Bayesian Graphical Models for Biomarker Relationships – Applications to Genomics Data

Bayesian Graphical models

One graph, two, and two hundred million...

2017 Nonclinical Biostatistics Conference

Yuan Ji, PhD

June 11, 2017

Collaborator & References

Collaborators

NorthShore	Yitan Zhu (Research Scientist)
UT Austin	Peter Müller (Profs)
Johns Hopkins	Yanxun Xu (Assist. Prof.)
U. of Louisville	Riten Mitra (Assist. Profs)
U. of Chicago	Lorenzo Pesce and Computational Institute and IGSB
Georgia Tech.	Peng Qiu (Assist. Prof.)

References

- ONE GRAPH :
  Mitra et al. (2013, JASA); Telesca et al. (2012, JASA)
- DIFFERENTIAL GRAPHS :
  Mitra et al. (2015a, Bayesian Analysis ; 2015b)
- TWO MILLION GRAPHS :
  Zhu et al. (2014, Nature Methods); Zhu et al. (2015, JNCI)

Website www.compgenome.org

ONE GRAPH

(Mitra et al., 2013; Telesca et al., 2013)

Bayesian Graphical Model – An overview

A class of Bayesian graphical hierarchical models

Bayesian paradigm:

$$\underbrace{\text{Prior Pathways } G_0}_{\textit{Graphical prior}} + \underbrace{\text{Data}}_{\textit{Likelihood}} \rightarrow \underbrace{\text{Posterior Pathways } G}_{\textit{Posterior knowledge}}$$

Bayesian Graphical Model – An overview

A class of Bayesian graphical hierarchical models

Bayesian paradigm:

$$\underbrace{\text{Prior Pathways } G_0}_{\textit{Graphical prior}} + \underbrace{\text{Data}}_{\textit{Likelihood}} \rightarrow \underbrace{\text{Posterior Pathways } G}_{\textit{Posterior knowledge}}$$

Graph is random Allow topology to change (add or remove edges); posterior distribution on different graphs

False discovery control FDR is estimated based on posterior probabilities of graphs and edges

Prior graph Prior knowledge can be incorporated (e.g., consensus network from KEGG, GeneGO, Ingenuity...)

General structure

Bayesian paradigm:

$$\underbrace{\text{Prior Pathways } \mathcal{G}_0}_{\textit{Graphical prior}} + \underbrace{\text{Data}}_{\textit{Likelihood}} \rightarrow \underbrace{\text{Posterior Pathways } \mathcal{G}}_{\textit{Posterior knowledge}}$$

General structure

Bayesian paradigm:

$$\underbrace{\text{Prior Pathways } \mathcal{G}_0}_{\textit{Graphical prior}} + \underbrace{\text{Data}}_{\textit{Likelihood}} \rightarrow \underbrace{\text{Posterior Pathways } \mathcal{G}}_{\textit{Posterior knowledge}}$$

Notation:

$\boldsymbol{Y}$: observed data y_{gt}, feature g, sample t

$\boldsymbol{e}$: latent indicators $e_{gt} \in \{-1, 0, 1\}$ for under-, over- and normal expression

$\mathcal{G}$: Graph – dependence structure (conditional independence)

$\mathbf{c}$: strength of dependence

Probability Model – 1. Priors on random graph $p(\mathcal{G})$

Let $G = (V, E)$ denote a graph

V : set of nodes in the graph (features)

E : set of edges between pairs of nodes (edges between features)

Probability Model – 1. Priors on random graph $p(\mathcal{G})$

Let $G = (V, E)$ denote a graph

V : set of nodes in the graph (features)

E : set of edges between pairs of nodes (edges between features)

Prior on G

- Informative prior around G_0 (consensus protein network): $p(G) \propto \tau^{d(G,G_0)}$
  - Can deal with a graph with moderate size (say, 50 nodes)
  - Need to have strong prior belief in G_0
  - Example: Cellular protein signaling pathways (Telesca et al., 2012); multi-platform molecular interation map – Zodiac (Zhu et al., 2015)

Probability Model – 1. Priors on random graph $p(\mathcal{G})$

Let $G = (V, E)$ denote a graph

V : set of nodes in the graph (features)

E : set of edges between pairs of nodes (edges between features)

Prior on G

- Informative prior around G_0 (consensus protein network): $p(G) \propto \tau^{d(G,G_0)}$
  - Can deal with a graph with moderate size (say, 50 nodes)
  - Need to have strong prior belief in G_0
  - Example: Cellular protein signaling pathways (Telesca et al., 2012); multi-platform molecular interation map – Zodiac (Zhu et al., 2015)
- Vague prior when a prior network is not known: $p(\mathcal{G}) \propto const$
  - Feasible only for graphs with relatively small size (e.g., 15 nodes), see Dobra et al. (2005)
  - For histone modifications, little prior knowledge is known about their dependence (Mitra et al. 2013)

Probability Model – 2. Joint prior of features presence given the graph $p(\boldsymbol{e} \mid \boldsymbol{\beta}, \mathcal{G})$

Presence of features : Define $\{e_{it} = 1\}$ the presence indicator of feature i in location t.

Joint distribution of $\boldsymbol{e}$ given G and $\boldsymbol{\beta}$ is defined as $p(\boldsymbol{e} \mid \boldsymbol{\beta}, \mathcal{G})$.

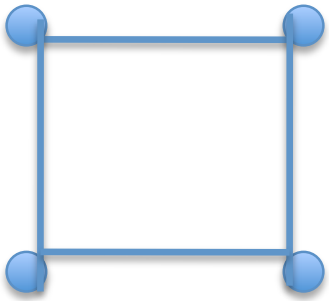
Besag (1974) shows that any joint $p(\boldsymbol{e} \mid \boldsymbol{\beta}, G)$ can be written as

$$p(\boldsymbol{e} \mid \boldsymbol{\beta}, G) = p(\mathbf{0} \mid \boldsymbol{\beta}, G) \times \exp\left\{ \sum_i \beta_i e_i + \sum_{i<j} \beta_{ij} e_i e_j + \sum_{i<j<k} \beta_{ijk} e_i e_j e_k + \ldots + \beta_{1\cdots m} e_1 \cdots e_m \right\}, \tag{1}$$

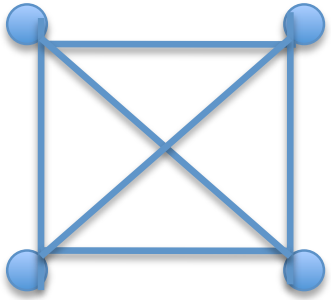
where $\beta_{i_1 \cdots i_k}$ is zero if and only if nodes $i_1, \ldots, i_k$ do not form a clique in the graph G.

Clique

A clique is a set of nodes of which all pairs in the set are connected.



Not a Clique



A Clique

Probability Model – 3. Sampling model $p(\boldsymbol{y} \mid \boldsymbol{e})$

We model y_{it} as random variable from a mixture distribution of Poisson and Log-normals.

$$p(y_{it} \mid e_{it}) \propto \begin{cases} \mathrm{Poi}(\lambda_i)\, I(y_{it} < c_i) & e_{it} = 0 \\ \pi_i \mathrm{LN}(\mu_{1i}, \sigma_{1i}^2) + (1 - \pi_i)\mathrm{LN}(\mu_{2i}, \sigma_{2i}^2) & e_{it} = 1 \end{cases} \tag{2}$$

Probability Model – 3. Sampling model $p(\boldsymbol{y} \mid \boldsymbol{e})$

We model y_{it} as random variable from a mixture distribution of Poisson and Log-normals.

$$p(y_{it} \mid e_{it}) \propto \begin{cases} \mathrm{Poi}(\lambda_i)\, I(y_{it} < c_i) & e_{it} = 0 \\ \pi_i \mathrm{LN}(\mu_{1i}, \sigma_{1i}^2) + (1 - \pi_i)\mathrm{LN}(\mu_{2i}, \sigma_{2i}^2) & e_{it} = 1 \end{cases} \tag{2}$$

The Poisson/log-normal mixture can be further replaced by introducing a trinary indicator $z_{it} \in \{-1, 0, 1\}$ with
$p(z_{it} \mid e_{it} = 0) = \delta_{-1}(z_{it})$ and
$p(z_{it} \mid e_{it} = 1) = \pi_i \delta_0(z_{it}) + (1 - \pi_i)\delta_1(z_{it})$. Then

$$p(y_{it} \mid e_{it}) = \begin{cases} \mathrm{Poi}(\lambda_i)\, I(y_{it} < c_i) & z_{it} = -1 \\ \mathrm{LN}(\mu_{1i}, \sigma_{1i}^2) & z_{it} = 0 \\ \mathrm{LN}(\mu_{2i}, \sigma_{2i}^2) & z_{it} = 1 \end{cases} \tag{3}$$

|

A fit of the mixture model (ChIP-Seq, Riten et al., 2013)

