**Statement of sharing, code and data dissemination, and collaboration**

**Code development**: All code developed as part of the CZI project will be disseminated through a Github repository through the PIs Github organization (FertigLab). The Github repository will be public to coordinate development with collaborators in the consortium and broader HCA network. The PI will monitor pull requests, discuss this development with collaborators through Slack, and include the GitHub app in Slack for these projects to facilitate such collaboration. At the time of publication of any new methods, algorithms developed on Github will be released in the CoGAPS Bioconductor package15 or new Bioconductor packages as best suits the project. The PI will coordinate these version control efforts and software release with collaborators and the CZI.

**Proposal:** The proposal has already been developed in collaboration with the collaborative network and shared publicly on <https://github.com/FertigLab/HCA>.

**Data sharing**: We will work closely with collaborators to ensure all raw sequence data used in this project is deposited in the HCA Data Coordination Platform. We will generate separate code repositories for applied analyses of these data (include results files and cache), consistent with the PI’s standard practice (e.g., <https://sourceforge.net/projects/psva/>, <https://github.com/FertigLab/EGFRFeedback)>.

**Protocols**: All protocols developed for bioinformatics analysis will be described fully in the Readme of the Github package and/or in the vignette of the Bioconductor software in which they are released. No experimental protocols are associated with this project.

**Manuscripts:** Methods will be published, and posted on Bioarxiv at the time of submission for publication. All publications will cite the grant number, as specified in the grant agreement.

**References**

1 Wagner A, Regev A & Yosef N. Revealing the vectors of cellular identity with single-cell genomics. *Nat Biotechnol* **34**, 1145-1160, (2016). PMC5465644.

2 Sibisi S & Skilling J. Prior distributions on measure space. *Journal of the Royal Statistical Society, B* **59**, 217-235, (1997). PMID PMCID not available.

3 Fertig EJ, Stein-O'Brien G, Jaffe A & Colantuoni C. Pattern identification in time-course gene expression data with the CoGAPS matrix factorization. *Methods Mol Biol* **1101**, 87-112, (2014). PMID 24233779, PMCID not available.

4 Hill SM, Heiser LM, Cokelaer T, Unger M, Nesser NK, Carlin DE, Zhang Y, Sokolov A, Paull EO, Wong CK, Graim K, Bivol A, Wang H, Zhu F, Afsari B, Danilova LV, Favorov AV, Lee WS, Taylor D, Hu CW, Long BL, Noren DP, Bisberg AJ, Consortium H-D, Mills GB, Gray JW, Kellen M, Norman T, Friend S, Qutub AA, Fertig EJ, Guan Y, Song M, Stuart JM, Spellman PT, Koeppl H, Stolovitzky G, Saez-Rodriguez J & Mukherjee S. Inferring causal molecular networks: empirical assessment through a community-based effort. *Nat Methods* **13**, 310-318, (2016). PMC4854847.

5 Ochs MF & Fertig EJ. Matrix Factorization for Transcriptional Regulatory Network Inference. *IEEE Symp Comput Intell Bioinforma Comput Biol Proc* **2012**, 387-396, (2012). PMC4212829.

6 Stein-O'Brien G, Kagohara LT, Li S, Thakar M, Ranaweera R, Ozawa H, Cheng H, Considine M, Favorov A, Danilova L, Califano JA, Izumchenko E, Gaykalova DA, Chung CH & Fertig EJ. *Integrated time-course omics analysis distinguishes immediate therapeutic response from acquired resistance* (Bioarxiv, 2017).

7 Fertig EJ, Ren Q, Cheng H, Hatakeyama H, Dicker AP, Rodeck U, Considine M, Ochs MF & Chung CH. Gene expression signatures modulated by epidermal growth factor receptor activation and their relationship to cetuximab resistance in head and neck squamous cell carcinoma. *BMC Genomics* **13**, 160, (2012). PMC3460736.

