Integrative Bayesian Modeling Approaches to Imaging Genetics

Michele Guindani

Department of Biostatistics
MD Anderson Cancer Center
Houston, TX

CCNS: Challenges in Functional Connectivity Modeling and Analysis
Samsi, April 8-10, 2016

Collaborators

Marina Vannucci, Rice

Fabrizio Leisen
University of Kent

Erik Erhardt,
MRN & UNM

Alberto Cassese
Maastricht University

Francesco Versace,
Stephenson Cancer Center

Vince Calhoun, Professor,
MRN & UNM

Francesco Stingo,
MDACC

Sharon Chiang,
Rice,
Keck Fellowship

Duncan Wadsworth,
Rice

Qiwei Li,
Rice

Ryan Warnick,
Rice
NSF fellowship

Eric Kook,
Rice

Ronaldo Guedes
University of Padua,
NYU Postdoc (soon)

Kim-Anh Do
MDACC

Thierry Chekouo,
Postdoc MDACC

Weixuan Zhu,
Universidad Carlos III, Madrid
Postdoc University of Sheffield

Linlin Zhang,
Phd from Rice

To keep updated:
http://www.micheleguindani.info
http://www.stat.rice.edu/~marina/

Imaging Genetics

- Imaging genetics refers to situations where imaging technologies are used as “phenotypic assays” in studies on subjects carrying genetic risk variants that relate to a psychiatric disorder (Silver, Montana & Nichols, 2010, NeuroIm).

- ✌ Overall idea is that individual differences in the genetic make-up lead to differences in brain wiring structure and intellectual function.

- ☞ Modeling the link between the imaging and genetic components could indeed lead to improved diagnostics and therapeutic interventions.

- ✎ Ex: Schizophrenia, a severe psychiatric disorder disrupting normal thinking, speech, and behavior.

Data, and data processing

- Data from the Mind Clinical Imaging consortium. $n_1 = 118$ healthy controls and $n_2 = 92$ schizophrenic patients.
- fMRI data, measuring brain activity as changes in blood flow, collected during a sensorimotor task:
  1. Atlas-based parcellation of the brain into p anatomical regions (ROIs - features).
  2. Data as ROI-based summaries of BOLD signal intensities

$$x_{i,j},\ i = 1, \ldots, n,\ j = 1, \ldots, p$$

for p features (ROIs) on n subjects.

- $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iR})^T$, R genetic covariates (SNPs implicated in schizophrenia) available on all subjects.

A Discriminative integrative model

Goal: Identify brain regions with discriminating activation patterns and SNPs relevant to explain such activations in either (or both) subgroups. We propose:

- Hierarchical mixture model with selection of discriminating features (e.g. ROIs)
- The model is a mixture of K components, each describing activations in K groups (e.g. cases and controls), and each depending on selected covariates (e.g. SNP)
- Network priors that capture structural dependencies among the features.

Mixture model with feature selection

We assume a general Gaussian mixture model with K groups (e.g., schizophrenic and healthy controls).

- Data from group k modeled as

$$(\mathbf{x}_i | g_i = k, \cdot) \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),$$

  with $k = 1, \ldots, K$ and $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the group-specificic mean and covariance matrix.

- Group assignments: $\mathbf{g} = (g_1, \ldots, g_n)'$, where $g_i = k$ if the i^{th} observation comes from group k and $w_k = P(g_i = k)$.

✎ Supervised setting (discriminant analysis): K, g known ($\hat{w}_k = n_k/n$). Model-based approach to classification.

We envision that only some of the features (ROIs) discriminate the n subjects.

- Introduce $\gamma = (\gamma_1, \ldots, \gamma_p)$ such that $\gamma_j = 1$ if j-th feature is discriminatory, $\gamma_j = 0$ otherwise.

  Indicate features indexed by $\gamma_j = 1$ as $\mathbf{X}_{(\gamma)}$, and those indexed by $\gamma_j = 0$ as $\mathbf{X}_{(\gamma^c)}$.

- Model becomes

$$\begin{aligned}(\mathbf{x}_{i(\gamma)}|g_i = k, \cdot) &\sim \mathcal{N}(\boldsymbol{\mu}_{k(\gamma)}, \boldsymbol{\Sigma}_{k(\gamma)}) \\ (\mathbf{x}_{i(\gamma^c)}|\cdot) &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{(\gamma^c)}),\end{aligned}$$

  with $g_i = k$ if the i-th sample belongs to group k.

Variable selection for mixture models, Tadesse et al (2005, JASA), Raftery & Dean (2006, JASA), Stingo et al. (2012, Sinica).

Network priors

- Use Markov Random Field prior on γ, capturing spatial dependencies among ROIs (proximity)

$$P(\gamma_j|\gamma_i, i \in N_j) = \frac{\exp(\gamma_j F(\gamma_j))}{1 + \exp(F(\gamma_j))},$$

where $F(\gamma_j) = e + f \sum_{i \in N_j}(2\gamma_i - 1)$ and N_j is the set of direct neighbors of ROI j in the network.

Parameter e controls sparsity. Higher values of f induce more neighbors to assume the same values.

☞ Favors *clusters of "relevant" ROIs*.

Covariate-dependent mixture components

We want to link imaging and genetic information in the participants' subgroups.

- ☞ Allow mixture components to depend on the covariates

$$\boldsymbol{\mu}_{ik(\gamma)} = \boldsymbol{\mu}_{0k(\gamma)} + \beta^{T}_{k(\gamma)} \mathbf{Z}_i, \quad k = 1, \ldots, K,$$

  where $\boldsymbol{\mu}_{0k(\gamma)}$ is a baseline process (see later).

- ✌ Obtain component-specific parameters determining how SNPs affect brain activities, given selected ROIs.

We want to identify different covariates (SNPs) affecting the individual mixture components.

- ☞ Use spike and slab priors on $\beta_{k(\gamma)}$

$$\beta_{rk(\gamma)} \sim \delta_{rk}\mathcal{N}(\mathbf{b}_{0k(\gamma)}, h\mathbf{\Sigma}_{k(\gamma)}) + (1-\delta_{rk})\mathcal{I}_0(\beta_{rk(\gamma)}),$$

with $\delta_{rk} = 1$ if r-th covariate relevant to explain measurements in k-th group.

Assume Bernoulli priors on δ_{rk}.

Variable selection approach to linear regression models of George and McCulloch (1997, Sinica) and Brown et al. (1998,2002, JRSSB).

Spatial dependencies

- ☞ Model component-specific dependencies via distribution of $\mu_{0k(\gamma)}$ (random effect)

$$\mu_{0k(\gamma)} \sim N_{p_\gamma}(\nu_{k(\gamma)}, h_1 \Gamma_{0k(\gamma)}), \quad k = 1, \ldots, K,$$

with $\Gamma_{0k(\gamma)} \sim IW(d_k, \mathbf{Q})$ and normal prior on $\nu_{k(\gamma)}$.

- ✏ This component captures correlation among distant ROIs (functional connectivity), and it is in addition to the local dependence captured by the network prior.

- ✏ Can also estimate component-specific networks among selected ROIs as

$$\mu_{0k(\gamma)} | G_{k(\gamma)} \sim N_{p_\gamma}(\nu_{k(\gamma)}, h_1 \Gamma_{0k(\gamma)}), \quad k = 1, \ldots, K,$$

with $G_{k(\gamma)}$ the graph encoding the relationships (Dobra et al, 2011).

MCMC for posterior inference

Want to select discriminating features (via γ) and important covariates (via δ). Also, inference on the dependence structure among the selected features ($\mu_{0k(\gamma)}$).

1. Metropolis-Hastings step on γ (add /delete/swap).
2. Metropolis-Hastings step for δ_k (add/delete/swap).
3. Random walk Metropolis-Hastings step on the $\mu_{0k(\gamma)}$'s:

$$\mu_{0kj}^{New} = \mu_{0kj}^{Old} + \epsilon, \quad \epsilon \sim N(0, v^2)$$

Posterior inference via marginal posterior probabilities of inclusion. Post-MCMC estimates of variance components and regression coefficients.

Use predictive distribution to classify new samples based on the selected features and covariates.

Case study on schizophrenia

- ✎ Participant recruitment and data collection by the Mind Clinical Imaging consortium (MCIC), a collaborative effort of teams from Boston, Iowa, Minnesota and New Mexico.
- ✎ fMRI data during a sensorimotor task for $n_1 = 118$ healthy controls and $n_2 = 92$ schizophrenic patients.
- ✎ Training set of 174 participants and validation set with 36 participants (balanced scheme).
- ✎ $R = 81$ genetic covariates (SNP) available for each participant in the study (implicated in schizophrenia).
- ☞ Use our unified modeling framework to relate brain activities in subjects with different conditions to the individuals' specific genetic characteristics.

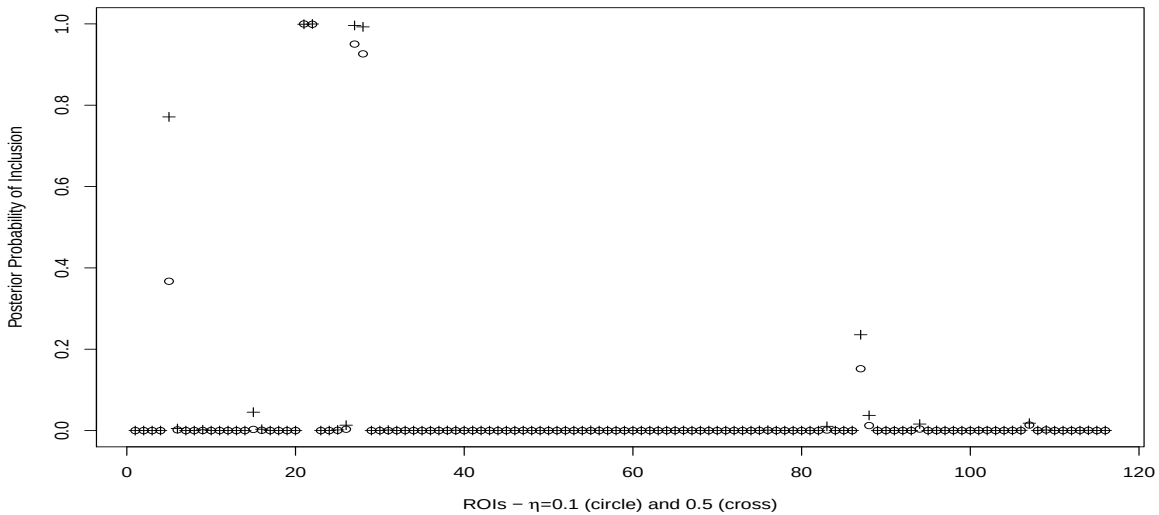
Processing of the fMRI data

- Imaging data preprocessed in SPM5, realigned, normalised, re-sliced and spatially smoothed.
- Data summarized in individual contrast images of ROI-based summary statistics:

1. Multiple regressions fit to the data from each participant, with regressors for stimulus and its temporal derivative plus intercept.
2. Resulting regression coefficients used to create contrast images –also called statistical parametric maps (Friston, 1995)– capturing the stimulus effect at each voxel.
3. Maps segmented into $p = 116$ regions of interest (ROIs) according to the MNI space Automated Anatomical Labeling (AAL) atlas and activations in each region summarised by median value for that region.

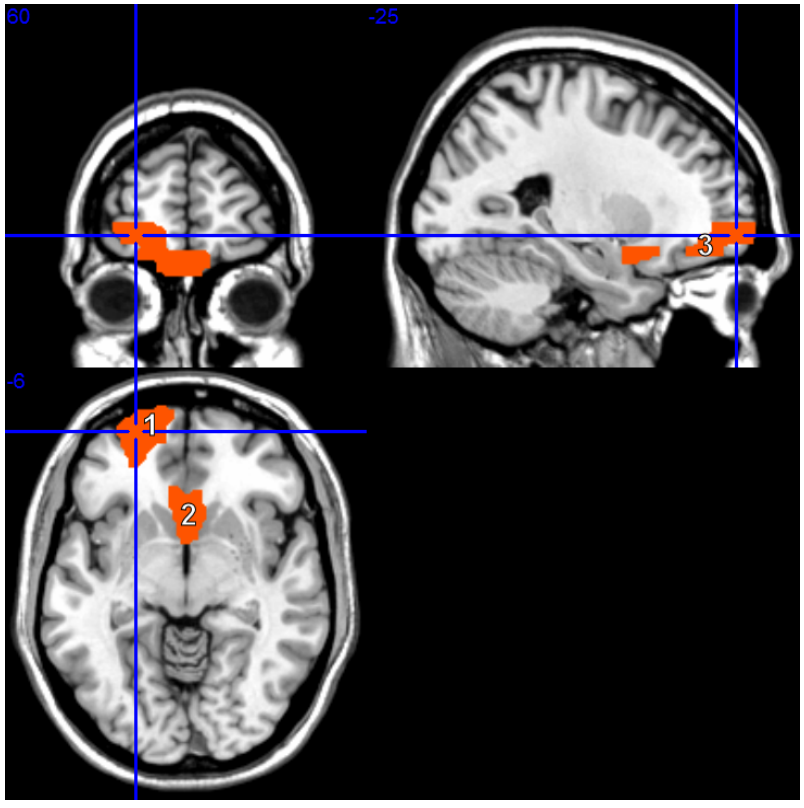
- For γ, set $e = -4$ (1% of total features, sparsity) and $f = 0.1$ and 0.5 (small to moderate neighborhood effect).
- For δ, set $w_{rk} = 0.1$ (10% of covariates).
- Vague prior specifications otherwise.
- MCMC chains with 200,000 iterations and a burn-in of 1,000 iterations.

Results: Selection of discriminating ROIs



ROI	Name	$p(\gamma_j \mid \boldsymbol{Z}, X)$ for $f = 0.1$	$p(\gamma_j \mid \boldsymbol{Z}, X)$ for $f = 0.5$
ROI 5	Frontal Sup Orb L	0.39	0.78
ROI 21	Olfactory L	1.00	1.00
ROI 22	Olfactory R	1.00	1.00
ROI 27	Rectus L	0.94	1.00
ROI 28	Rectus R	0.90	0.99

Increase in posterior prob of ROI 5 due to MRF prior, since ROI 5 is connected to ROIs 21, 27 and 28.



Orbital part of the superior frontal gyrus (ROI 5, coded as ’1’, spanning superior frontal gyrus, middle frontal gyrus, inferior frontal gyrus); olfactory cortex (ROIs 21&22, coded as ’2’, spanning subcallosal gyrus and anterior cingulate); gyrus rectus (ROIs 27&28, coded as ’3’, spanning medial frontal gyrus, rectal gyrus and superior frontal gyrus). Cross-hair identifies Brodmann area 10.

Results: Component-specific connectivity

Estimated correlation matrices for control and schizophrenic groups

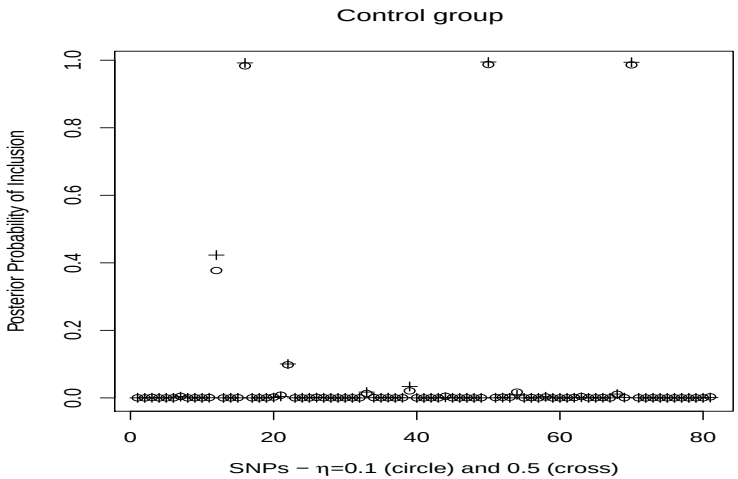
$$Corr_{\mu 01} = \begin{pmatrix} 1.0000 & 0.0149 & 0.0267 & 0.0295 & 0.0328 \\ 0.0149 & 1.0000 & 0.0246 & 0.0293 & 0.0235 \\ 0.0267 & 0.0246 & 1.0000 & 0.0373 & 0.0506 \\ 0.0295 & 0.0293 & 0.0373 & 1.0000 & 0.0539 \\ 0.0328 & 0.0235 & 0.0506 & 0.0539 & 1.0000 \end{pmatrix}$$

