Bachelor Thesis
in Bioinformatik

A Novel Approach for Finding Black Cats in Black Rooms

Der Kleine Bioinformatiker

LUDWIG-MAXIMILIANS-UNIVERSITÄT
TECHNISCHE UNIVERSITÄT MÜNCHEN

**Lehrstuhl für
Bioinformatik**

Bachelor Thesis
in Bioinformatik

A Novel Approach for Finding Black Cats in Black Rooms

Der Kleine Bioinformatiker

Aufgabensteller:	Prof. Dr. X
Betreuer:	Dr. Y
Abgabedatum:	15. Mai 2015

Ich versichere, dass ich diese Bachelor Thesis selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

15. Mai 2015 __________________________

Der Kleine Bioinformatiker

Abstract

An abstract abstracts the thesis...

Zusammenfassung

Eine Zusammenfassung fasst die Arbeit zusammen...

Contents

1 Background

The proteome is the entire set of proteins that is, or can be, expressed by a cell, tissue, or organism at a particular time and under a specific set of conditions. Unlike the genome, which is largely static, the proteome is highly dynamic, reflecting the cell's current functional state and its response to internal and external cues **Aebersold2016**. Proteins are the primary molecular machines that execute virtually all biological processes, including enzyme catalysis, cell signaling, and structural support **Alberts2017**.To significantly increase the functional diversity beyond the genetic code, proteins undergo Post-Translational Modifications (PTMs), which are chemical alterations that occur after a protein has been synthesized. PTMs are critical regulatory mechanisms, with over 400 types known, including phosphorylation, acetylation, and ubiquitination **Khoury2011**. These modifications act as molecular switches, profoundly influencing cellular processes by regulating a protein's activity (e.g., activating or deactivating enzymes), subcellular localization, stability, and interaction with other molecules **Hunter2012**; **Walsh2006**. Consequently, the accurate characterization of PTMs is essential for understanding cell biology and disease states.The standard method for large-scale protein analysis is Mass Spectrometry (MS)-based proteomics, specifically the bottom-up approach. In this workflow, the complex protein mixture is first enzymatically cleaved into smaller fragments called peptides. These peptides are then separated by liquid chromatography (LC) and introduced into the mass spectrometer. The MS instrument performs two main steps: in MS1, it measures the mass-to-charge ratio (m/z) of the intact peptide (the precursor ion); then, in MS/MS (MS2), it isolates a precursor ion, fragments it (e.g., via HCD), and measures the m/z of the resulting fragment ions. These fragment ion masses contain the sequence information of the peptide **Domon2015**.The most common computational method to identify the peptide sequence from the MS/MS spectrum is database searching **Yates1995**. A search algorithm compares the experimentally measured fragment ion spectrum against a comprehensive sequence database (like a FASTA file of all known proteins for an organism). The algorithm calculates the theoretical fragment masses for every possible peptide in the database, including predefined PTMs, and assigns the best-matching sequence (Peptide-Spectrum Match, PSM) by scoring the match quality **Mann2001**.However, this databasedependent strategy faces a critical limitation when dealing with PTMs. To identify a PTM, the modification must be specified in advance as a variable modification in the search. Adding even a small number of variable PTMs dramatically increases the size of the search space, leading to a combinatorial explosion of possible peptide sequences that must be tested **Tanner2005**. This greatly increases the computation time and, more critically, raises the False Discovery Rate (FDR), making confident identification difficult **Chalkley2019**. Furthermore, the method is inherently incapable of identifying novel or unexpected PTMs that are not included in the predefined search parameters, thereby limiting the scope of biological discovery.

2 Materials and Methods

3 Results and Discussion

4 Conclusion