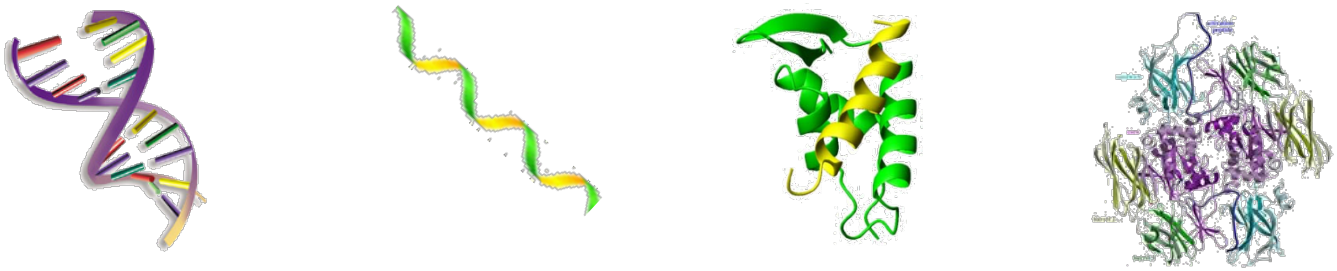
Proteogenomic Challenge Data Overview

Flow of Biological Information



DNA → RNA → Protein → Functional Protein

DNA Structure
Chromatin Organization

Transcriptional Control

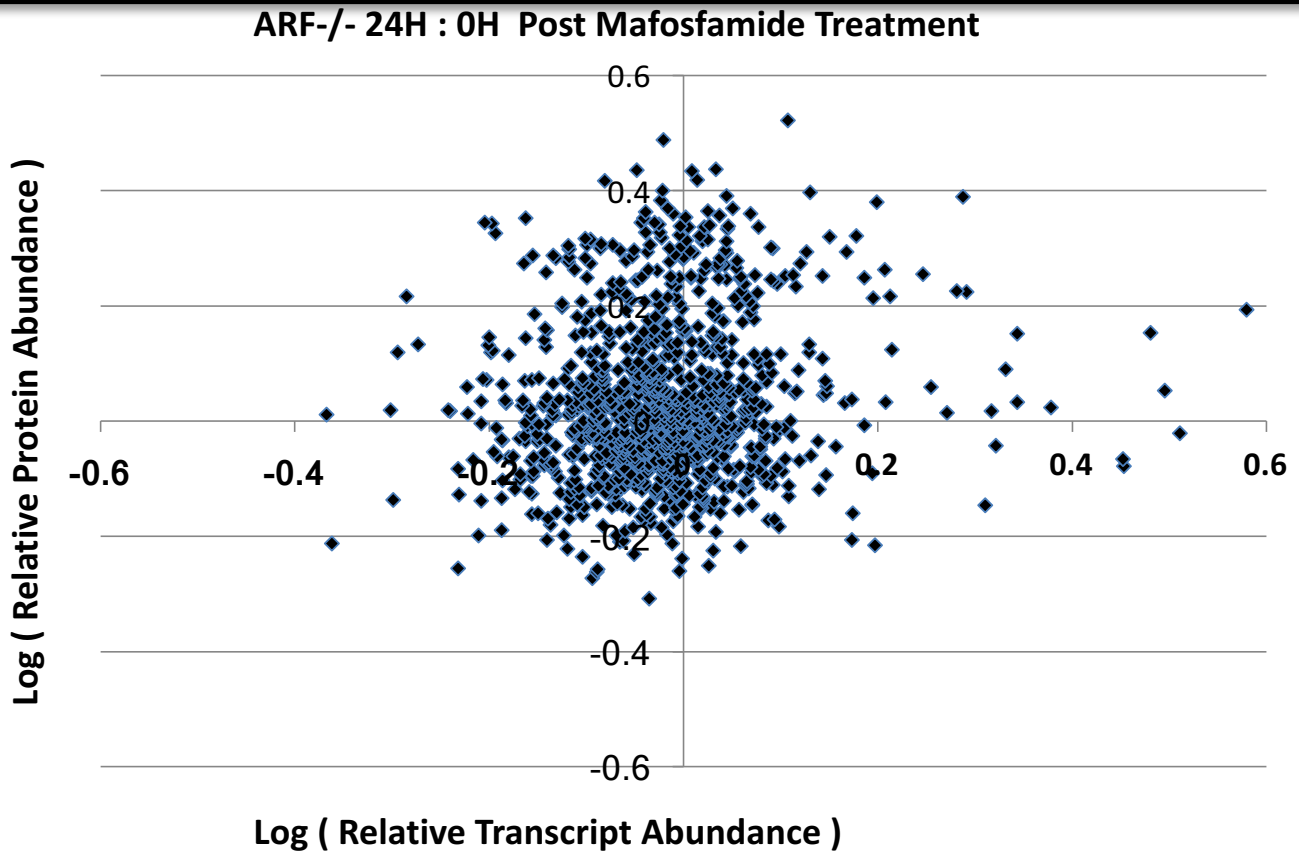
mRNA secondary structure
mRNA stability
uRNA