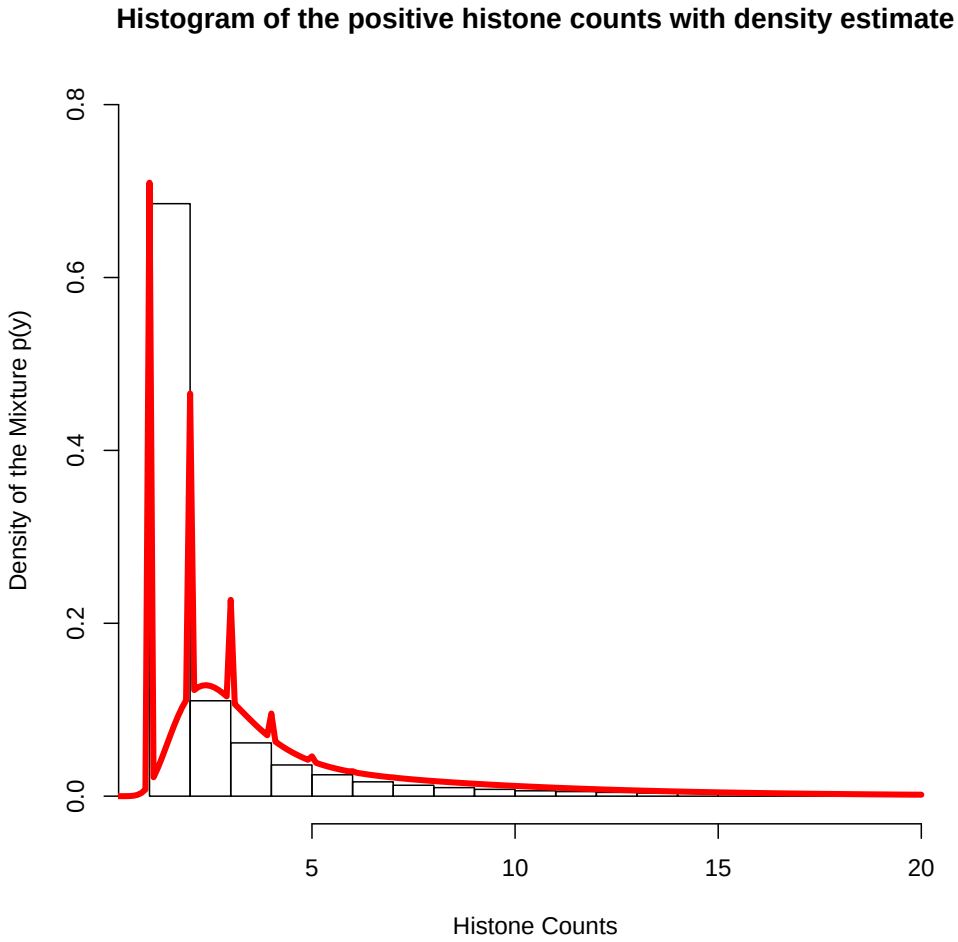


Figure: Fit of a Poisson/lognormal mixture model to the count data of a feature. The red (peaked) curve is the density of $0.5 \times Pois(1)I(y_{it} < 2) + 0.3 \times LN(1, 0.4) + 0.2 \times LN(2, 0.6)$. The

Joint Posterior

Let $\boldsymbol{\theta}$ be the parameter vector for the sampling model.

The joint posterior is given by

$$p(\boldsymbol{Y}, \boldsymbol{z}, \boldsymbol{e}, \boldsymbol{\theta}, G) \propto \underbrace{p(\boldsymbol{Y} \mid \boldsymbol{z}, \boldsymbol{\theta})}_{(3)}\, p(\boldsymbol{z} \mid \boldsymbol{e}, \boldsymbol{\theta})\, \underbrace{p(\boldsymbol{e} \mid \boldsymbol{\beta}, G)}_{(1)}\, p(\boldsymbol{\theta})\, p(\boldsymbol{\beta} \mid G)\, p(G) \tag{4}$$

MCMC and posterior inference

Posterior MCMC simulation proceeds by iterating over the following transition probabilities:

$$[\boldsymbol{e} \mid G, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{Y}],\ [\boldsymbol{z} \mid \boldsymbol{e}, \boldsymbol{\theta}, \boldsymbol{Y}],\ [\boldsymbol{\theta} \mid \boldsymbol{z}, \boldsymbol{Y}],\ [\boldsymbol{\beta} \mid \boldsymbol{e}, G],\ [G \mid \boldsymbol{\beta}, \boldsymbol{e}]$$

MCMC and posterior inference

Posterior MCMC simulation proceeds by iterating over the following transition probabilities:

$$[\boldsymbol{e} \mid G, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{Y}],\ [\boldsymbol{z} \mid \boldsymbol{e}, \boldsymbol{\theta}, \boldsymbol{Y}],\ [\boldsymbol{\theta} \mid \boldsymbol{z}, \boldsymbol{Y}],\ [\boldsymbol{\beta} \mid \boldsymbol{e}, G],\ [G \mid \boldsymbol{\beta}, \boldsymbol{e}]$$

- Updating $\boldsymbol{\beta}$ and G involves evaluating

$$c(\boldsymbol{\beta}, G) = 1/p(\mathbf{0} \mid \boldsymbol{\beta}, G) = \sum_{\boldsymbol{e}} \exp\left\{ \sum_i \beta_i e_i + \sum_{i<j} \beta_{ij}(e_i - \nu_i)(e_j - \nu_j) \right\} \tag{5}$$

MCMC and posterior inference

Posterior MCMC simulation proceeds by iterating over the following transition probabilities:

$$[\boldsymbol{e} \mid G, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{Y}],\ [\boldsymbol{z} \mid \boldsymbol{e}, \boldsymbol{\theta}, \boldsymbol{Y}],\ [\boldsymbol{\theta} \mid \boldsymbol{z}, \boldsymbol{Y}],\ [\boldsymbol{\beta} \mid \boldsymbol{e}, G],\ [G \mid \boldsymbol{\beta}, \boldsymbol{e}]$$

- Updating $\boldsymbol{\beta}$ and G involves evaluating

$$c(\boldsymbol{\beta}, G) = 1/p(\mathbf{0} \mid \boldsymbol{\beta}, G) = \sum_{\boldsymbol{e}} \exp\left\{\sum_i \beta_i e_i + \sum_{i<j} \beta_{ij}(e_i - \nu_i)(e_j - \nu_j)\right\} \tag{5}$$

step 1 Importance sampling to updated $\boldsymbol{\beta}$ (Chen and Shao, 1997; Che, Shao and Ibrahim, 2000)

- Approximate the M-H ratio by importance sampling

step 2 With step 1 and reversible jump, updating G.

CHIP-Seq Example

ChIP-Seq experiment for CD4 T Lymphocytes (Barski et al, 2007; Wang et al., 2008)

HM count data $[y_{it}]$ with 50,000 selected locations and 39 types of HMs.

Posterior inference is based on $\hat{P}_{ij}$, the posterior probability of including an edge $\{i, j\}$.

1. **Edge selection** is based on posterior expected FDR to determine a cutoff c

$$FDR_c = \frac{\sum_{i\,j} [(1 - \hat{P}_{ij}) I(\hat{P}_{ij} > c)]}{\sum_{i,j} I(\hat{P}_{ij} > c)},$$

