The American Journal of Human Genetics, Volume *102*

Supplemental Data

Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative

Lars G. Fritsche, Stephen B. Gruber, Zhenke Wu, Ellen M. Schmidt, Matthew Zawistowski, Stephanie E. Moser, Victoria M. Blanc, Chad M. Brummett, Sachin Kheterpal, Gonçalo R. Abecasis, and Bhramar Mukherjee

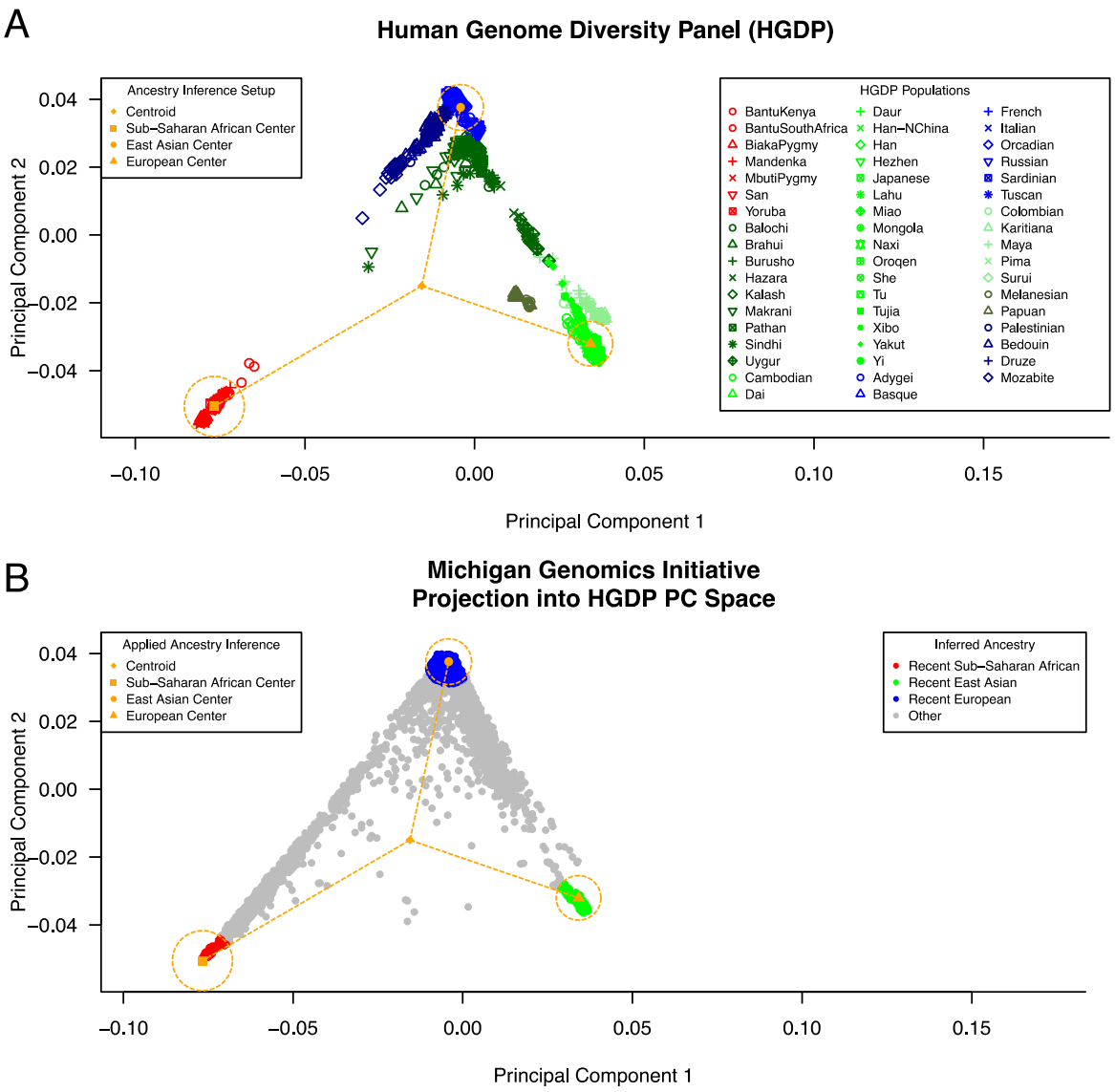
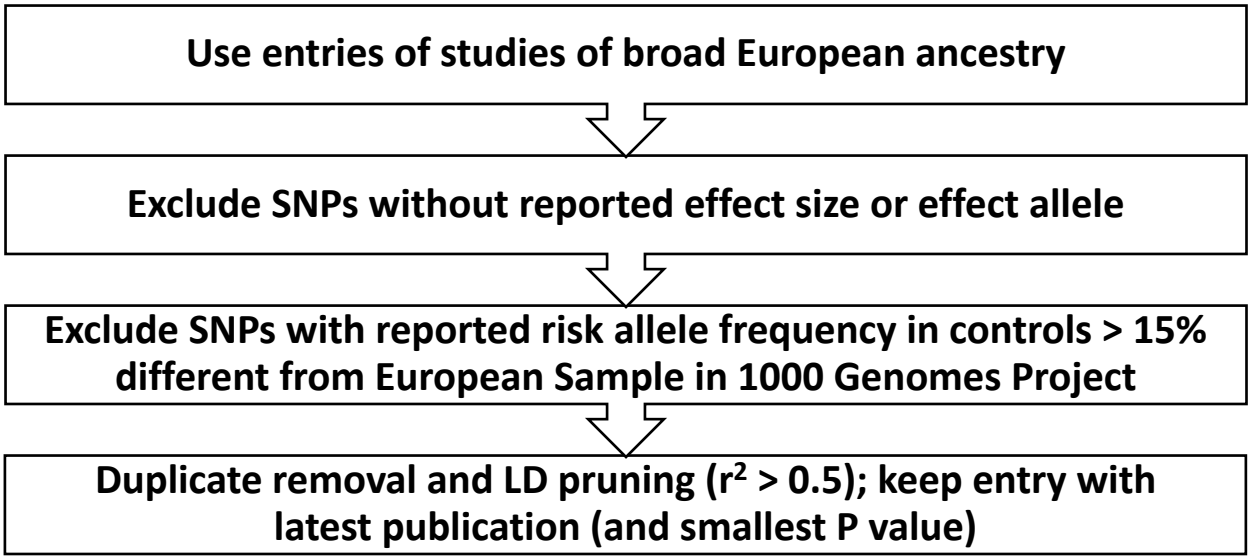


Figure S1 Principal Component Analysis. The Human Genome Diversity Panel (HGDP) (**A**) served as reference while the MGI samples (**B**) were projected into its PC space. The applied inference of recent Sub-Saharan African (sAFR, red), East Asian (eASN, green) and European (EUR, blue) ancestry based on the HGDP's first two principal components is shown in orange: the centers of the sAFR, eASN and EUR HGDP populations defined a centroid, while samples that fell within a circle around these centers with a radius of 1/8 of the distant between the corresponding center and the centroid were defined as recent sAFR, eASN or EUR ancestry while the remaining samples (grey) remained undefined.

A



B

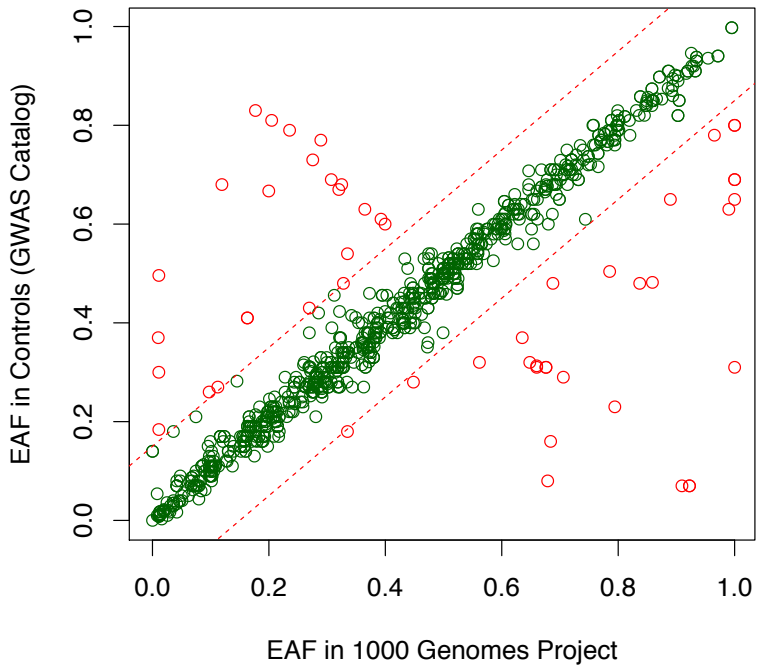


Figure S2 Quality control of NHGRI-EBI GWAS Catalog entries. (**A**) Flowchart of step-wise filtering of catalog entries for each cancer type. (**B**) Effect allele frequency (EAF) comparison of 22 cancer traits, catalog SNPs with 1000 Genomes Project variants with <= 15% and > 15 % frequency difference (dotted lines) are indicated in green (N = 838) and red (N = 53), respectively. Comparison was limited to entries with complete information (effect alleles, EAF, and presence in 1000 Genomes Project) and to samples of European ancestry.