Translational Control

Post translational Control

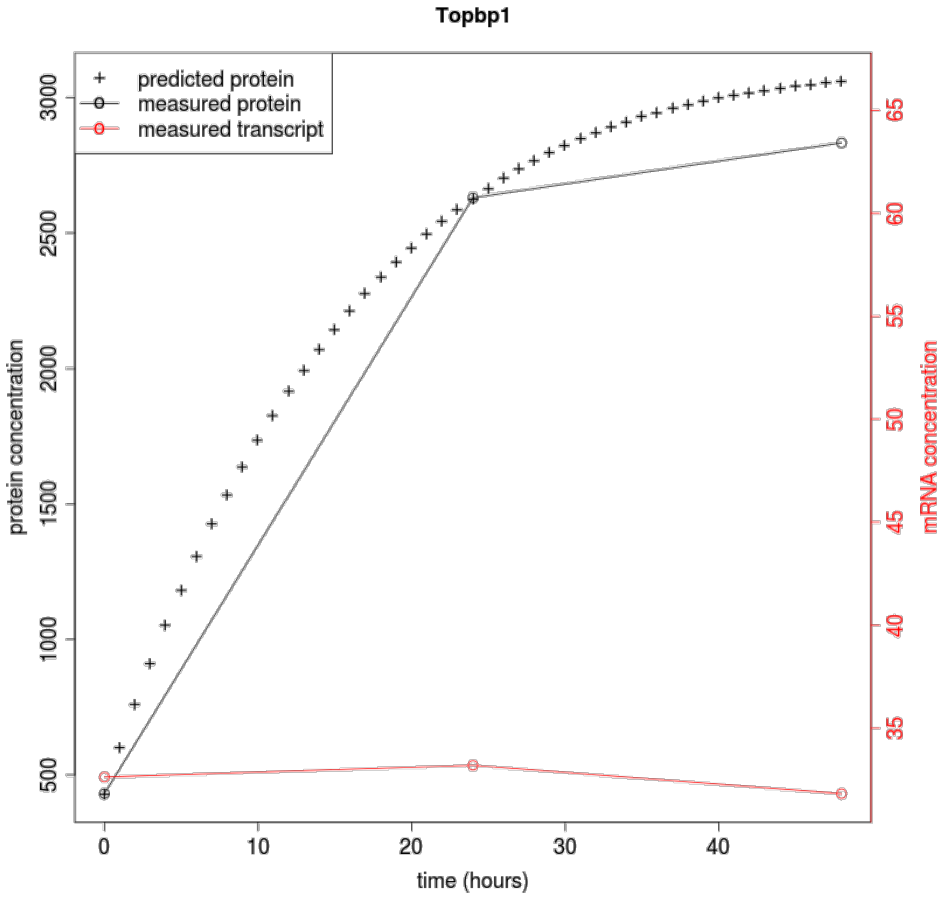
protein processing
protein modification
protein stability/turnover
protein complex formation

Comparing Changes in Protein and Transcript Abundances



No obvious correlation between transcript and protein changes in abundance

Predicting Protein Levels from Transcript Levels



Question:

- Given RNA Transcript Data – Predict Protein Abundance.