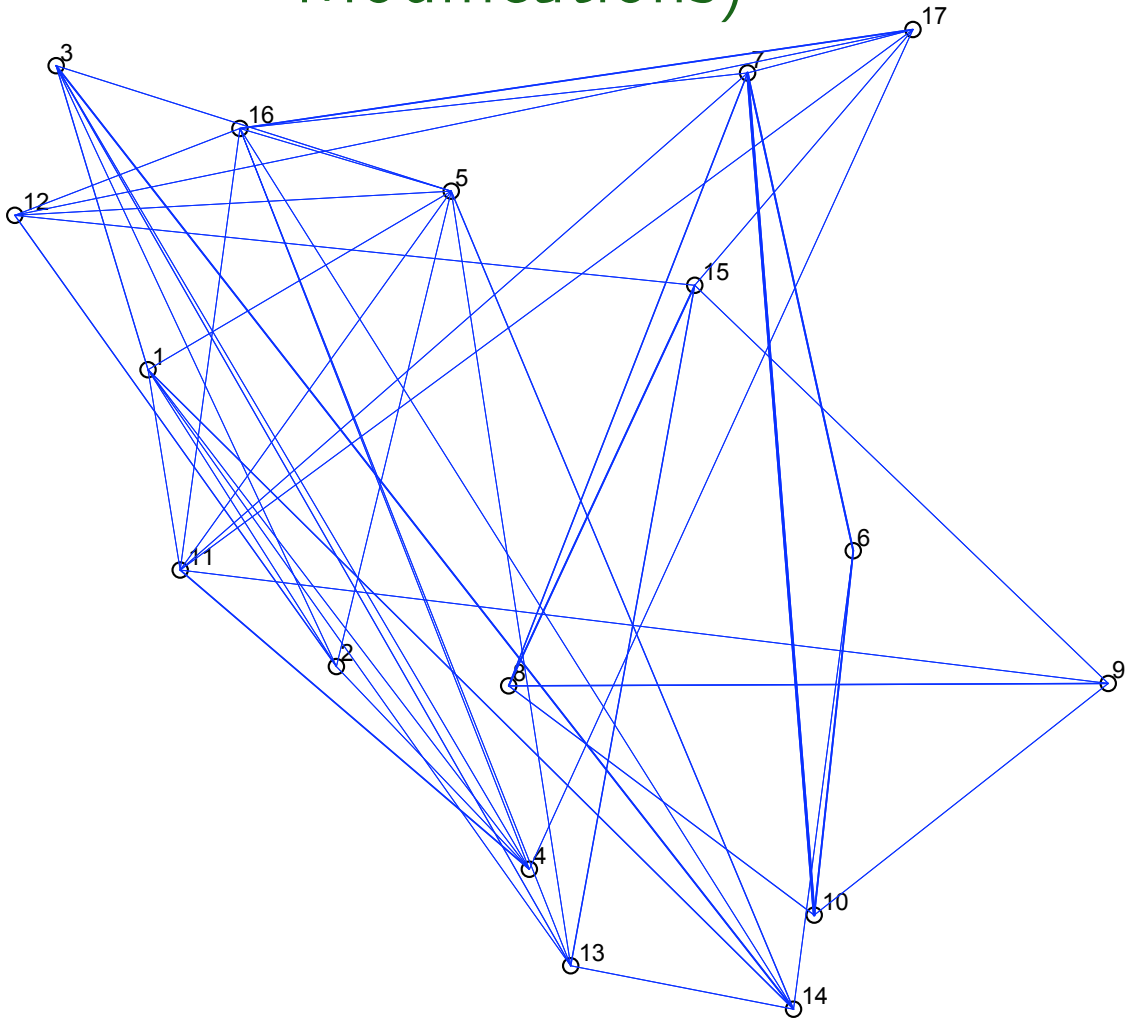
so that edges with $\hat{P}_{ij} > c$ are selected.

2. **Type of interaction** is based on

$$Pr(\beta_{ij} > 0 \mid \beta_{ij} \neq 0, \boldsymbol{y}) > 0.5$$

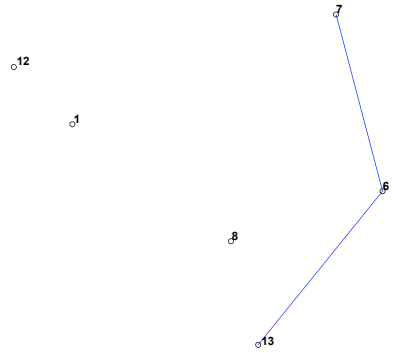
- Yes: positive
- No: negative

Results – 1: Point Estimate (ChIP-Seq on Histone Modifications)

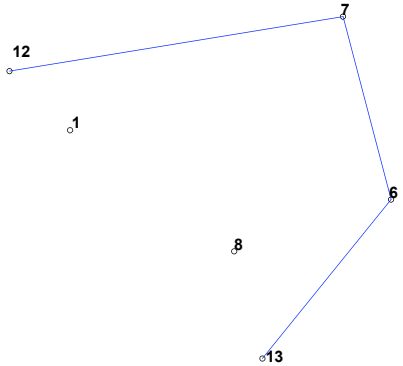


Posterior inference for the ChIP-Seq data on 17 HMs under a uniform prior $p(G)$. The thickness of the edges indicate the strength of the relationship and is a function of the posterior inclusion probabilities $\hat{P}_{ij}$.

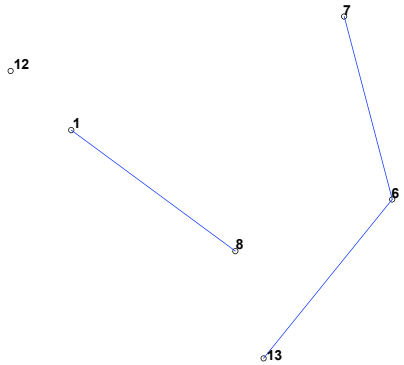
Results – 2: Variability Estimate



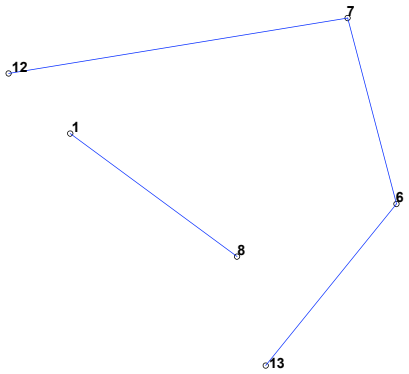
(a) 55



(b) 15



(c) 13



(d) 12

The four most frequent configurations (a through d) of a subgraph consisting of 4 edges. The posterior probabilities (in percent) are given below each subgraph.

DIFFERENTIAL GRAPHs (> 2 graphs)

(Mitra, Müller, Ji, 2014a; 2014b)

Differential Networks of

Assume an informative prior graph G_0. Inference on two graphs G^1 and G^2. Define $\delta_{ij} = |G^2_{ij} - G^1_{ij}|$ the differential edge indicator.

$$\begin{aligned} G^1 \mid G_0 &\sim U(G_0) \\ \delta_{ij} &\sim \text{Ber}(\pi), \;\; i < j \\ \pi &\sim Beta(a, b). \end{aligned} \tag{6}$$

Together G^1 and δ implicitly define G^2 by

$$G^2_{ij} = G^1_{ij}(1 - \delta_{ij}) + (1 - G^1_{ij})\delta_{ij}$$

for all edges $\{i, j\} \in E_0$.
We refer to (6) as the differential graph model, and refer to π as the global probability of similarity.

Differential Networks of

Assume an informative prior graph G_0. Inference on two graphs G^1 and G^2. Define $\delta_{ij} = |G^2_{ij} - G^1_{ij}|$ the differential edge indicator.

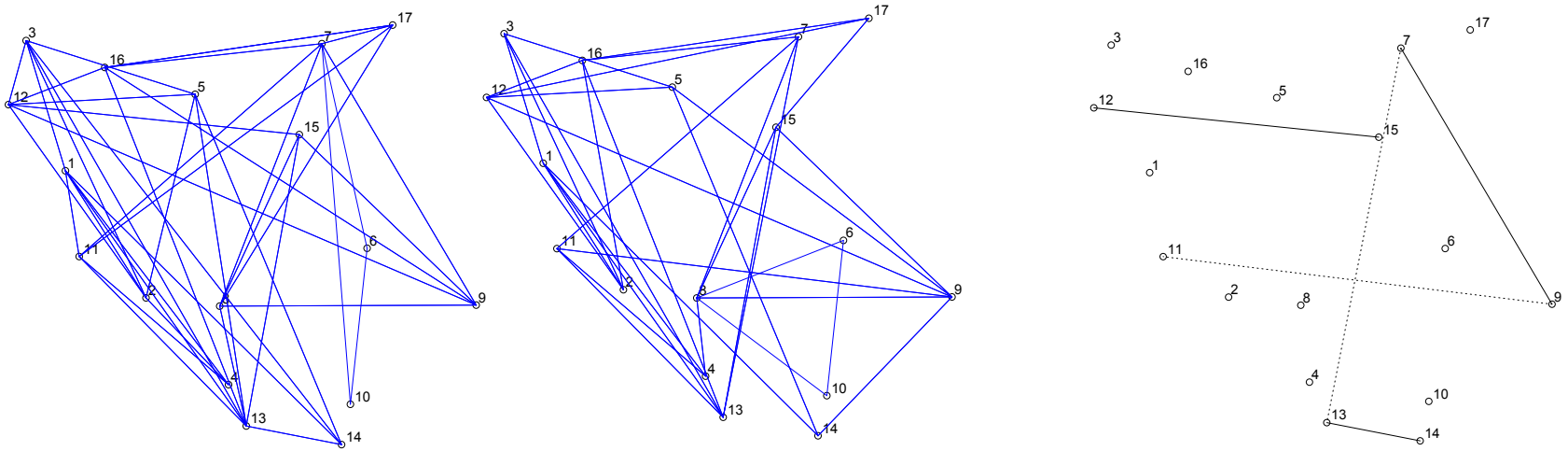
$$\begin{aligned} G^1 \mid G_0 &\sim U(G_0) \\ \delta_{ij} &\sim \text{Ber}(\pi), \;\; i < j \\ \pi &\sim Beta(a, b). \end{aligned} \tag{6}$$

Together G^1 and δ implicitly define G^2 by

$$G^2_{ij} = G^1_{ij}(1 - \delta_{ij}) + (1 - G^1_{ij})\delta_{ij}$$

for all edges $\{i, j\} \in E_0$.
We refer to (6) as the differential graph model, and refer to π as the global probability of similarity.