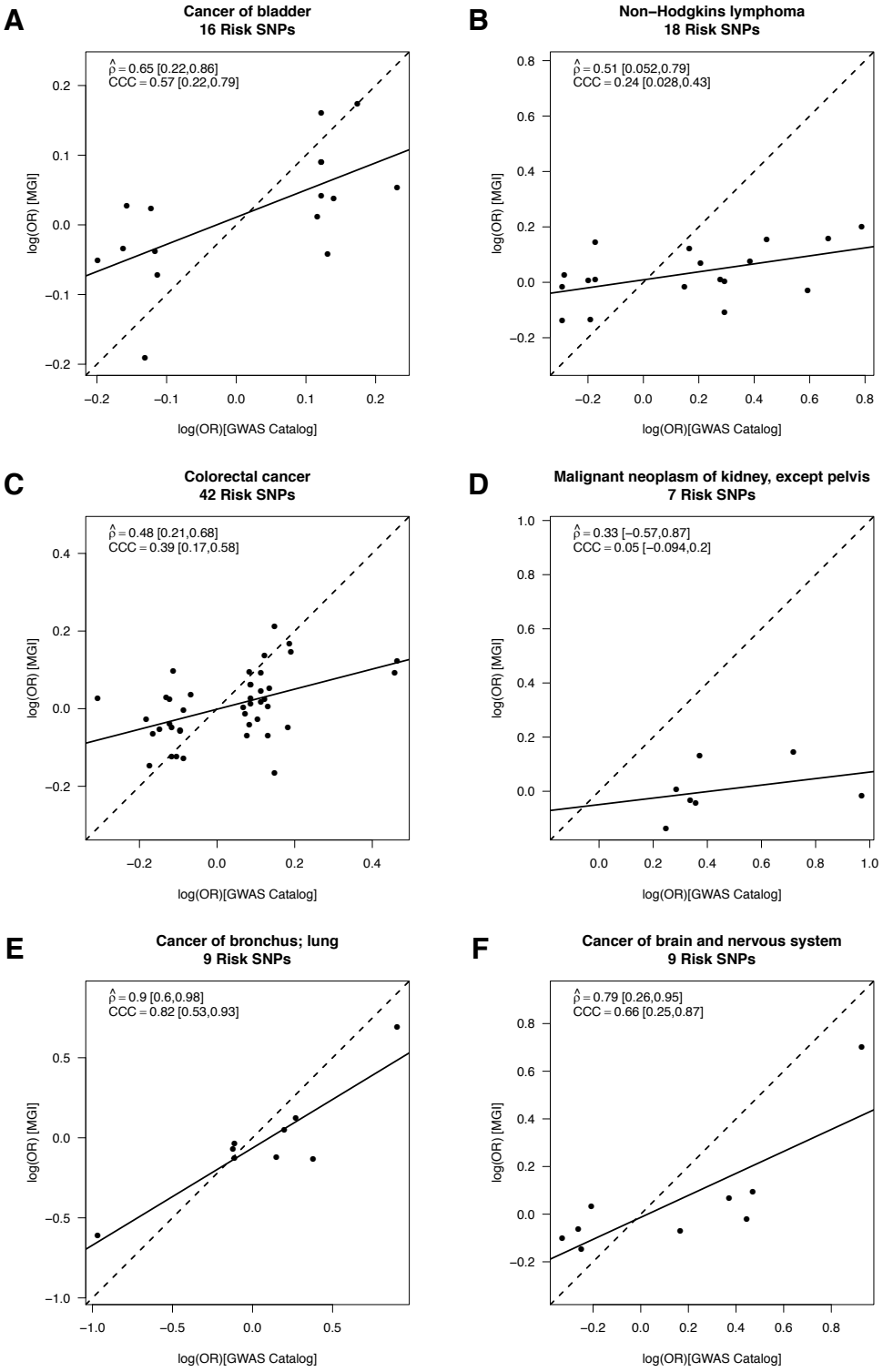


Figure S3. Calibration of association parameters between the MGI-GWAS and NHGRI-EBI GWAS derived effect estimates [log(OR)] for (**A**) cancer of bladder, (**B**) Non-Hodgkins lymphoma, (**C**) colorectal cancer, (**D**) Malignant neoplasm of kidney, except pelvis, (**E**) Cancer of bronchus; lung, and (**F**) Cancer of brain and nervous system. The agreement of two sets of SNP-specific beta coefficients (non-reference allele is the effect allele) and their Pearson Correlation (Coefficient ρ and *P*) are shown; dashed line: perfect concordance; solid line: fitted line.

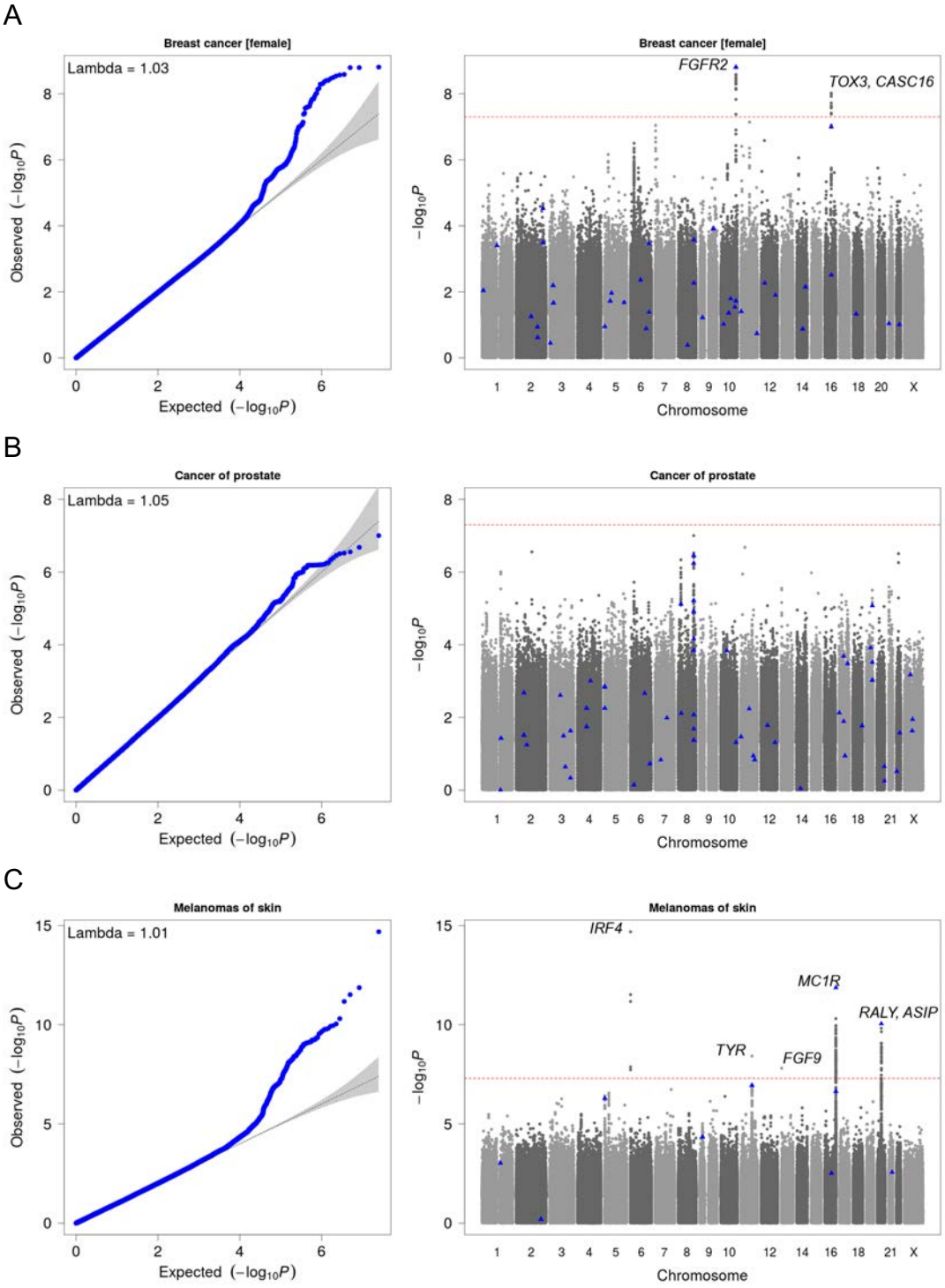
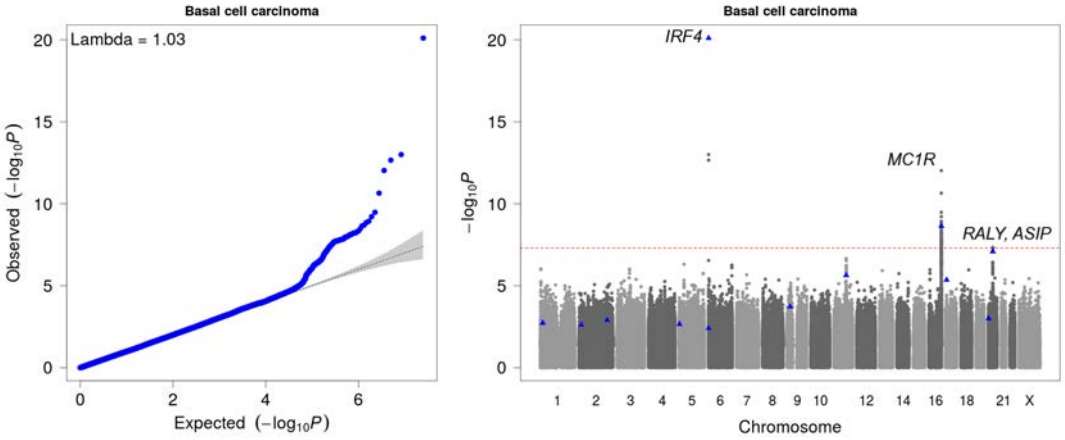
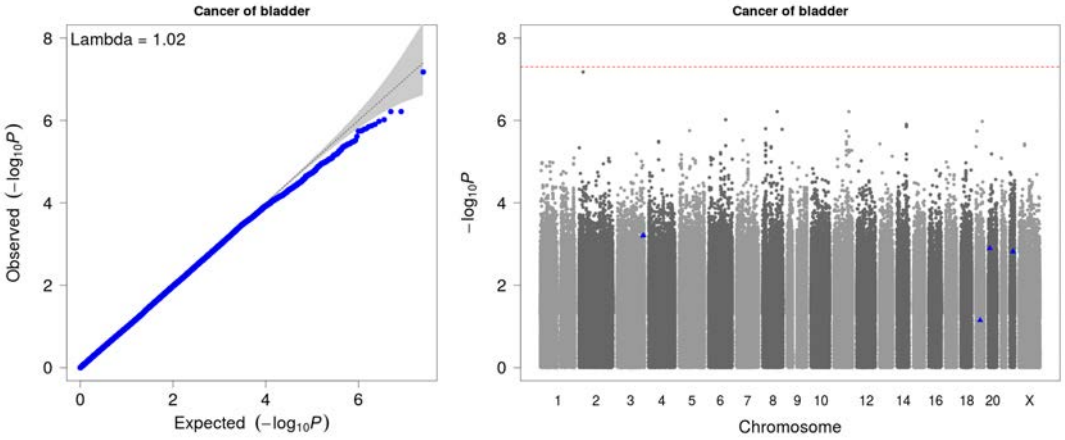


Figure S4. Genome-wide association studies of the 9 studied cancer traits revealed genome-wide significant associations: (**A**) Breast cancer [female], (**B**) cancer of prostate, (**C**) melanomas of skin, (**D**) basal cell carcinoma, (**E**) cancer of bladder, (**F**) squamous cell carcinoma, (**G**) cancer of bronchus; lung, (**H**) thyroid cancer, (**I**) cancer of brain and nervous system. Left: QQ plot; right: Manhattan plot, loci with genome-wide significant hits indicated with their previously reported gene names. The selected GWAS catalog variants (**Table S5**) of the corresponding trait are indicated by blue diamonds. Red line: genome-wide significance (p-value < 5 x 10^{-8}).

D



E



F

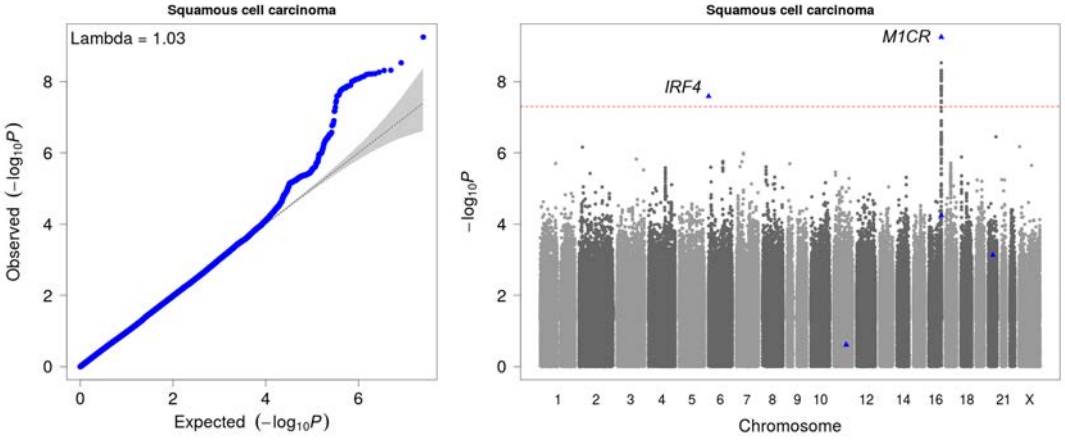
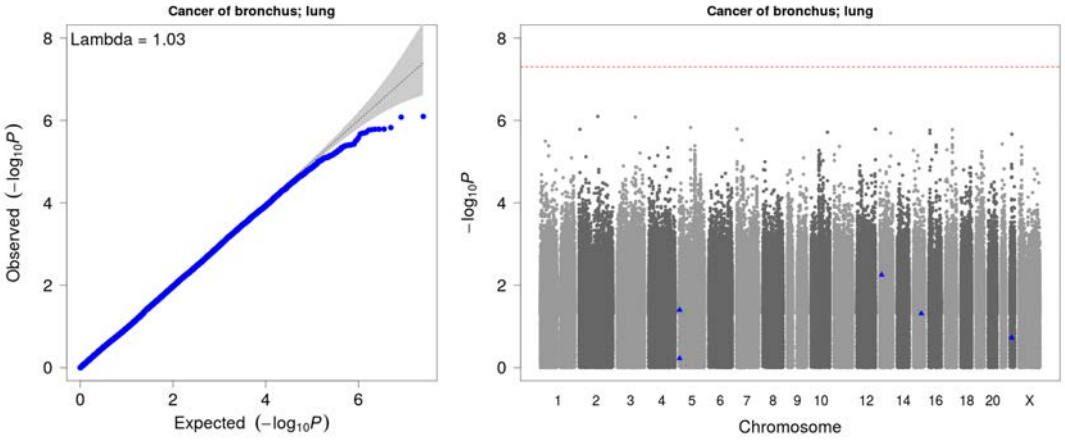
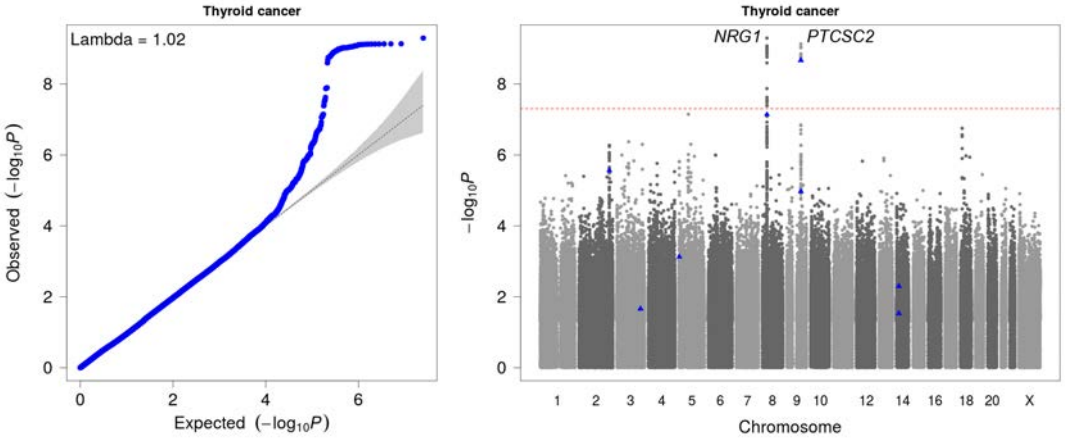