8 Stein-O'Brien GL, Carey JL, Lee WS, Considine M, Favorov AV, Flam E, Guo T, Li S, Marchionni L, Sherman T, Sivy S, Gaykalova DA, McKay RD, Ochs MF, Colantuoni C & Fertig EJ. PatternMarkers & GWCoGAPS for novel data-driven biomarkers via whole transcriptome NMF. *Bioinformatics*, (2017). PMID 28174896, PMCID not available.

9 Fertig EJ, Markovic A, Danilova LV, Gaykalova DA, Cope L, Chung CH, Ochs MF & Califano JA. Preferential activation of the hedgehog pathway by epigenetic modulations in HPV negative HNSCC identified with meta-pathway analysis. *PLoS One* **8**, e78127, (2013). PMC3817178.

10 Zakeri M, Srivastava A, Almodaresi F & Patro R. Improved data-driven likelihood factorizations for transcript abundance estimation. *Bioinformatics* **33**, i142-i151, (2017). PMID PMCID not available.

11 de Campos CP, Rancoita PM, Kwee I, Zucca E, Zaffalon M & Bertoni F. Discovering subgroups of patients from DNA copy number data using NMF on compacted matrices. *PLoS One* **8**, e79720, (2013). PMC3835832.

12 Tepper M & Sapiro G. Compressed Nonnegative Matrix Factorization Is Fast and Accurate. *IEEE Transactions on Signal Processing* **64**, 2269-2283, (2016). PMID PMCID not available.

13 Gaykalova DA, Zizkova V, Guo T, Tiscareno I, Wei Y, Vatapalli R, Hennessey PT, Ahn J, Danilova L, Khan Z, Bishop JA, Gutkind JS, Koch WM, Westra WH, Fertig EJ, Ochs MF & Califano JA. Integrative computational analysis of transcriptional and epigenetic alterations implicates DTX1 as a putative tumor suppressor gene in HNSCC. *Oncotarget* **8**, 15349-15363, (2017). PMID 28146432, PMCID not available.

14 Parker HS, Leek JT, Favorov AV, Considine M, Xia X, Chavan S, Chung CH & Fertig EJ. Preserving biological heterogeneity with a permuted surrogate variable analysis for genomics batch correction. *Bioinformatics* **30**, 2757-2763, (2014). PMC4173013.

15 Fertig EJ, Ding J, Favorov AV, Parmigiani G & Ochs MF. CoGAPS: an R/C++ package to identify patterns and biological process activity in transcriptomic data. *Bioinformatics* **26**, 2792-2793, (2010). PMC3025742.

16 van Dijk D, Nainys J, Sharma R, Kathail P, Carr AJ, Moon KR, Mazutis L, Wolf G, Krishnaswamy S & Pe'er D. MAGIC: A diffusion-based imputation method reveals gene-gene interactions in single-cell RNA-sequencing data. *bioRxiv*, (2017). PMID PMCID not available.

17 Fertig EJ, Ozawa H, Thakar M, Howard JD, Kagohara LT, Krigsfeld G, Ranaweera RS, Hughes RM, Perez J, Jones S, Favorov AV, Carey J, Stein-O'Brien G, Gaykalova DA, Ochs MF & Chung CH. CoGAPS matrix factorization algorithm identifies transcriptional changes in AP-2alpha target genes in feedback from therapeutic inhibition of the EGFR network. *Oncotarget* **7**, 73845-73864, (2016). PMC5342018.

18 Moon KR, Dijk Dv, Wang Z, Chen W, Hirn MJ, Coifman RR, Ivanova NB, Wolf G & Krishnaswamy S. *PHATE: A Dimensionality Reduction Method for Visualizing Trajectory Structures in High-Dimensional Biological Data* (2017).

19 Tan J, Doing G, Lewis KA, Price CE, Chen KM, Cady KC, Perchuk B, Laub MT, Hogan DA & Greene CS. Unsupervised Extraction of Stable Expression Signatures from Public Compendia with an Ensemble of Neural Networks. *Cell Syst* **5**, 63-71 e66, (2017). PMC5532071.