# *References*

# Multi-omics Boot Camp

## Analysis of Omics Data for Research Studies

*Below are the link to the Github Repo, some additional readings shared by instructors, and links to papers and websites shared during the training.*

**2024 GitHub Repository for Lab Session**: *https://github.com/ContiLab-usc*

## Polygenic risk and summary statistics (Section 1)

- Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A. and Yang, J., 2017. 10 years of GWAS discovery: biology, function, and translation. The American Journal of Human Genetics, 101(1), pp.5-22.

- Pasaniuc, B. and Price, A.L., 2017. Dissecting the genetics of complex traits using summary association statistics. Nature reviews genetics, 18(2), pp.117-127.

- Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T. and Kathiresan, S., 2018. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature genetics, 50(9), pp.1219-1224.

- Lambert, S.A., Gil, L., Jupp, S., Ritchie, S.C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman, M., Parkinson, H. and Danesh, J., 2021. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nature Genetics, 53(4), pp.420-425.

- Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J. and Im, H.K., 2015. A gene-based association method for mapping traits using reference transcriptome data. Nature genetics, 47(9), pp.1091-1098.

- Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., De Geus, E.J., Boomsma, D.I., Wright, F.A. and Sullivan, P.F., 2016. Integrative approaches for large-scale transcriptome-wide association studies. Nature genetics, 48(3), pp.245-252.

- Sanderson, E., Glymour, M.M., Holmes, M.V., Kang, H., Morrison, J., Munafò, M.R., Palmer, T., Schooling, C.M., Wallace, C., Zhao, Q. and Davey Smith, G., 2022. Mendelian randomization. Nature Reviews Methods Primers, 2(1), pp.1-21.
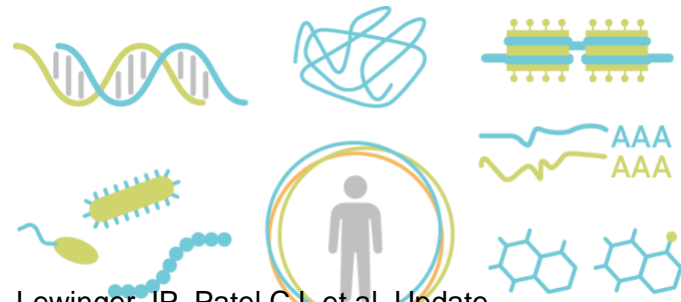
# GxE Interactions (Section 1)

**Review paper**
- Gauderman WJ, Mukherjee B, Aschard H, Hsu L, Lewinger JP, Patel CJ, et al. Update on the State of the Science for Analytical Methods for Gene-Environment Interactions. Am J Epidemiol. 2017;186(7):762-70.

**Two-step GxE testing**
*Disease trait (case-control studies)*
- DG screening: Kooperberg C, Leblanc M. Increasing the power of identifying gene x gene interactions in genome-wide association studies. Genet Epidemiol. 2008;32(3):255-63.
- EG screening: Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. Am J Epidemiol. 2009;169(2):219-26.
- EDGE (DG+EG) screening: Gauderman WJ, Zhang P, Morrison JL, Lewinger JP. Finding novel genes by testing G x E interactions in a genome-wide association study. Genet Epidemiol. 2013;37(6):603-13.
- Application to G x Alcohol for colorectal cancer: Jordahl KM, Shcherbina A, Kim AE, Su YR, Lin Y, Wang J, et al. Beyond GWAS of Colorectal Cancer: Evidence of Interaction with Alcohol Consumption and Putative Causal Variant for the 10q24.2 Region. Cancer Epidemiol Biomarkers Prev. 2022;31(5):1077-89.

