The Human Transcription Factor Proteome

Andrew Stergachis
Stamatoyannopoulos Lab
Dept. of Genome Sciences
University of Washington

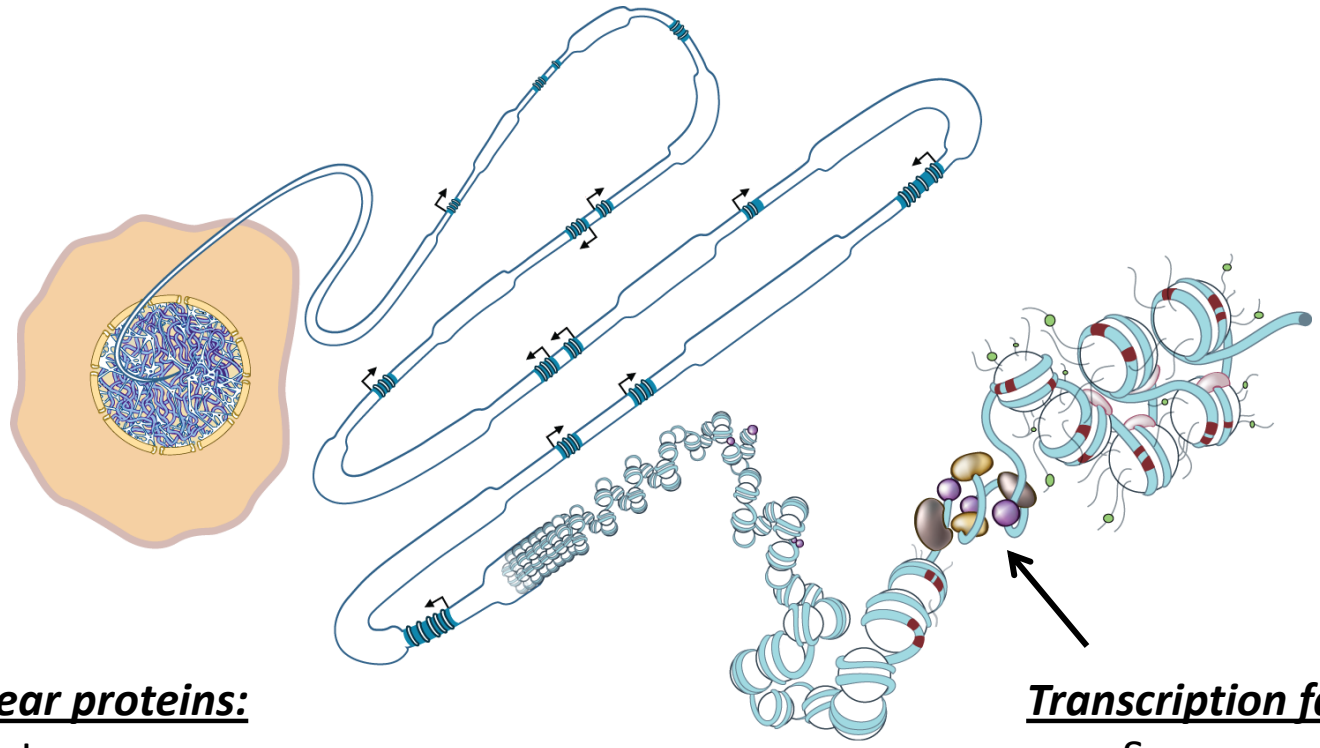
Overview

Part I – Brief background on transcription factors (TFs)

Part II – High-throughput generation of SRM methods

Part III – Compartmentalization of TFs within the nucleus

The nuclear proteome



Nuclear proteins:

Histones
Structural components
Ribosomal proteins
…

Millions of copies per nuclei

Transcription factors (TFs):

Sequence-specific DNA-binding proteins that mediate transcriptional regulation

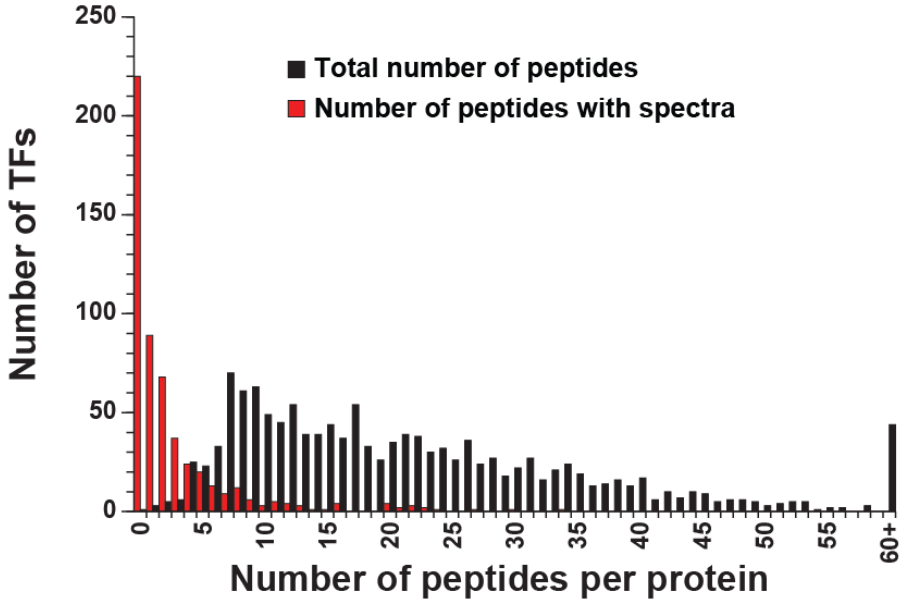
(~1,400 human TFs)

