Statistical methods for the analysis of complex omics data

Tissier, R.

Bibliography

Acharjee, A., B. Kloosterman, R. G. Visser, and C. Maliepaard (2016). Integration of multi-omics data for prediction of phenotypic traits using random forest. *BMC Bioinformatics 17 Suppl 5*.

Almasy, L. and J. Blangero (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *The American Journal of Human Genetics 62(5)*, 1198 – 1211.

Association., A. P. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.

Bahado-Singh, R., R. Ertl, R. Mandal, T. C. Bjorndahl, A. Syngelaki, B. Han, E. Dong, P. B. Liu, Z. Alpay-Savasan, D. S. Wishart, and K. H. Nicolaides (2014). Metabolomic prediction of fetal congenital heart defect in the first trimester. *American Journal of Obstetrics & Gynecology 211(3)*.

Balliu, B., R. Tsonaka, S. Boehringer, and J. Houwing-Duistermaat (2015). A retrospective likelihood approach for efficient integration of multiple omics factors in case-control association studies. *Genetic Epidemiology 39(3)*, 156 – 165.

Bas-Hoogendam, J. M., A. Harrewijna, R. L. M. Tissier, M. J. W. van der Molena, H. van Steenbergen, I. M., V. Vliet, C. G. Reichart, J. J. Houwing-Duistermaat, E. Slagboom, N. J. A. van der Wee, and M. Westenberg (2018). The Leiden Family Lab study on Social Anxiety Disorder: a multiplex, multigenerational family study on neurocognitive endophenotypes. *International Journal of Methods in Psychiatric Research (In press)*.

Beaty, T. H. and K. Y. Liang (1987). Robust inference for variance components models in families ascertained through probands: I. conditioning on proband's phenotype. *Genetic Epidemiology 4*, 203 – 210.

Boehnke, M. and D. A. Greenberg (2018). The leiden family lab study on social anxiety disorder: a multiplex, multigenerational family study on neurocognitive endophenotypes. *International Journal of Methods in Psychiatric Research In press*.

Bouhaddani, S., P. Houwing-Duistermaat, J. J. andd Salo, M. Perola, G. Jongbloed, and H.-W. Uh (2016). Evaluation of o2pls in omics data integration. *BMC Bioinformatics 17(Suppl 2)*, S11.

Bühlmann, P. and S. van de Geer (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications.* Berlin: Springer.

Cheung, C. Y., E. A. Thompson, and E. M. Wijsman (2013). Gigi: an approach to effective imputation of dense genotypes on large pedigrees. *The American Journal of Human Genetics 92(4)*, 504 – 516.

Chuang, H.-Y., E. Lee, Y.-T. Liu, D. Lee, and T. Ideker (2007). Network-based classification of breast cancer metastasis. *Molecular Systems Biology 3:140*.

Compier-de Block, L. H. C. G., L. R. A. Alink, M. Linting, L. J. M. van den Berg, B. M. Elzinga, A. Voorthuis, M. S. Tollenaar, and M. J. Bakermans-Kranenburg (2017). Parent-child agreement on parent-to-child maltreatment. *Journal of Family Violence 32*(2), 207–217.

de Andrade, M. and C. I. Amos (2000). Ascertainment issues in variance components models. *Genetic epidemiology 19*, 333 – 344.

de Jong, S., M. P. M. Boks, T. F. Fuller, E. Strengman, E. Janson, C. G. F. de Kovel, A. P. S. Ori, N. Vi, F. Mulder, J. D. Blom, B. Glenthøj, C. D. Schubart, W. Cahn, R. S. Kahn, S. Horvath, and R. A. Ophoff (2012). A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes. *PLoS One 7*(6), 1–10.

de Visser, M. C., R. van Minkelen, V. van Marion, M. den Heijer, J. Eikenboom, H. L. Vos, P. E. Slagboom, J. J. Houwing-Duistermaat, F. R. Rosendaal, and R. M. Bertina (2013). Genome-wide linkage scan in affected sibling pairs identifies novel susceptibility region for venous thromboembolism: Genetics in familial thrombosis study. *Journal of Thrombosis and Haemostasis 11(8)*, 1474–84.

