently, we were able to identify 25 cancer-specific peptides including phosphorylated sites. For further check, the identified peptides were aligned to the employed normal databases using NCBI BLAST. The alignment did not show any significant matches indicating that these peptides are specifically expressed in the HeLa S3 cancer cell-line. In future studies, we will apply the same approach in different cancers aiming to identify global cancer biomarkers and drug targets Figure 1.

Author details

¹Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan.

²Systems Biology Program, Graduate School of Media and Governance, Keio University, Japan.

Published: 11 October 2010

References

1. International network of cancer genome projects. *Nature* 464:993-998.
2. Helmy M, Tomita M, Ishihama Y: Novel features for rice genome revealed using proteogenomic analysis. *The 10th international conference on Systems Biology (ICSB2009)* Stanford University, CA, USA 2009.

doi:10.1186/gb-2010-11-S1-P17

Cite this article as: Helmy *et al.*: Onco-proteogenomics: a novel approach to identify cancer-specific mutations combining proteomics and transcriptome deep sequencing. *Genome Biology* 2010 **11**(Suppl 1): P17.

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

