Supplementary Methods

1958 Birth Cohort Controls (58BC) The 1958 Birth Cohort (also known as the National Child Development Study) includes all births in England, Wales and Scotland, during one week in 1958. From an original sample of over 17,000 births, survivors were followed up at ages 7, 11, 16, 23, 33 and 42 years (http://www.cls.ioe.ac.uk/studies.asp?section=000100020003). In a biomedical examination at 44-45 years (http://www.b58cgene.sgul.ac.uk/followup.php), 9377 cohort members were visited at home providing 7692 blood samples with consent for future EBV transformed cell lines. DNA samples extracted from 1500 cell lines of self-reported white ethnicity and representative of gender and each geographical region were selected for use as controls.

Ascetainment and Diagnosis of Primary Scan Samples

Ankylosing Spondylitis

AS was defined according to the modified New York diagnostic criteria[1]. The diagnosis of AS was made in all patients by a qualified rheumatologist. To confirm diagnosis all cases were either examined or interviewed by telephone by one of the investigators. In cases with atypical histories or where radiographs had not been previously performed, pelvic and lumbo-sacral spine radiographs were obtained, and attending physicians contacted to confirm the diagnosis. Cases were questioned regarding whether they had any history of IBD, and where this was reported, diagnoses were confirmed with caring physicians.

Auto-immune Thyroid Disease

Unrelated white Caucasian Graves' disease (GD) patients were recruited as part of the autoimmune thyroid disease (AITD) UK National Collection. Patients were recruited from centers across the U.K. including Birmingham, Bournemouth, Cambridge, Cardiff, Exeter, Leeds, Newcastle and Sheffield. All recruiting centers used standard clinical criteria to diagnose GD to avoid any clinical heterogeneity. These criteria have recently been well described and the dataset characterized by Manji[2]. Briefly, GD patients were defined by the presence of biochemical hyperthyroidism together with the presence of either: (1) a diffuse goitre on a scan, (2) positive autoantibodies to the thyrotropin receptor (TSHR), (3) diffuse goitre on palpation along with thyroglobulin or thyroid peroxidase autoantibodies, or (4) thyroid eye disease (NOSPECS classification score of 2-6).

Breast Cancer

We included 1053 samples from independently ascertained women with invasive breast cancer. 864 samples were from unrelated individuals with breast cancer, who each had a family history of breast cancer in relatives but no relative with ovarian cancer. The samples were from breast cancer families ascertained through Cancer Genetics clinics in the UK and families from non-UK ethnic groups were excluded. We quantified the extent of the family history of breast cancer using a Family History Score, which was defined as the number of relatives of the index case with breast cancer, weighted by their degree of relatedness to the index case to adjust

for the expected allele sharing (score = 1 for each affected 1st degree relatives, 0.5 for 2nd degree relatives and 0.25 for 3rd degree relatives; bilateral breast cancers score double). The minimum family history score was 1.5 (i.e. a women with breast cancer with one first degree and one second degree relative with breast cancer or equivalent). We screened all samples for mutations in the full coding sequence and intron/exon boundaries of BRCA1 and BRCA2 by Conformation Sensitive Gel Electrophores. We also performed Multiplex ligation-dependent Probe Amplification (MLPA) analysis on all samples using the P002 BRCA1 kit and P045 BRCA2 kit (MRC Holland) and the manufacturers' protocols. BRCA1/2 mutations were identified in 17 cases and were excluded in 847 cases. The remaining 189 samples were from women with breast cancer recruited to the Royal Marsden Hospital/Gloucester Oncology Centre Breast Radiotherapy Fractionation Trial and the Royal Marsden Hospital Breast Radiotherapy Dosimetry Trial, two closed trials aimed at investigating adverse effects of radiotherapy. None of these patients are known to carry BRCA1 or BRCA2 mutations. We obtained informed consent from all patients and the research was approved by the London Multicentre Research Ethics Committee (MREC/01/2/18).

Multiple Sclerosis

The 994 non-related individuals with multiple sclerosis considered in this study were recruited from across the UK. All cases were white and diagnosed according to standard criteria[3,4]. All individuals gave written informed consent and donated a venous blood sample from which DNA was extracted by standard methods. Clinical and demographic features are shown in Table 1. The Extended Disability Status Score (EDSS)[5] was estimated in 945 of these individuals at recruitment (95.1%) and was found to have a mean value of 4.6.

Table 1. Clinical and Demographic Features of MS cases

MS	Cases (n=994)
Age: Mean (range) / years	43.8 (17-79)
Age at onset: Mean (range) / years	30.3 (10-60)
Disease Duration: Mean (range) / years	13.5 (0-57)
Female: n (%)	716 (72.0)
Disease type: n (%)	
Relapsing-remitting	520 (52.3)
Secondary progressive	309 (31.1)
Primary progressive	115 (11.6)
undefined	50 (5.0)

Replication Datasets

The replication dataset for Ankylosing Spondylitis consisted of 471 white North American cases drawn from the Prospective Study of Outcomes in Ankylosing Spondylitis (PSOAS), an observational study of AS severity[6], who were recruited from the clinics of the investigators (JDR, MMW, JCD, MHW) or local rheumatologists, from patient support and advocacy groups, and from the community by advertisement. Enrollment occurred from 2002 to the present. All patients had a diagnosis of AS by the modified New York criteria. Informed consent was obtained in all cases, and this work was reviewed and approved by the respective Internal Review

Boards (IRB's) at the corresponding universities. These samples were genotyped by ABI Taqman assays. Further genotyping of the *IL23R* locus in British samples was performed using the iPLEX assay (MassArray, Sequenom).

For the assessments of shared loci between AS and inflammatory bowel disease, 1767 unrelated Caucasian IBD patients of north European origin attending IBD clinics in East Anglia, UK were recruited comprising 928 UC and 755 CD. Diagnosis was made using standard Lennard Jones criteria on notes review. The population was >99% non-Jewish. Median age at diagnosis was 36.7 and 26.1 years for UC and CD respectively. The 633 ethnically and geographically matched healthy controls were previously recruited in East Anglia for the European Prospective Investigation of Nutrition and Cancer (EPIC). Median age of controls was 60 years. Genotyping of CD cases and controls was performed using the Taqman biallelic discrimination system using an ABI 7900HT analyser; Applied Biosystems, Foster City, CA. Genotyping of UC cases was performed using a 1536 SNP Golden Gate bead array; Illumina, San Diega, CA. Genotype distributions in case and control populations were consistent with Hardy-Weinberg Equilibrium ($p>0.05$) for all SNPs. Ethics committee approval was obtained (Cambridge LREC 01/418; MREC 03/5/012) and written informed consent was obtained from all study subjects.

The AITD cohort was extended by 1500 Graves' disease patients from the greater Birmingham area, and 2500 independent 1958 British Birth Cohort controls. Eight novel SNPs were genotyped in order to tag the *FCRL3* and *FCRL5* region (two of which tagged the associated SNPs in the nsSNP study) using an alternative technology (i.e. ABI Taqman).

ADDITIONAL AUTHOR NOTES.

Membership of BRAGGS and BCSC

BRAGGS

• Anne Barton, arc Epidemiology Unit, University of Manchester, Oxford Rd, Manchester, M13 9PT
• John D. Isaacs, Department of Rheumatology, University of Newcastle-Upon-Tyne, Framlington Place, Newcastle-Upon-Tyne NE2 4HH
• Ann W. Morgan, Leeds Institute of Molecular Medicine, Section of Academic Unit of Musculoskeletal Disease Wellcome Trust Brenner Building, St. James's University Hospital, Beckett Street, Leeds LS9 7TF
• Gerry D. Wilson, Genomic Medicine, The University of Sheffield, Western Bank, Sheffield, S10 2TN

The Breast Cancer Susceptibility Collaboration (UK) consists of the following contributors:

A Ardern-Jones, J Berg, A Brady, N Bradshaw, C, Brewer, G Brice, B Bullman, J Campbell, B Castle, R Cetnarsryj, C Chapman, C Chu, N Coates, T Cole, R Davidson, A Donaldson, H Dorkins, F Douglas, D Eccles, R Eeles, F Elmslie, DG Evans, S Goff, S Goodman, D Goudie, J Gray, L Greenhalgh, H Gregory, SV Hodgson, T Homfray, RS Houlston, L Izatt, L Jackson, L Jeffers, V Johnson-Roffey, F Kavalier, C Kirk, F Lalloo, C Langman, I Locke, M Longmuir, J Mackay, A Magee, S Mansour, Z Miedzybrodzka, J Miller, P Morrison, V Murday, J Paterson, M Porteous, N Rahman, M Rogers, S Rowe, S Shanley, A Saggar, G Scott, L Side, L Snadden, M Steel, M Thomas, S Thomas

