Supplementary information for
From drug repositioning to target repositioning: prediction of therapeutic targets using genetically perturbed transcriptomic signatures

Satoko Namba, Michio Iwata, and Yoshihiro Yamanishi*

Department of Bioscience and Bioinformatics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Iizuka, Fukuoka 820-8502, Japan

*corresponding author: Yoshihiro Yamanishi (yamani@bio.kyutech.jp)
# Table of contents

**Supplementary Methods** ........................................................................................................ 1

1. Alternative SNP profiling methods ................................................................. 1
2. A formulation of trans-disease method ......................................................... 3
3. Hierarchical clustering of perturbed genes ................................................ 5
4. Assessing the distribution of genetically perturbed genes ......................... 5
5. GO and KEGG pathway analysis ................................................................. 5
6. Performance evaluation of data completion method .................................... 5

**Supplementary Results** ................................................................................................. 7

1. Performance comparison of alternative SNP profiling methods ................. 7
2. Performance evaluation of therapeutic target predictions on a cell-by-cell basis .... 8
3. Performance evaluation of data completion method .................................... 10

**Supplementary Discussion** .......................................................................................... 11

1. Toward the appropriate selection of disease-related cell lines ..................... 11

**Supplementary References** .......................................................................................... 12

**Supplementary Figures** ............................................................................................... 13

1. Performance comparison of alternative SNP profiling methods ................. 13
2. Hierarchical clustering of genetically perturbed genes ............................. 14
3. Classifications of genetically perturbed genes according to protein subfamily .......... 15
4. Performance evaluation for therapeutic target prediction using the inverse signature method ................................................................. 16
5. Performance evaluation for therapeutic target prediction using the trans-disease method .................................................................................. 17
6. Performance evaluation among the inverse signature and trans-disease methods with missing and completed data ........................................... 18
7. Distribution of enriched GO terms and pathways in the GO and pathway analyses of TSLP .......................................................... 19
8. Distribution of enriched GO terms and pathways in the GO and pathway analyses of IFNG ................................................................. 20
9. GO, pathway, and PPI network results for TAF1B as an inhibitory target .......................... 22
10. Distribution of enriched GO terms in GO analysis of TAF1B predicted as the inhibitory target .................................................................................................................. 23
11. Distribution of enriched pathways in the pathway analysis of TAF1B predicted as the inhibitory target .................................................................................................................. 24

**Supplementary Tables ................................................................. 25**

1. Cell line list of genetically perturbed gene expression profiles .............................. 25
2. Diseases that had at least one therapeutic target protein ........................................ 26
3. Leaf nodes belonging to the clusters formed by the hierarchical clustering for knock-down genes ................................................................................................................................... 27
4. Leaf nodes belonging to the clusters formed by the hierarchical clustering for over-expressed genes ................................................................................................................................ 33
5. The distribution of inhibitory targets repositioned from the original disease class to other disease classes .......................................................................................................................... 38
6. The distribution of activatory targets repositioned from the original disease class to other disease classes .......................................................................................................................... 49
7. Newly predicted 52 inhibitory targets repositioned from the original disease to other diseases ................................................................................................................................... 40
8. Newly predicted 46 activatory targets repositioned from the original disease to other diseases ................................................................................................................................... 42
9. Top 100 novel inhibitory target–disease pairs predicted using the trans-disease method with completed data .......................................................................................................................... 43
10. Top 100 novel activatory target–disease pairs predicted using the trans-disease method with completed data .......................................................................................................................... 46
11. Performance evaluation of data completion for the transcriptome data (knocked down genes, genes, and cell lines) .......................................................................................................................... 49
12. Performance evaluation of data completion for the transcriptome data (overexpressed genes, genes, and cell lines) .......................................................................................................................... 50
Supplementary Methods

Alternative SNP profiling methods

We compared five different approaches using SNP information to select a baseline method for comparison with the proposed methods. We referred to these approaches as “Mean p-value method,” “Gene based p-value method,” “Mean NES method,” “Min-max NES method,” and “CADD method,” respectively. The “Mean p-value method” used only SNPs within coding regions, while the other methods used SNPs within intergenic regions as well as coding regions. We obtained SNP data for the various diseases from the NHGRI-EBI Genome-wide Association Studies Catalog database (Buniello et al., 2019), and constructed disease-specific SNP profiles. Using the SNP profiles, we predicted therapeutic targets.

In the “Mean p-value method,” we predicted therapeutic targets focusing on the p-values of SNPs within coding regions to predict therapeutic targets that can be directly controlled by drugs. When a gene had multiple SNPs or was reported by multiple GWASs, we averaged the p-values for the gene. We constructed SNP profiles corresponding to the gold standard data by assigning the value 0 to genes with no SNP data. We used $-\log(P)$ values as predictive scores. Genes that had SNPs with significantly strong associations with disease were considered to be candidate therapeutic targets.

In the “Gene based p-value method,” we predicted therapeutic targets using weighted Fisher's statistic (Hou, 2005) as a meta-analysis approach. MAGMA (de Leeuw et al., 2015) was used as the GWAS meta-analysis tool. As input data, we used GWAS summary statistics for each disease obtained from the GWAS Catalog; MAGMA converted SNP-based $p$-values to gene-based $p$-values. For the gene-based $p$-value, we calculated $-\log(p$-value) and used it as a prediction score. We predicted genes with high scores as candidates for therapeutic targets of each disease.

In the “Mean NES method” and “Min-max NES method,” we predicted therapeutic targets using effect size to express strength of biological effect. We obtained normalized effect size (NES) for each SNP from “GTEx_Analysis_v8_eQTL” data obtained from Genotype-Tissue Expression (GTEx; Ardlie et al., 2015). NESs of tissues associated with each disease were extracted. We examined two possible strategies for the NES-based prediction score: i) taking the average and ii) taking the min-max. In the former strategy, we averaged the NES values when multiple SNPs were present for each gene, and referred to it as “Mean NES method.” In the latter strategy, we used the maximum and
minimum values of NESs for inhibitory target prediction and activatory target prediction, respectively, and referred to it as “Min-max NES method.” Genes with positive NES values should be inhibited for inhibitory target prediction, so we predicted genes with high NES values as candidate inhibitory targets. Genes with negative NES values were predicted to be activatory target candidates because genes with negative NES values should be activated.

In the “CADD method,” we predicted therapeutic targets based on functional annotation. We used FUMA (Watanabe et al., 2017) to perform functional annotation on SNPs. GWAS summary statistics were used as input data for FUMA. SNPs were mapped onto genes based on their genomic location by positional mapping, which is the default function of FUMA. Since the therapeutic targets must be controlled by drugs, the mapped genes were protein-coding genes, which is the default setting of FUMA. The Combined Annotation Dependent Depletion (CADD) score (Kircher et al., 2014; Rentzsch et al., 2019), which indicates the effect of the mutation on functions, was used as a prediction score. We predicted genes with high CADD scores as candidate therapeutic targets of each disease.
A formulation of trans-disease method

We address the problem of therapeutic indications prediction by focusing on therapeutic targets. Note that there are a number of candidates for diseases, and different diseases may have common characteristics in terms of molecular mechanisms. The same therapeutic targets are sometimes used for multiple diseases. Thus, we propose formulating the problem in the framework of supervised multiple label prediction.

Suppose that there are $M$ diseases and we are given $N$ targets. We consider predicting which diseases would be treated by a target, that is, the $i$-th target. Each target is represented by a $d$-dimensional feature vector as $\mathbf{x}_i$ in this study, where $\mathbf{x}_i$ was obtained by averaging the multiple signatures from different cell lines.

We constructed a learning set of target–disease pairs that are pairs given in target–disease associations (see the Methods section for more details). There are $M$ candidates for diseases, and each target in the learning set is assigned a binary class label representing the $m$-th disease ($m = 1, 2, \ldots, M$). Let $y_{m,i} \in \{0, 1\}$ be the class label for the $m$-th disease assigned to the $i$-th target, where $y_{m,i} = 1$ means that $i$-th target is used for the $m$-th disease, and $y_{m,i} = 0$ means that the $i$-th target is not used for the $m$-th disease.

We construct a predictive model to predict whether the $i$-th target would be used for the $m$-th disease ($m = 1, 2, \ldots, M$). Linear models are a useful tool to analyze extremely high-dimensional data for both prediction and feature extraction tasks. Thus, we adopt a linear function defined as $f_m = \mathbf{w}_m^T \mathbf{x}$, where $\mathbf{w}_m$ is a $d$-dimensional weight vector for the $m$-th disease. We represent a set of $M$ model weights by a $d \times M$ matrix defined as $\mathbf{W} := [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M]$ and estimate the weight matrix $\mathbf{W}$ by minimizing an objective function based on the learning set.

To overcome the scarcity of existing knowledge concerning relationships between targets and diseases, we propose learning individual predictive models $f_1, f_2, \ldots, f_M$ jointly, sharing information across $M$ diseases.

