Supplemental Methods

Overall design of CTG

The Hi-C contact map depicts a proximity network $G(V,E)$, where the vertices $V=\{v_1,v_2,...v_n\}$ denote the non-overlapping genomic regions and the edges $E = \{e_{i,j}\}$ denote the contact strength between pairwise connected genomic regions. Similar to diffusion-based methods for network denoising (1, 2), a Markov prosses (3) is used to describe the diffusion process on this network. $D_{i,i} = \sum_{j=1}^{n} e_{i,j}$, is the element of the diagonal degree matrix D for the network. The vector $P_i^{(1)} = \{P_{i,1}^{(1)}, P_{i,2}^{(1)}, \dots P_{i,n}^{(1)}\}$ is the conditional transition probability transiting from vertex v_i to $V=\{v_1,v_2,...v_n\}$ in one single step. Likewise, $P_i^{(k)} = \{P_{i,1}^{(k)}, P_{i,2}^{(k)}, \dots P_{i,n}^{(k)}\}$ is the conditional transition probability in k steps and $P_{i,j}^{(k)} = \sum_{p=1}^{n} P_{i,p}^{(k-1)} P_{p,j}^{(k-1)}$. With increasing k, the transition probability from v_i to v_j gradually integrates neighbor information and expand the inclusion of edges, since v_i and v_j may not be connected in one step but they can be connected in some finite steps as the network G is a connected graph. Taking k=2 and $P_{i,j}^{(2)} = \sum_{p=1}^{n} P_{i,p}^{(1)} P_{p,j}^{(1)}$ as an example, when the two pairs of vertices (v_i and v_p, v_j and v_p) are pairwise neighbors, which means $P_{i,p}^{(1)} \neq 0$ and $P_{p,j}^{(1)} \neq 0$, v_p contributes to $P_{i,j}^{(2)}$. $P_i^{(k)}$ converges to an invariant distribution for connected graph and the difference between $P_i^{(k-1)}$ and $P_i^{(k)}$ decreases.

It is thus appropriate to use the integrated information on $\{P_i^{(1)}, P_i^{(2)}, \dots P_i^{(k)}\}$ to describe the diffusion manner of vertex v_i within some given number of k steps, which can be infinite. In practice, we found that $P_i^{(k)}$ converges rapidly and therefore used the exponential decay to fit the convergence. $S_i^{(k)}$ is defined as the weighted summation of $P_i^{(t)}$ （$1 \leq t \leq k$）:

$$S_i^{(k)} = \sum_{t=1}^{k} \exp(-\alpha t)\, P_i^{(t)}$$

When k reaches infinity, $S_i^{(k)}$ converges to S_i (Supplementary note). As the weighted summation of $P_i^{(t)}$, S_i naturally integrates neighbor information of the connected graph and therefore alleviates in a physics-based manner the problems caused by the Hi-C

data sparsity. On the other hand, the exponential decay ensures that the integration does not eliminate the distinction of each vertex, taking the rapid convergence of $P_i^{(k)}$ into consideration.

The physical succession of the genomic structure suggests that the proximal genomic regions should share similar diffusion manners. The similarity between pairwise vertices v_i and v_j is quantified by L1 distance between S_i and S_j. L1 distance is used as a measure since it mitigates the impact of outliers caused by distance matrices of higher-order terms. A C_TG distance matrix is then constructed based on the Hi-C contact map.

Proof 1

Eigenvalues Λ of the P are within the range of [-1,1].

For any eigenvector X of P:

$$PX = \lambda X$$

The maximum element of X is denoted as x_{max}, and the minimum element of X is denoted as x_{min}. As the row summation of P is normalized to 1, and P is positive,

$$x_{min} \leq \lambda x_{min} \leq x_{max}$$
$$x_{min} \leq \lambda x_{max} \leq x_{max}$$

Therefore,

$$-1 \leq \lambda x_{max} \leq 1$$

Proof 2

When n approaches infinity, the transition propensity matrix $M^{(n)}$ is convergent.

P is diagonalizable:

$$P(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m) = (\lambda_1\vec{v}_1, \lambda_2\vec{v}_2, \dots, \lambda_m\vec{v}_m) = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m)\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix}$$

$$P^{(1)} = U^{-1}\Lambda U$$

$P^{(k)}$ can be written as:

$$P^{(k)} = P^k = U^{-1}\Lambda U$$

$S^{(n)}$ is the weighted summation of $P^{(k)}$:

$S^{(n)} = \sum_{k=1}^{n} \exp(-\alpha k)\, U^{-1}\Lambda^k\, U = \sum_{k=1}^{n} U^{-1}[\exp(-\alpha k)\, \Lambda^k]\, U$

According to the associative law of multiplication:

$$S^{(n)} = U^{-1}\sum_{k=1}^{n}[\exp(-\alpha k)\, \Lambda^k]\, U = U^{-1}[\sum_{k=1}^{n}\exp(-\alpha k)\, \Lambda^k]\, U$$

When *n* approaches infinity, we have

$$S = U^{-1}[\lim_{n\to\infty}\sum_{k=1}^{n}\exp(-\alpha k)\, \Lambda^k]U$$

In the above equation, $\exp(-\alpha k)\,\Lambda^k$ is a geometric progression, and

$$\lim_{n\to\infty} \exp(-\alpha k)\,\Lambda^k \;\to\; 0$$

Therefore, the summation over $\exp(-\alpha k)\Lambda^k$ is convergent when

$$\rho(P) < \exp(a), \rho(P) = \max|\lambda_i|$$

As $\rho(P) < 1$ and $\exp(a) > \exp(0) > 1$:

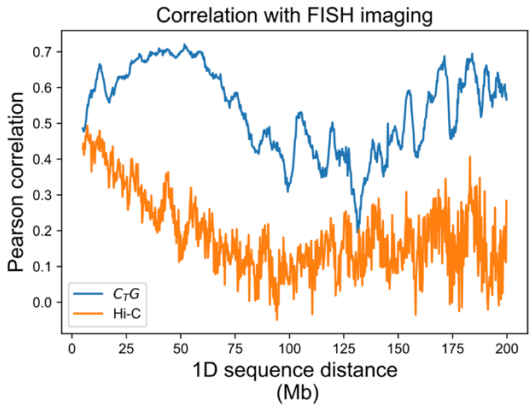
$$\lim_{n\to\infty} \sum_{k=1}^{n} \exp(-\alpha k)\,\Lambda^k = \Lambda\,[\exp(\alpha)I - \Lambda]^{-1}$$

I denotes the identity matrix.

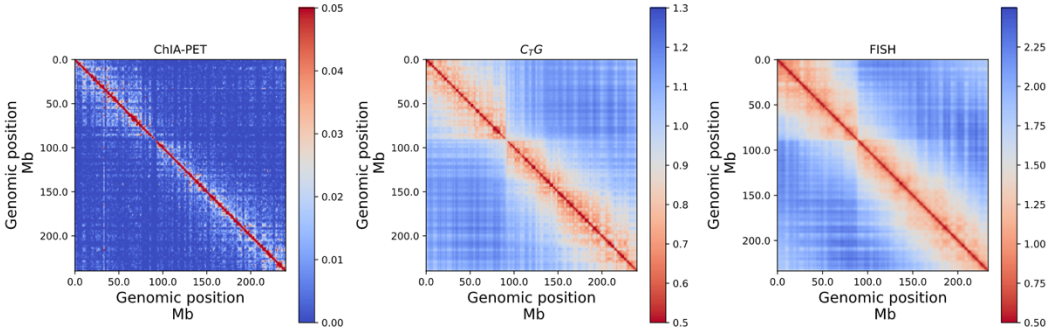
S is then also convergent and

$$S = U^{-1}\,\Lambda\,[\exp(\alpha)I - \Lambda]^{-1}U$$

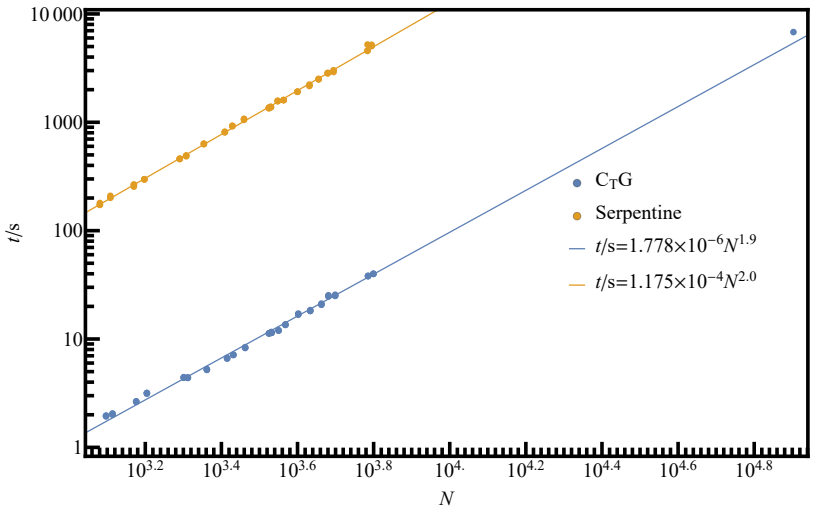
1. Wang,B., Pourshafeie,A., Zitnik,M., Zhu,J., Bustamante,C.D., Batzoglou,S. and Leskovec,J. (2018) Network enhancement as a general method to denoise weighted biological networks. *Nat. Commun.*, **9**.
2. Cao,M., Zhang,H., Park,J., Daniels,N.M., Crovella,M.E., Cowen,L.J. and Hescott,B. (2013) Going the Distance for Protein Function Prediction: A New Distance Metric for Protein Interaction Networks. *PLoS One*, **8**, e76339.
3. Stochastic Processes in Physics and Chemistry (2007) Elsevier.
4. Halverson,J.D., Smrek,J., Kremer,K. and Grosberg,A.Y. (2014) From a melt of rings to chromosome territories: The role of topological constraints in genome folding. *Reports Prog. Phys.*, **77**.
5. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O., *et al.* (2009) Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, **326**, 289–293.



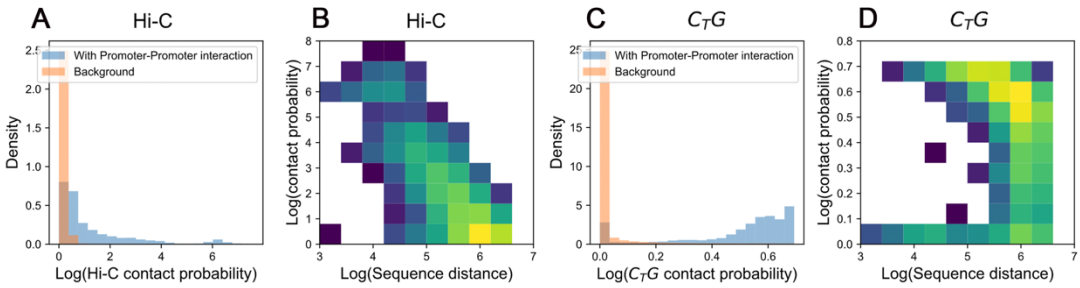
Supplemental Fig S1. The Pearson correlation coefficients between the imaging spatial distance matrix and C_TG distance (blue) and Hi-C (orange) at different 1D genomic distance



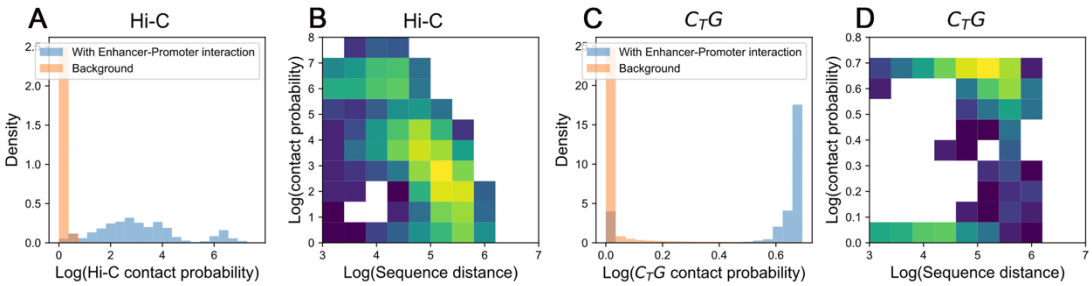
Supplemental Fig S2. The ChIA-PET contact map(left), C_TG distance matrix (middle) and the median spatial distance matrix (right) of chr2 (resolution of 50kb)



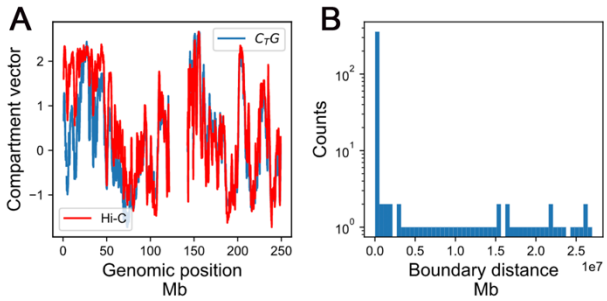
Supplemental Fig S3. The computational cost of C_TG and Serpentine. The C_TG algorithm is performed on NVIDIA GeForce RTX 3060 Laptop, Intel(R) Core(TM) i7-11800H @ 2.30 GHz. Serpentine is performed on PC workstation with 128G RAM and with Intel(R) Xeon(R) Silver 4210 CPU @2.20GHz. Serpentine also requires large memory usage (for Chr3, 100 G for Serpentine and 400 Mb for C_TG).



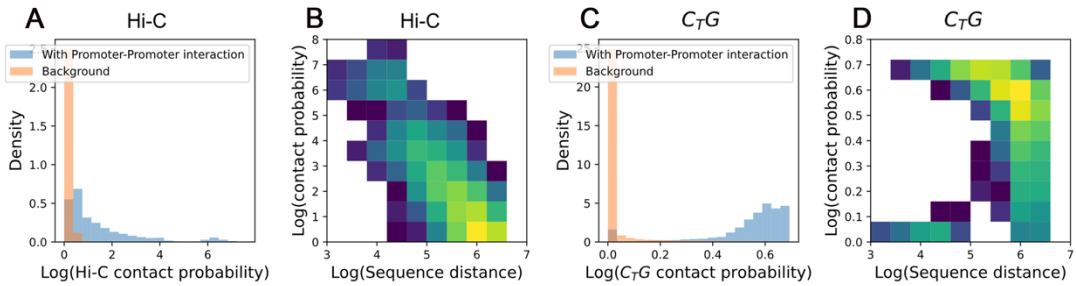
Supplemental Fig S4 (A) The distribution of Hi-C contact probability for genomic loci with related promoter-promoter interactions and the control background (using hg19 assembly). (B) The distribution of Hi-C contact probability at different 1D sequence distance (using hg19 assembly). (C) The distribution of C_TG contact probability for genomic loci with related promoter-promoter interactions and the control background (using hg19 assembly). **(D)** The distribution of C_TG contact probability at different 1D sequence distance (using hg19 assembly).



