

Big Data and Analytics of Genomics Data

Douglas Ruden, Ph.D.

Director of Epigenomics

Department of Obstetrics and Gynecology

Wayne State University

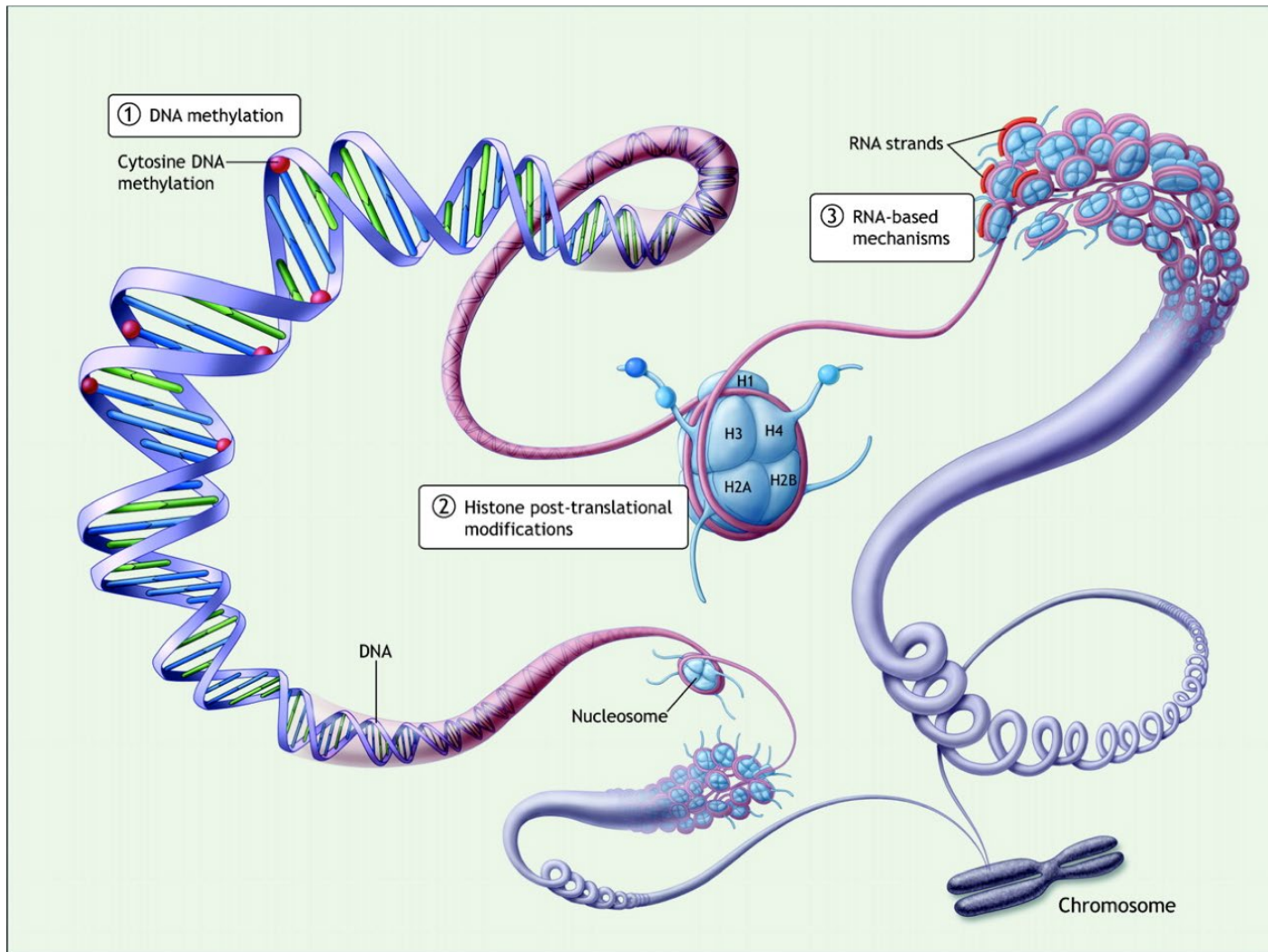
Detroit, MI



**STRENGTHENING THE BIG DATA
& ANALYTICS ECOSYSTEM**
WAYNE STATE UNIVERSITY

September 19, 2018

Big Data In Genomics



In Humans:

Each Cell has ~4 Billion Bases

Each Cell has ~20 Million mC's

Each Tissue has up to ~1 Trillion Cells

Each Human has over 500 cell types

Each Cell has over 10 Million type of proteins and small molecules



DNA Methylation and Histones Control Gene Expression

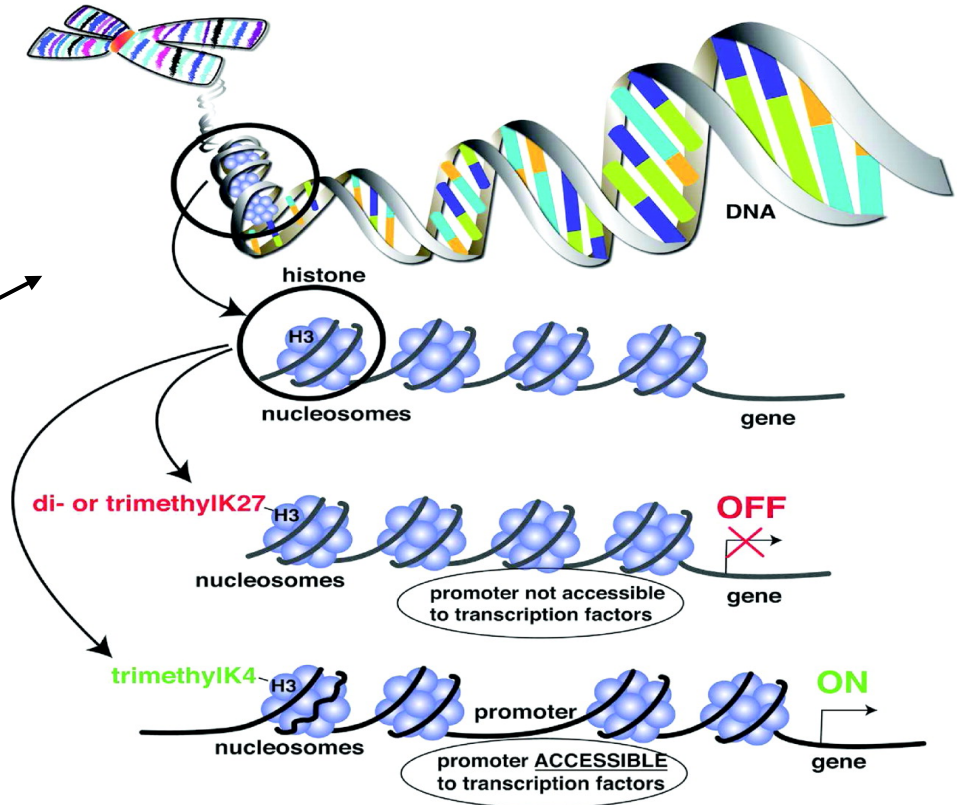


Each Cell has 4
Billion Bases

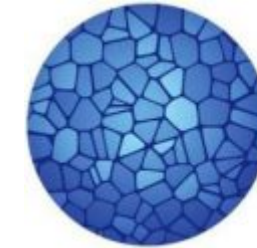
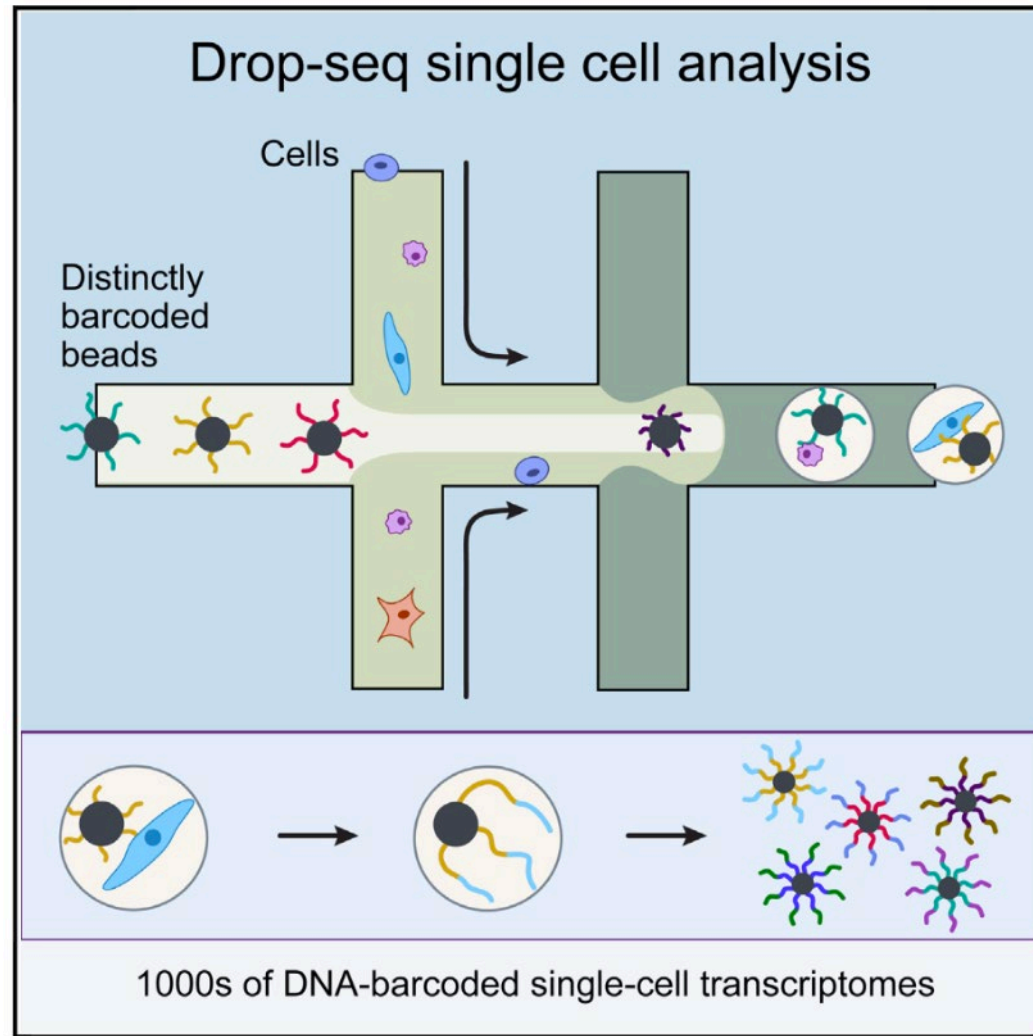
Each Cell has ~20
Million mC's