Figure 4 continued

G



H



I

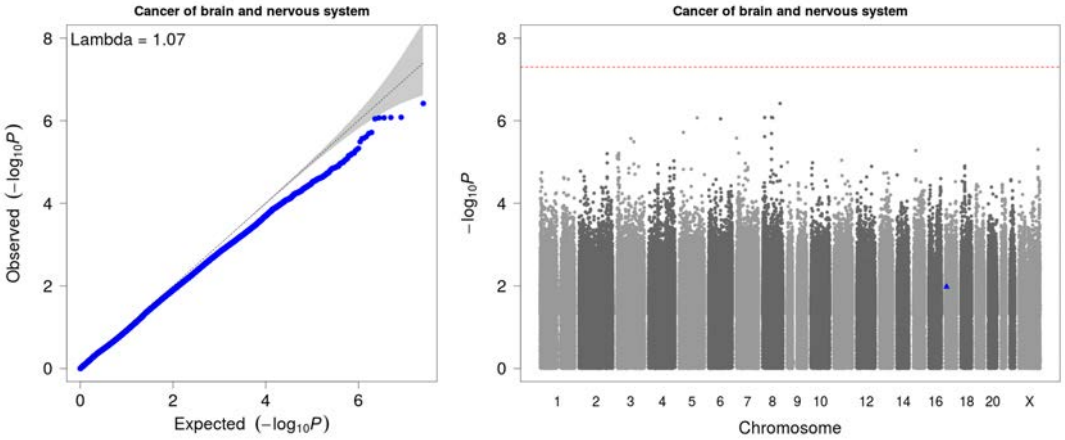


Figure 4 continued

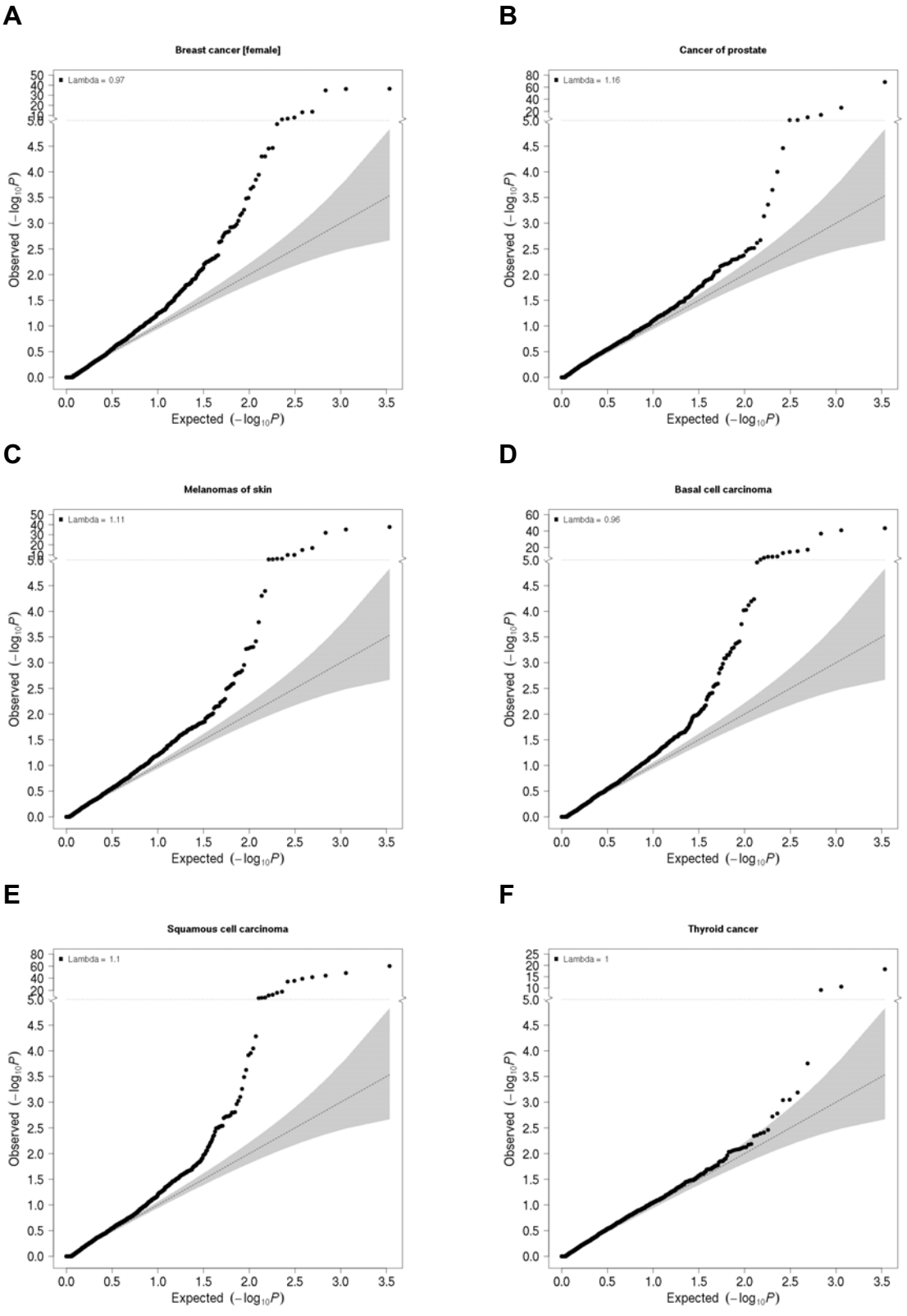


Figure S5 PRS PheWAS QQ plots (**A**) Breast cancer [female], (**B**) cancer of prostate, (**C**) melanomas of skin, (**D**) basal cell carcinoma, (**E**) squamous cell carcinoma, and (**F**) thyroid cancer.

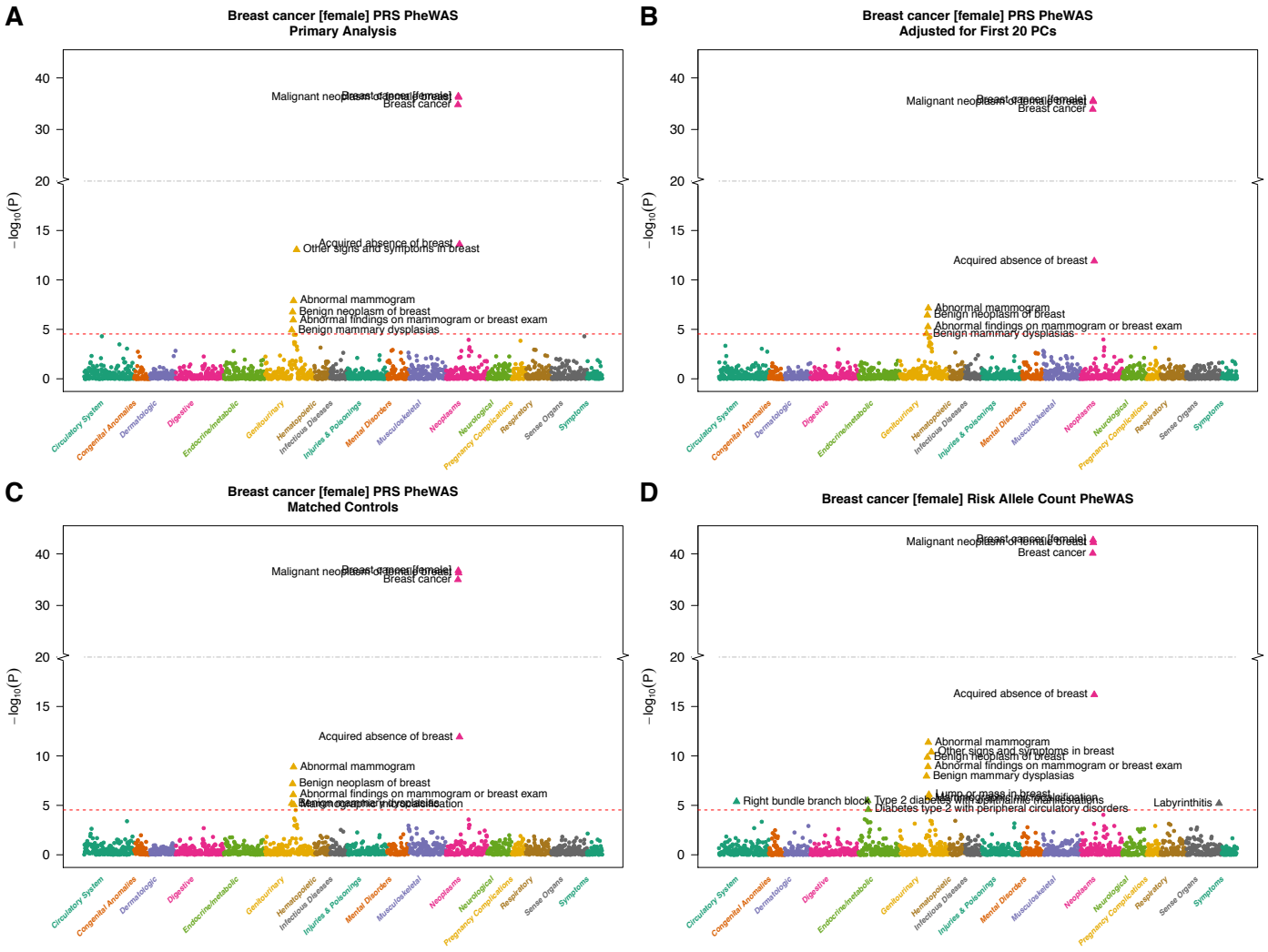


Figure S6 Breast cancer PRS PheWAS plots: primary analysis (**A**) and three sensitivity analyses: adjusted for 20 PCs (**B**), using matched controls (**C**) and using risk allele counts instead of PRS using effect sizes (**D**).