Differential graphs



(a) Promoters (G^1) (b) Insulators (G^2) (c) Differences $\delta_{ij} = |G^1_{ij} - G^2_{ij}|$

Figure: Panels (a) through (c) show posterior estimated networks in two regulatory regions and the posterior estimated differences between them. The solid lines denote the edges present in promoters, but not in insulators while dotted lines represent edges in insulators but not in promoters.

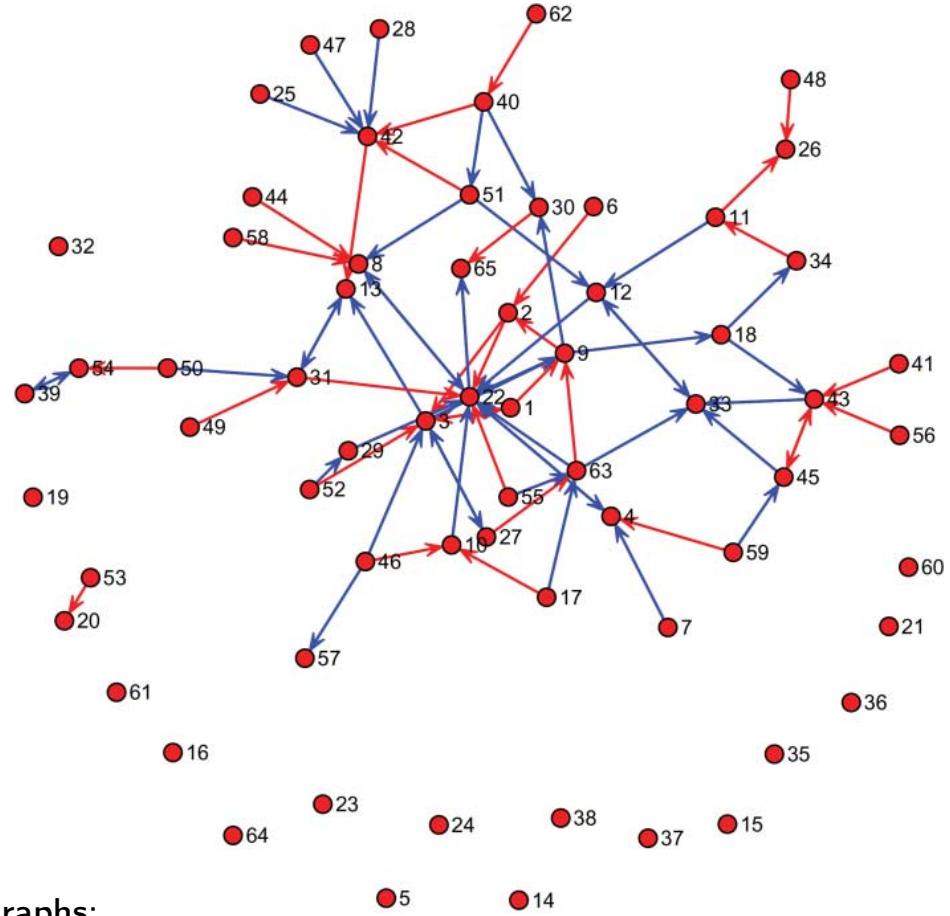
Extension to > 2 graphs

- A latent "baseline" graph G_0;
- Multiple graph model: For graph G^k, $k = 1, 2, ...K$,

$$p(G_{ij}^k = 1 \mid G_{ij}^0 = 1) = p_{11}^k \quad \text{and} \quad p(G_{ij}^k = 1 \mid G_{ij}^0 = 0) = p_{10}^k$$
$$p_{11}^k, p_{10}^k \sim Beta(a_1, b_1)$$
$$p(G_{ij}^0 = 1) = p_0; \quad p_0 \sim Beta(a_0, b_0)$$

Extension to Time-Course Proteomics Data

In Mitra et al. (2014), we consider a time-course data set from a functional proteomics experiment. About 66 proteins from PI3K pathway are measured over 8 time points. We consider a directed graph to estimate the joint dependence structure of these biomarkers.