Distel, M. A., J. M. Vink, G. Willemsen, C. M. Middeldorp, H. Merckelbach, and D. I. Boomsma (2008). Heritability of self-reported phobic fear. *Behavior Genetics 38*, 24 – 33.

Dubdbridge, F. (2003). Power and predictive accuracy of polygenic risk scores. *The American Journal of Psychiatry 160(4)*, 636 – 645.

Efron, B. (2004). Large-scale simultaneous hypothesis testing. *Journal of the American Statistical Association 99*, 96–104.

Elston, R. C. and J. Stewart (2013). A general model for the analysis of pedigree data. *Human Heredity 21*, 523–542.

Fisher, R. A. (1924). The distribution of the partial correlation coefficient. *Metron 3*, 329–332.

Friedman, J., T. Hastie, and R. Tibshirani (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics 9(3)*, 432–441.

Friedrichs, S., J. Manitz, P. Burger, C. I. Amos, A. Risch, and J. e. a. Chang-Claude (2017). Pathway-based kernel boosting for the analysis of genome-wide association studies. *Computational and Mathematical Methods in Medicine*, Article ID 6742763.

Fuady, A. M., R. Tissier, and J. J. Houwing-Duistermaat (2018). Genome-wide analysis in multiple-case families: assessing the relationship between triglyceride and methylation. *BMC Proceedings*.

Furmark, T. (2002). Social phobia: overview of community surveys. *Acta Psychiatrica Scandinavica 105*, 84 – 93.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics 1*, 141 – 150.

Ghosh, A., F. A. Wright, and F. Zou (2013). Unified analysis of secondary traits in case-control association studies. *Journal of the American Statistical Association 108(52)*, 140 – 151.

Ghosh, D. and A. M. Chinnaiyan (2005). Classification and selection of biomarkers in genomic data using lasso. *Journal of Biomedicine and Biotechnology 2005(2)*, 147–154.

Glahn, D. C., P. M. Thompson, and J. Blangero (2007). Neuroimaging endophenotypes: Strategies for finding genes influencing brain structure and function. *Human Brain Mapping 28*, 488 – 501.

Gottesman, I. I. and T. D. Gould (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry 160*, 636 – 645.

Gratten, J. and P. M. Visscher (2016). Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. *Genome Medicine 8*, 78.

Greenwood, T. A., D. L. Braff, G. A. Light, K. S. Cadenhead, M. E. Calkins, D. J. Dobie, R. Freedman, M. F. Green, R. E. Gur, R. C. Gur, J. Mintz, K. H. Nuechterlein, A. Olincy, A. D. Radant, L. J. Seidman, L. J. Siever, J. M. Silverman, W. S. Stone, N. R. Swerdlow, D. W. Tsuang, M. T. Tsuang, B. I. Turetsky, and N. J. Schork (2007). Initial heritability analyses of endophenotypic measures for schizophrenia: the consortium on the genetics of schizophrenia. *Archives of General Psychiatry 64(11)*, 1242 – 1250.

Ha, M. J. and W. Sun (2014). Partial correlation matrix estimation using ridge penalty followed by thresholding and re-estimation. *Biometrics 70(3)*, 765–773.

Hardin, J., S. R. Garcia, and D. Golan (2013). A method for generating realistic correlation matrices. *The Annals of Applied Statistics 7(3)*, 1733 – 1762.

Harrewijn, A., M. J. W. van der Molen, and P. M. Westenberg (2016). Putative eeg measures of social anxiety: Comparing frontal alpha asymmetry and deltaâĂŞbeta cross-frequency correlation. *Cognitive, Affective, & Behavioral Neuroscience 6*, 1086–1098.

Hastie, T. J., R. J. Tibshirani, and J. H. Friedman (2009). *The elements of statistical learning : data mining, inference, and prediction.* New York, Springer: Springer series in statistics.

He, J., H. Li, A. C. Edmonson, D. J. Rader, and M. Li (2011). A gaussian copula approach for the analysis of secondary phenotypes in case-control genetic association studies. *Biostatistics 13(3)*, 497 – 508.