References

1. Van der Linden, S., Valkenburg, H.A., Cats, A Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* **27**, 361-368 (1984).
2. Manji, N., Carr-Smith, J.D., Boelaert, K., Allahabadia, A., Armitage, M., Chatterjee, V.K., Lazarus, J.H., Pearce, S.H., Vaidya, B., Gough, S.C. & Franklyn, J.A. Influences of age, gender, smoking and family history on autoimmune thyroid disease phenotype. *J Clin Endocrinol Metab*; **91**, 4873-4880 (2006).
3. Poser, C.M., Paty, D.W., Scheinberg, L. et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* **13**, 227-231 (1983).
4. McDonald, W.I., Compston, A., Edan, G. et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol*. **50**, 121-127 (2001).
5. Kurtzke, J.F. Rating neurological impairment in multiple sclerosis: an expanded disability scale status. *Neurol*. **33**, 1444-1452 (1983).
6 Ward, M.M., Weisman, M.H., Davis, J.C., Jr. & Reveille, J.D. Risk factors for functional limitations in patients with long-standing ankylosing spondylitis. *Arthritis Rheum* **53**, 710-717 (2005).

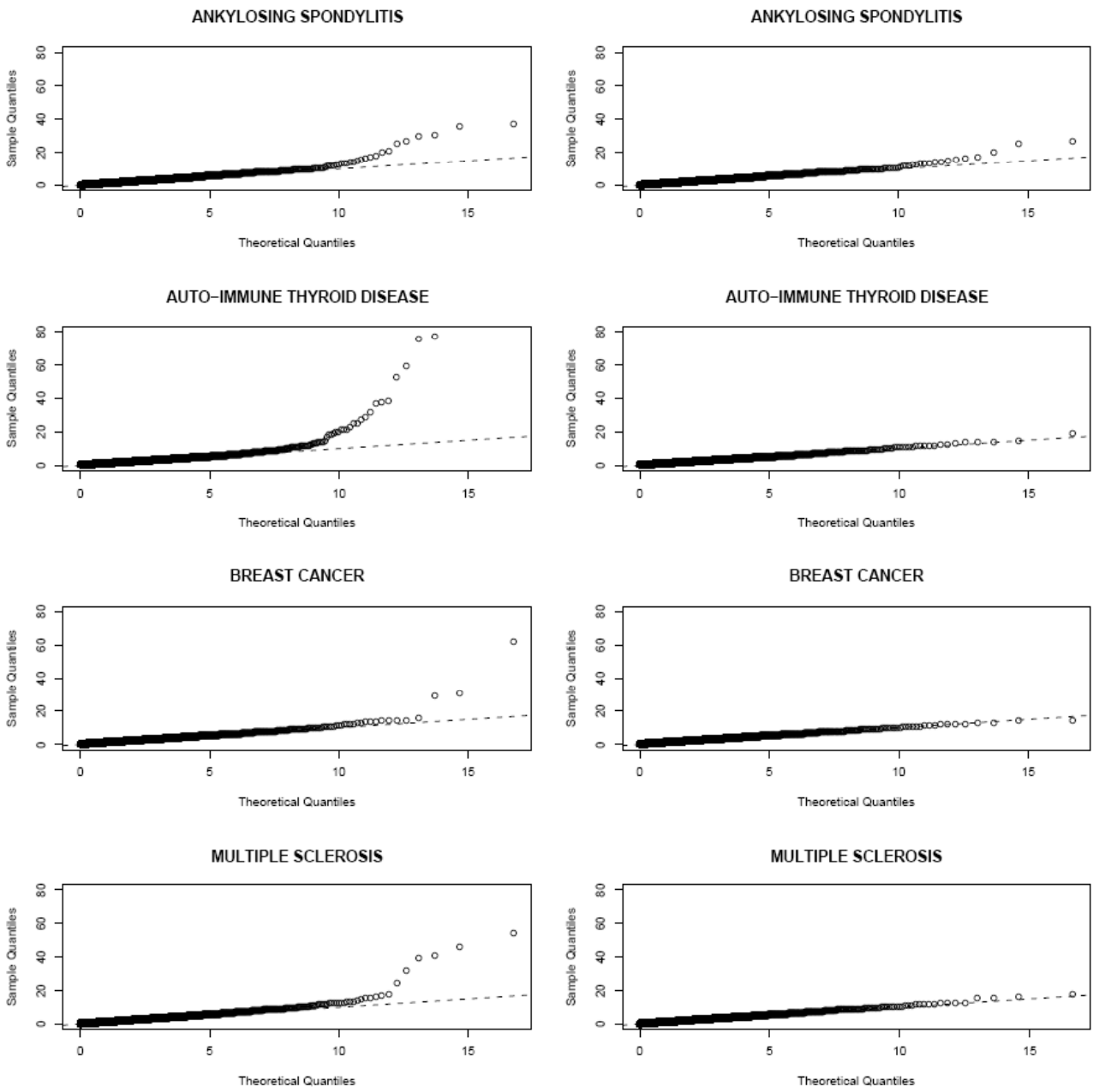
Membership of BRAGGS and BCSC

BRAGGS

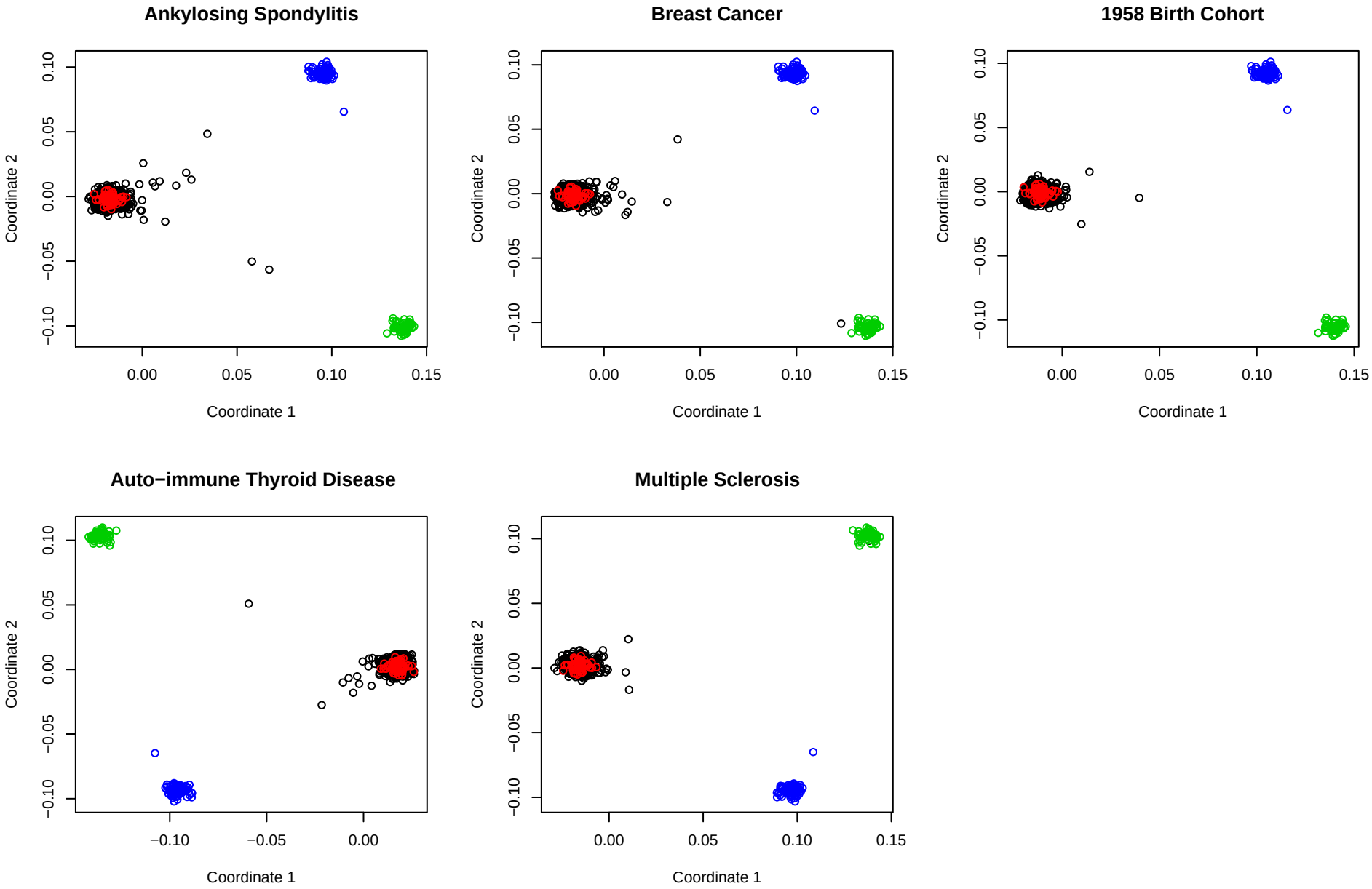
- *Anne Barton, arc Epidemiology Unit, University of Manchester, Oxford Rd, Manchester, M13 9PT*
- *John D Isaacs, Department of Rheumatology, University of Newcastle-Upon-Tyne, Framlington Place, Newcastle-Upon-Tyne NE2 4HH*
- *Ann W Morgan, Leeds Institute of Molecular Medicine, Section of Academic Unit of Musculoskeletal Disease Wellcome Trust Brenner Building, St. James's University of Hospital, Beckett Street, Leeds LS9 7TF*
- *Gerry D. Wilson, Genomic Medicine, The University of Sheffield, Western Bank, Sheffield, S10 2TN*