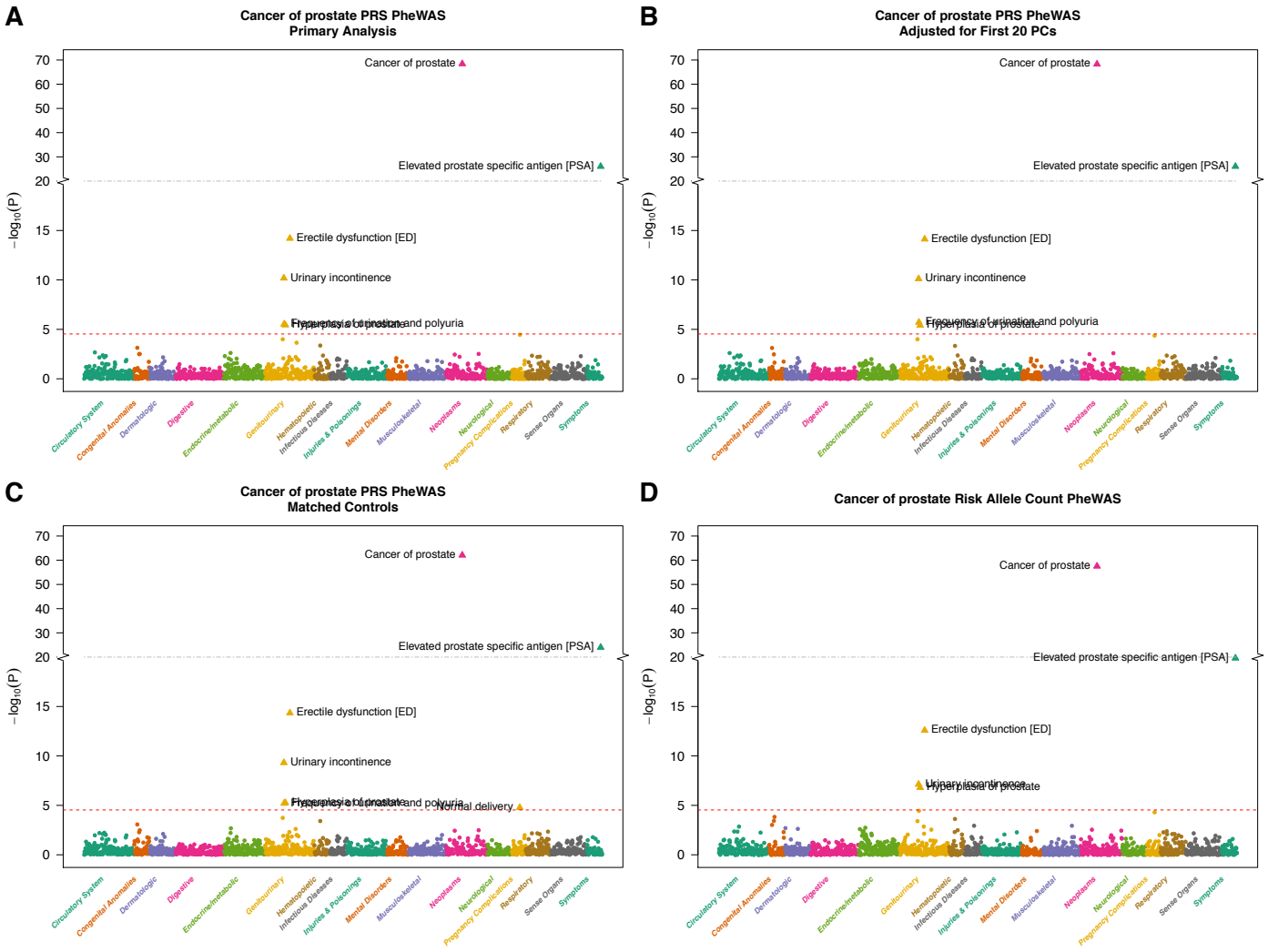


Figure S7 Cancer of prostate PRS PheWAS plots: primary analysis (**A**) and three sensitivity analyses: adjusted for 20 PCs (**B**), using matched controls (**C**) and using risk allele counts instead of PRS using effect sizes (**D**).

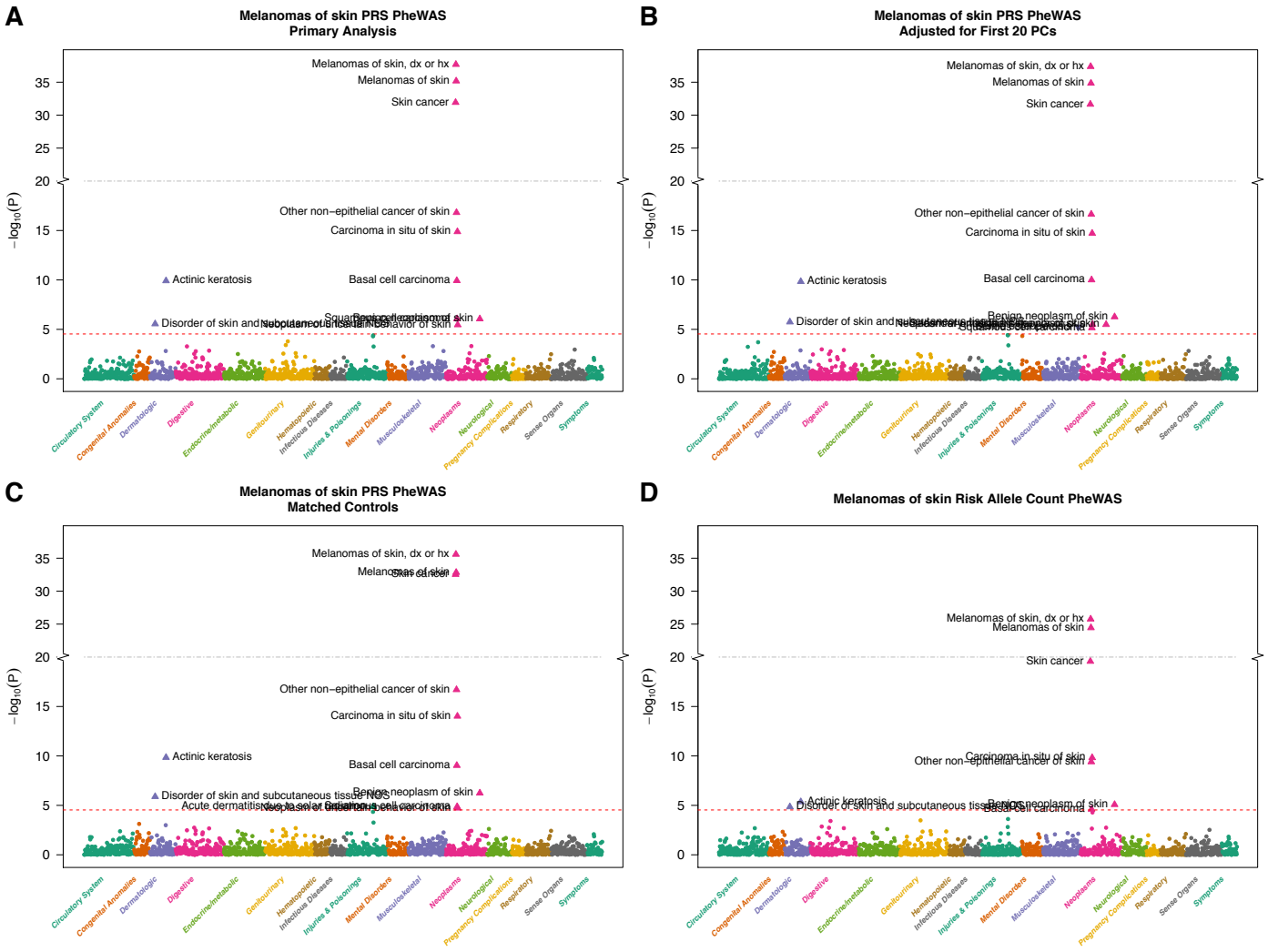


Figure S8 Melanoma PRS PheWAS plots: primary analysis (**A**) and three sensitivity analyses: adjusted for 20 PCs (**B**), using matched controls (**C**) and using risk allele counts instead of PRS using effect sizes (**D**).