We attempt to estimate all of the weight vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M$ jointly in the models by minimizing the logistic loss as follows:

$$R(\mathbf{W}) = \sum_{m=1}^{M} \sum_{i=1}^{N} \log(1 + \exp(-y_{m,i} \mathbf{w}_m^T \mathbf{x}_i)).$$
We introduce a regularization term $\Omega(W)$ to the loss function in order to enhance the generalization properties. Thus, the optimization problem is written as follows:

$$\min_W R(W) + \Omega(W).$$

(2)

Here we introduce two regularization terms. First, we use a standard ridge regularization term to avoid the over-fitting problem, which is defined as

$$\Omega_r := \frac{1}{2} Tr(WW^T).$$

(3)

Second, we design another regularization term reflecting the similarities among diseases. In this study we evaluate the similarity among diseases using the cosine similarity and construct an $M \times M$ similarity matrix $S$ for diseases in which each element $S_{i,j}$ is a similarity score between the $i$-th and $j$-th diseases. Then, we introduce the following regularization term:

$$\Omega_s(W) := \frac{1}{4} \sum_{l=1}^M \sum_{m=1}^M S_{l,m} \left\| \frac{w_l}{\sqrt{K_{l,l}}} - \frac{w_m}{\sqrt{K_{m,m}}} \right\| = \frac{1}{2} Tr(WL_sW^T),$$

(4)

where $\| \cdot \|$ is the Euclidean norm, $K$ is a diagonal matrix defined as $K_{l,l} := \sum_{m=1}^M S_{l,m}$, and $L_s$ is a symmetric normalized Laplacian defined as $K^{-1/2}(K - S)K^{-1/2}$. The regularization term $\Omega_s(W)$ has the effect of bringing the weight vectors $w_i$ and $w_j$ close to each other if $S_{i,m}$ is high.

Finally, we introduce the following regularization term in the optimization problem (2):

$$\Omega(W) := \lambda_s \Omega_s(W) + \lambda_r \Omega_r(W),$$

(5)

where $\lambda_s \geq 0$ and $\lambda_r \geq 0$ are hyper-parameters to control the strength of the regularization terms $\Omega_s$ and $\Omega_r$, respectively.

A grid search was performed to determine the optimal set of regularization parameters $\lambda_s$ and $\lambda_r$. We prepared a set of candidate values $\{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$ for $\lambda_s$ and $\lambda_r$ in advance. We calculated the AUC scores by performing a five-fold cross-validation experiment. We selected parameter values that achieved the highest AUC score. We used the parameters to predict applicable diseases for the therapeutic targets.
Hierarchical clustering of perturbed genes

To examine the features of genetically perturbed gene expression signatures, we performed hierarchical clustering of both the genetically perturbed genes and the L1000 genes. Hierarchical clustering was performed using the Python library SciPy (v 1.4.1, https://www.scipy.org/) with the following parameters: method = “ward” and metric = “euclidean.”

Assessing the distribution of genetically perturbed genes

To evaluate whether the set of genetically perturbed genes contained an adequate number of candidate therapeutic targets, we examined the distribution of these genes according to their protein family, biological processes, and biological pathways. Specifically, we classified these genes based on the protein families and subfamilies provided by PANTHER (Mi et al., 2013; Thomas et al., 2003), their associated GO biological process terms, and their linked KEGG pathways in the KEGG database (Kanehisa et al., 2002). “Global pathways” were excluded.

GO and KEGG pathway analysis

GO analysis is usually used to annotate genes and gene products according to their characteristic biological functions (Ashburner et al., 2000), whereas KEGG pathway analysis assigns genes to functional pathways (Kanehisa et al., 2002).

In this study, we performed GO and KEGG pathway analysis to analyze the mechanisms of the therapeutic indications of predicted inhibitory and activated targets. We extracted differentially expressed genes from the gene expression signatures after perturbation of these targets, and these genes were used for further analyses. A Cytoscape plug-in, ClueGO, was used to identify over-represented GO terms and KEGG pathways(Bindea et al., 2009).

Performance evaluation of data completion method

We evaluated the performance of the tensor decomposition algorithm (Yuan et al., 2017) used in this study in terms of the imputation accuracy. For comparison, we compared the performance of the algorithm with the CANDECOMP/PARAFAC (CP) decomposition (Acar et al., 2010). To evaluate
the performance of the imputation, we randomly added artificial missing values to the observed data and tested whether the tensor imputation algorithms could correctly recover these values.
Supplementary Results

Performance comparison of alternative SNP profiling methods

We compared five different approaches using SNP information to select a baseline method for comparison with the proposed methods; “Mean p-value method,” “Gene based p-value method,” “Mean NES method,” “Min-max NES method,” and “CADD method” (see Supplementary Methods section).

Supplementary Fig. S1 compares the predictive performances of these different SNP profiling methods. There was little difference between these methods when predicting inhibitory targets, as shown in Supplementary Fig. S1 (A). However, in the case of activatory target prediction, the prediction accuracy of the Mean p-value method was slightly higher than those of other methods, as shown in Supplementary Fig. S1 (B). The Mean p-value method uses only SNPs in the coding region, while the other methods also use SNPs in the intergenic region, indicating that the focus on SNPs within coding regions is slightly more accurate in terms of therapeutic target prediction. Since the mean p-value method uses the averaged p-values of SNPs, while the Gene based p-value method uses gene-based p-values converted by meta-analysis with MAGM, suggesting that the focus on using the averaged p-values is slightly more accurate in terms of therapeutic target prediction. The mean p-value method uses the significance of association, the Mean NES method and Min-max NES method use the NES score, which represents the strength of the biological effect, and the CADD method, which reflect the effect of the mutation on functions, suggesting that the using the significance of association is slightly more accurate in terms of therapeutic target prediction. These results suggest that the using the significance of association (averaged p-values) is slightly more accurate in terms of therapeutic target prediction.

Based on these results, we decided to use the slightly more accurate Mean p-value method for comparison with the proposed methods. In the main text, this method is referred to as the SNP profiling method.
Performance evaluation of therapeutic target predictions on a cell-by-cell basis

We compared predictive performance when using completed data vs. missing data. As shown in Fig. 4, the completion of genetically perturbed gene expression signature data improved the predictive performance of the methods, particularly when using the trans-disease method \((P = 5.01 \times 10^{-4}\) for inhibitory target–disease pairs; \(P = 1.09 \times 10^{-2}\) for activatory target–disease pairs; Wilcoxon signed-rank tests). In all cases, the trans-disease method with completed data outperformed the inverse signature method with missing data \((P = 1.47 \times 10^{-4}\) for inhibitory target–disease pairs; \(P = 2.53 \times 10^{-3}\) for activatory target–disease pairs). These results indicate that the trans-disease method with completed data is the most effective method for predicting therapeutic targets for diseases.

We evaluated the ability of the inverse signature method to predict therapeutic targets on a cell-by-cell basis. Supplementary Fig. S4 shows a comparison of the AUC scores for each cell line and the associated missing rate in the gene expression signatures for each cell line. Supplementary Fig. S4 (A) and (B) show evaluations of the inhibitory target prediction and activatory target prediction performances, respectively. For nearly all cell lines, the AUC scores for both the inhibitory and activatory target predictions were about 0.5, which shows that the predictions obtained using the inverse signature method were random inferences. This issue may have arisen because the inverse correlation method involves unsupervised learning. Because the number of known therapeutic target–disease associations is very limited, the accuracy of this method may be underestimated.

When using the trans-disease method, we evaluated the performance of therapeutic target predictions by performing five-fold cross-validation experiments. We calculated the AUC scores using the method described in the Results section. Supplementary Fig. S5 shows a comparison of the AUC scores for each cell line and the associated missing rate in the gene expression signatures for the same cell lines. Supplementary Fig. S5 (A) shows a performance evaluation for the inhibitory target prediction. For nearly all cell lines, inhibitory target predictions by the trans-disease method were superior to the predictions produced using the inverse signature method. In cell lines with high missing rates, the performance of predictions was poor. Supplementary Fig. S5 (B) shows a performance evaluation for the activatory target prediction. Performances differed depending on the cell line, which suggests that activatory target prediction by the trans-disease method performed well
in certain cell lines. Overall, these results indicate that the trans-disease method outperformed the inverse signature method when predicting therapeutic targets. However, in cell lines with high rates of missing data, the trans-disease method was not effective for inhibitory target predictions in particular.

Finally, we compared the predictive performances of the methods for each cell line. Supplementary Fig. S6 (A) shows a comparison of the AUC scores of inhibitory target prediction for each cell line as well as the associated missing rate in the gene expression signatures for each cell line. For nearly all cell lines, the use of gene expression signatures with completed data was more effective than the use of gene expression signatures with missing data; this finding was especially valid when using the trans-disease method. In cell lines with high rates of missing data, the usefulness of gene expression signatures with completed data was clearly shown. Notably, genetically perturbed gene expression signatures were largely unobserved (96.4% unobserved) in several cell lines, including SHSY5Y, SKL, HEK293T, and NCIH716, indicating that such cell lines included few perturbed genes. Supplementary Fig. S6 (B) shows a comparison of the AUC scores of activatory target prediction for each cell line as well as the associated missing rate in the gene expression signatures for each cell line. As with inhibitory target predictions, for nearly all cell lines, the use of gene expression signatures with completed data produced better results than those produced using gene expression signatures with missing data, particularly when using the trans-disease method. Thus, predictive performance can apparently be improved by imputing many missing values of genetically perturbed gene expression signatures.
Performance evaluation of data completion method

Supplementary Table S11 shows the result for the performance evaluation of the data completion for a third-ordered gene expression dataset comprising 4,345 knocked down genes, 978 genes and 17 cell lines, where the dataset contains artificial missing values. Similarly, Supplementary Table S12 shows the result for the performance evaluation of the data completion for a third-ordered gene expression dataset comprising 3,114 overexpressed genes, 978 genes and 10 cell lines, where the dataset contains artificial missing values. In these evaluations, the tensor decomposition algorithm used in this study worked better than the algorithm with the CP decomposition. These results suggest that the tensor decomposition algorithm can work well for data completion.
Supplementary Discussion

Toward the appropriate selection of disease-related cell lines

We evaluated the performance of our proposed methods on a cell-by-cell basis (Supplementary Results and Figs. S4–S6). It would be ideal to use only cellular data associated with disease organs; unfortunately, such data associated disease organs is not always available. In fact, the cellular data in this study was limited to cell lines because the cellular data for all diseases studied here was not available. However, we predicted new therapeutic target–disease associations using the proposed method based on the averaged cell line data. Note that our proposed method is applicable to new predictions on a cell-by-cell basis when genetically perturbed gene expression profiles are available for all cell lines that correspond to diseased organs. Accumulation of more data should solve this problem in the future.
Supplementary References

Acar,E. et al. (2010) Scalable tensor factorizations for incomplete data ☆.

Ardlie,K.G. et al. (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science, 348, 648–660.

Ashburner,M. et al. (2000) Gene ontology: Tool for the unification of biology. Nat. Genet., 25, 25–29.

Bindeia,G. et al. (2009) ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics, 25, 1091–1093.

Buniello,A. et al. (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res., 47, D1005–D1012.

Hou,C.D. (2005) A simple approximation for the distribution of the weighted combination of non-independent or independent probabilities. Stat. Probab. Lett., 73, 179–187.

Kanehisa,M. et al. (2002) The KEGG databases at GenomeNet. Nucleic Acids Res., 30, 42–46.

Kircher,M. et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet., 46, 310–315.

de Leeuw,C.A. et al. (2015) MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol., 11.

Mi,H. et al. (2013) PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res., 41.

Rentzsch,P. et al. (2019) CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res., 47, D886–D894.

Thomas,P.D. et al. (2003) PANTHER: A library of protein families and subfamilies indexed by function. Genome Res., 13, 2129–2141.

Watanabe,K. et al. (2017) Functional mapping and annotation of genetic associations with FUMA. Nat. Commun., 8.

Yuan,L. et al. (2017) Completion of high order tensor data with missing entries via tensor-train decomposition. In, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 222–229.
**Supplementary Figure S1.** Performance comparison of alternative SNP profiling methods. (A) Performance comparison on inhibitory target prediction. (B) Performance comparison on activatory target prediction. Mean p-value method used only SNPs within coding regions, while the other methods used SNPs within intergenic regions as well as coding regions. Mean p-value method is based on the significance of association ($p$-values), and corresponds to Fig. 4; Gene based p-value method is based on meta-analysis approach with MAGMA; Mean NES method is based on the normalized effect size (NES) with the average operation; Min-max NES method is based on NES with the min/max operation; CADD method is based on functional annotation with FUMA.
Supplementary Figure S2. Hierarchical clustering of (A) gene knock-down and (B) gene over-expression signatures. These dendrograms represent the results of hierarchical clustering for 4,345 knocked down and 3,114 overexpressed genes, respectively. Cluster node ids represent identifiers of clusters. Leaf nodes belonging to each cluster node id are shown in Supplementary Tables S3 and S4.
Supplementary Figure S3. Classifications of (A) knocked down genes and (B) overexpressed genes according to protein subfamily. These protein subfamilies belonged to the top 5 protein families in the top panel of the Fig. 3, and these protein subfamilies were obtained from the Protein Analysis Through Evolutionary Relationships (PANTHER; Mi et al., 2013; Thomas et al., 2003). Since the number of protein subfamilies was too large, the top 10 subfamilies in each protein family were displayed.
Supplementary Figure S4. Performance evaluation for therapeutic target prediction, using the inverse signature method. Performance evaluation for (A) inhibitory target prediction and (B) activatory target prediction. The top panel shows the area under the ROC curve score calculated using all prediction scores for all target–disease pairs. The bottom panel shows the rate of missing data in each cell line. Cell lines are listed in increasing order of the missing rate.
Supplementary Figure S5. Performance evaluation for therapeutic target prediction using the trans-disease method. Performance evaluation for (A) inhibitory target prediction and (B) activatory target prediction. The top panel shows the AUC score calculated using all prediction scores for all target–disease pairs. The bottom panel shows the rate of missing data in each cell line. Cell lines are listed in increasing order of missing rate.
Supplementary Figure S6. Performance comparison among the inverse signature and trans-disease methods with missing and completed data. (A) Performance comparison on inhibitory target prediction. (B) Performance comparison on activatory target prediction. The top panel shows the AUC score calculated using all prediction scores for all target–disease pairs. The bottom panel shows the rate of missing data in each cell line. Cell lines are listed in increasing order of missing rate.
Supplementary Figure S7. Distribution of enriched GO terms and pathways in the GO and pathway analyses of TSLP. The top panel shows the result of GO analysis of the biological process. The bottom panel shows the result of KEGG pathway analysis. These analyses were performed based on differentially expressed genes of gene expression signatures following TSLP knock-down. The x axis represents the ratio of associated genes to all genes in the term or pathway. The numbers next to the bars represent the number of genes associated with the term or pathway. The color represents the GO or pathway groups. Asterisks indicate the significance level as follows: * $p < 0.05$, ** $p < 0.01$. The functionally grouped network of these GO terms or pathways are shown in Fig. 7 (A).
Supplementary Figure S8. Distribution of enriched GO terms and pathways in the GO and pathway analyses of IFNG. The top panel shows the result of GO analysis of the biological process. The bottom panel shows the result of KEGG pathway analysis. These analyses
were performed based on differentially expressed genes of gene expression signatures following IFNG over-expression. The x axis represents the ratio of associated genes to all genes in the term or pathway. The numbers next to the bars represent the number of genes associated with the term or pathway. The color represents the GO or pathway groups. Asterisks indicate the significance level as follows: * $p < 0.05$, ** $P < 0.01$. The functionally grouped network of these GO terms or pathways are shown in Fig. 7 (B).
Supplementary Figure S9. (A) GO, (B) pathway, and (C) PPI network results for TAF1B as an inhibitory target. These analyses were performed using the genes differentially expressed following TAF1B knockdown. (A) The analysis of GO biological process terms and (B) KEGG pathway analysis, with circles representing GO terms and KEGG pathways, respectively. The edges denote term–term interactions and functional groups (GO groups) based on the genes shared between the terms. The node color represents the GO group. The node size represents the term significance; the biggest term from a group is the most significant and is highlighted. The GO terms and pathways in these networks are shown in Supplementary Figs. S10 and S11. (C) PPI network showing the protein (gene)–protein (gene) interaction network altered in response to TAF1B knockdown, as well as the results of network topology analysis. The circles in the network represent proteins. The edges denote protein–protein interactions. The node size represents degree. The edge size represents interaction confidence defined according to STRING.
Supplementary Figure S10. Distribution of enriched GO terms in GO analysis of TAF1B predicted as the inhibitory target. The bar chart shows the result of GO analysis of the biological process. GO analysis was performed based on differentially expressed genes of gene expression signatures following TAF1B knock-down. The x axis represents the ratio of associated genes to all genes in the term. The color represents the GO groups. Since the number of enriched GO terms was too large, the top 3 GO terms with high significance in each GO group were displayed. The functionally grouped network of these GO terms is shown in Supplementary Fig. 9 (A).
Supplementary Figure S11. Distribution of enriched pathways in the pathway analysis of TAF1B predicted as the inhibitory target. The bar chart shows the result of KEGG pathway analysis of the biological process. This analysis was performed based on differentially expressed genes of gene expression signatures following TAF1B knock-down. The x axis represents the ratio of associated genes to all genes in the pathway. The color represents the KEGG pathway groups. The functionally grouped network of these KEGG pathway terms is shown in Supplementary Fig. 9 (B).
**Supplementary Table S1.** Cell line list of genetically perturbed gene expression profiles. The gene knockdown profiles consist of data from 17 cell lines and the gene overexpression profiles from 10 cell lines. These cell lines are listed in ascending order of missing data, and the missing rate of each cell line is shown in the bottom panel of Supplementary Figs. S4-S6.

| Type of gene perturbation | Cell lines                                                  |
|---------------------------|------------------------------------------------------------|
| Gene knock-down           | PC3, MCF7, HA1E, A375, A549, HEPG2, VCAP, HT29, HCC515, HEKTE, NPC, SW480, ASC, SHSY5Y, SKL, HEK293T, NCIH716 |
| Gene over-expression      | PC3, A375, HEPG2, HT29, HA1E, MCF7, HCC515, A549, VCAP, HEK293T |
Supplementary Table S2. Diseases that had at least one therapeutic target protein. In total, 32 and 15 diseases had at least one inhibitory and activatory target protein, respectively. These diseases are listed in alphabetical order.