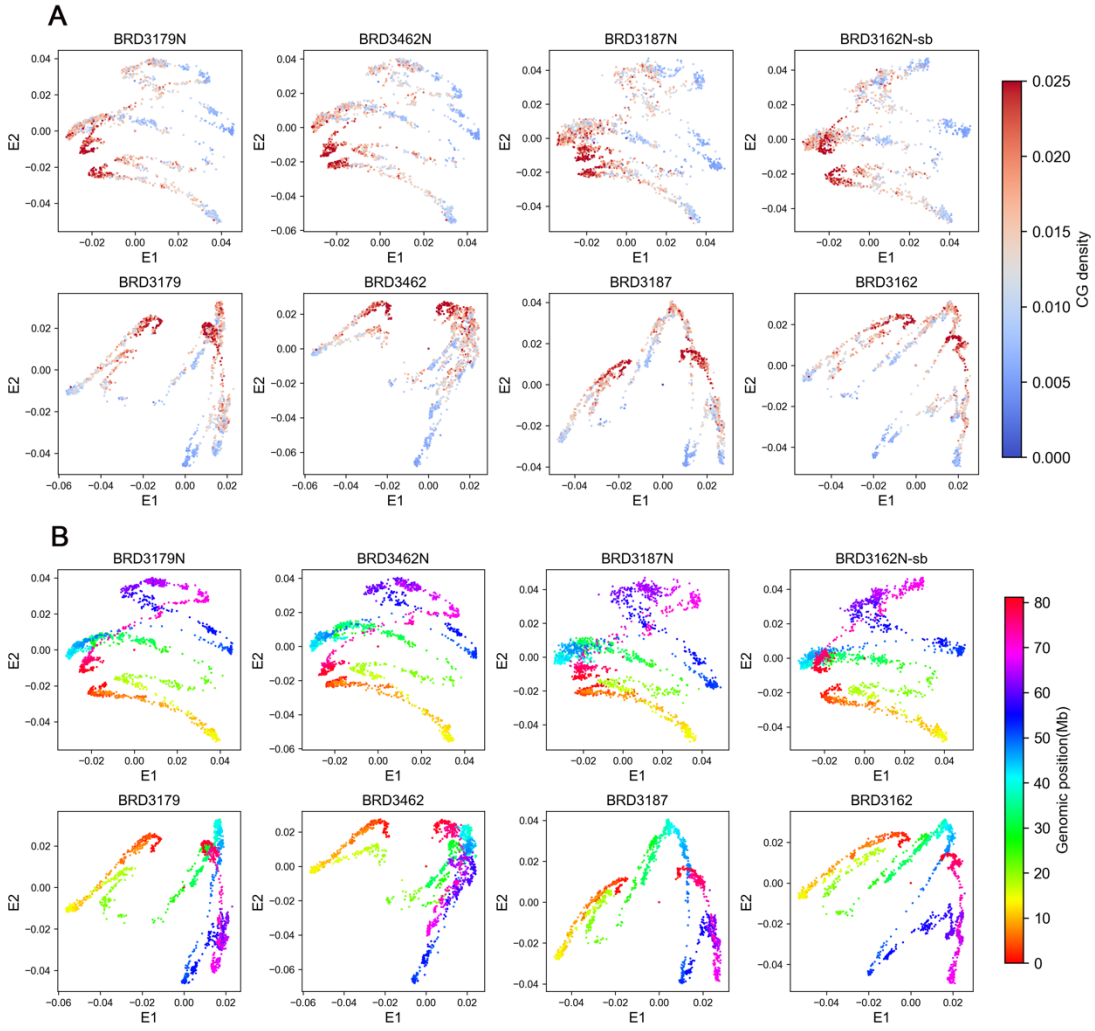
Supplemental Fig S5. (A) The distribution of Hi-C contact probability for genomic loci with related enhancer-promoter interactions and the control background (all genomic loci). (B) The distribution of Hi-C contact probability at different 1D sequence distance. (C) The distribution of C_TG contact probability for genomic loci with related enhancer-promoter interactions and the control background (all genomic loci). (D) The distribution of C_TG contact probability at different 1D sequence distance.



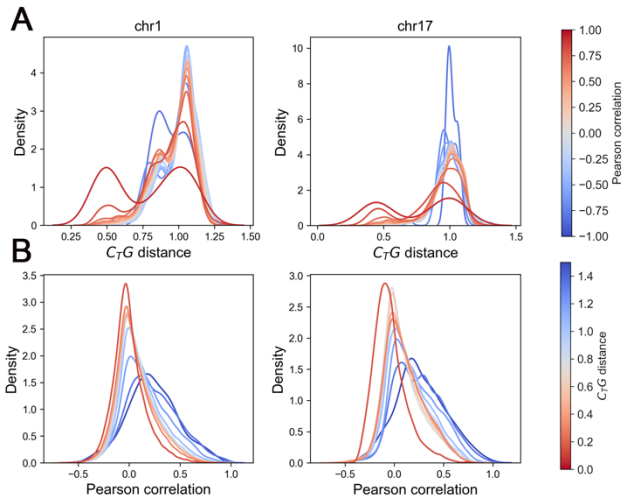
Supplemental Fig S6. (A)The compartment vector derived from C_TG distance matrix (blue) and Hi-C contact matrix (red) by performing PCA. The correlation coefficient reaches 0.890. (B) The genomic distance between TAD boundaries found in CTG distance matrix and those found in Hi-C contact matrix. We used KMeans to find TADs in C_TG distance matrix, compared with the TADs found by HiCExplore, the boundaries generated from two methods are very close.



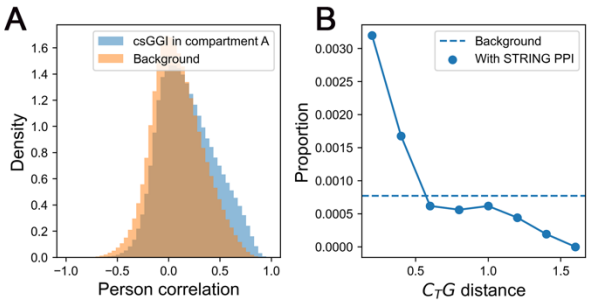
Supplemental Fig S7 (A) The distribution of Hi-C contact probability for genomic loci with related promoter-promoter interactions and the control background (using GRCh38 assembly). (B) The distribution of Hi-C contact probability at different 1D sequence distance (using GRCh38 assembly). (C) The distribution of C_TG contact probability for genomic loci with related promoter-promoter interactions and the control background (using GRCh38 assembly). **(D)** The distribution of C_TG contact probability at different 1D sequence distance (using GRCh38 assembly).



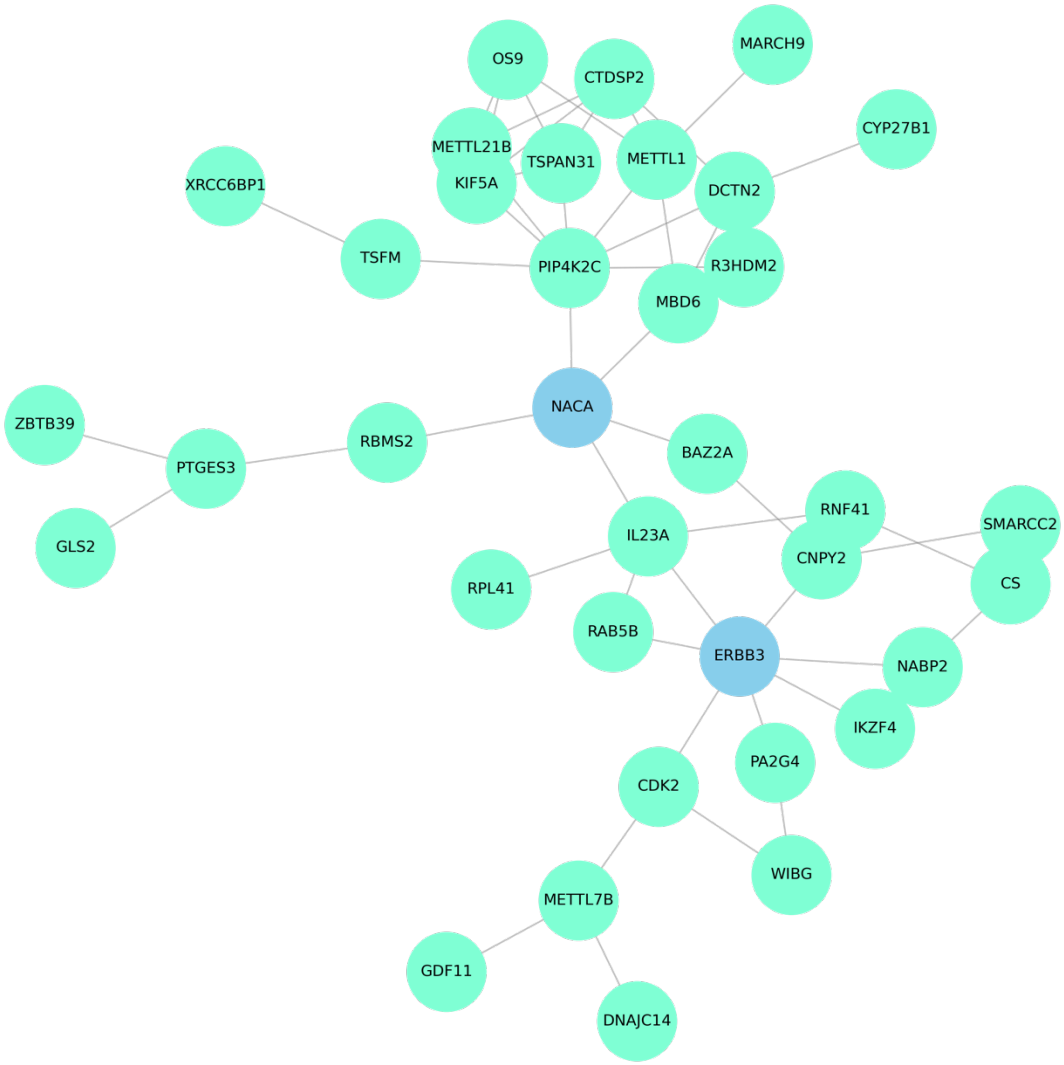
Supplemental Fig S8. 2D Laplacian eigenmap of colon cancer. (A) The 2D Laplacian Eigenmaps of C_TG distance matrices for pairwise normal (upper panel) and tumor (lower panel) colon samples. Each point represents a 40kb genomic region. The color is used to represent the CpG density of the corresponding genomic region. **(B)** The color is used to represent the genomic position of the corresponding genomic region.



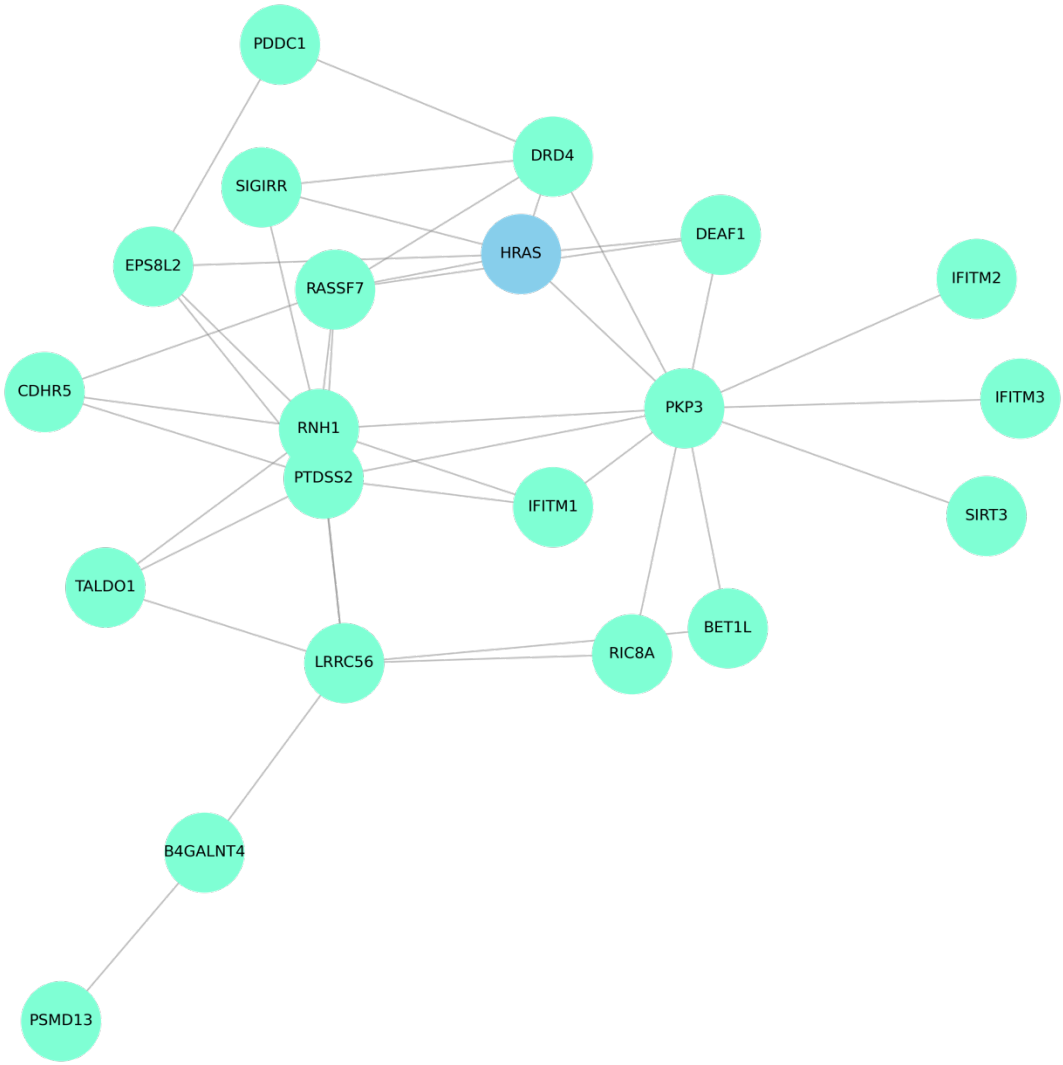
Supplemental Fig S9. Correspondence of gene-gene proximity and RNA co-regulation in leukemia. (**A**) The distribution of transcriptional Pearson correlation under different C_TG distance of chromosome 1(left) and chromosome 17(right), the color of each line indictes the corresponding C_TG distance. (**B**) The distribution of C_TG distance under different Pearson correlation of chromosome 1(left) and chromosome 17(right), the color of each line indicates corresponding Pearson correlation coefficient.



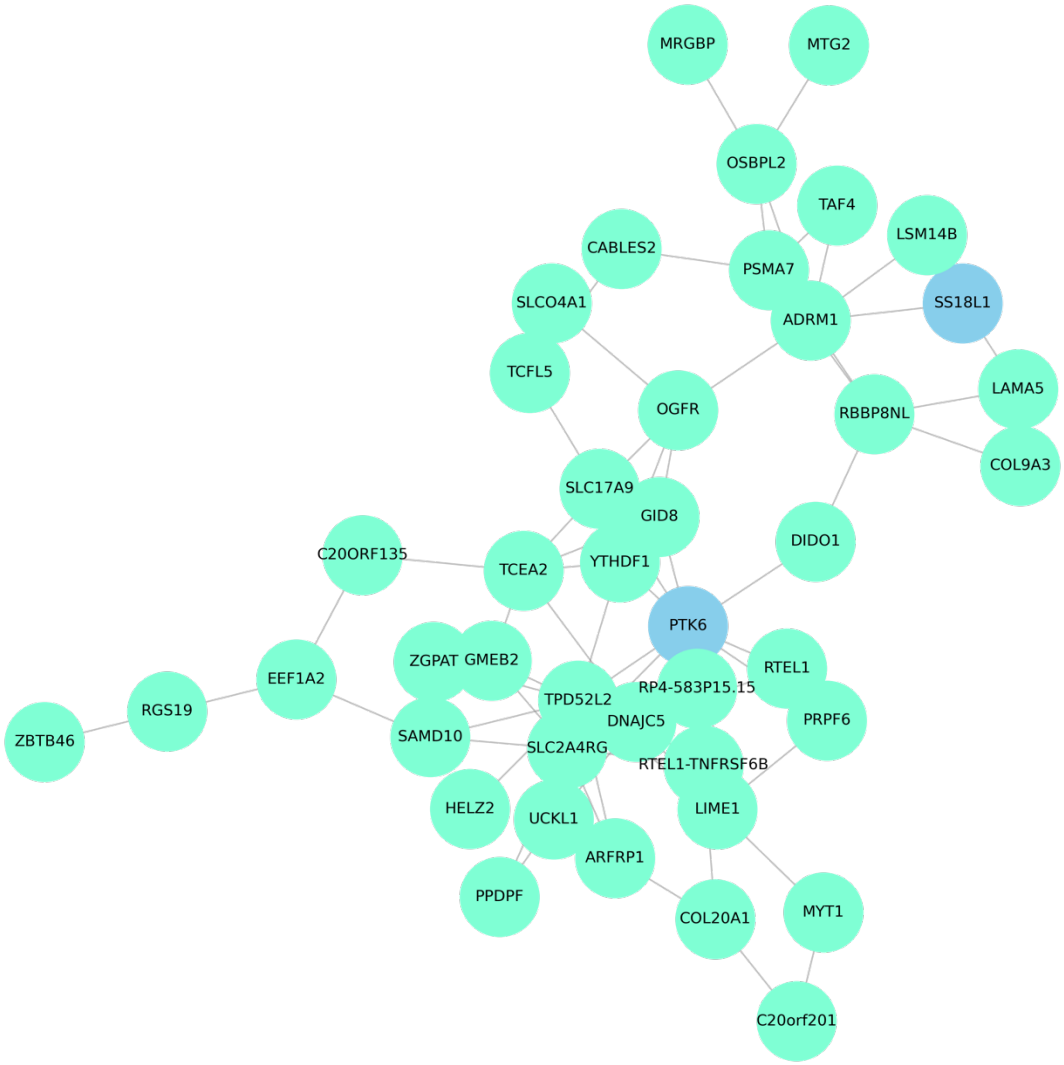
Supplemental Fig S10. (A) The distribution of correlation of gene pairs with csGGIs and overall background in compartment A. **(B)** The proportion of intra-chromosomal gene pairs with STRING PPI at different C_TG distances in compartment A. The proportion refers to number of gene pairs with PPIs at fixed C_TG distance /number of all gene pairs at fixed C_TG distance. The background refers to number of gene pairs with PPIs/number of all gene pairs at all C_TG distance.



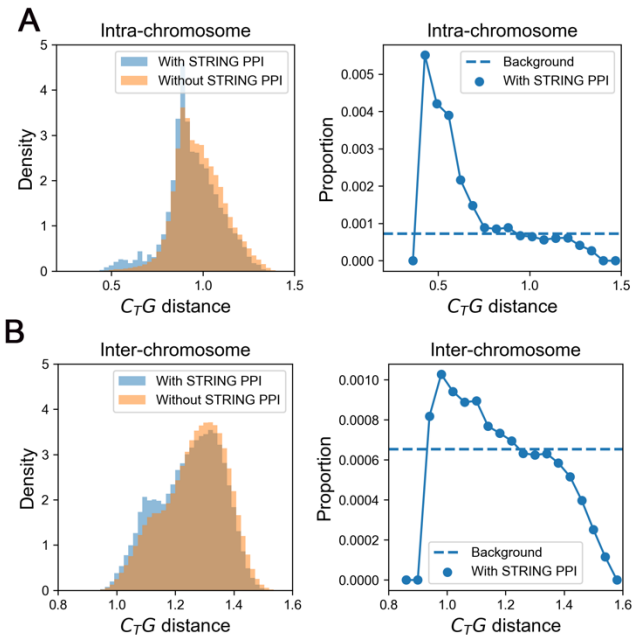
Supplemental Fig S11. Subgraph of csGGI network with *ERBB3*.



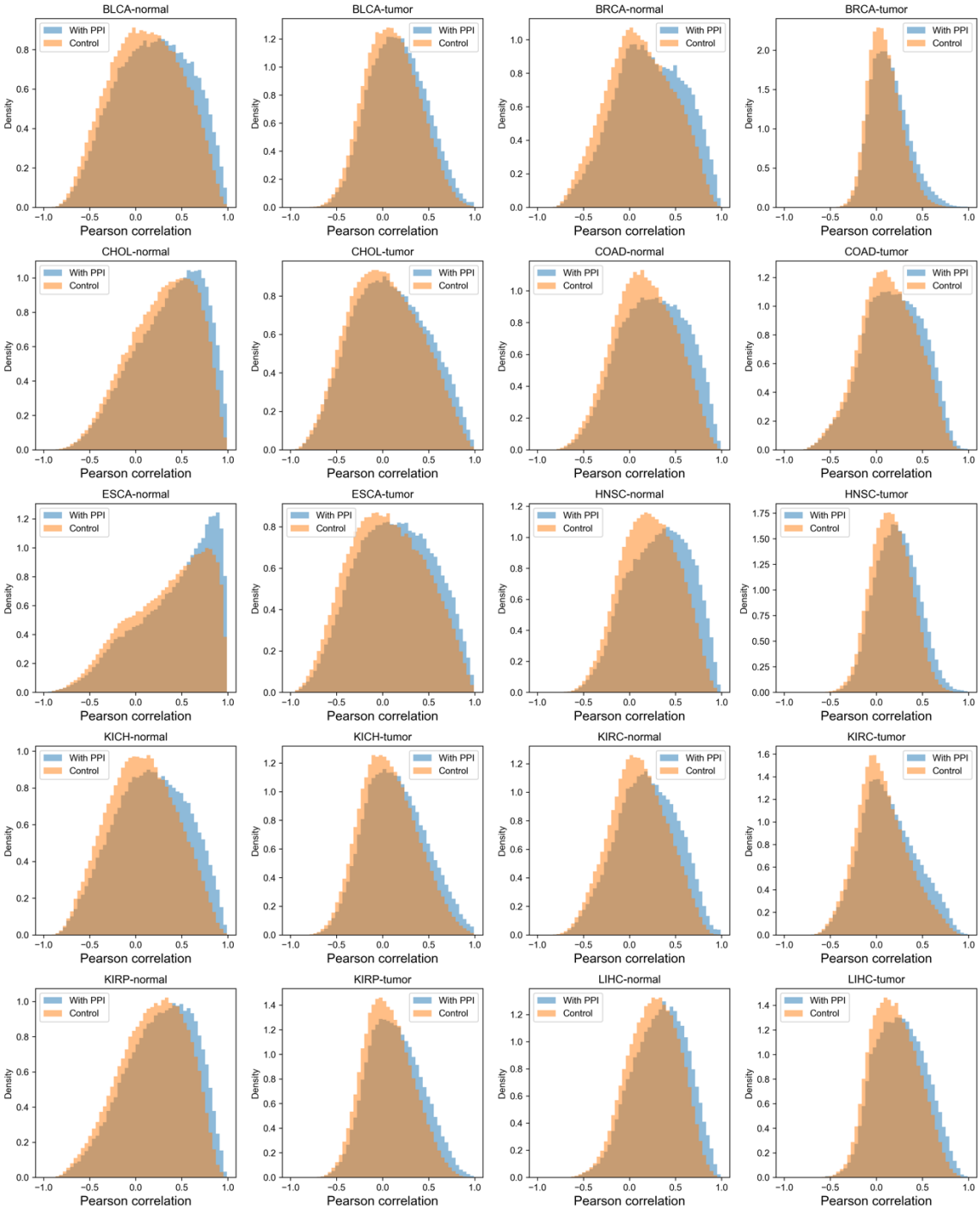
Supplemental Fig S12. Subgraph of csGGI network with *HRAS*.

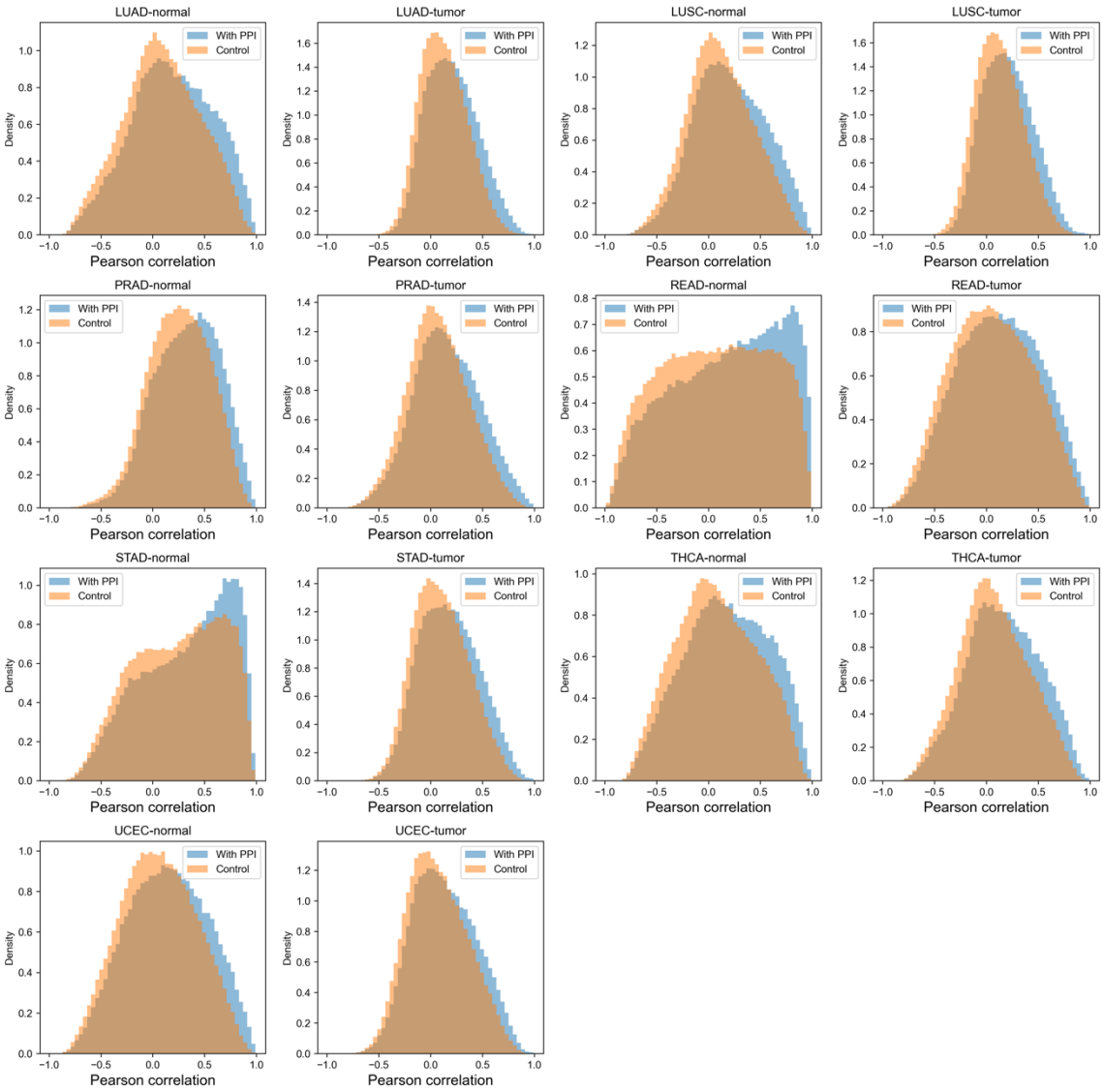


Supplemental Fig S13. Subgraph of csGGI network with *PTK6*.

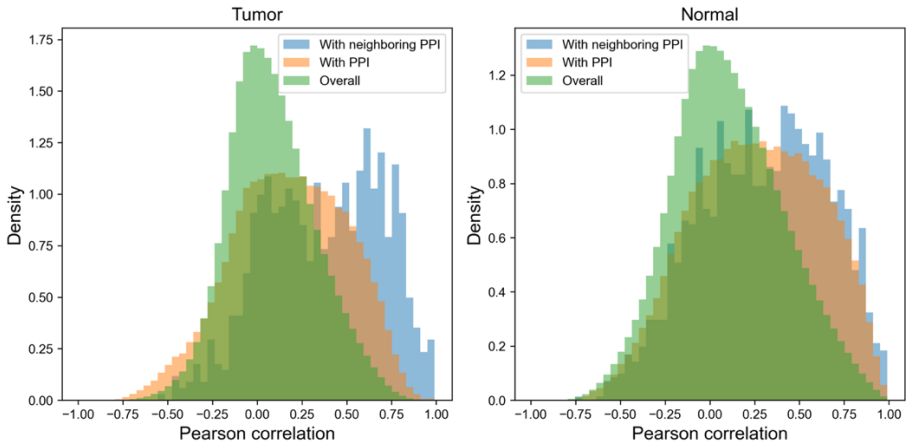


Supplemental Fig S14. Correspondence of gene-gene proximity and STRING PPIs in normal colon sample. (**A**) The distribution of C_TG distance between intra-chromosomal gene pairs with and without STRING PPIs for whole chromosome in tumor sample (left panel); The proportion of intra-chromosomal gene-pairs with STRING PPI at different C_TG distances in normal sample (right panel). (**B**) The distribution of C_TG distance between inter-chromosomal gene pairs with and without STRING PPIs for whole chromosome in normal sample (left panel); The proportion of inter-chromosomal gene-pairs with STRING PPI at different C_TG distances in normal sample (right panel).

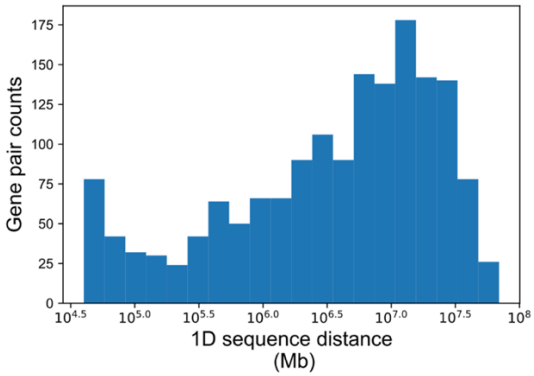

BLCA-normal
BLCA-tumor
BRCA-normal
BRCA-tumor
CHOL-normal
CHOL-tumor
COAD-normal
COAD-tumor
ESCA-normal
ESCA-tumor
HNSC-normal
HNSC-tumor
KICH-normal
KICH-tumor
KIRC-normal
KIRC-tumor
KIRP-normal
KIRP-tumor
LIHC-normal
LIHC-tumor
With PPI
Control
Density
Pearson correlation



Supplemental Fig S15. The distribution of RNA correlation with STRING PPI for 17 cancer types. The distributions of gene pairs with STRING PPI are colored blue and the control groups are colored orange. All samples show similar patterns that gene pairs with STRING PPI are more correlated at transcriptional level.



Supplemental Fig S16. The distribution of RNA correlation for colon cancer. Gene pairs with both STRING PPI and GGIs are colored blue and gene pairs with only STRING PPI are colored orange and the control group is colored green. Gene-pairs with both STRING PPI and GGIs are more correlated in transcriptional level for both tumor and normal colon samples.



Supplemental Fig S17. The distribution of 1D sequence distance of gene pairs integrating gene-gene interplay at DNA, RNA, and protein levels.

Table S1. Functional annotation clustering of colon csGGIs.

Annotation Cluster 1: Enrichment Score: 1.5327833580397092			
Term	Count	PValue	Benjamini
hsa01521: EGFR tyrosine kinase inhibitor resistance	7	4.63E-05	0.00920864
h_pyk2Pathway: Links between Pyk2 and Map Kinases	5	8.51E-05	0.00802485
h_at1rPathway: Angiotensin II mediated activation of JNK Pathway via Pyk2 dependent signaling	5	1.65E-04	0.00802485
hsa05219: Bladder cancer	5	3.55E-04	0.03383294
h_malPathway: Role of MAL in Rho-Mediated Activation of SRF	4	5.67E-04	0.01608635
hsa04662: B cell receptor signaling pathway	6	5.77E-04	0.03383294
hsa04012: ErbB signaling pathway	6	6.80E-04	0.03383294
h_rasPathway: Ras Signaling Pathway	4	0.00101341	0.0196602
hsa04140: Autophagy - animal	7	0.00108928	0.03764245
hsa05231: Choline metabolism in cancer	6	0.00129541	0.03764245
hsa04370: VEGF signaling pathway	5	0.00142489	0.03764245
hsa05205: Proteoglycans in cancer	8	0.00151326	0.03764245
h_sam68Pathway: Regulation of Splicing through Sam68	3	0.00163092	0.02636658
GO: 2000641~regulation of early endosome to late endosome transport	3	0.0017761	0.66330982
hsa04810: Regulation of actin cytoskeleton	8	0.00215053	0.04444882
h_erkPathway: Erk1/Erk2 Mapk Signaling pathway	4	0.00223416	0.03095906
hsa04664: Fc epsilon RI signaling pathway	5	0.00241051	0.04444882
hsa05211: Renal cell carcinoma	5	0.00254296	0.04444882
hsa04917: Prolactin signaling pathway	5	0.00268033	0.04444882
hsa05223: Non-small cell lung cancer	5	0.00297021	0.04546711
hsa05220: Chronic myeloid leukemia	5	0.00361302	0.05135645
h_metPathway: Signaling of Hepatocyte Growth Factor Receptor	4	0.00379947	0.03988317
hsa04650: Natural killer cell mediated cytotoxicity	6	0.00389495	0.05167298
h_fmlpPathway: fMLP induced chemokine gene expression in HMC-1 cells	4	0.00411167	0.03988317
h_integrinPathway: Integrin Signaling Pathway	4	0.00411167	0.03988317
hsa05210: Colorectal cancer	5	0.00561992	0.06989776
h_cdk5Pathway: Phosphorylation of MEK1 by cdk5/p35 down regulates the MAP kinase pathway	3	0.00586627	0.05172985
hsa04540: Gap junction	5	0.00609515	0.07134909
Annotation Cluster 2: Enrichment Score: 1.5241353790689698			
Term	Count	PValue	Benjamini
GO: 0039702~viral budding via host ESCRT complex	3	0.01072312	1

GO: 0036258~multivesicular body assembly	3	0.02069957	1
GO: 0043162~ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway	3	0.02069957	1
GO: 0090148~membrane fission	3	0.03330408	1
Annotation Cluster 3: Enrichment Score: 1.453480544316328			
Term	Count	PValue	Benjamini
KW-0653~Protein transport	12	0.01295629	0.24467061
GO: 0015031~protein transport	8	0.0389436	1
KW-0967~Endosome	9	0.08642541	0.79943504
Annotation Cluster 4: Enrichment Score: 1.444194122407839			
Term	Count	PValue	Benjamini
KW-0648~Protein biosynthesis	6	0.00775783	0.24467061
KW-0396~Initiation factor	3	0.07027195	0.71901399
GO: 0003743~translation initiation factor activity	3	0.08528948	1
Annotation Cluster 5: Enrichment Score: 1.3130484622630447			
Term	Count	PValue	Benjamini
GO: 0051402~neuron apoptotic process	5	0.001971	0.66330982
GO: 0043524~negative regulation of neuron apoptotic process	4	0.10479492	1
Annotation Cluster 6: Enrichment Score: 1.2661570708970376			
Term	Count	PValue	Benjamini
GO: 0097542~ciliary tip	4	0.00441052	0.19343661
GO: 0005813~centrosome	11	0.00465311	0.19343661
GO: 0042073~intraciliary transport	3	0.02328684	1
GO: 0005929~cilium	6	0.02622916	0.44898151
Annotation Cluster 7: Enrichment Score: 1.2170182899365116			
Term	Count	PValue	Benjamini
GO: 0019216~regulation of lipid metabolic process	4	0.00293104	0.66330982
GO: 0042752~regulation of circadian rhythm	3	0.08129563	1

Table S2. Functional annotation clustering of HCT116 structural-related intra-chromosomal PPIs.

Annotation Cluster 1: Enrichment Score: 18.886710115925517		
Term	PValue	Benjamini
GO: 0007156~homophilic cell adhesion via plasma membrane adhesion molecules	7.08E-39	1.38E-35
GO: 0007399~nervous system development	5.72E-17	5.59E-14
GO: 0007155~cell adhesion	1.66E-15	1.08E-12
GO: 0005509~calcium ion binding	1.07E-14	6.82E-12
GO: 0005887~integral component of plasma membrane	5.09E-12	2.37E-09
Annotation Cluster 2: Enrichment Score: 5.486720966179876		
Term	PValue	Benjamini
hsa05320: Autoimmune thyroid disease	8.92E-19	2.45E-16
GO: 0002323~natural killer cell activation involved in immune response	1.09E-10	5.32E-08
hsa05169: Epstein-Barr virus infection	4.37E-10	6.01E-08
GO: 0005132~type I interferon receptor binding	2.03E-09	6.44E-07
hsa05152: Tuberculosis	8.01E-09	3.96E-07
GO: 0033141~positive regulation of peptidyl-serine phosphorylation of STAT protein	1.02E-08	3.98E-06
hsa05164: Influenza A	1.78E-08	5.91E-07
hsa05163: Human cytomegalovirus infection	1.93E-08	5.91E-07
hsa05168: Herpes simplex virus 1 infection	5.62E-08	1.54E-06
GO: 0002286~T cell activation involved in immune response	7.08E-08	2.18E-05
GO: 0006959~humoral immune response	1.07E-07	2.61E-05
GO: 0019221~cytokine-mediated signaling pathway	5.31E-07	1.01E-04
hsa04623: Cytosolic DNA-sensing pathway	5.84E-07	1.46E-05
GO: 0042100~B cell proliferation	6.20E-07	1.01E-04
GO: 0060337~type I interferon signaling pathway	6.20E-07	1.01E-04
hsa05170: Human immunodeficiency virus 1 infection	6.38E-07	1.46E-05
hsa05167: Kaposi sarcoma-associated herpesvirus infection	6.93E-07	1.47E-05
GO: 0043330~response to exogenous dsRNA	7.02E-07	1.05E-04
hsa05165: Human papillomavirus infection	1.66E-06	3.26E-05
hsa04650: Natural killer cell mediated cytotoxicity	4.93E-06	8.47E-05
hsa05162: Measles	1.55E-05	2.37E-04
GO: 0051607~defense response to virus	2.19E-05	0.00237571
GO: 0030183~B cell differentiation	2.44E-05	0.0025043
hsa04622: RIG-I-like receptor signaling pathway	8.08E-05	0.00105761
hsa05161: Hepatitis B	8.57E-05	0.00107156
GO: 0002250~adaptive immune response	1.05E-04	0.00972708

hsa05200: Pathways in cancer	1.24E-04	0.00148135
GO: 0098586~cellular response to virus	1.42E-04	0.01258975
hsa04217: Necroptosis	2.67E-04	0.00271968
hsa04620: Toll-like receptor signaling pathway	3.80E-04	0.00373633
hsa05417: Lipid and atherosclerosis	5.11E-04	0.00484296
hsa05160: Hepatitis C	8.40E-04	0.00745575
hsa04936: Alcoholic liver disease	0.00123367	0.01028057
GO: 0005125~cytokine activity	0.00232746	0.21113399
hsa04630: JAK-STAT signaling pathway	0.00350412	0.02604415
hsa04060: Cytokine-cytokine receptor interaction	0.00466886	0.03292144
hsa04151: PI3K-Akt signaling pathway	0.00537873	0.0369788
hsa05171: Coronavirus disease - COVID-19	0.00802705	0.05133577
hsa04621: NOD-like receptor signaling pathway	0.00900205	0.05626278
GO: 0005126~cytokine receptor binding	0.01312996	0.75795707
Annotation Cluster 3: Enrichment Score: 4.299091183648573		
Term	PValue	Benjamini
hsa05320: Autoimmune thyroid disease	8.92E-19	2.45E-16
GO: 0071556~integral component of lumenal side of endoplasmic reticulum membrane	3.08E-10	7.16E-08
hsa05330: Allograft rejection	2.03E-09	1.86E-07
hsa05332: Graft-versus-host disease	6.58E-09	3.96E-07
hsa04940: Type I diabetes mellitus	8.64E-09	3.96E-07
GO: 0042605~peptide antigen binding	1.09E-08	2.30E-06
hsa04612: Antigen processing and presentation	1.11E-08	4.37E-07
GO: 0042613~MHC class II protein complex	5.37E-08	8.32E-06
GO: 0019882~antigen processing and presentation	7.82E-08	2.18E-05
GO: 0002504~antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	1.75E-07	3.79E-05
GO: 0032395~MHC class II receptor activity	2.70E-07	4.28E-05
GO: 0002503~peptide antigen assembly with MHC class II protein complex	8.60E-07	1.20E-04
GO: 0002381~immunoglobulin production involved in immunoglobulin mediated immune response	1.30E-06	1.70E-04
hsa05416: Viral myocarditis	3.00E-06	5.50E-05
GO: 0012507~ER to Golgi transport vesicle membrane	3.18E-06	2.95E-04
GO: 0019886~antigen processing and presentation of exogenous peptide antigen via MHC class II	3.77E-06	4.60E-04
hsa04145: Phagosome	9.83E-06	1.59E-04
hsa05166: Human T-cell leukemia virus 1 infection	1.99E-05	2.87E-04
GO: 0023026~MHC class II protein complex binding	3.50E-05	0.00444635

