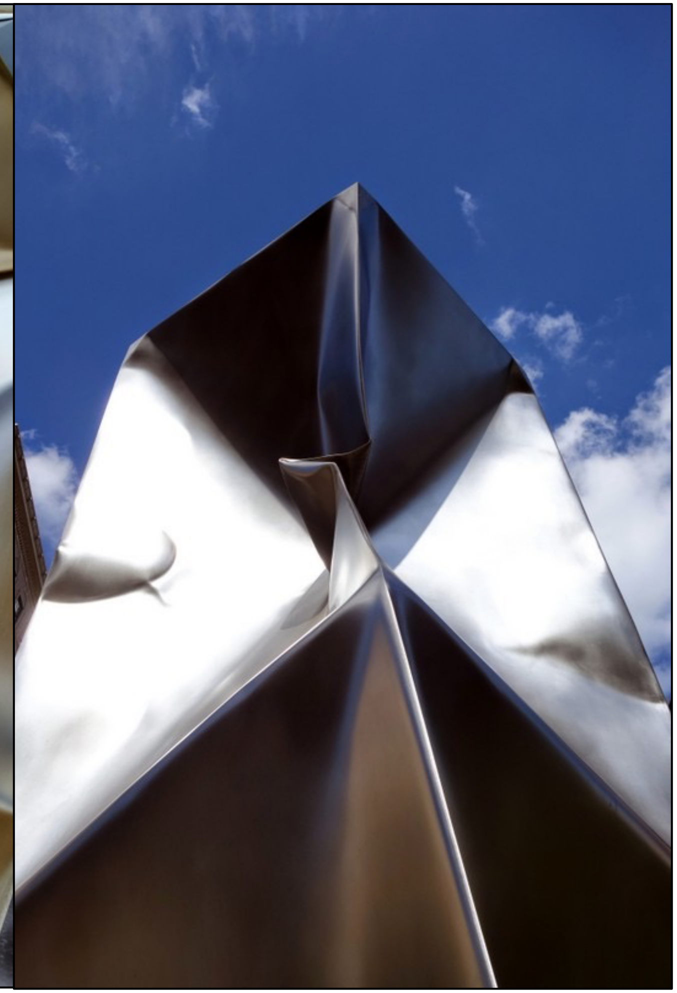
(Human Genome Analysis)

# Large-scale Transcriptome Mining: Building Interpretative Models while Protecting Individual Privacy

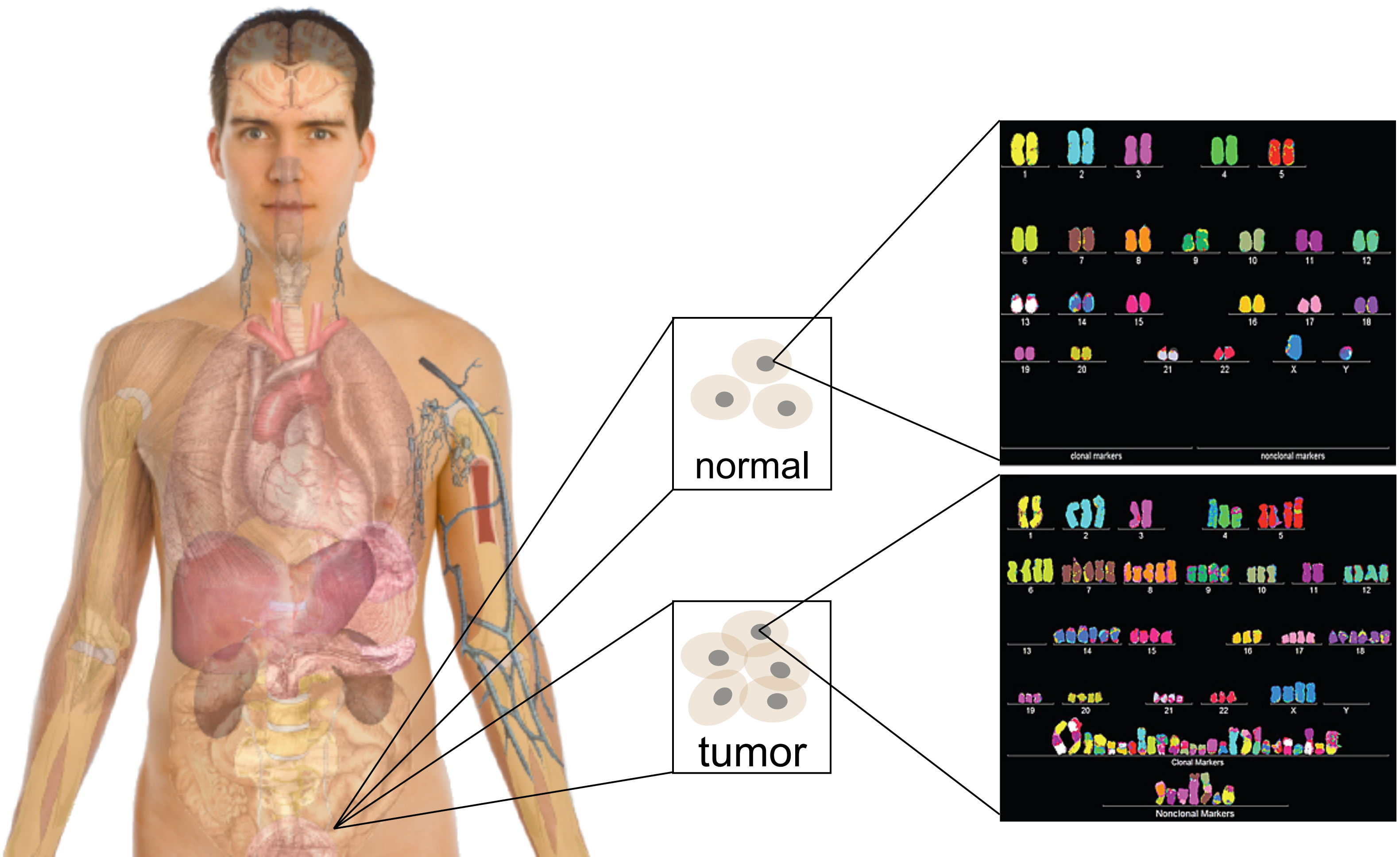


Mark Gerstein, Yale

Slides freely downloadable from Lectures.GersteinLab.org & “tweetable” (via @markgerstein). See last slide for more info.

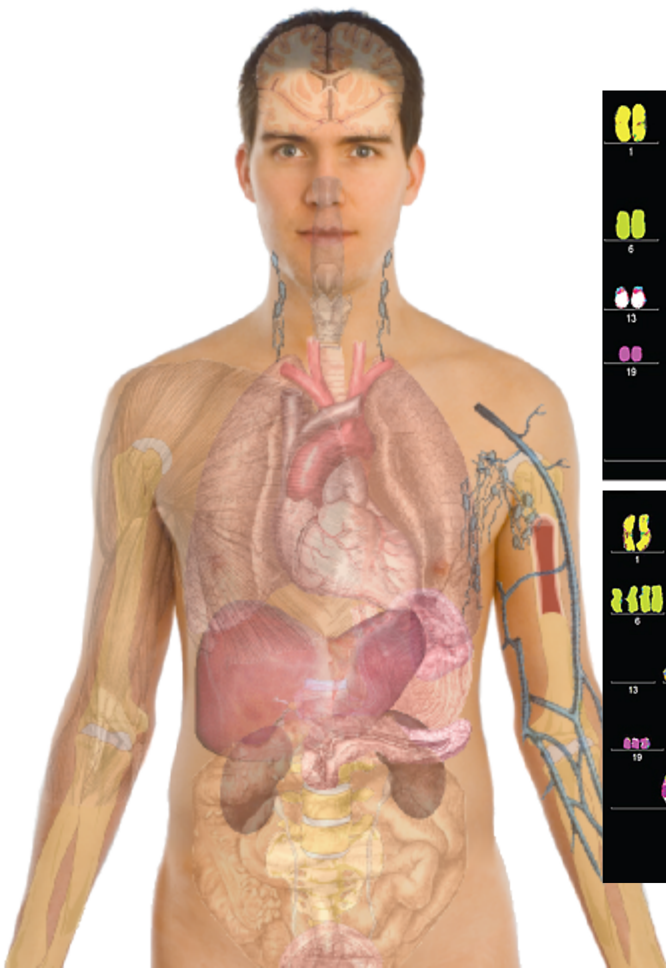
# Personal Genomics
# as a Gateway into Biology

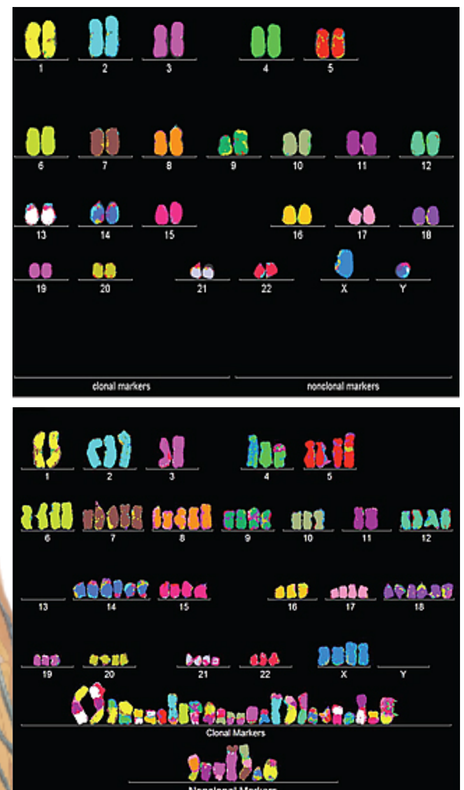
Personal genomes soon will become a commonplace part of medical research & eventually treatment (esp. for cancer). They will provide a primary connection for biological science to the general public.

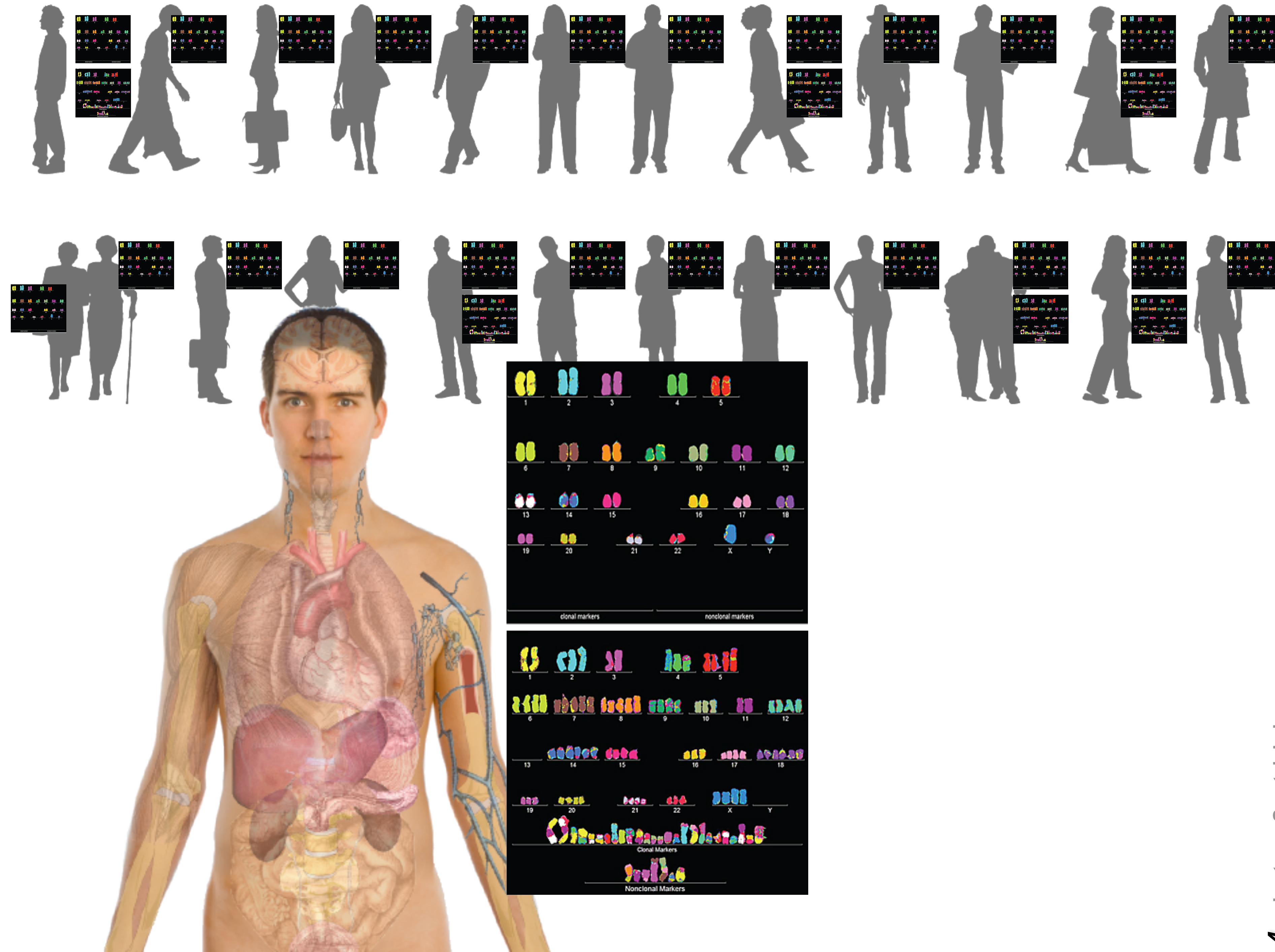


# Personal Genomics as a Gateway into Biology

Personal genomes soon will become a commonplace part of medical research & eventually treatment (esp. for cancer). They will provide a primary connection for biological science to the general public.






clonal markers
nonclonal markers
Clonal Markers
Nonclonal Markers

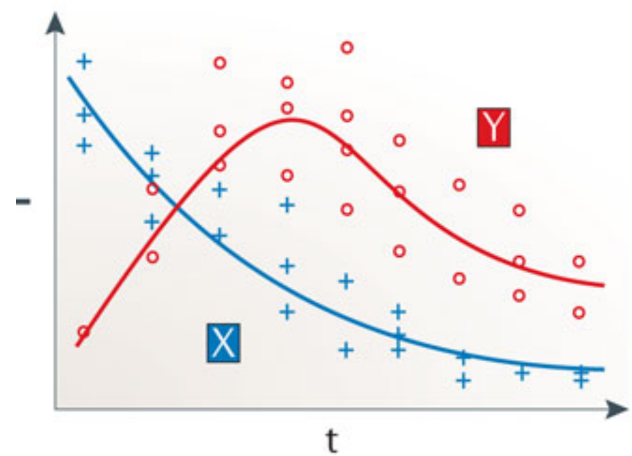
# Building Regulatory Models from Large-scale RNA-seq Data