Hoerl, A. E. and R. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12*, 55–67.

Hopper, J. L. and J. D. Mathews (1982). Extensions to multivariate normal models for pedigree analysis. *Annals of Human Genetics 46*, 373 – 383.

Houwing-Duistermaat, J. J., A. Callegaro, M. Beekman, R. G. Westendorp, P. E. Slagboom, and J. C. van Houwelingen (2009). Weighted statistics for aggregation and linkage analysis of human longevity in selected families: the Leiden Longevity Study. *Statistics in Medicine 28(1)*, 140 – 151.

Iacono, W. G., S. M. Malone, and S. I. Vrieze (2017). Endophenotype best practices. *International Journal of Psychophysiology 111*, 115 – 144.

Ibrahim-Verbaas, C. A., M. Fornage, J. C. Bis, S. H. Choi, B. M. Psaty, J. B. Meigs, M. Rao, M. Nalls, J. D. Fontes, and C. J. e. a. OâĂŹDonnell (2014). Predicting stroke through genetic risk functions: the charge risk score project. *Stroke 45 (2014)*, 403–412.

(IMSGC), I. M. S. G. C., W. S. Bush, S. J. Sawcer, P. L. de Jager, J. R. Oksenberg, J. L. McCauley, M. A. Pericak-Vance, and J. L. Haines (2010). Evidence for polygenic susceptibility to multiple sclerosisâĂŞthe shape of things to come. *The American Journal of Human Genetics 86(4)*, 421 – 425.

Inouye, M., J. Kettunen, P. Soininen, K. Silander, S. Ripatti, and L. S. e. a. Kumpula (2010). Metabonomic, transcriptomic, and genomic variation of a population cohort. *Molecular Systems Biology 21(6)*.

International Schizophrenia Consortium, ., P. S. M., N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'Donovan, P. F. Sullivan, and P. Sklar (2009, jul). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature 460(7256)*, 748 – 752.

Irvin, M. R., D. Zhi, R. Joehanes, M. Mendelson, S. Aslibekyan, S. A. Claas, K. S. Thibeault, N. Patel, K. Day, L. W. Jones, L. Liang, B. H. Chen, C. Yao, H. K. Tiwari, J. M. Ordovas, D. Levy, D. Absher, and D. K. Arnett (2014). Epigenome-wide association study of fasting blood lipids in the genetics of lipid-lowering drugs and diet network study: Clinical perspective. *Circulation 130*(7), 565–572.

Isomura, K., M. Boman, C. Ruck, E. Serlachius, H. Larsson, P. Lichtenstein, and D. Mataix-Cols (2015). Population-based, multi-generational family clustering study of social anxiety disorder and avoidant personality disorder. *Psychological Medicine 45*, 1581 – 1589.

Jacob, L., G. Obozinski, and J.-P. Vert (2009). Group lasso with overlap and graph lasso. *Proceedings of the International Conference on Machine Learning (ICML) ICML '09*.

Jolliffe, I. T. (2008). *Principal Component Analysis.* New York: Springer-Verlag.

Kendler, K. S., M. C. Neale, R. C. Kessler, A. C. Heath, and L. J. Eaves (1992). The genetic epidemiology of phobias in women - the interrelationship of agoraphobia, social phobia, situational phobia, and simple phobia. *Archives of General Psychiatry 49*, 273 – 281.

Kivelä, M., A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Proter (2014). Multilayer networks. *Journal of complex networks 2(3)*, 203–271.

Kraft, P., E. Schadt, J. Aten, and S. Horvath (2003, May). A family-based test for correlation between gene expression and trait values. *The American Journal of Human Genetics 72*(5), 1323–1330.

Kraft, P. and D. C. Thomas (2000). Bias and efficiency in family-based gene-characterization studies: Conditional, prospective, retrospective, and joint likelihoods. *The American Journal of Human Genetics 66*(3), 1119–1131.

Krumsiek, J., K. Suhre, T. Illig, J. Adamski, and F. J. Theis (2011). Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Systems Biology 5:21*.

