

# Heritability Estimation using a Regularized Regression Approach (HERRA): Applicable to Continuous, Dichotomous or Survival Outcome

Malka Gorfine<sup>1,\*</sup>, Sonja I Berndt<sup>2</sup>, Jenny Chang-Claude<sup>3</sup>, Michael Hoffmeister<sup>4</sup>, Loic Le Marchand<sup>5</sup>, John Potter<sup>6</sup>, Martha L Slattery<sup>7</sup>, Nir Keret<sup>1</sup>, Ulrike Peters<sup>6</sup>, Li Hsu<sup>6,\*</sup>

**1 Department of Statistics and Operation Research, Tel Aviv University, Tel Aviv, Israel**

**2 Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health**

**3 Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany**

**4 Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany**

**5 Epidemiology Program, University of Hawaii Cancer Center**

**6 Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA**

**7 Department of Internal Medicine, University of Utah Health Sciences Center**

**\* Correspondence: [gorfine@post.tau.ac.il](mailto:gorfine@post.tau.ac.il), [lih@fredhutch.org](mailto:lih@fredhutch.org)**

## S9 Text: Details on quality assurance and quality control of the colorectal cancer GWAS dataset

**Table S1. Quality assurance and quality control**

Study	Sample Exclusions <sup>1</sup>	Duplicate Concordance	Sample Call Rate	SNP Exclusions <sup>2</sup>	SNPs Passing QC	SNP Call Rate
	N	%	Mean	N	N	Mean
Colo2&3	42	100.0%	99.90%	30,446	252,176	99.95%
DACHS1	227	99.90%	99.93%	33,588	252,208	99.90%
DALS2	5	100.0%	99.94%	32,885	250,320	99.94%
MEC	61	100.0%	99.97%	34,494	259,364	99.96%
PLCO2	31	99.90%	99.80%	38,655	253,702	99.90%
VITAL	163	99.90%	99.81%	36,805	243,625	99.89%
WHI2	42	100.0%	99.96%	30,446	251,707	99.96%

<sup>1</sup>Samples were excluded for duplicate errors, gender errors, a call rate < 97%, outlying the CEU HapMap2 cluster in PCA, heterozygosity, and concordance with non-duplicates samples > 65%.

<sup>2</sup>Directly genotyped SNPs were excluded for a call rate < 98%, Hardy Weinberg Equilibrium (HWE) < 10<sup>-4</sup>, and minor allele frequency (MAF < 1%).