- Boolean logical model
- Continuous model

| Key | *gi,ga,gb,gc:* input genes | off |
|---|---|---|
| | *go:* output gene | on |

| | Logic | | Example | |
|---|---|---|---|---|
| | Operator | Definition | Vector Function | Model |
| Boolean | **NOT** | the output is off if the input is on | *go*: if **NOT** *gi*=1 then=1 else=0 | gi go |
| | **OR** | the output is on if at the least one of the inputs is on | *go*: if *ga*=1 **OR** *gb*=1 then=1 else=0 | ga gb go |
| | **AND** | the output is on only if both inputs are on | *go*: if *ga*=1 **AND** *gb*=1 then=1 else=0 | ga gb go |
| | **AND NOT** | the output is on if the first input is on and the second is off | *go*: if *ga*=1 **AND NOT** *gb*=1 then=1 else=0 | ga gb go |
| | **[ ]** | brackets for subsidiary functions | *go*: if *ga*=1 AND **[** *gb*=1 OR *gc*=1 **]** then=1 else=0 | ga gb gc go |
| | the vector equation can incorporate different module or functions | | *go*: if **Mod1** OR **Mod2** then=1 else=0<br>**Mod1** : if *ga*=1 then=1 else=0<br>**Mod2** : if *gb*=1 then=1 else=0 | ga gb go |

Istrail & Davidson, PNAS, '04




$$\frac{dx_i}{dt} = \sum_{j=1}^{n} a_{i,j} x_j$$

Nicolas Le Novère, Nature Reviews Genetics, '15

# Privacy Aspects of Large-scale RNA-seq Analysis

- Large magnitude of RNA-seq data generated
  - ENCODE, modENCODE, TCGA, GTEx, Roadmap, psychENCODE, etc.
- Mostly the data is about the phenotype (e.g., cancer gene expression), but the individual information often comes along as collateral
  - Maybe we can separate private info but couple it with the public presentation?

- **The Dilemma of Genomic Privacy**
  - Fundamental, inherited info that's very private vs the need for large-scale data-sharing to enable med. research
  - Current Social & Tech Approaches
    - Issues: burdensome security, inconsistencies + ways the solutions have been partially "hacked"
    - Strawman Hybrid Soc-Tech Proposal (Cloud Enclaves. Quantifying Leaks, & Closely Coupled priv.-public data)
- **RNA-seq: How to Publicly Share Some of it**
  - Removing SNVs in reads using MRF
  - Quantifying & removing variant info from expression levels + eQTLs
  - Linking Attack using extreme expression levels

# Large-scale Transcriptome Mining: Building Interpretative Models while Protecting Individual Privacy

- **Large-scale Mining of RNA-seq to Determine State Space Models**
  - Using dimensionality reduction to help determine internal & external drivers
  - Decoupling expression changes into those driven by worm-fly conserved genes vs species-specific ones. Also, Conserved genes have similar canonical patterns (iPDPs) in contrast to species specific ones (Ex of ribosomal v signaling genes)
  - In human cell cycle, only conserved genes show matching periodic pattern

# The Conundrum of Genomic Privacy: Is it a Problem?

## Yes

Genetic Exceptionalism :
genome is potentially very revealing about one's identity & characteristics

- Most discussion of Identification Risk but what about Characterization Risk?
  - Finding you were in study X vs identifying that you have trait Y from studying your identified genome

## No

**Shifting societal foci**

**No one really cares about *__your__* genes**

**__You__ might not care**


[Klitzman & Sweeney ('11), J Genet Couns 20:98l; Greenbaum & Gerstein ('09), New Sci. (Sep 23) ]

# Tricky Privacy Considerations in Personal Genomics

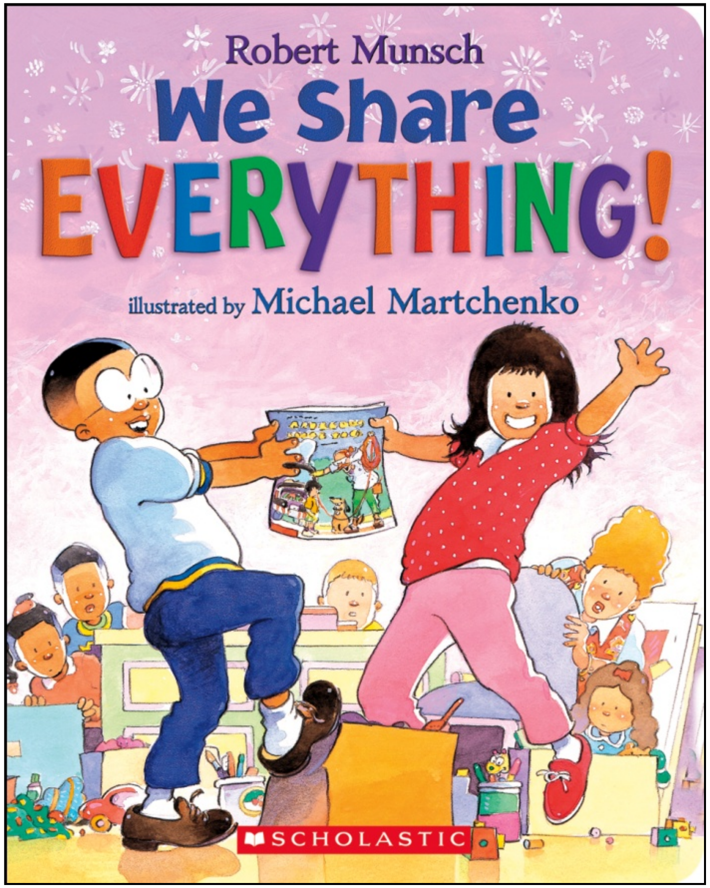
- Personal Genomic info. essentially meaningless currently but will it be in 20 yrs? 50 yrs?
  - Genomic sequence very revealing about one's children. Is true consent possible?
  - Once put on the web it can't be taken back
- Culture Clash: Genomics historically has been a proponent of "open data" but not clear personal genomics fits this
- Ethically challenged history of genetics

- Ownership of the data & what consent means (Hela)
  - Could your genetic data give rise to a product line?

[D Greenbaum & M Gerstein ('08). Am J. Bioethics; D Greenbaum & M Gerstein, Hartford Courant, 10 Jul. '08 ; SF Chronicle, 2 Nov. '08; Greenbaum et al. *PLOS CB* ('11) ; Greenbaum & Gerstein ('13), The Scientist; Photo from NY Times]

## The Other Side of the Coin: Why we should share

- Sharing helps speed research
  - Large-scale mining of this information is important for medical research
  - Privacy is cumbersome, particularly for big data
  - Sharing is important for reproducible research
- Sharing is useful for education




[Yale Law Roundtable ('10). Comp. in Sci. & Eng. 12:8; D Greenbaum & M Gerstein ('09). Am. J. Bioethics; D Greenbaum & M Gerstein ('10). SF Chronicle, May 2, Page E-4; Greenbaum et al. *PLOS CB* ('11)]

[Economist, 15 Aug '15]

## The Dilemma

- What is acceptable risk? What is acceptable data leakage? Can we quantify leakage?
- Cost Benefit Analysis: how helpful is identifiable data in genomic research v. potential harm from a breach?
- The individual (harmed?) v the collective (benefits)
  - But do sick patients care about their privacy?
- Maybe a we need a few "test pilots" (ala PGP)?
  - Sports stars & celebrities?

# Genomics has similar "Big Data" Dilemma in the Rest of Society

- Sharing & "peer-production" is central to success of many new ventures, with the same risks as in genomics
- We confront privacy risks every day we access the internet
- (...or is the genome more exceptional & fundamental?)

[Seringhaus & Gerstein ('09), *Hart. Courant* (Jun 5); Greenbaum & Gerstein ('11), *NY Times* (6 Oct)]

# Current Social & Technical Solutions

- Consents
- "Protected" distribution of data (dbGAP)
- Local computes on secure computer

- Issues
  - Non-uniformity of consents & paperwork
    - Different international norms, leading to confusion
  - Encryption & computer security creates burdensome requirements on data sharing & large scale analysis
  - Many schemes get "hacked"

[Greenbuam et al ('04), Nat. Biotech; Greenbaum & Gerstein ('13), The Scientist]

# Privacy Hacks

- Personalized genomic data generation is booming
- "Detection of genome in a mixture"
  - Individuals give consent to participate but request anonymity
    - HAPMAP, Personal genome project, 1000 Genomes…
- Larger and more datasets leads to more realistic risks of linking attacks, that may be much more damaging than detection of genome in a mixture attacks
- Main focus is on protecting variants

**Identifying Personal Genomes by Surname Inference**

Melissa Gymrek,[1,2,3,4] Amy L. McGuire,[5] David Golan,[6] Eran Halperin,[7,8,9] Yaniv Erlich[1*]

**Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays**

Nils Homer[1,2], Szabolcs Szelinger[1], Margot Redman[1], David Duggan[1], Waibhav Tembe[1], Jill Muehling[1], John V. Pearson[1], Dietrich A. Stephan[1], Stanley F. Nelson[2], David W. Craig[1*]

On Sharing Quantitative Trait GWAS Results in an Era of Multiple-omics Data and the Limits of Genomic Privacy

Hae Kyung Im,[1,*] Eric R. Gamazon,[2] Dan L. Nicolae,[2,3,4] and Nancy J. Cox[2,3,*]

**Identifying Participants in the Personal Genome Project by Name**

Latanya Sweeney, Akua Abu, Julia Winn

Harvard College
Cambridge, Massachusetts
latanya@fas.harvard.edu, aabu@college.harvard.edu, jwinn@post.harvard.edu

NETFLIX

IMDb

Earth's Biggest Movie Database™

# Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)

Arvind Narayanan and Vitaly Shmatikov

The University of Texas at Austin

February 5, 2008



**Abstract**

We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge.

We apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous movie ratings of 500,000 subscribers of Netflix, the world's largest online movie rental service. We demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, we successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.

**Cross correlated small set of identifiable IMDB movie database rating records with large set of "anonymized" Netflix customer ratings**

# Strawman Hybrid Social & Tech Proposed Solution?

- Fundamentally, researchers have to keep genetic secrets
  - Genetic Licensure & training for individuals (similar to medical license, drivers license)
- Technology to make things easier
  - Cloud computing & enclaves (eg solution of Genomics England)
- Technological barriers shouldn't create a social incentive for "hacking"

- **Quantifying Leakage & allowing a small amounts of it (eg photos of eye color)**
- **Careful separation & coupling of private & public data**
  - **Lightweight, freely accessible secondary datasets coupled to underlying variants**
  - Selection of stub & "test pilot" datasets for benchmarking
  - Develop programs on public stubs on your laptop, then move the program to the cloud for private production run

[D Greenbaum, M Gerstein ('11). Am J Bioeth 11:39. Greenbaum & Gerstein, The Scientist ('13)]

- **The Dilemma of Genomic Privacy**
  - Fundamental, inherited info that's very private vs the need for large-scale data-sharing to enable med. research
  - Current Social & Tech Approaches
    - Issues: burdensome security, inconsistencies + ways the solutions have been partially "hacked"
    - Strawman Hybrid Soc-Tech Proposal (Cloud Enclaves. Quantifying Leaks, & Closely Coupled priv.-public data)
- **RNA-seq: How to Publicly Share Some of it**
  - Removing SNVs in reads using MRF
  - Quantifying & removing variant info from expression levels + eQTLs
  - Linking Attack using extreme expression levels

## Large-scale Transcriptome Mining: Building Interpretative Models while Protecting Individual Privacy

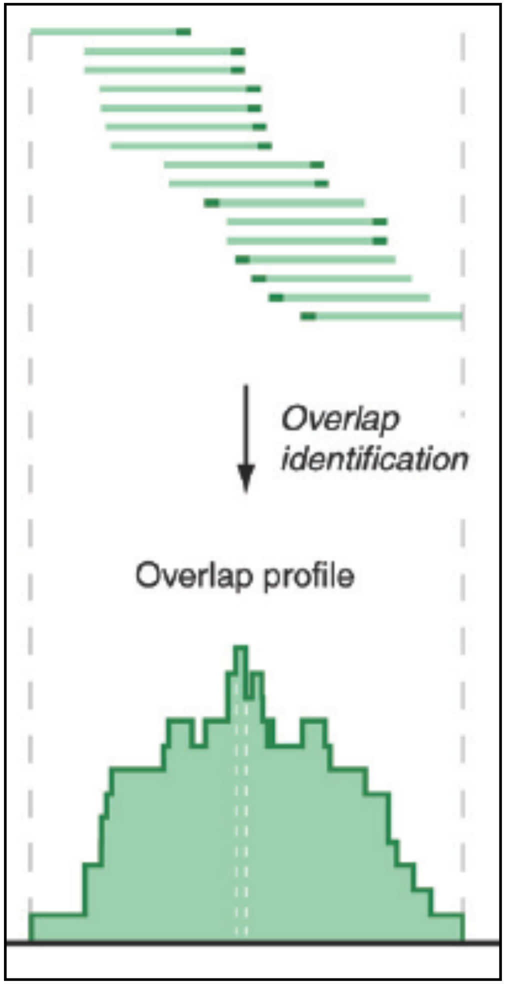
- **Large-scale Mining of RNA-seq to Determine State Space Models**
  - Using dimensionality reduction to help determine internal & external drivers
  - Decoupling expression changes into those driven by worm-fly conserved genes vs species-specific ones. Also, Conserved genes have similar canonical patterns (iPDPs) in contrast to species specific ones (Ex of ribosomal v signaling genes)
  - In human cell cycle, only conserved genes show matching periodic pattern

