**BIBLIOGRAPHY & REFERENCES CITED**

1. Danese S, Fiorino G, Peyrin-Biroulet L. (2017) Early intervention in Crohn’s disease: toward disease modification trials. *Gut* Sep. 5. [Epub ahead of print]

2. Jakimovski D, Weinstock-Guttman B, Ramanathan M, Kolb C, et al. (2017) Ocrelizumab: a B-cell depleting therapy for multiple sclerosis. *Expert Opin Biol Ther* 17(9):1163-1172.

3. Zink A, Albrecht K. (2017) Rheumatoid arthritis: The benefits of early treatment after decades. *Nat Rev Rheumatol.* 13(8):458-459.

4. Smolen JS, Landewe R, Bijlsma J, Burmester G, et al. (2016) EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 76(6):960-977.

5. Barkham N, Keen HI, Coates LC, O’Connor P, et al. (2009) Clinical and imaging efficacy of infliximab in HLA-B27-Positive patients with magnetic resonance imaging-determined early sacroiliitis. *Arthritis Rheum* 60(4):946-954.

6. Romanelli RJ, Chung S, Pu J, Nimbal V, et al. (2015) Comparative effectiveness of early versus delayed metformin in the treatment of type 2 diabetes*. Diabetes Res Clin Pract* 108(1):170-178.

7. Frydkjaer-Olsen U, Soegaard Hansen R, Simo R, Cunha-Vaz J, et al. (2016) Correlation between retinal vessel caliber and neurodegeneration in patients with type 2 diabetes mellitus in the European Consortium for the Early Treatment of Diabetic Retinopathy (EUROCONDOR). *Ophthalmic Res* 56(1):10-6.

8. Chaudhuri AA, Chabon JJ, Lovejoy AF, Newman AM, et al. (2017) Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer Discov* Sep 12 [Epub ahead of print]

9. Nair B, Wilt T, MacDonald R, Rutks I. (2002) Early versus deferred androgen suppression in the treatment of advanced prostatic cancer. *Cochrane Database Syst Rev* 1:CD003506.

10. Neal RD, Tharmanathan P, France B, Din NU, et al. (2015) Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer* 112(Suppl a):S92-S107.

11. Ford I, Murray H, Packard CJ, Shepherd J, et al. (2007) Long-term follow-up of the West of Scotland Coronary Prevention Study. *N Engl J Med* 357:1477-1486.

12. Lindstrom J, Peltonen M, Eriksson JG, Aunola S, et al. (2008) Determinants for the effectiveness of lifestyle intervention in the Finnish Diabetes Prevention Study. *Diabetes Care* 31(5):857-862.

13. Sacre JW, Jellis CL, Jenkins C, Haluska BA, et al. (2014) A six-month exercise intervention in subclinical diabetic heart disease: Effects on exercise capacity, autonomic and myocardial function. *Metabolism* 63(9):1104-1114.

14. Campbell F, Conti G, Heckman JJ, Moon SH, et al. (2014) Early childhood investments substantially boost adult health. *Science* 343(6178):1478-1485.

15. Berentzen NE, van Rossem L, Gehring U, Koppelman GH, et al. (2016) Overweight patterns throughout childhood and cardiometabolic markers in early adolescence*. Int J Obes* 40(1):58-64.

16. Hu Y, Bhupathiraju SN, de Koning L, Hu FB. (2014) Duration of obesity and overweight and risk of type 2 diabetes among US women. *Obesity* 22(10):2267-2273.

17. Carter TC, Rein D, Padberg I, Peter E, et al. (2016) Validation of a metabolite panel for early diagnosis of type 2 diabetes. *Metabolism* 65(9):1399-1408.

18. Herbst RS, Soria J-C, Kowanetz M, Fine GD, et al. (2014) Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* 515(7528):563-567.

19. Zethelius B, Berglund L, Sundstrom J, Ingelsson E, et al. (2008) Use of multiple biomarkers to improve the prediction of death from cardiovascular causes. *N Engl J Med* 358:2107-2116.

20. Schrag A, Siddiqui UF, Anastasiou Z, Weintraub N, Schott JM (2017) Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson’s disease: a cohort study. *Lancet Neurol* 16(1):66-75.

21. Schrodi SJ, Mukherkee S, Shan Y, Tromp G, et al. (2014) Genetic-based prediction of disease traits: prediction is very difficult, especially about the future. *Front Genet* 5:162.

22. Hebbring SJ, Slager SL, Epperla N, Mazza JJ, et al. (2013) Genetic evidence of PTPN22 effects on chronic lymphocytic leukemia. *Blood* 121(1):237-238.

23. Hebbring SJ, Schrodi SJ, Ye Z, Zhou Z, et al. (2013) A PheWAS approach in studying HLA-DRB1\*1501. *Genes Immun* 14(3):187-191.

