Supplementary Methods

Melanoma Discovery set in Harvard:

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 from among women who gave a blood sample and were free of diagnosed cancer (excluding non-melanoma skin cancer) up to and including the interval during 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. 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 reported an incident CHD event on the follow-up questionnaire were contacted for confirmation and permission to review medical records. Medical records were also

sought for deceased participants identified by families, postal officials, or through the National Death Index. Physicians blinded to the participants' 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.

5) Prostate cancer study nested within the HPFS (ADVCAP_HPFS): Prostate cancer cases were matched to controls on birth year (+/-1) and ethnicity. Controls were selected from those who were 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, and type of sample (blood or cheek cell).

7) PANSCAN

Subjects were followed prospectively with repeated assessment 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 as of the incidence date of the matched case (6).

Quality control (QC) procedures for eight GWAS of the discovery set:

1) Harvard

CGEMS: Detailed QC procedures were described 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 DNA samples 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 noted an average concordance rate of 99.985% (50,820,003 concordant genotype calls out of 50,827,468 comparisons).

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 ≥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 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, we had a missing call rate of ≥2%, more than one discordance, significant deviations from HWE ($P < 1 \times 10^{-4}$), and a minor allele

frequency (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.

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 the 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; and 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.

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 $p<10^{-4}$.

ADVCAP_HPFS: Samples were excluded if the genotyping call rate was ≤ 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.

GLAUCOMA: Samples were excluded if the genotyping call rate was ≤ 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.

PANSCAN: Samples were excluded if the genotyping call rate was ≤ 97% (no other filter).

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 $p<10^{-4}$.

After adjusting for the top five principal components of genetic variation from each cohort, the distribution of the observed *P* values did not suggest any inflation in Type 1 error rates due to population stratification or other sources of bias (Lambda=1.025) (Supplemental Figure 2 A). Sensitivity analysis restricted to the Illumina platform showed similar distribution (Lambda=1.028) (Supplemental Figure 2 B).

2) MD Anderson Cancer Center

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 identity by descent (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 patients with atypical melanocytic proliferation (AMP) 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 (9).

Melanoma replication set:

The NHSII (Nurses' Health Study II) was established in 1989, when 116,678 female registered nurses between the ages of 25 and 42 years and residing in the United States at the time of enrollment responded to an initial questionnaire on their medical histories and baseline health-related exposures. Updated information was obtained by questionnaire every 2 years. Details of this cohort have been described previously (10). Blood samples were collected from NHS2 participants in 1996-98. We collected cheek cell samples in 2005 from the NHS2 women who had not provided a blood sample in 1996-98.

Melanoma case-control study within the NHS II

All the cases and controls in our study were from sub-cohorts of the NHS II who had given a blood specimen or cheek cell sample. Eligible cases consisted of women in the NHS II with pathologically confirmed incident melanoma diagnosed any time after the baseline up to June 1, 2007 and without previously diagnosed cancer. One control per case was randomly selected from participants who were free of diagnosed melanoma up to and including the questionnaire cycle in which the cases were diagnosed. Controls were matched to cases by age (±1 year). The cases and their matched controls were selected in the same specimen cohort. All subjects were US non-Hispanic Caucasians.

Genotyping assay

We genotyped seven SNPs using the Taqman Assay (Applied Biosystems, CA). Laboratory personnel were blinded to case–control status, and blinded quality control samples were inserted to validate genotyping procedures; concordance for the blinded

samples was 100%. Primers, probes, and conditions for genotyping assays are available upon request.

Genotyping in replication set.

1) GenoMEL

The GenoMel data incorporated into the *in silico* replication came from a GWAS of 2,804 cases and 1,835 controls for melanoma collected at 11 different centers across Europe and Israel. These were genotyped in two phases: Phase 1 (Bishop et al., 20098) on the Illumina HumanHap300 BeadChip version 2 duo array (317k SNPs), and Phase 2 on the Illumina Human610 quad array (610k SNPs). WTCCC study controls were genotyped on the Illumina HumanHap1.2 million array. French controls were genotyped by the Centre National de Genotypage on the Illumina Humancnv370k array (11).

2) Australia

A total of 2,265 melanoma cases were available for genotyping, including a population-based sample of 1,697 cases from Queensland unselected for age at onset (Queensland Study of Melanoma: Environment and Genetic Associations; Q-MEGA1) and 568 cases from a population-based, case-control, family study of melanoma diagnosed before age 40, ascertained in Brisbane, Melbourne, and Sydney (Australian Melanoma Family Study 2; AMFS). Approval for these studies was obtained from the Human Research Ethics Committees of QIMR, the University of Sydney, the University of Melbourne, and cancer registries of NSW, Victoria, and Queensland. Informed consent was obtained from all participants.

DNA was extracted from peripheral blood or saliva samples. Australian twin and endometriosis sample controls were genotyped at deCODE Genetics (Reykjavik, Iceland) on the Illumina HumanHap610W Quad and Illumina HumanHap670 Quad Beadarrays, respectively. AMFS controls were genotyped by Illumina (San Diego) on Illumina Omni1-Quad arrays. Cases were genotyped by Illumina (San Diego) on Illumina Omni1-Quad (568 AMFS cases, 699 Q-MEGA cases) and HumanHap610W Quad arrays (998 Q-MEGA cases). All genotypes were called with the Illumina BeadStudio software (12).

Quality control (QC) procedures for the replication set:

1) GenoMEL

Individuals were excluded for low call rate (<97% on the array on which the sample was genotyped), non-European ethnicity (determined by PCA in a manner similar to that described above), sex discrepancy with recorded phenotype information, first-degree or closer relatedness with another sample or recommendation for exclusion by WTCCC. QC was applied to SNPs separately for each genotyping platform. SNPs were excluded for low call rate (<97%), Hardy-Weinberg equilibrium P-value < 10^{-20}, or recommendation for exclusion by WTCCC (usually on the basis of poor clustering).

2) Australia

SNPs with a mean BeadStudio GenCall score < 0.7 were excluded from the control data sets. All samples had successful genotypes for > 95% of SNPs. SNPs with call rates either < 0.95 (minor allele frequency, MAF > 0.05) or <0.99 (MAF > 0.01), Hardy-Weinberg equilibrium in controls P<10^{-6}, and/or MAF <0.01 were excluded.

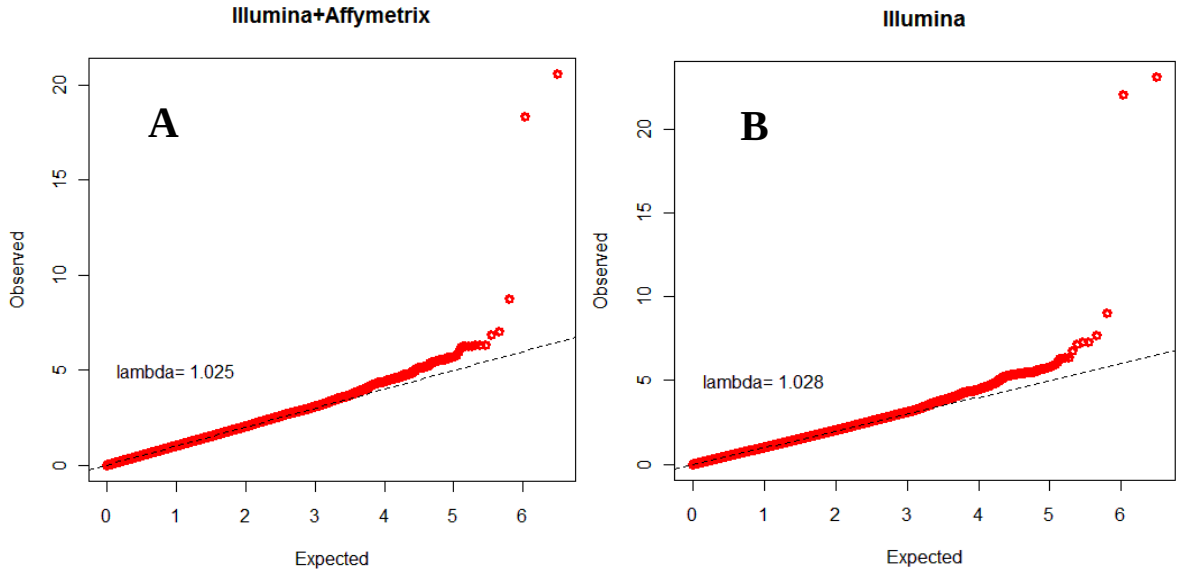
Supplemental Table 1: Replication of the top 69 SNPs in GenoMEL and Australia

SNP information				Harvard & MD Anderson			GenoMEL & Australia			Meta-analysis		
SNP	CHR	A1	A2	Beta	SE	P	Beta	SE	P	Beta	SE	P
rs12061304	1	a	g	-0.21	0.05	7.25e-06	-0.01	0.03	0.79	-0.06	0.02	0.015
rs1965402	1	a	g	-0.16	0.04	4.99e-05	-0.03	0.03	0.32	-0.06	0.02	0.003
rs641227	1	a	g	0.17	0.04	2.28e-05	-0.01	0.03	0.72	0.04	0.02	0.05
rs6424432	1	a	t	-0.17	0.04	3.86e-05	-0.02	0.03	0.47	-0.06	0.02	0.004
rs6684085	1	t	c	-0.22	0.05	1.12e-05	0.02	0.03	0.62	-0.05	0.03	0.06
rs10192348	2	t	c	-0.17	0.04	2.36e-05	0.06	0.03	0.01	-0.01	0.02	0.86
rs10193840	2	a	g	-0.39	0.09	2.30e-05	-0.01	0.05	0.79	-0.11	0.05	0.02
rs10931010	2	t	c	-0.27	0.06	2.21e-05	-0.08	0.04	0.06	-0.14	0.04	7.03e-05
rs11679247	2	t	g	-0.27	0.06	3.01e-05	0.01	0.05	0.78	-0.09	0.04	0.02
rs11901831	2	t	c	0.33	0.08	1.07e-05	0.12	0.05	0.03	0.19	0.04	1.48e-05
rs13404035	2	t	c	-0.17	0.04	3.70e-05	-0.08	0.03	0.008	-0.12	0.03	4.15e-06
rs17489442	2	t	c	-0.20	0.04	4.23e-06	-0.02	0.03	0.48	-0.08	0.02	0.002
rs3791470	2	t	c	0.25	0.06	4.84e-05	-0.02	0.03	0.49	0.03	0.03	0.22
rs4387818	2	a	g	-0.21	0.05	3.39e-05	0.04	0.03	0.23	-0.04	0.03	0.20
rs4671369	2	t	c	-0.17	0.04	3.21e-05	-0.01	0.03	0.57	-0.06	0.02	0.007
rs833755	2	t	g	0.21	0.05	1.37e-05	-0.01	0.03	0.84	0.05	0.02	0.037
rs1031925	3	t	c	0.25	0.06	2.06e-05	0.10	0.04	0.005	0.14	0.03	4.27e-06
rs12487800	3	a	g	-0.20	0.05	2.20e-05	-0.01	0.03	0.77	-0.07	0.03	0.007
rs13081536	3	t	c	0.17	0.04	1.86e-05	-0.02	0.03	0.59	0.06	0.03	0.03
rs13097028	3	t	c	-0.17	0.04	4.96e-05	-0.09	0.03	0.002	-0.11	0.02	9.99e-07
rs2352	3	a	g	-0.20	0.05	2.49e-05	0.04	0.04	0.26	-0.05	0.03	0.09
rs4279134	3	a	g	-0.17	0.04	4.47e-05	-0.02	0.03	0.42	-0.06	0.02	0.004
rs6443119	3	t	c	-0.30	0.06	3.34e-06	0.09	0.03	0.01	0.01	0.03	0.98
rs7647302	3	a	t	0.61	0.15	2.56e-05	NA	NA	NA	0.61	0.15	2.56e-05
rs11935777	4	a	g	0.23	0.06	4.10e-05	0.05	0.04	0.23	0.11	0.03	0.001
rs4698934	4	t	c	0.26	0.06	5.26e-06	0.10	0.04	0.018	0.15	0.03	5.21e-06
rs6533194	4	a	t	0.27	0.06	2.35e-06	0.04	0.04	0.28	0.12	0.03	0.0003
rs6835433	4	a	c	-0.60	0.13	9.07e-06	0.04	0.08	0.65	-0.13	0.07	0.06
rs1560550	5	a	g	-0.17	0.04	1.41e-05	0.03	0.02	0.21	-0.03	0.02	0.23
rs17196873	5	t	c	0.27	0.07	4.66e-05	-0.02	0.04	0.61	0.06	0.03	0.09
rs925203	5	t	g	-0.21	0.05	1.87e-05	-0.03	0.03	0.43	-0.07	0.03	0.004
rs1889497	6	a	t	-0.23	0.05	7.00e-06	-0.10	0.04	0.011	-0.14	0.03	2.45e-06
rs6597184	6	a	t	0.27	0.06	1.14e-05	-0.03	0.06	0.58	0.11	0.04	0.008
rs9381024	6	t	g	0.22	0.05	3.01e-05	0.02	0.05	0.68	0.11	0.03	0.002
rs9502429	6	t	c	0.19	0.05	3.20e-05	0.02	0.03	0.49	0.07	0.02	0.005

rs6460735	7	t	c	0.23	0.06	3.46e-05	0.02	0.03	0.64	0.08	0.03	0.009
rs6973378	7	a	g	-0.17	0.04	3.78e-05	0.02	0.03	0.47	-0.04	0.02	0.10
rs10081518	8	a	g	0.30	0.07	3.41e-05	-0.09	0.05	0.10	0.05	0.04	0.27
rs7015177	8	c	g	-0.19	0.04	1.95e-06	-0.01	0.04	0.77	-0.10	0.03	0.0005
rs754497	8	t	c	-0.20	0.05	3.54e-05	0.03	0.03	0.33	-0.03	0.02	0.19
rs10992715	9	a	g	0.34	0.08	3.56e-05	-0.01	0.05	0.88	0.08	0.04	0.05
rs12339003	9	c	g	-0.30	0.07	4.61e-05	-0.11	0.07	0.12	-0.20	0.05	8.27e-05
rs2123338*	9	t	c	-0.22	0.05	1.55e-06	-0.16	0.03	2.32e-07	-0.18	0.03	2.93e-12
rs923784	9	t	c	-0.18	0.04	1.33e-05	0.01	0.03	0.60	-0.04	0.02	0.06
rs11593549	10	t	c	-0.23	0.05	1.91e-05	0.03	0.04	0.44	-0.05	0.03	0.08
rs1444773	10	a	g	0.21	0.05	1.46e-05	0.04	0.03	0.29	0.10	0.03	0.0007
rs6481044	10	a	g	0.35	0.08	9.77e-06	-0.14	0.07	0.09	0.08	0.05	0.11
rs7090265	10	t	c	-0.18	0.04	6.84e-06	0.03	0.03	0.39	-0.05	0.02	0.04
rs10128702	11	t	g	-0.34	0.08	8.55e-06	0.06	0.04	0.12	-0.02	0.04	0.51
rs6606742	12	a	g	-0.18	0.04	1.56e-05	0.01	0.02	0.56	-0.04	0.02	0.08
rs1407584	13	a	c	-0.17	0.04	3.93e-05	0.02	0.03	0.45	-0.03	0.02	0.12
rs2329154	13	t	c	0.18	0.04	1.37e-05	-0.03	0.03	0.31	0.04	0.02	0.08
rs4773180	13	t	g	0.28	0.07	3.94e-05	0.10	0.05	0.044	0.16	0.04	5.69e-05
rs13379344	14	a	g	0.28	0.07	4.25e-05	0.03	0.05	0.51	0.11	0.04	0.005
rs1383595	14	t	c	-0.28	0.07	4.42e-05	0.02	0.04	0.61	-0.07	0.04	0.07
rs7157749	14	c	g	-0.20	0.05	2.43e-05	-0.03	0.03	0.34	-0.07	0.02	0.003
rs1406714	15	c	g	-0.18	0.04	9.20e-06	-0.05	0.03	0.10	-0.09	0.02	7.61e-05
rs2412524	15	t	g	0.20	0.04	8.98e-07	0.03	0.03	0.33	0.08	0.02	0.0005
rs11652580	17	t	c	0.21	0.05	5.42e-06	-0.03	0.03	0.36	0.04	0.03	0.09
rs12449476	17	t	c	0.22	0.05	5.67e-06	-0.01	0.03	0.66	0.05	0.03	0.04
rs12941384	17	a	g	0.17	0.04	3.82e-05	-0.02	0.03	0.50	0.04	0.02	0.10
rs12952218	17	t	c	0.17	0.04	3.06e-05	-0.02	0.02	0.49	0.03	0.02	0.11
rs225184	17	t	c	0.21	0.05	1.20e-05	0.01	0.03	0.89	0.06	0.03	0.02
rs1037418	18	t	c	0.22	0.05	1.62e-05	0.03	0.03	0.29	0.09	0.03	0.001
rs560634	19	t	g	0.29	0.07	4.28e-05	-0.03	0.04	0.47	0.06	0.04	0.13
rs2076687	20	a	c	-0.23	0.05	9.67e-06	0.05	0.04	0.22	-0.05	0.03	0.09
rs4622799	20	a	g	-0.20	0.05	2.75e-05	-0.02	0.05	0.62	-0.11	0.03	0.001
rs6118062	20	t	c	0.22	0.05	1.34e-05	-0.03	0.03	0.36	0.04	0.03	0.14
rs132985*	22	t	c	-0.17	0.04	2.80e-05	-0.11	0.03	6.75e-06	-0.13	0.02	1.59e-09

