Identifying potential cancer driver genes by genomic data integration

Yong Chen[1,2], Jingjing Hao[2], Wei Jiang[2], Tong He[3], Xuegong Zhang[2], Tao Jiang[2,4], Rui Jiang[2,*]

[1] National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

[2] MOE Key Laboratory of Bioinformatics and Bioinformatics Division, TNLIST/Department of Automation, Tsinghua University, Beijing 100084, China

[3] School of Applied Mathematics, Central University of Finance and Economics, Beijing 102206, China

[4] Department of Computer Science and Engineering, University of California, Riverside, CA 92521, USA

* Corresponding author

Tel.: 86-10-62795578-828

Fax: 86-10-62773552

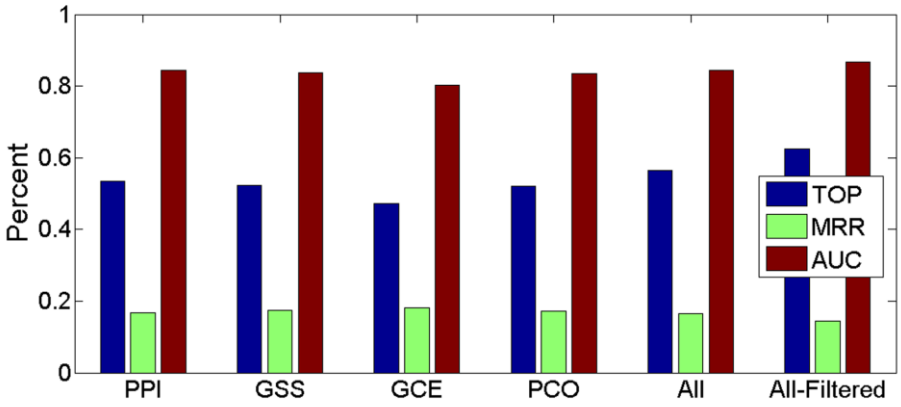
Rui Jiang:

E-mail: ruijiang@tsinghua.edu.cn

Supplementary figures and legends

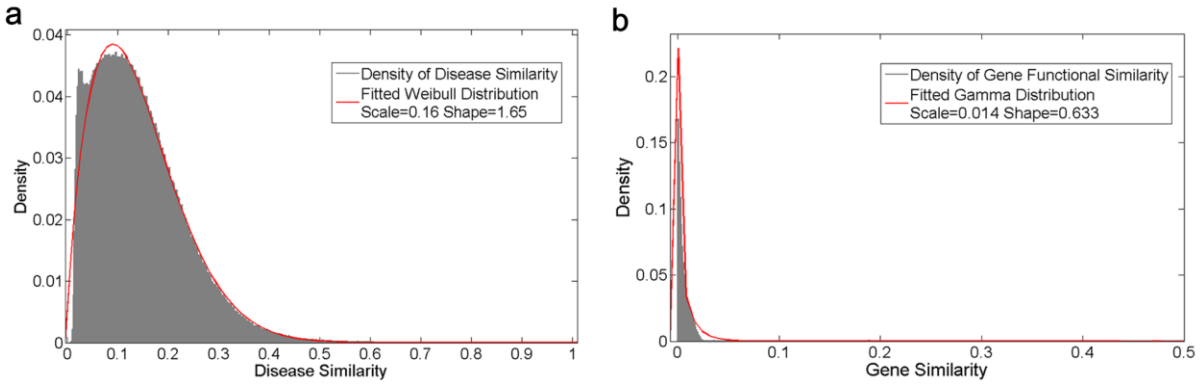
Supplementary Figure S1. The performances of each data source.

The results were obtained from the validations against random controls based on the heterogeneous network. PPI: Protein-protein interactions. GCE: Gene co-expression patterns. GSS: Gene sequence similarities. PCO: Pathway co-occurrence relationships. All: All the four types omics data integrated, but without noise filtering. All-Filtered: All the four types omics data integrated and noise filtered ($\beta = 0.25$ and $\gamma = 0.19$). TOP: Top one ranked ratio of positive cases. MRR: The averaged relative ranks of all positive cases. AUC: The area under the receiver operating characteristic curves (ROC).