Breast Cancer Susceptibility Collaboration (UK)

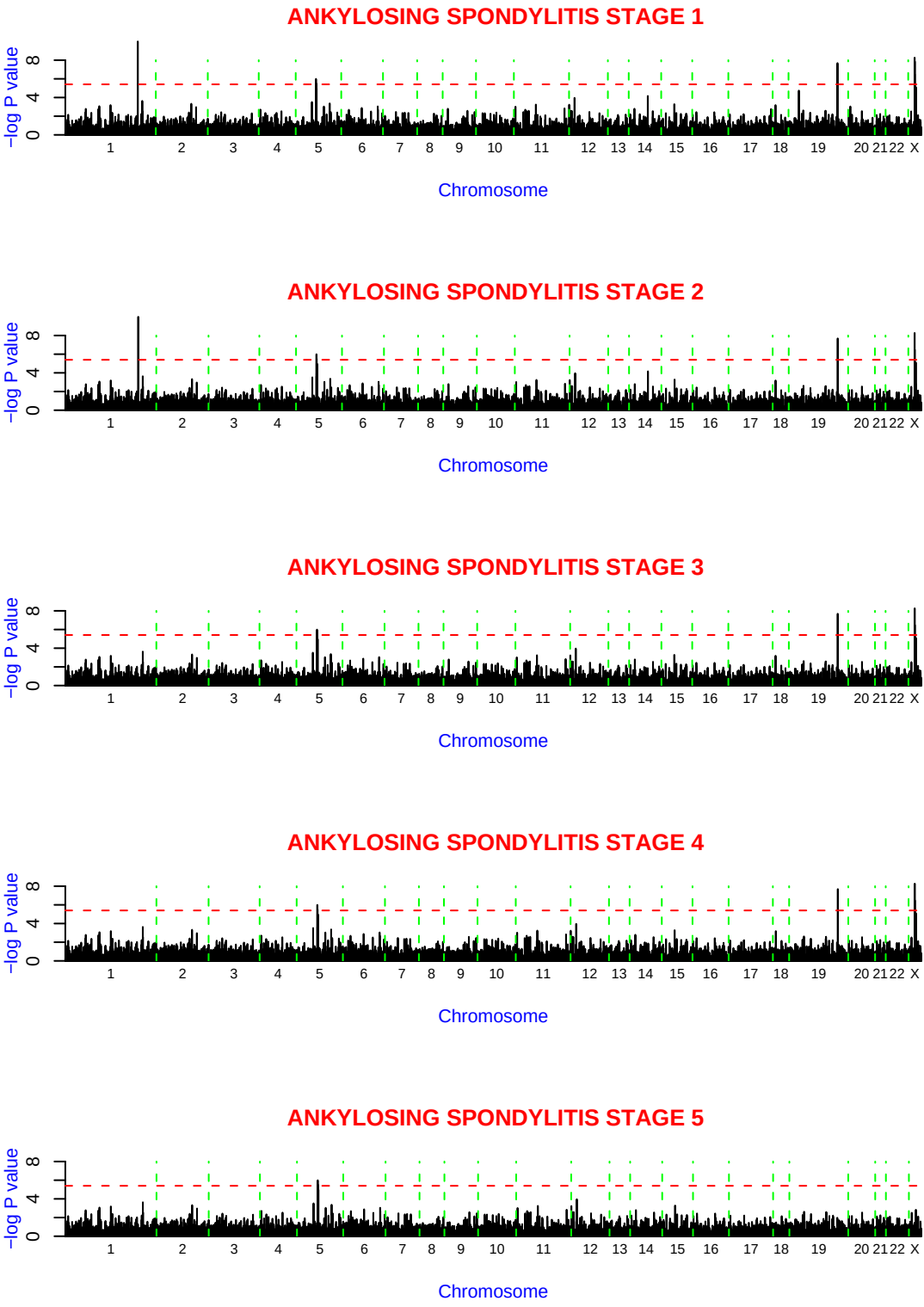
The Breast Cancer Susceptibility Collaboration (UK) consists of the following contributors: A Ardern-Jones, J Berg, A Brady, N Bradshaw, C Brewer, G Brice, B Bullman, J Campbell, B Castle, R Cetnarsryj, C Chapman, C Chu, N Coates, T Cole, R Davidson, A Donaldson, H Dorkins, F Douglas, D Eccles, R Eeles, F Elmslie, DG Evans, S Goff, S Goodman, D Goudie, J Gray, L Greenhalgh, H Gregory, SV Hodgson, T Homfray, RS Houlston, L Izatt, L Jackson, L Jeffers, V Johnson-Roffey, F Kavalier, C Kirk, F Lalloo, C Langman, I Locke, M Longmuir, J Mackay, A Magee, S Mansour, Z Miedzybrodzka, J Miller, P Morrison, V Murday, J Paterson, M Porteous, N Rahman, M Rogers, S Rowe, S Shanley, A Saggar, G Scott, L Side, L Snadden, M Steel, M Thomas, S Thomas



