

Supplemental Information

Genome-wide Identification of the Genetic Basis of Amyotrophic Lateral Sclerosis

Sai Zhang, Johnathan Cooper-Knock, Annika K. Weimer, Minyi Shi, Tobias Moll, Jack N.G. Marshall, Calum Harvey, Helia Ghahremani Nezhad, John Franklin, Cleide dos Santos Souza, Ke Ning, Cheng Wang, Jingjing Li, Allison A. Dilliot, Sali Farhan, Eran Elhaik, Iris Pasniceanu, Matthew R. Livesey, Chen Eitan, Eran Hornstein, Kevin P. Kenna, Project MinE Sequencing Consortium, Jan Veldink, Laura Ferraiuolo, Pamela J. Shaw, and Michael P. Snyder

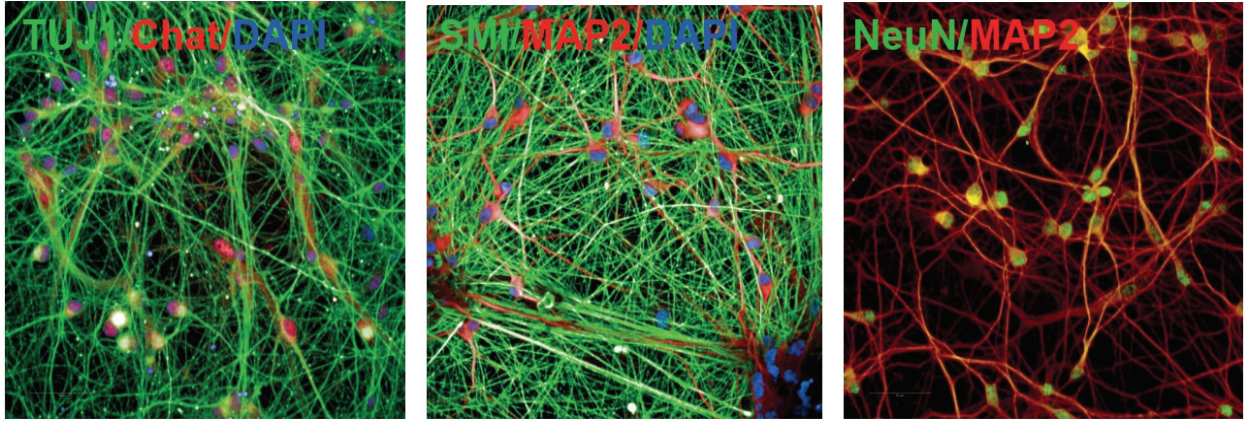
- **Figures S1-S7**
- **Supplemental Note.** Additional method details, Related to **STAR Methods**.

Figure S1

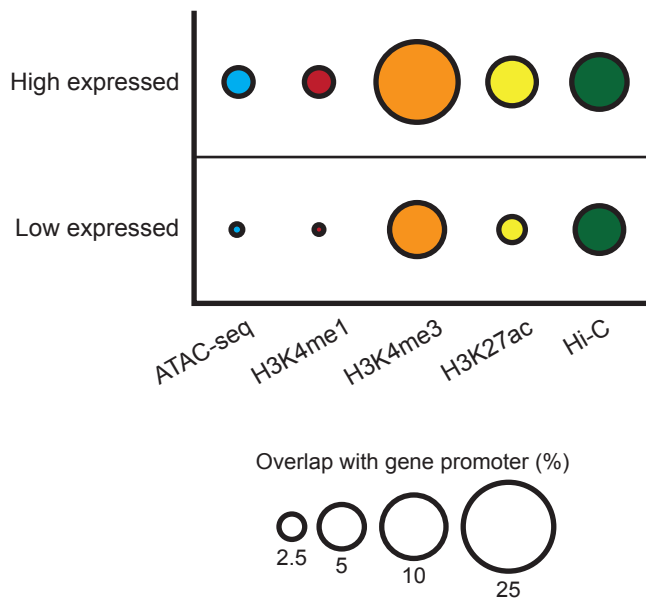
A

Cell Line	Short name	Source Tissue	Gender	Age at Sampling	Biobank
CS00iCTR-nxx	SMA	Fibroblast	Male	6	Cedars-Sinai
14iCTR-21nxx	CS14	Fibroblast	Female	52	Cedars-Sinai
GM23338	PGP	Fibroblast	Male	55	Coriell

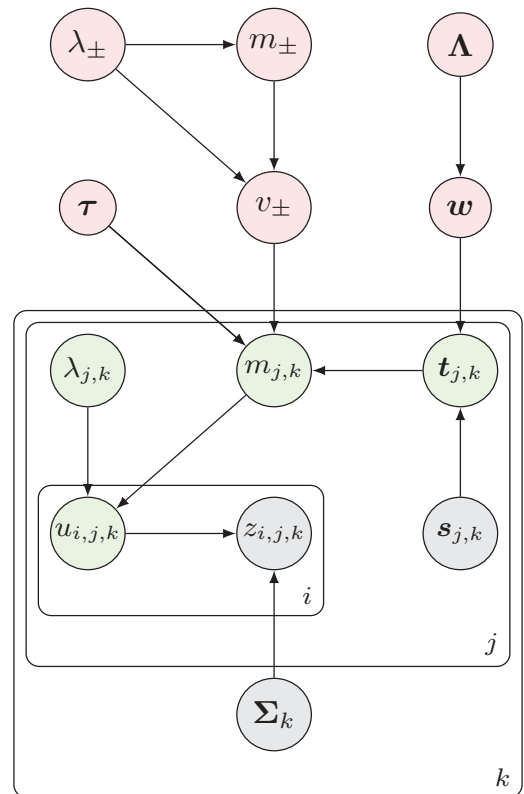
B



C



D



E

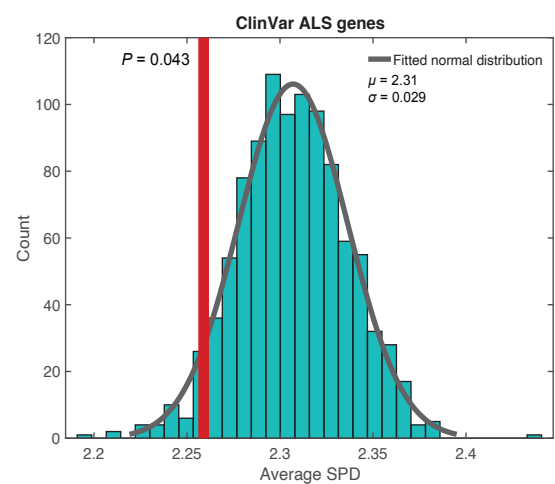
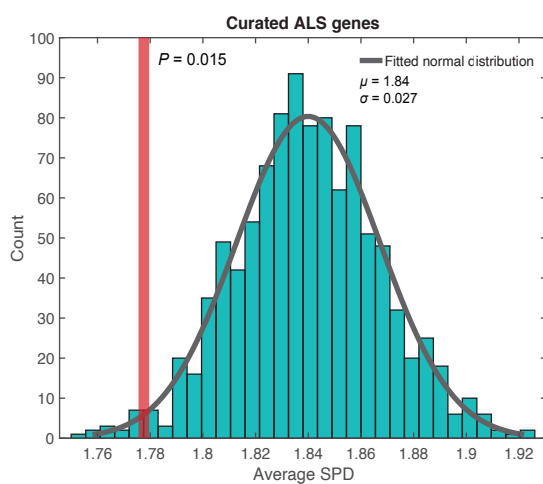
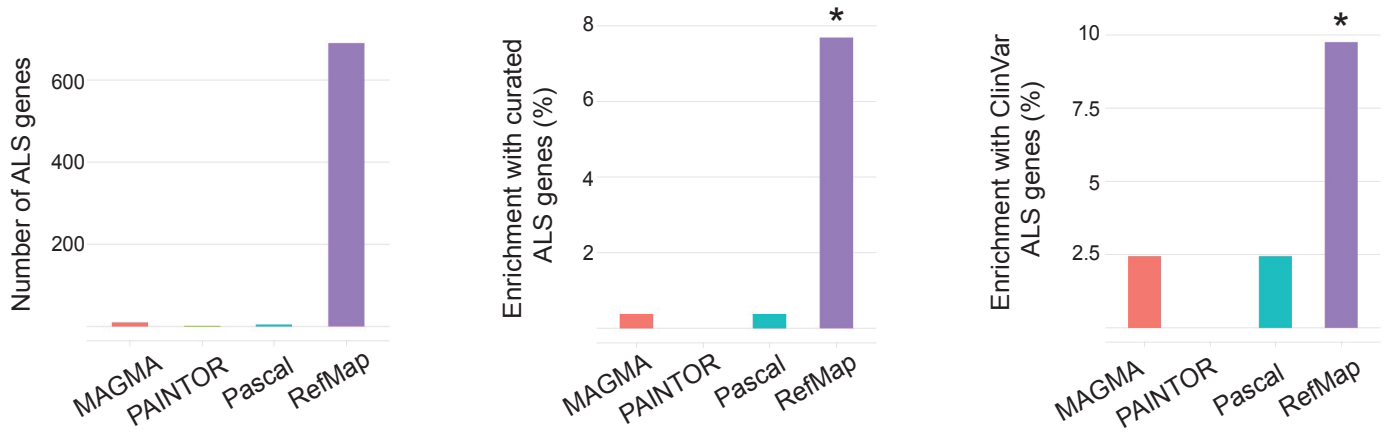


Figure S1. Donor information, iPSC differentiation, and RefMap model design and performance, Related to Figure 1

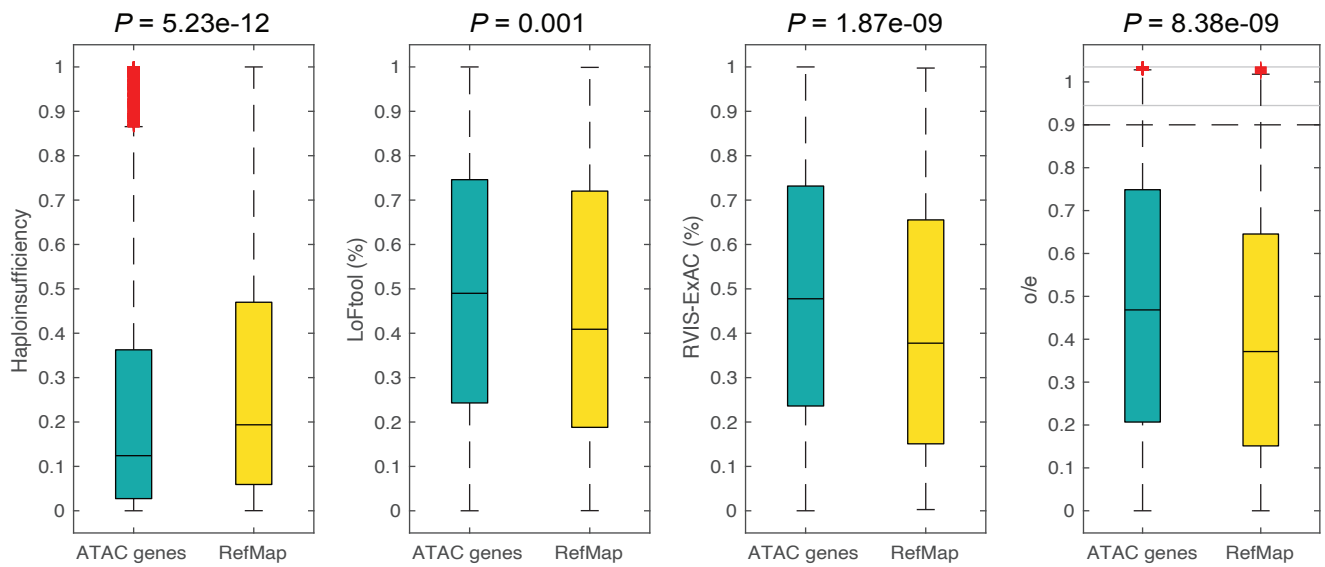
(A) iPSC cells were derived from control fibroblasts. (B) iPSC-derived motor neurons are morphologically consistent with lower motor neurons, including expression of appropriate markers. (C) Epigenetic profiling of iPSC-derived motor neurons is internally consistent. Markers of genomic activity are significantly enriched in promoter regions of highly-expressed genes compared to lowly-expressed genes. Circle area is proportional to the overlap percentage. Hi-C data was scaled by a factor of ten for clarity. (D) Graphical representation of RefMap. Observed variables are annotated in grey, local hidden variables are in green, and global latent variables are in pink. (E) PPI distance between RefMap ALS genes and known ALS genes. Two ALS gene sets, including curated ALS genes (left panel) and ClinVar ALS genes (right panel), were tested. The average shortest path distance (SPD) between novel RefMap ALS genes and known ALS genes was calculated and denoted by the red vertical line, and the average SPDs between randomly-selected (1,000 times) gene sets of equivalent size ($n=690$) and known ALS genes were calculated and denoted by histogram. The normal distribution was used to fit to the histogram (shown in black curves).

