# Twin Research and Human Genetics

# A Novel Approach for Pathway Analysis of GWAS Data Highlights Role of BMP Signalling and Muscle Cell Differentiation in Colorectal Cancer Susceptibility

# Aniket Mishra1, Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), Colorectal Cancer Family Registry (CCFR), and Stuart MacGregor1

# 1 Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia

# Supplementary text

## Supplementary text 1: Participating studies and data processing for colorectal cancer GWAS meta-analysis

The GWAS meta-analysis on colorectal cancer was reported previously(Peters et al., 2013). 11 different studies participated in meta-analysis totalling 10934 cases and 12328 controls of European ancestry. Participating studies were: French Association STudy Evaluating RISK (ASTERISK) (Kury et al., 2007); the Colon Cancer Family Registry (CCFR) (Newcomb, Baron, et al., 2007); Darmkrebs: Chancen der Verhütung durch Screening (DACHS) (Vossen, Hoffmeister, Chang-Claude, Rosendaal, & Brenner, 2011); Diet, Activity, and Lifestyle Study (DALS)(Slattery et al., 1997); Health Professionals Follow-up Study (HPFS) (Rimm et al., 1990); Nurses’ Health Study (NHS) (Belanger, Hennekens, Rosner, & Speizer, 1978); Physician’s Health Study (PHS)("Findings from the aspirin component of the ongoing Physicians' Health Study," 1988); Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) (Gohagan et al., 2000; Prorok et al., 2000); Post-Menopausal Hormones Supplemental Study to the CCFR (PMH-CCFR) (Newcomb, Zheng, et al., 2007); VITamins And Lifestyle (VITAL) study (White et al., 2004); and Women’s Health Initiative (WHI)("Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group," 1998; Hays et al., 2003).

For detailed description on sample selection and genotyping methods please refer to Peters *et al.* (Peters et al., 2013). Briefly, samples were excluded if the average call rate =< 97%. Unexpected duplicates, unexpected relative pairs or individuals with discrepancies in called and phenotypic sex were removed. Samples that did not cluster with the HapMap2 CEU population were also removed. The SNP exclusion criteria was: call rate =< 98%, Hardy Weinberg equilibrium p-value < 1 × 10-4, minor allele frequency < 5% and SNPs that did not perform consistently across platforms. Samples were phased using Beagle (Browning & Browning, 2007) and imputed to 1000 Genomes phase 1 data using Minimac(Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012). Imputation was done separately for each genotyping platform(Peters et al., 2013).

### References

Belanger, C. F., Hennekens, C. H., Rosner, B., & Speizer, F. E. (1978). The nurses' health study. *Am J Nurs, 78*(6), 1039-1040.

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet, 81*(5), 1084-1097. doi:10.1086/521987

Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. (1998). *Control Clin Trials, 19*(1), 61-109.

Findings from the aspirin component of the ongoing Physicians' Health Study. (1988). *N Engl J Med, 318*(4), 262-264. doi:10.1056/NEJM198801283180431

Gohagan, J. K., Prorok, P. C., Hayes, R. B., Kramer, B. S., Prostate, L. C., & Ovarian Cancer Screening Trial Project, T. (2000). The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. *Control Clin Trials, 21*(6 Suppl), 251S-272S.

Hays, J., Hunt, J. R., Hubbell, F. A., Anderson, G. L., Limacher, M., Allen, C., & Rossouw, J. E. (2003). The Women's Health Initiative recruitment methods and results. *Ann Epidemiol, 13*(9 Suppl), S18-77.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet, 44*(8), 955-959. doi:10.1038/ng.2354

Kury, S., Buecher, B., Robiou-du-Pont, S., Scoul, C., Colman, H., Lelievre, B., . . . Bezieau, S. (2007). The thorough screening of the MUTYH gene in a large French cohort of sporadic colorectal cancers. *Genet Test, 11*(4), 373-379. doi:10.1089/gte.2007.0029

Newcomb, P. A., Baron, J., Cotterchio, M., Gallinger, S., Grove, J., Haile, R., . . . Colon Cancer Family, R. (2007). Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer Epidemiol Biomarkers Prev, 16*(11), 2331-2343. doi:10.1158/1055-9965.EPI-07-0648

Newcomb, P. A., Zheng, Y., Chia, V. M., Morimoto, L. M., Doria-Rose, V. P., Templeton, A., . . . Potter, J. D. (2007). Estrogen plus progestin use, microsatellite instability, and the risk of colorectal cancer in women. *Cancer Res, 67*(15), 7534-7539. doi:10.1158/0008-5472.CAN-06-4275

Peters, U., Jiao, S., Schumacher, F. R., Hutter, C. M., Aragaki, A. K., Baron, J. A., . . . Epidemiology of Colorectal Cancer, C. (2013). Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. *Gastroenterology, 144*(4), 799-807 e724. doi:10.1053/j.gastro.2012.12.020

Prorok, P. C., Andriole, G. L., Bresalier, R. S., Buys, S. S., Chia, D., Crawford, E. D., . . . Ovarian Cancer Screening Trial Project, T. (2000). Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials, 21*(6 Suppl), 273S-309S.