Supplementary Figure 1. QQplots for each WTCCC nsSNP disease sample against the 1958 birth cohort controls. The panels on the left show the dispersion of chi-squared statistics before any QC filtering, and are contrasted with equivalent plots after data cleaning on the right. SNPs from the MHC have been excluded from this analysis.

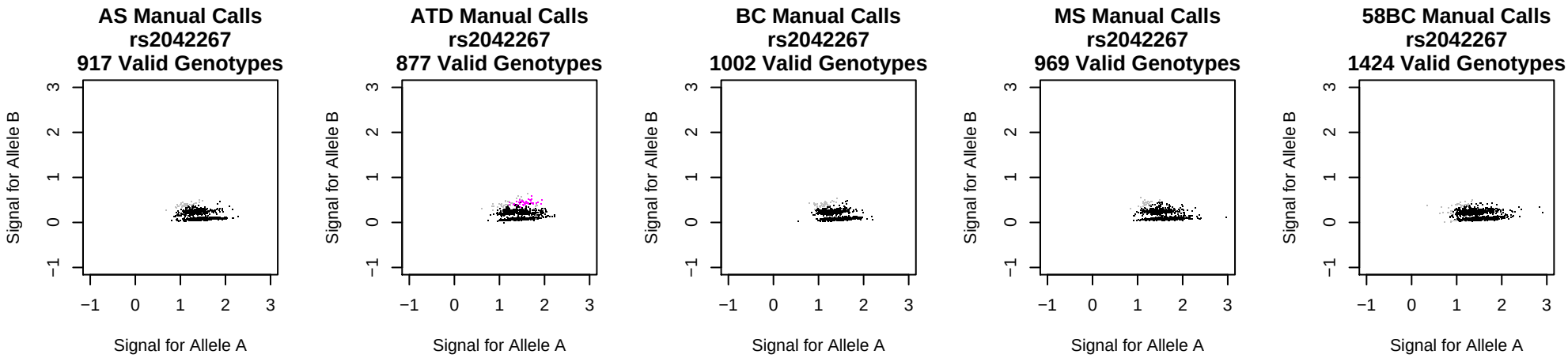


Supplementary Figure 2. Results of the principal coordinates analysis of the combined HapMap and WTCCC datasets.
Individuals' scores on the first two principal coordinates are plotted on the x and y axes respectively. CEPH individuals are plotted in red, YRI in green, CHB and JPT in blue, and WTCCC individuals in black. Clearly most WTCCC individuals cluster with the CEPH samples. However, several WTCCC individuals are displaced towards one of the other clusters. Follow up of these individuals revealed that many had ancestries which were not exclusively Northern European.

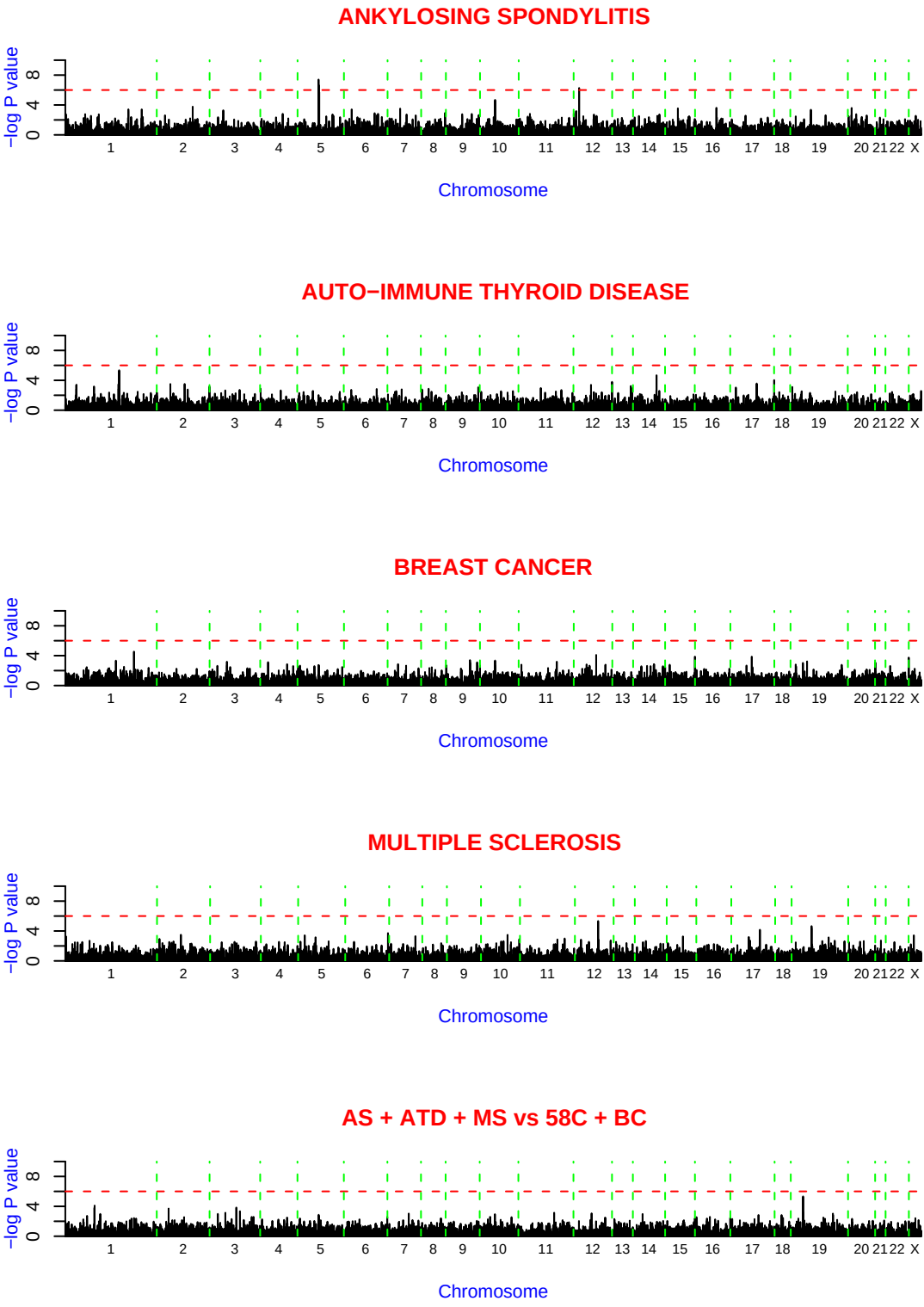


Supplementary Figure 3. Armitage significance tests after each stage of genotype filtering for Ankylosing Spondylitis (MHC SNPs not included.)

