Statistical Modeling to Quantitate the Central Dogma

Jingyi Jessica Li

Department of Statistics

University of California, Los Angeles

May 21, 2018

UCLA College

Statistics

What is Statistics?

16th-17th Century: Probability Theory (Mathematical Foundations of Statistics)

Jerome Cardan
1501-1576

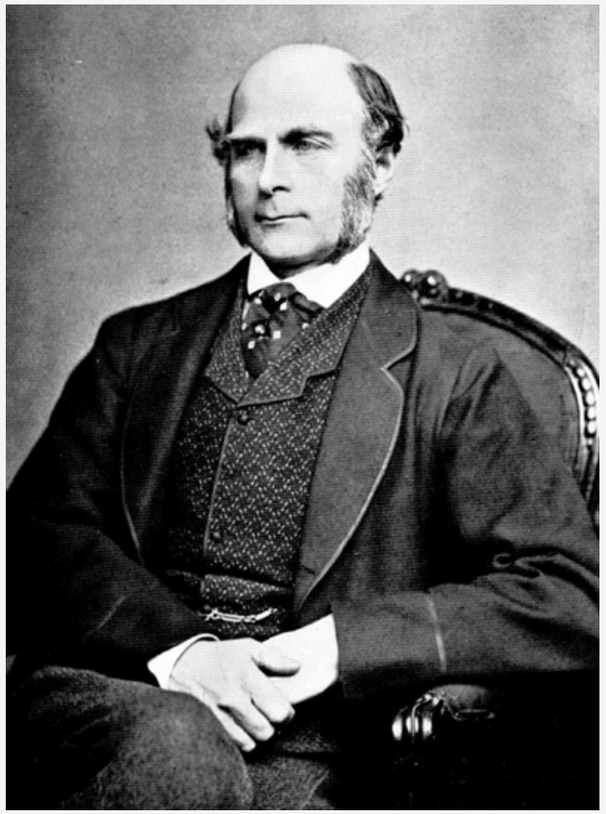
Blaise Pascal
1623-1662

Pierre de Fermat
1607-1665

From *Wikipedia*

What is Statistics?

19th-20th Century: Emergence and Development of Statistics



Francis Galton
1822-1911

Karl Pearson
1857-1936

Ronald Fisher
1890-1962

From *Wikipedia*

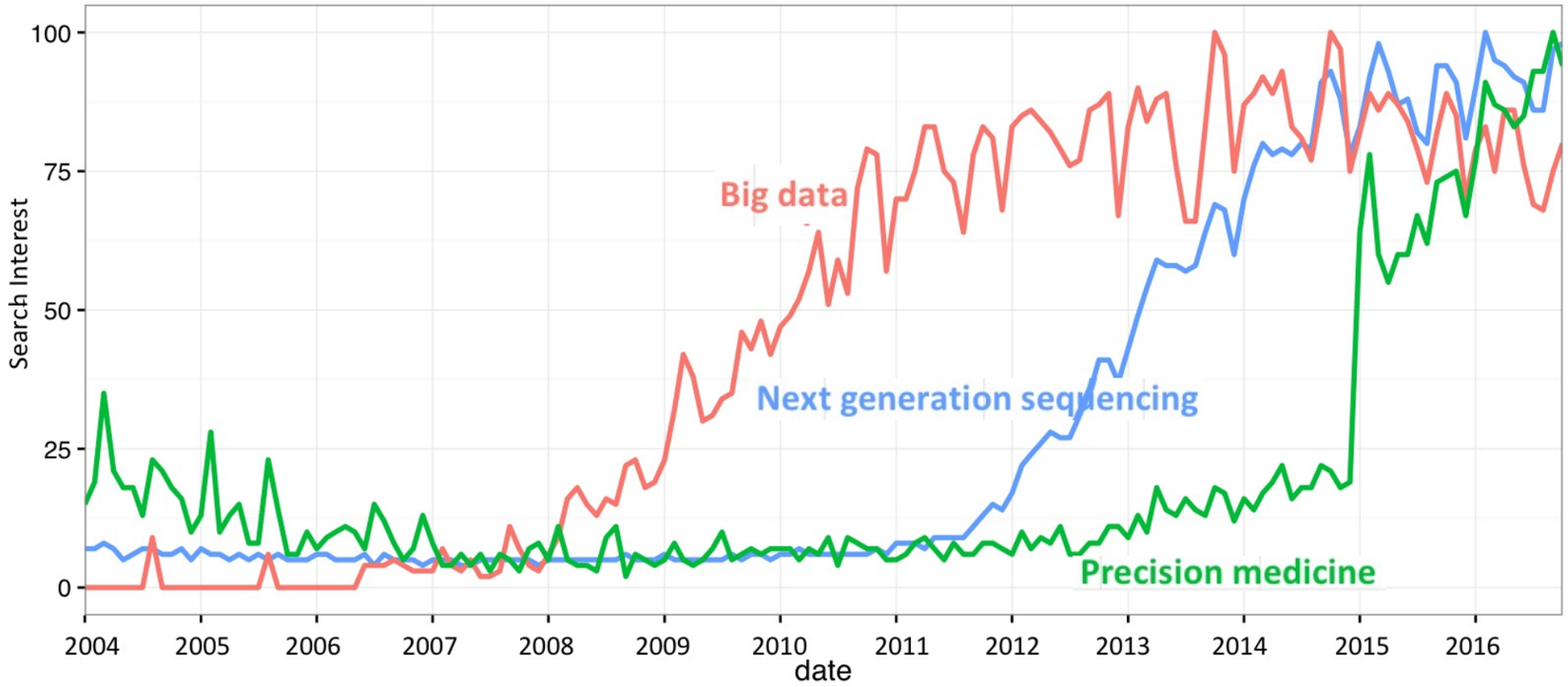
What is Statistics?