24. O’Brien SE, Schrodi SJ, Ye Z, Brilliant MH, et al. (2015) Differential lipid response to statins is associated with variants in the BUD13-APOA5 gene region. *J Cardiovasc Pharmacol* 66(2):183-188.

25. Marouli E, Graff M, Medina-Gomez C, Lo KS, et al. (2017) Rare and low-frequency coding variants alter human adult height. *Nature* 542(7640):186-190.

26. Schmidt AF, Swerdlow DI, Holmes MV, Patel RS, et al. (2017) PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomization study. *Lancet Diabetes Endocrinol* 5(2):97-105.

27. Fritsche LG, IgI W, Bailey JN, Grassman F, et al. (2016) A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet* 48(2):134-143.

28. Liu Y, Ye Z, Li X, Anderson JL, et al. (2017) Genetic and functional associations with decreased anti-inflammatory tumor necrosis factor alpha induced protein 3 in macrophages from subjects with axial spondyloarthritis. *Front Immunol* 8:860.

29. Maadooliat M, Bansal NK, Upadhya J, Farazi MR, et al. (2016) The decay of disease association with declining linkage disequilibrium: a fine mapping theorem. *Front Genet* 7:217.

30. Bansal NK, Maadooliat M, Schrodi SJ. Empirical Bayesian approach to multiple hypotheses with separate priors for left and right alternatives. (Manuscript Under review)

31. Hasin Y, Seldin M, Lusis A. (2017) Multi-omics approaches to disease. *Genome Biol* 18:83.

32. Miao R, Luo H, Zhou H, Li G, et al. (2014) Identification of prognostic biomarkers in hepatitis B virus-related hepatocellular carcinoma and stratification by integrative multi-omics analysis. *J Hepatol* 61(4):840-849.

33. Gjoneska E, Pfenning AR, Mathys H, Quon G, et al. (2015) Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer’s disease. *Nature* 518:365-369.

34. Zeevi D, Korem T, Zmora N, Israeli D, et al. (2015) Personalized nutrition by prediction of glycemic responses. *Cell* 163:1079-1094.

35. Colombo BM, Canevali P, Magnani O, Rossi E, et al. (2012) Defective expression and function of the leukocyte associated Ig-like receptor 1 in B lymphocytes from systemic lupus erythematosus patients. *PLoS One* 7(2):e31903.

36. Zhang Z, Liu J, Ding X, Bijma P, et al. (2010) Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS One* 5(9):e12648.

37. Meuwissen THE, Hayes BJ, Goddard ME. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.

38. VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91(11):4414-4423.

39. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, et al. (2011) Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 43:519-525.

40. de Los Campos G, Vazquez AI, Fernando R, Klimentidis YC, Sorensen D. (2013) Prediction of complex human traits using genomic best linear unbiased predictor. *PLoS Genet* 9(7):e1003608.

41. Yang J, Benyamin B, McEvoy BP, Gordon S, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42:565-569.

42. Maadooliat M, Zhou S, Najibi SM, Gao X, Huang JZ. (2016) Collective estimation of multiple bivariate density functions with application to angular-sampling-based protein loop modeling. *J Am Stat Assoc* 11(513):43-56.

43. Kim HM, Maadooliat M, Arellano-Valle RB, Genton MG. (2016) Skewed factor models using selection mechanism. *J Multivariate Analysis* 145:162-177.

44. Maadooliat M, Pourahmadi M, Huang JZ. (2013) Robust estimation of the correlation matrix of longitudinal data. *Stat Comp* 23(1):17-28.

45. Maadooliat M, Huang JZ, Hu J. (2012) Analyzing multiple-probe microarray: estimation and application of gene expression indexes. *Biometrics* 68(3):784-792.

46. Maadooliat M, Huang JZ, Hu J. (2015) Integrating data transformation in principal components analysis. *J Computational and Graphical Statistics*. 24(1):84-103.

47. Schrodi SJ. (2005) A probabilistic approach to large-scale association scans: a semi-Bayesian method to detect disease-predisposing alleles. *Stat Appl Genet Mol Biol* 4(1):Article 31.

48. Schrodi SJ, DeBarber A, He M, Ye Z, et al. (2015) Prevalence estimation for monogenic autosomal recessive diseases using population-based genetic data. *Hum Genet* 134(6):659-669.

49. He M,, Person TN, Hebbring SJ, Heinzen E, et al. (2015) SeqHBase: a big data toolset for family based sequencing data analysis. *J Med Genet* 52(4):282-288.

50. Schrodi SJ. (2016) The use of multiplicity corrections, order statistics and generalized family-wise statistics with application to genome-wide studies. *PLoS One* 11(4):e0154472.