*Quantitative trait*
- One trait: Zhang P, Lewinger JP, Conti D, Morrison JL, Gauderman WJ. Detecting Gene-Environment Interactions for a Quantitative Trait in a Genome-Wide Association Study. Genet Epidemiol. 2016;40(5):394-403
- Multivariate outcome: Majumdar A, Burch KS, Haldar T, Sankararaman S, Pasaniuc B, Gauderman WJ, et al. A two-step approach to testing overall effect of gene-environment interaction for multiple phenotypes. Bioinformatics. 2021.

*Survival trait*
- Kawaguchi ES, Li G, Lewinger JP, Gauderman WJ. Two-step hypothesis testing to detect gene-environment interactions in a genome-wide scan with a survival endpoint. Stat Med. 2022;41(9):1644-57.

*Weighted (bin-based) testing*
- Ionita-Laza I, McQueen MB, Laird NM, Lange C. Genomewide weighted hypothesis testing in family-based association studies, with an application to a 100K scan. Am J Hum Genet. 2007;81(3):607-14.
- Kawaguchi EK, Kim AK, Lewinger JP, Gauderman WJ. A novel significance-based approach to two-stage hypothesis testing for discovery of gene-environment interactions. *https://www.biorxiv.org/content/10.1101/2022.06.14.496154v1*

**Joint tests of GxE interaction**
- 2-df test: Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. Hum Hered. 2007;63(2):111-9.
- 3-df test: Gauderman WJ, Kim A, Conti DV, Morrison J, Thomas DC, Vora H, et al. A Unified Model for the Analysis of Gene-Environment Interaction. Am J Epidemiol. 2019;188(4):760-7.
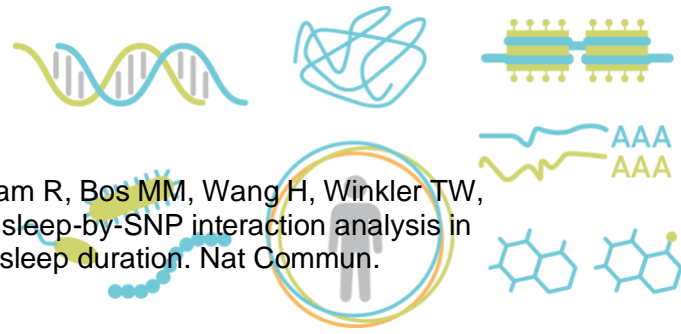
- Application of 2-df test for GWIS of lipids: Noordam R, Bos MM, Wang H, Winkler TW, Bentley AR, Kilpelainen TO, et al. Multi-ancestry sleep-by-SNP interaction analysis in 126,926 individuals reveals lipid loci stratified by sleep duration. Nat Commun. 2019;10(1):5121.

# Clustering (Section 3)

**Consensus PCA**
- Westerhuis, J. A.; Kourti, T.; Macgregor, J. F. Analysis of multiblock and hierarchical PCA and PLS moels J. Chemom. 1998, 12 (5) 301– 21 DOI: 10.1002/(SICI)1099-128X(199809/10)12:5<301::AID-CEM515>3.0.CO;2-S

**JIVE (like Consensus PCA)**
- https://pubmed.ncbi.nlm.nih.gov/23745156/
- Lock EF, Hoadley KA, Marron JS, Nobel AB. JOINT AND INDIVIDUAL VARIATION EXPLAINED (JIVE) FOR INTEGRATED ANALYSIS OF MULTIPLE DATA TYPES. Ann Appl Stat. 2013 Mar 1;7(1):523-542. doi: 10.1214/12-AOAS597. PMID: 23745156; PMCID: PMC3671601.

**mOcluster software for CPCA, RGCCA, MCIA:**
- https://pubmed.ncbi.nlm.nih.gov/26653205/
- J. Proteome Res. 2016, 15, 3, 755–765 Publication Date:December 14, 2015
- https://doi.org/10.1021/acs.jproteome.5b00824

**R.JIVE: Software for JIVE (~CPCA) (we didn't' study this variant, but it's used in the benchmark papers).**
- https://pubmed.ncbi.nlm.nih.gov/27273669/
- Michael J. O'Connell, Eric F. Lock, R.JIVE for exploration of multi-source molecular data, *Bioinformatics*, Volume 32, Issue 18, 15 September 2016, Pages 2877–2879, https://doi.org/10.1093/bioinformatics/btw324

**RGCCA software (we didn't study this variant, but it's used in many of the benchmark papers)**
- https://pubmed.ncbi.nlm.nih.gov/28536930/
- Tenenhaus M, Tenenhaus A, Groenen PJF. Regularized Generalized Canonical Correlation Analysis: A Framework for Sequential Multiblock Component Methods. Psychometrika. 2017 May 23. doi: 10.1007/s11336-017-9573-x. Epub ahead of print. PMID: 28536930.

**Method Comparisons:**
- https://pubmed.ncbi.nlm.nih.gov/30295871/
- Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Res. 2018 Nov 16;46(20):10546-10562. doi: 10.1093/nar/gky889. Erratum in: Nucleic Acids Res. 2019 Jan 25;47(2):1044. PMID: 30295871; PMCID: PMC6237755.

- https://pubmed.ncbi.nlm.nih.gov/31792509/

- Pierre-Jean M, Deleuze JF, Le Floch E, Mauger F. Clustering and variable selection evaluation of 13 unsupervised methods for multi-omics data integration. Brief Bioinform. 2020 Dec 1;21(6):2011-2030. doi: 10.1093/bib/bbz138. PMID: 31792509.

- https://pubmed.ncbi.nlm.nih.gov/33402734/
- Cantini L, Zakeri P, Hernandez C, Naldi A, Thieffry D, Remy E, Baudot A. Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. Nat Commun. 2021 Jan 5;12(1):124. doi: 10.1038/s41467-020-20430-7. PMID: 33402734; PMCID: PMC7785750.

- https://pubmed.ncbi.nlm.nih.gov/34383739/
- Duan R, Gao L, Gao Y, Hu Y, Xu H, Huang M, Song K, Wang H, Dong Y, Jiang C, Zhang C, Jia S. Evaluation and comparison of multi-omics data integration methods for cancer subtyping. PLoS Comput Biol. 2021 Aug 12;17(8):e1009224. doi: 10.1371/journal.pcbi.1009224. PMID: 34383739; PMCID: PMC8384175.

## Mediation (Section 4)

### General Mediation and Multiple Mediators

- Daniel, R.M., De Stavola, B.L., Cousens, S.N., and Vansteelandt, S. (2015). Causal mediation analysis with multiple mediators. Biometrics 71, 1-14.

- Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. Psychological Methods 15, 309-334.

- MacKinnon, D.P. (2008). Introduction to statistical mediation analysis. (New York: Lawrence Erlbaum Associates).

- Pearl, J. (2012). The Causal Mediation Formula—A Guide to the Assessment of Pathways and Mechanisms. Prev Sci 13, 426-436.

- Steen, J., Loeys, T., Moerkerke, B., and Vansteelandt, S. (2017). Flexible Mediation Analysis With Multiple Mediators. Am J Epidemiol 186, 184-193.

- VanderWeele, T.J., and Vansteelandt, S. (2014). Mediation Analysis with Multiple Mediators. Epidemiol Methods 2, 95-115.

- Vansteelandt, T.J.V., and Stijn. (2009). Conceptual issues concerning mediation, interventions and composition. 1-12.

### High Dimensional Mediation

- Huang, Y.-T. (2014). Integrative modeling of multi-platform genomic data under the framework of mediation analysis. Stat Med 34, 162-178.

- Huang, Y.-T., Liang, L., Moffatt, M.F., Cookson, W.O.C.M., and Lin, X. (2015). iGWAS: Integrative Genome-Wide Association Studies of Genetic and Genomic Data for Disease Susceptibility Using Mediation Analysis. Genet Epidemiol, n/a-n/a.
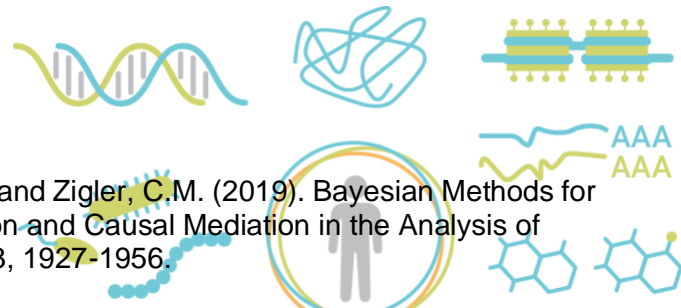
- Kim, C., Daniels, M.J., Hogan, J.W., Choirat, C., and Zigler, C.M. (2019). Bayesian Methods for Multiple Mediators: Relating Principal Stratification and Causal Mediation in the Analysis of Power Plant Emission Controls. Ann Appl Stat 13, 1927-1956.

- Song, Y., Zhou, X., Zhang, M., Zhao, W., Liu, Y., Kardia, S.L.R., Roux, A.V.D., Needham, B.L., Smith, J.A., and Mukherjee, B. (2020). Bayesian shrinkage estimation of high dimensional causal mediation effects in omics studies. Biometrics 76, 700-710.

- Zhang, H., Zheng, Y., Zhang, Z., Gao, T., Joyce, B., Yoon, G., Zhang, W., Schwartz, J., Just, A., Colicino, E., et al. (2016). Estimating and testing high-dimensional mediation effects in epigenetic studies. Bioinformatics 32, 3150-3154.

- Perera, C., Zhang, H., Zheng, Y., Hou, L., Qu, A., Zheng, C., Xie, K., and Liu, L. (2022). HIMA2: high-dimensional mediation analysis and its application in epigenome-wide DNA methylation data. BMC Bioinformatics 23, 296.

**Meet in the Middle**

- Assi, N., Fages, A., Vineis, P., Chadeau-Hyam, M., Stepien, M., Duarte-Salles, T., Byrnes, G., Boumaza, H., Knüppel, S., Kühn, T., et al. (2015). A statistical framework to model the meeting-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study. Mutagenesis 30, 743-753.

- Cadiou, S., Basagana, X., Gonzalez, J.R., Lepeule, J., Vrijheid, M., Siroux, V., and Slama, R. (2021). Performance of approaches relying on multidimensional intermediary data to decipher causal relationships between the exposome and health: A simulation study under various causal structures. Environ Int 153, 106509.

**Latent Mediation**

- Albert, J.M., Geng, C., and Nelson, S. (2016). Causal mediation analysis with a latent mediator. Biom J 58, 535-548.

- Derkach, A., Pfeiffer, R.M., Chen, T.H., and Sampson, J.N. (2019). High dimensional mediation analysis with latent variables. Biometrics 75, 745-756.

- Peng, C., Wang, J., Asante, I., Louie, S., Jin, R., Chatzi, L., Casey, G., Thomas, D.C., and Conti, D.V. (2019). A Latent Unknown Clustering Integrating Multi-Omics Data (LUCID) with Phenotypic Traits. Bioinformatics.

# Websites and Papers Shared During Boot Camp

**https://www.pgscatalog.org/**

**https://www.cog-genomics.org/plink/2.0/score**

**https://zzz.bwh.harvard.edu/plink/profile.shtml**

**Plink: https://www.cog-genomics.org/plink/2.0/**

**Plink Formats: https://www.cog-genomics.org/plink/2.0/formats**

**https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9081195/**

**https://www.ijcaonline.org/research/volume134/number7/irani-2016-ijca-907841.pdf**

**https://learn.saylor.org/mod/book/view.php?id=55626&chapterid=41483**

**https://www.bioconductor.org/packages/release/bioc/vignettes/scPCA/inst/doc/scpca_intro.html#sparse-pca**