



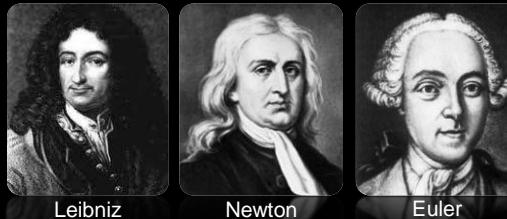
National Cancer Informatics Program

**MAKING SENSE OF DATA VIA MODELS**

# Golden Age of Natural Sciences

## ■ 1600-1700

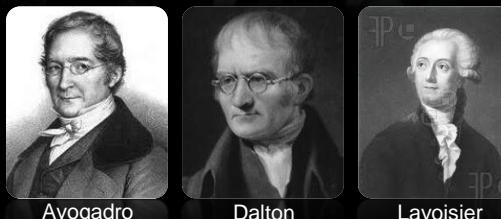
- Mathematics



Leibniz      Newton      Euler

## ■ 1800-1900

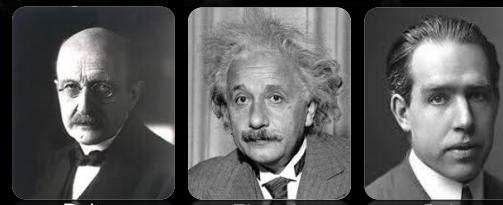
- Chemistry



Avogadro      Dalton      Lavoisier

## ■ 1850-1940

- Physics



Dalton      Einstein      Bohr

## ■ 1940-2000

- Information Science



Von Neumann      Minsky      Jobs

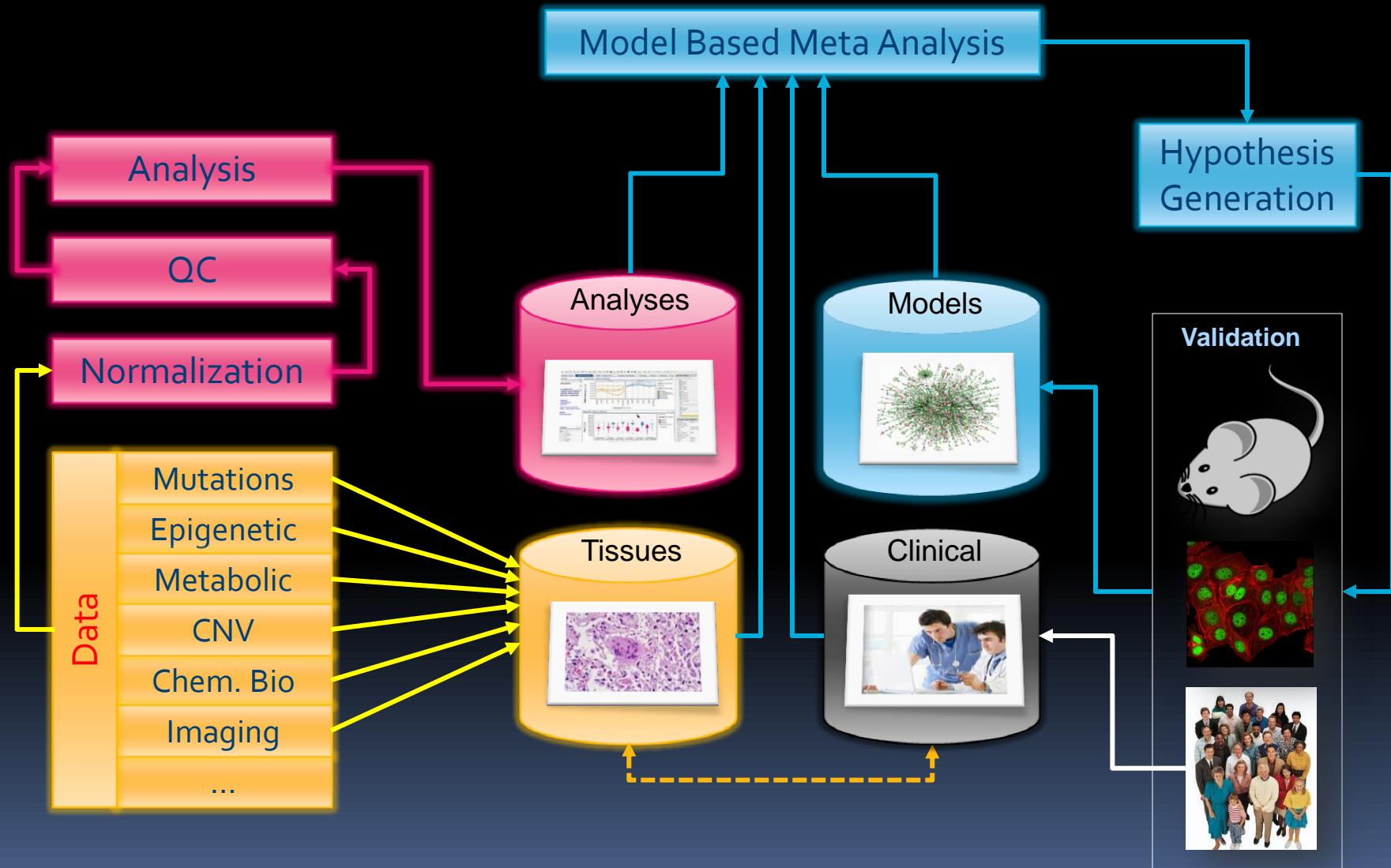
## ■ 1970-2100

- Biology



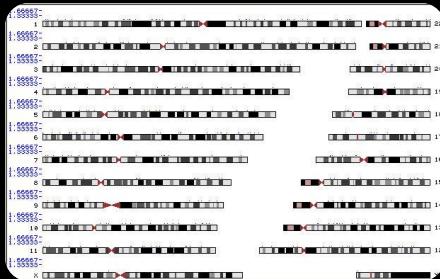
Watson      Brenner      Varmus

# Data and Model Driven Discovery

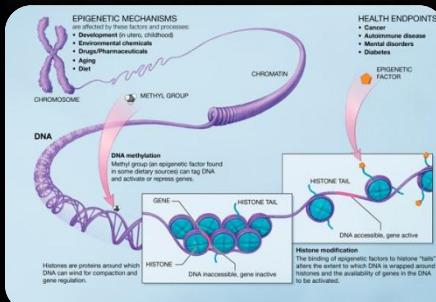


# Model-based Meta Analysis: Integrative genomics

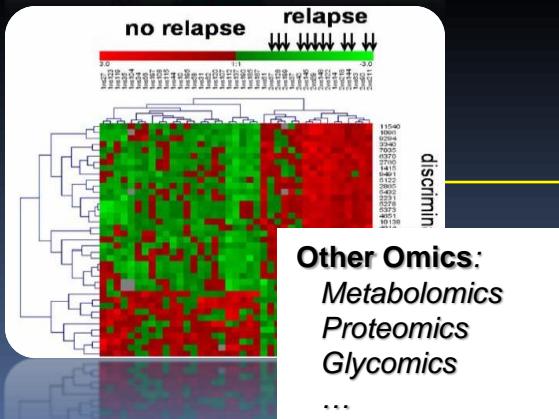
## Genomics



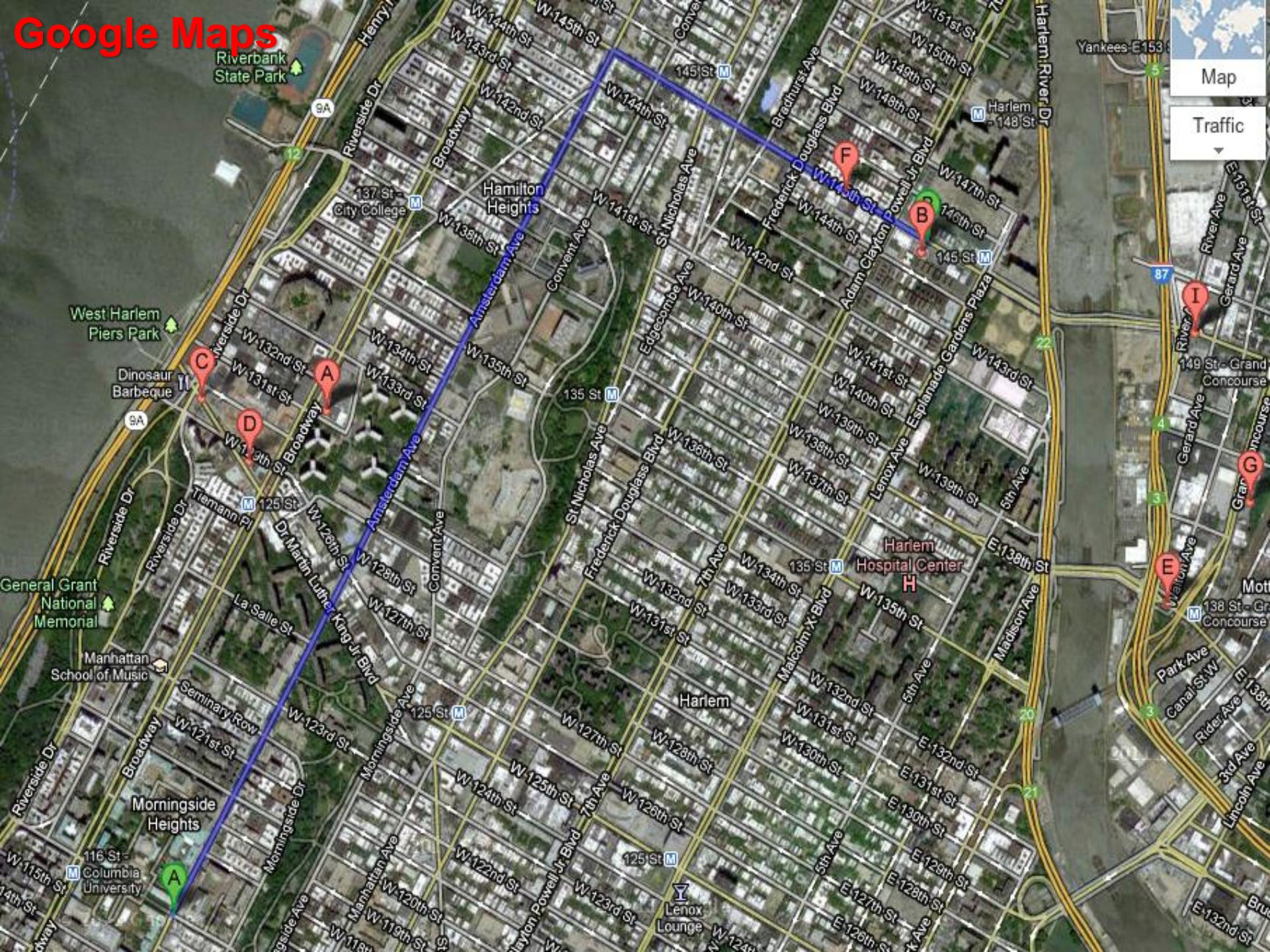
## Epigenomics



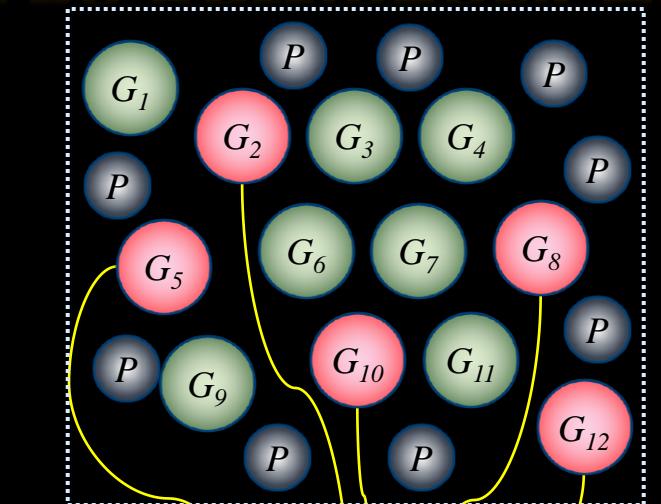
## Transcriptomics



# Google Maps

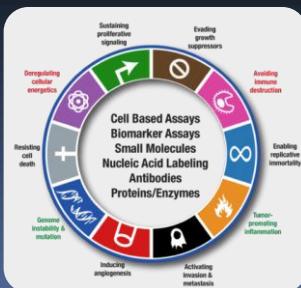
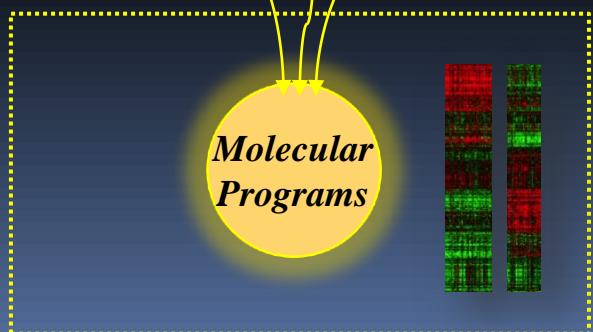


# The Bottleneck Hypothesis



Patient X

Genetic and  
Epigenetic  
Alterations



Glioblastoma:

- Carro MS et al. *Nature*. 2010 Jan 21;463(7279):318-25.

- Bottleneck: CEBP $\beta/\delta$  and Stat3

Diffuse Large B Cell Lymphoma:

- Compagno M et al. *Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma*. *Nature*. 2009 Jun 4;459(7247):717-21
- Bottleneck: Nf- $\kappa$ B

GC-Resistance in T-ALL:

- Real PJ et al. *Gamma-secretase inhibitors reverse glucocorticoid resistance in T cell acute lymphoblastic leukemia*. *Nat Med*. 2009 Jan;15(1):50-8.
- Bottleneck: NOTCH1 pathway

T-ALL Tumorigenesis:

- Dalla Gatta G et al. *Nature Medicine* 2012 Feb 26;18(3):436-40
- Bottleneck: TLX1, TLX3, RUNX1

# Six Principles for the golden age of cancer biology

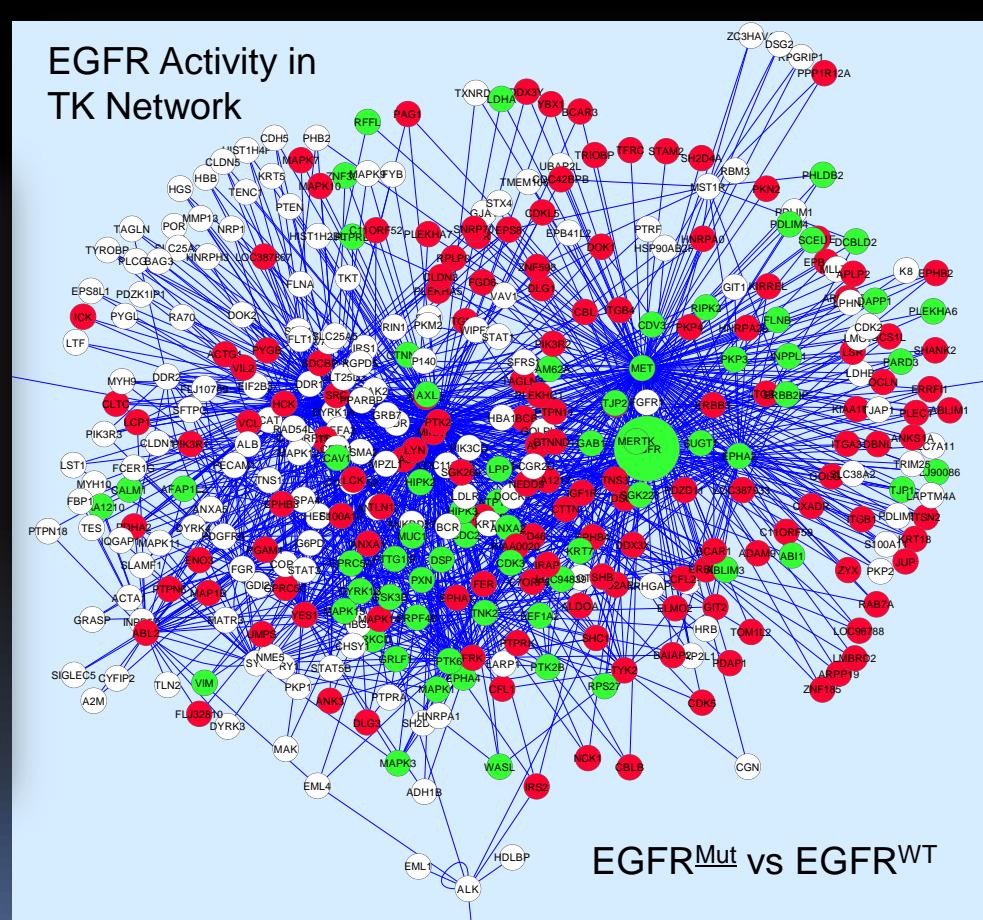
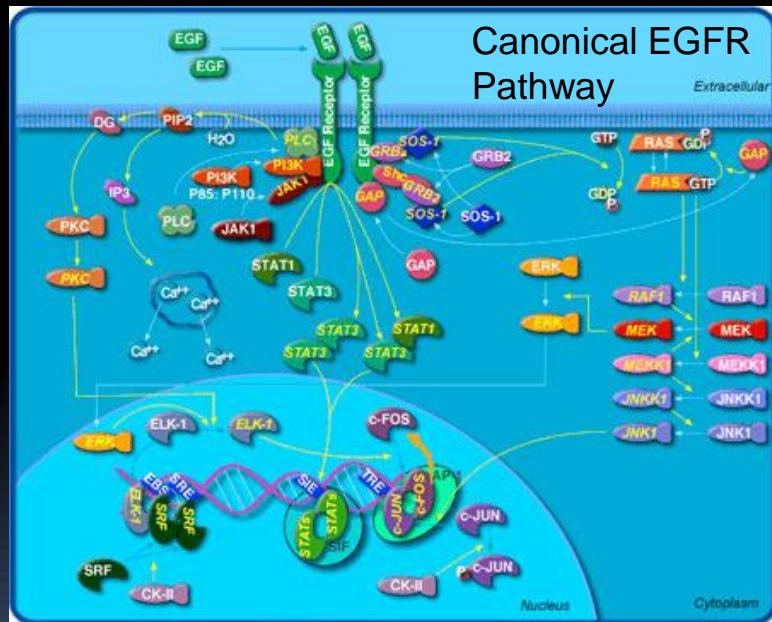
- 1. Change the culture:
  - Computational biology should be driven by key biological questions
  - Experimental biology should be data and model driven
  - Gradually achieve greater balance for funding, personnel, and resources
- 2. Create the necessary datasets
  - Computational sciences are intrinsically data rich and data dependent
  - The TCGA model is a critical stepping stone for computational modeling approaches
- 3. Support lightweight mechanisms for integrating diverse data modalities
  - Google and NSA process more unstructured data every day than the biological community has ever produced
  - Regulatory models and integrative frameworks are the Google Maps of cancer genomics

# **Six Principles for the golden age of cancer biology**

- **4. Bring the communities together:**
  - Embrace the Driving Biological Project (DBP) model from the NCBCs
  - Foster a community of integrated biology labs (humid model)
  - Create a portfolio of centers of excellence in Cancer Informatics
- **5. Support the Integration of molecular and clinical data:**
  - We must build appropriate frameworks to support integrative models
  - Framework must be driven by key translational questions (e.g., I-SPY)
- **6. Help build the infrastructure:**
  - Hybrid educational programs
  - Computational infrastructure (cycles on the cloud, petabyte storage)
  - Community dissemination
  - Crowd-sourcing
  - Support the software lifecycle