Thousands of copies per nuclei

Need a new experimental paradigm for TFs

Most human TFs lack good antibodies

Most human TFs have gone unseen by 'shotgun' proteomics approaches



Selected Reaction Monitoring (SRM) should be more sensitive

Need to identify for each TF protein:

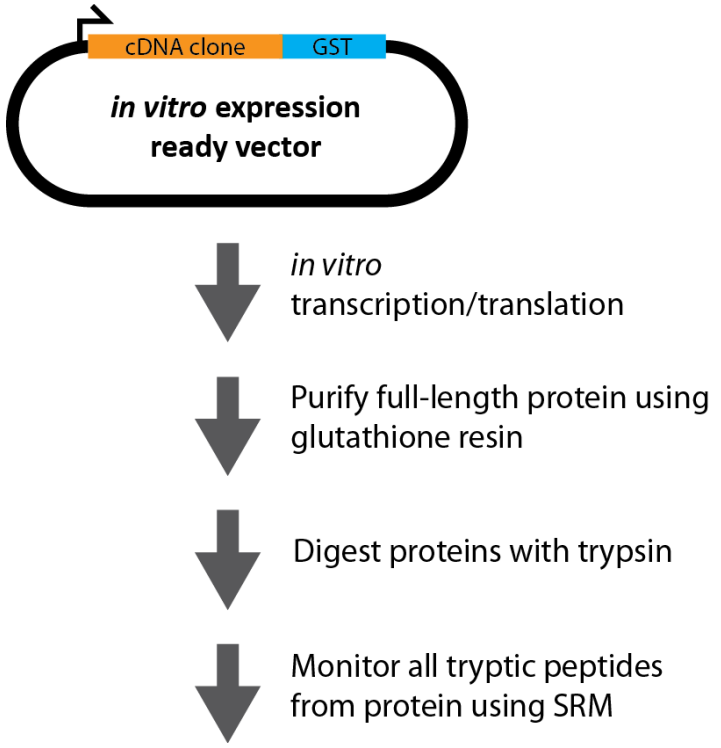
1. Best responding, 'proteotypic,' peptides
2. Fragmentation patterns of these 'proteotypic peptides

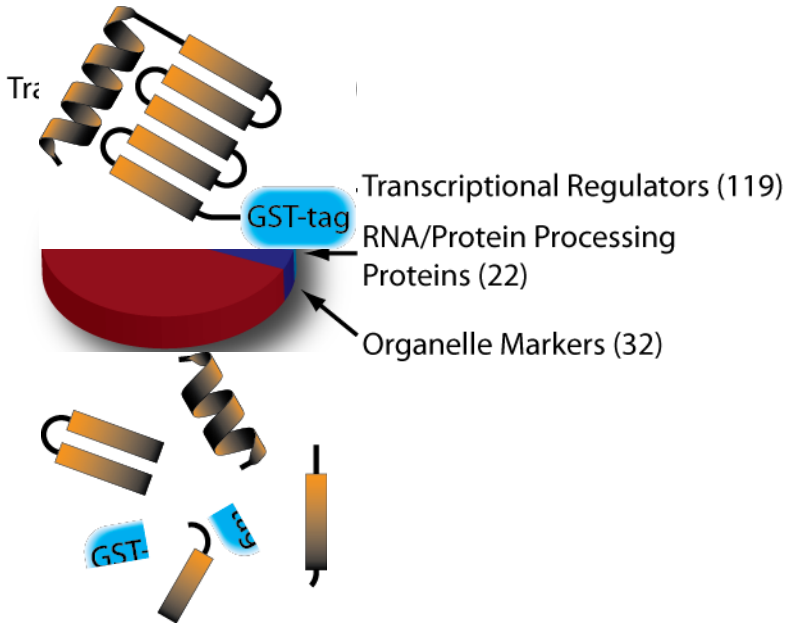
2011_05_26 NIST H. sapiens Ion Trap peptide spectral library