Rimm, E. B., Stampfer, M. J., Colditz, G. A., Chute, C. G., Litin, L. B., & Willett, W. C. (1990). Validity of self-reported waist and hip circumferences in men and women. *Epidemiology, 1*(6), 466-473.

Slattery, M. L., Potter, J., Caan, B., Edwards, S., Coates, A., Ma, K. N., & Berry, T. D. (1997). Energy balance and colon cancer--beyond physical activity. *Cancer Res, 57*(1), 75-80.

Vossen, C. Y., Hoffmeister, M., Chang-Claude, J. C., Rosendaal, F. R., & Brenner, H. (2011). Clotting factor gene polymorphisms and colorectal cancer risk. *J Clin Oncol, 29*(13), 1722-1727. doi:10.1200/JCO.2010.31.8873

White, E., Patterson, R. E., Kristal, A. R., Thornquist, M., King, I., Shattuck, A. L., . . . Potter, J. D. (2004). VITamins And Lifestyle cohort study: study design and characteristics of supplement users. *Am J Epidemiol, 159*(1), 83-93.

## Supplementary text 2: GECCO funding and acknowledgements

### Funding

*GECCO*: National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088; R01 CA059045). *ASTERISK*: a Hospital Clinical Research Program (PHRC) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC). *CCFR*: This work was supported by grant UM1 CA167551 from the National Cancer Institute and through cooperative agreements with the following CCFR centers: Australasian Colorectal Cancer Family Registry (U01 CA074778 and U01/U24 CA097735), Mayo Clinic Cooperative Family Registry for Colon Cancer Studies (U01/U24 CA074800), Ontario Familial Colorectal Cancer Registry (U01/U24 CA074783), Seattle Colorectal Cancer Family Registry (U01/U24 CA074794), University of Hawaii Colorectal Cancer Family Registry (U01/U24 CA074806), USC Consortium Colorectal Cancer Family Registry U01/U24 CA074799). The Colon CFR GWAS was supported by funding from the National Cancer Institute, National Institutes of Health (U01 CA122839 and R01 CA143237 to Graham Casey). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the CCFR. *DACHS*: German Research Council (Deutsche Forschungsgemeinschaft, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4 and CH 117/1-1), and the German Federal Ministry of Education and Research (01KH0404 and 01ER0814). *DALS*: National Institutes of Health (R01 CA48998 to M. L. Slattery);

*HPFS* is supported by the National Institutes of Health (P01 CA 055075, UM1 CA167552, R01 137178, R01 CA151993 and P50 CA127003), *NHS* by the National Institutes of Health (UM1 CA186107, R01 CA137178, P01 CA87969, R01 CA151993 and P50 CA127003,) and *PHS* by the National Institutes of Health (R01 CA042182). *PLCO*: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Additionally, a subset of control samples were genotyped as part of the Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer GWAS (Yeager, M et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 2007 May;39(5):645-9), Colon CGEMS pancreatic cancer scan (PanScan) (Amundadottir, L et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet*. 2009 Sep;41(9):986-90, and Petersen, GM et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet*. 2010 Mar;42(3):224-8), and the Lung Cancer and Smoking study (Landi MT, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet.* 2009 Nov;85(5):679-91). The prostate and PanScan study datasets were accessed with appropriate approval through the dbGaP online resource (http://cgems.cancer.gov/data/) accession numbers phs000207.v1.p1 and phs000206.v3.p2, respectively, and the lung datasets were accessed from the dbGaP website (http://www.ncbi.nlm.nih.gov/gap) through accession number phs000093.v2.p2. Funding for the Lung Cancer and Smoking study was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438. For the lung study, the GENEVA Coordinating Center provided assistance with genotype cleaning and general study coordination, and the Johns Hopkins University Center for Inherited Disease Research conducted genotyping. *PMH*: National Institutes of Health (R01 CA076366 to P.A. Newcomb). *VITAL*: National Institutes of Health (K05 CA154337). *WHI*: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

Acknowledgements

GECCO: The authors would like to thank all those at the GECCO Coordinating Center for helping bring together the data and people that made this project possible. The authors acknowledge Dave Duggan and team members at TGEN (Translational Genomics Research Institute), the Broad Institute, and the Génome Québec Innovation Center for genotyping DNA samples of cases and controls, and for scientific input for GECCO. ASTERISK: We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students. DACHS: We thank all participants and cooperating clinicians, and Ute Handte-Daub, Utz Benscheid, Muhabbet Celik and Ursula Eilber for excellent technical assistance. HPFS, NHS and PHS: We would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NHS, HPFS, and PHS under the supervision of Dr. Immaculata Devivo and Dr. David Hunter, Qin (Carolyn) Guo and Lixue Zhu who assisted in programming for NHS and HPFS, and Haiyan Zhang who assisted in programming for the PHS. We would like to thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-Up Study, for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data. PLCO: The authors thank Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff or the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Mr. Tom Riley and staff, Information Management Services, Inc., Ms. Barbara O’Brien and staff, Westat, Inc., and Drs. Bill Kopp and staff, SAIC-Frederick. Most importantly, we acknowledge the study participants for their contributions to making this study possible. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI. PMH: The authors would like to thank the study participants and staff of the Hormones and Colon Cancer study. WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf