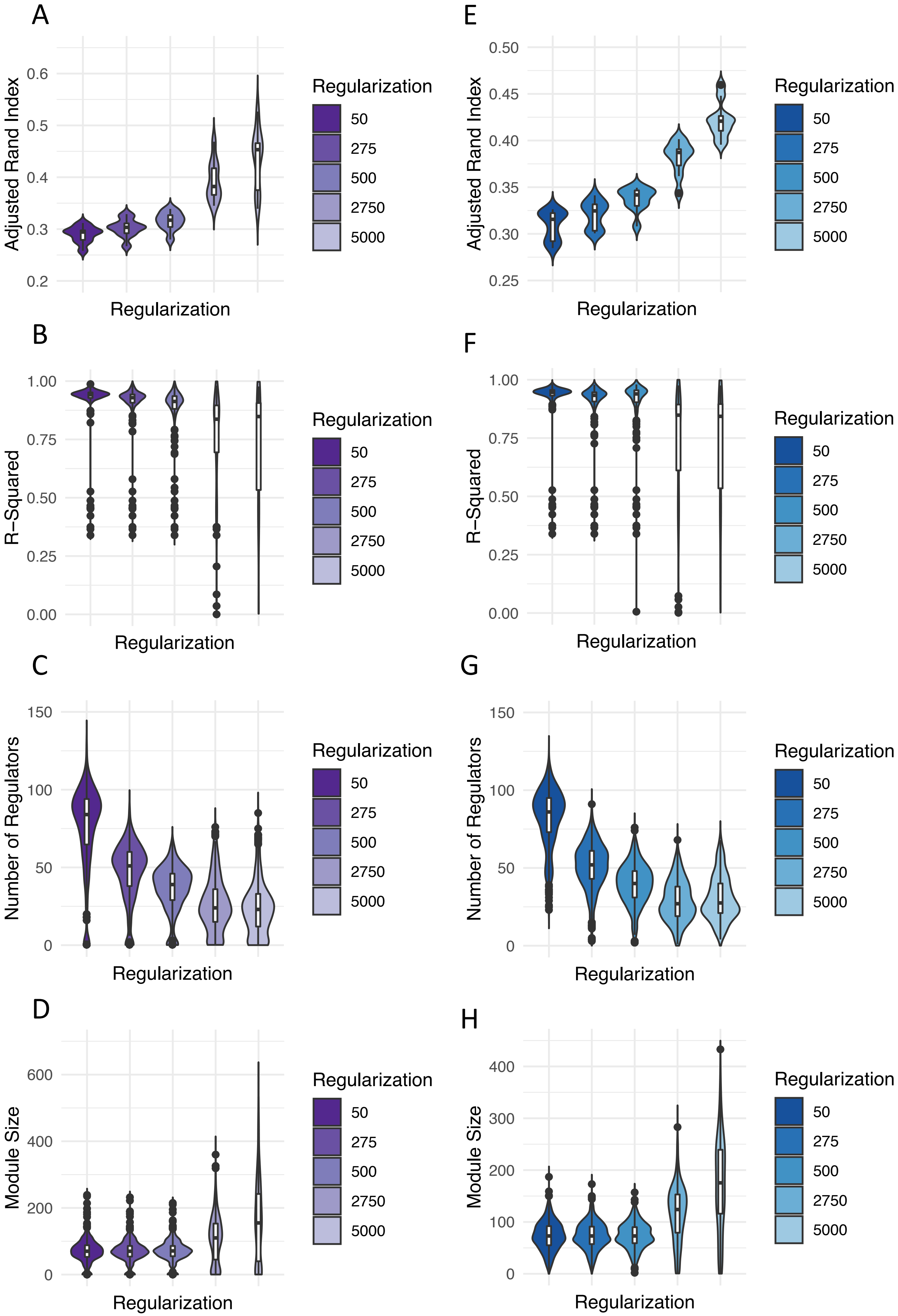
Cell Reports Methods, Volume *3*

Supplemental information

Identifying key multifunctional components shared by critical cancer and normal liver pathways via SparseGMM

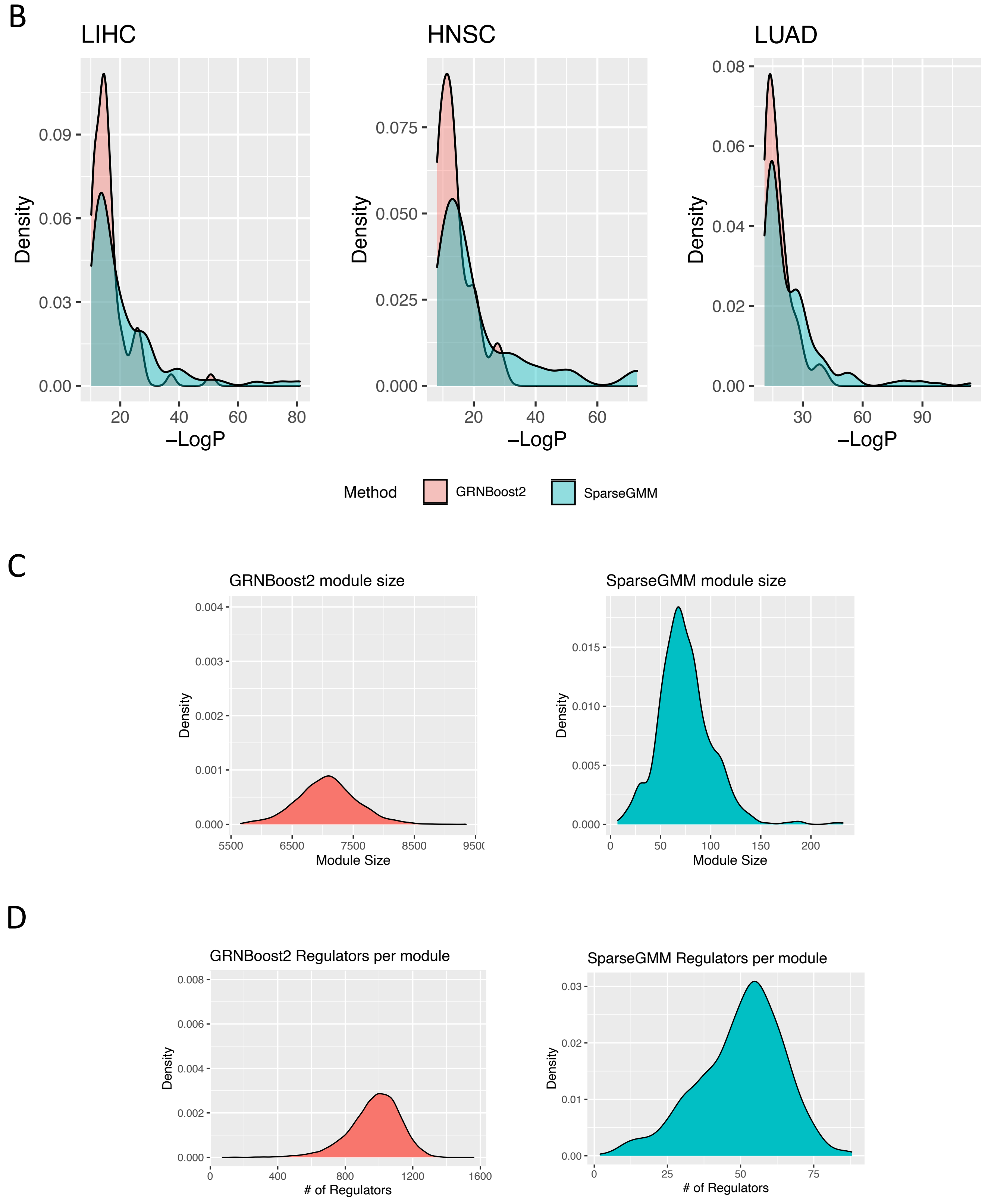
Shaimaa Bakr, Kevin Brennan, Pritam Mukherjee, Josepmaria Argemi, Mikel Hernaez, and Olivier Gevaert


A
Adjusted Rand Index
0.6
0.5
0.4
0.3
0.2
Regularization
Regularization
50
275
500
2750
5000
E
Adjusted Rand Index
0.50
0.45
0.40
0.35
0.30
0.25
Regularization
Regularization
50
275
500
2750
5000
B
R−Squared
1.00
0.75
0.50
0.25
0.00
Regularization
Regularization
50
275
500
2750
5000
F
R−Squared
1.00
0.75
0.50
0.25
0.00
Regularization
Regularization
50
275
500
2750
5000
C
Number of Regulators
150
100
50
0
Regularization
Regularization
50
275
500
2750
5000
G
Number of Regulators
150
100
50
0
Regularization
Regularization
50
275
500
2750
5000
D
Module Size
600
400
200
0
Regularization
Regularization
50
275
500
2750
5000
H
Module Size
400
300
200
100
0
Regularization
Regularization
50
275
500
2750
5000

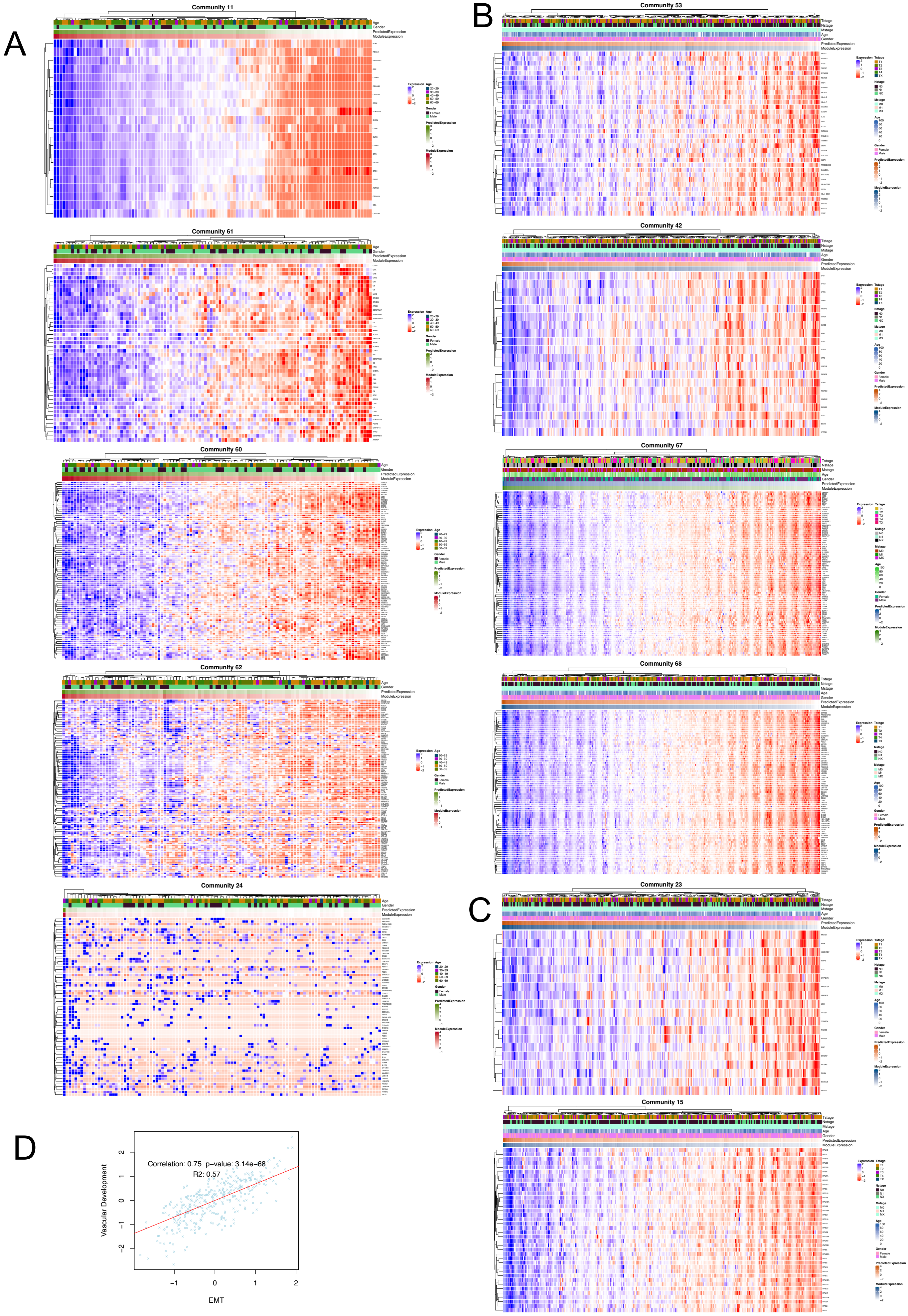
Supplementary Figure 1: Performance of SparseGMM at different regularization values. Related to Figure 2. Comparison shown for TCGA LUAD (A-D) and TCGA HNSC (E-H) data. (A,E) Robustness of clustering is evaluated using adjusted Rand index. (B,F) Validation of regulators is represented by adjusted R-squared. (C,G) Degree of sparsity is evaluated using statistics on the number of drivers. (D,H) Module size informs the choice of regularization parameter value.

A

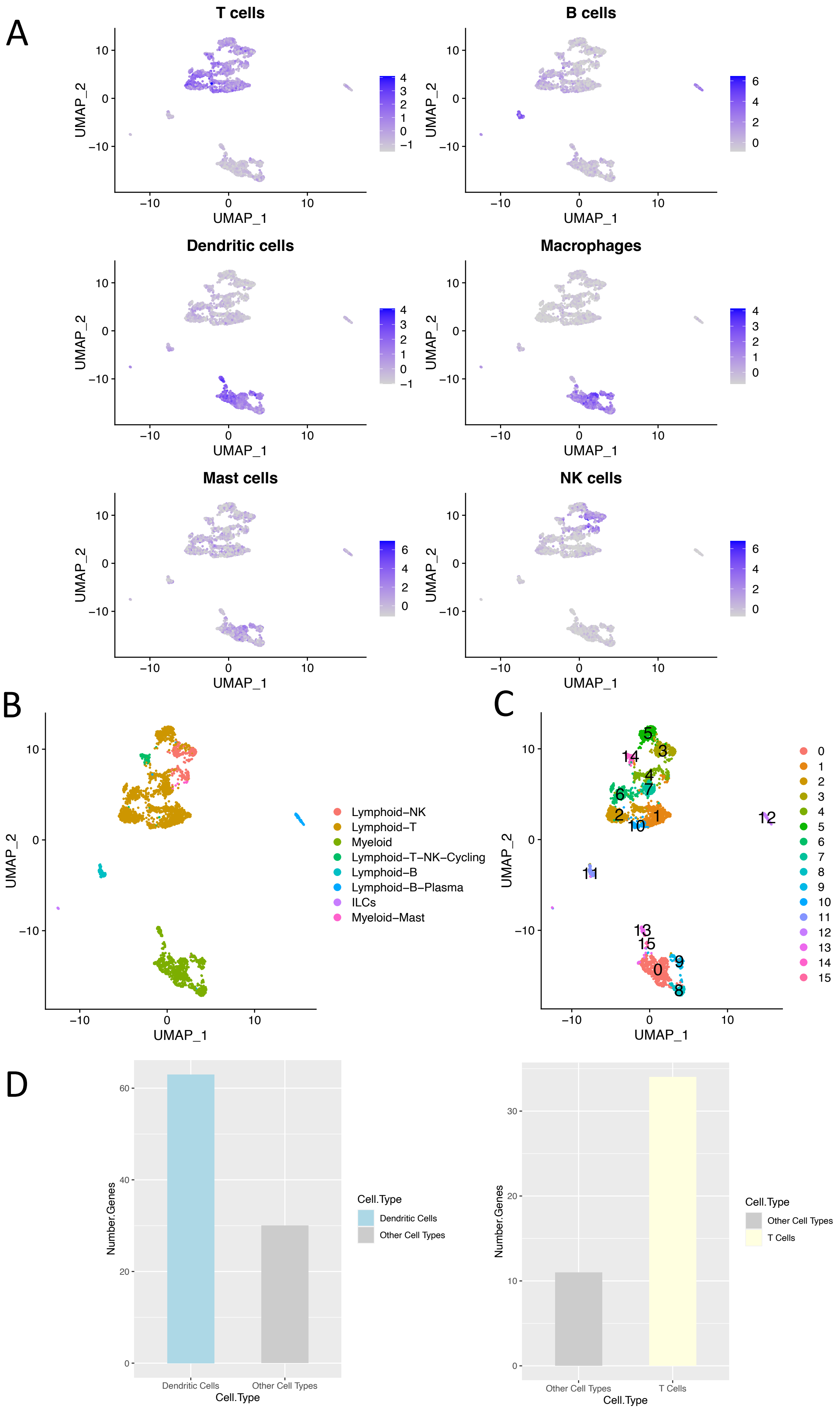
TCGA data set	# of Cell lines	Cell lines	GRNBoost2				SparseGMM			
			Total # of regulators	# Validated regulators	% Validated regulators	Runtime (sec)	Total # of regulators	# Validated regulators	% Validated regulators	Runtime (sec)
LUAD	7	A549, HCC515, dv90, h1299, NCIH2073, SKLU1, NCIH596	864	69	7.99	27258	***600***	***121***	***20.17***	***547***
HNSC	1	NPC	164	15	9.15	28692	***105***	***18***	***17.14***	***619***
LIHC	1	HEPG2	853	69	8.09	20493	***585***	***74***	***12.65***	***293***



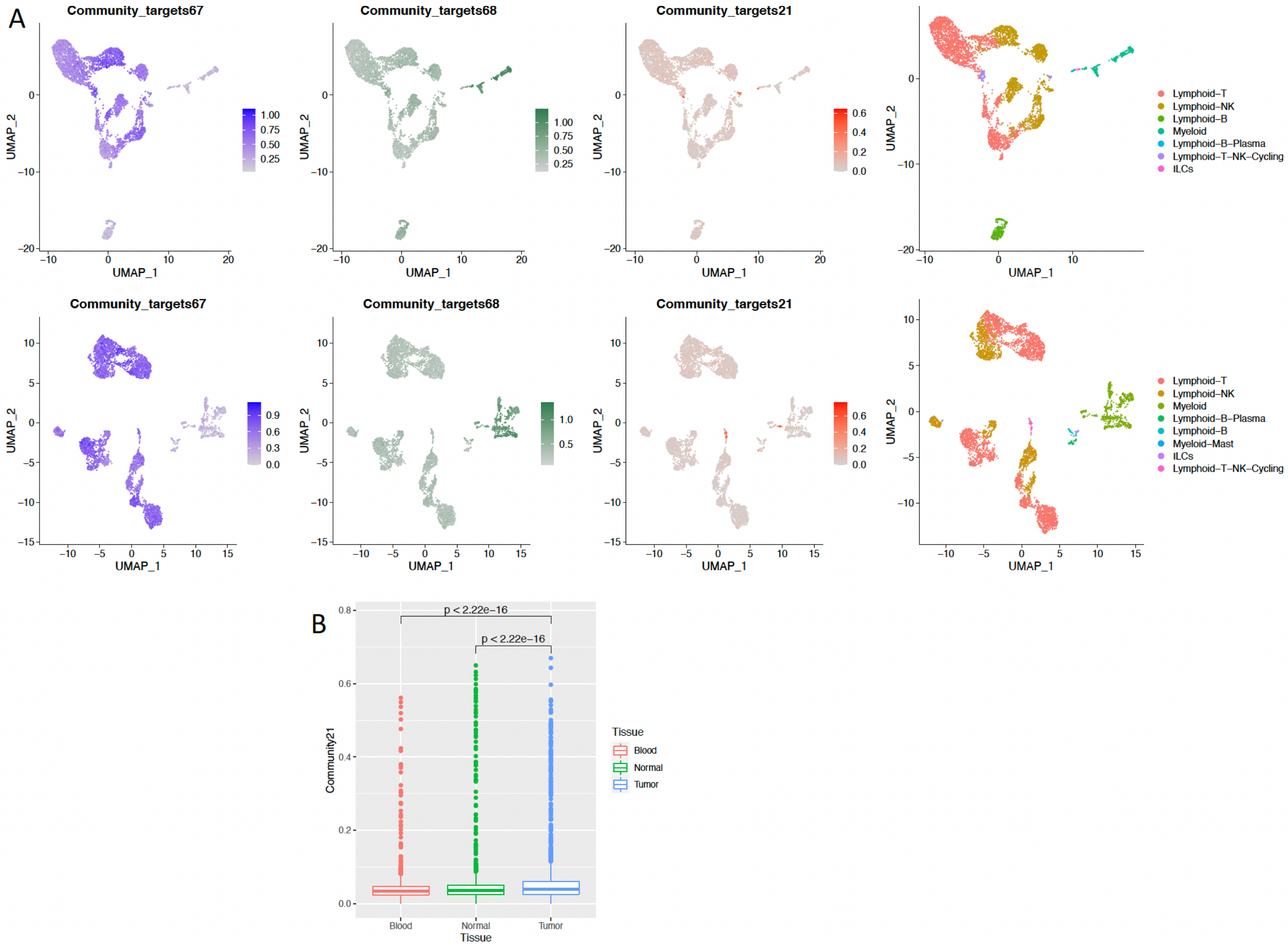
Supplementary Figure 2: SparseGMM achieves superior performance compared to GRNBoost2. Related to Figure 2. (A) Summary of comparison between SparseGMM and GRNBoost2. The total number of regulators represents the number of regulators selected by each method from the set of candidate regulators for which there are LINCS experimental data. (B) Distributions of p-value results from validation using the Fast Gene Set Enrichment Analysis tool. Results shown for both methods on the TCGA data sets: LIHC, HNSC and LUAD. (C) Distribution of modules size for SparseGMM and GRNBoost2. The number of target genes per regulators is used as a proxy for module size in GRNBoost2. (C) Distribution of the number of regulators per module for SparseGMM and GRNBoost2. The number of regulators per target gene is computed for GRNBoost2.


A
Community 11
Community 61
Community 60
Community 62
Community 24
B
Community 53
Community 42
Community 67
Community 68
C
Community 23
Community 15
Age
Gender
PredictedExpression
ModuleExpression
Tstage
Nstage
Mstage
Expression
20–29
30–39
40–49
50–59
60–69
Female
Male
T1
T2
T3
T4
TX
N0
N1
NX
M0
M1
MX
D
Correlation: 0.75 p-value: 3.14e−68
R2: 0.57
Vascular Development
EMT
−1
0
1
2
−2

Supplementary Figure 3: Co-expression patterns of target genes in highly robust communities. Related to Figure 4. (A) normal liver communities: Lipid and protein catabolism, complement, vesicle trafficking, myofibril formation, and FGFR1 signaling. (B) Cancer communities: antigen presentation, interferon signaling, T cell and myeloid. (C) Shared communities: cell cycle and ribosome - protein synthesis. (D) Correlation between vascular development and EMT communities in TCGA samples


A
T cells
B cells
Dendritic cells
Macrophages
Mast cells
NK cells
UMAP_1
UMAP_2
−10
0
10
4
3
2
1
0
−1
6
4
2
0
B
Lymphoid−NK
Lymphoid−T
Myeloid
Lymphoid−T−NK−Cycling
Lymphoid−B
Lymphoid−B−Plasma
ILCs
Myeloid−Mast
C
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
D
Number.Genes
Cell.Type
Dendritic Cells
Other Cell Types
T Cells
0
20
40
60
10
30

Supplementary Figure 4: Cell type identification based on Panglao DB markers: comparison with original data annotation and Seurat-based clustering. Related to Figure 5. (A) Average expression of different cell type Panglao DB markers. (B) Original cell type assignments. (C) Seurat clusters (D) Cell type specific expression of communities 21 (dendritic cells) and 60 (T cells).



Supplementary Figure 5: Comparison of gene expression in different tissues. Related to Figure 5. (A) Expression of cell-type specific robust communities. Similar expression patterns of T cell and myeloid communities occur in cancer, normal and blood tissue. Cell cycle genes are significantly more expressed in subpopulation of immune cell in tumor samples. Top, left to right expression in blood samples of target genes in T cell, myeloid and cell cycle community, and cell type assignment. Bottom, left to right normal samples of target genes in in T cell, myeloid and cell cycle community, and cell type assignment. (B) Average expression of cell cycle community target genes in blood, normal and cancer immune cells.

Supplementary Table 1: Comparison of SparseGMM to AMARETTO at different regularization values: 50, 500 and 5000. Robustness of clustering is evaluated using adjusted Rand index. Validation of regulators is represented by R-squared. Degree of sparsity is evaluated using statistics on the number of drivers. Module size informs the choice of regulatization parameter value. Standard Deviation (std dev). Related to Figure 2.

			Adjusted Rand Index	R-Squared	Module Size	Number of Drivers
GTEx	sparseGMM	**Mean**	0.29	0.97	72.02	37.87
	50	**std dev**	0.01	0.01	36.09	8.60
		p-value	3.99E-44	7.77E-32	1.00E+00	6.87E-222
	sparseGMM	**Mean**	0.34	0.91	97.68	21.27
	500	**std dev**	0.02	0.10	53.32	9.24
		p-value	5.96E-25	7.52E-05	9.57E-45	0.00E+00
	sparseGMM	**mean**	0.42	0.78	679.43	30.15
	5000	**std dev**	0.07	0.24	382.79	10.80
		p-value	1.30E-01	5.42E-13	3.57E-45	2.54E-185
	AMARETTO	**Mean**	0.40	0.93	72.02	84.14
		std dev	0.02	0.13	29.18	46.92
TCGA	sparseGMM	**Mean**	0.31	0.96	53.44	86.19
	50	**std dev**	0.01	0.02	23.54	18.63
		p-value	9.77E-32	6.82E-04	1.00E+00	6.53E-297
	sparseGMM	**Mean**	0.33	0.92	54.31	38.20
	500	**std dev**	0.02	0.08	21.65	12.64
		p-value	3.06E-30	1.21E-63	3.30E-01	0.00E+00
	sparseGMM	**Mean**	0.41	0.82	178.13	27.09
	5000	**std dev**	0.03	0.19	87.42	12.55
		p-value	3.47E-01	1.82E-48	5.96E-111	0.00E+00
	AMARETTO	**Mean**	0.40	0.97	53.44	196.33
		std dev	0.02	0.06	26.81	90.69

Supplementary Table 4: ReMap validation of robust normal liver and liver cancer communities. Robust communities were defined by having Jaccard Index >= 0.7. Main pathway of each community was revealed through gene set enrichment analysis of SparseGMM modules in GTEx and TCGA data against MSigDB collections. Validation of regulators is established with an adjusted p-value < 0.05. Related to Figure 4.

Community	**Main pathway/Gene set**	**Jaccard Index**	**ReMap-validated Regulators**
	Cancer Communities		
72	Blood coagulation	0.7	HNF4A
	Shared Communities		
23	Sterol biosynthesis	0.9	SREBF2
21	Cell Cycle/DNA replication	0.9	BRCA1, HMGXB4, HSF2, ZNF652

Methods S1: Model for Sparse Gaussian Mixtures

Related to STAR Methods

1 Model

We propose a Bayesian generative model to learning the regulatory relationships among genes. In the context of gene regulatory networks, we classify genes into one of two types: target genes and regulator genes. Regulator genes are genes undergoing genomic events that are relevant to cancer progression or tumor growth. Target genes are genes whose expression is controlled by regulator genes, and which contribute to the biological processes responsible for cancer progression. Each group of target genes is regulated by a small set of regulator genes.
This model can be formulated as follows: $X^T = [x_1 x_2 .. x_i .. x_N]$ is a gene expression matrix $X \in \mathbb{R}^{N \times M}$, where N is the number of target genes and M is the number of subjects, G is a regulator expression matrix $\in \mathbb{R}^{M \times P}$ where M is the number of subjects and P is the number of regulator genes. Finally, $\boldsymbol{\beta} \in \mathbb{R}^{P \times K}$ is a weight matrix, where K is the number of gene modules. The mean of each Gaussian component is a vector of weights passed through a constant regulator gene expression matrix:

$$\begin{aligned} z &\sim Cat(\boldsymbol{\pi}) \\ x_i | z_i = k &\sim \mathcal{N}(G\boldsymbol{\beta}_k, \sigma_k I), \end{aligned}$$

where z_i is the latent indicator of the mixture component that generated gene i. The expression of gene i is a sample from a Gaussian with mean equal to the weights, $\boldsymbol{\beta}_k$ passed through a constant regulator gene expression matrix G. σ_k is the variance of the Gaussian mixture component

k and $\boldsymbol{\pi}$ is the parameter of the categorical distribution. Thus, $\boldsymbol{\theta_k} = [\boldsymbol{\beta}_k, \sigma_k]$.

Our Bayesian approach combines Gaussian mixtures with $\ell1$-norm regularization to enforce sparsity on the regulator weights, resulting in a small set of regulators for each mixture component. We develop an *expectation-maximization* (EM)-based algorithm to obtain a *maximum a posteriori* (MAP) estimate the Gaussian mixture of parameters. This is detailed in the following sections.

2 Gaussian Mixtures for Gene Regulatory Networks

Mixture models are useful for representing data that are generated from different distributions, such as multimodal data. The data is assumed to be generated from a mixture of components, each with specific parameters that specify its distribution. The goal is to estimate these parameters using the observed data without observing the true component membership of the data points, which is a hidden or latent variable of our model. In a mixture model with K distributions $z_i \in \{1, \dots, K\}$, point $\boldsymbol{x_i}$ is generated from distribution k with likelihood $p(\boldsymbol{x_i}|z_i = k)$. z_i has the distribution $p(z_i) = Cat(\pi)$ and the K distributions are mixed as follows:

$$p(\boldsymbol{x}_i|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k p(\boldsymbol{x}_i|\boldsymbol{\theta}_k), \tag{1}$$

where $\boldsymbol{\theta}$ are the parameters to be estimated for $k = 1 : K$, $\boldsymbol{\theta}$ is $[\boldsymbol{\theta_1} \dots \boldsymbol{\theta_k} \dots \boldsymbol{\theta_K}]$. π_k is the mixing weight of base distribution k, $0 < \pi_k < 1$ and $\sum_{k=1}^{K} \pi_k = 1$. For example, a mixture of Gaussian distributions would be modeled as follows:

$$p(\boldsymbol{x}_i|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x_i}|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}). \tag{2}$$

Point i can then be assigned to a component using the MAP or ML estimate of the parameter $\boldsymbol{\theta}$ is needed.

To obtain this estimate, we fit the model for the data $\mathcal{D}$, using the iterative expectation-maximization (EM) algorithm applied to the likelihood function. The EM algorithm, consists of two steps. In the first (E) step the missing values are inferred using parameter estimates from the previous

iteration. In the second (M) step, the likelihood function is maximized with respect to model parameters, giving new parameter estimates, which are improved with each subsequent iteration until convergence.
Using this model to cluster the data involves calculating the posterior probability $p(z_i = k|\boldsymbol{x_i}, \boldsymbol{\theta}^{t-1})$, the posterior probability that point i is generated from distribution k or the responsibility of cluster k for point i:

$$r_{ik} \triangleq p(z_i = k|x_i, \boldsymbol{\theta}^{(t-1)}), \tag{3}$$

where t is the current iteration number. This can be expanded as:

$$r_{ik} = \frac{p(z_i = k|\boldsymbol{\theta}^{t-1})p(\boldsymbol{x_i}|z_i = k, \boldsymbol{\theta}^{t-1})}{\sum_{k'} p(z_i = k'|\boldsymbol{\theta}^{t-1})p(\boldsymbol{x_i}|z_i = k', \boldsymbol{\theta}^{t-1})}. \tag{4}$$

To derive the objective function, we first look at the complete log likelihood of the data, which is defined as:

$$\ell_c(\boldsymbol{\theta}) \triangleq \sum_{i=1}^{N} \log[p(\boldsymbol{x}_i, \boldsymbol{z}_i|\boldsymbol{\theta})]. \tag{5}$$

Since the cluster assignments, z_i are not observed the expected likelihood is used. This is defined as:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{t-1}) = \mathbb{E}[\ell_c(\boldsymbol{\theta})|\mathcal{D}, \boldsymbol{\theta}^{t-1}], \tag{6}$$

where we take the expectation to account for the fact that z_i is not observed.
Specifically in the case of GMM, this gives:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{t-1}) = \sum_{i=1}^{N}\sum_{k=1}^{K} r_{ik} \log[\pi_{ik} p(\boldsymbol{x}_i|\boldsymbol{\theta}_k)]. \tag{7}$$

Now, a MAP estimate can be performed on the above equation in the M step, obtaining $\boldsymbol{\theta}^t$. In the case of Gaussian mixtures, each class conditional density is a Gaussian distribution and $\boldsymbol{\theta}$ is made up of the mean and variance of each distribution and this estimate is iteratively improved . Upon convergence, the final iteration T gives the final estimate $\boldsymbol{\theta}^T$.
We can apply this model to gene regulatory networks. In this case, the average expression of target genes is the mean of the mixture component, which corresponds to a gene module. The mean of the component is a linear function of the regulator genes regulating that module. Equation (7) then becomes:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{t-1}) = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \log[\pi_{ik} \mathcal{N}(\boldsymbol{x}_i | G\boldsymbol{\beta}_k, \sigma_k^2)]. \tag{8}$$

3 MAP Estimation - ℓ1-norm Regularization and Sparse GMM

MAP estimation with the right prior can be useful when we would like to avoid over-fitting of parameter estimates, which can occur in the case of Maximum Likelihood Estimation (MLE). Adding parameter priors, (8) becomes:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{t-1}) = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \log[\pi_{ik} \mathcal{N}(\boldsymbol{x}_i | G\boldsymbol{\beta}_k, \sigma_k^2)] + log(p(\boldsymbol{\theta})). \tag{9}$$

The parameters of the GMM to be estimated are $\boldsymbol{\theta_k} = [\boldsymbol{\beta}_k, \sigma_k]$ and π_k for $k = 1 : K$. In our problem, we are more interested in discovering the regulatory relationships between regulator and target genes, so we use a zero-mean Laplace prior for the weights $\boldsymbol{\beta}_k$ and use uniform priors for σ_k and π_k. Uniform priors will give the same result as MLE estimates, while the Laplace prior will give a regularized MAP estimate. Specifically, a Laplace prior is commonly chosen where a sparse solution is desired as it corresponds to ℓ1-norm regularization. A sparse solution can improve our understanding of gene regulatory relationships, as we hypothesize that only a few regulator genes regulate each module.

The expected likelihood function from (9) is updated to be:

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{t-1}) = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \log[\pi_{ik} \mathcal{N}(\boldsymbol{x}_i | G\boldsymbol{\beta}_k, \sigma_k^2)] + \\ + \log[Lap(\boldsymbol{\beta}_k | 0, 1/\gamma_k)] + \log(p(\sigma_k) + \log(p(\pi_k)), \end{aligned} \tag{10}$$

where

$$log(p(\boldsymbol{\theta})) = \sum_{i=1}^{N} \sum_{k=1}^{K} \log[Lap(\boldsymbol{\beta}_k | 0, 1/\gamma_k)] + \log(p(\sigma_k) + \log(p(\pi_k)). \tag{11}$$

Thus, we define a sparse Gaussian Mixture model as a Gaussian mixture model, where the mean of each Gaussian component is a random vector of weights sampled from a Laplace distribution with zero mean and passed through a constant matrix.

4 Hierarchical Bayes modeling

Using a Laplace prior directly results in an $\ell 1$-norm, which does not give a closed form solution during optimization.
We follow an approach similar to the EM for lasso approach [S26]. We then utilize the representation of a Laplace distribution as a Gaussian Scale Mixture (GSM) [S27, S28].

$$Lap(\beta_p|0,1/\gamma) = \frac{\gamma}{2}e^{-\gamma|\beta_p|} = \int \mathcal{N}(\beta_p|0,\tau_p^2)Ga(\tau_p^2|1,\frac{\gamma^2}{2})d\tau_p^2. \tag{12}$$

This is an example of a hierarchical Bayes model, where we include a prior on the hyperparameter τ^2 of the prior distribution $p(\boldsymbol{\theta})$. In this case, the hyperprior is the Gamma distribution with scale parameter $\frac{\gamma^2}{2}$. The expected complete data log likelihood is given by:

$$\begin{aligned}Q(\boldsymbol{\theta},\boldsymbol{\theta}^{t-1}) &= \sum_{i=1}^{N}\sum_{k=1}^{K} r_{ik}\log[\pi_{ik}\mathcal{N}(\boldsymbol{x}_i|G\boldsymbol{\beta}_k,\sigma_k^2)] \\ &+ \int \log[\mathcal{N}(\boldsymbol{\beta}_k|0,\boldsymbol{D}_{\tau k})[\sum_{p} Ga(\tau_{kp}^2|1,\gamma^2/2)]]d\tau_{kp} + \log(p(\sigma_k) + \log(p(\pi_k)),\end{aligned} \tag{13}$$

where $D_{\tau k}$ is $diag(\tau_{kp}^2)$ for $p = 1 : K$. The objective function then becomes:

$$Q(\boldsymbol{\beta}_k,\sigma_k) = \sum_{i=1}^{N}[r_{ik}[-n\log\sigma_k - \frac{1}{2\sigma_k^2}||x_i - G\beta_k||_2^2 + \log(\pi_{ik})] - \frac{1}{2}\beta_k^T\Lambda_k\boldsymbol{\beta}_k)d\tau] + c, \tag{14}$$

where: π_{ik} is the marginal probability of component $z_i = k$, $\Lambda_k = diag(1/\tau_{kp}^2)$ for $p = 1 : K$ and $c = \sum_{k=1}^{K}\int[log(p(\tau_{kp}))]d\tau + log(p(\sigma_k) + log(p(\boldsymbol{\pi}))$.

5 EM Algorithm

E step: We evaluate: $E(\frac{1}{\tau^2})$ and r_{ik}. From the expected complete data likelihood equation (14), the expected value of Λ_k is

$$E[\Lambda_k|\hat{\beta}_k, x, \hat{\sigma}_k] = \gamma diag(|\hat{\beta}_{1k}^{-1}|...|\hat{\beta}_{Pk}^{-1}|). \tag{15}$$

and the responsibilities are:

$$r_{ik} = \frac{\pi_k p(x_i,\boldsymbol{\beta}_k,\sigma_k)}{\sum_{k'}\pi_{k'}p(x_i,\beta_k,\sigma_k)}, \tag{16}$$

where:

$$p(x_i|G, \boldsymbol{\beta}_k, \sigma_k) = \mathcal{N}(x_i|G\boldsymbol{\beta}_k, \sigma_k I_N), \tag{17}$$

and

$$p(\boldsymbol{\beta}_k|\tau_k) = \mathcal{N}(\boldsymbol{\beta}_k|\mathbf{0}, \boldsymbol{D}_{\tau k}). \tag{18}$$

M step: Using a sparse learning approach [S29]. We estimate model parameters $\pi_k, \boldsymbol{\beta_k}$ and σ_k by optimizing the expected complete likelihood function with respect to each of the parameters, after substituting $E(\frac{1}{\tau^2})$ and r_{ik} obtained in the E step, taking derivative wrt $\boldsymbol{\beta_k}$

$$\nabla_{\boldsymbol{\beta}_k} l_c = \sum_{i=1}^{N} r_{ik}[G^T x_i - G^T G\boldsymbol{\beta_k}] - \sigma_k \Lambda_k \boldsymbol{\beta_k}] = 0$$

$$(\sum_{i=1}^{N} r_{ik} G^T G + \sigma_k \Lambda_k)\boldsymbol{\beta_k} = \sum_{i=1}^{N} r_{ik}(G^T x_i)$$

$$[(\boldsymbol{r_k}^T \mathbf{1})G^T G + \sigma_k \Lambda_k]\boldsymbol{\beta_k} = (G^T X^T)\boldsymbol{r_k}$$

,

where r_k is the responsibility vector of component k $\in \mathbb{R}^N$

$$\hat{\boldsymbol{\beta}}_k = ((\boldsymbol{r_k}^T \mathbf{1})G^T G + \sigma_k \Lambda_k)^{-1}(G^T X^T \boldsymbol{r_k}). \tag{19}$$

Taking the derivative wrt σ_k yields

$$\hat{\sigma_k} = \frac{\sum_{i=1}^{N} r_{ik}||xi - G\hat{\boldsymbol{\beta_k}}||_2^2}{M(\boldsymbol{r_k}^T \mathbf{1})}. \tag{20}$$

Taking the derivative wrt π_k yields the same result as a GMM:

$$\hat{\pi}_k = \frac{\boldsymbol{r}_k^T \mathbf{1}}{N}. \tag{21}$$

6 Implementation for Numerical Stability

Since we expect most $\boldsymbol{\beta_k}$ to be equal to zero, and to make the matrix under the inverse numerically stable, we use the SVD decomposition of G as follows:

$$G = UDV^T, \tag{22}$$

and

$$\psi = diag(|\beta_{jk}|/\gamma). \tag{23}$$

Taking the derivative wrt $\boldsymbol{\beta}_k$:

$$\arg\max_{\boldsymbol{\beta_k}} \sum_{i=1}^{N} \frac{1}{2} r_{ik}[\frac{1}{\sigma_k}||x_i - UDV^T\boldsymbol{\beta}_k||_2^2] - \frac{1}{2}\beta_k^T \Lambda \boldsymbol{\beta_k} \tag{24}$$

$$\sum_{i=1}^{N} r_{ik}[-(UDV^T)^T(x_i - UDV^T\boldsymbol{\beta_k})] + \sigma_k \Lambda \boldsymbol{\beta}_k = 0 \tag{25}$$

$$\sum_{i=1}^{N} r_{ik}(UDV^T)^T x_i = \hat{\sigma_k}\Lambda\boldsymbol{\beta}_k + \sum_{i=1}^{N} r_{ik}[(UDV^T)^T(UDV^T)\boldsymbol{\beta}_k] \tag{26}$$

$$VDU^T X^T \boldsymbol{r_k} = [\hat{\sigma_k}\Lambda + (\boldsymbol{r_k}^T\mathbf{1})VD^2V^T]\boldsymbol{\beta}_k \tag{27}$$

$$\hat{\boldsymbol{\beta}}_k = \frac{1}{[\boldsymbol{r_k}^T\mathbf{1}]}(\frac{\hat{\sigma_k}\Lambda}{[\boldsymbol{r_k}^T\mathbf{1}]} + VD^2V^T)^{-1} VDU^T X^T \boldsymbol{r_k}. \tag{28}$$

$$\hat{\boldsymbol{\beta}}_k = \frac{1}{[\boldsymbol{r_k}^T\mathbf{1}]}(\frac{\hat{\sigma_k}\Lambda}{[\boldsymbol{r_k}^T\mathbf{1}]} + VD^2V^T)^{-1}(VD)^2(D^{-2}V^T)(VDU^T)X^T\boldsymbol{r_k}$$

$$= \frac{1}{[\boldsymbol{r_k}^T\mathbf{1}]}(\frac{\hat{\sigma_k}\Lambda}{[\boldsymbol{r_k}^T\mathbf{1}]} + VD^2V^T)^{-1}(VD^2)D^{-1}U^T X^T\boldsymbol{r_k}$$

$$= \frac{1}{[\boldsymbol{r_k}^T\mathbf{1}]}(\frac{\hat{\sigma_k}\Lambda VD^{-2}}{[\boldsymbol{r_k}^T\mathbf{1}]} + VD^2V^TVD^{-2})^{-1}D^{-1}U^T X^T\boldsymbol{r_k}.$$

Thus, we are able to remove Λ from the inverse:

$$= \frac{1}{[\boldsymbol{r_k}^T\mathbf{1}]}\psi VV^T\psi^{-1}(\frac{\hat{\sigma_k}\Lambda VD^{-2}}{[\boldsymbol{r_k}^T\mathbf{1}]} + V)^{-1}D^{-1}U^T X^T\boldsymbol{r_k}$$

$$= \frac{1}{[\boldsymbol{r_k}^T\mathbf{1}]}\psi V(\frac{\hat{\sigma_k}\Lambda VD^{-2}}{[\boldsymbol{r_k}^T\mathbf{1}]} + V^T\psi V)^{-1}D^{-1}U^T X^T\boldsymbol{r_k}. \tag{29}$$

$$\hat{\boldsymbol{\beta}}_k = \frac{1}{[\boldsymbol{r_k}^T\mathbf{1}]}\psi V(\frac{\hat{\sigma_k}\Lambda VD^{-2}}{[\boldsymbol{r_k}^T\mathbf{1}]} + V^T\psi V)^{-1}D^{-1}U^T X^T\boldsymbol{r_k}. \tag{30}$$

This computation of $\hat{\boldsymbol{\beta}}_k$ avoids numerical instability.

7 Target Gene Entropy

The model allows us to calculate the entropy of each target genes using the conditional distribution of the latent indicator z_i. This is given by:

$$H(Z_i) = \sum_{k=1}^{K} r_{ik} log_2(r_{ik}), \tag{31}$$

where

$$r_{ik} \triangleq p(z_i = k | x_i, \boldsymbol{\theta}). \tag{32}$$

Since entropy is a measure of uncertainty, we hypothesized that gene entropy, or uncertainty in assignment to a gene module, could be interpreted as a proxy for multiple module membership, and thus be used to unveil the elements of hidden crosstalk in cancer.