21st Century: Computer Age of Statistics

From *Google Images*

The Junction of Statistics and Biology

My Research Overview



Change in the popularity of search terms from January, 2004 to October, 2016 based on Google Search. Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. Likewise a score of 0 means the term was less than 1% as popular as the peak. Created using Google Trends: https://www.google.com/trends/.

The Junction of Statistics and Biology

Our Team

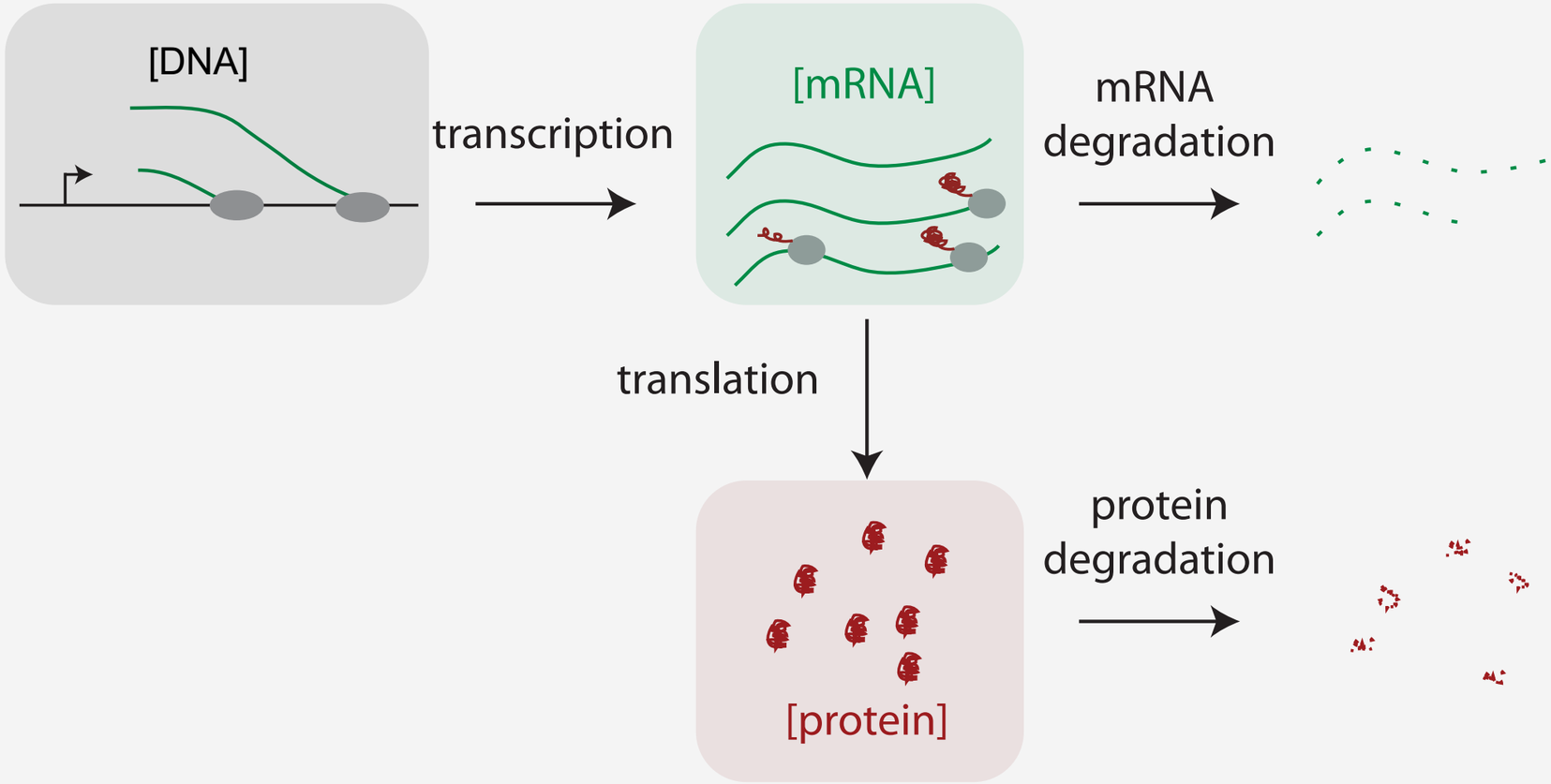
From *http://jsb.ucla.edu*

“The best thing about being a statistician is that you get to play in everyone's backyard.”