Figure S2

A



B



C

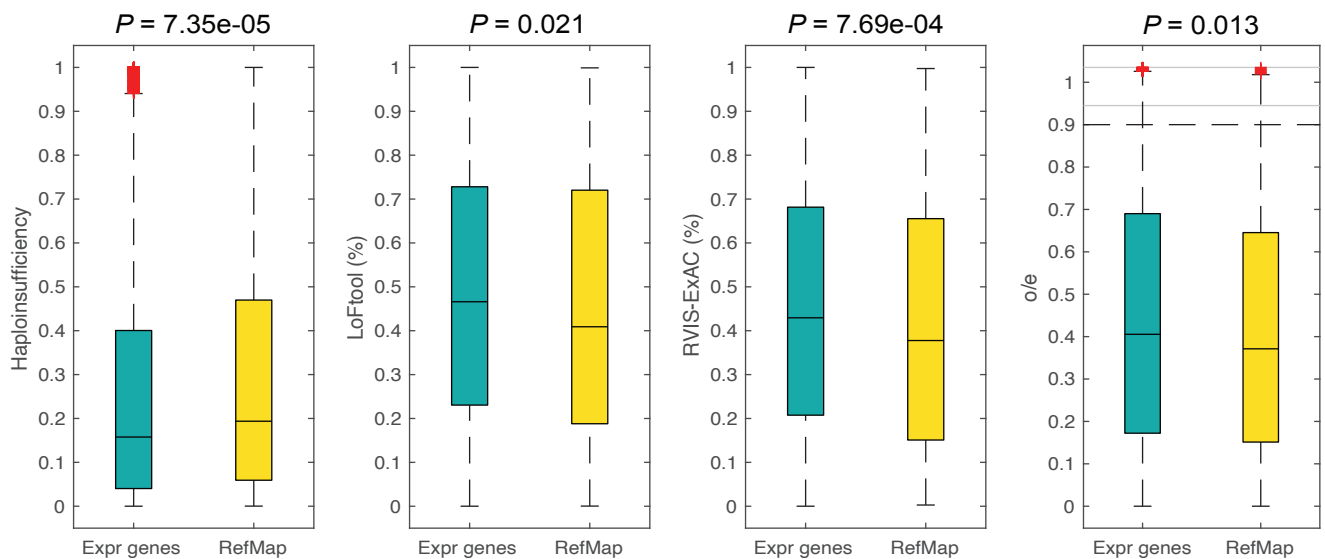
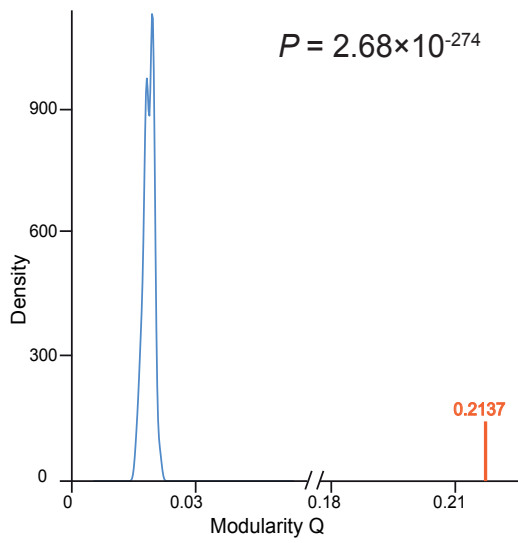


Figure S2. RefMap ALS genes are highly conserved and enriched with known ALS genes, Related to Figure 2

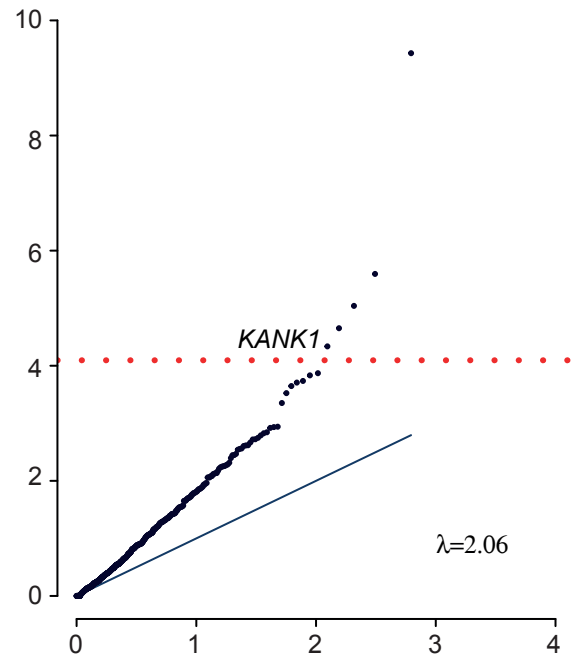
(A) Comparison of the number of discovered genes (left panel), enrichment with an curated ALS genes (middle panel), and enrichment with ClinVar ALS genes (right panel), for different methods. (B and C) Control conservation analysis. Conservation analysis based on active genes (B) with ATAC-seq peaks and expressed genes (C) in MNs. Expressed genes were defined as those with TPM>1 in our transcriptomic profiling of iPSC-derived MNs. Comparisons were performed using the one-sided Wilcoxon rank-sum test. The bottom and top of the boxes indicate the first and third quartiles, respectively, where the black line in between indicates the median. The whiskers denote the minimal value within 1.5 IQR of the lower quartile and the maximum value within 1.5 IQR of the upper quartile. Red symbols denote outliers. Black dashed lines indicate the lower and upper limits of the regions with regular scale. Outliers beyond the black dashed lines are visualized with a compressed scale in the regions denoted by gray lines.

Figure S3

A

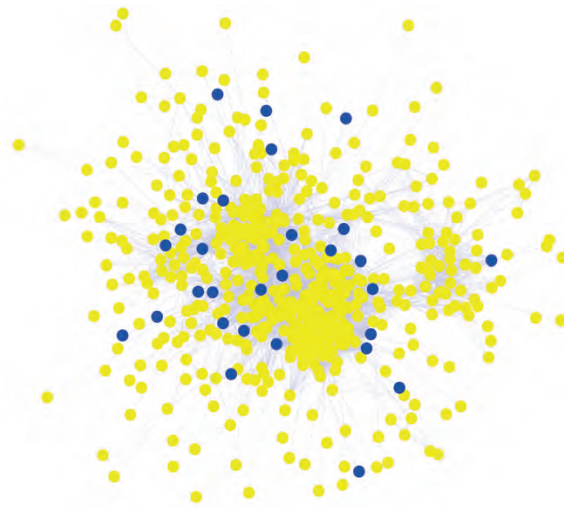


B



C

M826



D

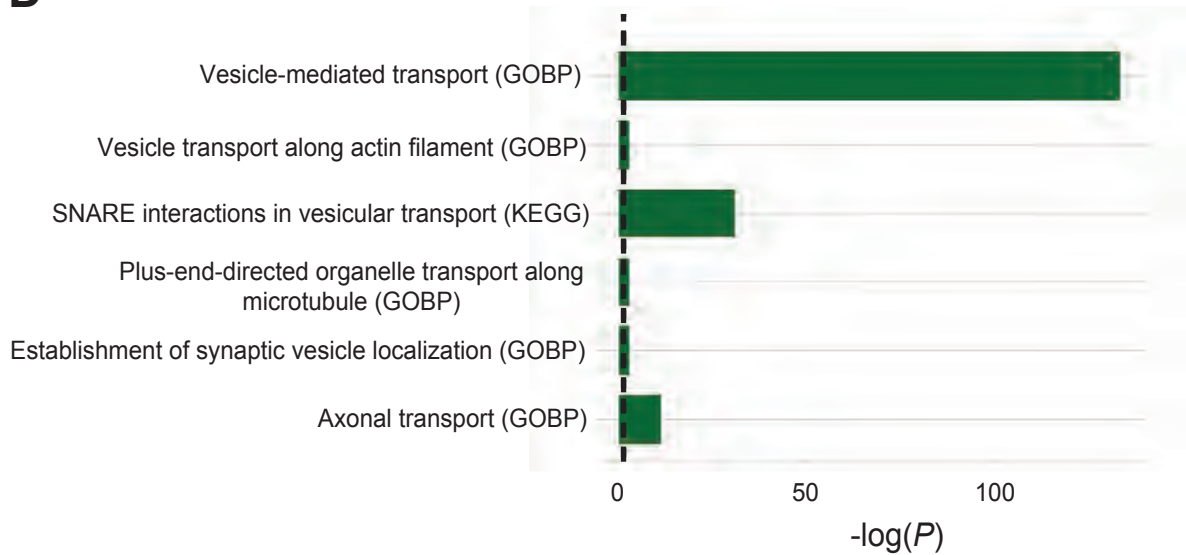
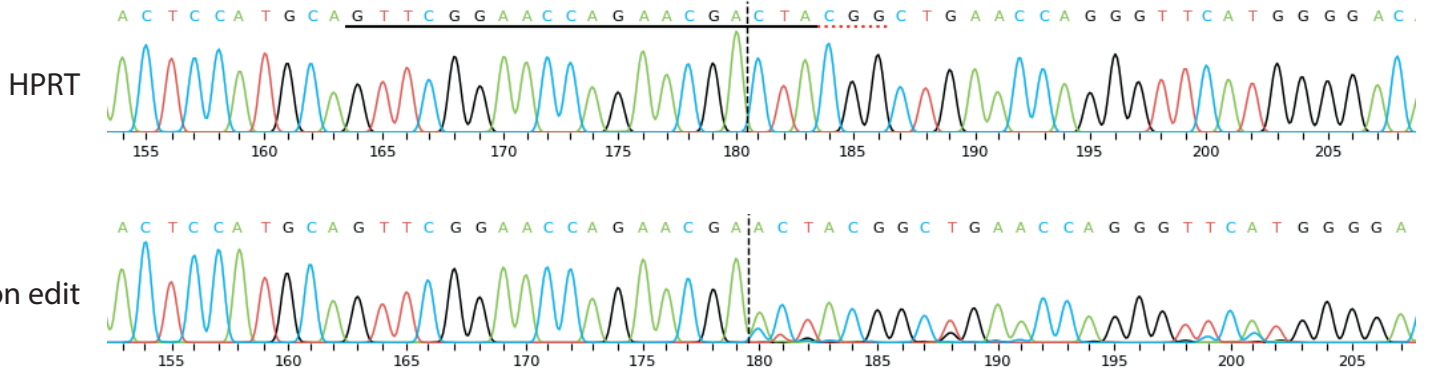


Figure S3. Additional results for network analysis and rare variant burden testing within an independent cohort, Related to Figure 4

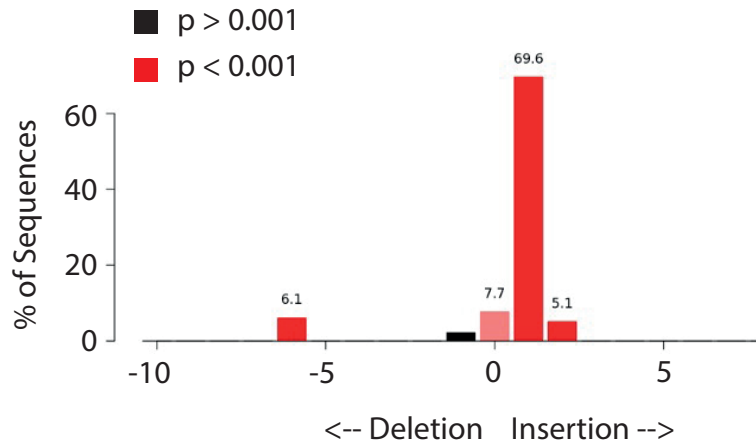
(A) Distribution of modularities after Louvain for the smoothed PPI network (red) and 100 randomized networks (blue). The modularity of our smoothed network is significantly shifted from the randomized network. (B) The set of 690 RefMap genes is enriched with ALS-associated rare missense variants in an independent cohort. Exome sequencing was conducted in 3,864 ALS patients and 7,839 controls. Rare missense and synonymous variants were identified with $MAF < 0.001\%$. Association testing was performed using Fisher's exact test. Red dotted line indicates Bonferroni multiple testing threshold. (C) M826 containing *KANK1*. M826 is enriched with RefMap genes ($P = 5.6 \times 10^{-3}$, hypergeometric test). (D) M826 is functionally enriched for vesicle transport within the motor neuron axon. GOBP, gene ontology biological process. Dashed line represents $P = 0.05$.

Figure S4

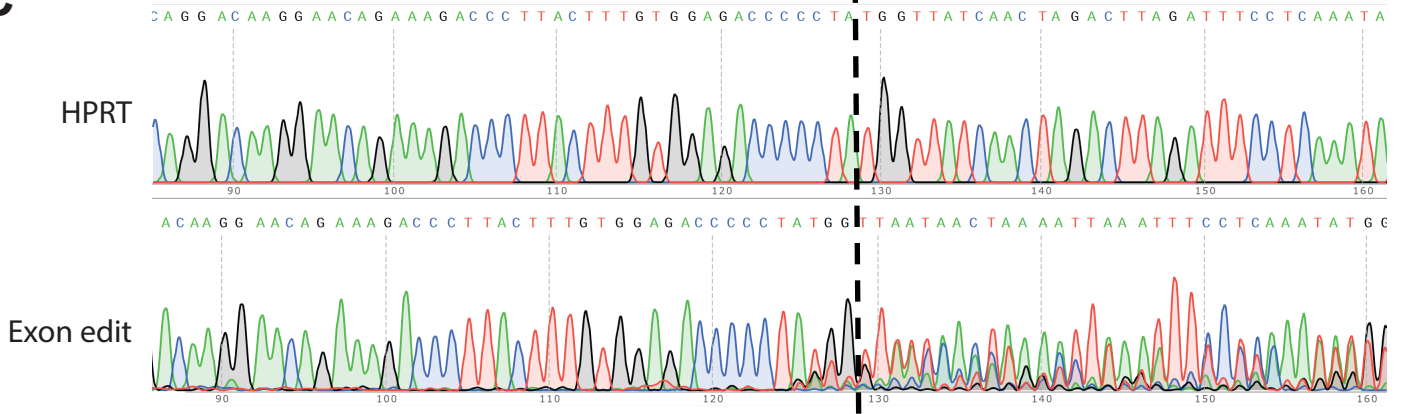
A



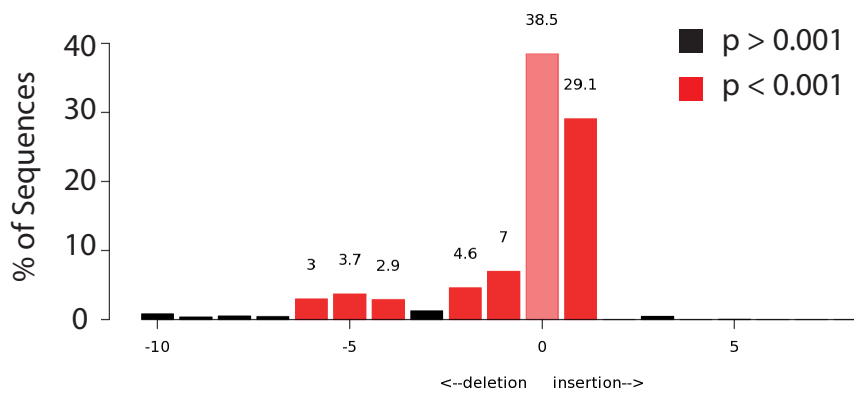
B



C



D



E

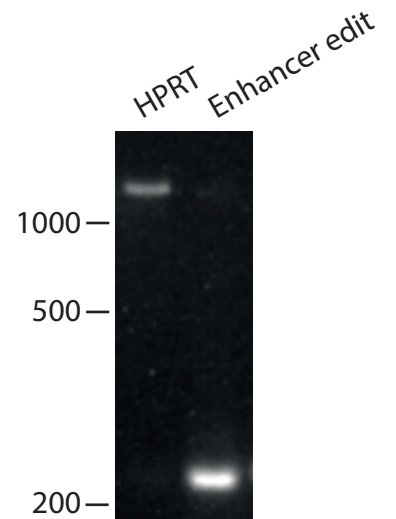
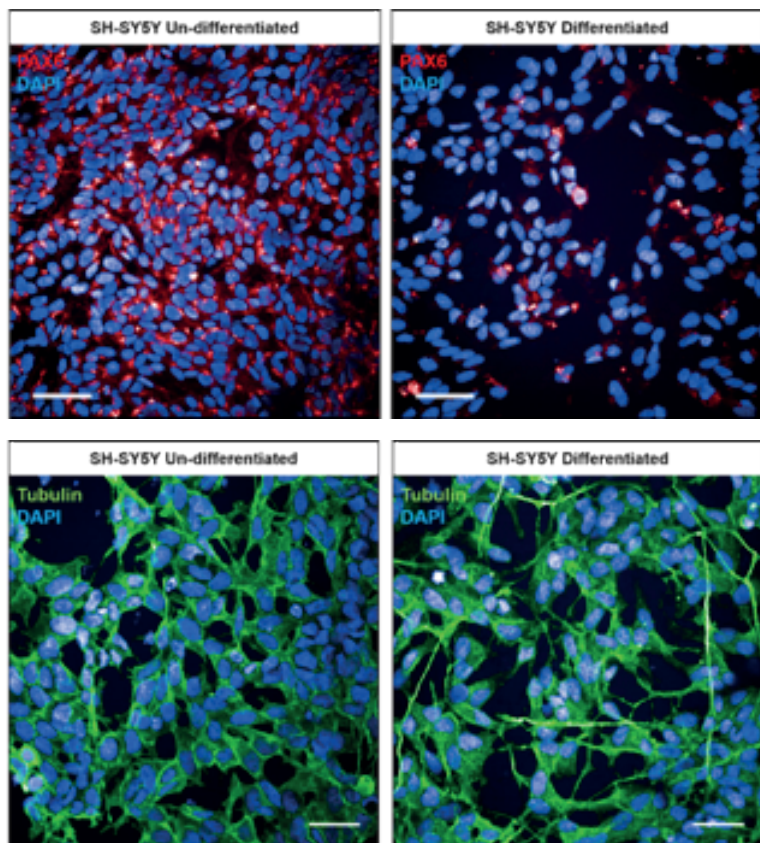


Figure S4. CRISPR-editing of *BNC2* and *KANK1* in SH-SY5Y cells, Related to Figure 6

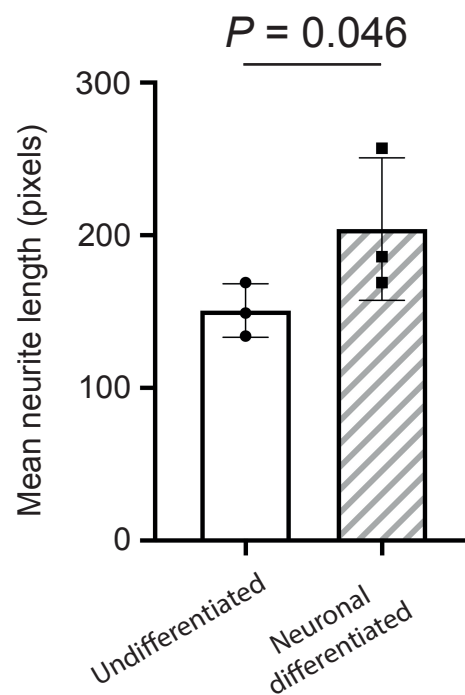
(A) Sanger sequencing traces demonstrating spCas9 cut site adjacent to PAM and subsequent waveform decomposition in *BNC2*-exon-edited cells. (B) Indel distribution of *BNC2*-exon-edited SH-SY5Y cells. (C) Sanger sequencing traces demonstrating spCas9 cut site adjacent to PAM and subsequent waveform decomposition in *KANK1*-exon-edited cells. (D) Indel distribution of *KANK1*-exon-edited SH-SY5Y cells. (E) PCR amplification of the relevant genomic segment in *KANK1*-enhancer-edited SH-SY5Y cells reveals that the chr9:663001-664000 region has been resected compared to *HPRT*-edited control cells.

Figure S5

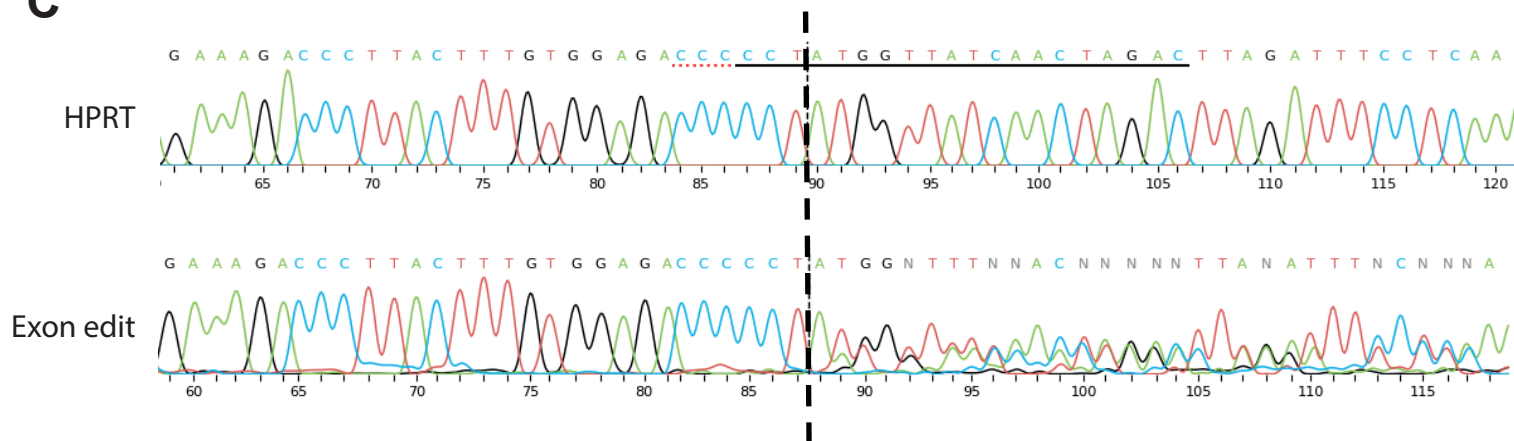
A



B



C



D

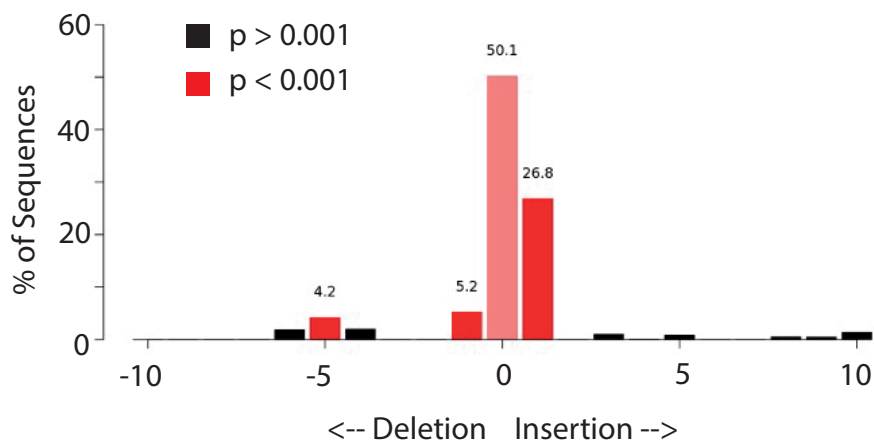


Figure S5. Neuronal differentiation of SH-SY5Y cells and CRISPR-editing of *KANK1* in iPSCs, Related to Figures 6 and 7

(A) Altered PAX6 expression and (B) increased dendrite length confirm the successful differentiation of SH-SY5Y cells. (C) Sanger sequencing traces demonstrating spCas9 cut site adjacent to PAM and subsequent waveform decomposition in *KANK1*-exon-edited cells. (D) Indel distribution of *KANK1*-exon-edited iPSCs.