Table 2	Spectral count data for 44 cell line samples, 95 tumor samples and 60 normal tissue samples.
Table 3	RNASeq data for 44 cell line samples, 87 tumor samples and 48 normal tissue samples.

Data Origins

- https://www.nature.com/articles/nature13438

Proteogenomic characterization of human colon and rectal cancer

Bing Zhang, Jing Wang, Xiaojing Wang, Jing Zhu, Qi Liu, Zhiao Shi, Matthew C. Chambers, Lisa J. Zimmerman, Kent F. Shaddox, Sangtae Kim, Sherri R. Davies, Sean Wang, Pei Wang, Christopher R. Kinsinger, Robert C. Rivers, Henry Rodriguez, R. Reid Townsend, Matthew J. C. Ellis, Steven A. Carr, David L. Tabb, Robert J. Coffey, Robbert J. C. Slebos, Daniel C. Liebler ✉ & the NCI CPTAC

•*Nature* **volume 513**, pages 382–387 (18 September 2014)

https://www.synapse.org/#!Synapse:syn8228304

RNASEQ Data

Gene Symbol	C125PM	C135	C70	CACO2	COLO201	COLO205	COLO320	DIFI	DLD1	GEO	GP5D	HCA7
A1BG	0	0	0	0	0	0	0	0	0	0	0	0
A1CF	0	1	0	4	1	2	0	3	0	0	0	0
A2M	0	0	2	10	1	0	0	1	0	0	1	3
A2ML1	0	0	0	0	0	1	0	0	0	0	0	0
AAAS	9	9	10	11	4	7	14	12	9	9	6	9
AACS	4	4	10	2	11	28	2	3	1	6	3	0
AADAT	0	0	0	0	0	0	0	0	0	0	0	2
AAED1	0	0	0	0	0	0	0	0	0	0	0	0
AAGAB	0	0	3	0	0	1	0	0	2	0	1	0
AAK1	0	12	0	1	1	0	0	0	10	0	0	1
AAMDC	0	0	0	3	0	2	7	1	0	0	0	0
AAMP	0	2	4	3	6	6	3	4	5	0	2	0
AAR2	0	1	0	1	5	5	2	2	0	1	0	2
AARD	1	3	3	3	4	3	1	4	2	1	7	2
AARS	18	43	51	59	43	69	86	31	25	59	30	4
AARS2	5	6	1	5	1	1	6	6	7	10	4	6

Proteomics Data

Gene Symbol	C125PM	C135	C70	CACO2	COLO201	COLO205	COLO320	DiFi	DLD1	GEO
61E3.4	15959	4161	4398	4766	4209	4807	14734	15289	5191	5268
A1BG	4	0	5	0	0	0	4	0	1	0
A1CF	54	3	234	1194	241	532	0	430	0	142
A2M	0	135	0	10	5	0	0	0	12	0
A2ML1	0	1	1	0	0	1	0	0	1	0
A4GALT	102	0	1	470	409	55	0	13	0	1
A4GNT	0	0	0	2	0	0	0	0	0	0
A-575C2.4	212	26	42	41	30	30	288	110	68	55
AAAS	987	1074	2098	1702	1693	1738	3977	2026	1477	1001
AACS	2387	890	5519	1210	3382	10859	1960	4949	3343	2120
AADAC	50	0	5	7	0	1	0	10	0	0
AADACL2	1	0	0	0	0	0	1	0	0	0
AADACL3	0	0	0	0	6	0	0	2	13	0
AADACL4	0	0	0	0	0	0	0	0	0	0