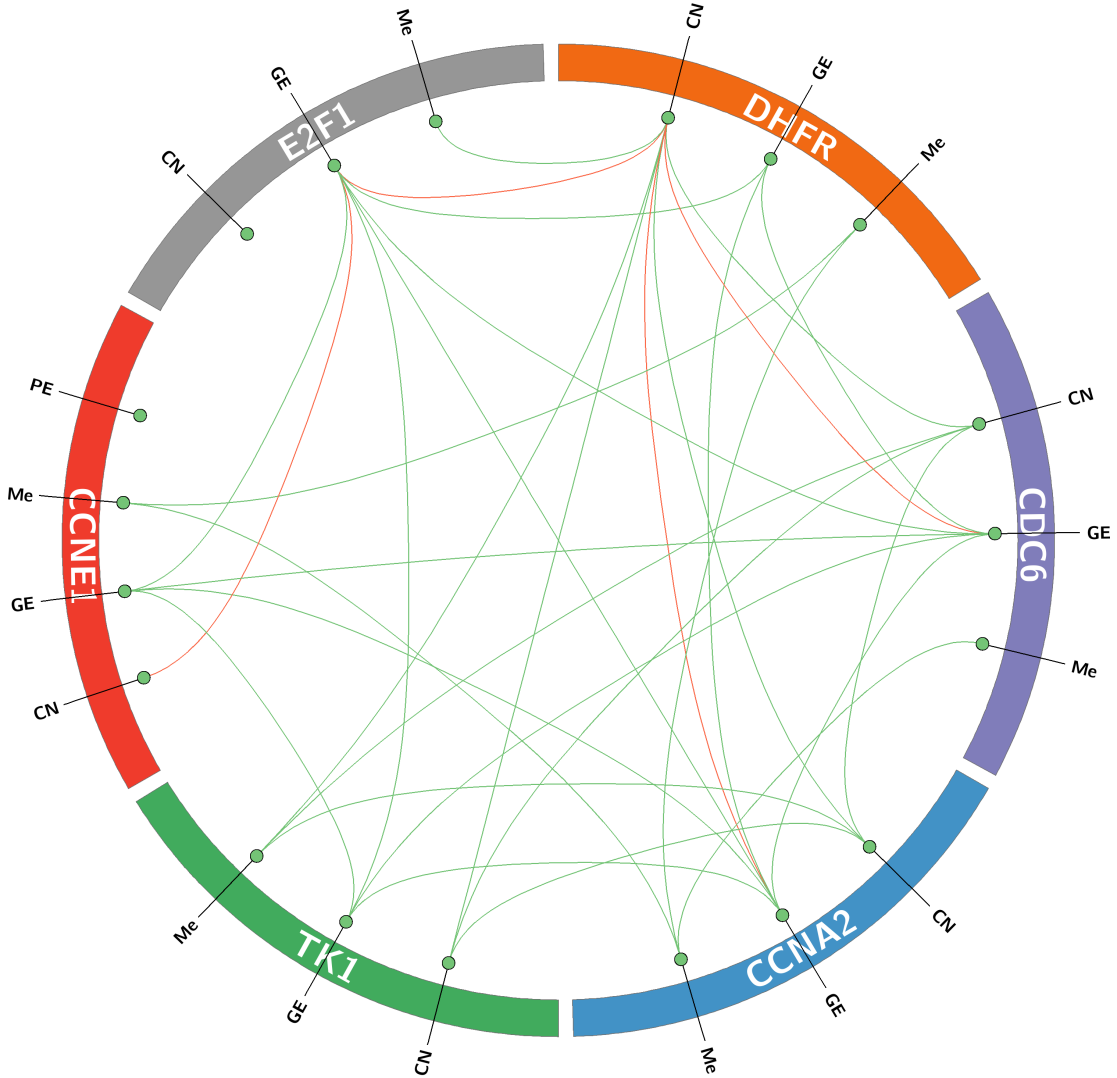


$200,000,000$ GRAPHS

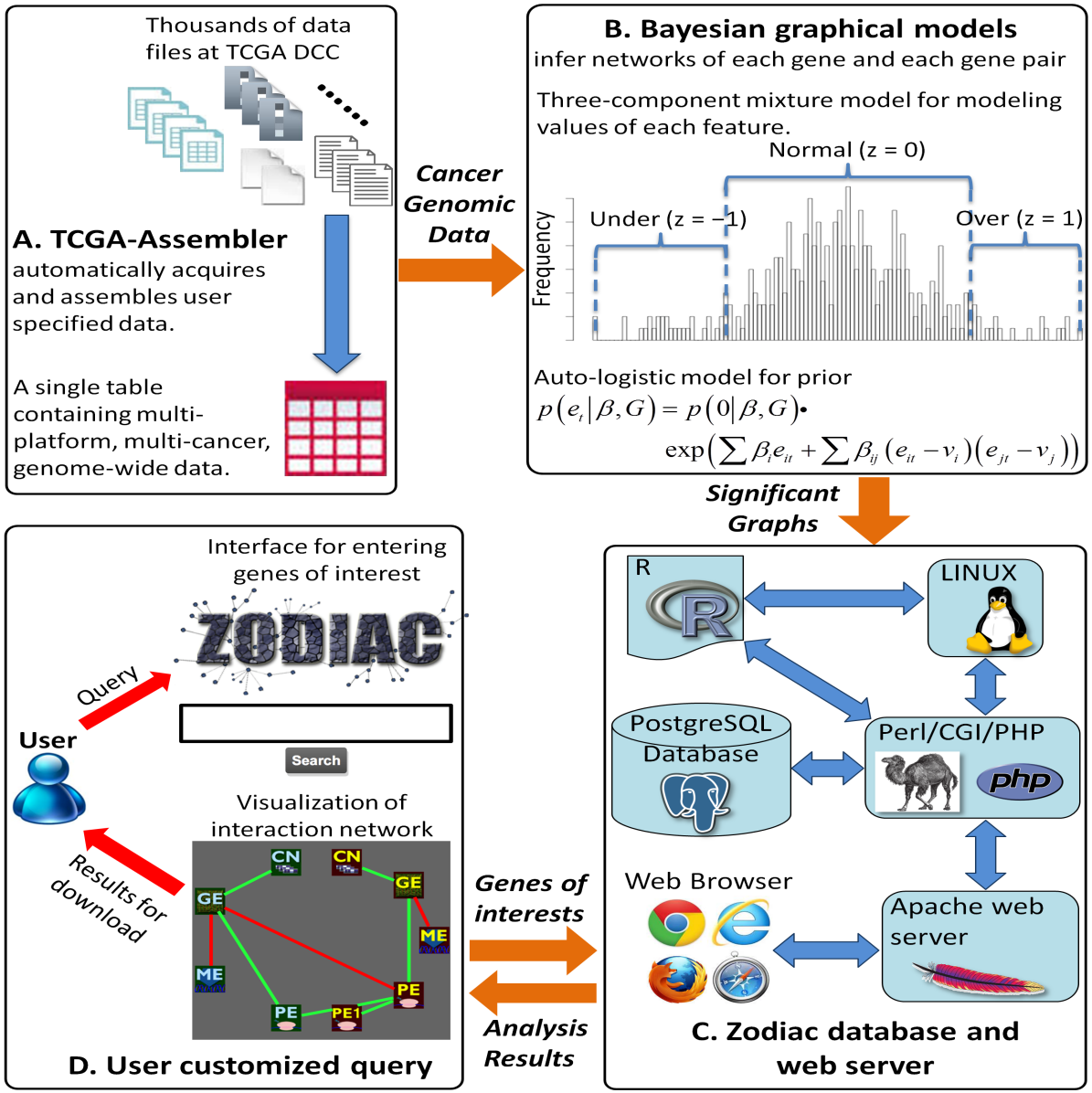
(Zhu et al., 2014; 2015)

Biological goal

Understand genetic interactions in cancer between different genomics features of different genes



Zodiac: Blueprint



The Cancer Genome Atlas (TCGA)

- An NCI/NHGRI pilot project (`cancergenome.nih.gov`), cost about $ 1 billion
- multiple cancer types (>25),
- Multiple -omics (copy number, mRNA, methylation, protein), whole genome, MATCHED samples!

Data Levels in TCGA

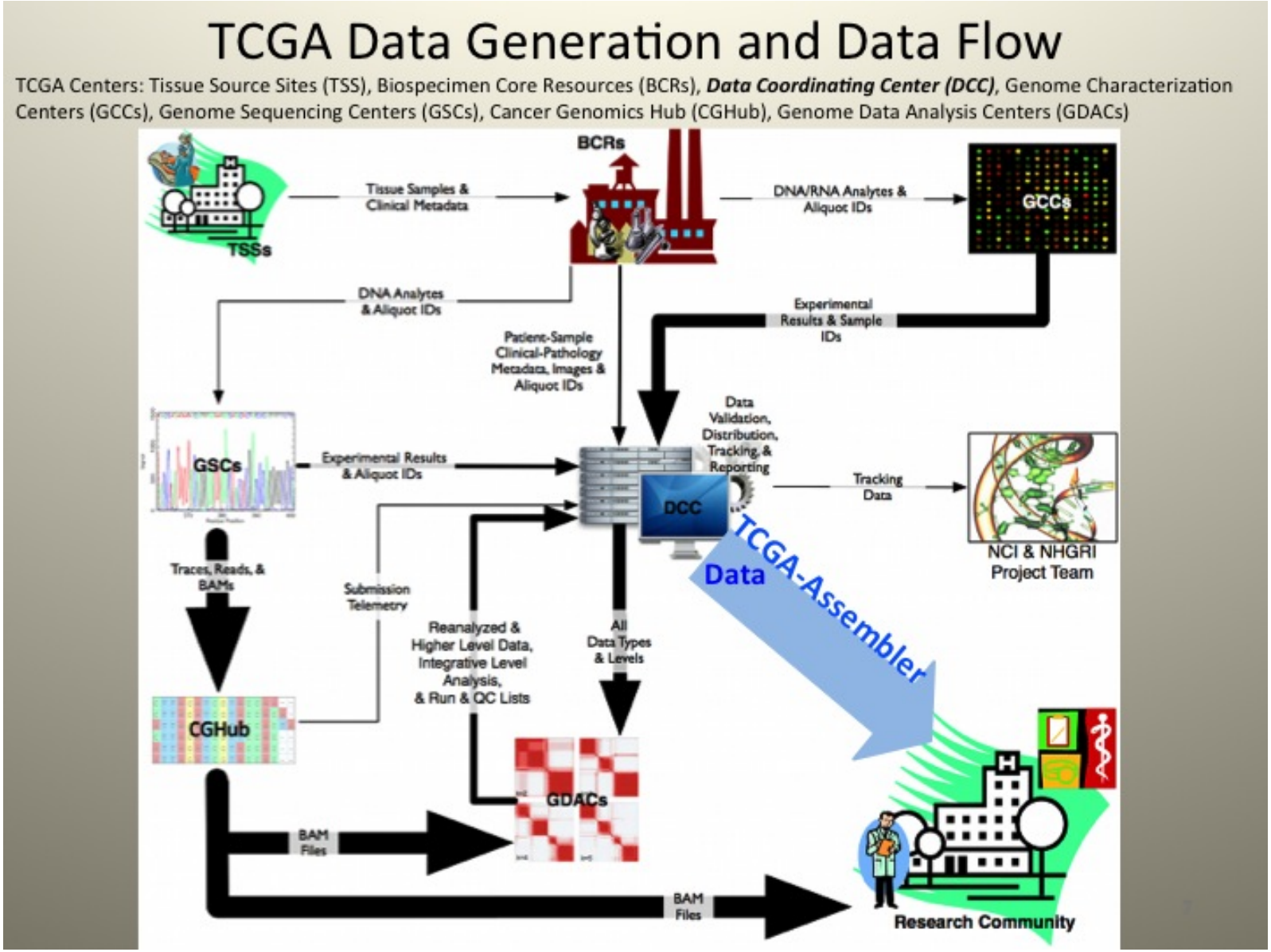
Restricted access | Publicly available data

