

Supplementary Methods

NHS & HPFS set:

1) Postmenopausal invasive breast cancer case-control study nested within the NHS

(CGEMS): Eligible cases in this study consisted of women with pathologically confirmed incident breast cancer from the sub-cohort who gave a blood specimen. Cases with a diagnosis after blood collection up to June 1, 2000 with no previously diagnosed cancer except for non-melanoma skin cancer were included. One control for each case was randomly selected among women who gave a blood sample and were free of diagnosed cancer (excluding non-melanoma skin cancer) up to and including the interval in which the case was diagnosed. Controls were matched to cases on year of birth, menopausal status, recent post-menopausal hormone (PMH) use, month of blood return, time of day of blood collection, and fasting status at blood draw (1).

2) Type 2 diabetes (T2D) case-control study nested within the NHS and HPFS

(T2D_NHS and T2D_HPFS): Diabetes cases were defined as self-reported incident diabetes confirmed by a validated supplementary questionnaire. For cases before 1998, diagnosis was made using criteria consistent with those proposed by the National Diabetes Data Group (NDDG). For cases during the 1998 and 2000 cycles, the American Diabetes Association's diagnostic criteria were used for the diagnosis of diabetes cases. The non-diabetic control subjects were matched to cases on age, month and year of blood draw, and fasting status (2).

3) Coronary heart disease (CHD) case-control study nested within the NHS and

HPFS (CHD_NHS and CHD_HPFS): In both the NHS and HPFS, participants who had reported an incident CHD event on the follow-up questionnaire were contacted for confirmation, and permission to review medical records was requested. Medical records for deceased participants were also sought for deaths that were identified by families and postal officials and through the National Death Index. Physicians blinded to the participant's questionnaire reports reviewed all medical records. Fatal CHD cases were identified primarily through review of medical records, as previously described (3).

Among participants who provided blood samples and who were without cardiovascular disease or cancer at blood draw, incident CHD cases occurring after blood draw were selected as cases. Controls were selected, in a 2:1 ratio matched to cases on age, smoking, and month of blood return.

4) Kidney stone study nested within the NHS and HPFS (KS_NHS_HPFS):

Participants from the KS_NHS_HPFS were individuals who performed a 24-hour urine collection; two-thirds had a history of incident nephrolithiasis. Details regarding the urine collection (4) and the confirmation of kidney stone disease were published previously (5).

The participants reported on the interval diagnosis of kidney stones every 2 years. Any study participant who reported a new kidney stone was sent an additional questionnaire to determine the date of occurrence and the symptoms produced by the stone.

5) Prostate cancer study nested within the HPFS (ADVCAP_HPFS): Prostate cancer cases are matched to controls on birth year (+/-1) and ethnicity. Controls are selected

from those who are cancer-free at the time of the case's diagnosis, and had a prostate-specific antigen test after the date of blood draw.

6) Glaucoma

Primary open angle glaucoma cases were defined as self-reported incident glaucoma confirmed to be primary open angle glaucoma with reproducible visual field loss confirmed with standardized visual field tests and validated supplementary questionnaires or medical records from diagnosing eye care providers. Cases were all Caucasian and matched to controls (roughly 1:1) on cohort, age, type of sample (blood or cheek cell).

7) PANSCAN

Subjects were followed prospectively with repeated assessments of lifestyle factors and ascertainment of cancer diagnoses. Each cohort study selected participants with blood or buccal cells collected before cancer diagnosis. One control was selected per case within each cohort. Controls were matched on year of birth (± 5 years), gender, self-reported race/ethnicity, and source of DNA (peripheral blood or buccal cells). Controls were alive without pancreatic cancer on the incidence date of the matched case (6).

Quality control (QC) procedures for eight GWAS of the NHS & HPFS set:

1) CGEMS: Detailed QC procedures were provided previously (1). Briefly, a total of 555,352 SNP genotype assays were attempted on the 2,494 DNA samples using the Illumina HumanHap550 chip. Whenever the completion rate for a sample was below 90%, the sample was assayed a second time. Samples that did not meet the 90%

completion threshold after a second attempt were excluded from further analysis. We excluded 59 samples from NHS (30 cases and 29 controls) from further analysis based on these criteria, which left 2,435 DNAs for the subsequent analyses.

A total of 8,706 SNPs (~1.57% overall) failed to provide accurate genotype results owing to either a lack of calls or low call rates (<90%). We performed further quality control analysis on the remaining 546,646 SNPs. The genotyping of the SNPs with high call rate on the 2,412 NHS DNAs with high completion rate generated 1.27 billion genotype calls. For this set of SNPs and samples, the percentage of missing data was <1%. The genotype concordance rate for SNP assays was evaluated using the 93 pairs of known duplicated DNAs from the NHS. These pairs of DNAs were separate aliquots from the same DNA preparation; all met quality control criteria required for the other DNAs, thereby providing reliable data for comparison. Analysis of the discrepancies within these pairs of DNA uncovered results similar to those of the Centre d'Etude du Polymorphisme Humain (CEPH) DNA duplicates reported in the prostate cancer CGEMS GWAS (7). We observed an average concordance rate of 99.985% (50,820,003 concordant genotype calls out of 50,827,468 comparisons).

2) T2D_NHS and T2D_HPFS: Detailed QC procedures were provided previously (2).

Briefly, the NHS and HPFS T2D GWA scans are a component of the Gene Environment-Association Studies (GENEVA) under the NIH Genes, Environment and Health Initiative (GEI). Genotypic data first passed Broad's initial QC which included SNP fingerprints for sample tracking and early detection of sample misidentification, missing call rates of

$\geq 5\%$, the use of a HapMap control to check genotype quality independent of study samples and tracking of reagent and instrumental performance.

Genotype data were subsequently released for further QC to the GENEVA Coordinating Center at the University of Washington. Relatedness was evaluated using pairwise identity-by descent estimation using 80k SNPs in a method of moments approach implemented in PLINK software (8). Of the 909,622 SNP probes on the array, 879,071 passed the Broad's technical QC standards for NHS samples, and 874,517 SNP probes passed this QC stage for HPFS samples. We applied the same QC parameters to both scans: excluding SNPs which were monomorphic, had a missing call rate of $\geq 2\%$, more than one discordance, significant deviations from HWE ($P < 1 \times 10^{-4}$) and an MAF of < 0.02 . Duplicate SNPs (assayed with different probes) were also removed. A total of 704,409 SNPs for NHS samples and 706,040 SNPs for HPFS samples passed QC.

3) CHD_NHS and CHD_HPFS: QC procedures for CHD_NHS and CHD_HPFS were similar to those used for T2D. Briefly, PLINK software (8) was used for data cleaning. The data were cleaned separately for NHS and HPFS according to thresholds recommended by the GENEVA consortia described above. We excluded SNPs that met any of following criteria: 1) MAF < 0.02 ; 2) call rate $< 95\%$; 3) P for Hardy-Weinberg equilibrium HWE < 0.00001 in control groups; 4) concordance rate $< 95\%$ among the duplicated QC samples; 5) significant difference in missing rates between cases and controls ($P < 0.00001$). After applying the QC filter, 721,316 SNPs remained for NHS and 724,881 for HPFS.

4) KS: All samples were highly genotyped. SNPs with MAF<1.0% in either cases or controls were removed, as were SNPs with less than 97% completion in cases or controls or any SNP with HWE-test $p < 10^{-4}$.

5) ADVCAP_HPFS: Samples were excluded if the genotyping call rate was $\leq 95\%$. Samples with autosomal heterozygosity < 0.25 or > 0.35 were excluded. SNPs with MAF<2.5% in either cases or controls were removed, as were SNPs with less than 95% completion in cases or controls or any SNP with HWE-test $p < 10^{-5}$. We had 519,982 SNPs after filtering.

6) GLAUCOMA: Samples were excluded if the genotyping call rate was $\leq 97\%$. SNPs with MAF<1.0% in either cases or controls were removed, as were SNPs with less than 95% completion in cases or controls.

7) PANSCAN: Samples were excluded if the genotyping call rate was $\leq 97\%$, no other filter.

8) MELANOMA: SNPs with MAF<1.0% in either cases or controls were removed, as were SNPs with less than 97% completion in cases or controls or any SNP with HWE-test $p < 10^{-4}$.