– Dr. John W. Tukey

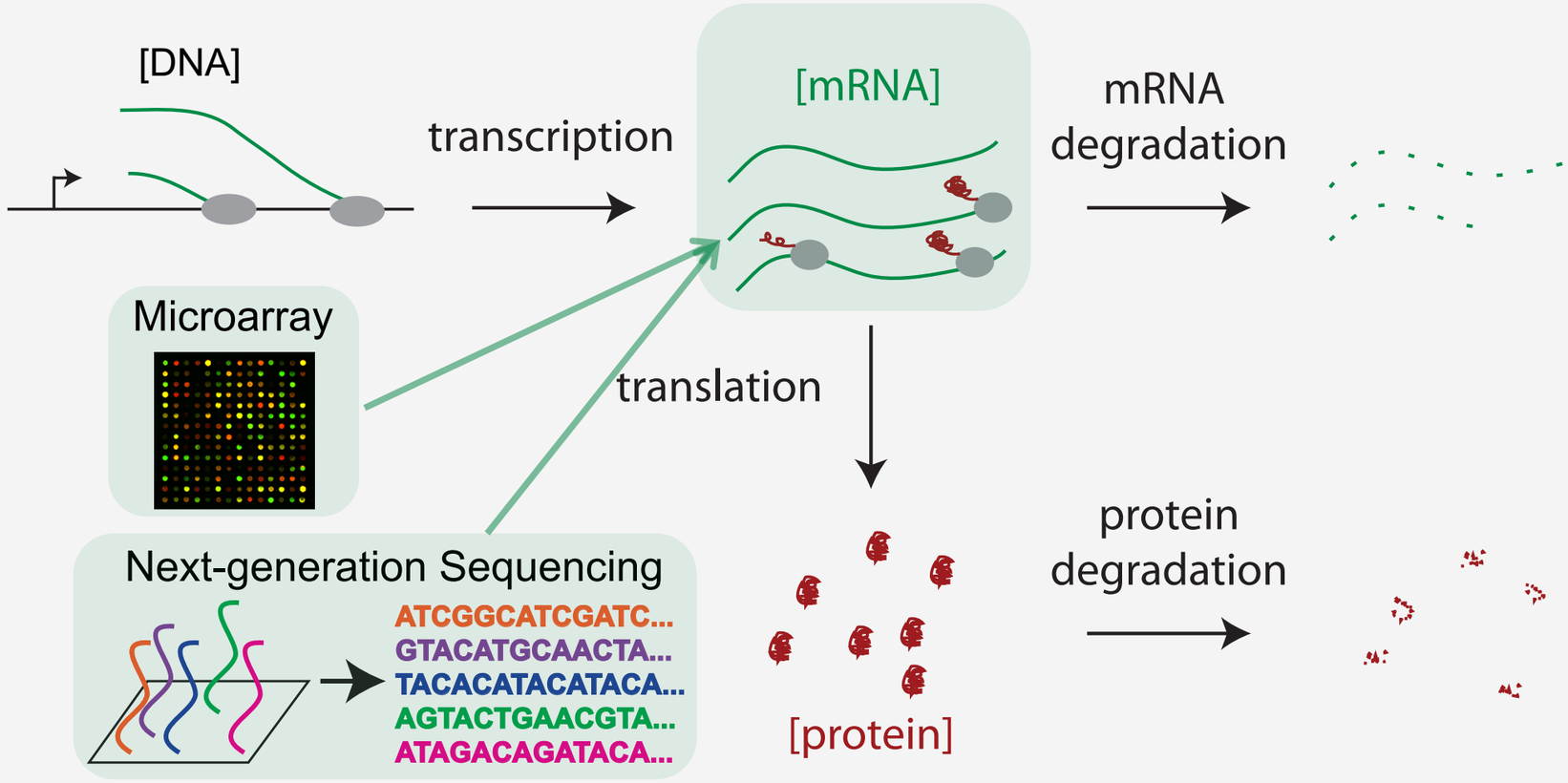
The Central Dogma

One of the Most Fundamental Principles of Modern Biological Sciences



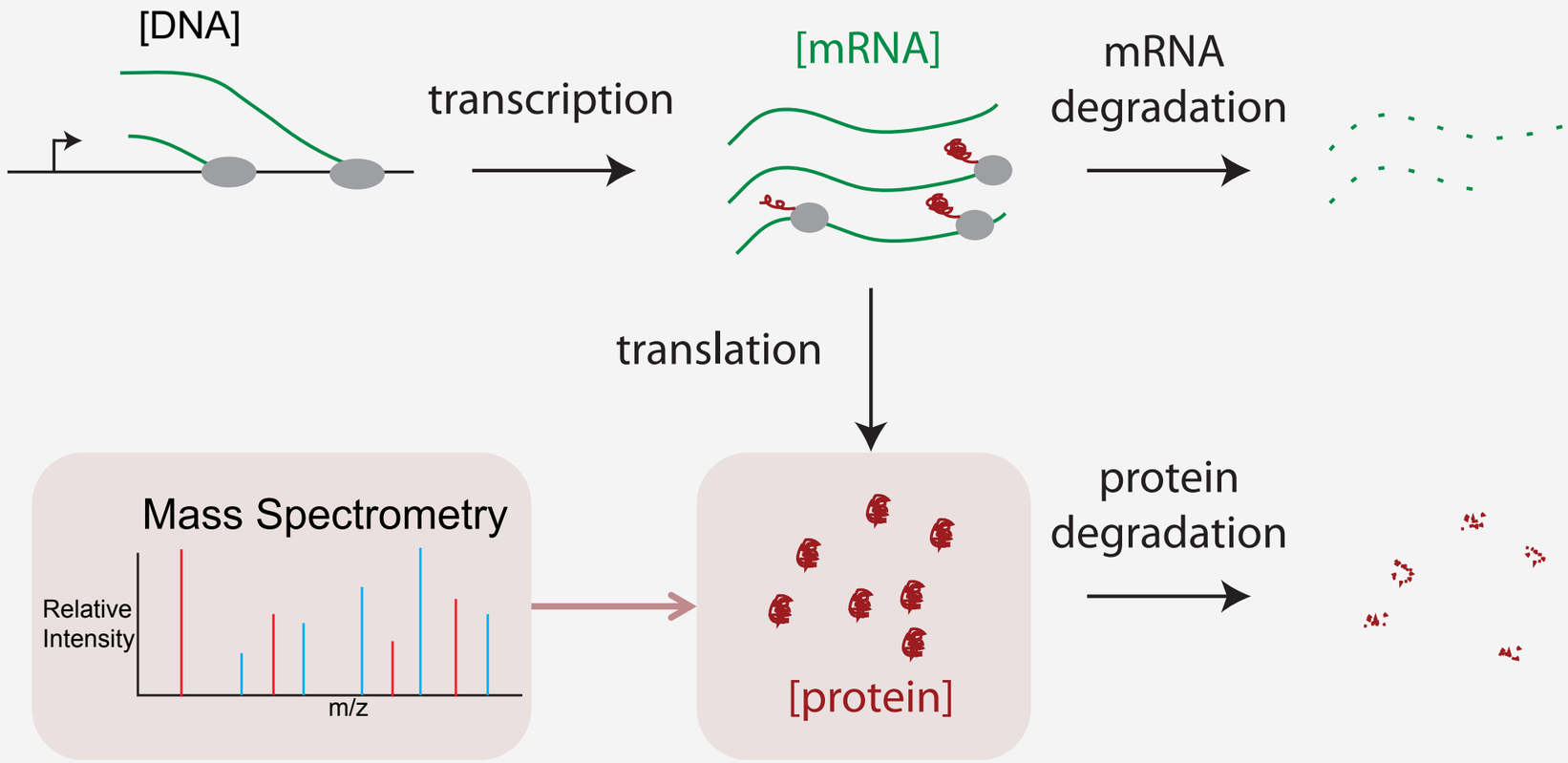
Quantifying the Central Dogma

High-throughput Experiments



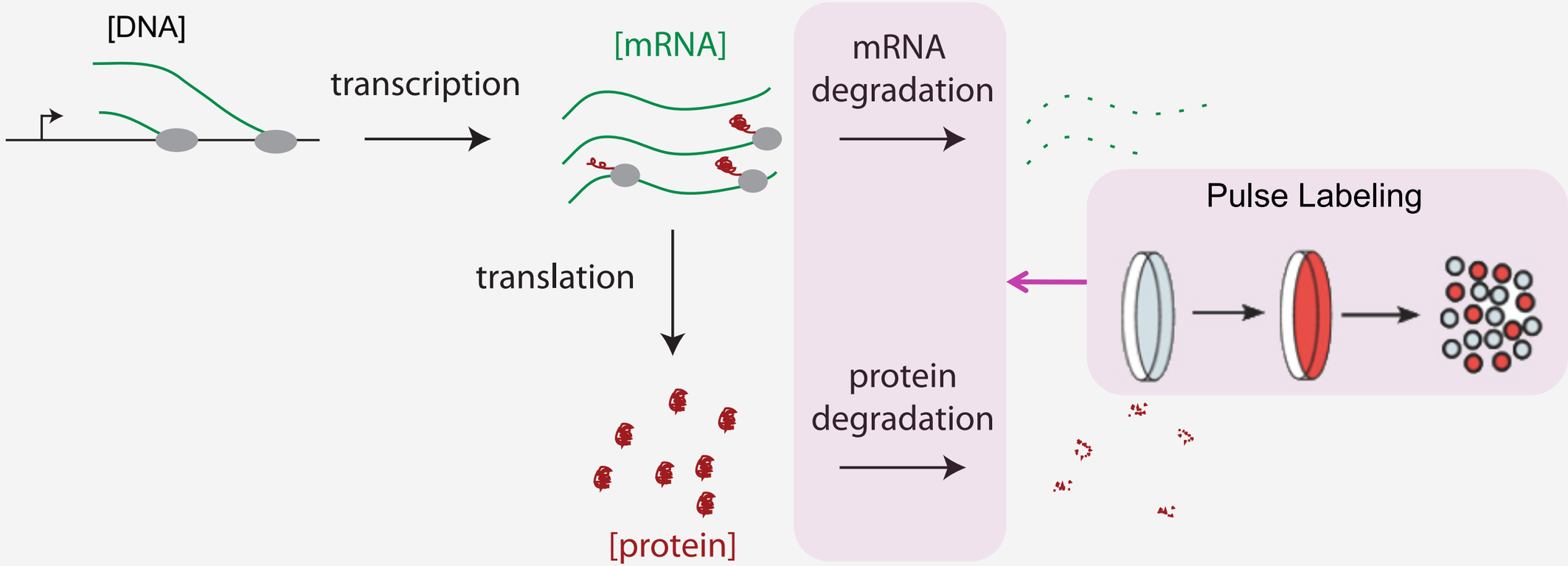
Quantifying the Central Dogma

High-throughput Experiments



Quantifying the Central Dogma

High-throughput Experiments



Important Steps that Determine Protein Levels?