Data Type	Level 1 (Raw Data)	Level 2 (Normalized/Processed)	Level 3 (Segmented/Interpreted)
Copy Number (CGH array)	Raw signals per probe	Normalized signals for copy number alterations of aggregated regions, per probe or probe set	Copy number alterations for aggregated/segmented regions, per sample
DNA Methylation	Raw signals per probe	Normalized signals per probe or probe set and allele calls	Methylated sites/genes per sample
Exon Expression	Raw signals per probe	Normalized signals per probe set	Expression calls for Exons/ Variants per sample
Gene Expression	Raw signals per probe	Normalized signals per probe or probe set	Expression calls for Genes per sample
miRNA Expression	Raw signals per probe	Normalized signals per probe or probe set	Expression calls for microRNAs per sample
Mutations	NA	Putative mutations	Validated somatic mutations
SNP	Raw signals per probe	Normalized signals per probe or probe set	NA

Data Size

300 TB 100 GB

TCGA-Assembler Retrieves Level-3 TCGA data



TCGA-Assembler Produces Mega-Data

Illustration of Combining Multi-modal Data for Integrative Analysis

Gene expression data file

Protein expression data file

miRNA expression data file

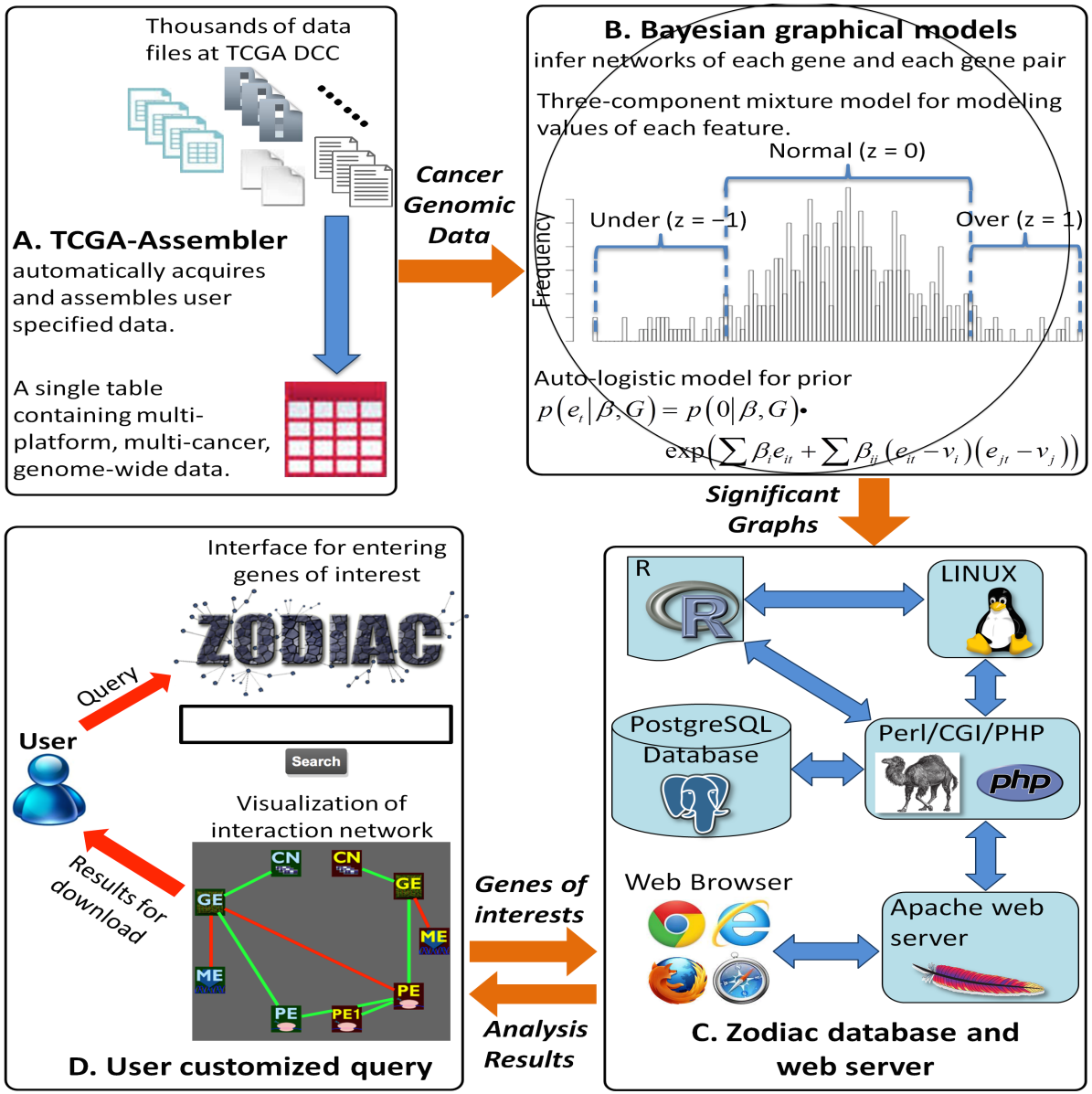
DNA copy number data file

DNA methylation data file

Single data table

Gene Symbol	Platform	Description	TCGA-EI-6506-01	TCGA-AG-4021-01	TCGA-AG-4022-01	TCGA-AG-3725-01
AKT1	GE	207	3109.227	4118.632	2905.794	4008.446
AKT1	PE	Akt-R-V	1.7805	2.0518	1.3533	2.0111
AKT1	PE	Akt_pS473-R-V	-1.621	-3.1844	-1.6175	-1.9758
AKT1	PE	Akt_pT308-R-V	-1.3476	-1.8019	-1.4822	-1.2898
AKT1	ME	Overall	0.720284	0.688232	0.680361	0.662689
AKT1	CN	CHR14-	-0.38	0.1423	-0.1192	-0.002
MIR200C	ME	Overall	0.189436	0.223844	0.183301	0.116829
MIR200C	CN	CHR12+	0.0079	-0.6209	0.1662	-0.0034
MIR200C	miRExp		16617.82	5761.941	11792.5	26984.18
MIR506	ME	Overall	0.771979	0.757992	0.700243	0.671736
MIR506	CN	CHRX-	0.0057	-0.1969	0.0017	0.0175
MIR506	miRExp		0.277389	1.212507	0.115049	0.06591
MTOR	GE	2475	1520.826	1496.095	1007.077	1298.564
MTOR	PE	mTOR-R-V	1.1394	1.0414	0.82713	1.2374
MTOR	PE	mTOR_pS2448-R-C	-1.9719	-2.3493	-1.9848	-1.8108
MTOR	ME	Overall	0.567012	0.585587	0.555973	0.549771
MTOR	CN	CHR1-	-0.1671	0.1284	-0.1071	-0.0109
PACS2	GE	23241	1141.753	1489.029	1041.575	1304.476
PACS2	ME	Overall	0.72097	0.702261	0.708845	0.695105
PACS2	CN	CHR14+	-0.38	0.1423	-0.1192	-0.002
TP53	GE	7157	3783.318	2123.094	2564.794	2444.257
TP53	ME	Overall	0.224788	0.233938	0.223865	0.227782
TP53	PE	p53-R-V	-2.059	-2.8108	-2.0793	-2.2214
TP53	CN	CHR17-	0.0047	-0.6397	-0.1182	-0.4899