A Histone is
Present every 250
bases

Each Histone can
have over 100
modifications

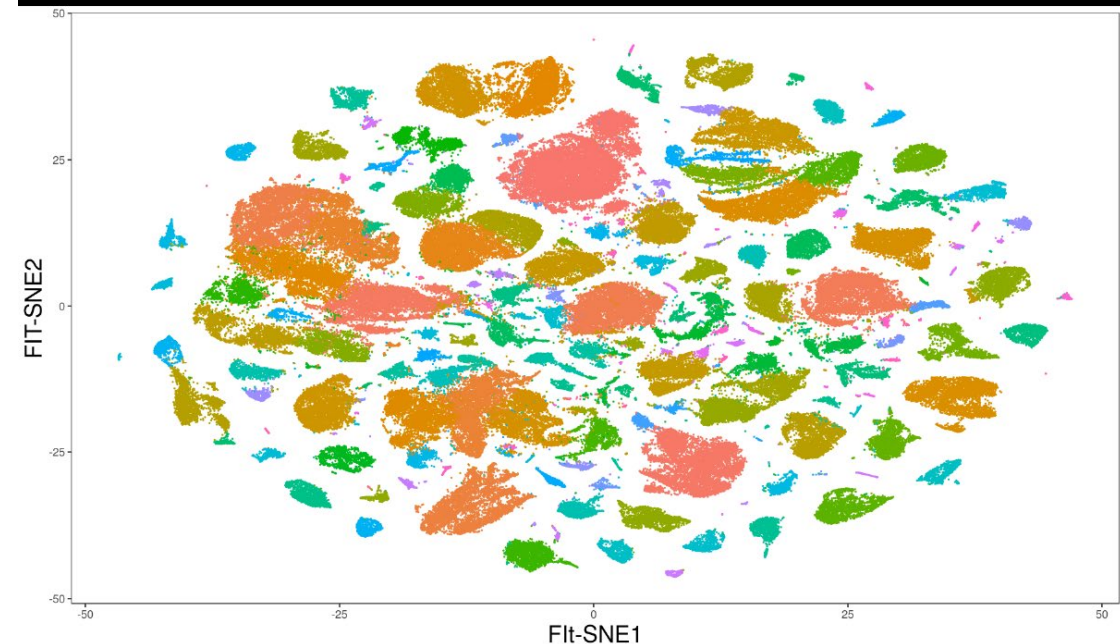


Single Cell Atlases of Every Organ and Tissue



HUMAN
CELL
ATLAS

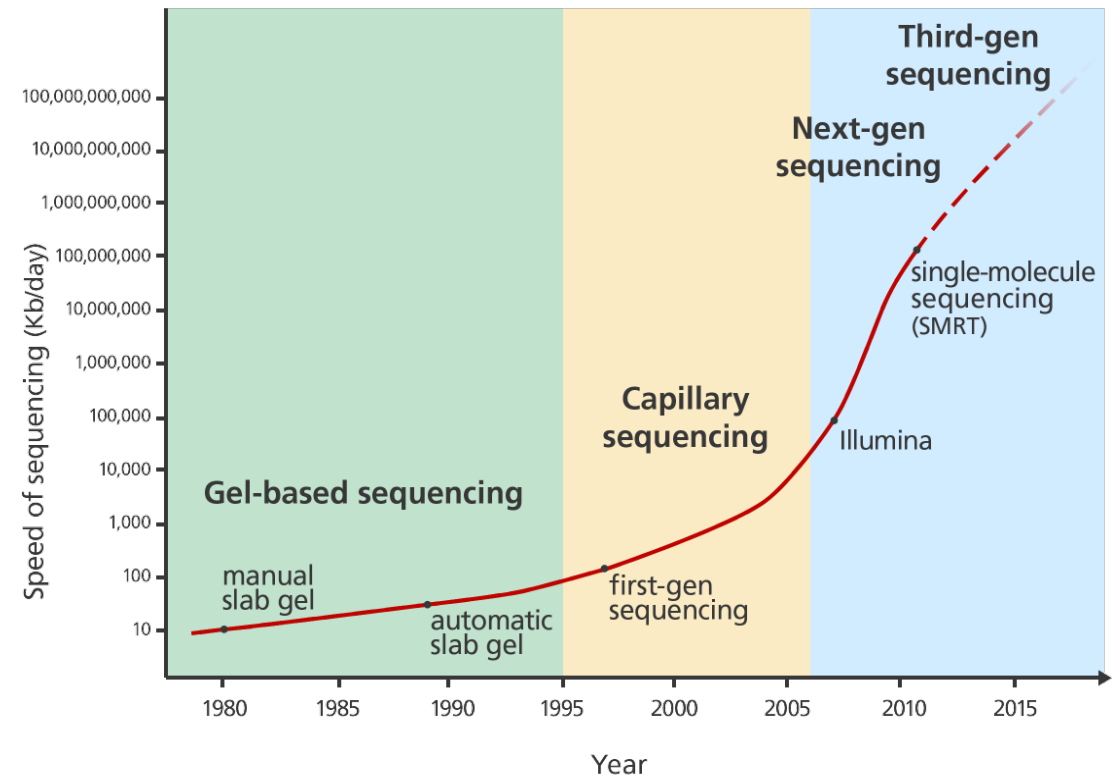
Vignette --- Mouse Cell Atlas, 250K cells



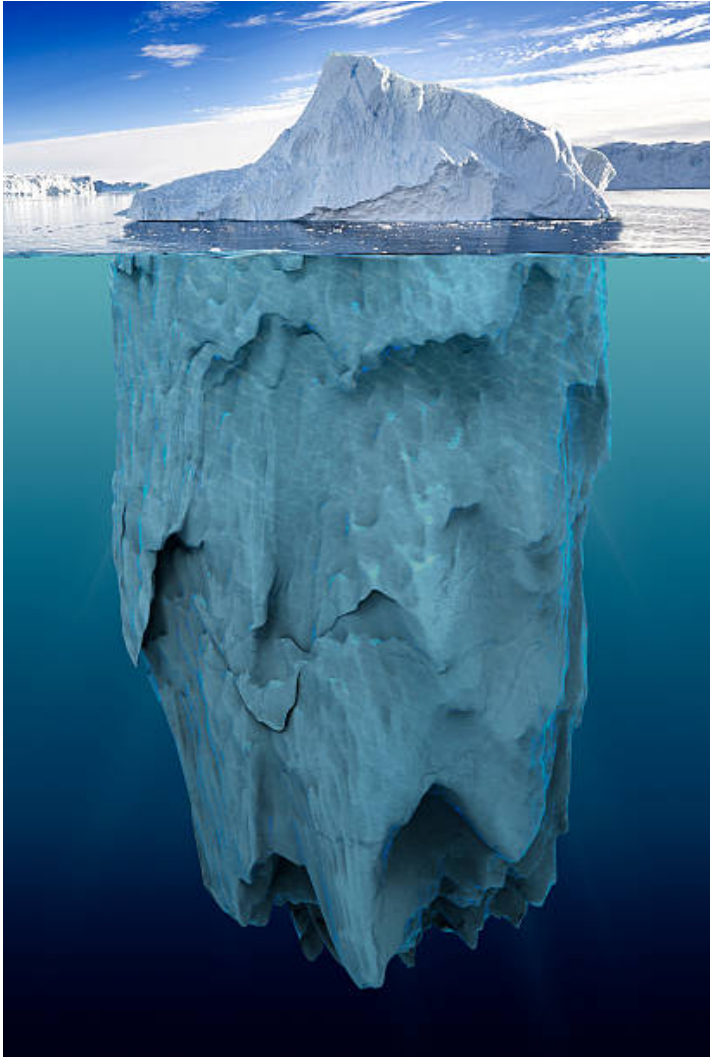
NovaSeq – DNA Sequencing Capabilities

WSU will purchase a NovaSeq600 in Fall 2018
This instrument can sequencing 4 trillion bases per day (100 people at 100x coverage!). 5-10 petabytes of data per year!

Moore's Law of DNA Sequencing –
10-fold increase in the speed of sequencing every 5 years since 1980.



Big Data in Biology – The Future



4 billion base genome (EVERYONE!)

100 trillion cells per person

- Single Cell Atlases
- Single Cell Transcriptome
- Single Cell Epigenome

10 million proteome and small molecule metabolome per cell