Claim 1: RNA levels do not well predict protein levels

nature
International journal of science

Altmetric: 118 Citations: 1942 More detail »

Article

Global quantification of mammalian gene expression control

Björn Schwanhäusser, Dorothea Busse, Na Li, Gunnar Dittmar, Johannes Schuchhardt, Jana Wolf, Wei Chen & Matthias Selbach

Nature **473**, 337–342 (19 May 2011)
doi:10.1038/nature10098
Download Citation

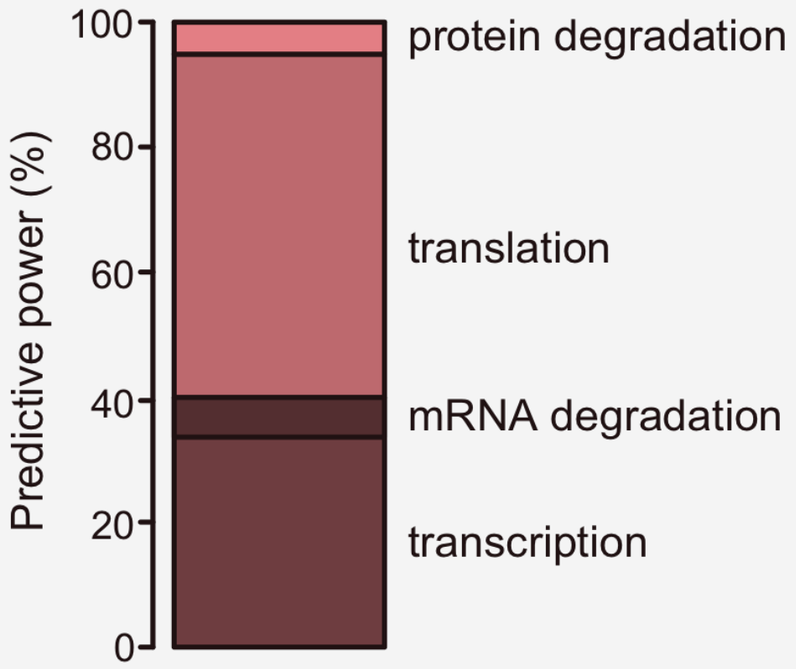
Epigenetics Gene expression
Gene regulation

Received: 16 November 2010
Accepted: 01 April 2011
Published: 18 May 2011

Corrigendum: 13 February 2013

Schwanhausser et al. (2011). *Nature.* 473: 337–342.

Schwanhausser *et al.* suggest that translation is the most important



Caution against Misuse of Statistics

Most Published Research Findings Are False?

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • https://doi.org/10.1371/journal.pmed.0020124

Mark Biggin, LBNL

Important Steps that Determine Protein Levels?

Claim 2: RNA levels predict protein levels well

Science

PERSPECTIVE | GENE EXPRESSION

Statistics requantitates the central dogma

Jingyi Jessica Li[1], Mark D. Biggin[2]

+ See all authors and affiliations

Science 06 Mar 2015:
Vol. 347, Issue 6226, pp. 1066-1067
DOI: 10.1126/science.aaa8332

Altmetric 189

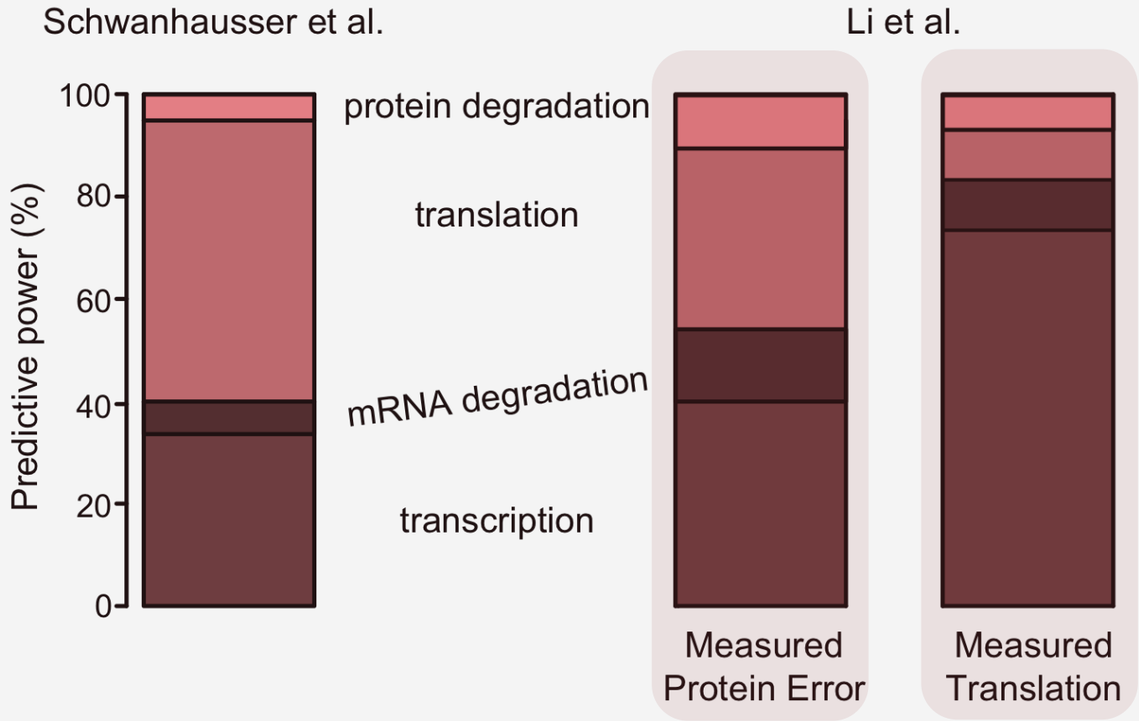
Summary

Mammalian proteins are expressed at ~10^3 to 10^8 molecules per cell (*1*). Differences between cell types, between normal and disease states, and between individuals are largely defined by changes in the abundance of proteins, which are in turn determined by rates of transcription, messenger RNA (mRNA) degradation, translation, and protein degradation. If the rates for one of these steps differ much more than the rates of the other three, that step would be dominant in defining the variation in protein expression. Over the past decade, system-wide studies have claimed that in animals, differences in translation rates predominate (*2*–*5*). On page 1112 of this issue, Jovanovic *et al.* (*6*), as well as recent studies by Battle *et al.* (*7*) and Li *et al.* (*1*), challenge this conclusion, suggesting that transcriptional control makes the larger contribution.

Li, Bickel and Biggin (2014). *PeerJ*. 2:e270.

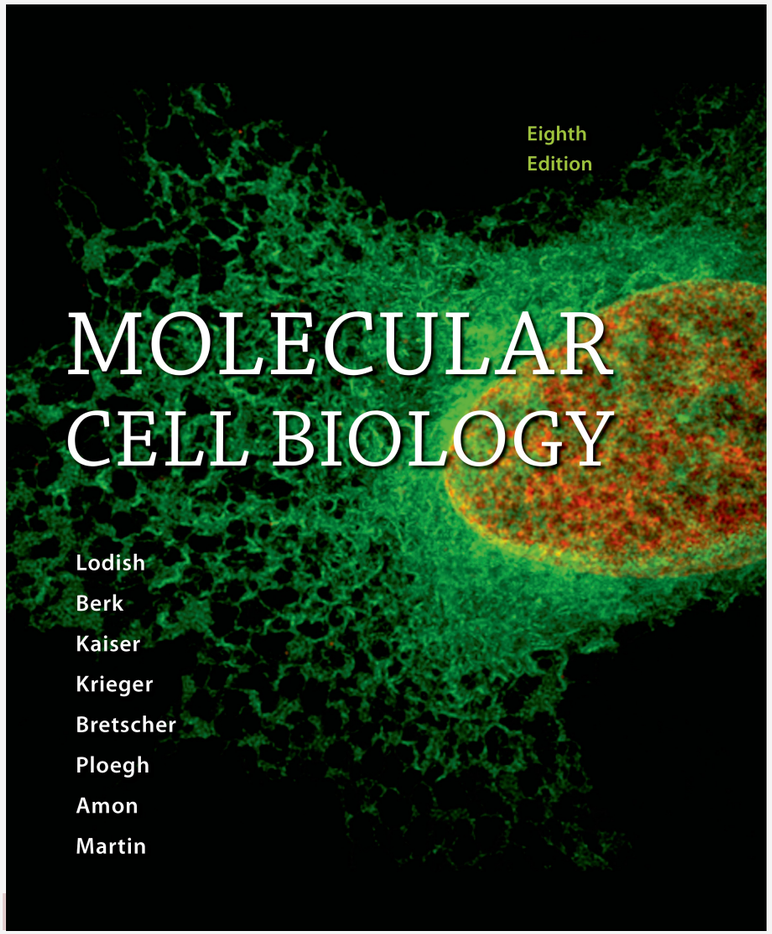
Li and Biggin. (2015). *Science*. 347(6226): 1066–1067.

Li *et al.* suggest that transcription is the still the most important



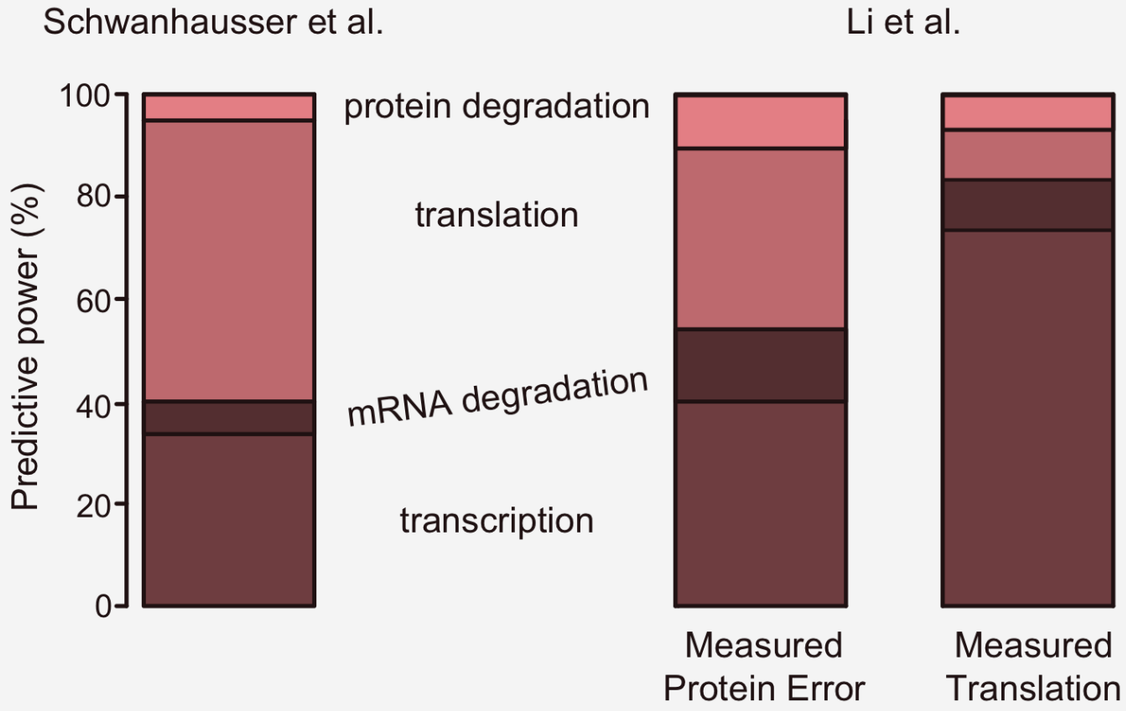
Important Steps that Determine Protein Levels?