Langfelder, P. and S. Horvath (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics 9*(1), 559.

Langfelder, P., B. Zhang, and S. Horvath (2008). Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r. *Journal of the American Statistical Association 99*, 96–104.

Lauritzen, S. L. (1996). *Graphical models.* Oxford, Clarendon Press: Oxford statistical science series.

Lee, A., L. McMurchy, and A. J. Scott (1997). Re-using data from case-control studies. *Statistics in Medicine 16(12)*, 1377 – 1389.

Lemesle, G., F. Maury, O. Beseme, L. Ovart, P. Amouyel, N. Lamblin, P. de Groote, C. Bauters, and F. Pinet (2015). Multimarker proteomic profiling for the prediction of cardiovascular mortality in patients with chronic heart failure. *PLoS One 10(4)*.

Li, C. and H. Li (2008). Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics 24(9)*, 1175–1182.

Li, H. and G. M. H. (2012). Efficient adaptively weighted analysis of secondary phenotypes in case-control genome-wide association studies. *Human Heredity 73*, 159 – 173.

Lin, D. Y. and D. Zeng (2009). Proper analysis of secondary phenotype data in case-control association studies. *Genetic Epidemiology 33*, 256 – 265.

Liu, J., T.-Z. Huang, Z. Xu, and X.-G. Lv (2013). High-order total variation-based multiplicative noise removal with spatially adapted parameter selection. *Journal of the Optical Society of America A 30*, 1956–1966.

Mertens, B. J. A., M. E. de Noo, R. A. E. M. Tollenaar, and A. M. Deelder (2006). Mass spectrometry proteomic diagnosis: enacting the double crossvalidatory paradigm. *Journal of Computational Biology 13*, 1591–1605.

Mertens, B. J. A., Y. E. M. van der Burgt, B. Velstra, W. E. Mesker, R. A. E. M. Tollenaar, and A. M. Deelder (2011). On the use of double crossvalidation for the combination of proteomic mass spectral data for enhanced diagnosis and prediction. *Statistics and Probability Letters 81*, 759–766.

Miller, G. A. and B. Rockstroh (2013). Endophenotypes in psychopathology research: where do we stand? *Annual Review of Clinical Psychology 9*, 177 – 213.

Miwa, T., A. J. Hayer, and S. Kuriki (2003). The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*, 223 – 234.

Monsees, G. M., R. M. Taqmimi, and P. Kraft (2009). Genome-wide association scans for secondary traits using case-control samples. *Genetic Epidemiology 33*, 717 – 728.

Mooijaart, S. P., D. van Heems, R. Noordman, M. P. Rozing, C. A. Wijsman, A. J. M. de Craen, R. G. J. Westendorp, M. Beekman, and E. P. Slagboom (2010). Polymorphisms associated with type 2 diabetes in familial longevity: The Leiden Longevity Study. *Aging 3*, 55 – 62.

Mootha, V. K., C. M. Lindgren, K.-F. Eriksson, A. Subramanian, S. Sihag, and J. e. a. Lehar (2003). Pgc-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics 34*, 267–273.

Najita, J. S., Y. Li, and P. J. Catalano (2009). A novel application of a bivariate regression model for binary and continuous outcomes to studies of fetal toxicity. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 58(4)*, 555 – 573.

Noah, S., J. Friedman, T. Hastie, and R. Tibshirani (2013). A sparse-group lasso. *Journal of Computational and Graphical Statistics 22:2*, 231–245231–245.

Oldham, M., S. Horvath, and D. Geschwind (2006). Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proceedings of the National Academy of Sciences 103(47)*, 17973 – 17978.

Oldham, M., G. Konopka, K. Iwamoto, P. Langfelder, T. Kato, S. Horvath, and D. Geschwind (2008). Functional organization of the transcriptome in human brain. *Nature Neuroscience 11(11)*, 1271 – 1282.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine 2(11)*, 559 – 572.

Pena, M. J., A. Heinzel, P. Rossing, H. Parving, G. Dallma, and K. e. a. Rossing (2016). Serum metabolites predict response to angiotensin ii receptor blockers in patients with diabetes mellitus. *Journal of Translational Medicine 14*, 203.