M.D. Anderson set

The study participants for the discovery analysis were from a hospital-based case-control study of melanoma, for which cases were recruited from among non-Hispanic white patients and controls at MD Anderson between March 1998 and August 2008. Samples and data were available from 931 melanoma patients and 1,026 cancer-free controls (friends of other patients reporting to clinics), who were frequency-matched on age and sex, completed a comprehensive skin lifestyle questionnaire, and passed quality control filters for genotyping. This questionnaire was administered by an interviewer to 70% of patients and controls and was self-administered for the remaining 30%. An additional case series comprising 873 individuals presenting for treatment for melanoma at MD Anderson was also included, bringing the total number of melanoma patients to 1,804. The study protocols were approved by the Institutional Review Board at MD Anderson, and informed consent was obtained from all participants.

Tissue samples were collected as whole blood, with various DNA extraction methods (including Gentra, Qiagen, and phenol/chloroform). DNA samples for the first-stage genome-wide association study were genotyped using the Illumina Omni1-Quad array and were called using the BeadStudio algorithm, at the John Hopkins University Center for Inherited Disease Research (CIDR).

QC procedures for the M.D. Anderson set:

Mean call rate for all samples was 99.86%. Only 41 failed genotyping with >10% missing rate across all SNPs, and 11 samples had identity problems that could not be

resolved. For this study, the IBD coefficients were estimated using 116,002 autosomal SNPs in PLINK15. In total, 126 duplicated, related (IBD), or outliers identified by PCA were excluded from the study. Following these exclusions there were 1,952 cases and 1,026 controls. Among 2,978 total cases and controls passing quality control, 138 in situ cases were subsequently removed from the study for indeterminate phenotype. Ten atypical melanocytic proliferation (AMPs) patients were also excluded as not having invasive cancers. Finally, we analyzed data from 1,804 cases and 1,026 controls available for the association study of melanoma susceptibility.

PD GWASs

The 37 significant PD SNPs were selected from 11 PD GWASs (9-19).

References

1. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat Genet.* 2007;39:870-4.
2. Qi L, Cornelis MC, Kraft P, Stanya KJ, Linda Kao WH, Pankow JS, et al. Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum Mol Genet.* 2010;19:2706-15.
3. Rimm EB, Giovannucci EL, Willett WC, Colditz GA, Ascherio A, Rosner B, et al. Prospective study of alcohol consumption and risk of coronary disease in men. *Lancet.* 1991;338:464-8.
4. Curhan GC, Taylor EN. 24-h uric acid excretion and the risk of kidney stones. *Kidney Int.* 2008;73:489-96.
5. Taylor EN, Stampfer MJ, Curhan GC. Obesity, weight gain, and the risk of kidney stones. *JAMA.* 2005;293:455-62.
6. Wolpin BM, Kraft P, Gross M, Helzlsouer K, Bueno-de-Mesquita HB, Steplowski E, et al. Pancreatic cancer risk and ABO blood group alleles: results from the pancreatic cancer cohort consortium. *Cancer research.* 2010;70:1015-23.
7. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet.* 2007;39:645-9.

8. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559-75.
9. Spencer CC, Plagnol V, Strange A, Gardner M, Paisan-Ruiz C, Band G, et al. Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21. *Hum Mol Genet.* 2011;20:345-53.
10. Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet.* 2009;41:1308-12.
11. Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet.* 2009;41:1303-7.
12. Saad M, Lesage S, Saint-Pierre A, Corvol JC, Zelenika D, Lambert JC, et al. Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. *Hum Mol Genet.* 2011;20:615-27.
13. Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin UM, Saad M, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet.* 2011;377:641-9.
14. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, et al. High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet.* 2005;77:685-93.
15. Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Yearout D, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nat Genet.* 2010;42:781-5.
16. Fung HC, Scholz S, Matarin M, Simon-Sanchez J, Hernandez D, Britton A, et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol.* 2006;5:911-6.
17. Edwards TL, Scott WK, Almonte C, Burt A, Powell EH, Beecham GW, et al. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. *Ann Hum Genet.* 2010;74:97-109.
18. Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, et al. Web-based genome-wide association study identifies two novel Loci and a substantial genetic component for Parkinson's disease. *PLoS Genet.* 2011;7:e1002141.
19. IPDGC and WTCCC2. A two-stage meta-analysis identifies several new Loci for Parkinson's disease. *PLoS Genet.* 2011;7:e1002142.