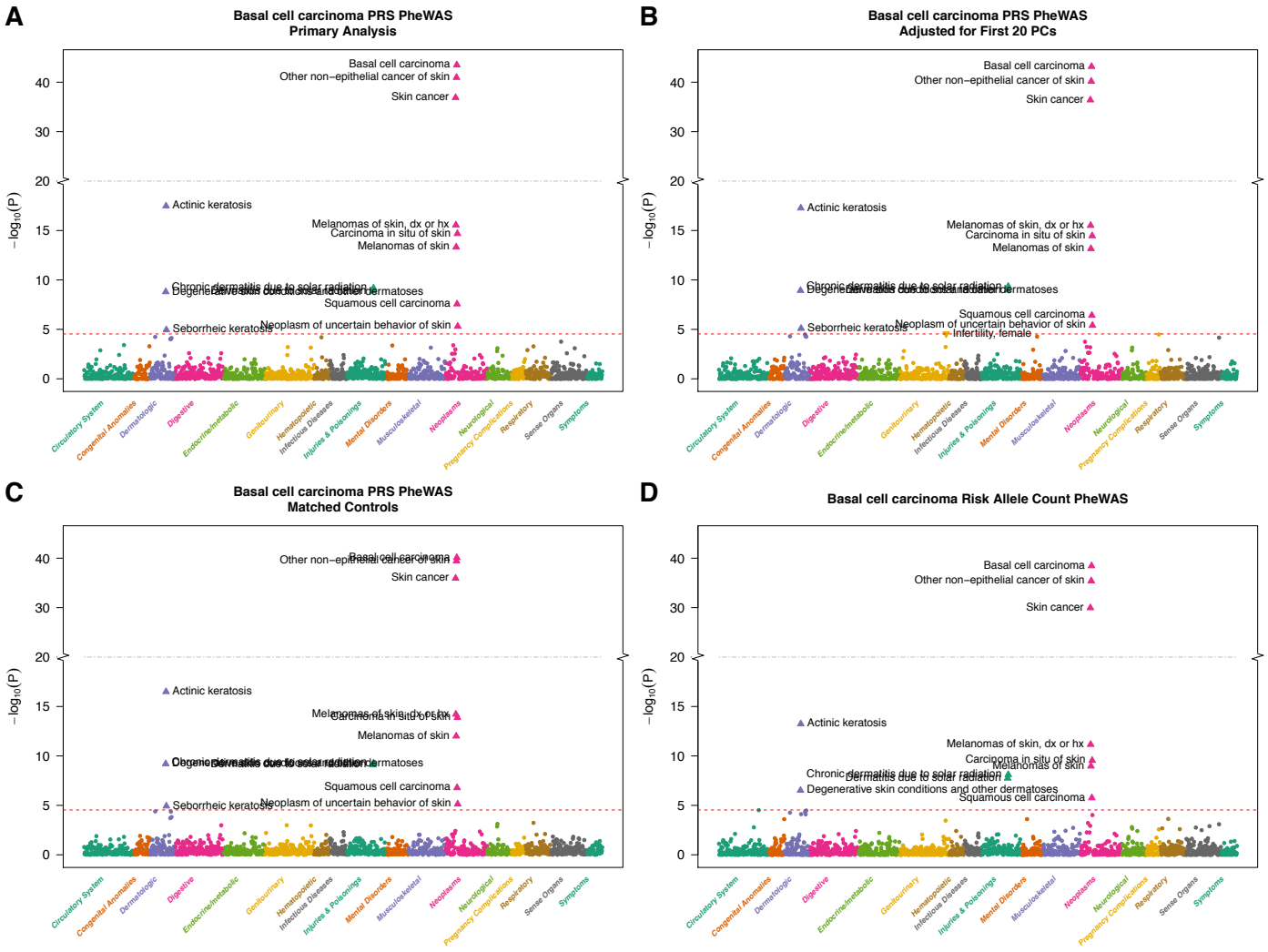


Figure S9 Basal cell carcinoma PRS PheWAS plots: primary analysis (**A**) and three sensitivity analyses: adjusted for 20 PCs (**B**), using matched controls (**C**) and using risk allele counts instead of PRS using effect sizes (**D**).

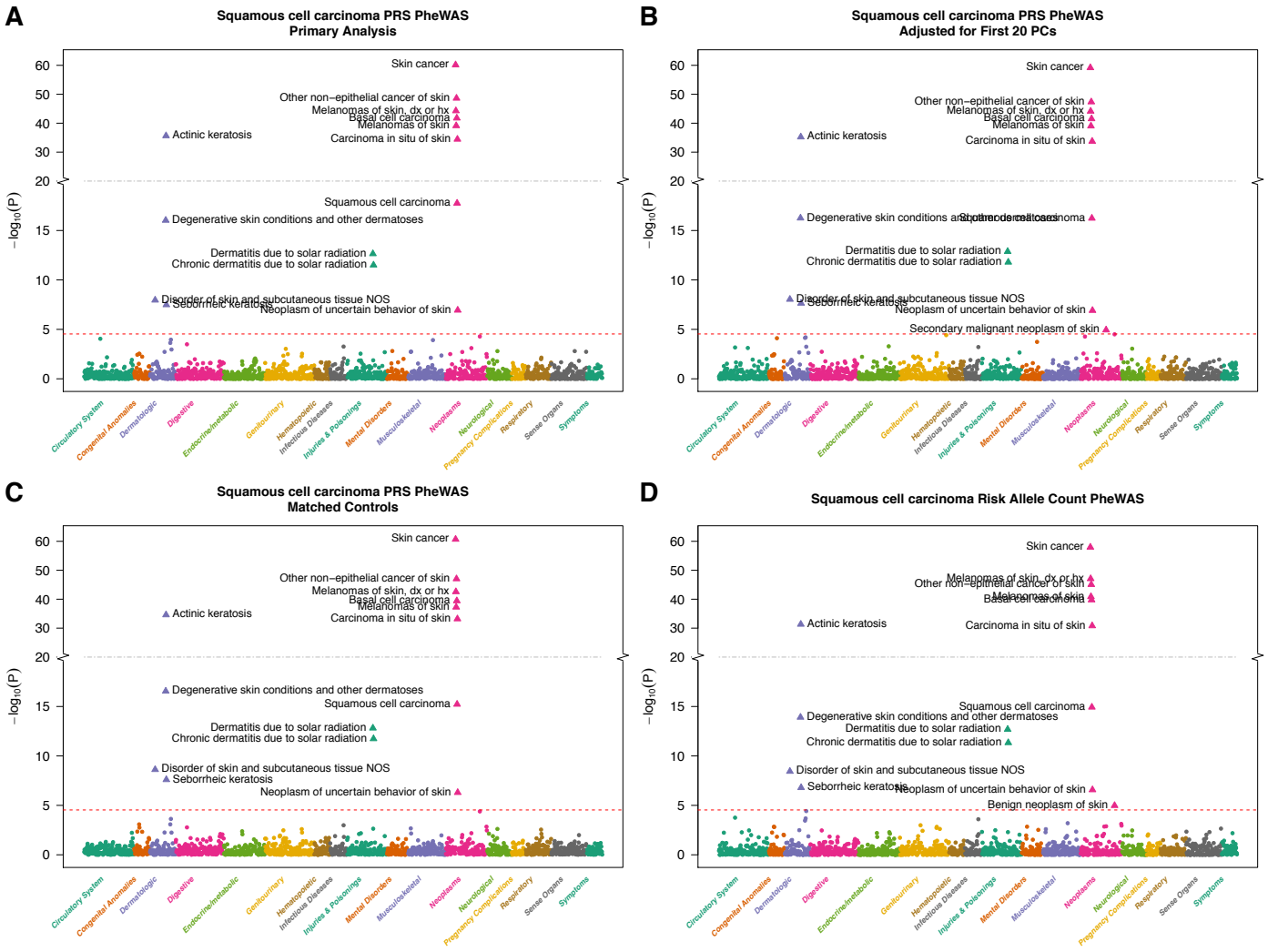


Figure S10 Squamous cell carcinoma PRS PheWAS plots: primary analysis (**A**) and three sensitivity analyses: adjusted for 20 PCs (**B**), using matched controls (**C**) and using risk allele counts instead of PRS using effect sizes (**D**).