# RNA-seq

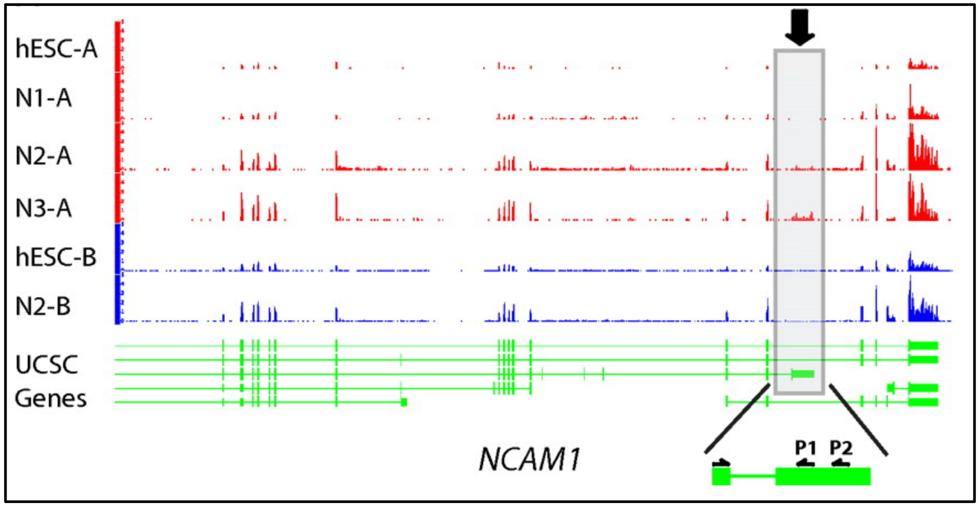
RNA-seq uses next-generation sequencing technologies to reveal RNA presence and quantity within a biological sample.

```
ATACAAGCAAGTATAAGTTCGTATGCCGTCTT
GGAGGCTGGAGTTGGGGACGTATGCGGCATAG
TACCGATCGAGTCGACTGTAAACGTAGGCATA
ATTCTGACTGGTGTCATGCTGATGTACTTAAA
```

Reads (fasta)
- Quality scores (fastq)
- Mapping (BAM)
- Contain variant information in transcribed regions




Reads => Signal

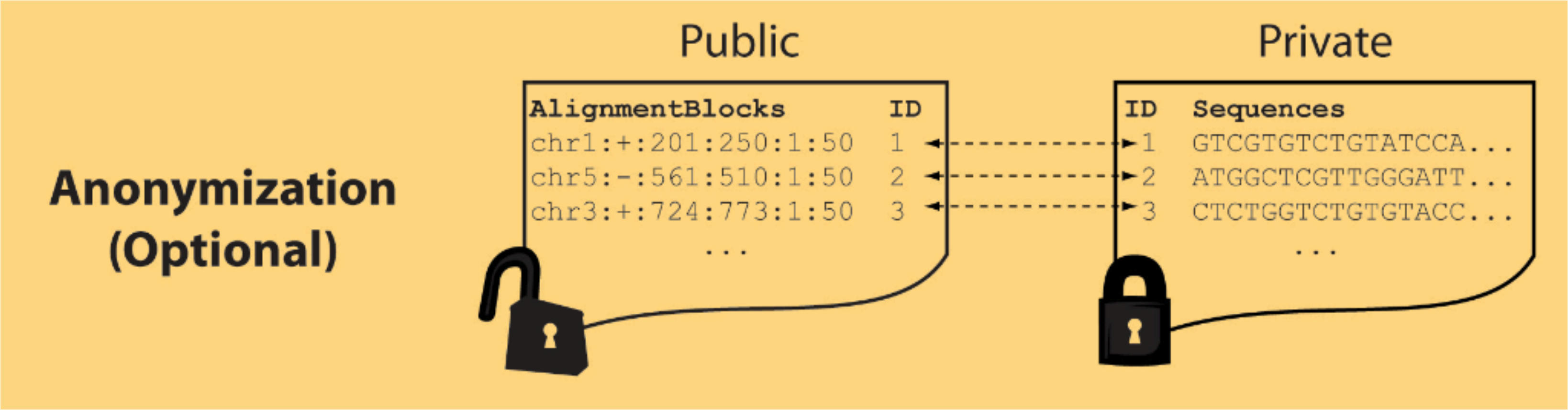


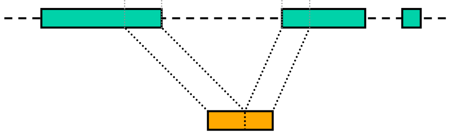

Quantitative information from RNA-seq signal: average signals at exon level (RPKMs)

[*PLOS CB* 4:e1000158; *PNAS* 4:107: 5254 ; *IJC* 123:569 ]

# Light-weight formats

- Some lightweight format clearly separate public & private info., aiding exchange
- Files become much smaller
- Distinction between formats to compute on and those to archive with – become sharper with big data



**Mapping coordinates without variants (MRF)**

**Reads (linked via ID, 10X larger than mapping coord.)**

[*Bioinformatics* 27: 281]

# MRF Examples

**10X Compression Ex.**

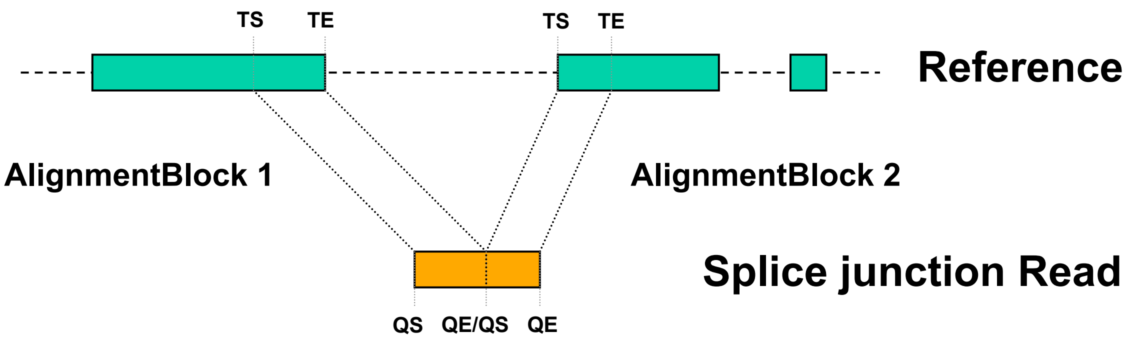
**Raw ELAND** export file has uncompressed file size: ~4 GB; total number of reads: ~20 million; number of mapped reads: ~12 million .

**MRF file** is significantly smaller (~400 MB uncompressed, ~130 MB compressed with gzip).

**BAM file**
has a size of ~1.2 GB.

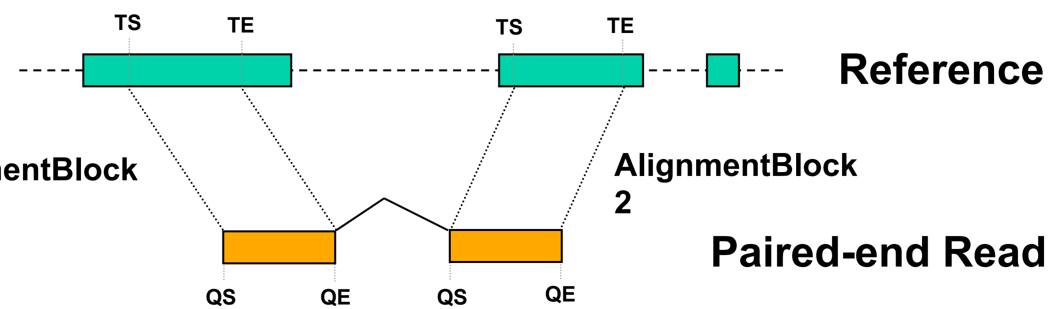
Reference based compression (ie CRAM) is similar but it stores actual variant beyond just position of alignment block

`chr2:+:601:630:1:30,chr2:+:921:940:31:50`




Legend: TS = TargetStart, TE = TargetEnd, QS = QueryStart, QE = QueryEnd

`chr9:+:431:480:1:50|chr9:+:945:994:1:50`



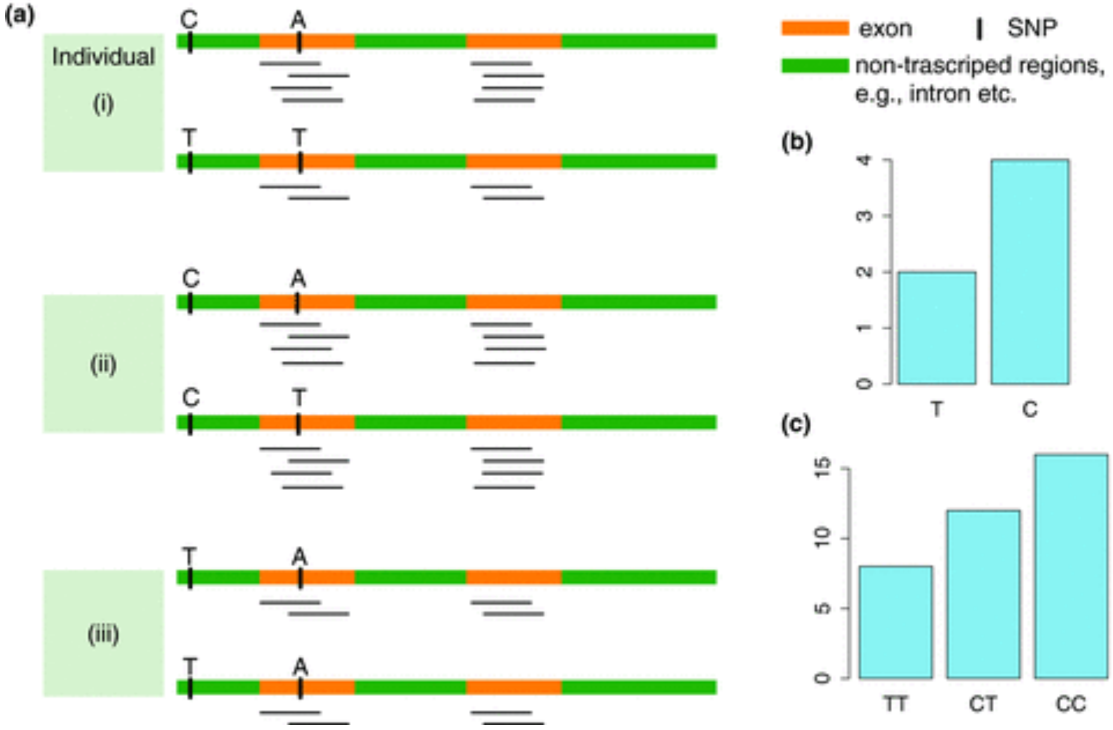

Legend: TS = TargetStart, TE = TargetEnd, QS = QueryStart, QE = QueryEnd

[Habegger et al., Bioinformatics ('11)]

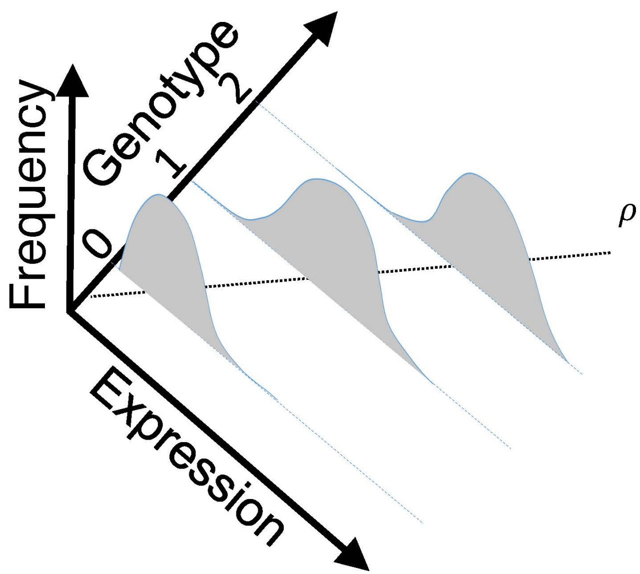
- **The Dilemma of Genomic Privacy**
  - Fundamental, inherited info that's very private vs the need for large-scale data-sharing to enable med. research
  - Current Social & Tech Approaches
    - Issues: burdensome security, inconsistencies + ways the solutions have been partially "hacked"
    - Strawman Hybrid Soc-Tech Proposal (Cloud Enclaves. Quantifying Leaks, & Closely Coupled priv.-public data)
- **RNA-seq: How to Publicly Share Some of it**
  - Removing SNVs in reads using MRF
  - Quantifying & removing variant info from expression levels + eQTLs
  - Linking Attack using extreme expression levels

## Large-scale Transcriptome Mining: Building Interpretative Models while Protecting Individual Privacy

- **Large-scale Mining of RNA-seq to Determine State Space Models**
  - Using dimensionality reduction to help determine internal & external drivers
  - Decoupling expression changes into those driven by worm-fly conserved genes vs species-specific ones. Also, Conserved genes have similar canonical patterns (iPDPs) in contrast to species specific ones (Ex of ribosomal v signaling genes)
  - In human cell cycle, only conserved genes show matching periodic pattern



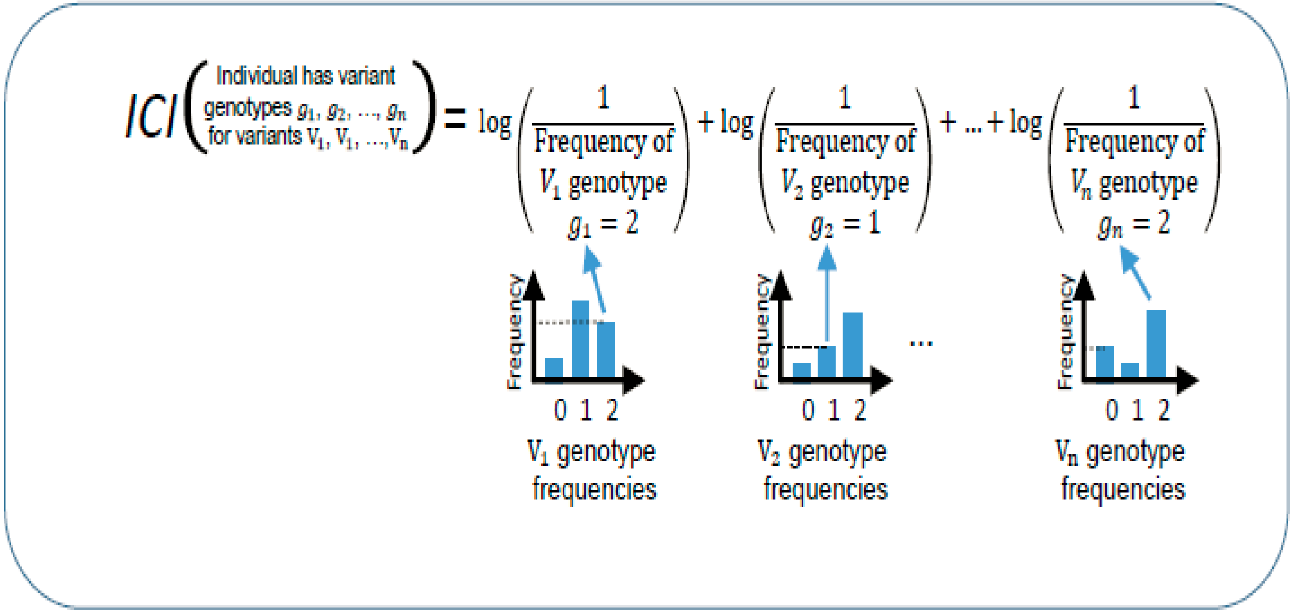

[Biometrics 68(1) 1–11]