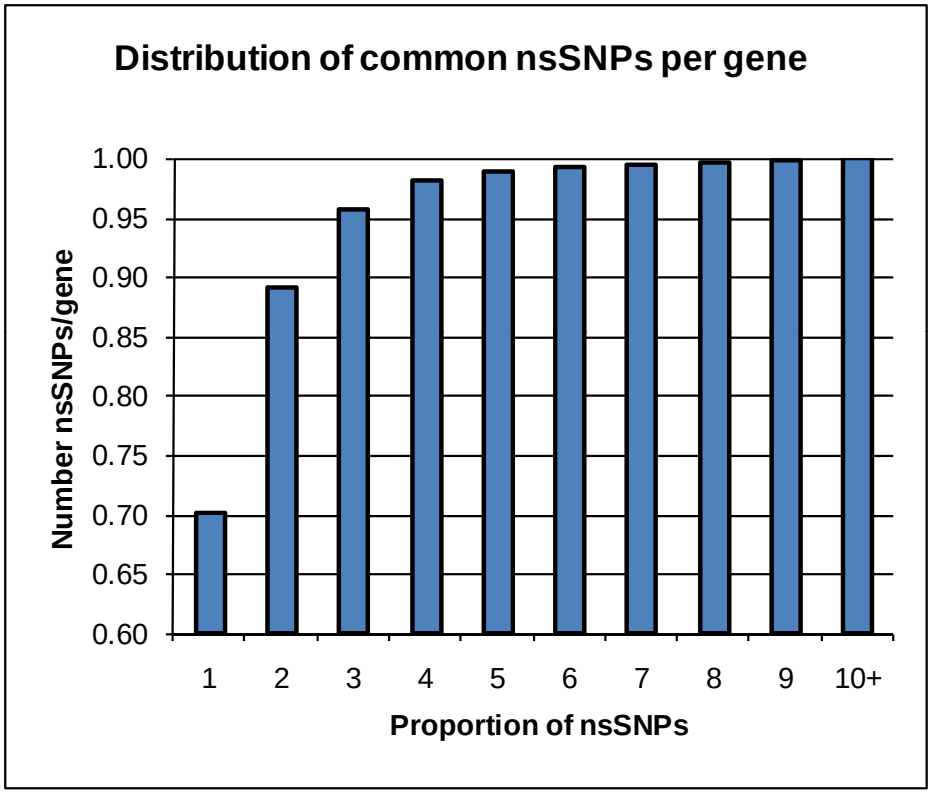
The filters employed are Stage 1: no SNPs removed from analyses; Stage 2: SNPs with > 10% missing genotypes removed from analyses; Stage 3: SNPs failing Hardy−Weinberg at $p < 10^{-7}$ in control individuals removed; Stage 4: SNPs that differ in missing rate between cases and controls at $p < 10^{-4}$ removed from analyses. Stage 5: Upon manual inspection of the raw genotype intensities, SNPs that poorly cluster removed from subsequent analyses.



Supplementary Figure 4. An example of a marker that had poor clustering properties, but which initially passed quality control.
The existence of these markers highlights the need to manually check significant results using the raw intensity data.



Supplementary Figure 5. Minus log10 p values for the Armitage test of trend for genome−wide association scans involving combined controls.
The spacing between SNPs on the plot is uniform and does not reflect distances between the SNPs. The vertical dashed lines reflect chromosomal boundaries. The horizontal dashed lines display the cutoff for $p = 10^{-6}$.



Supplementary Figure 6. Cumulative distribution of nsSNPs per gene. Most of the nSNPs in the present study occur either individually or together with 1-2 additional nsSNPs per gene. This distribution reflects the present availability of common variants and underrepresentation of rare variants.