

Table SII. Supplementary references.

Domain	Description	Reference
<b>Databases</b>	Epithelial-Mesenchymal Transition Gene Database 2.0 (dbEMT 2.0, <a href="http://dbemt.bioinfo-minzhao.org/">http://dbemt.bioinfo-minzhao.org/</a> )	(1)
	Database of Genomic Variation and Phenotype in Humans using Ensembl Resources (DECIPHER)	(2)
	Harmonized Cancer Datasets - Genomic Data Commons (GDC) Data Portal ( <a href="https://portal.gdc.cancer.gov/">https://portal.gdc.cancer.gov/</a> ) (3)	(4)
	DAVID v6.8 bioinformatics resources	(5,6)
	The Human Protein Atlas ( <a href="http://www.proteinatlas.org/">http://www.proteinatlas.org/</a> )	(7)
	GeneMANIA prediction server	(8)
	Cytoscape v.3.8.2	(9)
	R v4.0.2 ( <a href="https://www.R-project.org/">https://www.R-project.org/</a> )	(10)
	Bioconductor v3.12 packages ( <a href="http://bioconductor.org/">http://bioconductor.org/</a> )	(11,12)
<b>Softwares, packages, and scripts</b>	Perl script to load the CNV segment data files generated by ChAS ( <a href="https://github.com/BiosLS/MakeReport">https://github.com/BiosLS/MakeReport</a> ) for each sample, compare the files to build a table of genes that contains events types (gains or losses), frequencies of altered regions, including chromosomes and cytogenetic bands, and Online Mendelian Inheritance in Man (OMIM) information, and incorporate additional information from different databases: haploinsufficiency information (2,13), genes reported in Epithelial-Mesenchymal Transition Gene DataBase (14), and genes affected in gastric adenocarcinoma from Harmonized Cancer Datasets	(15)
	Chromosome Analysis Suite v3.2 software package (ChAS; Affymetrix, 2017), using NetAff 33 Libraries based on the construction of the hg19 genome (Feb 2009) as a reference model. The Hidden Markov Model (HMM) available in ChAS was used to determine the copy number state (CN-state) and their breakpoints.	(16)
	KaryoploteR to create the karyotype	(17)
	BSgenome.Hsapiens.UCSC.hg19 v1.4.0	(18)
	jvenn server for Venn diagrams	(19)
	gplots for heatmaps ( <a href="https://cran.r-project.org/web/packages/gplots">https://cran.r-project.org/web/packages/gplots</a> )	(20)
	ClusterProfiler v3.16.1 packages (21) for enrichment and gene ontology (GO) analysis, including:	(21)

	org.Hs.eg.db v3.11.4	(22)
	enrich plot v1.8.1	(23)
	GOpplot v1.0.2	(24)
<b>Protocols</b>	Modified DNA extraction protocol	(25)
	High-density whole-genome microarray analysis with Affymetrix® CytoScan™ microarrays ( <a href="http://media.affymetrix.com/support/technical/datasheets/cytoscan_hd_datasheet.pdf">http://media.affymetrix.com/support/technical/datasheets/cytoscan_hd_datasheet.pdf</a> )	(26)

---

## References

1. Zhao M, Liu Y, Zheng C and Qu H: dbEMT 2.0: An updated database for epithelial-mesenchymal transition genes with experimentally verified information and precalculated regulation information for cancer metastasis. *J Genet Genomics* 46: 595–597, 2019.
2. Firth HV, Richards SM, Bevan AP, *et al.*: DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* 84: 524–533, 2009.
3. NIH National Cancer Institute: Harmonized Cancer Datasets - Genomic Data Commons Data Portal. GDC Data Portal
4. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA and Staudt LM: Toward a Shared Vision for Cancer Genomic Data. *New England Journal of Medicine* 375: 1109–1112, 2016.
5. Huang DW, Sherman BT and Lempicki RA: Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1–13, 2009.
6. Huang DW, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44–57, 2009.
7. Uhlen M, Zhang C, Lee S, *et al.*: A pathology atlas of the human cancer transcriptome. *Science* 357, 2017.
8. Warde-Farley D, Donaldson SL, Comes O, *et al.*: The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 38: W214–W220, 2010.
9. Shannon P, Markiel A, Ozier O, *et al.*: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504, 2003.
10. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2021.
11. Gentleman R, Carey V, Huber W, Dudoit S and Irizarry R: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, 2005.
12. Huber W, Carey VJ, Gentleman R, *et al.*: Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 12: 115–121, 2015.
13. Huang N, Lee I, Marcotte EM and Hurles ME: Characterising and Predicting Haploinsufficiency in the Human Genome. *PLoS Genet* 6: e1001154, 2010.
14. Zhao M, Kong L, Liu Y and Qu H: dbEMT: an epithelial-mesenchymal transition associated gene resource. *Sci Rep* 5: 1438, 2015.
15. Larios-Serrato V: MakeReport. GitHub repository, 2020.
16. Chromosome Analysis Suite 3.2 (ChAS 3.2) User Guide. Thermo Fisher Scientific
17. Gel B and Serra E: karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* 33: 3088–3090, 2017.
18. Team TBD: BSgenome.Hsapiens.UCSC.hg19: Full genome sequences for Homo sapiens (UCSC version hg19, based on GRCh37.p13)., 2020.
19. Bardou P, Mariette J, Escudié F, Djemiel C and Klopp C: jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* 15: 293, 2014.
20. Warnes GR, Bolker B, Bonebakker L, *et al.*: gplots: Various R Programming Tools for Plotting Data., 2019.
21. Yu G, Wang L-G, Han Y and He Q-Y: clusterProfiler: an R Package for Comparing Biological Themes

- Among Gene Clusters. OMICS 16: 284–287, 2012.
22. Carlson M: org.Hs.eg.db: Genome wide annotation for Human., 2019.
  23. Yu, Guangchuang: enrichplot: Visualization of Functional Enrichment Result., 2019.
  24. Walter W, Sánchez-Cabo F and Ricote M: GOpplot: an R package for visually combining expression data with functional analysis. Bioinformatics 31: 2912–2914, 2015.
  25. Fischer I, Cunliffe C, Bollo RJ, *et al.*: Glioma-like proliferation within tissues excised as tubers in patients with tuberous sclerosis complex. Acta Neuropathol 116: 67–77, 2008.
  26. Affymetrix Support by Product for CytoScan® HD Array Kit and Reagent Kit Bundle. CytoScan® HD Array Kit and Reagent Kit Bundle - Support Materials, 2017.