

Supplementary Acknowledgements

We wish to acknowledge all of the investigators and funding agencies that enabled the deposition of data in dbGaP that we used in this study:

Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1 through dbGaP accession number phs000092.v1.p1.

The CGEMS Breast Cancer study was supported by the National Cancer Institute. Details of the study can be found at <http://cgems.cancer.gov/data> and in reference 28. The data was obtained from dbGaP under accession phs000147v1.

The CGEMS Prostate Cancer study was supported by the National Cancer Institute. Details of the study can be found at <http://cgems.cancer.gov/data> and in reference 29. The data was obtained from dbGaP under accession phs000207v1.

The CIDR: Genome Wide Association Study in Familial Parkinson Disease study was funded by the National Institute of Neurological Disorders and Stroke. The dataset analysed here was obtained from dbGaP at under accession number phs000126v1.

The Study of Irish Amyotrophic Lateral Sclerosis was funded by the National Institute of Neurological Disorders and Stroke. The dataset analysed here was obtained from dbGaP at under accession number phs000127v1. Details of the study can be found in ref. 26.

Funding support for the GWAS of Lung Cancer and Smoking was provided through the NIH Genes, Environment and Health Initiative [GEI] (Z01 CP 010200). The human subjects participating in the GWAS derive from The Environment and Genetics in Lung Cancer Etiology (EAGLE) case-control study and the Prostate, Lung Colon and Ovary Screening Trial and these studies are supported by intramural resources of the National Cancer Institute. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was

provided by the Gene Environment Association Studies, GENEVA Coordinating Center (U01HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438). The datasets used for the analyses described in this manuscript were obtained from dbGaP at <http://www.ncbi.nlm.nih.gov/gap> through dbGaP accession number phs000093v2.

The PanScan study was supported by the National Cancer Institute. Details of the study can be found at <http://cgems.cancer.gov/data> and in references 11 and 12. The data was obtained from dbGaP under accession phs000206v1.