| Type of therapeutic target protein | Diseases                                                                 |
|-----------------------------------|--------------------------------------------------------------------------|
| Inhibitory target protein         | acute myeloid leukemia, adult T-cell leukemia, Alzheimer’s disease, aplastic anemia, asthma, atopic dermatitis, breast cancer, cervical cancer, chronic lymphocytic leukemia, chronic myeloid leukemia, colorectal cancer, congenital muscular dystrophies, Crohn’s disease, endometrial cancer, gastric cancer, hepatitis C, hypercholesterolemia, idiopathic pulmonary fibrosis, inflammatory bowel disease, malignant melanoma, multiple myeloma, ovarian cancer, pancreatic cancer, Parkinson disease, renal cell carcinoma, rheumatoid arthritis, small cell lung cancer, systemic lupus erythematosus, testicular cancer, tuberculosis, type I diabetes mellitus, type II diabetes mellitus |
| Activatory target protein         | acute myeloid leukemia, breast cancer, chronic granulomatous disease, chronic myeloid leukemia, colorectal cancer, Crohn’s disease, inflammatory bowel disease, multiple myeloma, ovarian cancer, Parkinson disease, renal cell carcinoma, Rett syndrome, rheumatoid arthritis, type I diabetes mellitus, type II diabetes mellitus |
**Supplementary Table S3.** Leaf nodes belonging to the clusters formed by the hierarchical clustering for knocked down genes. ‘Cluster node id’ corresponds to the cluster node ids in the Supplementary Fig. S2 (A). The nodes of the clusters are listed in the ‘Knocked down genes’ column. The number of knocked down genes belonging to the clusters are shown in the ‘Count’ column represents.
| Gene Symbols |
|--------------|
| RAF1, RASSF7, RB1, RDH11, RFX2, RGS4, RHEF, RIMS3, RIPK2, RIPK4, RNASE4, RNF3, RN1F45, RPP18, RPS6KA1, RPS6KB1, RRAGC, Rsu1, RYR2, S100A1, S100P, S1PR2, SACM1L, SAR1BP, SBNO1, SCUBE1, SEPT1, SERPINB6, SERPINB7, SERTAD1, SEPT7, SF1, SGPL1, SHANK3, SIAH2, SIK1, SIK2, SIRT1, SIRT2, SIX2, SLC12A7, SLC16A5, SLC22A18, SLC25A22, SLC25A46, SLC2A1, SLC30A8, SLC36A4, SLC7A5, SMAD6, SMARCBl, SMARCC2, SMOT, SOC52, SOCS4, SOCS5, SORD, SORT1, SOX11, SOX4, SP110, SPA17, SPRY1, SRK, SRPK1, ST3GAL5, ST6GAN, STAT2, STAT3A, STK17A, STK24, STUB1, STXB1, STXB2, STX1A1, STX1A2, TAAR6, TAAR9, TAF15, TAF51, TAL2, TAPI, TARBP2, TAPC, TCEAL4, TCF3, TFIIB, TFIID, TIN2M2, TMM50A, TNSFRS13B, TNSFRS19, TNSFRS68, TNSFRS69, TP53BP1, TP53TG1, TPTM, TPKF1, TRAF6, TRAF4, TRERF1, TRIB1, TRIM2, TSLC, TSP1, TAP1, UB2E3D, UBEB1, UBEL, UBP1, UFRM1, UGT1A16, UGT1A9, VAV3, VCAN, VNR2, WASF3, WNT9A, ZAPTF, ZNF175, ZNF195, ZNF200, ZNF454, ZNF486, ZNF51, ZNF514, ZNF546, ZNF559, ZNF583, ZNF586, ZNF619, ZNF767, ZNF778, ZNF855, ZSNCA9, ZSNCA14, ZW10 |
| Cluster node id | Overexpressed genes | Count |
|----------------|----------------------|-------|
| 1              | {KSR, MAP4K4}        | 2     |
| 2              | {CAMKV, MAP4K2, PIDK1L, PIM3, PRKCG, PT2B, RPS6KA2, SGK2} | 8     |
| 3              | {ACVR2B, ALPK1, CSF1R, GS3K3} | 4     |
| 4              | {FBXL10, HDAC6, POLDI, PRMT7, SETMAR, SIRT1, SYMYD4} | 7     |

Supplementary Table S4. Leaf nodes belonging to the clusters formed by the hierarchical clustering for overexpressed genes. ‘Cluster node id’ corresponds to the cluster node ids in the Supplementary Fig. S2 (B). The nodes of the clusters are listed in the ‘Overexpressed genes’ column. The number of over-expressed genes belonging to the clusters are shown in the ‘Count’ column represents.
DPK2, 'DARS', 'DBI', 'DEK', 'DEPCD5', 'DET1', 'DHCR24', 'DHX8', 'DIDO1', 'DLGAP5', 'DLX2', 'DMRTA1', 'DNAJB6', 'DNM1L', 'DPF2', 'DSE', 'DTNA', 'DTXL1', 'DTX4', 'DUT', 'DYNLT3', 'EBF4', 'EIF2AK1', 'EIF4A2', 'ENST00000553644', 'EPHA3', 'EPHX2', 'EPSP', 'ERBB4', 'ERBB4_E872K', 'EYA1', 'EYFL1', 'F3', 'FAM14A2', 'FBXO10', 'FBXO11', 'FBXO3', 'FBXO4', 'FBXO40', 'FERD3L', 'FEZ2', 'FGF3R3', 'FIRG', 'FKBP13', 'FOXC2', 'FOXN3', 'FOXR1', 'FTL', 'FZD7', 'GAA', 'GABARAPL1', 'GFOD1', 'GJA1', 'GLI3', 'GNA11', 'GNA15', 'GNAI1', 'GNAO', 'GNAZ', 'GNB2', 'GOT1', 'GPR128', 'GPR83', 'GRHL3', 'GRPR', 'GS2G', 'GSTM1', 'GT1F2', 'GTHDA', 'GTHDA', 'HL', 'HBP1', 'Hdac11', 'Hhex1', 'Hsthi1a', 'Hsthi1b', 'HSF5', 'Icam1', 'Icpt', 'IGF2', 'IIL10', 'INPP1', 'Irf3', 'Irf9', 'Irs1', 'Kcnna6', 'Kcnkp', 'Kdct', 'Klf10', 'Klhl12', 'Klhl12', 'Klhl30', 'Lamtor3', 'Lars2', 'Lbr', 'Lcor', 'Left1', 'Lepr', 'Lgals8', 'Lias', 'Lonrf3', 'Lsm6e', 'Lyk5', 'Lypla2', 'Maelli', 'Magea10', 'Maoa', 'Mapk3', 'Mapk2', 'Mapk5', 'Mapk8', 'Mapk15', 'Mapk3', 'Mas', 'Mast2', 'Mct7', 'Mettl14', 'Met_t1252c', 'Mtx3d', 'Mgc16169', 'Mlna', 'Mnkn1', 'Mnk2', 'Mhl1', 'Mzp1', 'Mrgrp3x', 'Mtrfr', 'Naa85', 'Nagk', 'Nat1', 'Ncaoa1', 'Ncoa3', 'Nedd4l', 'Nfats3', 'Nmi', 'Nqo2', 'Nrobi', 'Nrp3', 'Nrp1', 'Nudt4', 'Nudt6', 'Nup62', 'Oggi', 'Ovol1', 'Paq7', 'Paq7', 'Paq7', 'Pax7', 'Pck', 'Pck2', 'Pctk3', 'Pdgfrb', 'Phda1', 'Pdk4', 'Pemt', 'Pfn2', 'Pftk2', 'Thf19', 'Pias1', 'Pigk', 'Pik3cb', 'Pik3cd', 'Pik3r1', 'Pip3na', 'Plag12a2', 'Pnn', 'Pola2', 'Pole', 'Pole4', 'Ppox', 'Pprrc2', 'Pprrc3c', 'Ppt1', 'Prmd14', 'Prkaa1', 'Prkaca', 'Prkcb', 'Prkcg', 'Prkkrk', 'Prnp', 'Prpff19', 'Prpff6', 'Prs15l1', 'Prss52', 'Ps4h1', 'Psma3', 'Psmb10', 'Ptck2', 'Pttna3', 'Pvrl2', 'Rab11a', 'Rab11f2p', 'Rab17', 'Raba5', 'Rala', 'Rad14', 'Rassf1', 'Rbbp7', 'Rbm15', 'Rcor3', 'Rgln1', 'Rnf3', 'Rnf31', 'Ros', 'Rpl22', 'Rpl9l1', 'Rps15a', 'Rps3', 'Rps6ka3', 'Rps6ka5', 'Rrp3', 'Rufy1', 'Ruvbl1', 'S100a1', 'S100a13', 'Scaffe', 'Scyl3', 'Shmt4', 'Sirt4', 'Sirt5', 'Sloc2a18', 'Sloc2a13', 'Sloc2a3', 'Sloc3a8', 'Slnn1', 'Smad7', 'Smarca2', 'Smu1', 'Soat1', 'Socs6', 'Sos2', 'Sox1', 'Sp1c', 'Squ', 'Srpk3', 'Sxc3', 'St18', 'Stk17a', 'Stk24', 'S tad3', 'Taf15', 'Tbic198b', 'Tbk1', 'Tcf7l2', 'Tdrd3', 'Terf2ip', 'Tesk1', 'Tfdp1', 'Thap5', 'Tig4d', 'Tig7d', 'Timm59', 'Tmed9', 'Tmem110', 'Tmem174', 'Tnsfsf6b', 'Tnsfsf6c', 'Tnka', 'Too', 'Tox', 'Tpx1', 'Traf11', 'Trim21', 'Trim23', 'Trim23', 'Trim36', 'Trim4', 'Tsc22d4', 'Tspa6', 'Tsp510', 'Tssk6', 'Tuhb48', 'Ube2e3', 'Usp1', 'Usp15', 'Utp14a', 'Vapb', 'Wnt7b', 'Yes1', 'Yhtdfd1', 'Ywhaa', 'Zbtb45', 'Zchc3', 'Zmi3', 'Zkscan1', 'Zkscan2', 'Znf175', 'Znf192', 'Znf195', 'Znf206', 'Znf263', 'Znf274', 'Znf343', 'Znf350', 'Znf415', 'Znf449', 'Znf45', 'Znf50', 'Znf505', 'Znf511', 'Znf513', 'Znf577', 'Znf585a', 'Znf585b', 'Znf585c', 'Znf589', 'Znf777', 'Znf785', 'Znf8197', 'Zscan29'
**Supplementary Table S5.** The distribution of inhibitory targets repositioned from the original disease class to other disease classes. Each element represents the number of inhibitory targets repositioned. The rows indicate the original ICD disease chapters, and the number in parentheses represents the number of known inhibitory targets belonging to the disease chapters. The columns indicate the newly predicted ICD disease chapters. The chapters are as follows: Chapter I: certain infectious or parasitic dis-eases; Chapter II: neoplasms; Chapter III: diseases of the blood or blood-forming organs; Chapter IV: diseases of the immune system; Chapter V: endocrine, nutritional, or metabolic diseases; Chapter VI: mental, behavioral, or neurodevelopmental disorders; Chapter VII: sleep–wake disorders; Chapter VIII: diseases of the nervous system; Chapter IX: diseases of the visual system; Chapter X: diseases of the ear or mastoid process; Chapter XI: diseases of the circulatory system; Chapter XII: diseases of the respiratory system; Chapter XIII: diseases of the digestive system; Chapter XIV: diseases of the skin; Chapter XV: diseases of the musculoskeletal system or connective tissue; Chapter XVI: diseases of the genitourinary system; Chapter XVII: conditions related to sexual health; Chapter XVIII: pregnancy, childbirth, or the puerperium; Chapter XIX: certain conditions originating in the perinatal period; Chapter XX: developmental anomalies; Chapter XXI: symptoms, signs, or clinical findings, not else-where classified; and Chapter XXII: injury, poisoning, or certain other consequences of external causes. A network visualizing this matrix is shown in Fig. 5 (A).