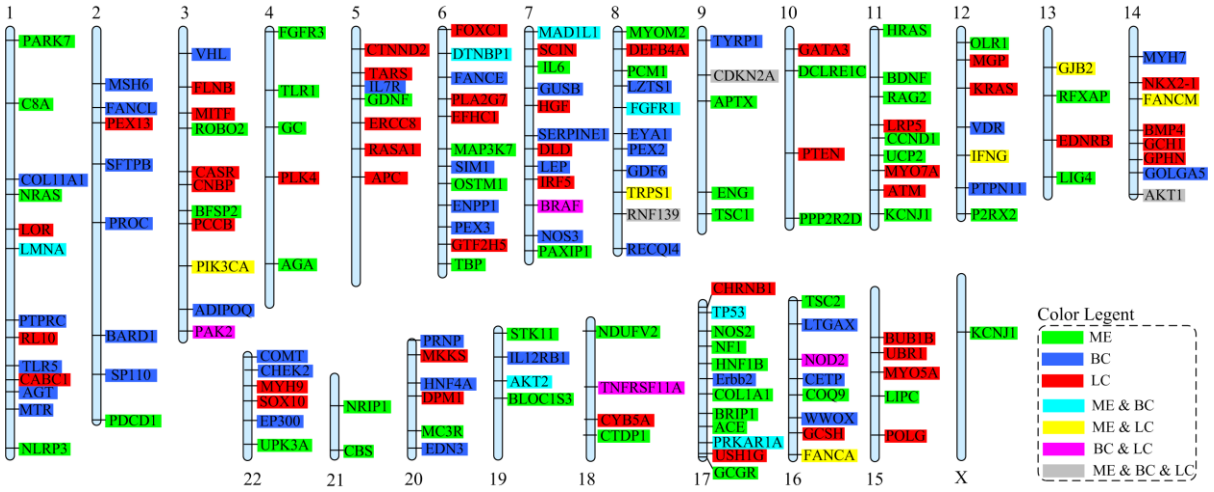
Supplementary Figure S2. Statistical analysis of and disease phenotypic similarities and fused gene functional similarities.

Both distributions were fitted by using MATLAB Fitting ToolBox. (**a**) The density and fitted Weibull distributions of disease phenotypic similarities. (**b**) The density and fitted Gamma distributions of fused gene functional similarities.

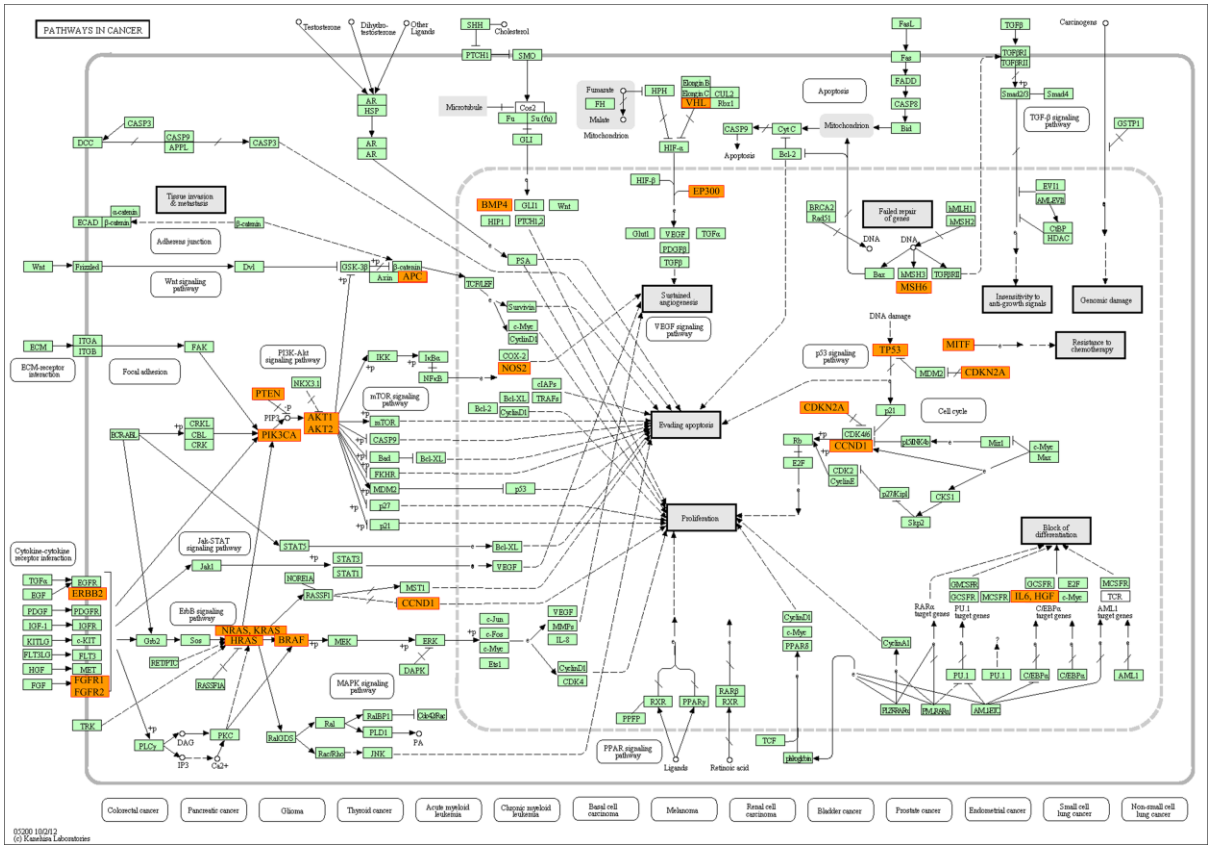


Supplementary Figure S3. The chromosomal localizations of 164 potential driver genes.

The genes of same cancers are marked in the same colours. Genes are marked on chromosomes according to their relative positions on the chromosomes. BC: Breast Cancer. ME: Melanoma. LC: Liver Carcinoma.



Supplementary Figure S4. The distribution of 23 predicted driver genes (red) in the cancer pathway (KEGG: hsp05200).



Supplementary Figure S5. Relationship analysis of 164 genes.

The data were analyzed and outputted by using STRING database.