Part II

High-throughput empirical generation of SRM methods

Empirical identification of proteotypic peptides



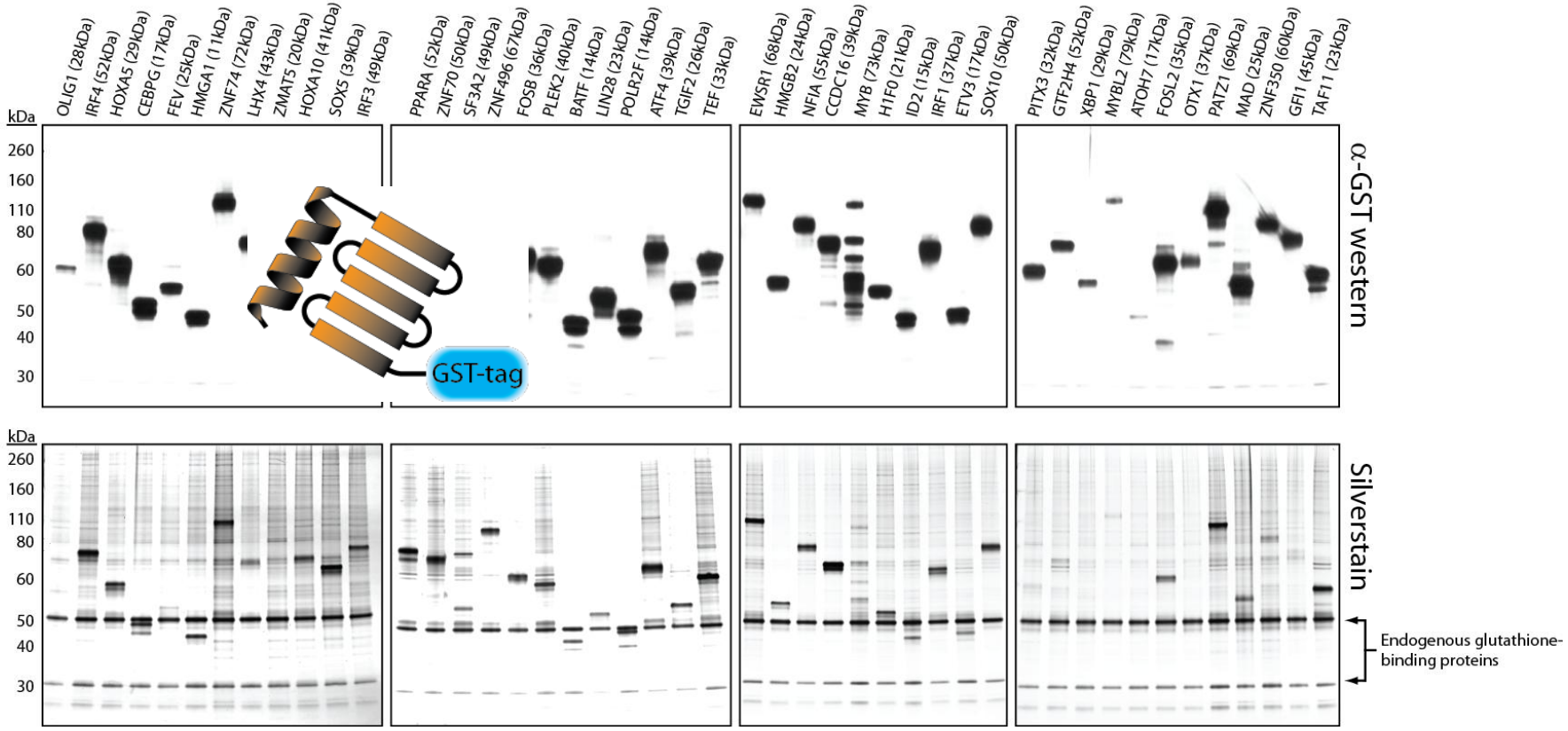


Able to identify:

1. Best responding, 'proteotypic,' peptides
2. Fragmentation patterns of these 'proteotypic peptides

Stergachis et al., *Nature Methods* **8**, 1041 (2011).

Rapid production of enriched full-length proteins



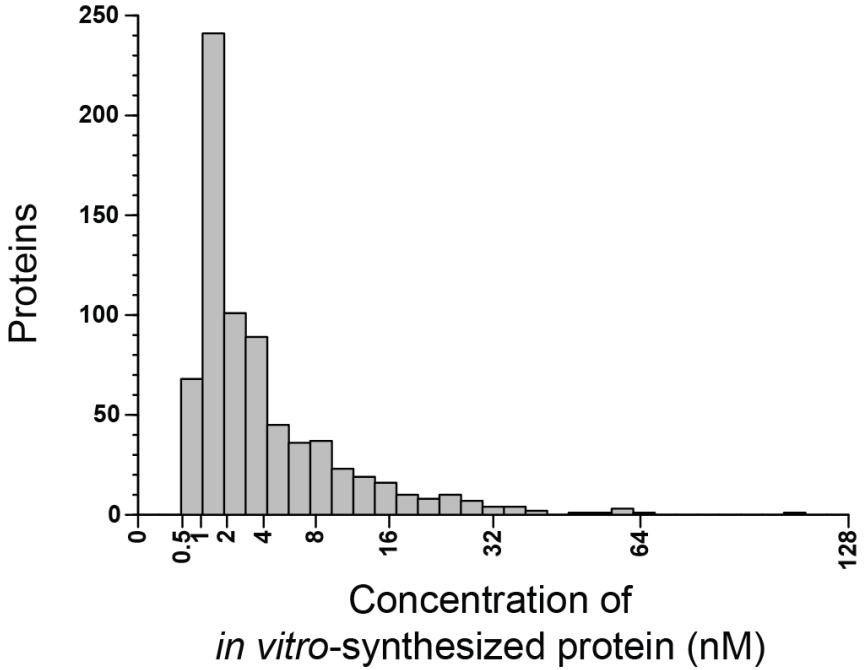
Rapid production of enriched full-length proteins

Performed absolute quantification on all 730 *in vitro*-synthesized proteins

GST peptides quantified:

LLLEYLEEK

IEAIPQIDK

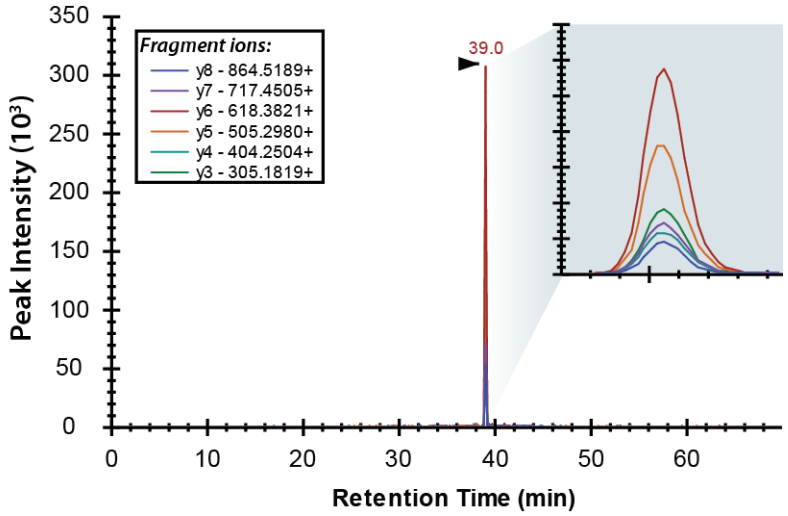


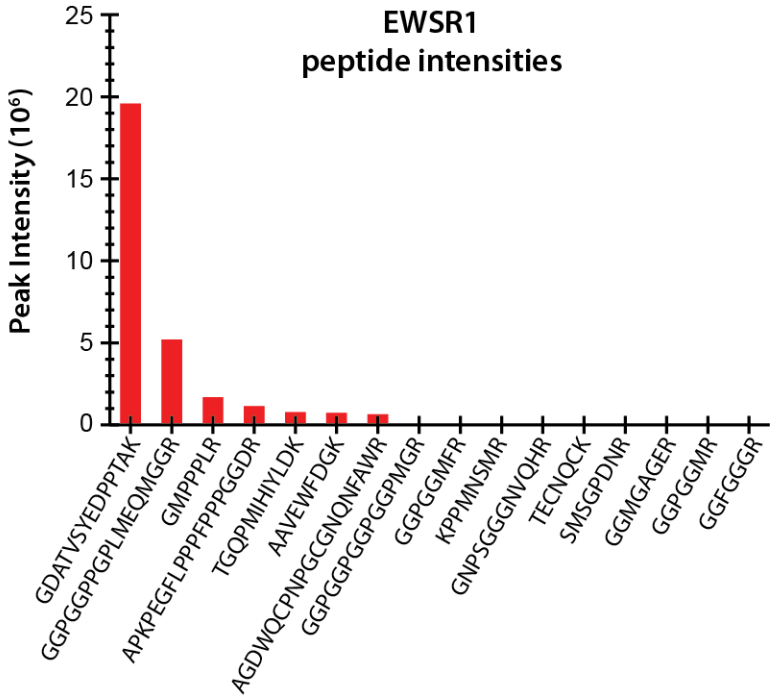
More details are available through the ***Absolute Quantification*** tutorial posted on the Skyline website

Identification of proteotypic peptides

Using Skyline, we monitored for each protein:

- Fully tryptic peptides
- 7-23 amino acids in length (+2 charge state monoisotopic)
- y_3 to y_{n-1} product ions (+1 charge state monoisotopic)
- In total, we monitored >100,000 product ions





Panorama (WP407) – Vagisha Sharma

Identification of proteotypic peptides

- Data was acquired for 12,344 peptides
- Annotated each peptide to identify those of high quality (quality score 1 or 2)
- 4,927 peptides were identified with a quality score of 1 or 2

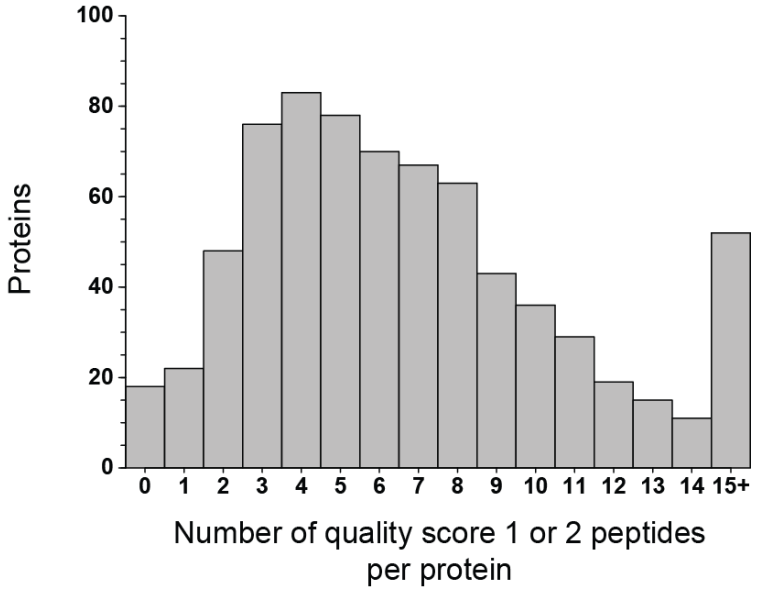
Criterion used to determine peptide quality:

A) A prominent chromatographic peak with a signal intensity of at least 60,000

B) Two or more data points were collected across the peak

C) Three or more product ions not including y_3 co-eluted to contribute to this peak signal

D) The chromatographic peak had a Gaussian elution profile

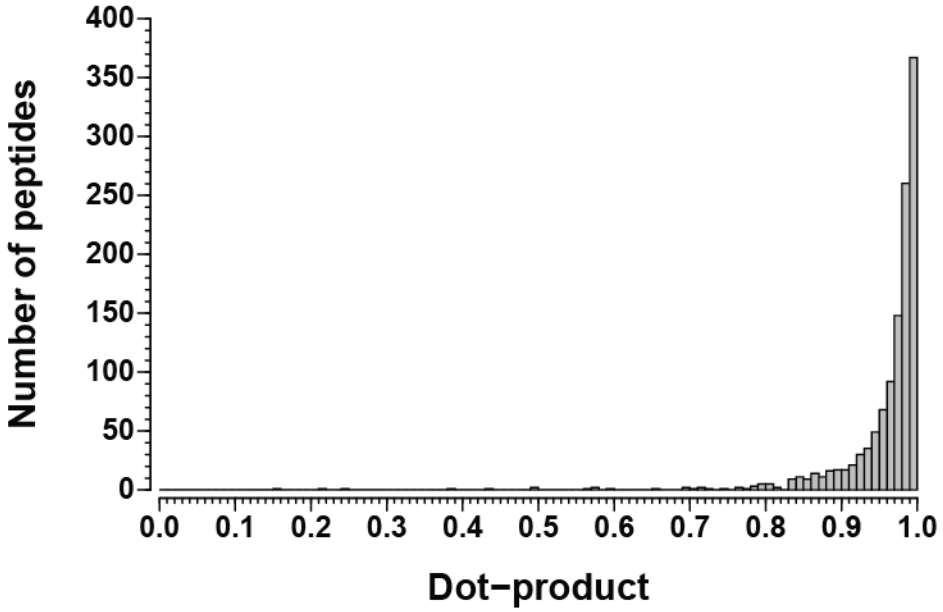


Stergachis et al., *Nature Methods* **8**, 1041 (2011).

Correspondence with spectral databases

Dot-product: Measure similarity between our SRM observed fragmentation patterns and database fragmentation patterns for the same peptide (1 = perfect match)

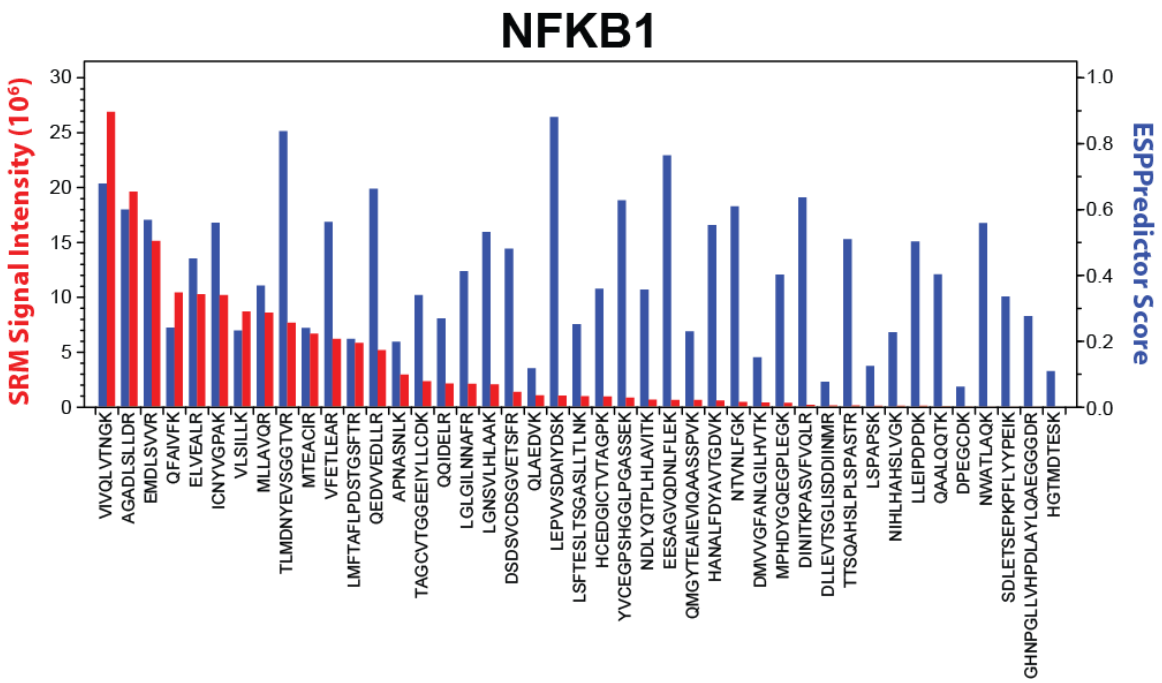
22% (1,093/4,927) of the quality score 1 and 2 peptides in our data were represented in NIST