Claim 2: RNA levels predict protein levels well



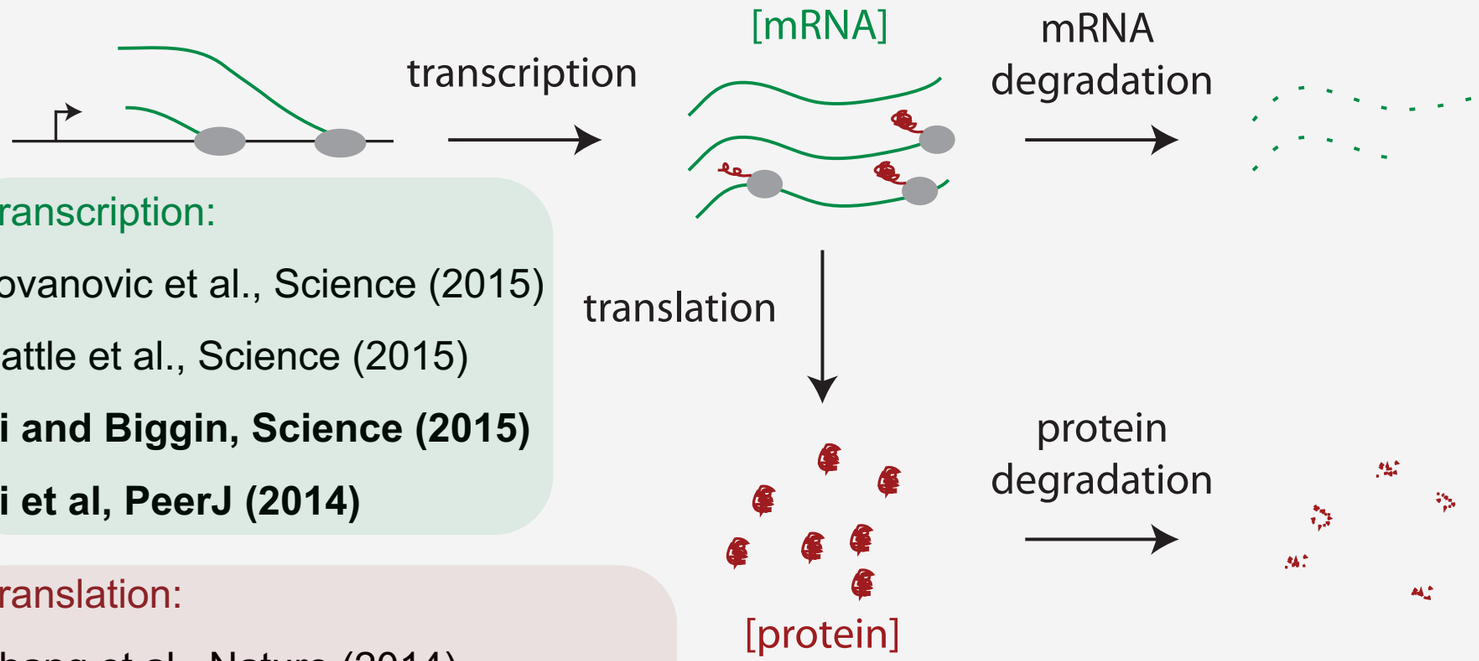
Lodish et al. (2016). Molecular Cell Biology (Eighth Edition).

Li *et al.* suggest that transcription is the still the most important



The Debate: Translation vs. Transcription

Which is more important?



Transcription:

Jovanovic et al., Science (2015)

Battle et al., Science (2015)

Li and Biggin, Science (2015)

Li et al, PeerJ (2014)

Translation:

Zhang et al., Nature (2014)

Kristensen et al, Mol. Syst. Biol. (2013)

Ghazalpour et al., PLOS Genet. (2011)

Schwanhäusser et al., Nature (2011)

SPECIES
S. cerevisae (yeast)
S. pombe (yeast)
A. thaliana (plant)
D. melanogaster (fruit fly)
M. musculus (mouse)
H. sapiens (human)

Li and Biggin. (2015). *Science.* 347(6226): 1066–1067.

The Junction of Statistics and Biology

Educational Goals

THEORETICAL STATISTICS	APPLIED STATISTICS
Decision Theory	Data Processing
Large Sample Theory	Data Visualization
High Dimensional Theory	Data Analysis
Optimization	Computer Programming
Information Theory	Data Interpretation
…	…

Caution against Misuse of Statistics

UC Berkeley Graduate Admission Gender Bias (1973)

Overall

	APPLICANTS	ADMITTED
MEN	8442	44%
WOMEN	4321	35%

By Department

DEPARTMENT	MEN		WOMEN	
	APPLICANTS	ADMITTED	APPLICANTS	ADMITTED
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	373	6%	341	7%

Simpson’s Paradox

P.J. Bickel, E.A. Hammel and J.W. O'Connell (1975). "Sex Bias in Graduate Admissions: Data From Berkeley". *Science*. 187 (4175): 398–404.

Data from *Wikipedia*

“The best thing about being a statistician is that you get to play in everyone's backyard.” – Dr. John W. Tukey

Thank you.

Jingyi Jessica Li
Department of Statistics
University of California, Los Angeles
jli@stat.ucla.edu
http://jsb.ucla.edu

UCLA College
Statistics