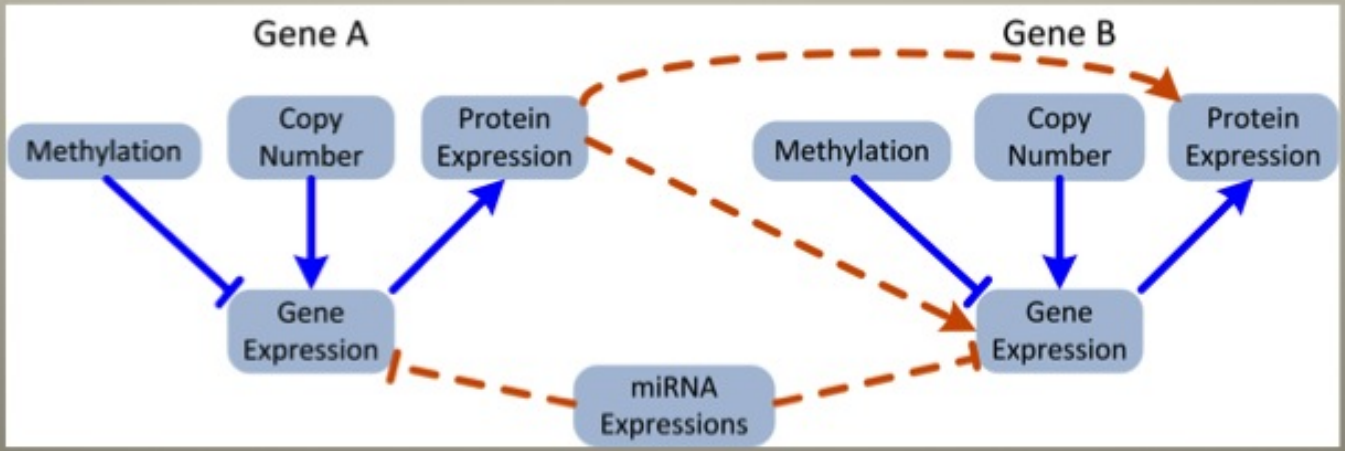
Bayesian Graphical Models



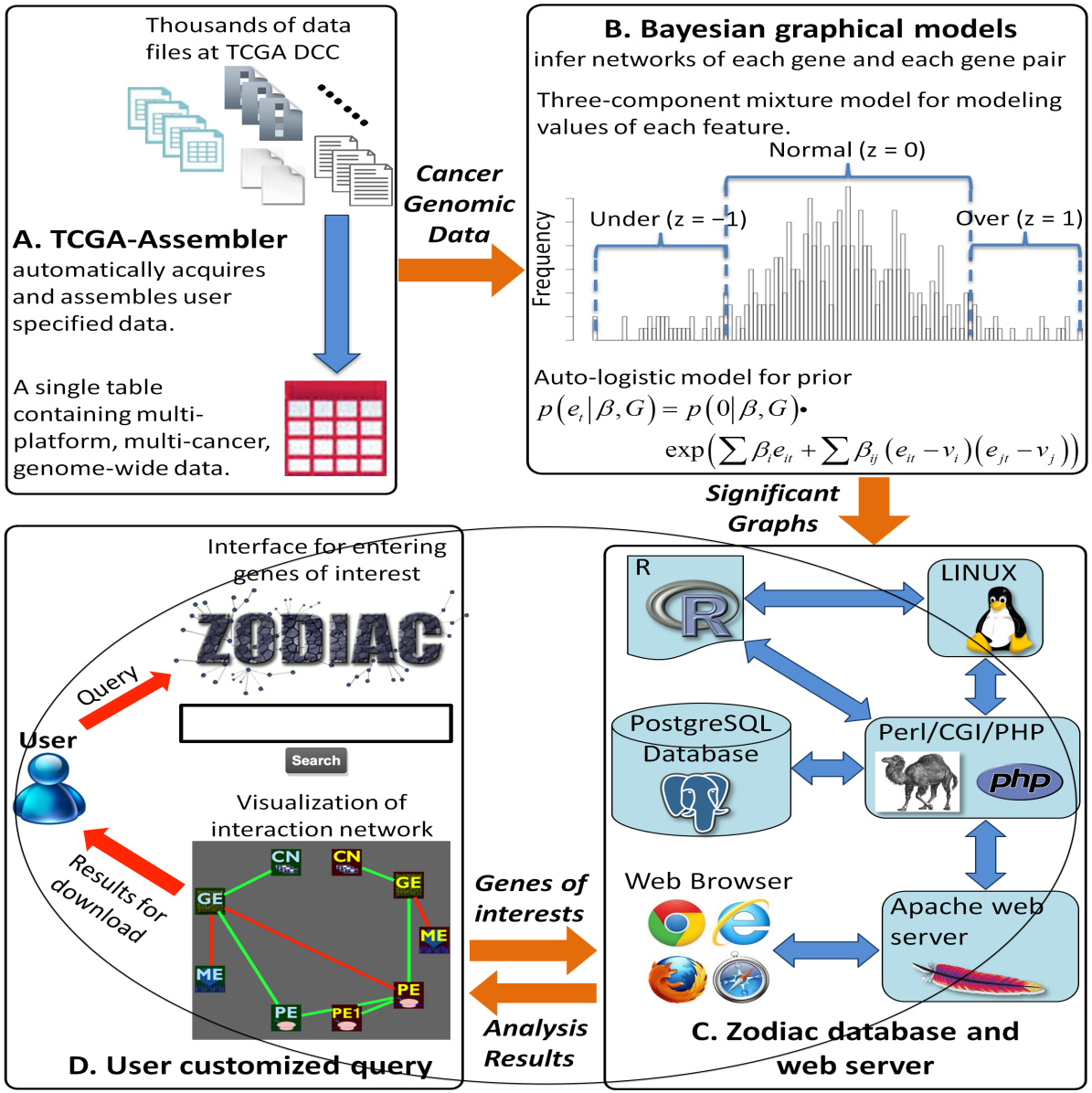
Multi-omics Molecular Interaction Map

Inference of Intragenic and Intergenic Interactions

- Integrate data from multiple genomic/epigenomic/proteomic assay platforms to infer interaction mechanisms.
  - Within and across cancer types
- Intragenic interactions of each gene (~20,000 genes).
- Intergenic interactions between each pair of genes (~200,000,000 pairs).



Big-Data Computation and Visualization



Massive Parallel Computation

- Analysis of one gene pair takes ~47 seconds.
- Total required computation time is ~2,459,455 CPU hours.
- Analysis was conducted on **Beagle**, a super computer with > 17000 CPUs in University of Chicago and Argonne National Laboratory.
- Size of analysis results (~800 GB)
  - 19,411 intragenic interaction networks
  - ~200 million intergenic interaction networks

Overlap with Existing Databases of Genomic Regulations

KEGG pathways used for validation of inferred interactions

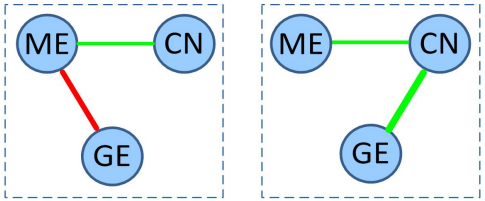
Cancers Overview	Pathways in cancer Transcriptional misregulation in cancer Proteoglycans in cancer
Signal Transduction	MAPK signaling pathway PI3K-Akt signaling pathway Notch signaling pathway mTOR signaling pathway Wnt signaling pathway TGF-beta signaling pathway ErbB signaling pathway VEGF signaling pathway Jak-STAT signaling pathway NF-kappa B signaling pathway
Cell Growth and Death	Cell cycle Apoptosis p53 signaling pathway

Overlaps between KEGG and Zodiac

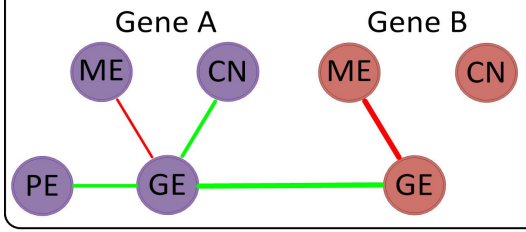
KEGG Relationship (Corresponding Zodiac relationship)	Enrichment Fold	Enrichment P-value
Gene Expression Activation (Positive PE-GE or GE-GE)	2.38	2.92E-18
Protein Phosphorylation (Positive PE-PE(phos) or GE-PE(phos))	14.17	4.93E-14
Multi-unit Protein and Protein Complex (Positive GE-GE or PE-PE)	3.10	1.29E-312

Results-1: Intra-genic transcription regulation

A-i Intragenic interactions within a gene



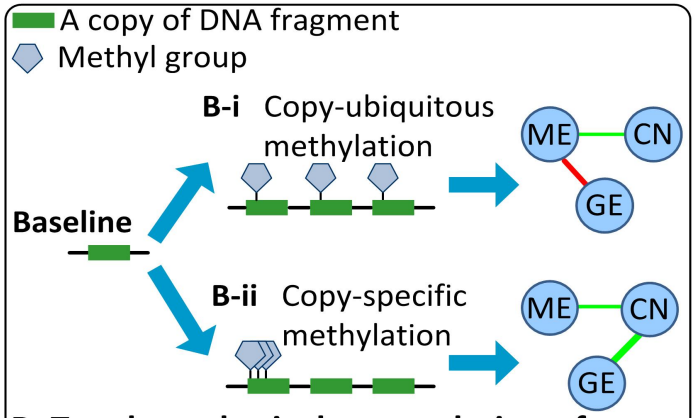
A-ii Intragenic and intergenic interactions in a pair of genes