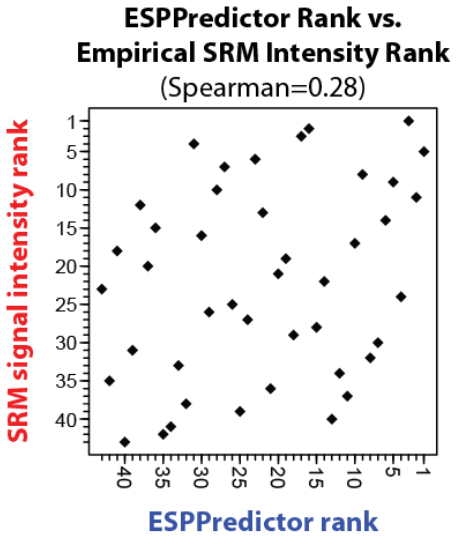


2011_05_26 NIST H. sapiens Ion Trap peptide spectral library

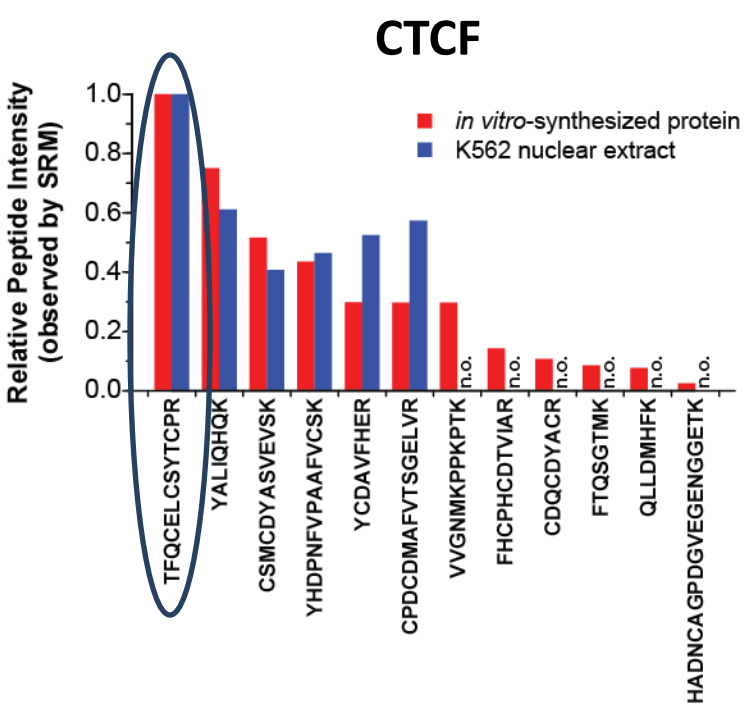
Relationship with other ranking systems

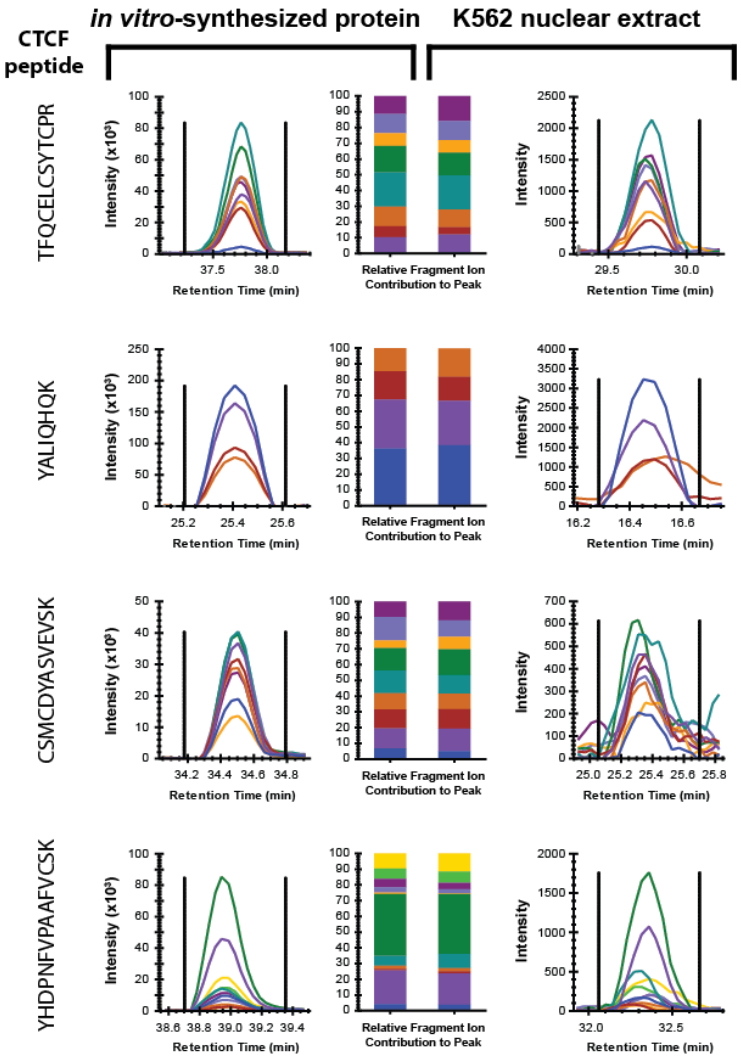
Proteins show an average Spearman correlation of 0.47 (range -0.45 to 0.85)