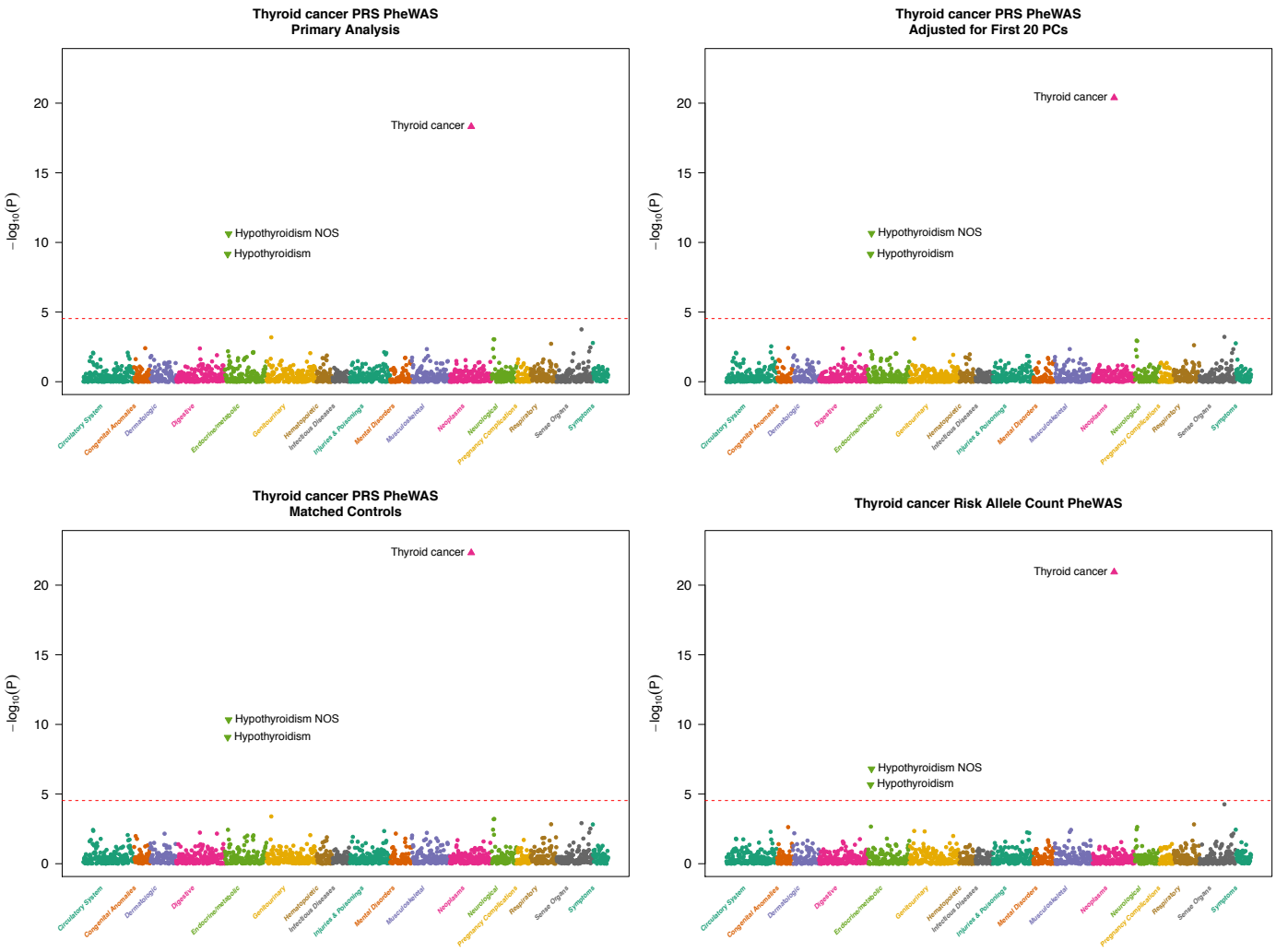


Figure S11 Thyroid cancer PRS PheWAS plots: primary analysis (**A**) and three sensitivity analyses: adjusted for 20 PCs (**B**), using matched controls (**C**) and using risk allele counts instead of PRS using effect sizes (**D**).

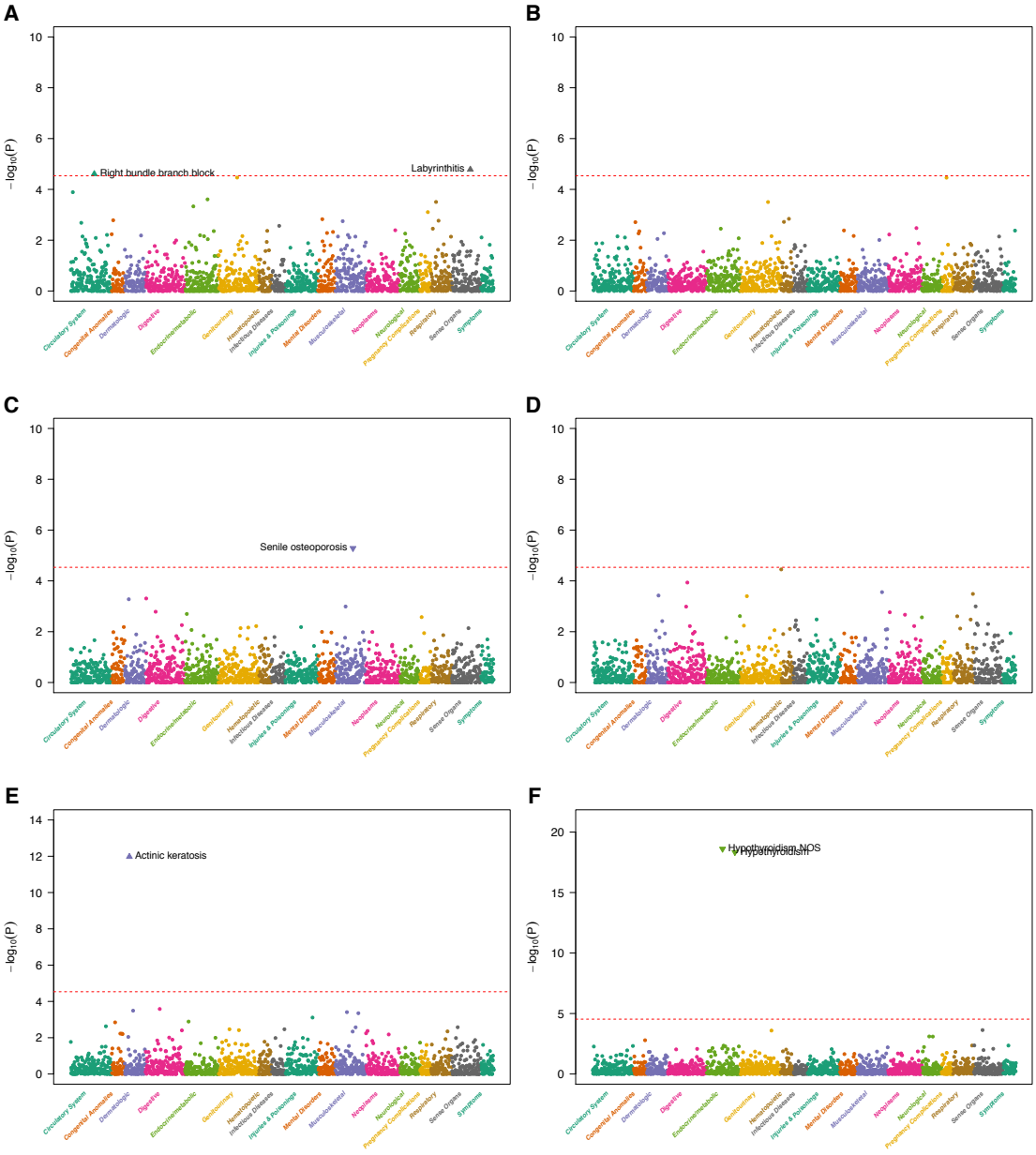


Figure S12. PRS "Exclusion" PheWAS plots: (**A**) female breast cancer PRS after excluding breast cancer cases, (**B**) cancer of prostate PRS after excluding prostate cancer cases, (**C**) melanoma PRS after excluding skin cancer cases, (**D**) basal cell carcinoma PRS after excluding skin cancer cases, (**E**) squamous cell carcinoma PRS after excluding skin cancer cases, and (**F**) thyroid cancer PRS after excluding thyroid cancer cases. 1,711 traits are grouped into 16 color-coded categories as shown on the horizontal axis; the p-values for testing the associations of PRS with the traits are minus log-base-10-transformed and shown on the vertical axis. Triangles indicate phenome-wide significant associations with their effect orientation (up-pointing = risk increasing; down-pointing = risk decreasing). PRS upon multiplicity adjustment (see ***Methods***). The solid horizontal line for P=2.9 x10^{-5} cut-off.

