Quantitative High-Resolution Analysis of Cell Cycle Regulated Protein and Phosphorylation Abundance in Non-Transformed Cells

By

Paul Anderson

A Master’s Thesis

Barts and The London School of Medicine and Dentistry

Queen Mary University of London

August 2025

Word count:

# Acknowledgements

I would like to express my sincere thanks to Doctor Tanya Soliman of the Barts Cancer Institute for providing the needed data for the project and for her helpful advice and supervision throughout.

Table of Contents

[Acknowledgements 2](#_Toc203464294)

[Abstract 3](#_Toc203464295)

[Introduction 4](#_Toc203464296)

[Aims and Objectives 5](#_Toc203464297)

[Materials and Methods 6](#_Toc203464298)

[Results 7](#_Toc203464299)

[Discussion 8](#_Toc203464300)

[Conclusion 9](#_Toc203464301)

[References 10](#_Toc203464302)

# Abstract

The cell cycle consists of a highly regulated sequence of events that result in the cell splitting into two daughter cells. It consists of several stages, each of which is covered by changes in protein and phosphorylation abundance.

Various papers have addressed changes in these abundance levels in cancerous cells. This paper addresses changes in abundance TODO – copy from Tanya. and with a higher greater temporal resolution than the bulk of other investigations.

# Introduction

# Aims and Objectives

The aim of the study was to identify proteins and phosphorylation events relevant to the cell cycle process.

# Materials and Methods

The data consists of two spreadsheets of abundances, one for proteins and one for phosphopeptides. Each entry has readings for ten timepoints within the cell cycle. These data are analysed by software written in Python, pytest, Django, Postgres, Docker and D3.js. The steps in the analysis consist of:

Clean and import protein abundances by timepoint from proteomics spreadsheet.

Clean and import phosphopeptide abundances by timepoint from phosphopeptides spreadsheet.

Calculate medians for raw protein and phosphopeptide data.

Calculate log2 means and various metrics (ANOVA and Fischer p value and q values TODO - confirm) for proteins and phosphopeptides.

Calculate protein and phosphopeptide oscillations.

Identify relevant proteins and phosphopeptides – those with low q values (TODO – confirm ideal q values) and high curve fold changes (> 1.2)

Classify by group, complex and subcellular localisation (TODO page 6 ‘Biological Role of cell-cycle dependent proteins – function written for CORUM).

Identify those of above that are high/low variability and oscillation.

Focus on three categories: protein kinases, components of the ubiquitin-proteasome system (UPS) and transcription factors (TFs). (TODO maybe? End of page 6).

Determine link between CCDs and transcription (TODO page 8 paragraph 2).

integrated protein half-life data (TODO page 8 paragraph 2 end).

Connection between CCD proteins and mean gene effect (TODO page 8, paragraph 3)

Percentage of phosphorylation accessibility and disorder sites, phosphorylation site conservation (TODO end of page 8, Phosphorylation dynamics driving cell cycle progression – rewrite ICR code)

Taxonomic range of phosphorylation events (TODO page 9)

Predict protein kinases responsible for phosphorylation events (TODO end of page 9 – rewrite ICR code)

Phosphorylation oscillation relevance (TODO page 10 paragraph 2)

Degrons (TODO page 10 ‘Degradation mechanisms of CCD proteins’)

Generated various charts from the data (Fig TODO)

SEE ALSO page 14, Defining cell cycle-dependent (CCD) proteins and phosphorylation sites

Outline how the curve fold threshold of 1.2 was determined, if done (Page 15, Protein curve fold change cut-offs were defined by quantifying theenrichment of the‘cell cycle’ term).

# Results

Where the phosphorylation sites are and whether they’re disordered or accessible (end of page 8).

# Discussion

Page 12

# Conclusion

# Additional Information

# References