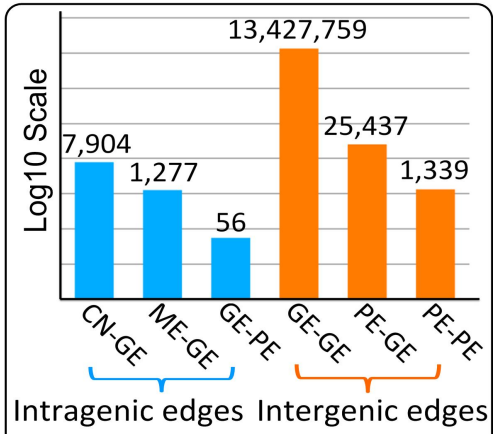
A-iii Top 10 genes having strongest, significant (Bayesian FDR <= 0.01) positive GE-GE interactions with transcription factor gene EZH2

Gene Symbol	Mean of Strength Coefficient	Gene Symbol	Mean of Strength Coefficient
HIST1H2BH	15.92	SKP2	10.49
LEFTY1	14.65	TBC1D31	10.49
KDM3A	11.16	SCLT1	10.46
MTIF2	10.93	NEK2	10.31
NEURL	10.71	FASTKD1	10.21

A. Illustration of statistical inference on intragenic and intergenic interaction networks

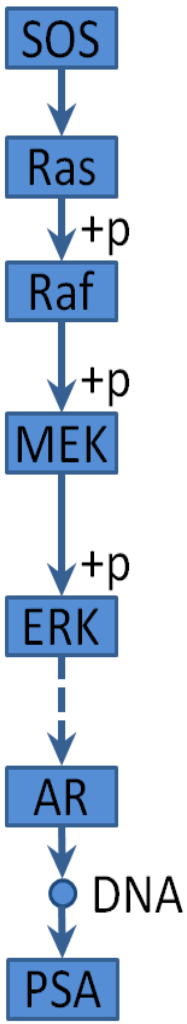


B. Two hypothetical co-regulation of DNA methylation and copy number variation on gene expression.

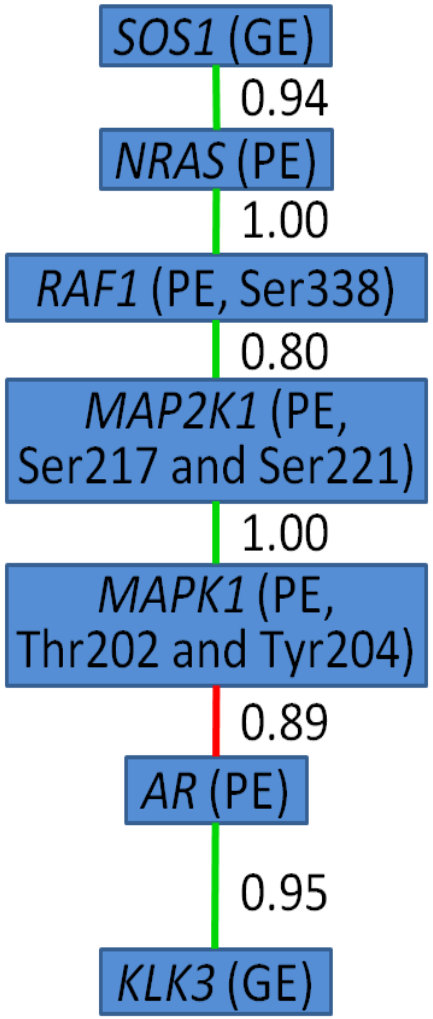


C. Numbers of significant edges inferred by analyses

Results-2: Entire Pathway



A. Signaling cascade in KEGG prostate cancer pathway



B. Posterior network inferred by BGM analysis

Results-3: Predictive markers for anti-PD-1 immune treatment

Mismatch-repair deficiency predicts response of solid tumors to PD-1 blockade

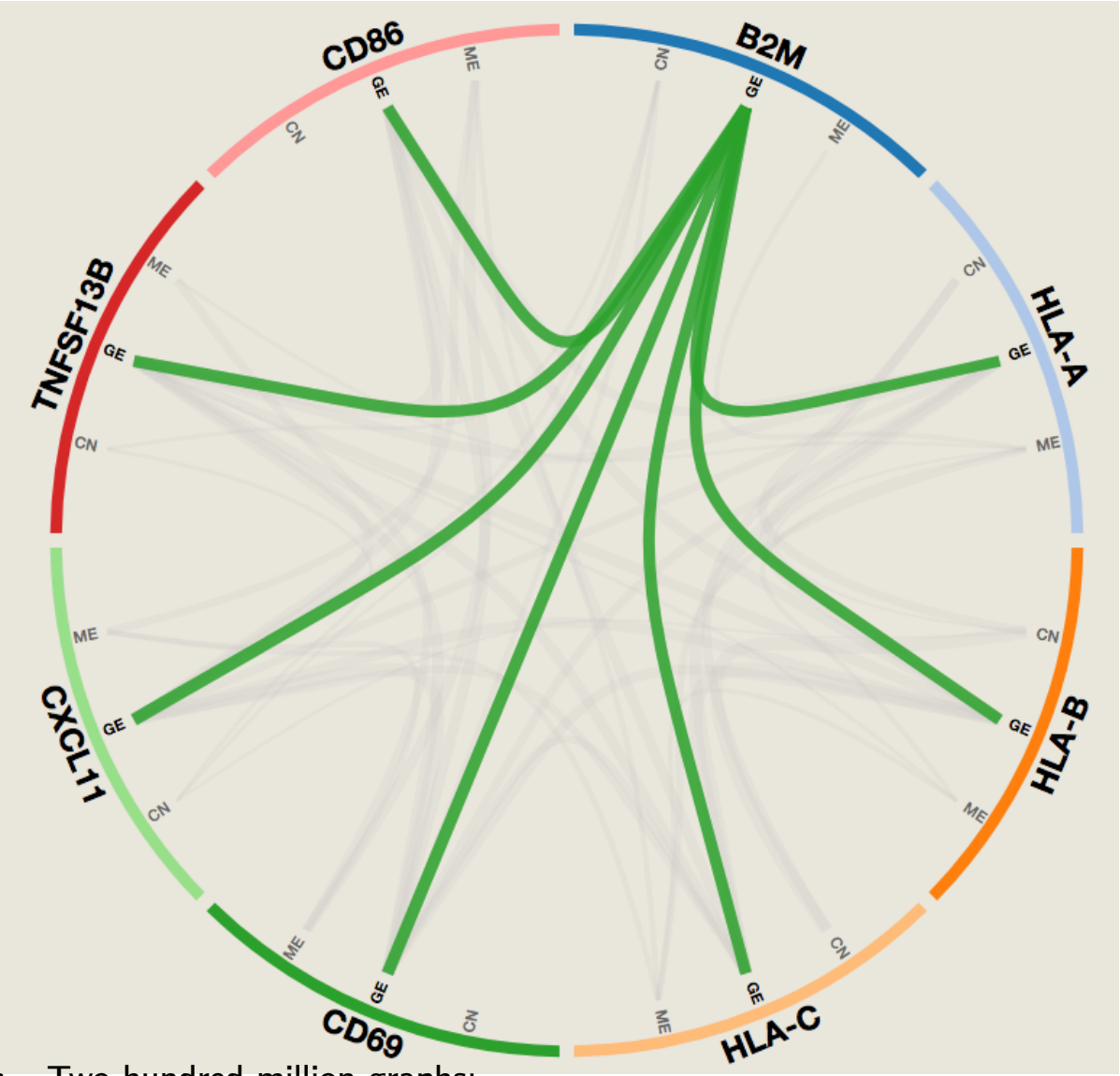
Dung T. Le[1,2,3], Jennifer N. Durham[1,2,3,*], Kellie N. Smith[1,3,*], Hao Wang[3,*], Bjarne R. Bartlett[2,4,*], Laveet K. Aulakh[2,4], St...
+ See all authors and affiliations

Science 08 Jun 2017:
eaan6733
DOI: 10.1126/science.aan6733

Peer Reviewed
← see details

Dung et al. (2017, *Science*) discussed predictive biomarkers for anti-PD-1 blockade in treating cancer patients.
B2M is a gene that predicted worse outcome when mutated

Results-3: Predictive markers for anti-PD-1 immune treatment



The HLA gene family provides instructions for making a group of related proteins known as the human leukocyte antigen (HLA) complex. The HLA complex helps the immune system distinguish the body's own proteins from proteins made by foreign invaders such as viruses and bacteria. – Cancer too?

Zodiac Website:
`http://www.compgenome.org/zodiac`

Zodiac Blog:
`http://compgenome.wordpress.com`

Thank you!

Zodiac 2 – to be continued...

- Patient subgroups defined by different pathway architecture
- Status of pathway activation for individual patient (allowing for precision therapeutic decisions)
- Update existing cancer pathways using TCGA
- Tissue-specific pathways