Identifying CTCF peptides *in vivo*

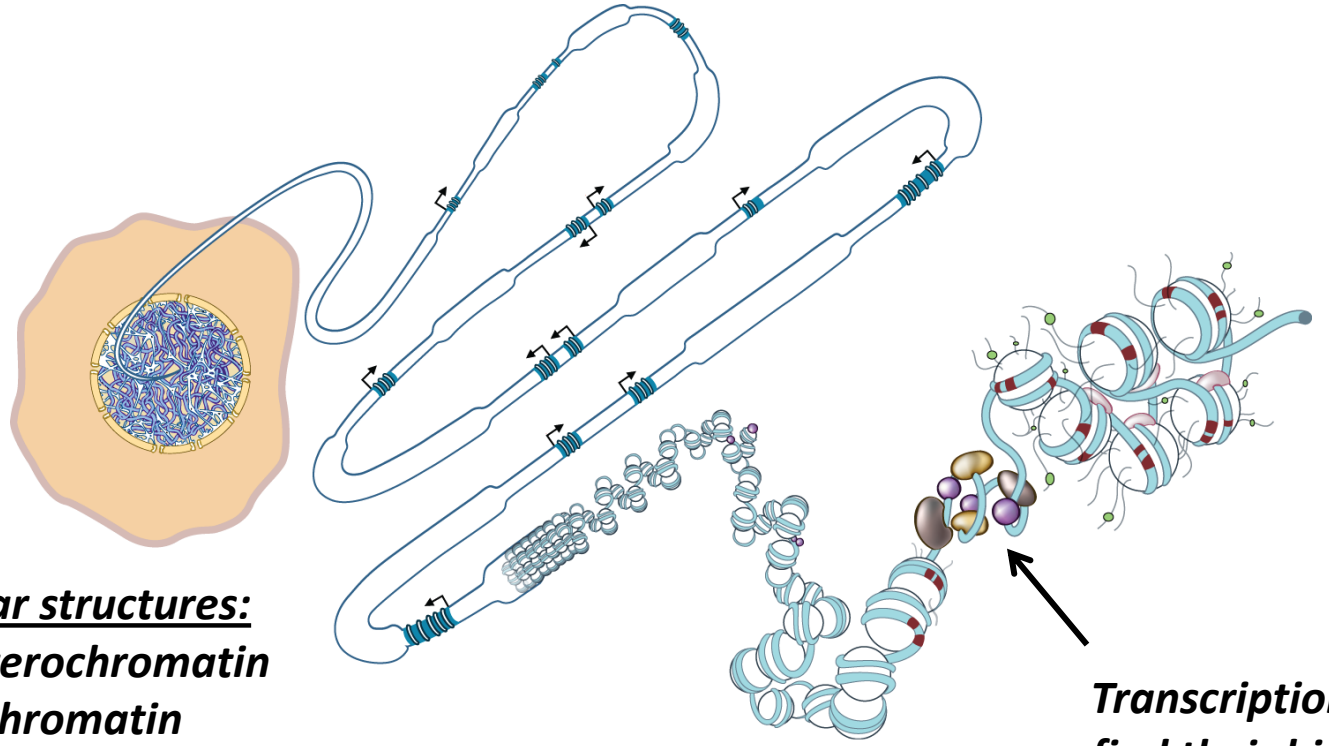




Part III

Compartmentalization of human TFs within the nucleus

Compartmentalization of the nuclear proteome

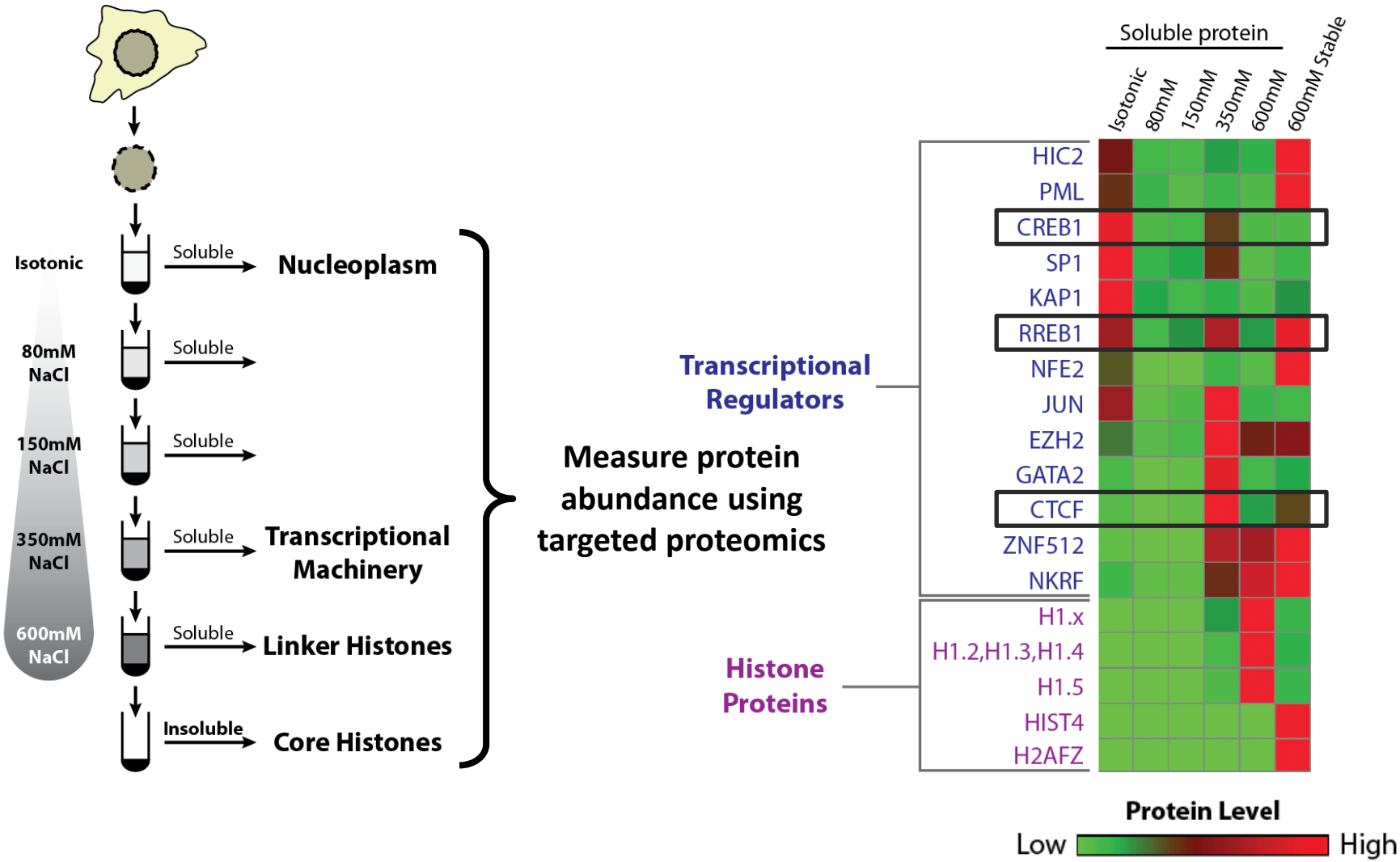


Nuclear structures:

- ***Heterochromatin***
- ***Euchromatin***
- ***Nucleoli***
- ***Splicing factories***
- ***…***

Transcription factors (TFs) must find their binding sites and recruit appropriate co-regulators

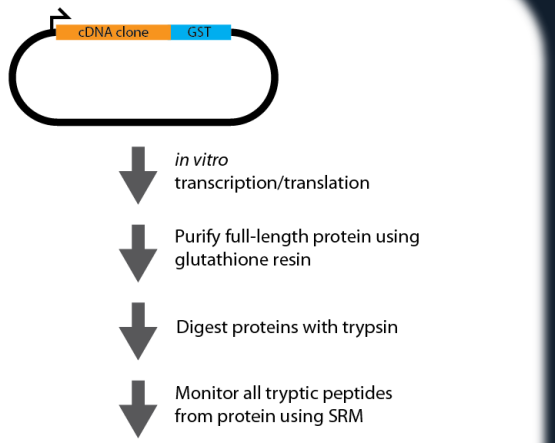
Distribution of TFs across nuclear chromatin



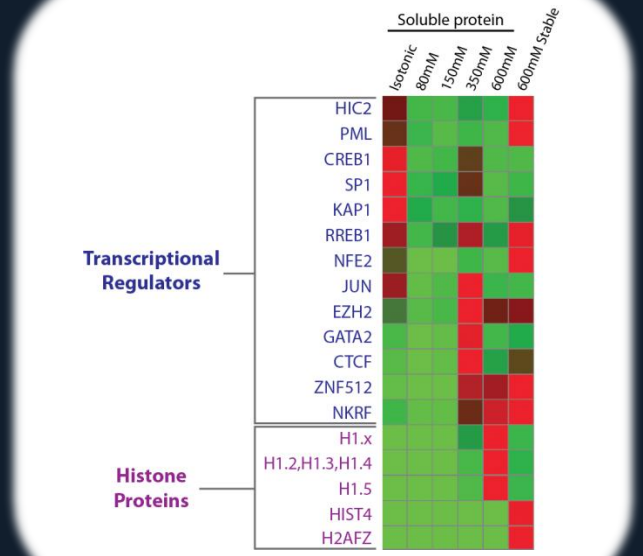
Quantification of the compartmentalization ~100 TFs in K562 nuclei (WP256)

Summary

High-throughput empirical generation of SRM methods



Compartmentalization of human TFs within the nucleus



Acknowledgments

Proteomics

Michael MacCoss (UW)
Brendan MacLean (UW)
Kristen Lee (UW)
Priska von Haller (UWPR)
Daniela Tomazela (UW)
Michael Bereman (UW)
Eric Hommema (Thermo)
John Rogers (Thermo)

Chromatin Biology

John Stamatoyannopoulos (UW)
Hao Wang (UW)
Matt Maurano (UW)

Funding

University of Washington's Proteomics Resource (UWPR)
Thermo Scientific Pierce Human In Vitro Translation Research Grant
NIDDK F30 fellowship

Posters to see!

WP407 – Panorama: A repository of targeted proteomics assays for Skyline

WP256 – Functional assortment of human transcription factors into defined chromatin niches