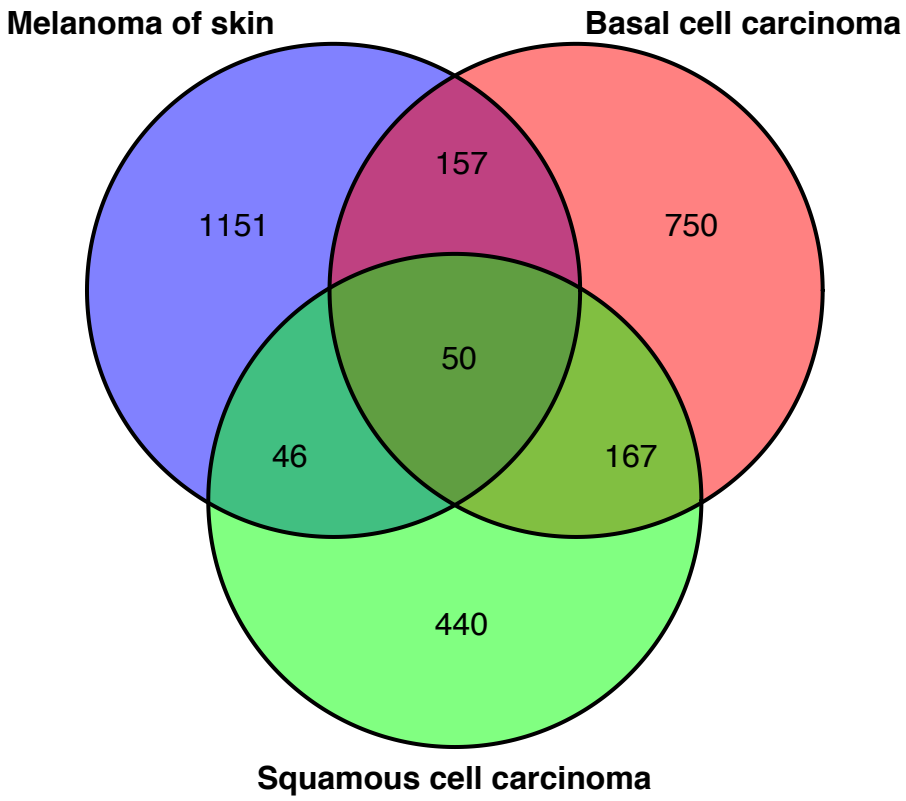


Figure S13. Sample overlap between skin cancer subtypes: Venn diagram showing the overlapping numbers of individuals between melanoma, basal cell carcinoma, or squamous cell carcinoma

Cancer Trait	N Cases	N Unmatched Controls	Continous PRS *P**	Maximal Number of Matched Controls Per Case Average Control to Case Ratio N Matched Controls Continous PRS P									
				1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Breast cancer [female]	1,827	11,073	3.59×10^{-37}	1 1,827 1.4×10^{-24}	2 3,653 3.4×10^{-29}	3 5,478 1.9×10^{-33}	4 7,288 1.7×10^{-36}	4.8 8,822 1.0×10^{-36}	5.1 9,239 7.8×10^{-37}	5.2 9,522 4.9×10^{-37}	5.3 9,717 1.4×10^{-37}	5.4 9,857 1.6×10^{-37}	5.5 9,960 1.8×10^{-37}
Cancer of prostate	1,425	9,793	3.78×10^{-69}	1 1,425 1.5×10^{-34}	2 2,849 4.0×10^{-47}	2.9 4,163 1.2×10^{-52}	3.5 5,037 6.0×10^{-57}	3.9 5,601 1.5×10^{-59}	4.2 6,009 4.6×10^{-61}	4.4 6,326 1.8×10^{-61}	4.6 6,577 1.4×10^{-62}	4.8 6,768 6.3×10^{-63}	4.9 6,932 7.5×10^{-63}
Melanomas of skin	1,404	23,798	6.66×10^{-36}	1 1,403 8.9×10^{-19}	2 2,802 7.1×10^{-23}	3 4,182 2.8×10^{-28}	3.9 5,494 1.3×10^{-31}	4.8 6,778 2.1×10^{-31}	5.7 8,024 2.4×10^{-31}	6.6 9,210 1.3×10^{-32}	7.4 10,339 3.0×10^{-33}	8.1 11,426 2.6×10^{-33}	8.9 12,487 1.5×10^{-33}
Basal cell carcinoma	1,124	23,798	3.30×10^{-44}	1 1,119 8.5×10^{-18}	2 2,229 2.3×10^{-28}	3 3,318 1.1×10^{-33}	3.9 4,393 4.7×10^{-34}	4.9 5,448 1.4×10^{-36}	5.7 6,434 8.7×10^{-39}	6.6 7,371 5.3×10^{-39}	7.3 8,233 8.7×10^{-40}	8.0 9,039 5.6×10^{-40}	8.7 9,805 7.7×10^{-41}
Cancer of bladder	978	26,748	4.91×10^{-6}	1 978 0.014	2 1,956 0.0011	3 2,933 0.00047	4 3,910 0.00027	5 4,871 1.8×10^{-5}	5.9 5,783 9.1×10^{-6}	6.8 6,651 5.8×10^{-6}	7.6 7,433 6.5×10^{-6}	8.4 8,164 2.6×10^{-6}	9.1 8,851 1.5×10^{-6}
Squamous cell carcinoma	703	23,798	1.75×10^{-18}	0.99 697 2.0×10^{-9}	1.97 1,382 1×10^{-11}	2.9 2,060 3.3×10^{-13}	3.9 2,731 7.8×10^{-15}	4.8 3,372 2.8×10^{-15}	5.7 3,987 1.1×10^{-15}	6.5 4,599 6.6×10^{-16}	7.4 5,206 2.1×10^{-15}	8.2 5,793 1.3×10^{-15}	9.1 6,368 6.0×10^{-16}
Thyroid cancer	472	26,692	4.80×10^{-19}	1 472 1.2×10^{-11}	2 944 1.1×10^{-15}	3 1,416 4.1×10^{-18}	4 1,887 7.9×10^{-18}	5 2,358 9.7×10^{-20}	6 2,829 3.0×10^{-20}	7 3,299 1.3×10^{-20}	8 3,769 1.8×10^{-22}	9 4,239 1.9×10^{-22}	10 4,709 4.8×10^{-23}

* P value from Firth's penalized-likelihood logistic regression

Table S6. Case Control Matching. Exploration of different maximal matched control numbers for six cancer PRS that reached phenome-wide significant associations (P <= 0.05/1711) with their primary cancer trait in the corresponding unmatched case control study. Obtained total number of controls as well as association of each cancer with their continuous PRS are shown. P-values are obtained by fitting conditional logistic regression.

Cancer Trait	N Cases	N Controls	N Risk SNPs used for PRS	Replicated at P < (0.05 / N risk SNPs)*
Breast cancer [female]	1,827	11,073	78	8
Cancer of prostate	1,425	9,793	93	14
Melanomas of skin	1,404	23,798	16	9
Basal cell carcinoma	1,124	23,798	19	11
Cancer of bladder	978	26,748	16	3
Squamous cell carcinoma	703	23,798	5	4
Cancer of bronchus; lung	570	27,596	9	0
Thyroid cancer	472	26,692	8	6
Cancer of brain and nervous system	321	27,069	9	0

* with consistent effect orientation as discovery GWAS study

Table S7 Replication of discovery GWAS findings in MGI

Cancer Trait	N Cases	N Controls	Reported GWAS Catalog SNPs in Top 1% of MGI GWAS Results
Breast cancer [female]	1,827	11,073	15 of 78 SNPs with P < 0.0089
Cancer of prostate	1,425	9,793	27 of 93 SNPs with P < 0.0083
Melanomas of skin	1,404	23,798	9 of 16 SNPs with P < 0.003
Basal cell carcinoma	1,124	23,798	12 of 19 SNPs with P < 0.004
Cancer of bladder	978	26,748	3 of 16 SNPs with P < 0.0015
Squamous cell carcinoma	703	23,798	4 of 5 SNPs with P < 0.00074
Cancer of bronchus; lung	570	27,596	1 of 9 SNPs with P < 0.0056
Thyroid cancer	472	26,692	6 of 8 SNPs with P < 0.005
Cancer of brain and nervous system	321	27,069	1 of 9 SNPs with P < 0.01

Table S8. Ranking of NHGRI-EBI GWAS Catalog variants in the GWAS analyses of the MGI data set (> 12.5 million variants with MAF >= 0.1%)