51. Ye Z, Tafti AP, He KY, Wang K, He MM. (2016) Sparktext: Biomedical text mining on big data framework. *PLoS One* 11(9):e0162721.

52. Mosley JD, Witte JS, Larkin EK, Bastarache L, et al. (2016) Identifying genetically driven clinical phenotypes using linear mixed models. *Nat Commun* 7:11433.

53. Fisher B, Redmond C, Fisher ER, Caplan R. (1988) Relative worth of estrogen or progesterone receptor and pathologic characteristics of differentiation as indicators of prognosis in node negative breast cancer patients: findings from National Surgical Adjuvant Breast and Bowel Project Protocol B-06. *J Clin Oncology* 6:1076-1087.

54. Fisher B, Costantino J, Redmond C, Poisson R, et al. (1989) A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. *N Engl J Med* 320: 479-484.

55. Paik S, Shak S, Tang G, Kim C, et al. (2004) A multigene assay to predict recurrence of tamoxifen-treated node-negative breast cancer. *N Engl J Med* 351:2817-2826.

56. Madeira S, Oliveira AL (2004) Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM transactions on Computational Biology and Bioinformatics (TCBB)* 1(1):24-45.

57. Pang Y, Tao D, Yuan Y, Li X. (2008) Binary two-dimensional PCA. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 38(4):1176-1180.

58. Fan J, Lv J. (2008) Sure independence screening for ultrahigh dimensional feature space. *J R Statist Soc B* 70(5):849-911.

59. Friedman N, Geiger D, Goldszmidt M (1997) Bayesian network classifiers. *Machine Learning* 29(2-3):131-163.

60. Jordan MI, Ghahramani Z, Jaakkola TS, Saul LK (1999) An introduction to variational methods for graphical models. *Machine Learning* 37(2):183-233.

61. Hastie T, Tibshirani R, Friedman JH (2009) *The elements of statistical learning: data mining, inference, and prediction* (2nd ed.). Springer-Verlag, New York.

62. Schrodi SJ. The impact of diagnostic code misclassification on optimizing the experimental design of genetic association studies *J Healthc Eng* (in Press)

63. McCarty CA, Wilke RA, Giampietro PF, Wesbrook SD, Caldwell MD. (2005) Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large population-based biobank. *Per Med* 2:49-79.

64. <https://emerge.mc.vanderbilt.edu/>

65. Holzinger ER, Verma SS, Moore CB, Hall M, et al. (2017) Discovery and replication of SNP-SNP interactions for quantitative lipid traits in over 60,000 individuals. *BioData Min* 10:25.

66. El-Gabalawy H, Robinson D, Smolik I, Hart D, et al. (2012) Familial clustering of the serum cytokine profile in the relatives of rheumatoid arthritis patients. *Arthritis Rheum* 64(6):1720-1729.

67. Egli A, Santer DM, O’Shea D, Barakat K, et al. (2014) IL-28B is a key regulator of B- and T-cell vaccine responses against influenza. *PLoS Pathog* 10(12):e1004556.

68. Heard BJ, Rosvold JM, Fritzler MJ, El-Gabalawy H, et al. (2014) A computational method to differentiate normal individuals, osteoarthritis and rheumatoid arthritis patients using serum biomarkers. J R Soc Interface 11(97):20140426.

69. Sundaresh S, Doolan DL, Hirst S, My Y, et al. (2006) Identification of humoral immune responses in protein microarrays using DNA microarray data analysis techniques. *Bioinformatics* 22:1760-1766.

70. Vigil A, Davies DH, Felgner PL (2010) Defining the humoral immune response to infectious agents using high-density protein microarrays. *Future Microbiol* 5:241-251.

71. Lee SJ, Liang L Juarez S, Nanton MR, et al. (2012) Identification of a common immune signature in murine and human systemic Salmonellosis. *Proc Natl Acad Sci USA* 109(13):4998-5003.

72. Kunnath-Velayudhan S, Salamon H, Wang HY, Davidow AL, et al. (2010) Dynamic antibody responses to the Mycobacterium tuberculosis proteome. *Proc Natl Acad Sci USA* 107(33):14703-14708.

73. Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am J Hum Genet* 81:1084-1097.

74. Ma S, Huang J. (2008) Penalized feature selection and classification in bioinformatics. *Briefing in Bioinformatics*, 9(5):392–403.

75. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Royal Stat Society. Series B (Methodological)*. 58(1):267-288.

76. Hui Zou. (2006) The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476):1418-1429.

77. Lee S, Huang JZ and Hu J. (2010) Sparse logistic principal component analysis for binary data. *Annuals of Applied Statistics*, 4(3):1579-1601.

78. Cheng Y, Church GM. (2000) Biclustering of expression data, *ISMB 8*, 93-103.