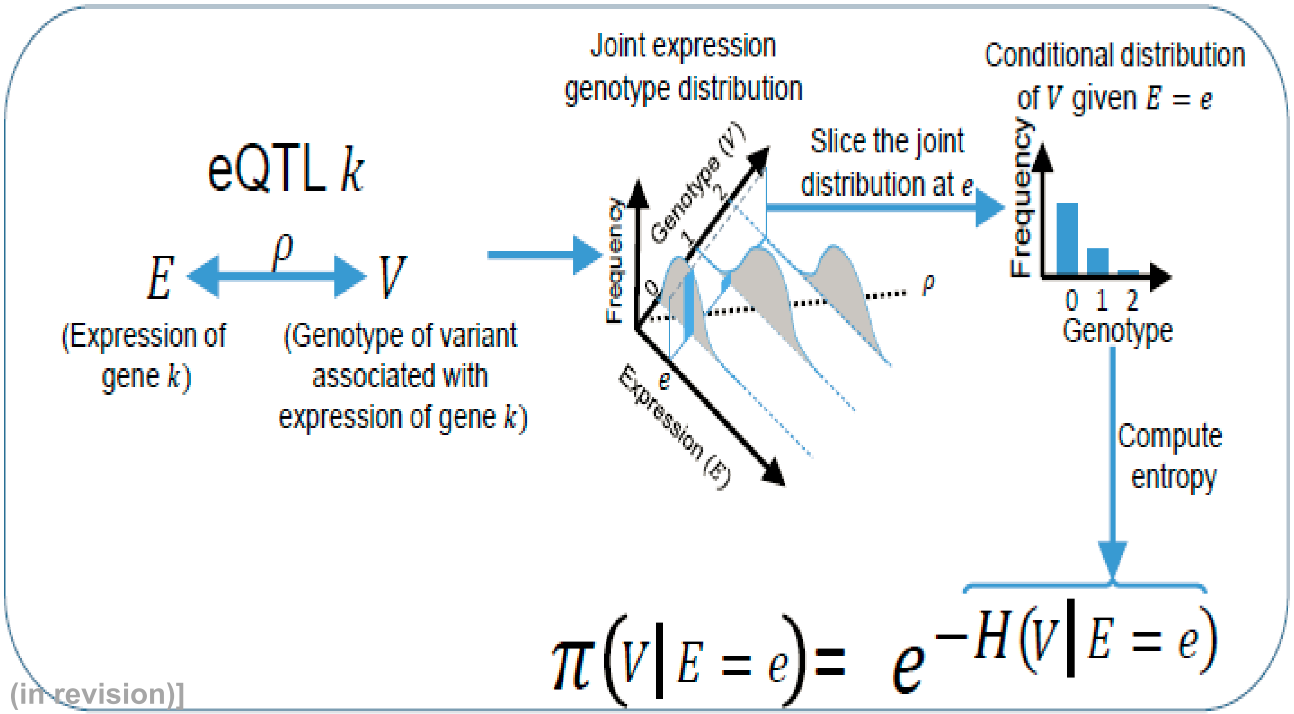



# eQTL Mapping Using RNA-Seq Data

- eQTLs are genomic loci that contribute to variation in mRNA expression levels
- eQTLs provide insights on transcription regulation, and the molecular basis of phenotypic outcomes
- eQTL mapping can be done with RNA-Seq data

# Information Content and Predictability



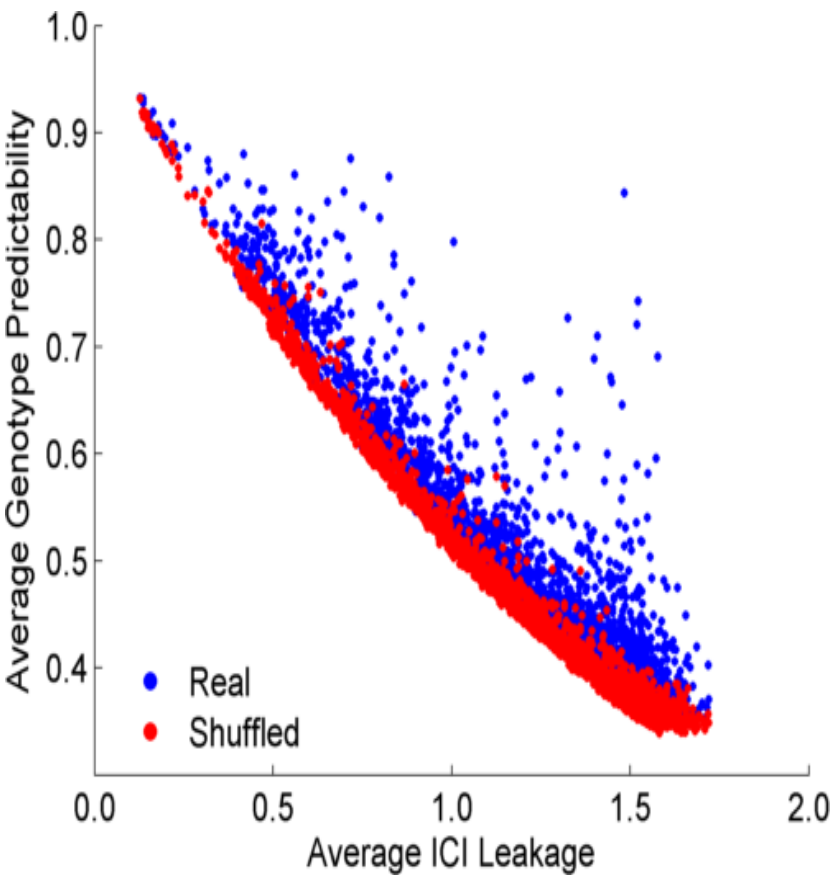




[Harmanciet al. Nat. Meth. (in revision)]

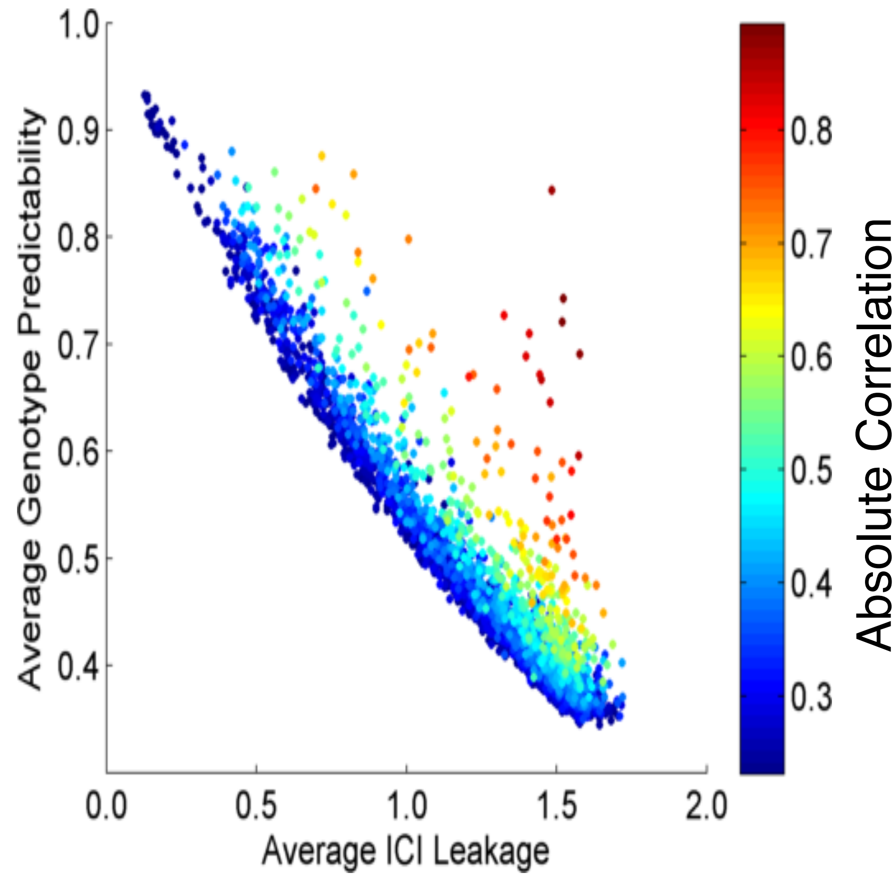
# Representative Expression, Genotype, eQTL Datasets

- mRNA sequencing for 462 individuals
  - Publicly availableQuantification for protein coding genes
- Approximately 3,000 cis-eQTL (FDR<0.05)
- Genotypes are available from the 1000 Genomes Project