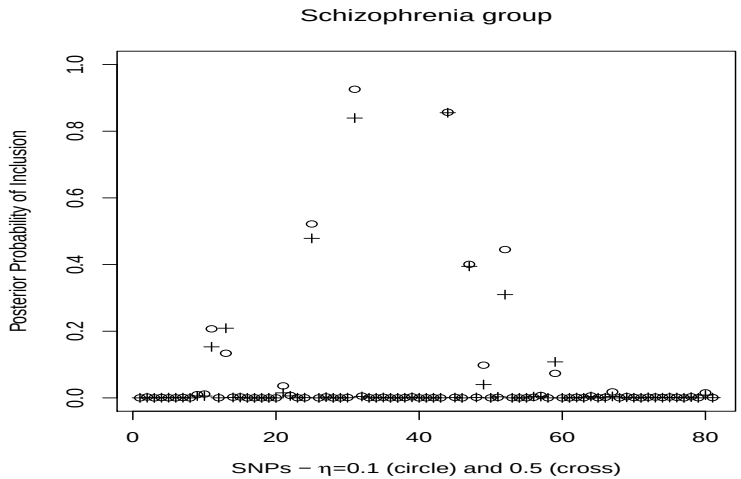
and

$$Corr_{\mu_{02}} = \begin{pmatrix} 1.0000 & 0.3532 & 0.3403 & 0.3310 & 0.3562 \\ 0.3532 & 1.0000 & 0.4509 & 0.4193 & 0.4227 \\ 0.3403 & 0.4509 & 1.0000 & 0.3617 & 0.4024 \\ 0.3310 & 0.4193 & 0.3617 & 1.0000 & 0.3818 \\ 0.3562 & 0.4227 & 0.4024 & 0.3818 & 1.0000 \end{pmatrix}$$

Finding consistent with work in fMRI, less unique brain activity in cases versus controls, supporting a generalized cognitive deficit in schizophrenic patients, Calhoun *et al.* (2006).

Results: Selection of SNPs





Schizophrenia			
SNP	Name	$p(\delta_{2l}\|\boldsymbol{Z}, X)$ for $f = 0.1$	$p(\delta_{2l}\|\boldsymbol{Z}, X)$ for $f = 0.5$
SNP 25	rs1934909	0.49	0.47
SNP 31	rs875462	0.92	0.83
SNP 44	rs17101921	0.84	0.85
Control			
SNP	Name	$p(\delta_{1l}\|\boldsymbol{Z}, X)$ for $f = 0.1$	$p(\delta_{1l}\|\boldsymbol{Z}, X)$ for $f = 0.5$
SNP 16	rs6794467	0.98	0.99
SNP 50	rs2421954	0.98	0.99
SNP 70	rs2270641	0.98	0.99

- ☞ Selected SNPs relate to genes DISC1 and DTNBP1, implicated in schizophrenia. Colantuoni et al. (2008) report age-related changes in the expression of these genes in the human prefrontal cortex, including Brodmann area 10.
- ☞ Selected SNPs in the control group are implicated in the functioning of the central nervous system (CNS) that controls behavior.
  - ✌ Post-MCMC estimates of the regression coefficients inform us on the effects of the selected SNPs on the activations of the discriminating ROIs we selected.
- ☞ Our setting allows individual covariates to have differential effects $(\beta_{r1(\gamma)}, \ldots, \beta_{rK(\gamma)})$ on the selected features.

Results: Inference on selected regression coefficients

Interestingly, while effects are all significant across selected ROIs in the control group, differential effects are indicated in the schizophrenia group (SNP 25 - in gene DISC1- has a significant effect on the Rectus L only and SNP 31 - in gene DTNBP1- on the Olfactory ROIs).

ROI	Name	Schizophrenia group	Control Group
		SNP 25	*SNP 16*
ROI 5	Frontal Sup Orb L	0.0646 (-0.0861,0.2153)	-0.1801 (-0.3123,-0.0478)
ROI 21	Olfactory L	0.0635 (-0.1053,0.2322)	-0.2821 (-0.4446,-0.1195)
ROI 22	Olfactory R	0.0644 (-0.1060,0.2348)	-0.2783 (-0.4176,-0.1389)
ROI 27	Rectus L	0.2297 (0.0401,0.4193)	-0.2719 (-0.4400,-0.1038)
ROI 28	Rectus R	0.1649 (-0.0215,0.3514)	-0.2919 (-0.4350,-0.1487)
		SNP 31	*SNP 50*
ROI 5	Frontal Sup Orb L	0.0125 (-0.0698,0.0949)	0.2100 (0.0584,0.3615)
ROI 21	Olfactory L	0.1392 (0.0470,0.2314)	0.3273 (0.1411,0.5135)
ROI 22	Olfactory R	0.1373 (0.0442,0.2304)	0.2468 (0.0872,0.4064)
ROI 27	Rectus L	0.0978 (-0.0057,0.2014)	0.2240 (0.0313,0.4166)
ROI 28	Rectus R	0.0740 (-0.0279,0.1759)	0.2446 (0.0806,0.4087)

Predictions and comparisons

- Using all the selected ROIs and the selected SNPs, we correctly classify 67% of the validation set.
- We compare our joint estimation strategy with two-step approaches:

  (1) first classify subjects based on the imaging data (ROIs) data only

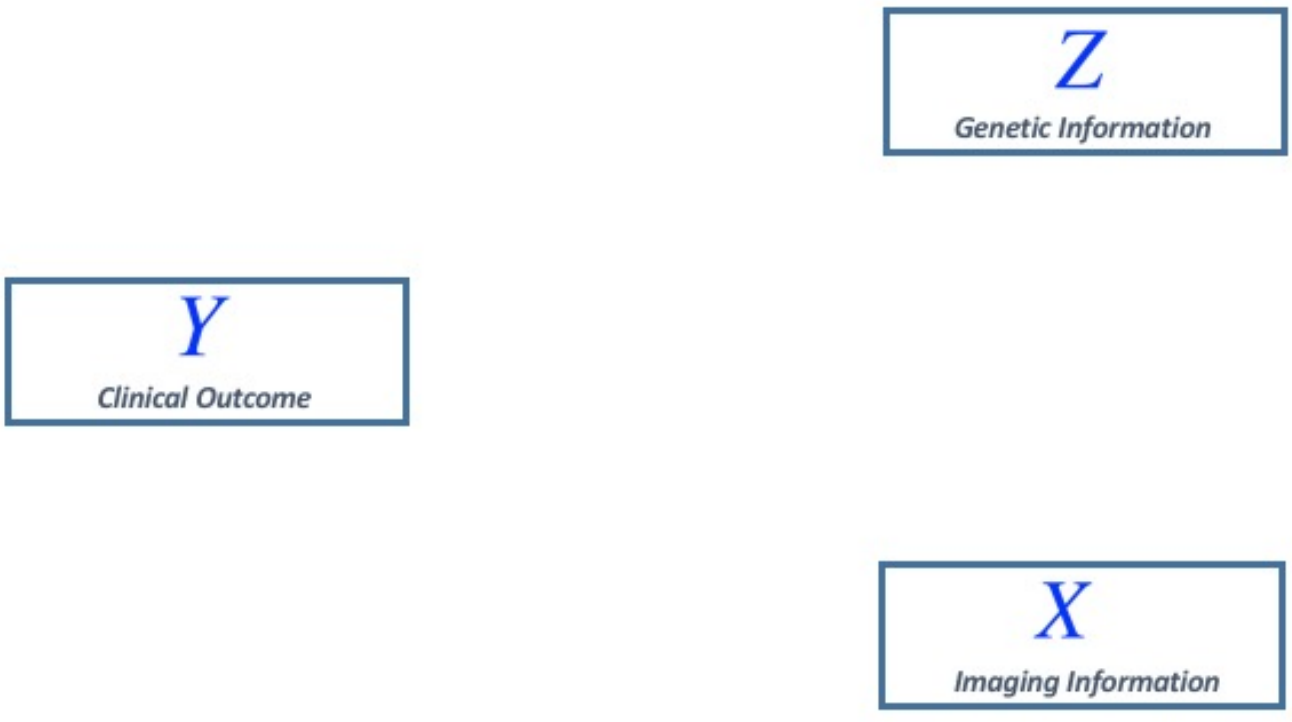
  (2) then apply variable selection in linear models that regress the individual ROIs on the SNPs.

  In step (1) Bayesian variable selection method for probit models of Sha et al. (2004, Biometrics) and support vector machine (SVM) gave classifications very similar to ours.

  In step (2), Guan and Stephens (2011, AOAS) selected none of the SNPs in the control group, and SNP9 for ROI5, SNP47 for ROI21 and SNP21 for ROI22 in schizophrenia.

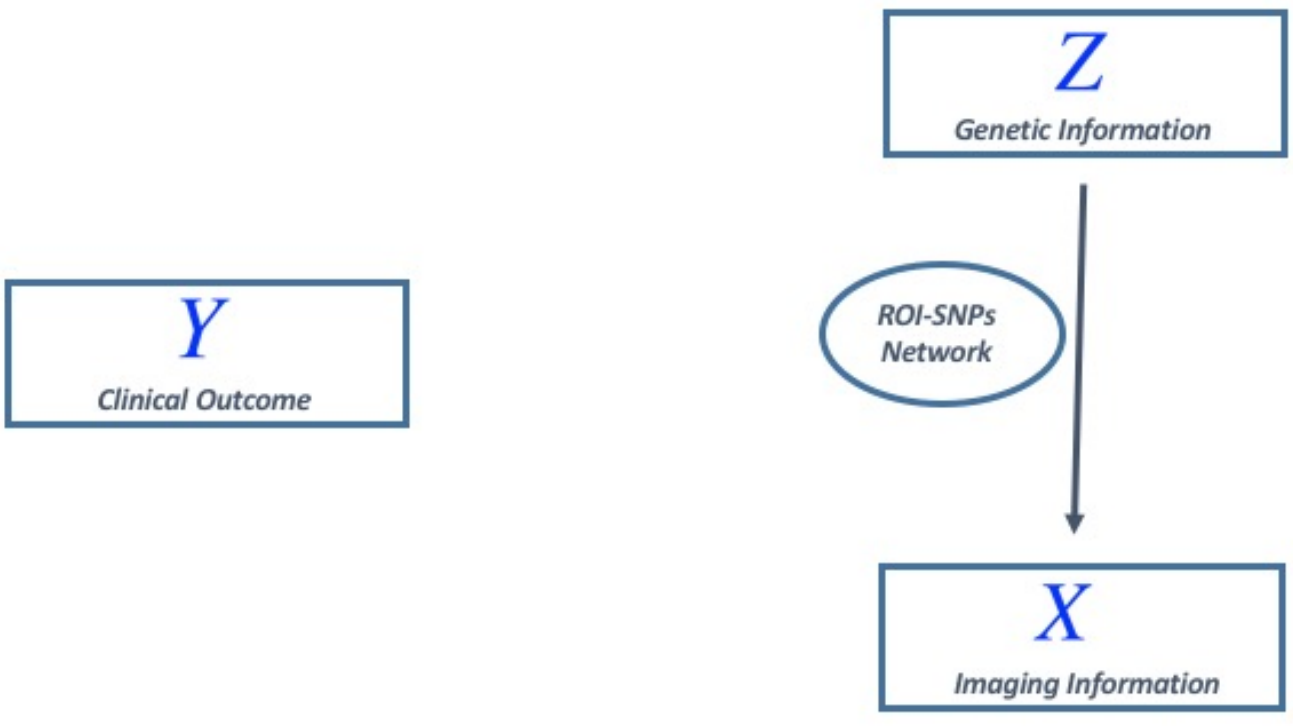
Alternative predictive model

A risk predictive model for disease status that takes into account direct associations between the SNPs/ROIs information and the disease status, as well as the indirect associations captured by a ROI-SNPs network



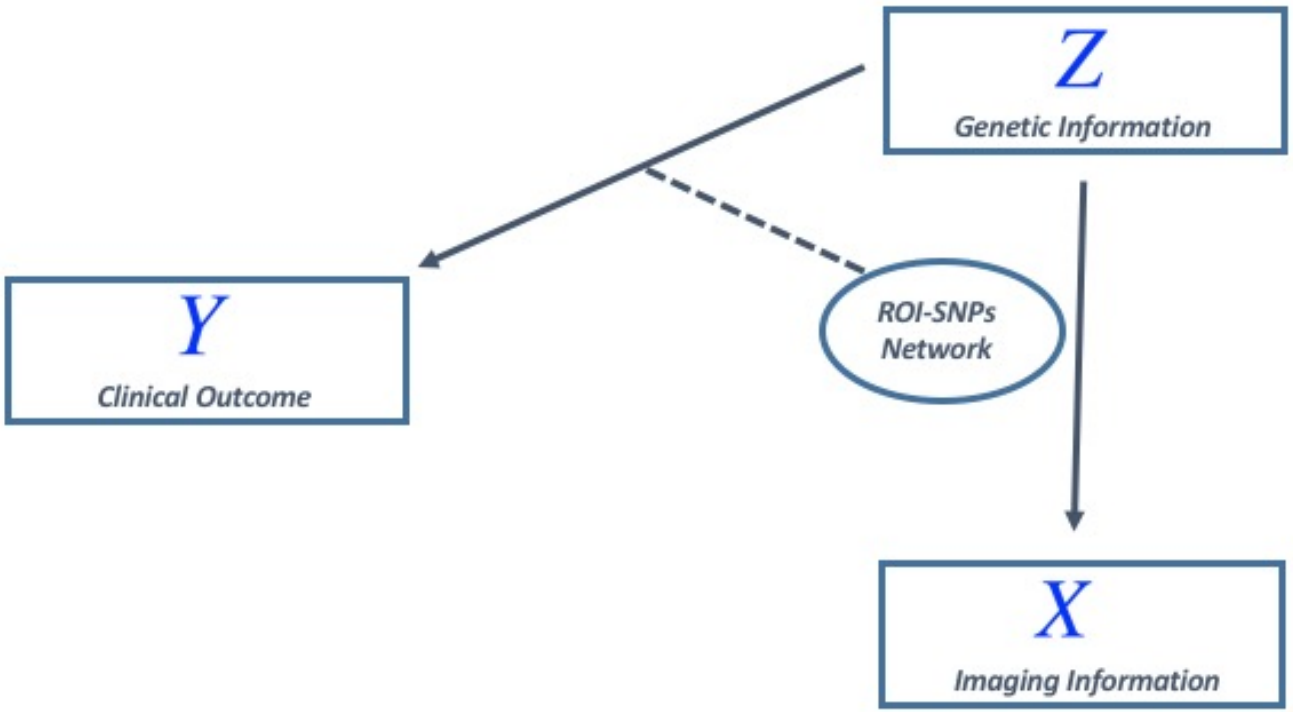
Alternative predictive model

A regulatory network in which SNPs can affect ROI intensities



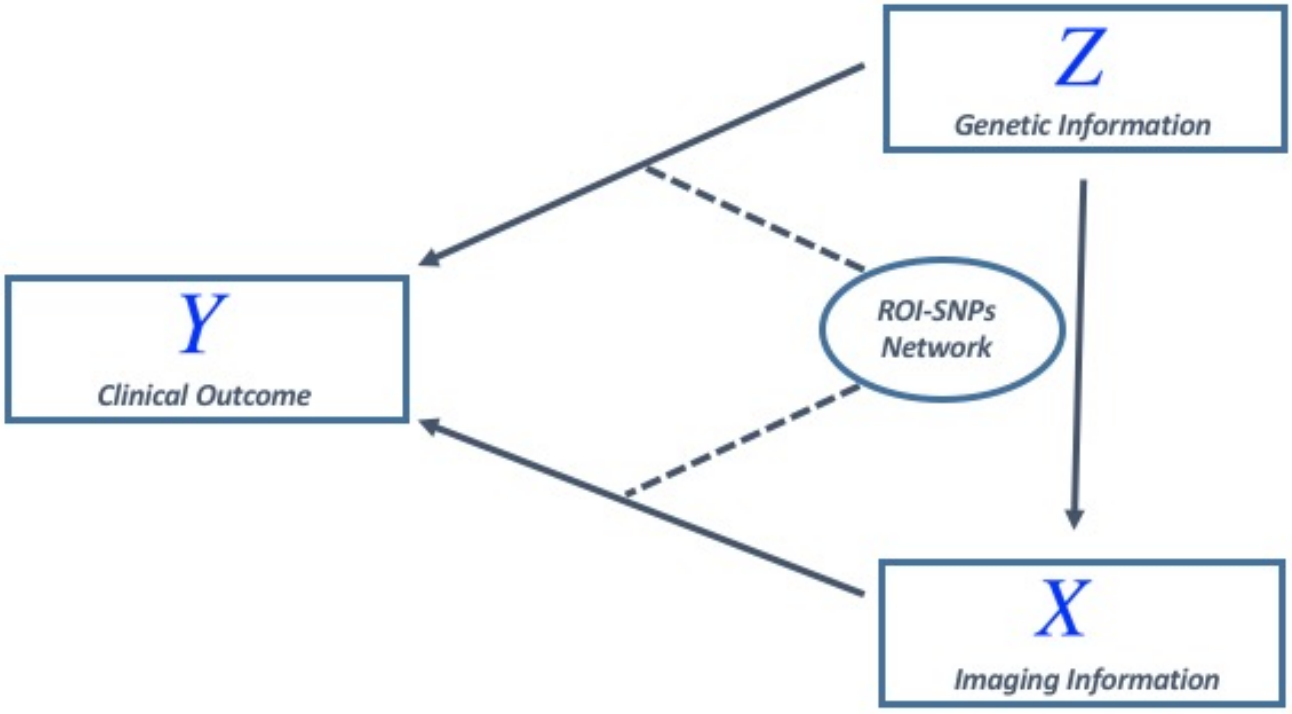
Alternative predictive model

The selection of discriminatory SNPs is informed by the ROI-SNP network (since SNPs involved in the regulatory network are more likely to be significantly associated with the clinical outcome).