References

1. Pahlman, L.I., Morgelin, M., Kasetty, G., Olin, A.I., Schmidtchen, A., and Herwald, H. (2013). Antimicrobial activity of fibrinogen and fibrinogen-derived peptides--a novel link between coagulation and innate immunity. Thromb Haemost *109*, 930-939.
2. Endo, Y., Nakazawa, N., Iwaki, D., Takahashi, M., Matsushita, M., and Fujita, T. (2010). Interactions of ficolin and mannose-binding lectin with fibrinogen/fibrin augment the lectin complement pathway. J Innate Immun *2*, 33-42.
3. Hoppe, B. (2014). Fibrinogen and factor XIII at the intersection of coagulation, fibrinolysis and inflammation. Thromb Haemost *112*, 649-658.
4. Kajander, T., Lehtinen, M.J., Hyvarinen, S., Bhattacharjee, A., Leung, E., Isenman, D.E., Meri, S., Goldman, A., and Jokiranta, T.S. (2011). Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement. Proc Natl Acad Sci U S A *108*, 2897-2902.
5. Blaum, B.S., Hannan, J.P., Herbert, A.P., Kavanagh, D., Uhrin, D., and Stehle, T. (2015). Structural basis for sialic acid-mediated self-recognition by complement factor H. Nat Chem Biol *11*, 77-82.
6. Groom, J.R., and Luster, A.D. (2011). CXCR3 in T cell function. Exp Cell Res *317*, 620-631.
7. van Gisbergen, K.P., Kragten, N.A., Hertoghs, K.M., Wensveen, F.M., Jonjic, S., Hamann, J., Nolte, M.A., and van Lier, R.A. (2012). Mouse Hobit is a homolog of the transcriptional repressor Blimp-1 that regulates NKT cell effector differentiation. Nat Immunol *13*, 864-871.
8. Mackay, L.K., Minnich, M., Kragten, N.A., Liao, Y., Nota, B., Seillet, C., Zaid, A., Man, K., Preston, S., Freestone, D*., et al.* (2016). Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. Science *352*, 459-463.
9. Bird, L. (2016). Lymphocyte responses: Hunker down with HOBIT and BLIMP1. Nat Rev Immunol *16*, 338-339.
10. Fang, G., Yu, H., and Kirschner, M.W. (1998). Direct binding of CDC20 protein family members activates the anaphase-promoting complex in mitosis and G1. Mol Cell *2*, 163-171.
11. Fang, G., Yu, H., and Kirschner, M.W. (1998). The checkpoint protein MAD2 and the mitotic regulator CDC20 form a ternary complex with the anaphase-promoting complex to control anaphase initiation. Genes Dev *12*, 1871-1883.
12. Kramer, E.R., Gieffers, C., Holzl, G., Hengstschlager, M., and Peters, J.M. (1998). Activation of the human anaphase-promoting complex by proteins of the CDC20/Fizzy family. Curr Biol *8*, 1207-1210.
13. Jelluma, N., Brenkman, A.B., van den Broek, N.J., Cruijsen, C.W., van Osch, M.H., Lens, S.M., Medema, R.H., and Kops, G.J. (2008). Mps1 phosphorylates Borealin to control Aurora B activity and chromosome alignment. Cell *132*, 233-246.
14. Klein, U.R., Nigg, E.A., and Gruneberg, U. (2006). Centromere targeting of the chromosomal passenger complex requires a ternary subcomplex of Borealin, Survivin, and the N-terminal domain of INCENP. Mol Biol Cell *17*, 2547-2558.
15. Gassmann, R., Carvalho, A., Henzing, A.J., Ruchaud, S., Hudson, D.F., Honda, R., Nigg, E.A., Gerloff, D.L., and Earnshaw, W.C. (2004). Borealin: a novel chromosomal passenger required for stability of the bipolar mitotic spindle. J Cell Biol *166*, 179-191.

16. Sampath, S.C., Ohi, R., Leismann, O., Salic, A., Pozniakovski, A., and Funabiki, H. (2004). The chromosomal passenger complex is required for chromatin-induced microtubule stabilization and spindle assembly. Cell *118*, 187-202.
17. Wesche, H., Gao, X., Li, X., Kirschning, C.J., Stark, G.R., and Cao, Z. (1999). IRAK-M is a novel member of the Pelle/interleukin-1 receptor-associated kinase (IRAK) family. J Biol Chem *274*, 19403-19410.
18. Li, N., Jiang, J., Fu, J., Yu, T., Wang, B., Qin, W., Xu, A., Wu, M., Chen, Y., and Wang, H. (2016). Targeting interleukin-1 receptor-associated kinase 1 for human hepatocellular carcinoma. J Exp Clin Cancer Res *35*, 140.
19. Cheng, B.Y., Lau, E.Y., Leung, H.W., Leung, C.O., Ho, N.P., Gurung, S., Cheng, L.K., Lin, C.H., Lo, R.C., Ma, S.*, et al.* (2018). IRAK1 Augments Cancer Stemness and Drug Resistance via the AP-1/AKR1B10 Signaling Cascade in Hepatocellular Carcinoma. Cancer Res *78*, 2332-2342.
20. Pedrelli, M., Davoodpour, P., Degirolamo, C., Gomaraschi, M., Graham, M., Ossoli, A., Larsson, L., Calabresi, L., Gustafsson, J.A., Steffensen, K.R.*, et al.* (2014). Hepatic ACAT2 knock down increases ABCA1 and modifies HDL metabolism in mice. PLoS One *9*, e93552.
21. Lu, M., Hu, X.H., Li, Q., Xiong, Y., Hu, G.J., Xu, J.J., Zhao, X.N., Wei, X.X., Chang, C.C., Liu, Y.K.*, et al.* (2013). A specific cholesterol metabolic pathway is established in a subset of HCCs for tumor growth. J Mol Cell Biol *5*, 404-415.
22. Weng, M., Zhang, H., Hou, W., Sun, Z., Zhong, J., and Miao, C. (2020). ACAT2 Promotes Cell Proliferation and Associates with Malignant Progression in Colorectal Cancer. Onco Targets Ther *13*, 3477-3488.
23. Che, L., Chi, W., Qiao, Y., Zhang, J., Song, X., Liu, Y., Li, L., Jia, J., Pilo, M.G., Wang, J.*, et al.* (2020). Cholesterol biosynthesis supports the growth of hepatocarcinoma lesions depleted of fatty acid synthase in mice and humans. Gut *69*, 177-186.
24. Calvisi, D.F., Wang, C., Ho, C., Ladu, S., Lee, S.A., Mattu, S., Destefanis, G., Delogu, S., Zimmermann, A., Ericsson, J.*, et al.* (2011). Increased lipogenesis, induced by AKT-mTORC1-RPS6 signaling, promotes development of human hepatocellular carcinoma. Gastroenterology *140*, 1071-1083.
25. Gabitova, L., Gorin, A., and Astsaturov, I. (2014). Molecular pathways: sterols and receptor signaling in cancer. Clin Cancer Res *20*, 28-34.
26. Murphy, K.P. (2012). Machine learning: a probabilistic perspective (MIT press).
27. Andrews, D.F., and Mallows, C.L. (1974). Scale mixtures of normal distributions. Journal of the Royal Statistical Society: Series B (Methodological) *36*, 99-102.
28. West, M. (1987). On scale mixtures of normal distributions. Biometrika *74*, 646-648.
29. Figueiredo, M.A. (2003). Adaptive sparseness for supervised learning. IEEE transactions on pattern analysis and machine intelligence *25*, 1150-1159.