Data Analysis Project 2022

Report

Institute of Pharmacy and Molecular Biotechnology

Topic 03: Proteome-wide Screen for RNA-dependent Proteins

Sub-project 2: HeLa Cells Synchronized in Mitosis

Supervisor: Dr. Maïwen Caudron-Herger Tutor: Niklas Engel

Students: Michel Tarnow, Michèle Bennek, Lennart Müller, Sebastian Rickert

Date: 20.07.2022

Table Of Contents

Introduction

RNA-binding proteins (RBPs) are a group of proteins who share the characteristic of directly binding to RNA molecules (Gebauer et al., 2021). Thus formed complexes are termed ribonucleoprotein particles (RNPs), prominent examples of which include the spliceosom (Corley et al. 2020), the signal recognition particle (Faoro and Ataide, 2021) and the RNA-induced silencing complex (Pratt and MacRae, 2009). RBPs play key regulatory functions in the life cycle of RNA molecules, including transcription, splicing, modification, intracellular trafficking, translation and degradation (Gebauer et al., 2021).

RBPs are known to bear discrete domains that are responsible for the molecular interactions between amino acid residues and RNA nucleotides that lead to RNA binding, and which for this reason are defined as RNA-binding domains (RBDs), such as the RNA recognition motif, the K-homology domain or DEAD/DEAH helicase and zinc-finger domains (Corley et al., 2020). Although the occurrence of one or more of these domains is characteristic of RBPs, more recent research has identified RBPs that lack known RBDs but instead bind RNA through intrinsically disordered regions (Gebauer et al., 2021), indicating the heterogeneity of this group of proteins.

Since RBPs are key players in RNA metabolism, their malfunction has been associated with diverse pathological phenotypes, including cancer (Zhang et al., 2020) and neurodegenerative disorders (Maziuk et al., 2017). Therefore, identifying new RBPs may contribute to point out novel drug targets for already known conditions on the one hand and to shed new light on the diseased mechanisms of yet unresolved conditions on the other.

While the study and quantification of RBPs has received much attention in previous research, the exact number of mammalian RBPs remains a matter of debate. Techniques for proteome-wide screening of RBPs in the past heavily relied on pull-down assays of polyadenylated RBPs (Beckmann et al., 2015; Castello et al., 2012), protease digestion (Mullari et al., 2017) or UV cross-linking (Urdaneta et al., 2019), each method having its advantages and shortcomings. Thus, current estimations about the number of RBPs remain largely inconsistent across studies, leading only to a small consensus set of around 200 proteins (Caudron-Herger et al.,

2019).

While not devoid of its own limitations, the here presented method provides a novel approach to the proteome-wide screening of RBPs with the goal of validating the core set of RBPs previously defined while also contributing to adding new proteins to the number of possible RBP candidates.
For this purpose, the notion of RNA dependence was introduced, defining a protein as RNA dependent when its interactome depends on RNA, while binding to RNA is sufficient but not necessary to qualify as RNA dependent. Thus, every RNA binding protein is RNA dependent, but not every RNA dependent protein is RNA binding (Caudron-Herger et al., 2019).

The method here employed relies on the differential migration pattern of proteins in a sucrose density gradient ultracentrifugation in the presence or absence of RNA.
In density gradient centrifugation, macromolecules such as proteins are forced through a density gradient until they find a density equal to their own (Farrell, 2010). Thereby the rate of sedimentation is function of size, shape and density of the macromolecules, as well as density and viscosity of the gradient and applied centrifugal force (Raschke et al., 2009).
Here, triplicates of native (control) and RNase-treated cell lysates from HeLa cells synchronized in mitosis were loaded onto a 5% to 50% sucrose density gradient. Upon ultracentrifugation, the gradient was divided into 25 fractions and the protein amount of individual proteins per fraction determined by quantitative mass spectrometry. Because RNA-dependent proteins directly or indirectly depend on the presence of RNA, they are expected to show a different migration pattern between control and RNase-treated groups (Caudron-Herger et al., 2019).

In this report, we seek to draw a qualified conclusion about the RNA-dependence status of the 7160 proteins investigated in the provided data set using bioinformatics and statistics. For this purpose, we will define criteria and parameters the given proteins need to fulfill to qualify as RNA-dependent, and will evaluate our given data with respect to aforementioned aspects using the programming language R. Eventually, we will discuss the significance and possible limitations of our methods and findings.