Alternative predictive model

ROIs highly connected in the ROI-SNP network are more likely associated with the clinical outcome;
and clusters of adjacent ROIs



Outcome Predictive model

- We consider a binary outcome (e.g., disease status) ⇨ $n \times 1$ binary vector $\boldsymbol{y}$
- Bayesian Probit regression ⇨ auxiliary latent variables

$$\boldsymbol{y}^{\star} = \mathbf{1}_n \beta_0 + \boldsymbol{Z}\boldsymbol{\beta}^{(1)} + \boldsymbol{X}\boldsymbol{\beta}^{(2)} + \boldsymbol{\nu}, \quad \boldsymbol{\nu} \sim N(\mathbf{0}, \boldsymbol{I_n})$$

where

$$y_i = \begin{cases} 1 & \text{if } y_i^{\star} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

- Bayesian Variable Section: selection indicators

$\boldsymbol{\gamma^{(1)}} = (\gamma_1^{(1)}, ..., \gamma_M^{(1)})$ with $\gamma_m^{(1)} = 1$ if SNP m is included

$\boldsymbol{\gamma^{(2)}} = (\gamma_1^{(2)}, ..., \gamma_G^{(2)})$ with $\gamma_g^{(1)} = 1$ if ROI g is included

(George & McCulloch (1997); Stingo and Vannucci (2011))

Outcome Predictive model

- We consider a binary outcome (e.g., disease status) ⇨ $n \times 1$ binary vector $\boldsymbol{y}$
- Bayesian Probit regression ⇨ auxiliary latent variables

$$\boldsymbol{y}^{\star} = \mathbf{1}_n \beta_0 + \boldsymbol{Z}\boldsymbol{\beta}^{(1)} + \boldsymbol{X}\boldsymbol{\beta}^{(2)} + \boldsymbol{\nu}, \quad \boldsymbol{\nu} \sim N(\mathbf{0}, \boldsymbol{I_n})$$

where

$$y_i = \begin{cases} 1 & \text{if } y_i^{\star} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

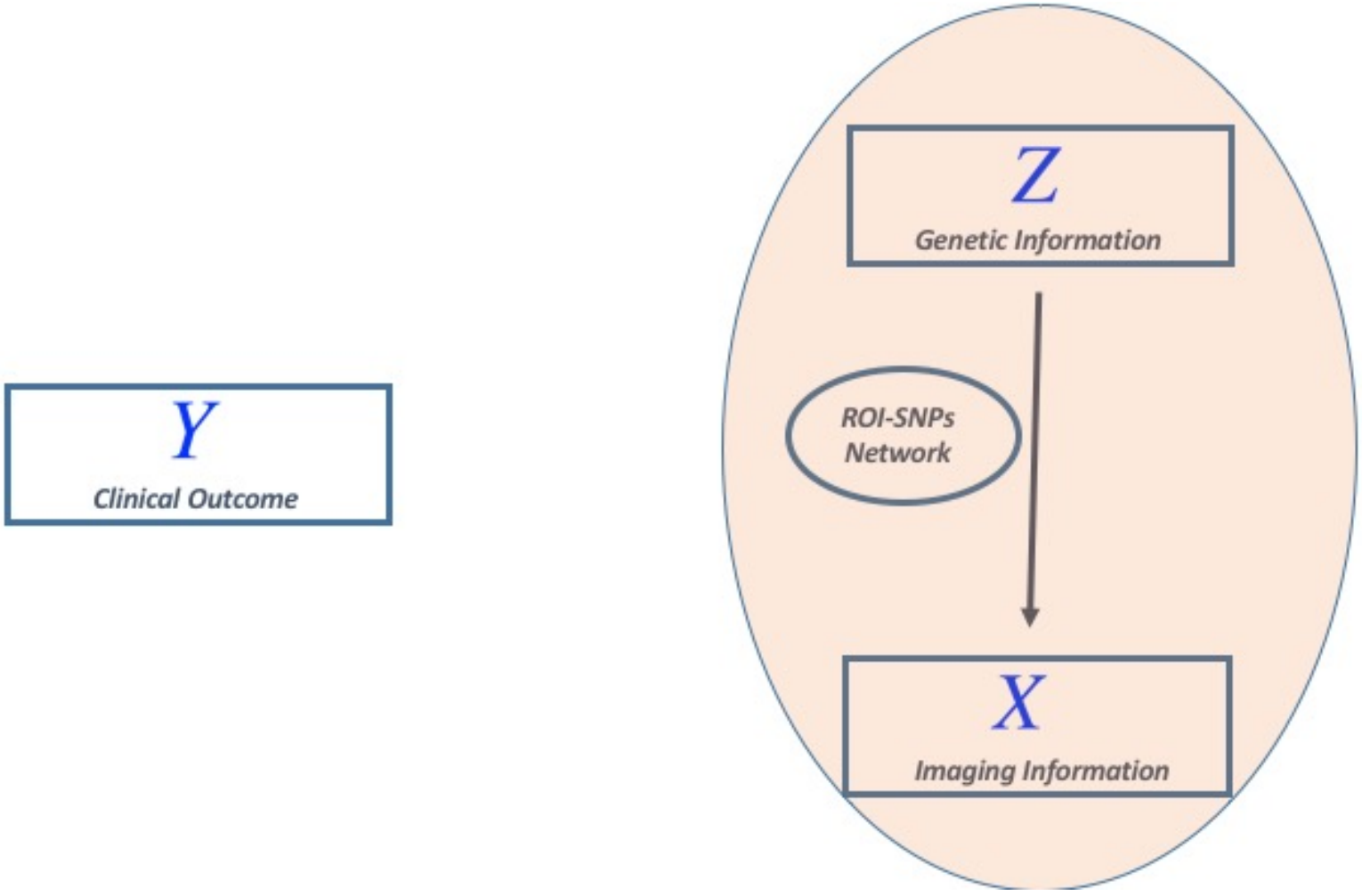
- Bayesian Variable Section: selection indicators

$\boldsymbol{\gamma^{(1)}} = (\gamma_1^{(1)}, ..., \gamma_M^{(1)})$ with $\gamma_m^{(1)} = 1$ if SNP m is included

$\boldsymbol{\gamma^{(2)}} = (\gamma_1^{(2)}, ..., \gamma_G^{(2)})$ with $\gamma_g^{(1)} = 1$ if ROI g is included

(George & McCulloch (1997); Stingo and Vannucci (2011))

Modeling the ROI-SNPs Network



Modeling the ROI-SNPs Network

- We model the ROI-SNP network as a DAG, to model ROIs potentially affected by the SNPs

$$\boldsymbol{x_g} = \mathbf{Z}\beta_{\boldsymbol{g}}^{(3)} + \epsilon_g, \qquad g = 1, ..., G,$$

with $\boldsymbol{\epsilon_g} = (\epsilon_{1g}, ..., \epsilon_{ng})^T \sim N(0, \sigma_g \boldsymbol{I_n})$

☞ *Conditional independence assumption:* $\boldsymbol{x_g} \perp\!\!\!\perp \boldsymbol{x_{g'}} \mid \boldsymbol{Z}$

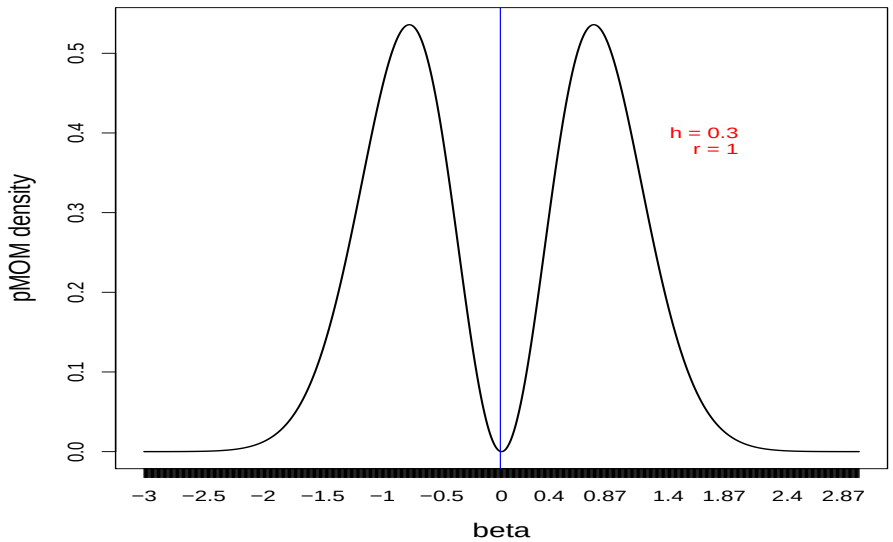
- Mixture prior (Spike-and-slab prior) on the β_{gm}'s

$$\beta_{gm}^{(3)} \sim \gamma_{gm}^{(3)}\ PM(0, r, \tau, \sigma^2) + (1 - \gamma_{gm}^{(3)})\ \delta_0, \quad m = 1, \ldots, M$$

If a SNP does not affect ROI g, then $\beta_{gm} = 0$
If a SNP affects ROI, then $\beta_{gm} \sim PM(0, r, \tau, \sigma^2)$.

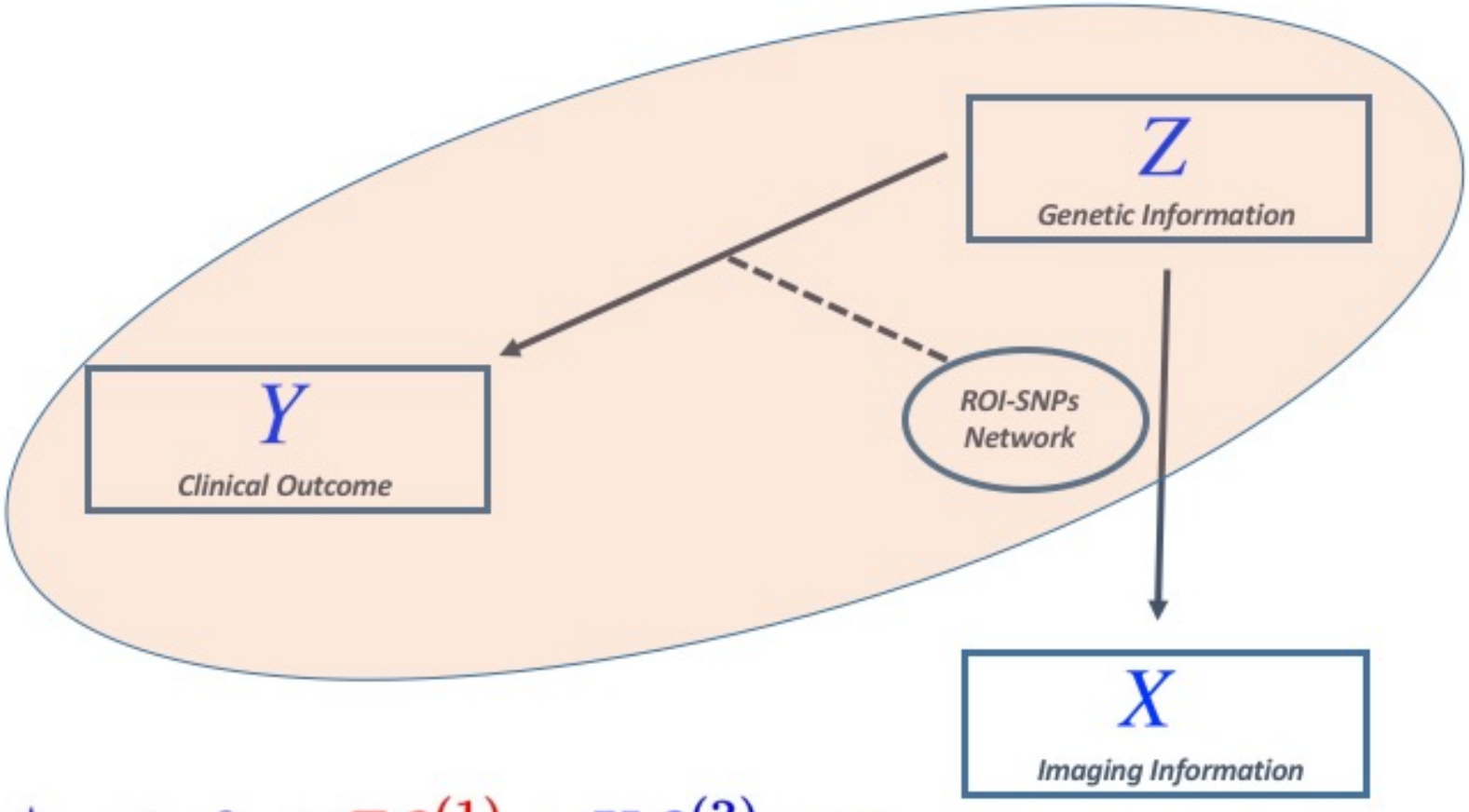
Product Moment prior $PM(0, r, \tau, \sigma^2)$ (Johnson & Rossell, 2012)

- Symmetric at zero
- Low prior probability to coefficients close to 0 ⇨ large effect sizes *Non local prior* .



Parameters r, h, σ^2: r characterizes the order of the distribution and h determines the dispersion around zero. (↑ h ⇨ ↑ effects).

Selection of discriminatory SNPs



$$y^{\star} = 1_n \beta_0 + Z\beta^{(1)} + X\beta^{(2)} + \nu$$

Selection of discriminatory SNPs

- Spike-and-slab prior on the $\beta_m^{(1)}$'s

$$\beta_m^{(1)} \sim \gamma_m^{(1)}\ PM(0, r, \tau, \sigma^2) + (1 - \gamma_m^{(1)})\ \delta_0, \quad m = 1, \ldots, M$$

- We model the SNP selection indicators $\gamma_m^{(1)}$ as a function of the inferred ROI-SNPs network:

$$P(\gamma_m^{(1)} = 1 | \mathbf{\Gamma^{(3)}}, \nu_1, \tau_1) = \frac{\exp\left(\nu_1 + \tau_1 \sum_{g=1}^{G} \gamma_{gm}^{(3)}\right)}{1 + \exp\left(\nu_1 + \tau_1 \sum_{g=1}^{G} \gamma_{gm}^{(3)}\right)}.$$

  - ν_1 sparsity parameter

Selection of discriminatory SNPs

- Spike-and-slab prior on the $\beta_m^{(1)}$'s

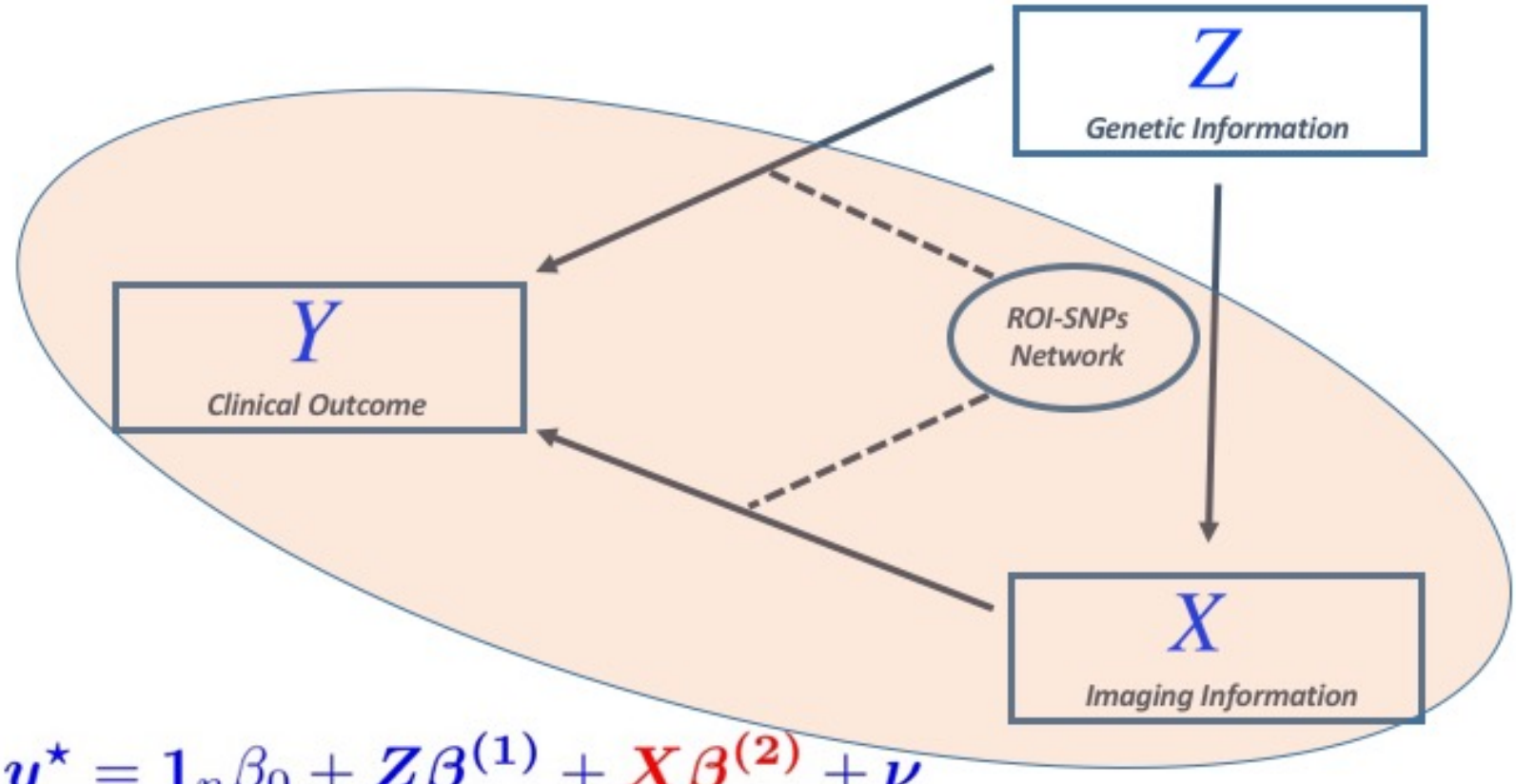
$$\beta_m^{(1)} \sim \gamma_m^{(1)}\ PM(0, r, \tau, \sigma^2) + (1 - \gamma_m^{(1)})\ \delta_0, \quad m = 1, \ldots, M$$

- We model the SNP selection indicators $\gamma_m^{(1)}$ as a function of the inferred ROI-SNPs network:

$$P(\gamma_m^{(1)} = 1 | \mathbf{\Gamma^{(3)}}, \nu_1, \tau_1) \propto \exp\left(\nu_1 + \tau_1 \sum_{g=1}^{G} \gamma_{gm}^{(3)}\right)$$

  - τ_1 controls the effect of the ROI-SNP network on the SNP selection
  - ☞ increasing function of the number of ROIs connected to each SNP

Selection of discriminatory ROIs



$$y^{\star} = 1_n \beta_0 + Z\beta^{(1)} + X\beta^{(2)} + \nu$$

Selection of discriminatory ROIs

- Spike-and-slab prior on the $\beta_g^{(2)}$'s

$$\beta_g^{(2)} \sim \gamma_g^{(2)}\ PM(0, r, \tau, \sigma^2) + (1 - \gamma_g^{(2)})\ \delta_0, \quad g = 1, \ldots, G$$

- Spatial dependencies via a covariate-dependent MRF:

$$P(\gamma_g^{(2)} | \mathbf{\Gamma^{(3)}}, (\gamma_{g'}^{(2)})_{g' \in N_g}) \propto \exp\left(\nu_2 \gamma_g^{(2)} + \tau_2 \sum_{m=1}^{M} \gamma_{gm}^{(3)} \gamma_g^{(2)} + \right.$$
$$\left. + 2\eta_2 \sum_{g' \in N_g} b_{gg'} \mathcal{I}(\gamma_g^{(2)} = \gamma_{g'}^{(2)}) \right).$$

  - ν_2 general sparsity parameter

Selection of discriminatory ROIs

- Spike-and-slab prior on the $\beta_g^{(2)}$'s

$$\beta_g^{(2)} \sim \gamma_g^{(2)}\ PM(0, r, \tau, \sigma^2) + (1 - \gamma_g^{(2)})\ \delta_0, \quad g = 1, \ldots, G$$

- Spatial dependencies via a covariate-dependent MRF:

$$P(\gamma_g^{(2)} | \mathbf{\Gamma^{(3)}}, (\gamma_{g'}^{(2)})_{g' \in N_g}) \propto \exp\left(\nu_2 \gamma_g^{(2)} + \tau_2 \sum_{m=1}^{M} \gamma_{gm}^{(3)} \gamma_g^{(2)} + \right.$$

$$\left. + 2\eta_2 \sum_{g' \in N_g} b_{gg'} \mathcal{I}(\gamma_g^{(2)} = \gamma_{g'}^{(2)}) \right).$$

  - τ_2 controls the effect of the number of SNPs connected to the ROIs;

Selection of discriminatory ROIs

- Spike-and-slab prior on the $\beta_g^{(2)}$'s

$$\beta_g^{(2)} \sim \gamma_g^{(2)}\ PM(0, r, \tau, \sigma^2) + (1 - \gamma_g^{(2)})\ \delta_0, \quad g = 1, \ldots, G$$

- Spatial dependencies via a covariate-dependent MRF:

$$P(\gamma_g^{(2)} | \mathbf{\Gamma^{(3)}}, (\gamma_{g'}^{(2)})_{g' \in N_g}) \propto \exp\left(\nu_2 \gamma_g^{(2)} + \tau_2 \sum_{m=1}^{M} \gamma_{gm}^{(3)} \gamma_g^{(2)} + \right.$$

$$\left. + 2\,\eta_2 \sum_{g' \in N_g} b_{gg'} \mathcal{I}(\gamma_g^{(2)} = \gamma_{g'}^{(2)}) \right).$$

- $b_{gg'} = \exp\{-\frac{d(g,g')^2}{2\sigma_r^2}\}$ if $g' \in N_g$ and 0 otherwise.
  η_2 is a smoothness parameter: $\uparrow \eta_2 \Rightarrow \uparrow \#\{\gamma_{g^{(2)}} = 1\}$

Prediction: classification of future cases

- Given imaging and genetic measurements $\boldsymbol{X_{new}}$ and $\boldsymbol{Z_{new}}$ for new subjects. we can predict $\boldsymbol{y_{new}}$.
- The latent variables $\boldsymbol{y^{\star}_{new}}$ are predicted using a Bayesian model averaging approach (Sha et al, 2004):

$$\hat{\boldsymbol{y}}^{\star}_{\boldsymbol{new}} = \sum_{(\gamma^{(1)},\gamma^{(2)})} (\mathbf{1}_n\tilde{\beta}_0+\boldsymbol{Z_{new}}\tilde{\boldsymbol{\beta}}^{(1)}+\boldsymbol{X_{new}}\tilde{\boldsymbol{\beta}}^{(2)})p(\gamma^{(1)},\gamma^{(2)}|\hat{\boldsymbol{y}}^{\star},\boldsymbol{X},\boldsymbol{Z},\hat{\theta}),$$

where

- $\hat{\theta} = (\hat{\tau}_1, \hat{\tau}_2, \hat{\boldsymbol{\Gamma}}^{(3)})$ and $\hat{\Theta} = (\tilde{\beta}_0, \tilde{\boldsymbol{\beta}}^{(1)T}, \tilde{\boldsymbol{\beta}}^{(2)T})^T$ are MCMC posterior estimates
- The latent variable $\boldsymbol{y}^{\star}$ is set to the mean $\hat{\boldsymbol{y}}^{\star}$ of the $\boldsymbol{y}^{\star}$'s, sampled during the MCMC algorithm.
- The predictive probabilities of disease status can be computed as $\hat{p}(y_i = 1|\boldsymbol{X},\boldsymbol{Z}) \approx \Phi(\hat{y}_i^{\star})$