hsa05150: Staphylococcus aureus infection	3.91E-05	5.38E-04
GO: 0050870~positive regulation of T cell activation	8.58E-05	0.00838156
hsa05310: Asthma	1.29E-04	0.00148193
GO: 0030666~endocytic vesicle membrane	1.55E-04	0.01200662
hsa05145: Toxoplasmosis	1.61E-04	0.00177432
hsa05140: Leishmaniasis	1.71E-04	0.00180795
GO: 0005765~lysosomal membrane	2.22E-04	0.0147493
GO: 0006955~immune response	3.83E-04	0.03118785
GO: 0002486~antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP-independent	5.93E-04	0.046337
GO: 0042612~MHC class I protein complex	7.87E-04	0.04053694
hsa04514: Cell adhesion molecules	8.40E-04	0.00745575
GO: 0000139~Golgi membrane	0.00109335	0.04621875
hsa04640: Hematopoietic cell lineage	0.00111535	0.00958507
hsa05321: Inflammatory bowel disease	0.00148522	0.01201284
hsa04672: Intestinal immune network for IgA production	0.00165846	0.01303076
GO: 0030658~transport vesicle membrane	0.00190334	0.06808101
hsa05322: Systemic lupus erythematosus	0.00299623	0.02288785
GO: 0001916~positive regulation of T cell mediated cytotoxicity	0.00301056	0.22613968
GO: 0030670~phagocytic vesicle membrane	0.00555792	0.15684631
hsa04659: Th17 cell differentiation	0.00734264	0.04807682
hsa04658: Th1 and Th2 cell differentiation	0.01032726	0.06173908
hsa05323: Rheumatoid arthritis	0.01092655	0.06393195
GO: 0030669~clathrin-coated endocytic vesicle membrane	0.01841431	0.42813264
GO: 0010008~endosome membrane	0.05188268	0.86361666
GO: 0031901~early endosome membrane	0.05200272	0.86361666
GO: 0055038~recycling endosome membrane	0.07280561	1
GO: 0032588~trans-Golgi network membrane	0.07280561	1
GO: 0050852~T cell receptor signaling pathway	0.24866213	1
hsa05203: Viral carcinogenesis	0.28823311	1
hsa04218: Cellular senescence	0.38971633	1
hsa04144: Endocytosis	0.49025631	1
Annotation Cluster 4: Enrichment Score: 2.5719744078830575		
Term	PValue	Benjamini
GO: 0016339~calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules	4.69E-06	5.39E-04
GO: 0007416~synapse assembly	1.79E-04	0.015208
GO: 0007268~chemical synaptic transmission	0.12291758	1

GO: 0045202~synapse	0.49932314	1
Annotation Cluster 5: Enrichment Score: 1.3966018993019305		
Term	PValue	Benjamini
hsa00480: Glutathione metabolism	0.00361814	0.02618394
GO: 0004364~glutathione transferase activity	0.00366572	0.29096645
hsa05204: Chemical carcinogenesis - DNA adducts	0.00922531	0.05637689
GO: 0006749~glutathione metabolic process	0.02149396	1
hsa00982: Drug metabolism - cytochrome P450	0.04022589	0.22124239
hsa01524: Platinum drug resistance	0.04229977	0.22808699
hsa05207: Chemical carcinogenesis - receptor activation	0.05243462	0.27729846
hsa00980: Metabolism of xenobiotics by cytochrome P450	0.0536251	0.27824345
hsa00983: Drug metabolism - other enzymes	0.05860336	0.29301679
GO: 0006805~xenobiotic metabolic process	0.07181652	1
GO: 0042178~xenobiotic catabolic process	0.09474298	1
hsa05225: Hepatocellular carcinoma	0.15260351	0.68796663
hsa05418: Fluid shear stress and atherosclerosis	0.15821423	0.69977247
hsa05208: Chemical carcinogenesis - reactive oxygen species	0.23136469	0.92210563
Annotation Cluster 6: Enrichment Score: 1.3570585896445353		
Term	PValue	Benjamini
GO: 0045869~negative regulation of single stranded viral RNA replication via double stranded DNA intermediate	0.00518358	0.33745123
GO: 0010529~negative regulation of transposition	0.01639338	0.89908234
GO: 0070383~DNA cytosine deamination	0.02040899	0.99646883
GO: 0047844~deoxycytidine deaminase activity	0.02262817	1
GO: 0009972~cytidine deamination	0.02905456	1
GO: 0016554~cytidine to uridine editing	0.02905456	1
GO: 0004126~cytidine deaminase activity	0.03216004	1
GO: 0080111~DNA demethylation	0.08096562	1
hsa03250: Viral life cycle - HIV-1	0.49555146	1
GO: 0000932~P-body	0.62879617	1
Annotation Cluster 7: Enrichment Score: 1.2134687565396018		
Term	PValue	Benjamini
GO: 0045324~late endosome to vacuole transport	0.01657295	0.89908234
GO: 0097352~autophagosome maturation	0.06918992	1
GO: 0016236~macroautophagy	0.1995957	1

Table S3. Functional annotation clustering of HCT116 structural-related intra-chromosomal PPIs.

Annotation Cluster 1: Enrichment Score: 2.943527846030514		
Term	PValue	Benjamini
GO: 0000502~proteasome complex	1.47E-06	1.74E-04
hsa03050: Proteasome	6.92E-04	0.043804
GO: 0022624~proteasome accessory complex	0.0066303	0.1380662
hsa05017: Spinocerebellar ataxia	0.24932119	0.99411035
Annotation Cluster 2: Enrichment Score: 2.5665692443489947		
Term	PValue	Benjamini
GO: 0042765~GPI-anchor transamidase complex	9.14E-04	0.03595721
GO: 0016255~attachment of GPI anchor to protein	0.00185364	0.55531848
hsa00563: Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	0.01178264	0.22179084
Annotation Cluster 3: Enrichment Score: 2.4423007751503065		
Term	PValue	Benjamini
GO: 0008380~RNA splicing	2.54E-06	0.00457121
hsa03040: Spliceosome	1.25E-04	0.01331879
GO: 0000398~mRNA splicing, via spliceosome	0.00136178	0.49401419
GO: 0005681~spliceosomal complex	0.00175958	0.05662648
GO: 0071013~catalytic step 2 spliceosome	0.00260104	0.07082841
GO: 0071005~U2-type precatalytic spliceosome	0.00771117	0.15598591
GO: 0071007~U2-type catalytic step 2 spliceosome	0.01130191	0.20517311
GO: 0000375~RNA splicing, via transesterification reactions	0.11039709	1
GO: 0005682~U5 snRNP	0.11623505	0.86190502
GO: 0046540~U4/U6 x U5 tri-snRNP complex	0.17059402	0.9309836
Annotation Cluster 4: Enrichment Score: 2.3608515879841923		
Term	PValue	Benjamini
GO: 0030141~secretory granule	1.78E-04	0.0104911
GO: 0004252~serine-type endopeptidase activity	0.01734523	0.95171215
GO: 0008236~serine-type peptidase activity	0.02680996	1
Annotation Cluster 5: Enrichment Score: 2.1495169142708916		
Term	PValue	Benjamini
GO: 0071051~polyadenylation-dependent snoRNA 3'-end processing	1.51E-04	0.1354518
GO: 0034475~U4 snRNA 3'-end processing	2.90E-04	0.20444477

GO: 0000177~cytoplasmic exosome (RNase complex)	2.93E-04	0.01480803
GO: 0045006~DNA deamination	3.98E-04	0.20444477
GO: 0000178~exosome (RNase complex)	5.91E-04	0.02791445
GO: 0101019~nucleolar exosome (RNase complex)	7.58E-04	0.033306
GO: 0034427~nuclear-transcribed mRNA catabolic process, exonucleolytic, 3'-5'	8.07E-04	0.36271775
GO: 0000176~nuclear exosome (RNase complex)	0.00165804	0.05662648
GO: 0016075~rRNA catabolic process	0.00558333	0.8519673
GO: 0043928~exonucleolytic nuclear-transcribed mRNA catabolic process involved in deadenylation-dependent decay	0.02071858	1
GO: 0006401~RNA catabolic process	0.02274463	1
GO: 0035327~transcriptionally active chromatin	0.02293376	0.31837454
GO: 0071028~nuclear mRNA surveillance	0.06364649	1
hsa03018: RNA degradation	0.06393924	0.60178111
GO: 0000175~3'-5'-exoribonuclease activity	0.07306224	1
GO: 0000791~euchromatin	0.16498087	0.9309836
GO: 0090503~RNA phosphodiester bond hydrolysis, exonucleolytic	0.22541777	1
GO: 0006396~RNA processing	0.99805831	1
Annotation Cluster 6: Enrichment Score: 1.9218679321185945		
Term	PValue	Benjamini
GO: 0044183~protein binding involved in protein folding	0.00113455	0.23666785
GO: 0051082~unfolded protein binding	0.03069127	1
GO: 0006457~protein folding	0.04926703	1
Annotation Cluster 7: Enrichment Score: 1.7774390184940738		
Term	PValue	Benjamini
hsa05020: Prion disease	0.00703128	0.20136586
hsa05014: Amyotrophic lateral sclerosis	0.00818049	0.20136586
hsa05012: Parkinson disease	0.00975062	0.22157805
hsa05022: Pathways of neurodegeneration - multiple diseases	0.01038647	0.22157805
hsa05016: Huntington disease	0.02612139	0.34828525
hsa05010: Alzheimer disease	0.14225106	0.8128632