# Per eQTL and ICI Cumulative Leakage versus Genotype Predictability

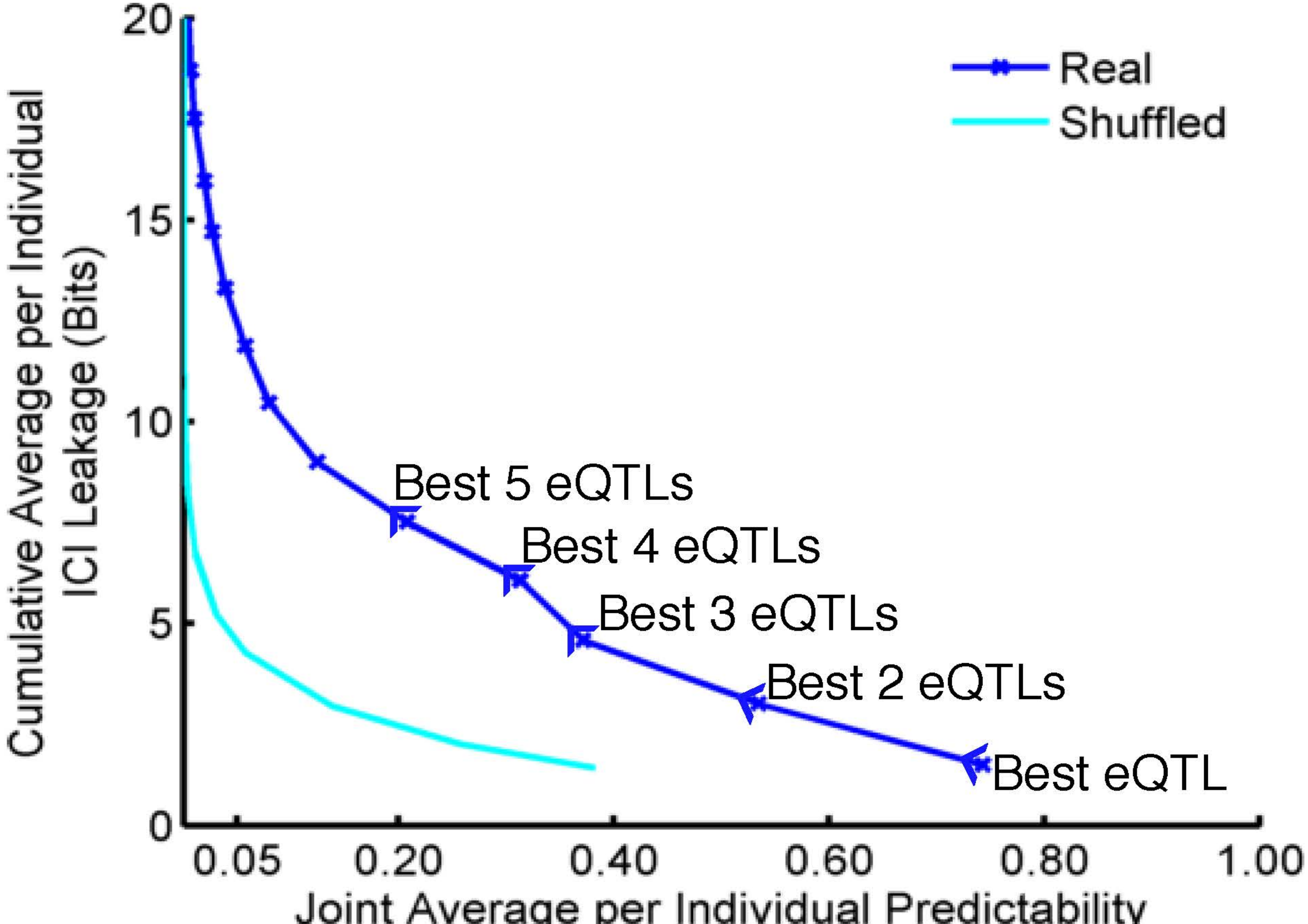



Colors by absolute correlation



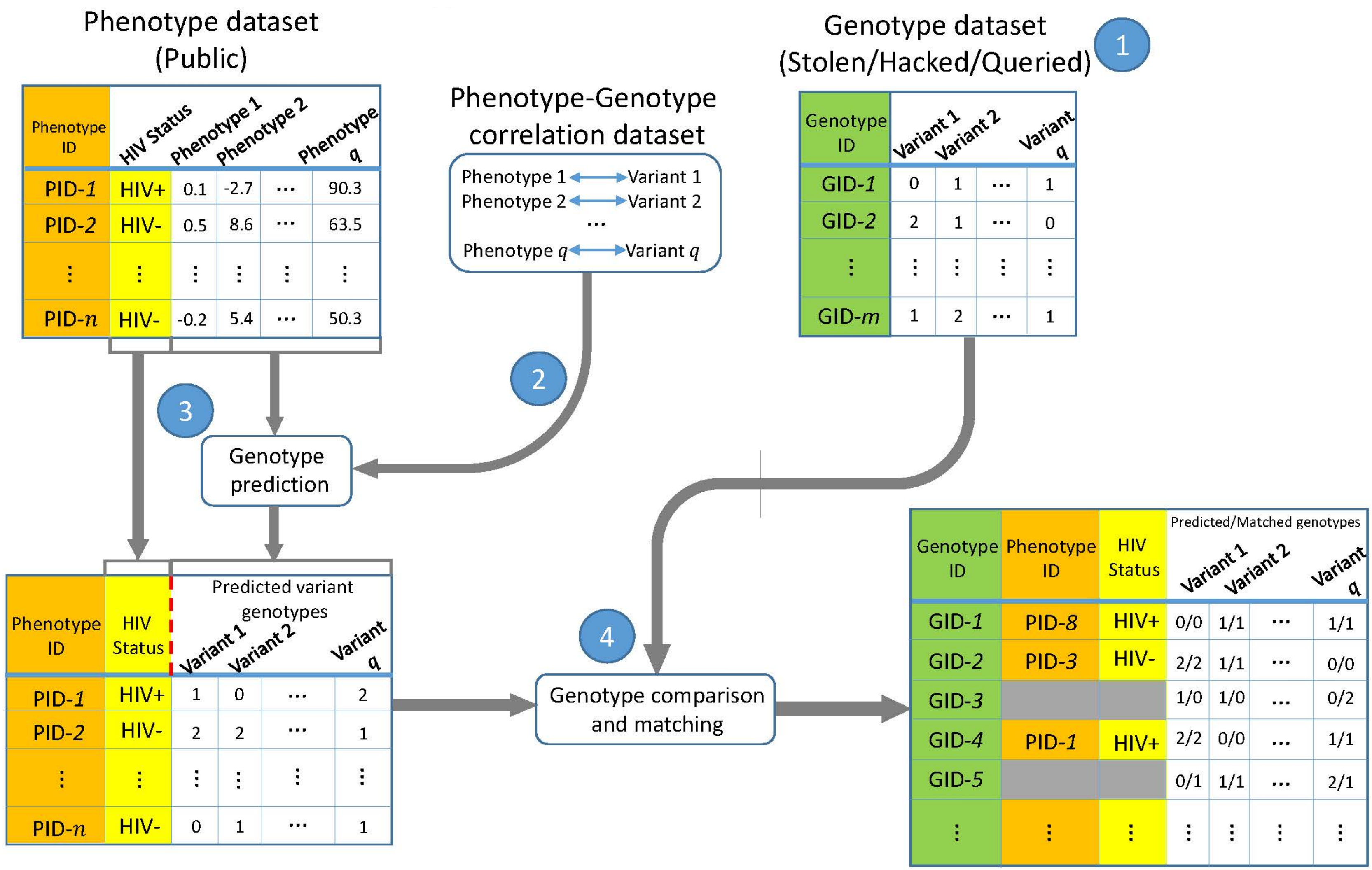

[Harmanciet al. Nat. Meth. (in revision)]

# Cumulative Leakage versus Joint Predictability



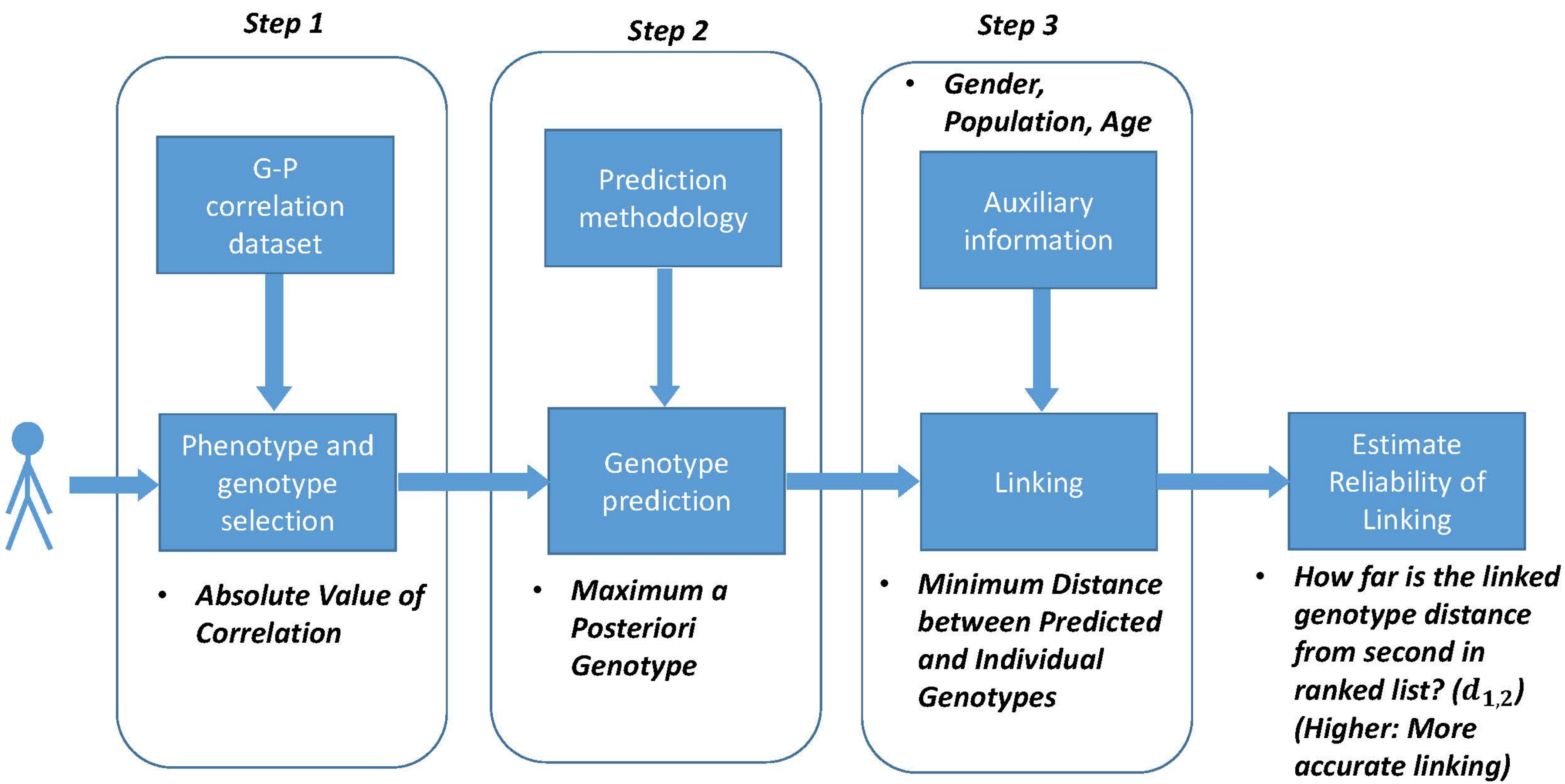

[Harmanciet al. Nat. Meth. (in revision)]

# Linking Attack Scenario

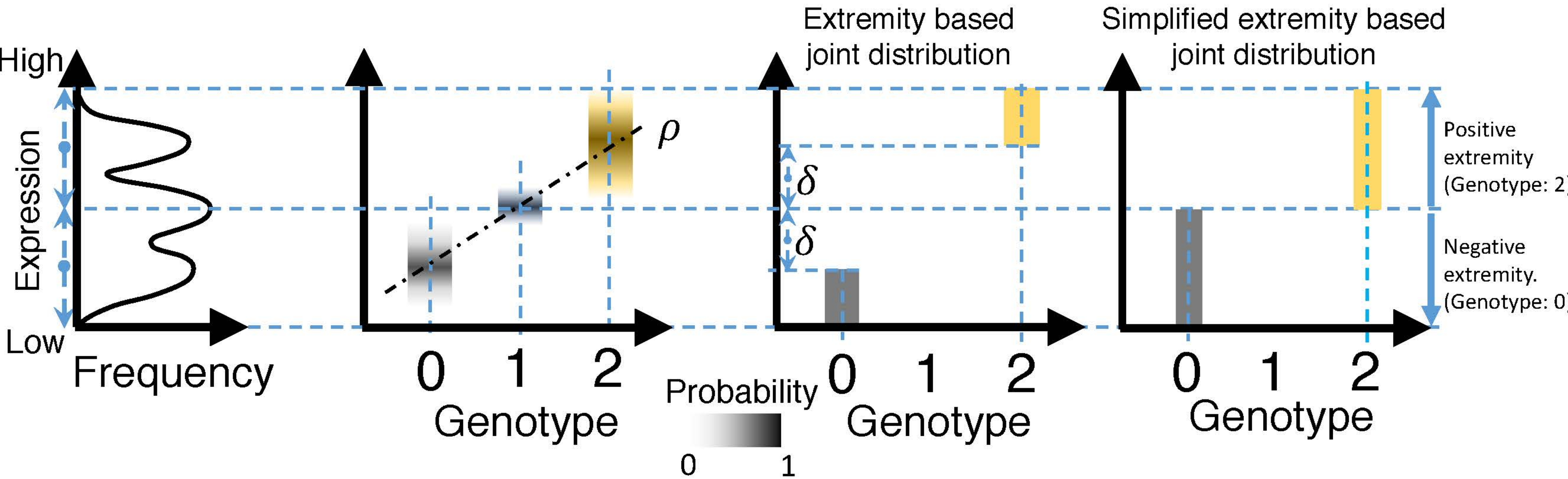



[Harmanciet al. Nat. Meth. (in revision)]

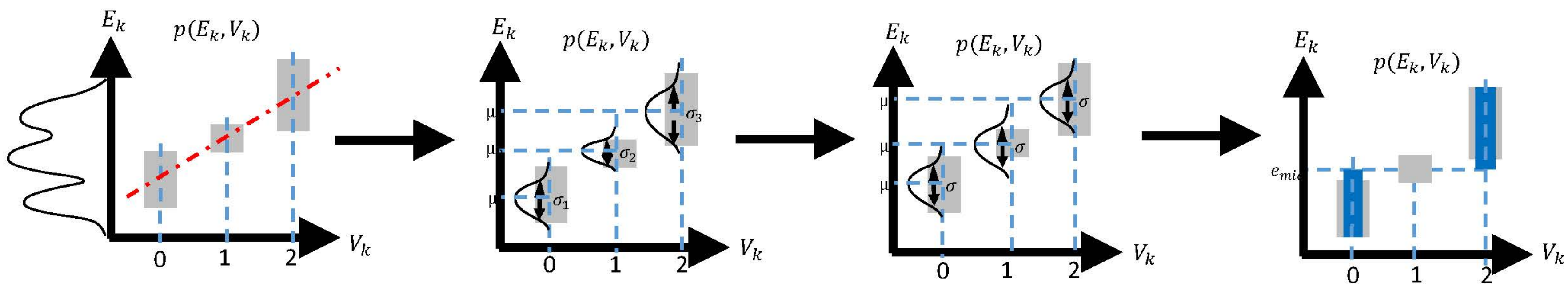
# Steps in Instantiation of a (Mock) Linking Attack




[Harmanciet al. Nat. Meth. (in revision)]

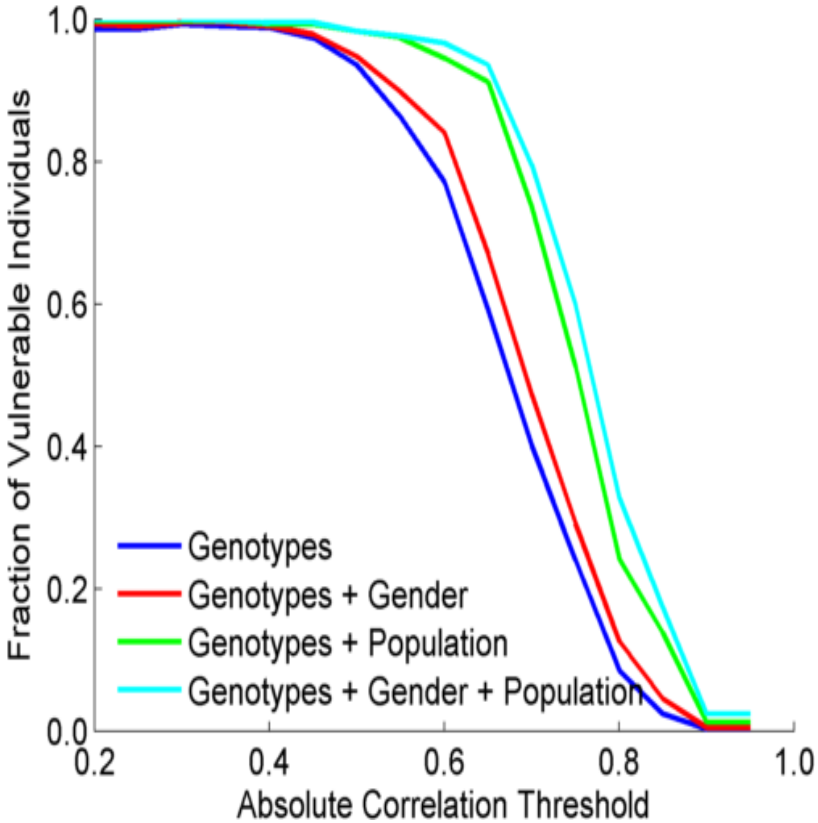



Levels of Expression-Genotype Model Simplifications:

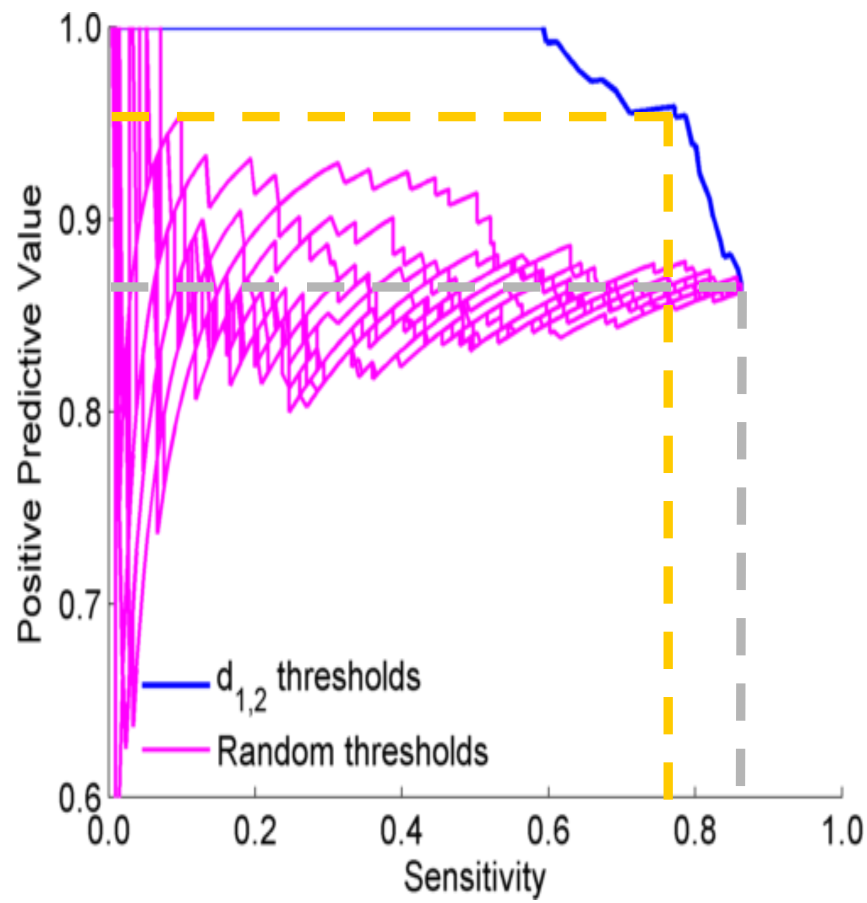



[Harmanciet al. Nat. Meth. (in revision)]

# Extremity based linking with homozygous genotypes




# Attacker can estimate the reliability of linkings




Sensitivity: Fraction of correctly linked Individuals among all individuals

PPV: Fraction of correctly linked individuals among selected individuals

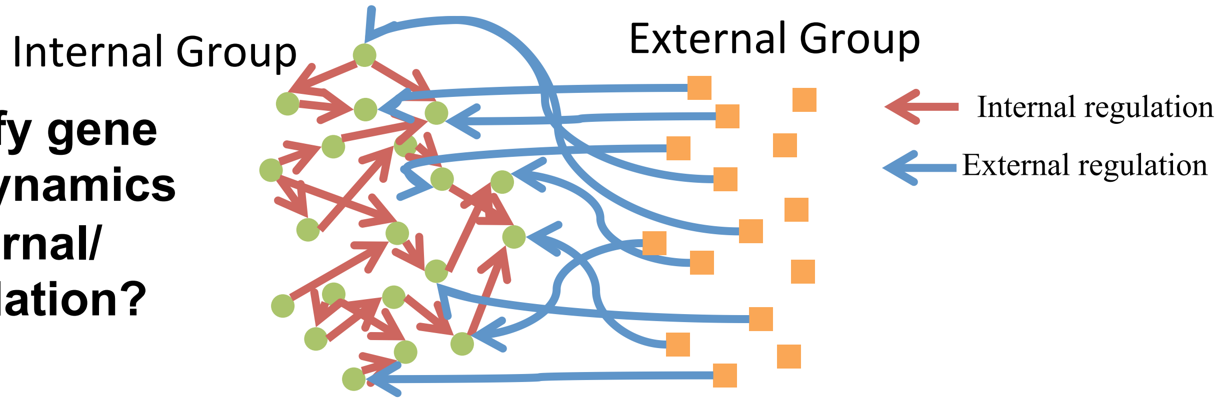
[Harmanciet al. Nat. Meth. (in revision)]

- **The Dilemma of Genomic Privacy**
  - Fundamental, inherited info that's very private vs the need for large-scale data-sharing to enable med. research
  - Current Social & Tech Approaches
    - Issues: burdensome security, inconsistencies + ways the solutions have been partially "hacked"
    - Strawman Hybrid Soc-Tech Proposal (Cloud Enclaves. Quantifying Leaks, & Closely Coupled priv.-public data)
- **RNA-seq: How to Publicly Share Some of it**
  - Removing SNVs in reads using MRF
  - Quantifying & removing variant info from expression levels + eQTLs
  - Linking Attack using extreme expression levels

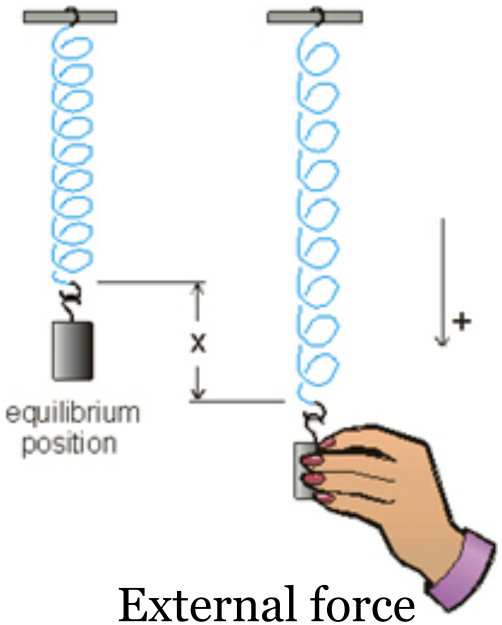
## Large-scale Transcriptome Mining: Building Interpretative Models while Protecting Individual Privacy

- **Large-scale Mining of RNA-seq to Determine State Space Models**
  - Using dimensionality reduction to help determine internal & external drivers
  - Decoupling expression changes into those driven by worm-fly conserved genes vs species-specific ones. Also, Conserved genes have similar canonical patterns (iPDPs) in contrast to species specific ones (Ex of ribosomal v signaling genes)
  - In human cell cycle, only conserved genes show matching periodic pattern

# Internal and external gene regulatory networks




**How to identify gene expression dynamics driven by internal/ external regulation?**

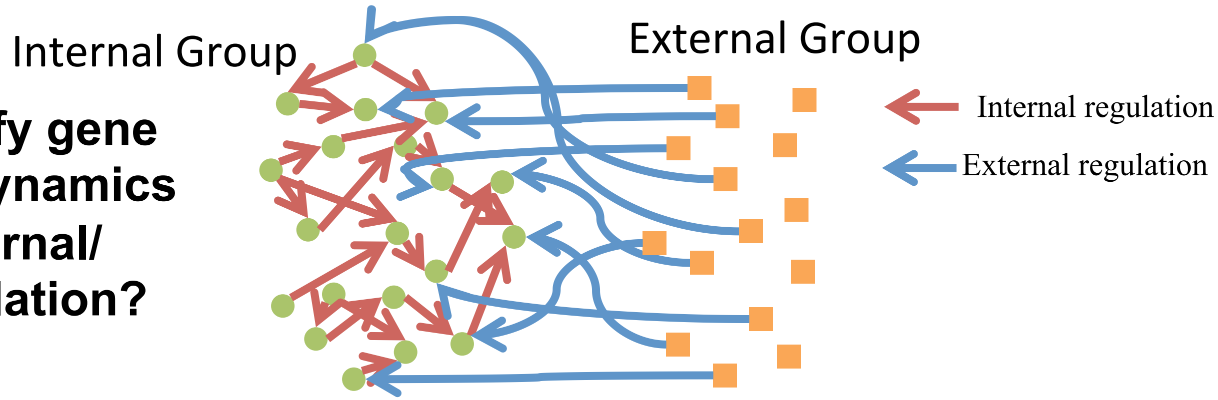



| Interested system | Internal regulatory network | External regulatory network |
|---|---|---|
| Cross-species conserved genes | Conserved transcriptional factors (TFs) | Non-conserved TFs |
| Protein-coding genes | TFs | micro-RNAs |
| Individual's protein coding genes | Wild-type TFs | Somatic mutated TFs |
| Protein-coding genes in brain | Commonly expressed TFs | Brain-specific expressed TFs |
| Protein-coding genes in development | House-keeping TFs | Developmental TFs |

[Wang et al. *PLOS CB* (in revision, '15)]

# State-space model for internal and external gene regulatory networks

**How to identify gene expression dynamics driven by internal/ external regulation?**




State space model

$$X_{t+1} = A \; X_t + B \; U_t$$

Control: Gene expression vector of external factors at time $t$

$B_{kl}$ captures temporal casual influence from external factor $k$ to Gene $l$ in internal group

State: Gene expression vector of Group $X$ at time $t$+1

$A_{ij}$ captures temporal casual influence from Gene $i$ to Gene $j$ in internal group

State: Gene expression vector of internal group at time $t$

[Wang et al. *PLOS CB* (in revision, '15)]

# Effective state space model for meta-genes

Not enough data to estimate state space model for genes

(e.g., 25 time points per gene to estimate 4 million elements of $A$ or $B$ for 2000 genes)

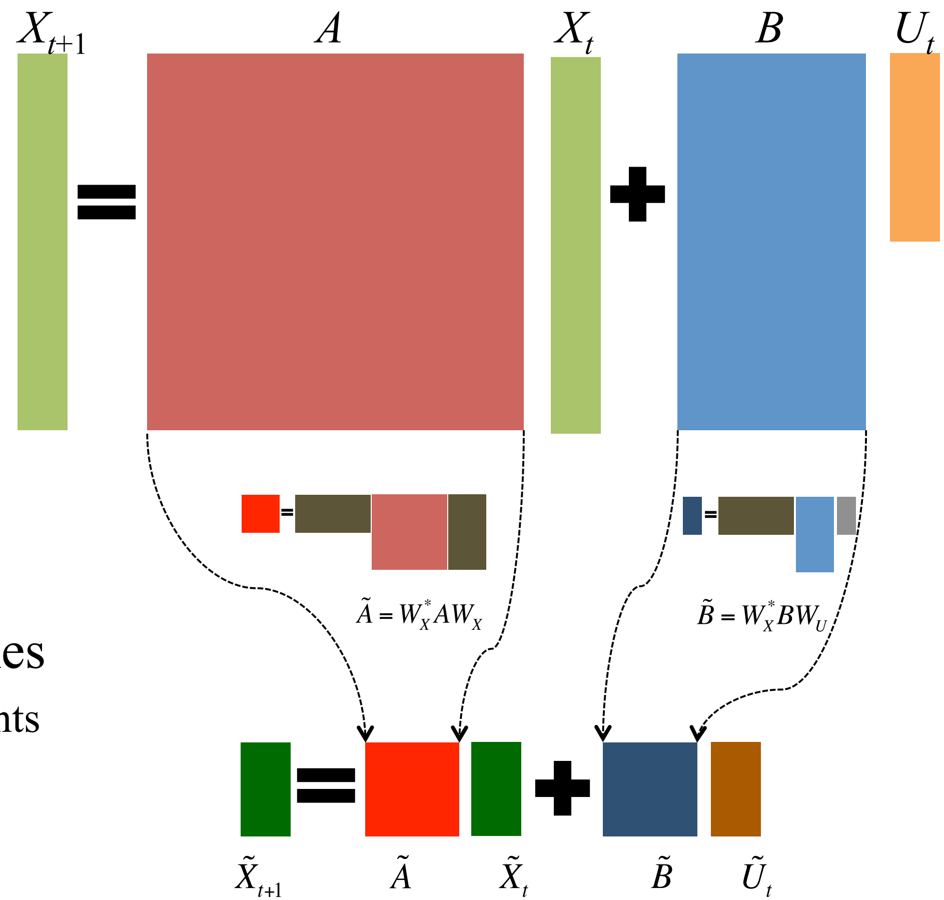
$$X_{t+1} = AX_t + BU_t$$

Dimensionality reduction from genes to meta-genes (e.g., SVD)

Effective state space model for meta-genes

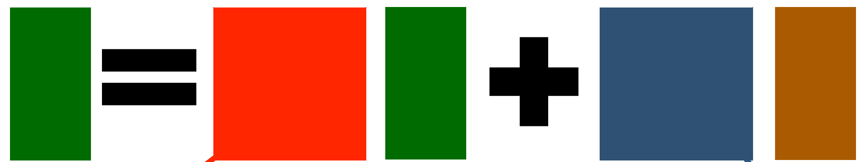
(e.g., 250 time points to estimate 50 matrix elements if 5 meta-genes)

$$\boxed{\tilde{X}_{t+1} = \tilde{A}\tilde{X}_t + \tilde{B}\tilde{U}_t}$$




[Wang et al. *PLOS CB* (in revision, '15)]

# Canonical temporal expression trajectories from effective state space model



$$\tilde{X}_{t+1} = \tilde{A}\,\tilde{X}_t + \tilde{B}\,\tilde{U}_t$$

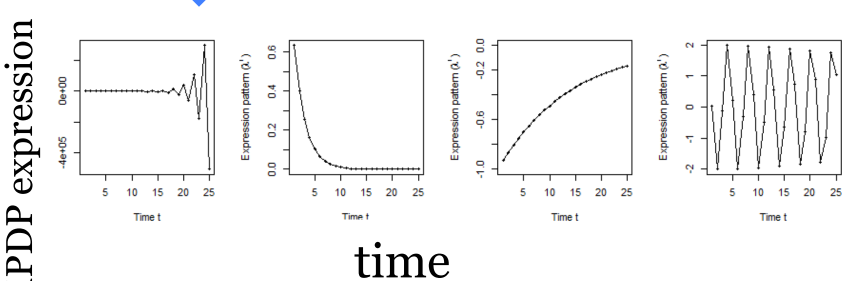
Internal driven dynamics

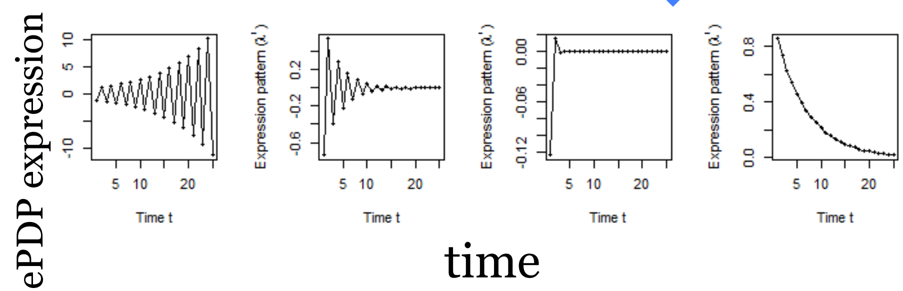
$p^{th}$ internal principal dynamic pattern (iPDP): $[\lambda_p^{\,1}, \lambda_p^{\,2}, \ldots, \lambda_p^{\,T}]$,
where $\lambda_p$ is $p^{th}$ eigenvalue of $\tilde{A}$.

Externally driven dynamics

$q^{th}$ external principal dynamic pattern (ePDP): $[\sigma_q^{\,1}, \sigma_q^{\,2}, \ldots, \sigma_q^{\,T}]$,
where $\sigma_q$ is $q^{th}$ eigenvalue of $\tilde{B}$.