* SNPs rs2123338 and rs132985 were located in known loci reported to be associated with melanoma risk.

Supplemental Figure 1: Log Quantile-Quantile (Q-Q) P value plot for melanoma GWAS in Harvard.



References

1 Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A. *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nature genetics*, **39**, 870-874.

2 Qi, L., Cornelis, M.C., Kraft, P., Stanya, K.J., Linda Kao, W.H., Pankow, J.S., Dupuis, J., Florez, J.C., Fox, C.S., Pare, G. *et al.* (2010) Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Human molecular genetics*, **19**, 2706-2715.

3 Rimm, E.B., Giovannucci, E.L., Willett, W.C., Colditz, G.A., Ascherio, A., Rosner, B. and Stampfer, M.J. (1991) Prospective study of alcohol consumption and risk of coronary disease in men. *Lancet*, **338**, 464-468.

4 Curhan, G.C. and Taylor, E.N. (2008) 24-h uric acid excretion and the risk of kidney stones. *Kidney Int*, **73**, 489-496.

5 Taylor, E.N., Stampfer, M.J. and Curhan, G.C. (2005) Obesity, weight gain, and the risk of kidney stones. *JAMA*, **293**, 455-462.

6 Wolpin, B.M., Kraft, P., Gross, M., Helzlsouer, K., Bueno-de-Mesquita, H.B., Steplowski, E., Stolzenberg-Solomon, R.Z., Arslan, A.A., Jacobs, E.J., Lacroix, A. *et al.* (2010) Pancreatic cancer risk and ABO blood group alleles: results from the pancreatic cancer cohort consortium. *Cancer research*, **70**, 1015-1023.

7 Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N. *et al.* (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet*, **39**, 645-649.

8 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, **81**, 559-575.

9 Amos, C.I., Wang, L.E., Lee, J.E., Gershenwald, J.E., Chen, W.V., Fang, S., Kosoy, R., Zhang, M., Qureshi, A.A., Vattathil, S. *et al.* (2011) Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. *Human molecular genetics*, **20**, 5012-5023.

10 Bertone-Johnson, E.R., Hankinson, S.E., Johnson, S.R. and Manson, J.E. (2009) Timing of alcohol use and the incidence of premenstrual syndrome and probable premenstrual dysphoric disorder. *J Womens Health (Larchmt)*, **18**, 1945-1953.

11 Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C. and Zhang, Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129-1133.

12 Jankowska, A.M., Szpurka, H., Tiu, R.V., Makishima, H., Afable, M., Huh, J., O'Keefe, C.L., Ganetzky, R., McDevitt, M.A. and Maciejewski, J.P. (2009) Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. *Blood*, **113**, 6403-6410.