Figure S6

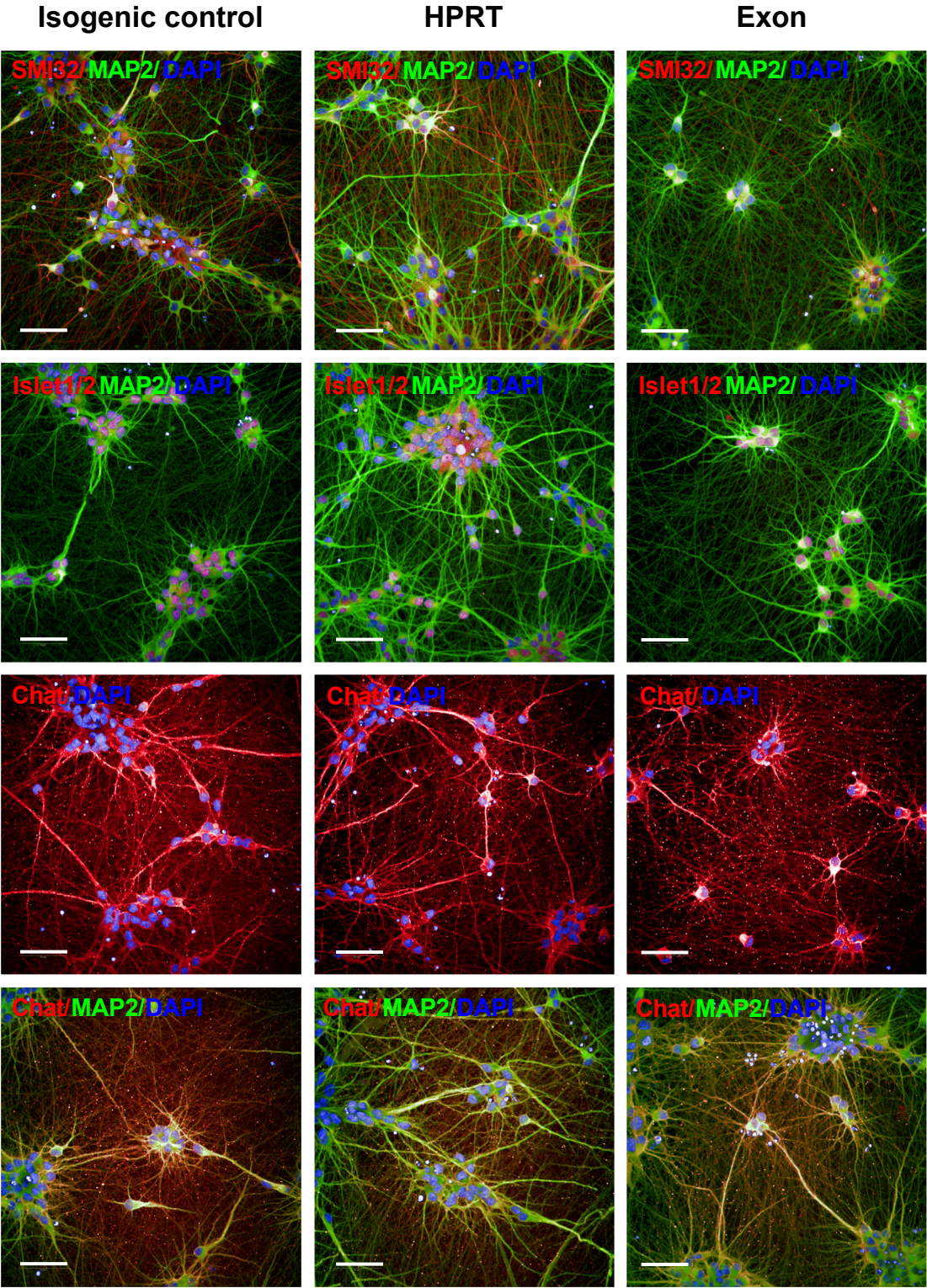
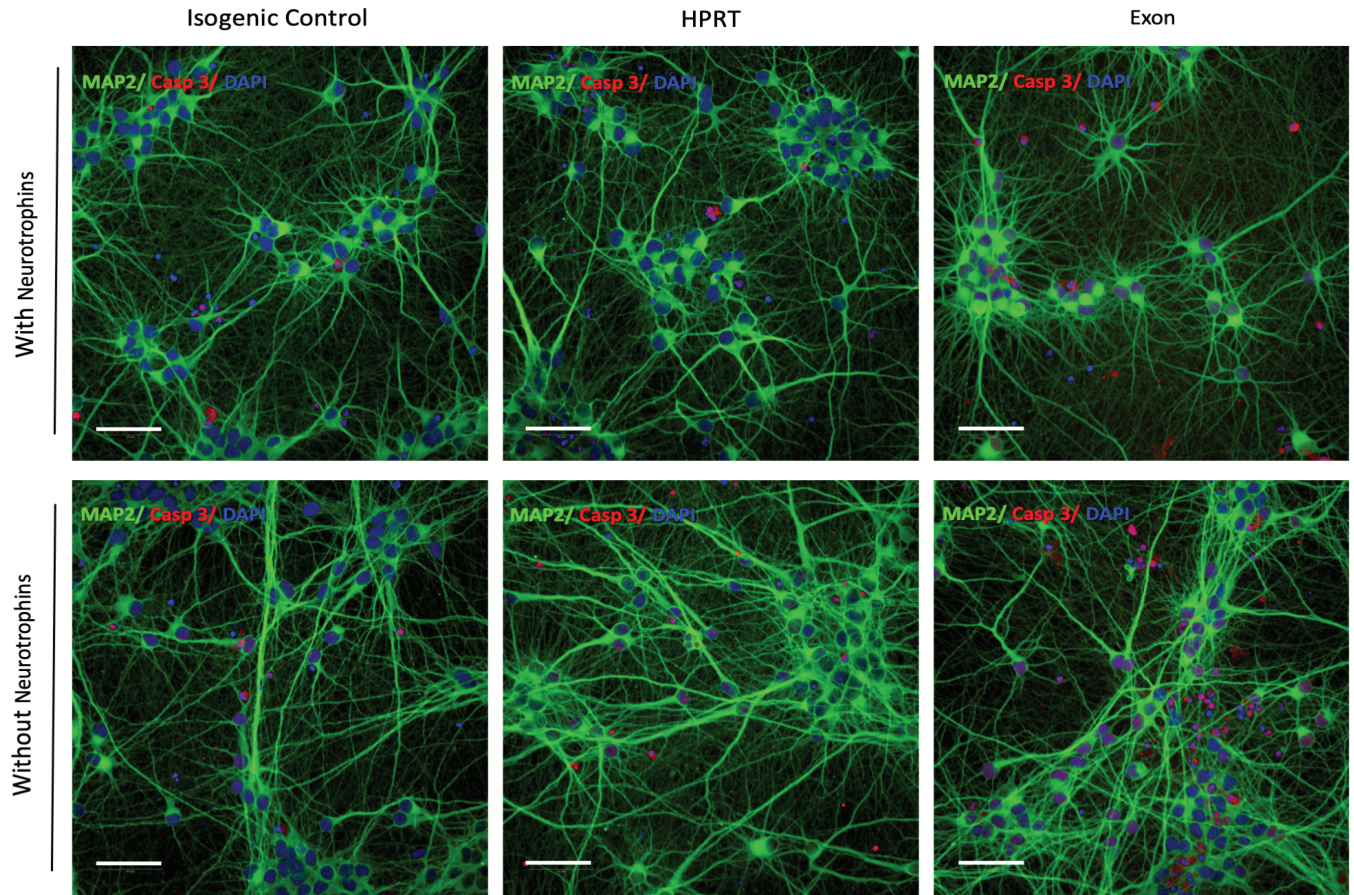


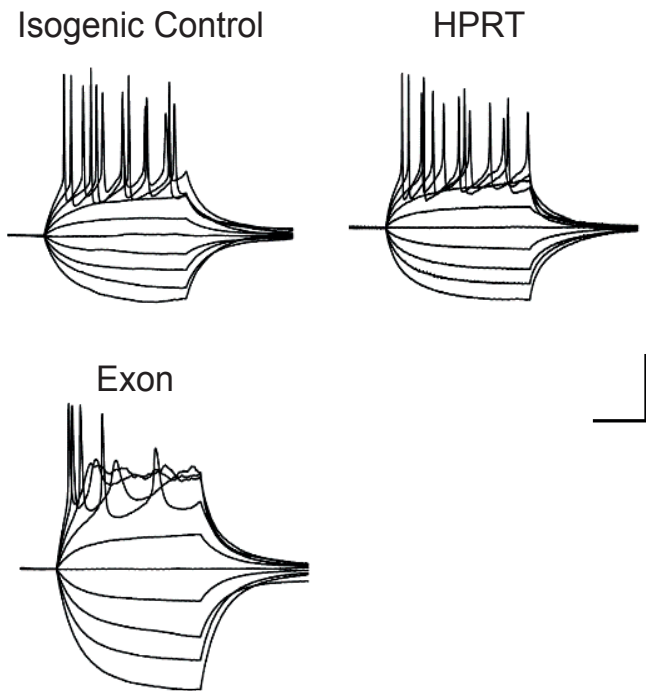
Figure S6. Differentiation of iPSCs into mature motor neurons, Related to Figure 7
Derived motor neurons expressed markers confirm the successful differentiation of iPSCs.

Figure S7

A



B



C

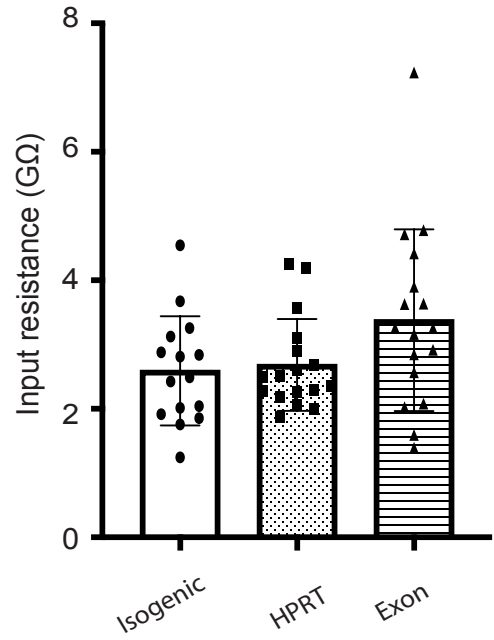


Figure S7. *KANK1*-edited motor neurons display increased apoptosis and evidence of electrophysiological dysfunction within the axon, Related to Figure 7

(A) Staining for cleaved-caspase-3 reveals excessive apoptosis in *KANK1*-edited neurons with neurotrophin withdrawal. (B) Raw traces reveal deficient axon potential firing in *KANK1*-edited motor neurons. (C) Comparison of input resistance between *KANK1*-edited motor neurons and controls.