Canonical temporal expression trajectories
(e.g., degradation, growth, damped oscillation, etc.)

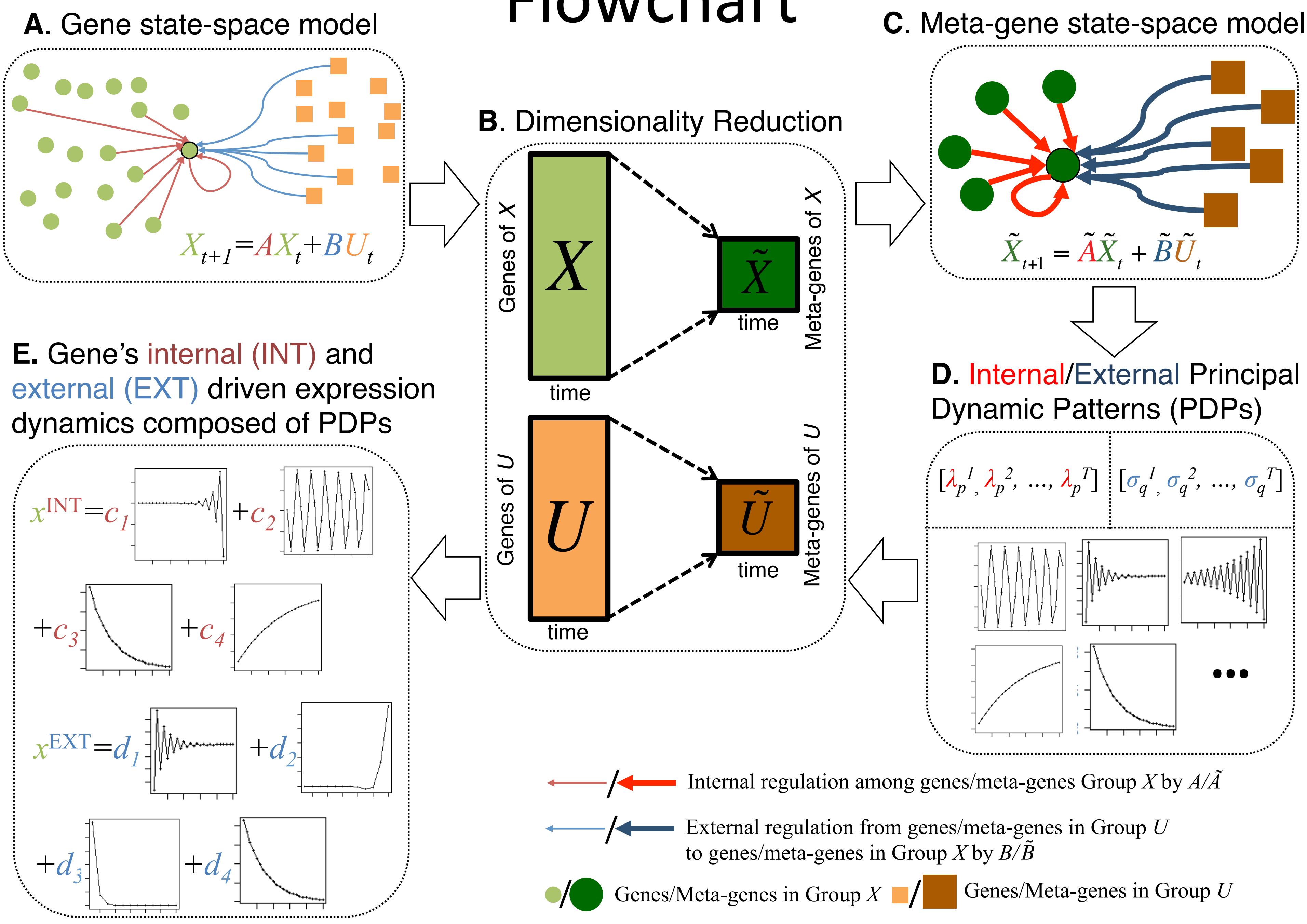


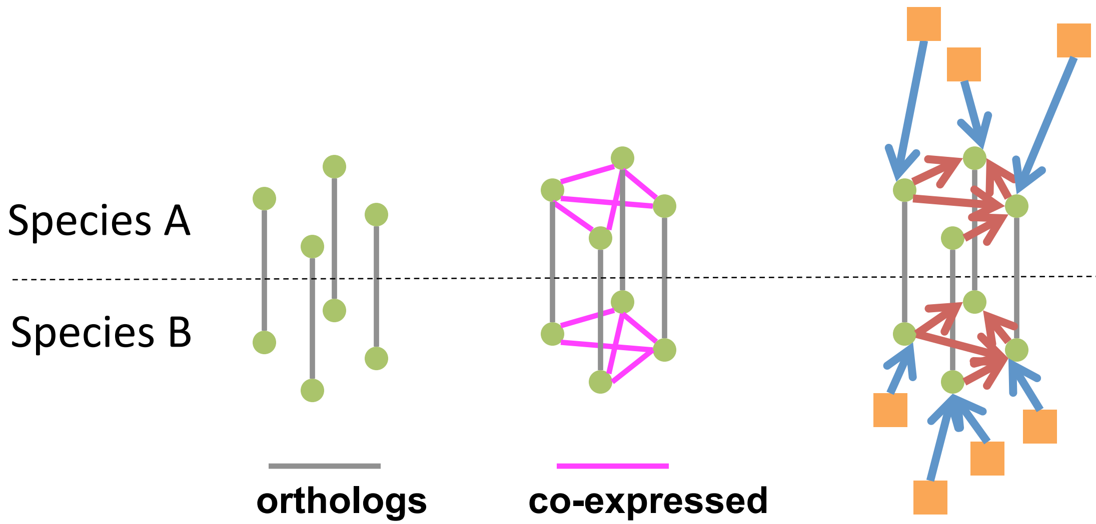

[Wang et al. *PLOS CB* (in revision, ‘15)]

[Wang et al. *PLOS CB* (in revision, '15)]

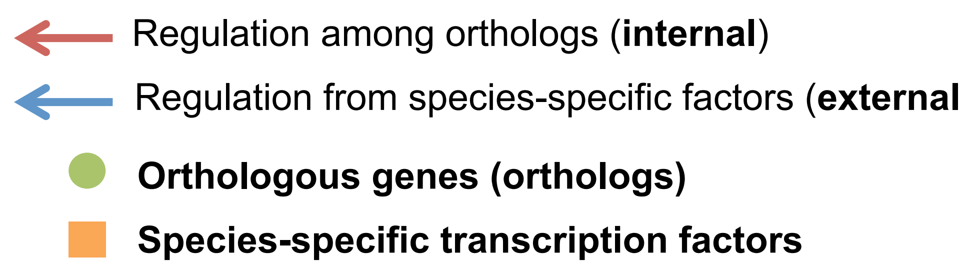
# Flowchart



# Are gene regulations among orthologs conserved across species?



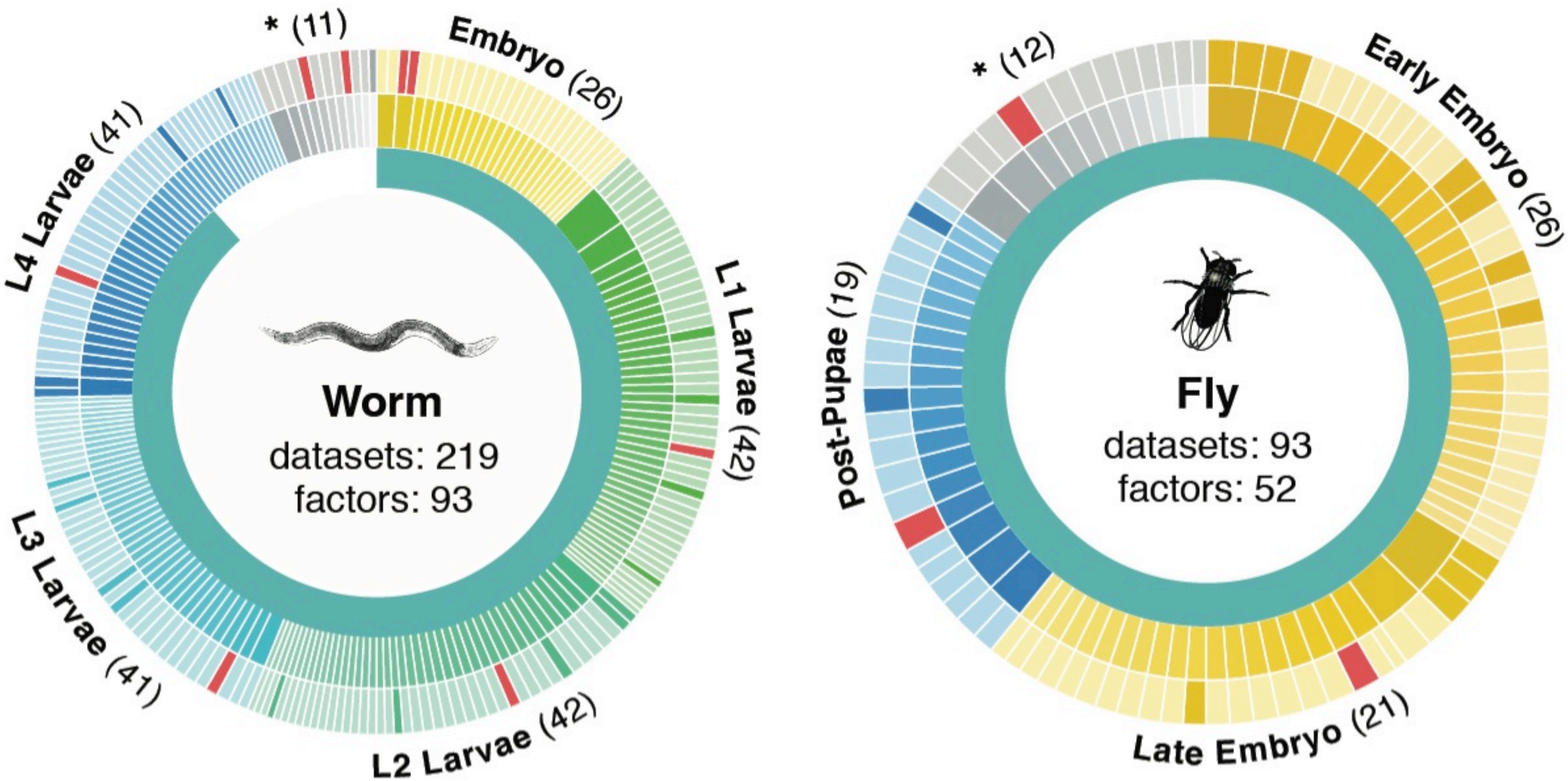

*Are gene regulatory networks among orthologs conserved across species?*




**To what degree can't ortholog expression levels be predicted due to species-specific regulation**

[Wang et al. *PLOS CB* (in revision, '15)]

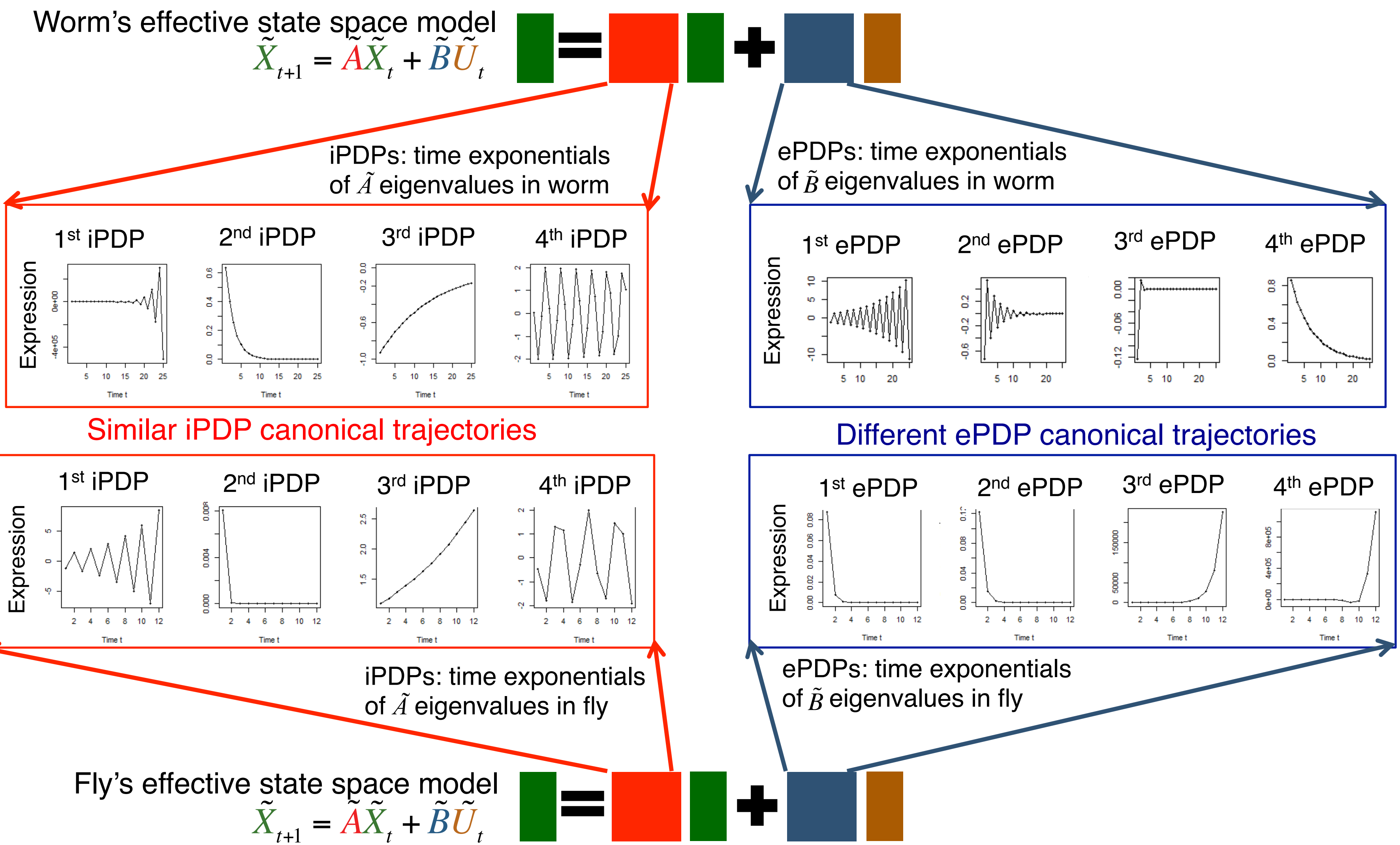
# Time-course gene expression data of worm & fly development




| Organism | Major developmental stages |
|---|---|
| worm (*C. elegans*) | 33 stages: 0, 0.5, 1, …, 12 hours, L1, L2, L3, L4, …, Young Adults, Adults |
| fly (*D. mel.*) | 30 stages: 0, 2, 4, 6, 8,…, 20, 22 hours, L1-L4, Pupaes, Adults |