Plaisier, C. L., S. Horvath, A. Huertas-Vazquez, I. Cruz-Bautista, M. F. Herrera, T. Tusie-Luna, C. Aguilar-Salinas, and P. Pajukanta (2009). A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. *PLoS Genetics 5*(9), 1–10.

Rao, K. R. and S. Lakshminarayanan (2007). Partial correlation based variable selection approach for multivariate data classification methods. *Chemometrics and Intelligent Laboratory Systems 86(1)*, 68–81.

Reis, M. S. (2013). Applications of a new empirical modelling framework for balancing model interpretation and prediction accuracy through the incorporation of clusters of functionally related variables. *Chemometrics and Intelligent Laboratory Systems 127*, 7– 16.

Richardson, D. B., P. Rzehak, J. Klenk, and S. K. Weiland (2007). Analyses of case-control data for additional outcomes. *Epidemiology 8(4)*, 441 – 445.

Rodríguez-Girondo, M., J. Deelen, P. E. Slagboom, and J. J. Houwing-Duistermaat (2018). Survival analysis with delayed entry in selected families with application to human longevity. *Statistical Methods in Medical Research 27*(3), 933–954.

Rodríguez-Girondo, M., P. Salo, T. Burzykowsky, M. Perola, J. J. Houwing-Duistermaat, and B. Mertens (2018). Sequential double cross-validation for augmented prediction assessment in high-dimensional omic applications. *Annals of Applied Statistics*.

Sásik, R., E. Calvo, and J. Corbeil (2002). Statistical analysis of high-density oligonucleotide arrays: a multiplicative noise model. *Bioinformatics 18(12)*, 1633–1640.

Schade, D. S. and P. R. Eaton (1974). Role of insulin and glucagon in obesity. *Diabetes 23(8)*, 657–661.

Schäfer, J. and K. Strimmer (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology 4*, Art. 32.

Schifano, E. D., L. Li, D. C. Christiani, and X. Lin (2013). Genome-wide association analysis for multiple continuous secondary phenotypes. *The American Journal of Human Genetics 92(5)*.

Schoenmaker, M., A. J. M. de Craen, P. H. E. M. de Meijer, M. Beekman, G. J. Blauw, P. E. Slagboom, and R. G. J. Westendorp (2006). Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *European Journal Of Human Genetics 14*, 79 EP –.

Shahabi, A., J. P. Lewinger, J. Ren, C. April, A. E. Sherrod, and J. G. e. a. Hacia (2006). Novel gene expression signature predictive of clinical recurrence after radical prostatectomy in early stage prostate cancer patients. *Prostate 76(14)*, 1239–1256.

Shamai, L., E. Lurix, M. Shen, G. M. Novaro, S. Szomstein, and R. e. a. Rosenthal (2011). Association of body mass index and lipid profiles: evaluation of a broad spectrum of body mass index patients including the morbidly obese. *Obesity Surgery 21(1)*, 42–47.

Shim, J. E., C. Bang, S. Yang, T. Lee, S. Hwang, and C. Y. e. a. Kym (2017). Gwab: a web server for the network-based boosting of human genome-wide association data. *Nucleic Acids Research 45(1)*, W154–W161.

Simon, N., J. Friedman, T. Hastie, and R. Tibshirani (2013). A sparse-group lasso. *Journal of Computational and Graphical Statistics 22(2)*, 231–245.

Simonson, M. A., A. G. Wills, M. C. Keller, and M. B. McQueen (2011). Recent methods for polygenic analysis of genome-wide data implicate an important effect of common variants on cardiovascular disease risk. *BMC Medical Genetics 12*, 146.

Skytthe, A., S. Valensin, B. Jeune, E. Cevenin, F. Balard, M. Beekman, V. Bezrukov, H. Blanch, L. Bolund, K. Broczek, C. Carru, K. Christensen, L. Christiansen, J. C. Collerton, and R. Cotichini (2011). Design, recruitment, logistics, and data management of the GEHA (genetics of healthy ageing) project. *Experimental Gerontology 46(11)*, 934–945.