Supplemental Note for “Genome-wide identification of the genetic basis of amyotrophic lateral sclerosis”

1 Mathematical foundation of RefMap

Here, we provide a mathematical theory to justify Eq. 1 in the Method section of the main text. To facilitate the development of the theory, we first describe a universal discriminative framework that models the relationship between the genotype and phenotype, and then deduce a general distribution over summary statistics from this framework. Based on this result, a flexible probabilistic model that characterizes summary statistics with various prior structures can be developed, which generalizes multiple previous studies (Pasaniuc and Price 2016; B. Bulik-Sullivan et al. 2015; Kichaev et al. 2014; B. K. Bulik-Sullivan et al. 2015; Joo et al. 2016; Finucane et al. 2015; Han, Kang, and Eskin 2009). In particular, Equation 1 of RefMap follows directly after assuming a linear relation between the genotype and phenotype. In the following, we will develop the framework in both cases of quantitative trait and case-control studies.

1.1 Quantitative trait studies

We start from considering a general genotype-phenotype model for continuous traits, i.e.,

$$y_n = F(\mathbf{x}_n, \mathbf{w}) + \epsilon_n, \quad n = 1, \dots, N, \quad (1)$$

in which N is the sample size, \mathbf{x}_n and y_n are the genotypes and phenotype for the n th sample, respectively, F is an unknown (usually non-linear) function with parameters \mathbf{w} determining personal phenotype from his/her genotypes, and ϵ_n is the random noise following

$$\epsilon_n \sim \mathcal{N}(0, \sigma_\epsilon^2). \quad (2)$$

Note that as a routine procedure, genotypes are first standardized by

$$x_{ni} = \frac{g_{ni} - 2p_i}{\sqrt{2p_i(1 - p_i)}}, \quad i = 1, \dots, M, \quad (3)$$

where M is the number of alleles, g_{ni} is the genotype of the i th allele for the n th sample, and p_i is the frequency of the i th allele in the study cohort. After standardization, the sample mean and sample variance of each allele are 0 and 1, respectively. Moreover, we adopt a general setting and treat both genotypes and function parameters as random variables, yielding

$$y_n \mid \mathbf{x}_n, \mathbf{w}, \sigma_\epsilon \sim \mathcal{N}(F(\mathbf{x}_n, \mathbf{w}), \sigma_\epsilon^2). \quad (4)$$

Following the conventional annotation in the genome-wide association study (GWAS), the estimated effect sizes $\hat{\beta}_i$ for individual alleles are the most widely-used summary statistics, which are closely related to χ^2 and Z -score. Given the genotype standardization, we have

$$\hat{\beta}_i = \frac{\mathbf{x}_i^\top \mathbf{y}}{N}, \quad (5)$$

where \mathbf{x}_i is the genotype vector for the i th allele and $\mathbf{y} = y_{1:N}$. With matrix representation, we have

$$\hat{\boldsymbol{\beta}} = \frac{1}{N} \mathbf{X}^\top \mathbf{y} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n y_n, \quad (6)$$

where $\mathbf{X} = (x_{ni}) \in \mathbb{R}^{N \times M}$. Indeed, we have the following theorem characterizing the asymptotic distribution of $\sqrt{N} \hat{\boldsymbol{\beta}}$.

Theorem 1. *Given the definitions in Eqs. 1, 2 and 5, when the sample size N is large enough, we have*

$$\sqrt{N} \hat{\boldsymbol{\beta}} \mid \mathbf{X}, \mathbf{w}, \sigma_\epsilon \sim \mathcal{N} \left(\sqrt{N} \boldsymbol{\mu}(\mathbf{X}, F, \mathbf{w}), \sigma_\epsilon^2 \boldsymbol{\Sigma}_{\text{LD}} \right), \quad (7)$$

where $\boldsymbol{\Sigma}_{\text{LD}}$ is the in-sample linkage disequilibrium (LD) matrix quantifying SNP correlations, and $\boldsymbol{\mu}(\mathbf{X}, F, \mathbf{w})$ is a quantity depending on the genotypes and the discriminative function F .

Proof. We first show that $\sqrt{N} \hat{\boldsymbol{\beta}}$ follows a normal distribution asymptotically. In fact, according to Eq. 6, given the genotypes and the discriminative function, $\sqrt{N} \hat{\boldsymbol{\beta}}$ can be computed by the sum of $\mathbf{x}_n y_n$, which are independent with each other but with different expectations. On the other hand, the variance of $\mathbf{x}_n y_n$ is given by

$$\begin{aligned} \text{Var} [\mathbf{x}_n y_n \mid \mathbf{x}_n, \mathbf{w}, \sigma_\epsilon] &= \text{Var} [\mathbf{x}_n (F(\mathbf{x}_n, \mathbf{w}) + \epsilon_n) \mid \mathbf{x}_n, \mathbf{w}, \sigma_\epsilon] \\ &= \text{Var} [\mathbf{x}_n \epsilon_n \mid \mathbf{x}_n, \mathbf{w}, \sigma_\epsilon] \\ &= \mathbb{E} [\epsilon_n^2 \mathbf{x}_n \mathbf{x}_n^\top \mid \mathbf{x}_n, \mathbf{w}, \sigma_\epsilon] \\ &= \mathbf{x}_n \mathbf{x}_n^\top \sigma_\epsilon^2, \end{aligned} \quad (8)$$

yielding

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n \mathbf{x}_n^\top \sigma_\epsilon^2 &= \lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{X}^\top \mathbf{X} \cdot \sigma_\epsilon^2 \\ &= \sigma_\epsilon^2 \hat{\boldsymbol{\Sigma}}_{\text{LD}} \\ &\approx \sigma_\epsilon^2 \boldsymbol{\Sigma}_{\text{LD}}, \end{aligned} \quad (9)$$

in which the estimated LD matrix $\hat{\boldsymbol{\Sigma}}_{\text{LD}} = (\hat{r}_{ij})$ is given by

$$\hat{r}_{ij} = \frac{\mathbf{x}_i^\top \mathbf{x}_j}{\sqrt{\mathbf{x}_i^\top \mathbf{x}_i} \sqrt{\mathbf{x}_j^\top \mathbf{x}_j}} = \frac{1}{N} \mathbf{x}_i^\top \mathbf{x}_j, \quad (10)$$

and the last approximation is guaranteed by $\mathbb{E}[\hat{r}_{ij}] = r_{ij} = \mathbb{E}[x_i x_j]$. Therefore, according to the multivariate Lindeberg-Feller central limit theorem (CLT), we conclude that $\sqrt{N} \hat{\boldsymbol{\beta}} =$

$1/\sqrt{N} \sum_{i=1}^N \mathbf{x}_n y_n$ asymptotically follows a normal distribution with covariance $\sigma_\epsilon^2 \boldsymbol{\Sigma}_{LD}$, whose expectation is given by

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{n=1}^N \mathbb{E}[\mathbf{x}_n y_n \mid \mathbf{x}_n, \mathbf{w}, \sigma_\epsilon] &= \frac{1}{\sqrt{N}} \sum_{n=1}^N \mathbb{E}[\mathbf{x}_n (F(\mathbf{x}_n, \mathbf{w}) + \epsilon_n) \mid \mathbf{x}_n, \mathbf{w}, \sigma_\epsilon] \\ &= \frac{1}{\sqrt{N}} \sum_{n=1}^N \mathbf{x}_n F(\mathbf{x}_n, \mathbf{w}) \\ &= \sqrt{N} \boldsymbol{\mu}(\mathbf{X}, F, \mathbf{w}), \end{aligned} \quad (11)$$

where $\boldsymbol{\mu}(\cdot)$ is defined as

$$\boldsymbol{\mu}(\mathbf{X}, F, \mathbf{w}) = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n F(\mathbf{x}_n, \mathbf{w}). \quad (12)$$

This completes the proof. \square

Note that if we use Z -scores computed by GWAS as the approximation of $\sqrt{N} \hat{\boldsymbol{\beta}}/\sigma_\epsilon$, i.e., dividing $\hat{\beta}_i$ by its estimated standard error, we have

$$\hat{\mathbf{z}} \mid \mathbf{X}, \mathbf{w} \sim \mathcal{N}\left(\sqrt{N} \boldsymbol{\mu}(\mathbf{X}, F, \mathbf{w}), \boldsymbol{\Sigma}_{LD}\right), \quad (13)$$

in which σ_ϵ is absorbed into $\boldsymbol{\mu}(\cdot)$ for annotation brevity.

1.2 Case-control studies

We state the analysis for case-control studies using a Bernoulli distribution over case-control status, i.e.,

$$y_n \mid \pi_n \sim \text{Bernoulli}(\pi_n), \quad n = 1, \dots, N, \quad (14)$$

whose logit is defined similarly as Eq. 1 but without random noise, i.e.,

$$\log \frac{\pi_n}{1 - \pi_n} = F(\mathbf{x}_n, \mathbf{w}). \quad (15)$$

After a few calculations we can easily get

$$\pi_n = \sigma(F(\mathbf{x}_n, \mathbf{w})), \quad (16)$$

where $\sigma(\cdot)$ is the sigmoid function defined by $\sigma(x) = 1/(1 + \exp(-x))$.

To facilitate the following analysis, here we illustrate the standardization procedure in more detail, i.e.,

$$x_{ni} = \frac{g_{ni} - 2\hat{p}_i}{\sqrt{2\hat{p}_i(1 - \hat{p}_i)}}, \quad (17)$$

where g_{ni} is the genotype coded by 0, 1 and 2, and \hat{p}_i is the in-sample allele frequency. Therefore, suppose we have the same number ($N/2$) of cases and controls in the study cohort, the widely-used Z -scores for case-control studies defined as

$$\hat{\mathbf{z}}_i = \frac{\sqrt{N}(\hat{p}_i^+ - \hat{p}_i^-)}{\sqrt{2\hat{p}_i(1 - \hat{p}_i)}} \quad (i = 1, \dots, M) \quad (18)$$

can be written as

$$\hat{\mathbf{z}} = \frac{1}{\sqrt{N}} \sum_{n=1}^N (2\mathbb{1}\{y_n = 1\} - 1)\mathbf{x}_n. \quad (19)$$

Again, utilizing the multivariate Lindeberg-Feller CLT, we can derive the asymptotic conditional distribution of $\hat{\mathbf{z}}$, which is approximately the same as that in the quantitative trait studies (Eq. 13). In particular, we have the following result.

Theorem 2. *Given the definitions in Eqs. 14, 15 and 19, when the sample size N is large enough, we have*

$$\hat{\mathbf{z}} \mid \mathbf{X}, \mathbf{w} \sim \mathcal{N}\left(\sqrt{N}\boldsymbol{\mu}(\mathbf{X}, F, \mathbf{w}), \boldsymbol{\Sigma}_{\text{LD}}\right), \quad (20)$$

where $\boldsymbol{\Sigma}_{\text{LD}}$ is the in-sample LD matrix, and $\boldsymbol{\mu}(\mathbf{X}, F, \mathbf{w})$ is a quantity depending on the genotypes and the discriminative function.

Proof. Conditioned on \mathbf{X} and \mathbf{w} , the variance of $(2\mathbb{1}\{y_n = 1\} - 1)\mathbf{x}_n$ can be calculated as

$$\begin{aligned} \text{Var}[(2\mathbb{1}\{y_n = 1\} - 1)\mathbf{x}_n] &= \mathbf{x}_n \mathbf{x}_n^\top - \mathbb{E}[2\mathbb{1}\{y_n = 1\} - 1]^2 \mathbf{x}_n \mathbf{x}_n^\top \\ &= 4\mathbf{x}_n \mathbf{x}_n^\top \mathbb{P}[y_n = 1] (1 - \mathbb{P}[y_n = 1]) \\ &= 4\mathbf{x}_n \mathbf{x}_n^\top \text{Var}[y_n], \end{aligned} \quad (21)$$

where the conditions are omitted for brevity. In fact, as $\text{Var}[y_n \mid \mathbf{x}_n, \mathbf{w}] < 1$, we conclude that the average of variance $1/N \sum_{n=1}^N \text{Var}[(2\mathbb{1}\{y_n = 1\} - 1)\mathbf{x}_n \mid \mathbf{x}_n, \mathbf{w}]$ converges as $N \rightarrow \infty$, whose limit is denoted as $\boldsymbol{\Sigma}_\infty$. According to the multivariate Lindeberg-Feller CLT, the asymptotic conditional distribution of $\hat{\mathbf{z}}$ is a normal distribution with covariance matrix $\boldsymbol{\Sigma}_\infty$.