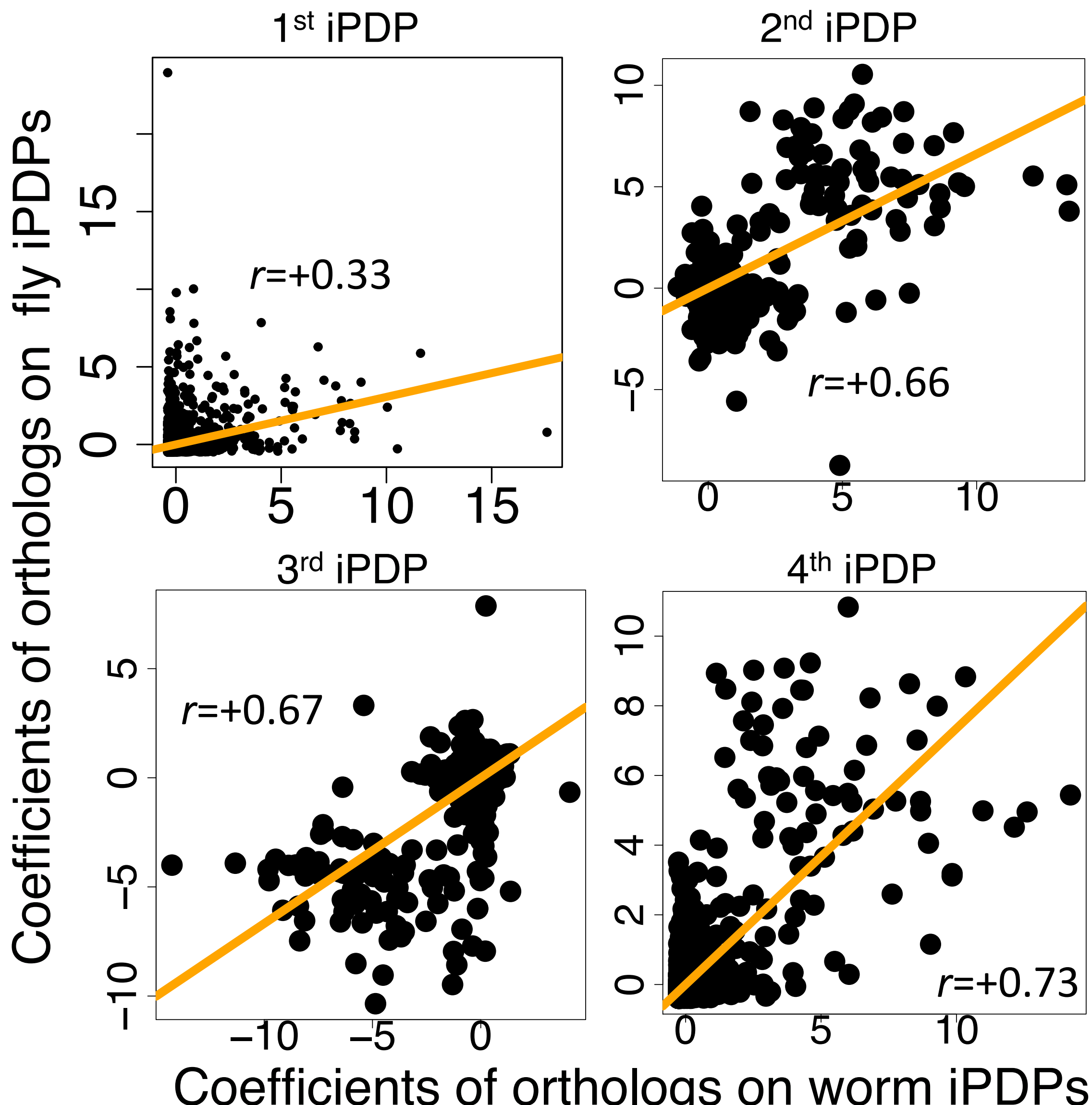
[*Nature* 512:445 ('14); doi: 10.1038/nature13424]

# Orthologs have similar internal but different external dynamic patterns during embryonic development




[Wang et al. *PLOS CB* (in revision, '15)]

# Orthologs have correlated iPDP coefficients




[Wang et al. *PLOS CB* (in revision, '15)]

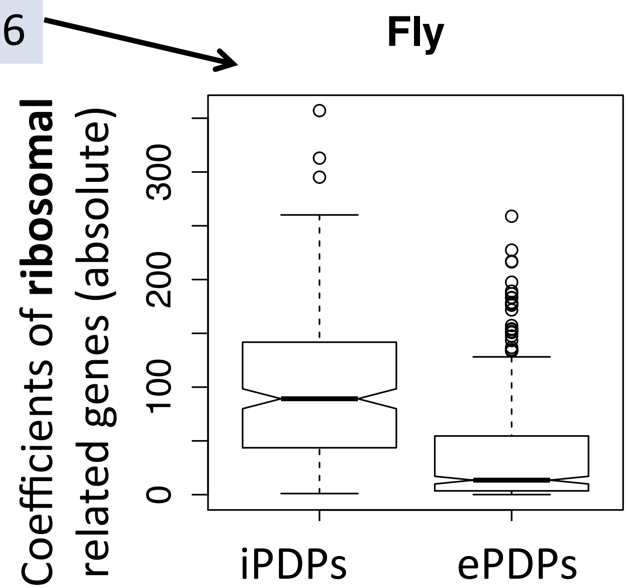
# Evolutionarily conserved and younger genes exhibit the opposite internal and external PDP coefficients

| iPDP coeffs > ePDP coeffs | Worm | Fly |
|---|---|---|
| Ribosomal genes | *p*<0.001 | *p*<2.2e-16 |

Ribosomal genes have significantly larger coefficients for the internal than external PDPs, but signaling genes exhibit the opposite trend

| iPDP coeffs < ePDP coeffs | Worm | Fly |
|---|---|---|
| Signaling genes | *p*<7e-4 | *p*<6e-4 |

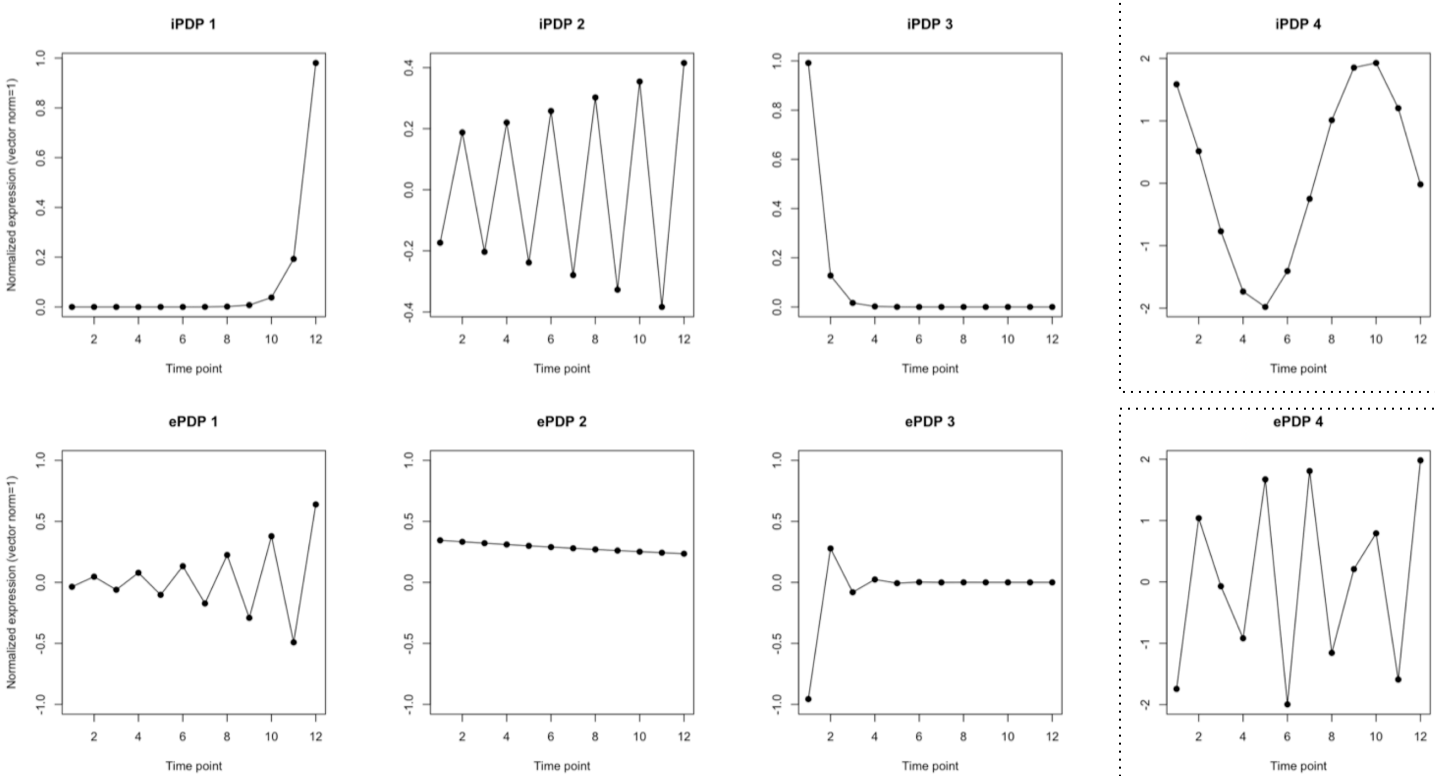
* *p*-values from KS-test




[Wang et al. *PLOS CB* (in revision, '15)]

# Breast cancer cell cycle under hormonal stimulation

| Dataset | Group *X* (internal) | Group *U* (external) | Time samples of a full cell cycle |
|---|---|---|---|
| Human breast cancer cell cycle under hormonal stimulation | 1132 metazoan conserved genes incl. 150 orthologous TFs | 1870 non-conserved metazoan transcription factors | T=12 time points: 0, 4, 6, 8, 12, …, 28, 32 hours |




Oscillated iPDP by conserved TFs

a full cell cycle

Oscillated ePDP by non-conserved TFs

faster cycle due to hormone

[Wang et al. *PLOS CB* (in revision, '15)]

- **The Dilemma of Genomic Privacy**
  - Fundamental, inherited info that's very private vs the need for large-scale data-sharing to enable med. research
  - Current Social & Tech Approaches
    - Issues: burdensome security, inconsistencies + ways the solutions have been partially "hacked"
    - Strawman Hybrid Soc-Tech Proposal (Cloud Enclaves. Quantifying Leaks, & Closely Coupled priv.-public data)
- **RNA-seq: How to Publicly Share Some of it**
  - Removing SNVs in reads using MRF
  - Quantifying & removing variant info from expression levels + eQTLs
  - Linking Attack using extreme expression levels

## Large-scale Transcriptome Mining: Building Interpretative Models while Protecting Individual Privacy

- **Large-scale Mining of RNA-seq to Determine State Space Models**
  - Using dimensionality reduction to help determine internal & external drivers
  - Decoupling expression changes into those driven by worm-fly conserved genes vs species-specific ones. Also, Conserved genes have similar canonical patterns (iPDPs) in contrast to species specific ones (Ex of ribosomal v signaling genes)
  - In human cell cycle, only conserved genes show matching periodic pattern

# Acknowledgements

**DREISS**.gersteinlab.org

**D Wang**, F He, S Maslov

papers.gersteinlab.org/subject/**privacy**

**D Greenbaum**

**PrivaSeq.**gersteinlab.org

**A Harmanci**

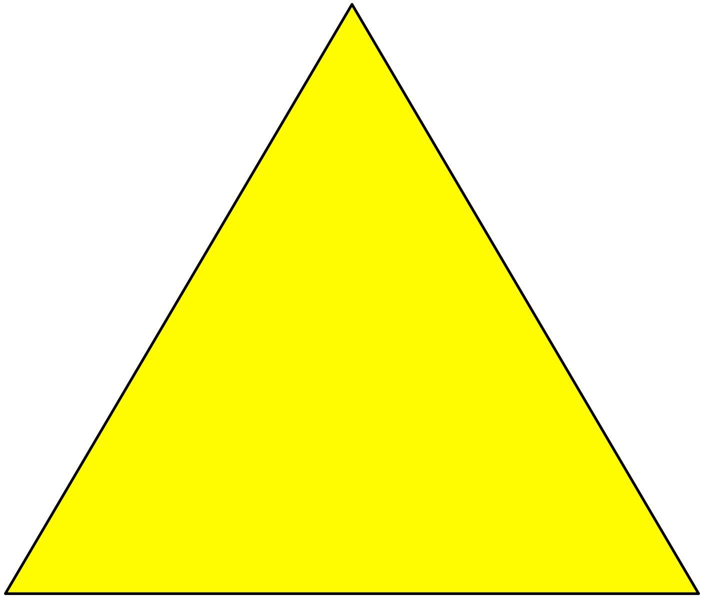
**RSEQtools.**gersteinlab.org

L Habegger, A Sboner, TA Gianoulis,
J Rozowsky, A Agarwal, M Snyder

**Hiring Postdocs. See gersteinlab.org/jobs !**

# Default Theme

- Default Outline Level 1
  - Level 2



# More Information on this Talk

SUBJECT: Networks

DESCRIPTION:

NOTES:
This PPT should work on mac & PC. Paper references in the talk were mostly from Papers.GersteinLab.org.



PHOTOS & IMAGES. For thoughts on the source and permissions of many of the photos and clipped images in this presentation see http://streams.gerstein.info . In particular, many of the images have particular EXIF tags, such as kwpotppt , that can be easily queried from flickr, viz: http://www.flickr.com/photos/mbgmbg/tags/kwpotppt .