| ICD chapter (# of targets) | I  | II | III | IV | V  | VI | VII | VIII | IX | XI | XII | XIII | XIV | XV  | XVI | XVII | XVIII | XIX | XX  | XXI | XXII |
|---------------------------|----|----|-----|----|----|----|-----|-------|----|----|-----|-------|-----|-----|------|-------|------|-----|-----|-----|------|
| chapter I (4)             | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter II (145)          | 0  | 34 | 8   | 0  | 0  | 0  | 15  | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 7    | 4     | 0    | 0   | 0   | 0   | 0    |
| chapter III (6)           | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter IV (20)           | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 1    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter V (31)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter VI (4)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter VII (4)           | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter VIII (13)         | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter IX (0)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter X (0)             | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XI (0)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XII (26)          | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XIII (20)         | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XIV (16)          | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XV (47)           | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XVI (0)           | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XVII (0)          | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XVIII (0)         | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XIX (0)           | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XX (0)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XXI (0)           | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
| chapter XXII (0)          | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0     | 0  | 0  | 0   | 0     | 0   | 0   | 0    | 0     | 0    | 0   | 0   | 0   | 0    |
**Supplementary Table S6.** The distribution of activatory targets repositioned from the original disease class to other disease classes. Each element represents the number of activatory targets repositioned. The rows indicate the original ICD disease chapters, and the number in parentheses represents the number of known activatory targets belonging to the disease chapters. The columns indicate the newly predicted ICD disease chapters. The chapters are as follows: Chapter I: certain infectious or parasitic dis-eases; Chapter II: neoplasms; Chapter III: diseases of the blood or blood-forming organs; Chapter IV: diseases of the immune system; Chapter V: endocrine, nutritional, or metabolic diseases; Chapter VI: mental, behavioral, or neurodevelopmental disorders; Chapter VII: sleep–wake disorders; Chapter VIII: diseases of the nervous system; Chapter IX: diseases of the visual system; Chapter X: diseases of the ear or mastoid process; Chapter XI: diseases of the circulatory system; Chapter XII: diseases of the respiratory system; Chapter XIII: diseases of the digestive system; Chapter XIV: diseases of the skin; Chapter XV: diseases of the musculoskeletal system or connective tissue; Chapter XVI: diseases of the genitourinary system; Chapter XVII: conditions related to sexual health; Chapter XVIII: pregnancy, childbirth, or the puerperium; Chapter XIX: certain conditions originating in the perinatal period; Chapter XX: developmental anomalies; Chapter XXI: symptoms, signs, or clinical findings, not else-where classified; and Chapter XXII: injury, poisoning, or certain other consequences of external causes. A network visualizing this matrix is shown in Fig. 5 (B).

| ICD chapter (# of targets) | I  | II | III | IV | V  | VI | VII | VIII | IX | XI | XII | XIII | XIV | XV  | XVI | XVII | XVIII | XIX | XX | XI | XXII |
|----------------------------|----|----|-----|----|----|----|-----|------|----|----|-----|------|-----|-----|-----|------|------|-----|----|-----|------|
| chapter I (0)              | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter II (9)             | 0  | 2  | 0   | 2  | 4  | 0  | 0   | 4    | 0  | 0  | 0   | 0    | 0   | 0   | 4   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter III (0)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter IV (1)             | 0  | 2  | 0   | 0  | 1  | 0  | 0   | 1    | 0  | 0  | 0   | 0    | 0   | 1   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter V (16)             | 0  | 8  | 0   | 1  | 4  | 0  | 0   | 5    | 0  | 0  | 0   | 0    | 0   | 0   | 4   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter VI (0)             | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter VII (0)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter VIII (2)           | 0  | 2  | 0   | 1  | 2  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 1   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter IX (0)             | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter X (0)              | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XI (0)             | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XII (0)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XIII (1)           | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XIV (0)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XV (1)             | 0  | 2  | 0   | 0  | 0  | 0  | 0   | 1    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XVI (0)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XVII (0)           | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XVIII (0)          | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XIX (0)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XX (1)             | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XXI (0)            | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
| chapter XXII (0)           | 0  | 0  | 0   | 0  | 0  | 0  | 0   | 0    | 0  | 0  | 0   | 0    | 0   | 0   | 0   | 0    | 0    | 0   | 0  | 0   | 0    |
**Supplementary Table S7.** Newly predicted 52 inhibitory targets repositioned from the original disease to other diseases. These inhibitory target–disease associations are predicted by only the trans-disease method with the completed data. Inhibitory targets are listed in decreasing order of predicted scores. Predicted scores represent the therapeutic targetability for the diseases.

| Rank | Diseases                                 | Inhibitory targets | Predicted scores |
|------|------------------------------------------|--------------------|------------------|
| 1    | congenital muscular dystrophies (CMD/MDC)| MAPK11             | 1                |
| 2    | congenital muscular dystrophies (CMD/MDC)| TNFSF13B          | 0.847            |
| 3    | congenital muscular dystrophies (CMD/MDC)| TSLP              | 0.784            |
| 4    | congenital muscular dystrophies (CMD/MDC)| CSF1R             | 0.739            |
| 5    | congenital muscular dystrophies (CMD/MDC)| PARP1             | 0.639            |
| 6    | congenital muscular dystrophies (CMD/MDC)| AURKC             | 0.454            |
| 7    | cervical cancer                          | MAPK11            | 0.414            |
| 8    | congenital muscular dystrophies (CMD/MDC)| MUC1              | 0.363            |
| 9    | congenital muscular dystrophies (CMD/MDC)| TOP2B            | 0.325            |
| 10   | systemic lupus erythematosus             | MAPK11            | 0.32             |
| 11   | ovarian cancer                           | TNFSF13B          | 0.271            |
| 12   | cervical cancer                          | TNFSF13B          | 0.27             |
| 13   | chronic myeloid leukemia (CML)           | MAPK11            | 0.25             |
| 14   | ovarian cancer                           | TSLP              | 0.242            |
| 15   | congenital muscular dystrophies (CMD/MDC)| TUBB2A           | 0.235            |
| 16   | cervical cancer                          | TSLP              | 0.228            |
| 17   | ovarian cancer                           | CSF1R             | 0.215            |
| 18   | systemic lupus erythematosus             | TSLP              | 0.2              |
| 19   | cervical cancer                          | CSF1R             | 0.192            |
| 20   | systemic lupus erythematosus             | CSF1R             | 0.175            |
| 21   | chronic myeloid leukemia (CML)           | TNFSF13B          | 0.164            |
| 22   | atopic dermatitis                        | MAPK11            | 0.149            |
| 23   | renal cell carcinoma                     | MAPK11            | 0.142            |
| 24   | chronic myeloid leukemia (CML)           | TSLP              | 0.134            |
| 25   | systemic lupus erythematosus             | PARP1             | 0.12             |
| 26   | atopic dermatitis                        | TNFSF13B          | 0.091            |
| 27   | gastric cancer                           | MAPK11            | 0.082            |
| 28   | breast cancer                            | MAPK11            | 0.079            |
| 29   | rheumatoid arthritis                     | MAPK11            | 0.076            |
| 30   | renal cell carcinoma                     | TNFSF13B          | 0.071            |
| 31   | chronic myeloid leukemia (CML)           | PARP1             | 0.061            |
| 32   | atopic dermatitis                        | CSF1R             | 0.056            |
| 33   | renal cell carcinoma                     | TSLP              | 0.05             |
| 34   | breast cancer                            | TNFSF13B          | 0.049            |
| 35   | multiple myeloma                         | TNFSF13B          | 0.045            |
| 36   | ovarian cancer                           | AURKC             | 0.044            |
| 37   | breast cancer                            | TSLP              | 0.044            |
| 38   | breast cancer                            | CSF1R             | 0.041            |
| 39   | multiple myeloma                         | TSLP              | 0.035            |
| 40   | rheumatoid arthritis                     | TSLP              | 0.031            |
| 41   | colorectal cancer                        | MAPK11            | 0.029            |
| 42   | gastric cancer                           | TNFSF13B          | 0.027            |
| 43   | multiple myeloma                         | CSF1R             | 0.026            |
| 44   | atopic dermatitis                        | PARP1             | 0.023            |
| 45   | pancreatic cancer                        | MAPK11            | 0.022            |
| 46   | rheumatoid arthritis                     | CSF1R             | 0.019            |
| 47   | systemic lupus erythematosus             | AURKC             | 0.018            |
| 48   | gastric cancer                           | TSLP              | 0.013            |
| 49   | multiple myeloma                         | PARP1             | 0.009            |
| 50   | gastric cancer                           | CSF1R             | 0.001            |
| 26 | chronic myeloid leukemia (CML) | CSF1R | 0.113 |
|----|-------------------------------|-------|-------|
| 52 | breast cancer                 | AURKC | 0     |
**Supplementary Table S8.** Newly predicted 46 activatory targets repositioned from the original disease to other diseases. These activatory target–disease associations are predicted by only the trans-disease method with the completed data. Activatory targets are listed in decreasing order of predicted scores. Predicted scores represent the therapeutic targetability for the diseases.

| Rank | Diseases                                | Activatory targets | Predicted scores |
|------|-----------------------------------------|--------------------|------------------|
| 1    | acute myeloid leukemia (AML)            | TH                 | 1                |
| 2    | rheumatoid arthritis                    | TH                 | 0.981            |
| 3    | type I diabetes mellitus                | TH                 | 0.981            |
| 4    | chronic granulomatous disease           | IFNG               | 0.929            |
| 5    | acute myeloid leukemia (AML)            | IFNG               | 0.871            |
| 6    | rheumatoid arthritis                    | IFNG               | 0.855            |
| 7    | type I diabetes mellitus                | IFNG               | 0.855            |
| 8    | acute myeloid leukemia (AML)            | SOCS2              | 0.799            |
| 9    | rheumatoid arthritis                    | SOCS2              | 0.789            |
| 10   | type I diabetes mellitus                | SOCS2              | 0.789            |
| 11   | acute myeloid leukemia (AML)            | PRKAB1             | 0.625            |
| 12   | rheumatoid arthritis                    | PRKAB1             | 0.607            |
| 13   | type I diabetes mellitus                | PRKAB1             | 0.607            |
| 14   | chronic granulomatous disease           | SOCS2              | 0.591            |
| 15   | breast cancer                           | TH                 | 0.545            |
| 16   | acute myeloid leukemia (AML)            | SOCS4              | 0.503            |
| 17   | rheumatoid arthritis                    | SOCS4              | 0.496            |
| 18   | type I diabetes mellitus                | SOCS4              | 0.496            |
| 19   | acute myeloid leukemia (AML)            | FFAR1              | 0.46             |
| 20   | breast cancer                           | IFNG               | 0.45             |
| 21   | rheumatoid arthritis                    | FFAR1              | 0.446            |
| 22   | type I diabetes mellitus                | FFAR1              | 0.446            |
| 23   | acute myeloid leukemia (AML)            | CTLA4              | 0.367            |

| Rank | Diseases                                | Activatory targets | Predicted scores |
|------|-----------------------------------------|--------------------|------------------|
| 24   | Parkinson's disease (PD)                | IFNG               | 0.35             |
| 25   | type I diabetes mellitus                | SYK                | 0.328            |
| 26   | rheumatoid arthritis                    | SYK                | 0.328            |
| 27   | Parkinson's disease (PD)                | PRKAB1             | 0.32             |
| 28   | Parkinson's disease (PD)                | FFAR1              | 0.29             |
| 29   | Parkinson's disease (PD)                | SOCS2              | 0.287            |
| 30   | breast cancer                           | PRKAB1             | 0.265            |
| 31   | chronic granulomatous disease           | PRKAB1             | 0.264            |
| 32   | Parkinson's disease (PD)                | SOCS4              | 0.23             |
| 33   | Parkinson's disease (PD)                | SYK                | 0.225            |
| 34   | Parkinson's disease (PD)                | CTLA4              | 0.225            |
| 35   | Parkinson's disease (PD)                | PRKAG3             | 0.179            |
| 36   | breast cancer                           | FFAR1              | 0.143            |
| 37   | acute myeloid leukemia (AML)            | PRKAG3             | 0.132            |
| 38   | Parkinson's disease (PD)                | RAPGEF4            | 0.131            |
| 39   | type I diabetes mellitus                | PRKAG3             | 0.122            |
| 40   | rheumatoid arthritis                    | PRKAG3             | 0.122            |
| 41   | type II diabetes mellitus               | DRD1               | 0.116            |
| 42   | chronic granulomatous disease           | SOCS4              | 0.076            |
| 43   | breast cancer                           | CTLA4              | 0.062            |
| 44   | acute myeloid leukemia (AML)            | RAPGEF4            | 0.002            |
| 45   | type I diabetes mellitus                | RAPGEF4            | 0                |
| 46   | rheumatoid arthritis                    | RAPGEF4            | 0                |
**Supplementary Table S9.** Top 100 novel inhibitory target–disease pairs predicted using the trans-disease method with completed data. These predicted targets include not only existing therapeutic targets, but also completely new therapeutic targets that have not any known target–disease associations. Inhibitory targets are listed in decreasing order of predicted scores. Predicted scores represent the therapeutic targetability for the diseases.

| Rank | Diseases                          | Inhibitory targets | Predicted scores |
|------|-----------------------------------|--------------------|------------------|
| 1    | adult T-cell leukemia              | TAF1B              | 1.000            |
| 2    | endometrial Cancer                 | TAF1B              | 0.896            |
| 3    | adult T-cell leukemia              | USP9X              | 0.871            |
| 4    | type I diabetes mellitus           | TAF1B              | 0.868            |
| 5    | hepatitis C                        | TAF1B              | 0.867            |
| 6    | endometrial Cancer                 | USP9X              | 0.832            |
| 7    | testicular cancer                  | TAF1B              | 0.831            |
| 8    | type I diabetes mellitus           | USP9X              | 0.826            |
| 9    | adult T-cell leukemia              | TLK1               | 0.820            |
| 10   | hepatitis C                        | USP9X              | 0.819            |
| 11   | tuberculosis                       | TAF1B              | 0.816            |
| 12   | adult T-cell leukemia              | AATF               | 0.816            |
| 13   | adult T-cell leukemia              | ADA                | 0.814            |
| 14   | hypercholesterolemia               | TAF1B              | 0.813            |
| 15   | type I diabetes mellitus           | CD3D               | 0.813            |
| 16   | chronic lymphocytic leukemia (CLL) | TAF1B              | 0.810            |
| 17   | type I diabetes mellitus           | TLK1               | 0.809            |
| 18   | type I diabetes mellitus           | ABCC1              | 0.807            |
| 19   | hepatitis C                        | EPCAM              | 0.807            |
| 20   | endometrial Cancer                 | TLK1               | 0.807            |
| 21   | type I diabetes mellitus           | A2M                | 0.807            |
| 22   | tuberculosis                       | HRH1               | 0.806            |
| 23   | testicular cancer                  | USP9X              | 0.806            |
| 24   | endometrial Cancer                 | FGFR2              | 0.806            |
| 25   | endometrial Cancer                 | AATF               | 0.806            |
| 26   | acute myeloid leukemia (AML)       | TAF1B              | 0.803            |
| 27   | small cell lung cancer             | SHH                | 0.801            |
| 28   | hepatitis C                        | TLK1               | 0.800            |
| 29   | hepatitis C                        | AATF               | 0.800            |
| 30   | congenital muscular dystrophies    | MAP3K12            | 0.800            |
| 31   | aplastic anemia (AA)               | PPP3CC             | 0.800            |
| 32   | cervical cancer                    | TGFBR2             | 0.800            |
| 33   | type I diabetes mellitus           | TNFSF4             | 0.800            |
| 34   | small cell lung cancer             | MMP14              | 0.800            |
| 35   | cervical cancer                    | ERBB3              | 0.800            |
| 36   | small cell lung cancer             | MMP1               | 0.800            |
| 37   | adult T-cell leukemia              | ABAT               | 0.800            |
| 38   | congenital muscular dystrophies    | SLC2A4RG           | 0.800            |
| 39   | congenital muscular dystrophies    | XYL1               | 0.799            |
| 40   | hepatitis C                        | ABAT               | 0.799            |
| Rank | Diseases                        | Inhibitory targets | Predicted scores |
|------|---------------------------------|--------------------|------------------|
| 41   | cervical cancer                 | TYMS               | 0.799            |
| 42   | congenital muscular dystrophies | ADCK5              | 0.799            |
| 43   | Hypercholesterolemia            | USP9X              | 0.799            |
| 44   | congenital muscular dystrophies | GLYCTK             | 0.799            |
| 45   | congenital muscular dystrophies | EPHA10             | 0.799            |
| 46   | asthma                          | IL4R               | 0.799            |
| 47   | endometrial Cancer              | ABAT               | 0.799            |
| 48   | tuberculosis                    | USP9X              | 0.799            |
| 49   | congenital muscular dystrophies | PLA2G12A           | 0.799            |
| 50   | inflammatory bowel disease      | PPP3CC             | 0.799            |
| 51   | congenital muscular dystrophies | P2RY2              | 0.799            |
| 52   | congenital muscular dystrophies | NPR1               | 0.799            |
| 53   | malignant melanoma              | MITF               | 0.799            |
| 54   | type I diabetes mellitus        | ABCB1              | 0.799            |
| 55   | congenital muscular dystrophies | PDGFRL             | 0.799            |
| 56   | congenital muscular dystrophies | MAPK11             | 0.799            |
| 57   | congenital muscular dystrophies | ASPM               | 0.799            |
| 58   | type I diabetes mellitus        | ABCA5              | 0.799            |
| 59   | congenital muscular dystrophies | STX16              | 0.799            |
| 60   | congenital muscular dystrophies | PLK3               | 0.799            |
| 61   | congenital muscular dystrophies | MAGI1              | 0.798            |
| 62   | congenital muscular dystrophies | CERKL              | 0.798            |
| 63   | congenital muscular dystrophies | PIPI4K2A           | 0.798            |
| 64   | endometrial Cancer              | ABCB1              | 0.798            |
| 65   | congenital muscular dystrophies | EPHB3              | 0.798            |
| 66   | congenital muscular dystrophies | SMARCA2            | 0.798            |
| 67   | congenital muscular dystrophies | MST4               | 0.798            |
| 68   | cervical cancer                 | BRCA1              | 0.798            |
| 69   | congenital muscular dystrophies | ZBTB20             | 0.798            |
| 70   | congenital muscular dystrophies | RELN               | 0.798            |
| 71   | testicular cancer               | TUBB6              | 0.798            |
| 72   | congenital muscular dystrophies | ICK                | 0.798            |
| 73   | congenital muscular dystrophies | ACSL6              | 0.798            |
| 74   | congenital muscular dystrophies | AATK               | 0.798            |
| 75   | Crohn's disease                 | ITGB2              | 0.798            |
| 76   | congenital muscular dystrophies | CDK12              | 0.798            |
| 77   | congenital muscular dystrophies | TAO1               | 0.798            |
| 78   | congenital muscular dystrophies | MAP3K10            | 0.798            |
| 79   | congenital muscular dystrophies | ZNF169             | 0.798            |
| 80   | congenital muscular dystrophies | GSC                | 0.798            |
| 81   | congenital muscular dystrophies | CD2                | 0.798            |
| 82   | type I diabetes mellitus        | TRIB3              | 0.798            |
| 83   | congenital muscular dystrophies | RNIF133            | 0.798            |
| 84   | congenital muscular dystrophies | NQO2               | 0.798            |
| 85   | congenital muscular dystrophies | PDK1L              | 0.798            |
| 86   | congenital muscular dystrophies | LOC388259          | 0.798            |
| 87   | congenital muscular dystrophies | LOC441777          | 0.798            |
| 88   | congenital muscular dystrophies | NUMB               | 0.798            |
| 89   | congenital muscular dystrophies | MORN2              | 0.798            |
| 90   | inflammatory bowel disease      | IL23R              | 0.798            |
| Rank | Diseases                              | Inhibitory targets | Predicted scores |
|------|---------------------------------------|--------------------|-----------------|
| 91   | congenital muscular dystrophies       | NRXN3              | 0.798           |
| 92   | malignant melanoma                    | BRAF               | 0.798           |
| 93   | asthma                                | TBX A2R            | 0.798           |
| 94   | congenital muscular dystrophies       | RBM8A              | 0.798           |
| 95   | congenital muscular dystrophies       | PSAT1              | 0.798           |
| 96   | congenital muscular dystrophies       | CDK14              | 0.798           |
| 97   | congenital muscular dystrophies       | MAP3K6             | 0.797           |
| 98   | congenital muscular dystrophies       | PTPN5              | 0.797           |
| 99   | congenital muscular dystrophies       | BMPR2              | 0.797           |
| 100  | congenital muscular dystrophies       | PRK Y              | 0.797           |
**Supplementary Table S10.** Top 100 novel activatory target–disease pairs predicted using the trans-disease method with completed data. These predicted targets include not only existing therapeutic targets, but also completely new therapeutic targets that have not any known target–disease associations. Activatory targets are listed in decreasing order of predicted scores. Predicted scores represent the therapeutic targetability for the diseases.

| Rank | Diseases                        | Activatory targets | Predicted scores |
|------|---------------------------------|--------------------|------------------|
| 1    | Parkinson's disease (PD)        | OLIG3              | 1.000            |
| 2    | chronic myeloid leukemia (CML)  | IFNAR2             | 0.999            |
| 3    | multiple myeloma                | IFNAR2             | 0.998            |
| 4    | Parkinson's disease (PD)        | NDUFC2             | 0.998            |
| 5    | ovarian cancer                  | FBXW7              | 0.995            |
| 6    | Rett syndrome                   | MECP2              | 0.992            |
| 7    | type II diabetes mellitus       | RPS6KA2            | 0.992            |
| 8    | type II diabetes mellitus       | SIRT1              | 0.990            |
| 9    | Parkinson's disease (PD)        | FOXO4              | 0.990            |
| 10   | type II diabetes mellitus       | PDK1L              | 0.989            |
| 11   | type II diabetes mellitus       | MEF2C              | 0.989            |
| 12   | type II diabetes mellitus       | SETMAR             | 0.988            |
| 13   | Parkinson's disease (PD)        | MYO3B              | 0.987            |
| 14   | colorectal cancer               | TP53               | 0.987            |
| 15   | type II diabetes mellitus       | PPARG              | 0.986            |
| 16   | Crohn's disease                 | IL10               | 0.984            |
| 17   | inflammatory bowel disease      | IL10               | 0.984            |
| 18   | Parkinson's disease (PD)        | BCR-ABL            | 0.983            |
| 19   | Parkinson's disease (PD)        | ZNF384             | 0.983            |
| 20   | Parkinson's disease (PD)        | FLJ25006           | 0.983            |

| Rank | Diseases                        | Activatory targets | Predicted scores |
|------|---------------------------------|--------------------|------------------|
| 21   | Parkinson's disease (PD)        | HMGB4              | 0.983            |
| 22   | Parkinson's disease (PD)        | DDAH2              | 0.983            |
| 23   | type II diabetes mellitus       | BCR-ABL            | 0.982            |
| 24   | type II diabetes mellitus       | CLK1               | 0.981            |
| 25   | Parkinson's disease (PD)        | TRIM27             | 0.981            |
| 26   | Parkinson's disease (PD)        | MAGEA10            | 0.981            |
| 27   | Parkinson's disease (PD)        | MATK               | 0.981            |
| 28   | chronic granulomatous disease   | SYF2               | 0.981            |
| 29   | type II diabetes mellitus       | HEY1               | 0.981            |
| 30   | type II diabetes mellitus       | POLD1              | 0.981            |
| 31   | Parkinson's disease (PD)        | PRKAR2A            | 0.980            |
| 32   | Parkinson's disease (PD)        | RAF1               | 0.980            |
| 33   | type II diabetes mellitus       | CDIP1              | 0.979            |
| 34   | Parkinson's disease (PD)        | ETV5               | 0.979            |
| 35   | Parkinson's disease (PD)        | ZNF582             | 0.979            |
| 36   | Parkinson's disease (PD)        | DLX3               | 0.979            |
| 37   | type II diabetes mellitus       | FBXL10             | 0.979            |
| 38   | type II diabetes mellitus       | EGFR               | 0.979            |
| 39   | chronic granulomatous disease   | ASPHD1             | 0.979            |
| 40   | Parkinson's disease (PD)        | PRDM14             | 0.979            |
| Rank | Diseases                        | Activatory targets | Predicted scores |
|------|--------------------------------|--------------------|------------------|
| 41   | type II diabetes mellitus       | BRAF               | 0.979            |
| 42   | type II diabetes mellitus       | GPR119             | 0.979            |
| 43   | type II diabetes mellitus       | TP53RK             | 0.978            |
| 44   | Parkinson's disease (PD)        | CREB3              | 0.978            |
| 45   | type II diabetes mellitus       | NAP1L1             | 0.978            |
| 46   | Parkinson's disease (PD)        | ZNF268             | 0.978            |
| 47   | Parkinson's disease (PD)        | NDUF8              | 0.978            |
| 48   | type II diabetes mellitus       | GMEB1              | 0.978            |
| 49   | Parkinson's disease (PD)        | ZNF592             | 0.978            |
| 50   | type II diabetes mellitus       | EHM72              | 0.978            |
| 51   | chronic granulomatous disease   | C2                 | 0.977            |
| 52   | Parkinson's disease (PD)        | ATOH1              | 0.977            |
| 53   | Parkinson's disease (PD)        | FOXN2              | 0.977            |
| 54   | type II diabetes mellitus       | ESR2               | 0.977            |
| 55   | chronic granulomatous disease   | ZNF22              | 0.977            |
| 56   | Parkinson's disease (PD)        | NEK9               | 0.977            |
| 57   | Parkinson's disease (PD)        | IRF5               | 0.977            |
| 58   | Parkinson's disease (PD)        | DEPC1              | 0.977            |
| 59   | type II diabetes mellitus       | FARSA              | 0.977            |
| 60   | Parkinson's disease (PD)        | NR1H3              | 0.977            |
| 61   | Parkinson's disease (PD)        | MAP3K6             | 0.977            |
| 62   | type II diabetes mellitus       | SGK2               | 0.977            |
| 63   | chronic granulomatous disease   | WDC7C              | 0.976            |
| 64   | Parkinson's disease (PD)        | ALK                | 0.976            |
| 65   | Parkinson's disease (PD)        | BID                | 0.976            |
| 66   | Parkinson's disease (PD)        | ZIK1               | 0.976            |
| 67   | renal cell carcinoma            | IL2RB              | 0.976            |
| 68   | Parkinson's disease (PD)        | LRR2               | 0.976            |
| 69   | type II diabetes mellitus       | CUL5               | 0.976            |
| 70   | Parkinson's disease (PD)        | CDC42SE1           | 0.976            |
| 71   | Parkinson's disease (PD)        | PCK2               | 0.976            |
| 72   | Parkinson's disease (PD)        | CREM               | 0.976            |
| 73   | chronic granulomatous disease   | RXFP1              | 0.976            |
| 74   | type II diabetes mellitus       | HESX1              | 0.976            |
| 75   | chronic granulomatous disease   | RG9MTD3            | 0.976            |
| 76   | acute myeloid leukemia (AML)    | SYF2               | 0.976            |
| 77   | Parkinson's disease (PD)        | AP1S2              | 0.976            |
| 78   | chronic granulomatous disease   | TIMM17B            | 0.976            |
| 79   | Parkinson's disease (PD)        | RFX3               | 0.976            |
| 80   | type II diabetes mellitus       | ZNF22              | 0.976            |
| 81   | Parkinson's disease (PD)        | ZNF74              | 0.976            |
| 82   | type II diabetes mellitus       | PPARD              | 0.976            |
| 83   | Parkinson's disease (PD)        | MAP3K8             | 0.975            |
| 84   | renal cell carcinoma            | IFNGR1             | 0.975            |
| 85   | type II diabetes mellitus       | PHF23              | 0.975            |
| 86   | rheumatoid arthritis            | SYF2               | 0.975            |
| 87   | type I diabetes mellitus        | SYF2               | 0.975            |
| 88   | chronic granulomatous disease   | TXNIP              | 0.975            |
| 89   | Parkinson's disease (PD)        | NEU1               | 0.975            |
| 90   | type II diabetes mellitus       | GOT1               | 0.975            |
| Rank | Diseases                          | Activatory targets | Predicted scores |
|------|-----------------------------------|--------------------|------------------|
| 91   | Parkinson's disease (PD)          | STK32C             | 0.975            |
| 92   | type II diabetes mellitus         | FOXP4              | 0.975            |
| 93   | Parkinson's disease (PD)          | MXI1               | 0.975            |
| 94   | acute myeloid leukemia (AML)      | ASPHD1             | 0.975            |
| 95   | chronic granulomatous disease     | TUBB4B             | 0.975            |
| 96   | type II diabetes mellitus         | HSD17B4            | 0.974            |
| 97   | Parkinson's disease (PD)          | TFDP1              | 0.974            |
| 98   | rheumatoid arthritis              | ASPHD1             | 0.974            |
| 99   | type I diabetes mellitus          | ASPHD1             | 0.974            |
| 100  | Parkinson's disease (PD)          | SMYD1              | 0.974            |
Supplementary Table S11. Performance evaluation of data completion by tensor decomposition algorithms for third-order transcriptome data (knocked down genes, genes, and cell lines) with different rates of artificial missing values. Missing values are generated by the “random missing” strategy. Relative standard errors (RSEs) between the original and reconstructed data from tensor decomposition are calculated for (a) all values and (b) missing values only. The tensor decomposition method used in this study and the baseline method are denoted as TT (Yuan et al., 2017) and CP (Acar et al., 2010), respectively. The optimized tensor ranks are shown for each method. Artificially generated missing rates: 10%, 50%, and 90% are tested. Cell lines are listed in order of increasing original missing rates.

| Artificial missing rate | 10% | 20% | 30% | 50% | 70% | 90% |
|-------------------------|-----|-----|-----|-----|-----|-----|
| CP (baseline)           |     |     |     |     |     |     |
| CP-ranks (this study)   |     |     |     |     |     |     |
| TT-ranks (this study)   |     |     |     |     |     |     |
| CP (baseline)           |     |     |     |     |     |     |
| CP-ranks (this study)   |     |     |     |     |     |     |
| TT-ranks (this study)   |     |     |     |     |     |     |
| CP (baseline)           |     |     |     |     |     |     |
| CP-ranks (this study)   |     |     |     |     |     |     |
| TT-ranks (this study)   |     |     |     |     |     |     |

(a) RSEs for all values

| Total cell lines | CP-ranks | TT-ranks |
|------------------|----------|----------|
| 0.0476           | 0.0455   | 0.0496   |
| 0.0613           | 0.0599   | 0.0629   |
| 0.0620           | 0.0597   | 0.0637   |
| 0.0632           | 0.0579   | 0.0652   |
| 0.0676           | 0.0634   | 0.0696   |
| 0.0662           | 0.0615   | 0.0683   |
| 0.0616           | 0.0590   | 0.0634   |
| 0.0607           | 0.0589   | 0.0624   |
| 0.0612           | 0.0592   | 0.0632   |
| 0.0528           | 0.0515   | 0.0550   |
| 0.0241           | 0.0238   | 0.0269   |
| 0.0367           | 0.0350   | 0.0386   |
| 0.0183           | 0.0178   | 0.0223   |
| 0.0316           | 0.0313   | 0.0341   |
| 0.0123           | 0.0117   | 0.0173   |
| 0.0209           | 0.0206   | 0.0243   |
| 0.0061           | 0.0047   | 0.0137   |
| 0.0043           | 0.0016   | 0.0138   |

(b) RSEs for missing values

| Total cell lines | CP-ranks | TT-ranks |
|------------------|----------|----------|
| 0.0699           | 0.0668   | 0.0719   |
| 0.0656           | 0.0641   | 0.0674   |
| 0.0675           | 0.0650   | 0.0693   |
| 0.0686           | 0.0628   | 0.0709   |
| 0.0732           | 0.0687   | 0.0756   |
| 0.0721           | 0.0669   | 0.0747   |
| 0.0698           | 0.0667   | 0.0721   |
| 0.0672           | 0.0650   | 0.0689   |
| 0.0721           | 0.0669   | 0.0747   |
| 0.0698           | 0.0674   | 0.0721   |
| 0.0698           | 0.0687   | 0.0726   |
| 0.0706           | 0.0786   | 0.0801   |
| 0.0759           | 0.0733   | 0.0775   |
| 0.0764           | 0.0761   | 0.0771   |
| 0.0727           | 0.0727   | 0.0744   |
| 0.0741           | 0.0747   | 0.0750   |
| 0.0727           | 0.0722   | 0.0746   |
| 0.0827           | 0.0831   | 0.0856   |
| 0.0505           | 0.0498   | 0.0531   |
Supplementary Table S12. Performance evaluation of data completion by tensor decomposition algorithms for third-order transcriptome data (overexpressed genes, genes, and cell lines) with different rates of artificial missing values. Missing values are generated by the “random missing” strategy. Relative standard errors (RSEs) between the original and reconstructed data from tensor decomposition are calculated for (a) all values and (b) missing values only. The tensor decomposition method used in this study and the baseline method are denoted as TT (Yuan et al., 2017) and CP (Acar et al., 2010), respectively. The optimized tensor ranks are shown for each method. Artificially generated missing rates: 10%, 50%, and 90% are tested. Cell lines are listed in order of increasing original mis

| Artificial missing rate | 10% | 50% | 90% |
|-------------------------|-----|-----|-----|
| CP (baseline)           |     |     |     |
| TT (this study)         |     |     |     |
| CP (baseline)           |     |     |     |
| TT (this study)         |     |     |     |
| CP (baseline)           |     |     |     |
| TT (this study)         |     |     |     |
| CP (baseline)           |     |     |     |
| TT (this study)         |     |     |     |

(a) RSEs for all values

| Total cell lines | 10% | 50% | 90% |
|------------------|-----|-----|-----|
| CP (baseline)    | 0.0820 | 0.0768 | 0.0796 |
| PC3 (this study) | 0.0843 | 0.0807 | 0.0843 |
| A375 (this study) | 0.0877 | 0.0811 | 0.0880 |
| HEPG2 (this study) | 0.0887 | 0.0835 | 0.0888 |
| HT29 (this study) | 0.0834 | 0.0783 | 0.0835 |
| HA1E (this study) | 0.0796 | 0.0718 | 0.0796 |
| MCF7 (this study) | 0.0795 | 0.0748 | 0.0795 |
| HCC515 (this study) | 0.0854 | 0.0805 | 0.0855 |
| A549 (this study) | 0.0858 | 0.0790 | 0.0859 |
| VCAP (this study) | 0.0751 | 0.0706 | 0.0751 |
| HEK293T (this study) | 0.0683 | 0.0649 | 0.0689 |

(b) RSEs for missing values

| Total cell lines | 10% | 50% | 90% |
|------------------|-----|-----|-----|
| CP (baseline)    | 0.1013 | 0.0951 | 0.0987 |
| PC3 (this study) | 0.0980 | 0.0940 | 0.0957 |
| A375 (this study) | 0.1016 | 0.0943 | 0.1025 |
| HEPG2 (this study) | 0.1056 | 0.0995 | 0.1063 |
| HT29 (this study) | 0.1001 | 0.0939 | 0.1001 |
| HA1E (this study) | 0.0948 | 0.0859 | 0.0954 |
| MCF7 (this study) | 0.0945 | 0.0892 | 0.0954 |
| HCC515 (this study) | 0.1015 | 0.0956 | 0.1023 |
| A549 (this study) | 0.1118 | 0.1032 | 0.1123 |
| VCAP (this study) | 0.0981 | 0.0921 | 0.0983 |
| HEK293T (this study) | 0.1085 | 0.1041 | 0.1101 |

50