To get a clearer structure of $\boldsymbol{\Sigma}_\infty$, we now apply a few approximations for Eq. 21. In particular, we have

$$\begin{aligned} \boldsymbol{\Sigma}_\infty &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \text{Var}[(2\mathbb{1}\{y_n = 1\} - 1)\mathbf{x}_n \mid \mathbf{x}_n, \mathbf{w}] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N 4\mathbf{x}_n \mathbf{x}_n^\top \text{Var}[y_n \mid \mathbf{x}_n, \mathbf{w}] \\ &= \mathbb{E}[4\mathbf{x}_n \mathbf{x}_n^\top \text{Var}[y_n \mid \mathbf{x}_n, \mathbf{w}]] \\ &\approx 4\boldsymbol{\Sigma}_{\text{LD}} \mathbb{E}[\text{Var}[y_n \mid \mathbf{x}_n, \mathbf{w}]] \\ &= 4\boldsymbol{\Sigma}_{\text{LD}} (\text{Var}[y_n \mid \mathbf{w}] - \text{Var}[\mathbb{E}[y_n \mid \mathbf{x}_n, \mathbf{w}]]) \\ &< 4\boldsymbol{\Sigma}_{\text{LD}} \text{Var}[y_n \mid \mathbf{w}] \\ &\approx 4\boldsymbol{\Sigma}_{\text{LD}} \mathbb{E}[\text{Var}[y_n \mid \mathbf{w}]] \\ &= 4\boldsymbol{\Sigma}_{\text{LD}} (\text{Var}[y_n] - \text{Var}[\mathbb{E}[y_n \mid \mathbf{w}]]) \\ &< 4\boldsymbol{\Sigma}_{\text{LD}} \text{Var}[y_n] \\ &= \boldsymbol{\Sigma}_{\text{LD}}, \end{aligned} \quad (22)$$

in which the third and the fourth “=” come from the law of total variance, the first and the second “<” are implied by the positivity of variance, and $\boldsymbol{\Sigma}_{\text{LD}}$ is the in-sample LD matrix. For the last “=”, we argue that the expectation and variance in Eq. 22 are taken over the sampling space in case-control studies, rather than the general population. Under

the assumption of equal number of cases and controls, the sampling disease prevalence is 0.5, yielding $\text{Var}[y_n] = 0.25$.

Furthermore, the expectation of the asymptotic conditional distribution can be calculated as

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{n=1}^N \mathbb{E}[(2\mathbb{1}\{y_n = 1\} - 1)\mathbf{x}_n \mid \mathbf{x}_n, \mathbf{w}] &= \frac{1}{\sqrt{N}} \sum_{n=1}^N \mathbf{x}_n (2\sigma(F(\mathbf{x}_n, \mathbf{w})) - 1) \\ &= \sqrt{N} \boldsymbol{\mu}(\mathbf{X}, F, \mathbf{w}), \end{aligned} \quad (23)$$

where we define

$$\boldsymbol{\mu}(\mathbf{X}, F, \mathbf{w}) = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n (2\sigma(F(\mathbf{x}_n, \mathbf{w})) - 1). \quad (24)$$

This completes the proof. \square

1.3 A linear model for RefMap

We consider a linear model that underlies the design of RefMap. Specifically, in the quantitative trait studies, we define

$$F(\mathbf{x}_n, \mathbf{w}) = w_0 + \sum_{i=1}^M w_i x_{ni}. \quad (25)$$

Note that this linear model has been widely used in traditional GWAS studies (B. Bulik-Sullivan et al. 2015; B. K. Bulik-Sullivan et al. 2015; Finucane et al. 2015), and w_i is called the *effect size* of the i th allele. The linear model for case-control studies can be developed similarly by considering the approximation of sigmoid function using its Taylor expansion. Therefore, the expectation of the asymptotic distribution of Z -scores can be calculated as

$$\begin{aligned} \sqrt{N} \boldsymbol{\mu}(\mathbf{X}, F, \mathbf{w}) &= \frac{1}{\sqrt{N}} \sum_{n=1}^N \mathbf{x}_n (\mathbf{x}_n^\top \mathbf{w} + w_0) \\ &= \sqrt{N} \hat{\boldsymbol{\Sigma}}_{\text{LD}} \mathbf{w}, \end{aligned} \quad (26)$$

indicating that the expected Z -score for each allele is determined by its effect size as well as its strongly-associated neighbors. By absorbing \sqrt{N} into \mathbf{w} , we eventually get Eq. 1 in the RefMap model.

2 Inference for RefMap

The RefMap model was defined in Eqs. 1 to 18 in the Method section of the main text. Here, we are interested in the posterior $p(\mathbf{T} \mid \mathbf{Z}, \mathbf{S})$, whose exact calculation is intractable. Therefore, we seek for approximate inference based on the mean-field variational inference (MFVI). Basically, we first assume that the approximate posterior over latent variables factorizes, indicating conditional independence across latent variables, and then perform approximate

inference by optimizing the *evidence lower bound* (ELBO) with respect to factorized proposal distributions, i.e.,

$$q(\lambda_{j,k}, \lambda, \boldsymbol{\tau}, \mathbf{v}, \mathbf{w}, \mathbf{M}, \mathbf{T}, \mathbf{U}, \boldsymbol{\Lambda}) = \max_q \mathbb{E}_q \left[\log \left(\frac{p(\mathbf{Z}, \lambda_{j,k}, \lambda, \boldsymbol{\tau}, \mathbf{v}, \mathbf{w}, \mathbf{M}, \mathbf{T}, \mathbf{U}, \boldsymbol{\Lambda} \mid \mathbf{S})}{q(\lambda_{j,k}, \lambda, \boldsymbol{\tau}, \mathbf{v}, \mathbf{w}, \mathbf{M}, \mathbf{T}, \mathbf{U}, \boldsymbol{\Lambda})} \right) \right], \quad (27)$$

which can be shown to be equivalent to minimizing the Kullback-Leibler (KL) divergence between the true posterior and its proposal.

In the following, we will first introduce several specific techniques we used in MFVI, and then summarize the update rules for different variational parameters. At last, a coordinate ascent-based VI algorithm will be given.

2.1 Rectification nonlinearity

We impose non-negativity on v_{-1} and v_{+1} using the technique of rectification nonlinearity proposed in Harva and Kabán (2007). This technique relaxes the sparsity constraint over factors and meanwhile enjoys tractable variational inference.

We first note that the approximate posterior $q(r_{-1})$ from MFVI follows the free-form solution

$$q(r_{-1}) = \frac{1}{\tilde{Z}_{-1}} \prod_{k=1}^K \prod_{j=1}^{J_k} \mathcal{N}(\mathbb{E}[m_{j,k}] \mid -v_{-1}, \mathbb{E}[\tau_{-1}]^{-1})^{\mathbb{E}[t_{j,k}^{(-1)}]} \times \mathcal{N}(r_{-1} \mid \mathbb{E}[m_{-1}], \mathbb{E}[\lambda_{-1}]^{-1}), \quad (28)$$

where \tilde{Z}_{-1} is the normalization term to be computed later. Moreover, it can be easily shown that Eq. 28 can be written as $q(r_{-1}) = q_p(r_{-1}) + q_n(r_{-1})$ with the form

$$q_p(r_{-1}) = \frac{\tilde{w}_p^{(-1)}}{\tilde{Z}_{-1}} \mathcal{N}\left(r_{-1} \mid \tilde{\mu}_p^{(-1)}, \left(\tilde{\lambda}_p^{(-1)}\right)^{-1}\right) u(r_{-1}), \quad (29)$$

$$q_n(r_{-1}) = \frac{\tilde{w}_n^{(-1)}}{\tilde{Z}_{-1}} \mathcal{N}\left(r_{-1} \mid \tilde{\mu}_n^{(-1)}, \left(\tilde{\lambda}_n^{(-1)}\right)^{-1}\right) u(-r_{-1}), \quad (30)$$

in which

$$\tilde{\mu}_p^{(-1)} = \left(-\mathbb{E}[\tau_{-1}] \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbb{E}[t_{j,k}^{(-1)}] \mathbb{E}[m_{j,k}] + \mathbb{E}[\lambda_{-1}] \mathbb{E}[m_{-1}] \right) \left(\tilde{\lambda}_p^{(-1)} \right)^{-1}, \quad (31)$$

$$\tilde{\mu}_n^{(-1)} = \mathbb{E}[m_{-1}], \quad (32)$$

$$\tilde{\lambda}_p^{(-1)} = \mathbb{E}[\tau_{-1}] \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbb{E}[t_{j,k}^{(-1)}] + \mathbb{E}[\lambda_{-1}], \quad (33)$$

$$\tilde{\lambda}_n^{(-1)} = \mathbb{E}[\lambda_{-1}], \quad (34)$$

and $u(\cdot)$ is the standard step function. With Eqs. 31 to 34, $\tilde{w}_p^{(-1)}$ and $\tilde{w}_n^{(-1)}$ can be computed by integrating Eqs. 28, 29 and 30 with respect to r_{-1} . Then the normalization term is given

by

$$\tilde{Z}_{-1} = \frac{\tilde{w}_n^{(-1)}}{2} \operatorname{erfc} \left(\tilde{\mu}_n^{(-1)} \sqrt{\tilde{\lambda}_n^{(-1)}/2} \right) + \frac{\tilde{w}_p^{(-1)}}{2} \operatorname{erfc} \left(-\tilde{\mu}_p^{(-1)} \sqrt{\tilde{\lambda}_p^{(-1)}/2} \right). \quad (35)$$

The moments for posteriors are obtained by

$$\mathbb{E}[r_{-1}] = \tilde{M}_p^{(-1)} + \tilde{M}_n^{(-1)}, \quad (36)$$

$$\mathbb{E}[r_{-1}^2] = \tilde{M}_p^{(-2)} + \tilde{M}_n^{(-2)}, \quad (37)$$

$$\mathbb{E}[v_{-1}] = \tilde{M}_p^{(-1)}, \quad (38)$$

$$\mathbb{E}[v_{-1}^2] = \tilde{M}_p^{(-2)}, \quad (39)$$

where

$$\tilde{M}_p^{(-0)} = \frac{\tilde{w}_p^{(-1)}}{2\tilde{Z}_{-1}} \operatorname{erfc} \left(-\tilde{\mu}_p^{(-1)} \sqrt{\tilde{\lambda}_p^{(-1)}/2} \right), \quad (40)$$

$$\tilde{M}_p^{(-1)} = \frac{\tilde{w}_p^{(-1)}}{2\tilde{Z}_{-1}} \left\{ \operatorname{erfc} \left(-\tilde{\mu}_p^{(-1)} \sqrt{\tilde{\lambda}_p^{(-1)}/2} \right) \tilde{\mu}_p^{(-1)} + \sqrt{\frac{2}{\pi \tilde{\lambda}_p^{(-1)}}} \frac{1}{\exp \left(\tilde{\lambda}_p^{(-1)} \left(\tilde{\mu}_p^{(-1)} \right)^2 / 2 \right)} \right\}, \quad (41)$$

$$\tilde{M}_p^{(-2)} = \frac{\tilde{w}_p^{(-1)}}{2\tilde{Z}_{-1}} \left\{ \operatorname{erfc} \left(-\tilde{\mu}_p^{(-1)} \sqrt{\tilde{\lambda}_p^{(-1)}/2} \right) \left(\left(\tilde{\mu}_p^{(-1)} \right)^2 + \frac{1}{\tilde{\lambda}_p^{(-1)}} \right) + \sqrt{\frac{2}{\pi \tilde{\lambda}_p^{(-1)}}} \frac{\tilde{\mu}_p^{(-1)}}{\exp \left(\tilde{\lambda}_p^{(-1)} \left(\tilde{\mu}_p^{(-1)} \right)^2 / 2 \right)} \right\}, \quad (42)$$

$$\tilde{M}_n^{(-0)} = \frac{\tilde{w}_n^{(-1)}}{2\tilde{Z}_{-1}} \operatorname{erfc} \left(\tilde{\mu}_n^{(-1)} \sqrt{\tilde{\lambda}_n^{(-1)}/2} \right), \quad (43)$$

$$\tilde{M}_n^{(-1)} = \frac{\tilde{w}_n^{(-1)}}{2\tilde{Z}_{-1}} \left\{ \operatorname{erfc} \left(\tilde{\mu}_n^{(-1)} \sqrt{\tilde{\lambda}_n^{(-1)}/2} \right) \tilde{\mu}_n^{(-1)} - \sqrt{\frac{2}{\pi \tilde{\lambda}_n^{(-1)}}} \frac{1}{\exp \left(\tilde{\lambda}_n^{(-1)} \left(\tilde{\mu}_n^{(-1)} \right)^2 / 2 \right)} \right\}, \quad (44)$$

$$\tilde{M}_n^{(-2)} = \frac{\tilde{w}_n^{(-1)}}{2\tilde{Z}_{-1}} \left\{ \operatorname{erfc} \left(\tilde{\mu}_n^{(-1)} \sqrt{\tilde{\lambda}_n^{(-1)}/2} \right) \left(\left(\tilde{\mu}_n^{(-1)} \right)^2 + \frac{1}{\tilde{\lambda}_n^{(-1)}} \right) - \sqrt{\frac{2}{\pi \tilde{\lambda}_n^{(-1)}}} \frac{\tilde{\mu}_n^{(-1)}}{\exp \left(\tilde{\lambda}_n^{(-1)} \left(\tilde{\mu}_n^{(-1)} \right)^2 / 2 \right)} \right\}. \quad (45)$$

Similar to $q(r_{-1})$, the posterior $q(r_{+1})$ also follows a free-form solution given by

$$q(r_{+1}) = \frac{1}{\tilde{Z}_{+1}} \prod_{k=1}^K \prod_{j=1}^{J_k} \mathcal{N} \left(\mathbb{E}[m_{j,k}] \mid v_{+1}, \mathbb{E}[\tau_{+1}]^{-1} \right)^{\mathbb{E}[t_{j,k}^{(+1)}]} \times \mathcal{N} \left(r_{+1} \mid \mathbb{E}[m_{+1}], \mathbb{E}[\lambda_{+1}]^{-1} \right), \quad (46)$$

where \tilde{Z}_{+1} is the normalization term. Equation 46 can also be written as $q(r_{+1}) = q_p(r_{+1}) +$

$q_n(r_{+1})$ with the form

$$q_p(r_{+1}) = \frac{\tilde{w}_p^{(+1)}}{\tilde{Z}_{+1}} \mathcal{N} \left(r_{+1} \mid \tilde{\mu}_p^{(+1)}, \left(\tilde{\lambda}_p^{(+1)} \right)^{-1} \right) u(r_{+1}), \quad (47)$$

$$q_n(r_{+1}) = \frac{\tilde{w}_n^{(+1)}}{\tilde{Z}_{+1}} \mathcal{N} \left(r_{+1} \mid \tilde{\mu}_n^{(+1)}, \left(\tilde{\lambda}_n^{(+1)} \right)^{-1} \right) u(-r_{+1}), \quad (48)$$

in which

$$\tilde{\mu}_p^{(+1)} = \left(\mathbb{E}[\tau_{+1}] \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbb{E} \left[t_{j,k}^{(+1)} \right] \mathbb{E}[m_{j,k}] + \mathbb{E}[\lambda_{+1}] \mathbb{E}[m_{+1}] \right) \left(\tilde{\lambda}_p^{(+1)} \right)^{-1}, \quad (49)$$

$$\tilde{\mu}_n^{(+1)} = \mathbb{E}[m_{+1}], \quad (50)$$

$$\tilde{\lambda}_p^{(+1)} = \mathbb{E}[\tau_{+1}] \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbb{E} \left[t_{j,k}^{(+1)} \right] + \mathbb{E}[\lambda_{+1}], \quad (51)$$

$$\tilde{\lambda}_n^{(+1)} = \mathbb{E}[\lambda_{+1}]. \quad (52)$$

After computing $\tilde{w}_p^{(+1)}$ and $\tilde{w}_n^{(+1)}$, the normalization term is given by

$$\tilde{Z}_{+1} = \frac{\tilde{w}_n^{(+1)}}{2} \operatorname{erfc} \left(\tilde{\mu}_n^{(+1)} \sqrt{\tilde{\lambda}_n^{(+1)}/2} \right) + \frac{\tilde{w}_p^{(+1)}}{2} \operatorname{erfc} \left(-\tilde{\mu}_p^{(+1)} \sqrt{\tilde{\lambda}_p^{(+1)}/2} \right). \quad (53)$$

The moments for posteriors are obtained by

$$\mathbb{E}[r_{+1}] = \tilde{M}_p^{(+1)} + \tilde{M}_n^{(+1)}, \quad (54)$$

$$\mathbb{E}[r_{+1}^2] = \tilde{M}_p^{(+2)} + \tilde{M}_n^{(+2)}, \quad (55)$$

$$\mathbb{E}[v_{+1}] = \tilde{M}_p^{(+1)}, \quad (56)$$

$$\mathbb{E}[v_{+1}^2] = \tilde{M}_p^{(+2)}, \quad (57)$$

in which

$$\tilde{M}_p^{(+0)} = \frac{\tilde{w}_p^{(+1)}}{2\tilde{Z}_{+1}} \operatorname{erfc} \left(-\tilde{\mu}_p^{(+1)} \sqrt{\tilde{\lambda}_p^{(+1)}/2} \right), \quad (58)$$

$$\tilde{M}_p^{(+1)} = \frac{\tilde{w}_p^{(+1)}}{2\tilde{Z}_{+1}} \left\{ \operatorname{erfc} \left(-\tilde{\mu}_p^{(+1)} \sqrt{\tilde{\lambda}_p^{(+1)}/2} \right) \tilde{\mu}_p^{(+1)} + \sqrt{\frac{2}{\pi \tilde{\lambda}_p^{(+1)}}} \frac{1}{\exp \left(\tilde{\lambda}_p^{(+1)} \left(\tilde{\mu}_p^{(+1)} \right)^2 / 2 \right)} \right\}, \quad (59)$$

$$\tilde{M}_p^{(+2)} = \frac{\tilde{w}_p^{(+1)}}{2\tilde{Z}_{+1}} \left\{ \operatorname{erfc} \left(-\tilde{\mu}_p^{(+1)} \sqrt{\tilde{\lambda}_p^{(+1)}/2} \right) \left(\left(\tilde{\mu}_p^{(+1)} \right)^2 + \frac{1}{\tilde{\lambda}_p^{(+1)}} \right) + \sqrt{\frac{2}{\pi \tilde{\lambda}_p^{(+1)}}} \frac{\tilde{\mu}_p^{(+1)}}{\exp \left(\tilde{\lambda}_p^{(+1)} \left(\tilde{\mu}_p^{(+1)} \right)^2 / 2 \right)} \right\}, \quad (60)$$

$$\tilde{M}_n^{(+0)} = \frac{\tilde{w}_n^{(+1)}}{2\tilde{Z}_{+1}} \operatorname{erfc} \left(\tilde{\mu}_n^{(+1)} \sqrt{\tilde{\lambda}_n^{(+1)}/2} \right), \quad (61)$$

$$\tilde{M}_n^{(+1)} = \frac{\tilde{w}_n^{(+1)}}{2\tilde{Z}_{+1}} \left\{ \operatorname{erfc} \left(\tilde{\mu}_n^{(+1)} \sqrt{\tilde{\lambda}_n^{(+1)}/2} \right) \tilde{\mu}_n^{(+1)} - \sqrt{\frac{2}{\pi\tilde{\lambda}_n^{(+1)}}} \frac{1}{\exp \left(\tilde{\lambda}_n^{(+1)} \left(\tilde{\mu}_n^{(+1)} \right)^2 / 2 \right)} \right\}, \quad (62)$$

$$\tilde{M}_n^{(+2)} = \frac{\tilde{w}_n^{(+1)}}{2\tilde{Z}_{+1}} \left\{ \operatorname{erfc} \left(\tilde{\mu}_n^{(+1)} \sqrt{\tilde{\lambda}_n^{(+1)}/2} \right) \left(\left(\tilde{\mu}_n^{(+1)} \right)^2 + \frac{1}{\tilde{\lambda}_n^{(+1)}} \right) - \sqrt{\frac{2}{\pi\tilde{\lambda}_n^{(+1)}}} \frac{\tilde{\mu}_n^{(+1)}}{\exp \left(\tilde{\lambda}_n^{(+1)} \left(\tilde{\mu}_n^{(+1)} \right)^2 / 2 \right)} \right\}. \quad (63)$$

2.2 Local variational method

We adopt the local variational method to tackle the intractability of MFVI for \mathbf{w} due to the introduction of the sigmoid function (Eq. 16 in the Method section). In particular, we have the following result regarding Eq. 15 in the Method section:

$$\begin{aligned} (0.5\pi_{j,k})^{t_{j,k}^{(-1)}} (1 - \pi_{j,k})^{t_{j,k}^{(0)}} (0.5\pi_{j,k})^{t_{j,k}^{(+1)}} &\propto \pi_{j,k}^{t_{j,k}^{(-1)} + t_{j,k}^{(+1)}} (1 - \pi_{j,k})^{t_{j,k}^{(0)}} \\ &= \exp \left\{ \mathbf{w}^\top \mathbf{s}_{j,k} \left(t_{j,k}^{(-1)} + t_{j,k}^{(+1)} \right) \right\} \sigma \left(-\mathbf{w}^\top \mathbf{s}_{j,k} \right) \\ &\geq \exp \left\{ \mathbf{w}^\top \mathbf{s}_{j,k} \left(t_{j,k}^{(-1)} + t_{j,k}^{(+1)} \right) \right\} \sigma(\xi_{j,k}) \exp \left\{ -\frac{1}{2} \left(\mathbf{w}^\top \mathbf{s}_{j,k} + \xi_{j,k} \right) - \chi(\xi_{j,k}) \left(\left(\mathbf{w}^\top \mathbf{s}_{j,k} \right)^2 - \xi_{j,k}^2 \right) \right\}, \end{aligned} \quad (64)$$

where

$$\chi(\xi) = \frac{1}{2\xi} \left(\sigma(\xi) - \frac{1}{2} \right). \quad (65)$$

Then we can perform standard MFVI with respect to the lower bound of Eq. 64, which yields

$$\begin{aligned} \ln q(\mathbf{w}) &\propto \mathbb{E}_{-\mathbf{w}} \left[\sum_{k=1}^K \sum_{j=1}^{J_k} \mathbf{w}^\top \mathbf{s}_{j,k} \left(t_{j,k}^{(-1)} + t_{j,k}^{(+1)} \right) - \frac{1}{2} \mathbf{w}^\top \mathbf{s}_{j,k} - \chi(\xi_{j,k}) \left(\mathbf{w}^\top \mathbf{s}_{j,k} \right)^2 - \frac{1}{2} \mathbf{w}^\top \mathbf{\Lambda} \mathbf{w} \right] \\ &= -\frac{1}{2} \mathbf{w}^\top \left(\mathbb{E}[\mathbf{\Lambda}] + 2 \sum_{k=1}^K \sum_{j=1}^{J_k} \chi(\xi_{j,k}) \mathbf{s}_{j,k} \mathbf{s}_{j,k}^\top \right) \mathbf{w} + \mathbf{w}^\top \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbf{s}_{j,k} \left(\mathbb{E} \left[t_{j,k}^{(-1)} \right] + \mathbb{E} \left[t_{j,k}^{(+1)} \right] - \frac{1}{2} \right). \end{aligned} \quad (66)$$

This indicates that $q(\mathbf{w})$ follows a normal distribution given by

$$q \left(\mathbf{w}; \tilde{\boldsymbol{\mu}}_w, \tilde{\boldsymbol{\Lambda}}_w \right) = \mathcal{N} \left(\tilde{\boldsymbol{\mu}}_w, \tilde{\boldsymbol{\Lambda}}_w \right), \quad (67)$$

in which

$$\tilde{\boldsymbol{\mu}}_w = \tilde{\boldsymbol{\Lambda}}_w^{-1} \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbf{s}_{j,k} \left(\mathbb{E} \left[t_{j,k}^{(-1)} \right] + \mathbb{E} \left[t_{j,k}^{(+1)} \right] - \frac{1}{2} \right), \quad (68)$$

$$\tilde{\boldsymbol{\Lambda}}_w = \mathbb{E}[\mathbf{\Lambda}] + 2 \sum_{k=1}^K \sum_{j=1}^{J_k} \chi(\xi_{j,k}) \mathbf{s}_{j,k} \mathbf{s}_{j,k}^\top. \quad (69)$$

2.3 Update rules for other variational parameters

For other latent variables in RefMap besides v_{-1} , v_{+1} and \mathbf{w} , we carry out the naive MFVI and obtain

$$q(\mathbf{u}_k; \tilde{\boldsymbol{\mu}}_k, \tilde{\boldsymbol{\Lambda}}_k) = \mathcal{N}(\mathbf{u}_k; \tilde{\boldsymbol{\mu}}_k, \tilde{\boldsymbol{\Lambda}}_k^{-1}), \quad (70)$$

$$q(m_{j,k}; \tilde{\mu}_{j,k}, \tilde{\lambda}_{j,k}) = \mathcal{N}(m_{j,k}; \tilde{\mu}_{j,k}, \tilde{\lambda}_{j,k}^{-1}), \quad (71)$$

$$q(\lambda_{j,k}; \tilde{a}_{j,k}, \tilde{b}_{j,k}) = \text{Gamma}(\lambda_{j,k}; \tilde{a}_{j,k}, \tilde{b}_{j,k}), \quad (72)$$

$$q(\tau_{-1}; \tilde{a}_{-1}, \tilde{b}_{-1}) = \text{Gamma}(\tau_{-1}; \tilde{a}_{-1}, \tilde{b}_{-1}), \quad (73)$$

$$q(\tau_{+1}; \tilde{a}_{+1}, \tilde{b}_{+1}) = \text{Gamma}(\tau_{+1}; \tilde{a}_{+1}, \tilde{b}_{+1}), \quad (74)$$

$$q(\tau_0; \tilde{a}_0, \tilde{b}_0) = \text{Gamma}(\tau_0; \tilde{a}_0, \tilde{b}_0), \quad (75)$$

$$q(m_{-1}, \lambda_{-1}; \tilde{\mu}_{-1}, \tilde{\beta}_{-1}, \tilde{c}_{-1}, \tilde{d}_{-1}) = \mathcal{N}(m_{-1}; \tilde{\mu}_{-1}, (\tilde{\beta}_{-1}\lambda_{-1})^{-1}) \text{Gamma}(\lambda_{-1}; \tilde{c}_{-1}, \tilde{d}_{-1}), \quad (76)$$

$$q(m_{+1}, \lambda_{+1}; \tilde{\mu}_{+1}, \tilde{\beta}_{+1}, \tilde{c}_{+1}, \tilde{d}_{+1}) = \mathcal{N}(m_{+1}; \tilde{\mu}_{+1}, (\tilde{\beta}_{+1}\lambda_{+1})^{-1}) \text{Gamma}(\lambda_{+1}; \tilde{c}_{+1}, \tilde{d}_{+1}), \quad (77)$$

$$q(\mathbf{t}_{j,k}; \tilde{\boldsymbol{\pi}}_{j,k}) = \tilde{\boldsymbol{\pi}}_{j,k}^{\mathbf{t}_{j,k}}, \quad (78)$$

$$q(\boldsymbol{\Lambda}; \tilde{\mathbf{W}}_{\Lambda}, \tilde{\nu}_{\Lambda}) = \mathcal{W}(\tilde{\mathbf{W}}_{\Lambda}, \tilde{\nu}_{\Lambda}), \quad (79)$$

in which

$$\tilde{\boldsymbol{\mu}}_k = \tilde{\boldsymbol{\Lambda}}_{u_k}^{-1} (\sqrt{N} \mathbf{z}_k + \mathbb{E}[\boldsymbol{\Lambda}_k] \mathbb{E}[\mathbf{m}_k]), \quad (80)$$

$$\tilde{\boldsymbol{\Lambda}}_k = N \boldsymbol{\Sigma}_k + \mathbb{E}[\boldsymbol{\Lambda}_k], \quad (81)$$

$$\tilde{\mu}_{j,k} = \left(\mathbb{E}[\lambda_{j,k}] \sum_{i=1}^{I_{j,k}} \mathbb{E}[u_{i,j,k}] - \mathbb{E}[v_{-1}] \mathbb{E}[\tau_{-1}] \mathbb{E}[t_{j,k}^{(-1)}] + \mathbb{E}[v_{+1}] \mathbb{E}[\tau_{+1}] \mathbb{E}[t_{j,k}^{(+1)}] \right) \tilde{\lambda}_{j,k}^{-1}, \quad (82)$$

$$\tilde{\lambda}_{j,k} = I_{j,k} \mathbb{E}[\lambda_{j,k}] + \mathbb{E}[t_{j,k}^{(-1)}] \mathbb{E}[\tau_{-1}] + \mathbb{E}[t_{j,k}^{(0)}] \mathbb{E}[\tau_0] + \mathbb{E}[t_{j,k}^{(+1)}] \mathbb{E}[\tau_{+1}], \quad (83)$$

$$\tilde{a}_{j,k} = a_0 + \frac{I_{j,k}}{2}, \quad (84)$$

$$\tilde{b}_{j,k} = b_0 + \frac{1}{2} \sum_{i=1}^{I_{j,k}} \mathbb{E}[u_{i,j,k}^2] + \frac{I_{j,k}}{2} \mathbb{E}[m_{j,k}^2] - \mathbb{E}[m_{j,k}] \sum_{i=1}^{I_{j,k}} \mathbb{E}[u_{i,j,k}], \quad (85)$$

$$\tilde{a}_{-1} = a_0 + \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbb{E}[t_{j,k}^{(-1)}], \quad (86)$$

$$\tilde{b}_{-1} = b_0 + \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbb{E}[t_{j,k}^{(-1)}] (\mathbb{E}[m_{j,k}^2] + \mathbb{E}[v_{-1}^2] + 2\mathbb{E}[m_{j,k}] \mathbb{E}[v_{-1}]), \quad (87)$$

$$\tilde{a}_{+1} = a_0 + \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbb{E} \left[t_{j,k}^{(+1)} \right], \quad (88)$$

$$\tilde{b}_{+1} = b_0 + \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbb{E} \left[t_{j,k}^{(+1)} \right] (\mathbb{E} [m_{j,k}^2] + \mathbb{E} [v_{+1}^2] - 2\mathbb{E} [m_{j,k}] \mathbb{E} [v_{+1}]), \quad (89)$$

$$\tilde{a}_0 = a_0 + \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbb{E} \left[t_{j,k}^{(0)} \right], \quad (90)$$

$$\tilde{b}_0 = b_0 + \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^{J_k} \mathbb{E} [m_{j,k}^2] \mathbb{E} \left[t_{j,k}^{(0)} \right], \quad (91)$$

$$\tilde{\mu}_{-1} = \frac{\beta_0 \mu_0 + \mathbb{E}[r_{-1}]}{\beta_0 + 1}, \quad (92)$$

$$\tilde{\beta}_{-1} = \beta_0 + 1, \quad (93)$$

$$\tilde{c}_{-1} = a_0 + \frac{1}{2}, \quad (94)$$

$$\tilde{d}_{-1} = b_0 + \frac{1}{2} \beta_0 \mu_0^2 + \frac{1}{2} \mathbb{E} [r_{-1}^2] - \frac{1}{2(\beta_0 + 1)} (\beta_0 \mu_0 + \mathbb{E}[r_{-1}])^2, \quad (95)$$

$$\tilde{\mu}_{+1} = \frac{\beta_0 \mu_0 + \mathbb{E}[r_{+1}]}{\beta_0 + 1}, \quad (96)$$

$$\tilde{\beta}_{+1} = \beta_0 + 1, \quad (97)$$

$$\tilde{c}_{+1} = a_0 + \frac{1}{2}, \quad (98)$$

$$\tilde{d}_{+1} = b_0 + \frac{1}{2} \beta_0 \mu_0^2 + \frac{1}{2} \mathbb{E} [r_{+1}^2] - \frac{1}{2(\beta_0 + 1)} (\beta_0 \mu_0 + \mathbb{E}[r_{+1}])^2, \quad (99)$$

$$\tilde{\pi}_{j,k}^{(i)} = \frac{\exp \left\{ \tilde{\rho}_{j,k}^{(i)} \right\}}{\exp \left\{ \tilde{\rho}_{j,k}^{(-1)} \right\} + \exp \left\{ \tilde{\rho}_{j,k}^{(0)} \right\} + \exp \left\{ \tilde{\rho}_{j,k}^{(+1)} \right\}} \quad (i = -1, 0, +1), \quad (100)$$

$$\tilde{\nu}_\Lambda = \nu_0 + 1, \quad (101)$$

$$\tilde{\mathbf{W}}_\Lambda = (\mathbf{W}_0^{-1} + \mathbb{E} [\mathbf{w}\mathbf{w}^\top])^{-1}, \quad (102)$$

and we define

$$\tilde{\rho}_{j,k}^{(-1)} = \frac{1}{2} \mathbb{E} [\ln \tau_{-1}] - \frac{1}{2} \mathbb{E} [\tau_{-1}] (\mathbb{E} [m_{j,k}^2] + \mathbb{E} [v_{-1}^2] + 2\mathbb{E} [m_{j,k}] \mathbb{E} [v_{-1}]) + \mathbb{E} [\ln \pi_{j,k}] - \ln 2, \quad (103)$$

$$\tilde{\rho}_{j,k}^{(+1)} = \frac{1}{2} \mathbb{E} [\ln \tau_{+1}] - \frac{1}{2} \mathbb{E} [\tau_{+1}] (\mathbb{E} [m_{j,k}^2] + \mathbb{E} [v_{+1}^2] - 2\mathbb{E} [m_{j,k}] \mathbb{E} [v_{+1}]) + \mathbb{E} [\ln \pi_{j,k}] - \ln 2, \quad (104)$$

$$\tilde{\rho}_{j,k}^{(0)} = \frac{1}{2} \mathbb{E} [\ln \tau_0] - \frac{1}{2} \mathbb{E} [\tau_0] \mathbb{E} [m_{j,k}^2] + \mathbb{E} [\ln(1 - \pi_{j,k})]. \quad (105)$$

Algorithm 1: MFVI for RefMap

Input : Z -scores $z_{i,j,k}$, epigenome features $\mathbf{s}_{j,k}$ and LD matrices Σ_k .

Output : Posteriors q and local variational parameters $\xi_{j,k}$.

1 Initialize variational parameters.

2 **while** *not converged* **do**

3 Update global variational parameters based on Eqs. 31, 32, 33, 34, 49, 50, 51, 52, 68, 69, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, and 102.

4 Update local variational parameters based on Eq. 107.

5 Calculate ELBO (details omitted).

6 **end**

2.4 Update rules for local variational parameters

One needs to maximize the lower bound on marginal likelihood in Eq. 64 with respect to $\xi_{j,k}$ to rationalize the local variational inference. In particular, we have the following optimization problem

$$Q(\boldsymbol{\xi}, \boldsymbol{\xi}^{\text{old}}) \propto \sum_{k=1}^K \sum_{j=1}^{J_k} \ln \sigma(\xi_{j,k}) - \frac{1}{2} \xi_{j,k} - \chi(\xi_{j,k}) ((\mathbf{w}^\top \mathbf{s}_{j,k})^2 - \xi_{j,k}^2). \quad (106)$$

Solving the above problem with respect to each $\xi_{j,k}$ gives its update rule

$$\xi_{j,k}^{\text{new}} = \sqrt{\mathbf{s}_{j,k}^\top \mathbb{E}[\mathbf{w}\mathbf{w}^\top] \mathbf{s}_{j,k}}. \quad (107)$$

2.5 Coordinate ascent algorithm for MFVI

With the above update rules we can construct a coordinate ascent algorithm to update variational parameters iteratively until convergence (i.e., the change of ELBO falls below a threshold which was set to be 10^{-6} in our study). The inference algorithm is summarized in Algorithm 1.