# A charter project

- We have enumerated the components of the cell machinery
- Our understanding of the actual machinery is spotty at best



# The Encode+ Project for Cancer

- Elucidate all layers of regulatory interactions in a specific cellular context (e.g. Non-small cell lung cancer)
  - Transcriptional Interactions (TF targets + combinatorial regulation)
  - miR targets, miR activity modulation, miR biogenesis
  - Signal transduction (Kinase, ubiquitin ligase, acetylation,... substrates)
  - Stable protein complexes
  - miR-mediated mRNA-mRNA interactions
  - Metabolic reactions
  - Chromatin modification enzymes
  - ...
- 80/20 Rule
  - Reconstruct 80% of relevant interactions with fewer than 20% false positives from experimental validation.

# Collection

Integrative CNV,  
Methylation, and  
mutation analysis

Wang et al., in review  
Ayetes et al., in review  
Compagno et al., 2009 Nature  
Mani et al., 2008 Mol Sys Bio

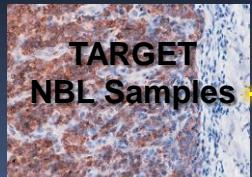
Signal Transduction  
Network Analysis

Zhao X et al., 2009 Dev Cell.  
Wang et al., 2008 Nat Biotech

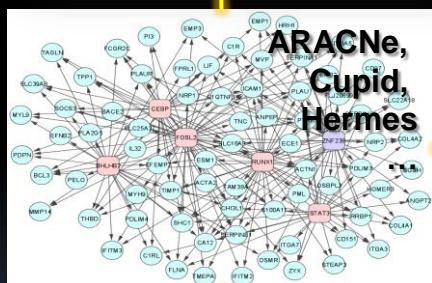
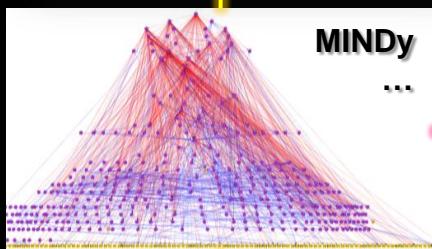
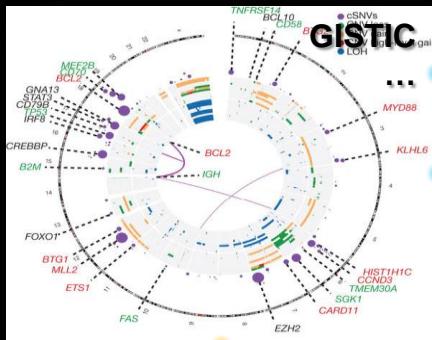
Transcriptional (TF)  
Post-transcr. (miR)  
Network Analysis

Sumazin et al. 2011 Cell  
Carro et al. 2010 Nature  
Basso et al., 2005 Nat Genet

Tissue collection

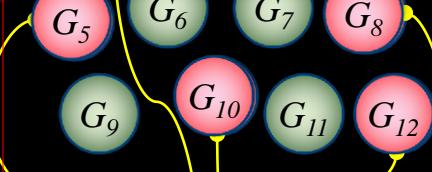


# Reconstruction



# Interrogation

Modulators



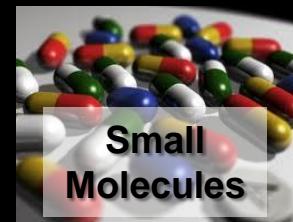
MRs



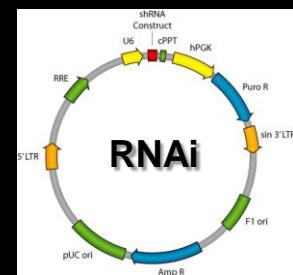
Signatures

# Validation

Methods



Small  
Molecules



Contexts



In Vitro