Solovieff, N., C. Cotsapas, P. H. Lee, S. M. Purcell, and J. W. Smoller (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics 14*, 483–495.

Stuart, J. M., E. Segal, D. Koller, and S. K. Kim (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science 302(5643)*, 249 – 255.

Subramaniana, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Eberta, and M. A. e. a. Gillette (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences 21(1)*, 15545–15550.

Therneau, T. M. (2018). *coxme: Mixed Effects Cox Models*. R package version 2.2-7.

Thomas, D. C. (2004). *Statistical Methods in Genetic Epidemiology.* Oxford: Oxford University Press.

Thompson, E. A. (2008). The ibd process along four chromosomes. *Theoretical Population Biology 73*(3), 369 – 373.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B 58(1)*, 267–288.

Tissier, R., J. J. Houwing-Duistermaat, and M. Rodríguez-Girondo (2018). Improving stability of prediction models based on correlated omics data by using network approaches. *PLoS One 13(2)*, e0192853.

Tissier, R., R. Tsonaka, S. P. Mooijaart, E. Slagboom, and J. J. Houwing-Duistermaat (2017). Secondary phenotype analysis in ascertained family designs: application to the leiden longevity study. *Statistics in Medicine 36(14)*, 2288 – 2301.

Trygg, J. and S. Wold (2003). O2-pls, a two-block (xâĂŞy) latent variable regression (lvr) method with an integral osc filter. *Journal of Chemometrics 7(1)*, 53–64.

Tsonaka, R., M. C. H. de Visser, and J. J. Houwing-Duistermaat (2013). Estimation of genetic effects in multiple cases family studies using penalized maximum likelihood methodology. *Biostatistics 14(2)*, 220 – 231.

Tsonaka, R., D. van der Woude, and J. J. Houwing-Duistermaat (2015). Marginal genetic effects estimation in family and twin studies using random-effects models. *Biometrics 71(4)*, 1130 – 1138.

Turetsky, B. I., T. A. Greenwood, A. Olincy, A. D. Radant, D. L. Braff, K. S. Cadenhead, D. J. Dobie, R. Freedman, M. F. Green, R. E. Gur, R. C. Gur, G. A. Light, J. Mintz, K. H. Nuechterlein, N. J. Schork, L. J. Seidman, L. J. Siever, J. M. Silverman, W. S. Stone, N. R. Swerdlow, D. W. Tsuang, M. T. Tsuang, and M. E. Calkins (2015). Abnormal auditory n100 amplitude: a heritable endophenotype in first-degree relatives of schizophrenia probands. *Biol Psychiatry 64(12)*, 1051 – 1059.

van de Wiel, M. A., T. G. Lien, W. Verlaat, W. N. van Wieringen, and S. M. Wilting (2014). Better prediction by use of co-data: adaptive group-regularized ridge regression. *Statistics in medicine 35(3)*.

Welter, D., J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, and H. Parkinson (2014). The nhgri gwas catalog, a curated resource of snp-trait associations. *Nucleic Acids Research 42(Database Issue)*, D1001–D1006.

Winter, C., G. Kristiansen, S. Kersting, J. Roy, D. Aust, and T. e. a. Knösel (2012). Google goes cancer: Improving outcome prediction for cancer patients by network-based ranking of marker genes. *PLoS Computational Biology*.

Yip, A. M. and S. Horvath (2007). The generalized topological overlap matrix for detecting modules in gene networks. *BMC Bioinformatics 8(22)*.

Yuan, M. and Y. Lin (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B 68(1)*, 49 – 67.

Zemmour, C., F. Bertucci, P. Finetti, B. Chetrit, T. Filleron, and J. M. Boher (2015). Prediction of early breast cancer metastasis from dna microarray data using high-dimensional cox regression models. *Cancer Informatics 14(Suppl 2)*, 129–138.

Zhang, B. and S. Horvath (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology 4*, Article17.

Zhu, Y., X. Shen, and W. Pan (2009). Network-based support vector machine for classification of microarray samples. *BMC Bioinformatics 10(Suppl I)*, S21.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B 67*, 301–320.