Results: ROI-SNP network

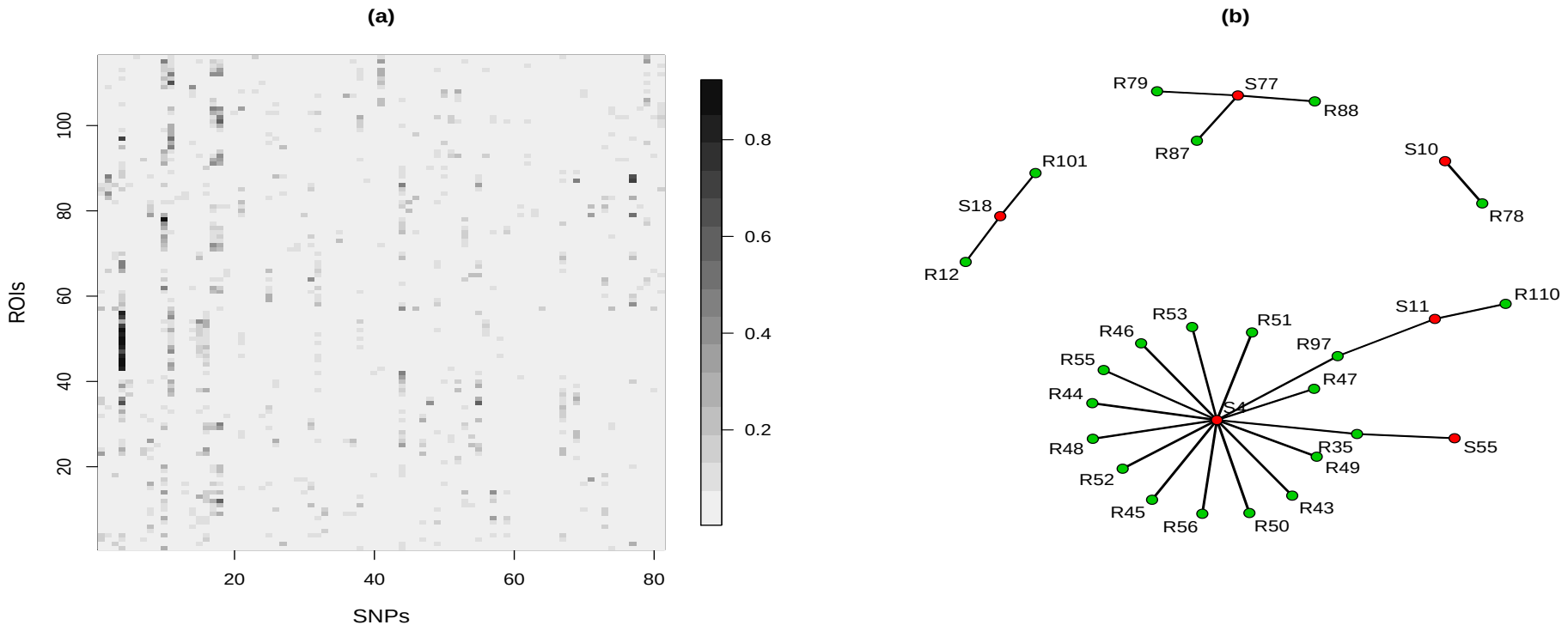
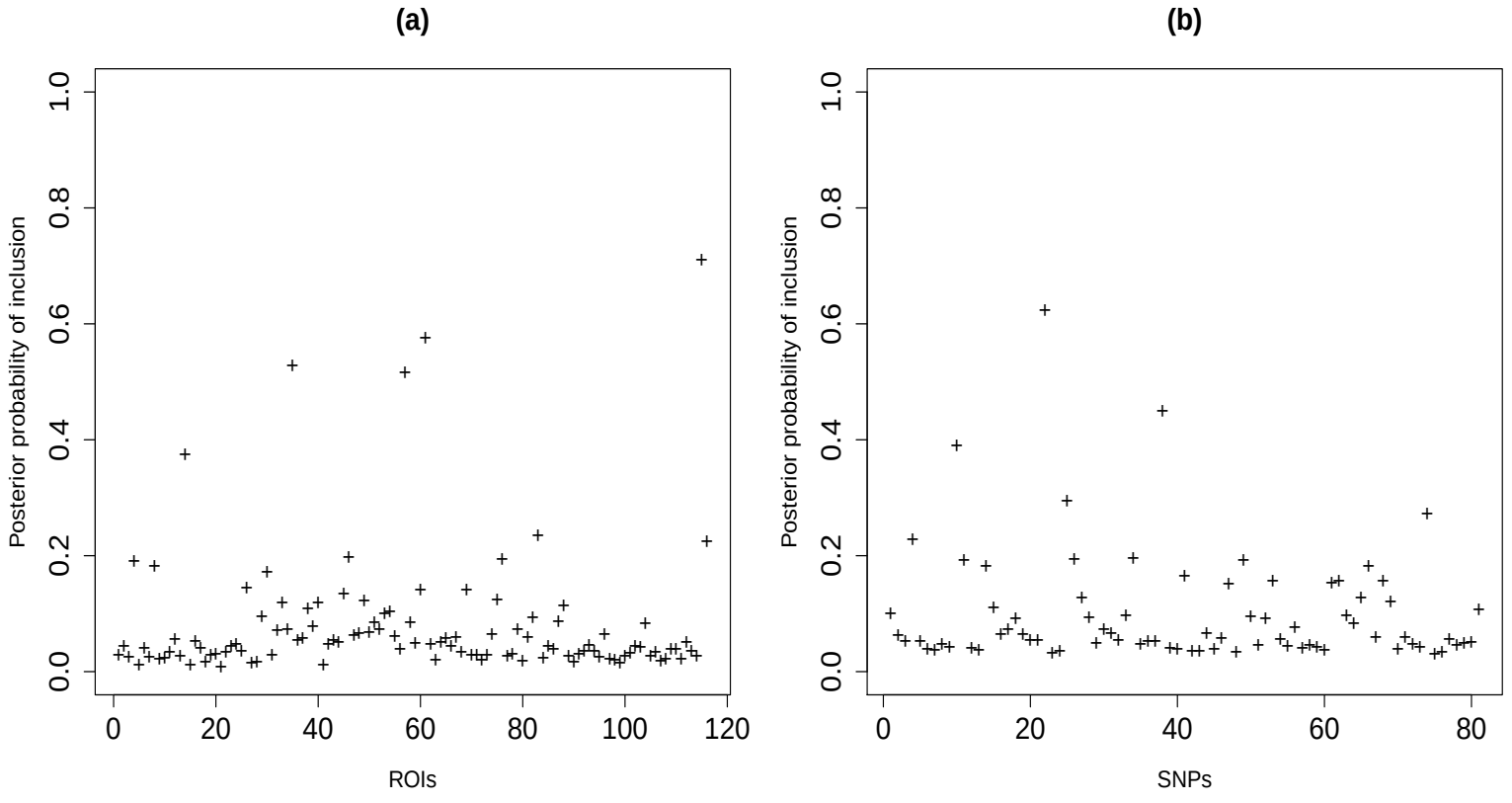


Figure: (a) ROI-SNP marginal posterior probabilities; (b) ROI-SNP network. Red nodes correspond to SNPs and green nodes correspond to ROIs.

Results: Selection of discriminatory ROIs and SNPS

Marginal posterior probabilities for ROIs (left) and SNPs (right)



Conclusions

References:

- Stingo, F.C., Guindani, M., Vannucci, M. and Calhoun, V. (2013). An Integrative Bayesian Modeling Approach to Imaging Genetics. *Journal of the American Statistical Association*, **108**, 876-891.
- Chekou, T, Stingo, F.C., Guindani, M. and Do, K. A Bayesian predictive model for imaging genetics with an application to schizophrenia. Under Invited Revision.

- ☞ Bayesian hierarchical modeling for the analysis of data that arise in imaging genetics.
- ☞ Identify brain regions (ROIs) with discriminating activations between schizophrenic patients and healthy controls and corresponding selection of SNPs.

Drawbacks and future work

- We use ROI-based summary statistics (point estimates):
  - implicit assumptions of stationarity
  - loss of temporal information
  - loss of power
- We have considered healthy controls and schizophrenic patients, based on clinical, symptom-based, categories:
  - Schizophrenia is a complex disease, and symptom-based categories are increasingly seen inadequate to represent such complexity:
  - Unsupervised model based clustering is necessary to identify important subgroups of the population
  - Available information can be incorporated in the clustering selection in a purely Bayesian framework.

Collaborators

Marina Vannucci, Rice

Fabrizio Leisen
University of Kent

Erik Erhardt,
MRN & UNM

Alberto Cassese
Maastricht University

Francesco Versace,
Stephenson Cancer Center

Vince Calhoun, Professor,
MRN & UNM

Francesco Stingo,
MDACC

Sharon Chiang,
Rice,
Keck Fellowship

Duncan Wadsworth,
Rice

Qiwei Li,
Rice

Ryan Warnick,
Rice
NSF fellowship

Eric Kook,
Rice

Ronaldo Guedes
University of Padua,
NYU Postdoc (soon)

Kim-Anh Do
MDACC

Thierry Chekouo,
Postdoc MDACC

Weixuan Zhu,
Universidad Carlos III, Madrid
Postdoc University of Sheffield

Linlin Zhang,
Phd from Rice

To keep updated:
http://www.micheleguindani.info
http://www.stat.rice.edu/~marina/