SUPPLEMENTARY INFORMATION

MATERIALS AND METHODS

Sample Collection

Samples were collected by a variety of means. One of the most successful approaches was to identify a region of interest, book a venue and then publicise the event through the local media. Another productive approach was to set up stands at agricultural shows and recruit volunteers who came to the stand. Again the local media were invaluable in helping to publicise the project. Organisations such as the Women's Institute, Rotary International, local Family History Societies and the Church of England were also very helpful and talks given at local meetings often led to recruitment of volunteers. A number of samples were also collected remotely by sending to the volunteer the forms, vacutainer collection tubes and needle, specific tubes for transporting vacutainers containing blood and a return envelope. The volunteer then went to his/her GP and the GP took consent, took a blood samples and the volunteer then returned the sample in the post.

1.

During the period of sample collection, consent for genotyping has broadened. After an initial collection period (2,608 samples), consent was extended to allow for whole genome sequencing (about 1,400 additional samples), with the further subsequent extension to allow the remaining samples to be used as part of the 1,000 Genome Project. There are plans to continue recruiting on a small scale from all regions as and when individuals volunteer and, more specifically, samples will be collected from the west coast of Scotland, which is currently underrepresented.

3.

Samples were transported at room temperature within two days of collection to the laboratory. Peripheral blood lymphocytes were separated on a Ficoll-Hypaque gradient (Lymphoprep, Nycomed AS., Oslo, Norway) under sterile conditions[32]. If samples arrived three or four days after the collection, 0.5ml RosetteSep (Stem Cell

Technologies SARL, Grenoble, France) was added to enhance collection of viable PBLs. The lymphocyte layer was harvested and washed twice in RPMI 1640 and the PBL pellets were resuspended in 1-2ml 10% DMSO (in Foetal Bovine Serum) and then frozen either on dry ice or in a -80^{o}C freezer, before being transferred to storage under liquid Nitrogen.

DNA was prepared from the 10ml of blood residue remaining after sterile separation. using either magnetic beads (GeneCatcher, Invitrogen) or spin columns (either QIAamp DNA Blood Maxi Kit, Qiagen or Macherey-Nagel Nucleospin Blood XL, Thermo Fisher Scientific). The DNA concentration was determined using PICO Green[33]. Sets of DNA samples were normalised for genotyping to 25ng/ul using PICO Green.

Samples

The birthplaces provided with the samples were geocoded in a series of steps. First, all the place names were standardised by correcting misspellings and formats, and assigning to each a unique identifier. Standardised places of birth were then linked to an Ordnance Survey place name Gazetteer, which gives the latitude and longitude of each place name, using a flexible text string matching procedure through a series of database queries. In cases where more than one match was found, the correct assignment was confirmed manually, using contextual information on the county or area of sample collection. In cases where no matches were found the place name was searched in an alternative Gazetteer provided by <http://geonames.org>. If it was then still not possible to geocode by using the gazetteers, Google Maps (maps.google.co.uk) was used, utilising the available LatLng Tool Tip, which informs the user of the latitude and longitude of the location the cursor is pointing to. Finally, a table was created in which all place names had an assigned latitude and longitude using geographical coordinates as the unit of measurement.

From these coordinates, the mean distance (MD) between the known grandparental birthplaces of each volunteer who gave details of all four grandparents was calculated using the haversine formula (a method for calculating great-circle distances between two points on a sphere from their latitudes and longitudes, www.movable-type.co.uk/scripts/gis-faq-5.1.html). This is an approximation as the Earth is not perfectly spherical but within a region the size of the UK, errors are negligible (www.movable-type.co.uk/scripts/gis-faq-5.1.html).

Genotyping

The samples were genotyped by three methods – a customised Illumina GoldenGate assay, a multiplex assay using the Sequenom MassARRAY platform, both following the manufacturer's protocol, and an in-house ARMS-PCR method[34], detected by AMDI[35]. *HLA* was typed exclusively by an in-house method, based on the 12th International Histocompatibility Workshop, and the most common European alleles of *HLA-A*, *-B*, *-C*, *-DRB1*, *-DQB1*[36] were typed at low-medium resolution (Table 2, Supplementary Table 1).

Assessment of allele frequency differences and calculation of F_{ST}

Fisher's exact test was used to assess allele frequency differences using 2x2 tables of allele counts to split the data in three ways: 1) by geographic region, 2) by geographic region restricted to samples with either local or non-local surnames as described above and 3) within geographic regions, comparing counts of alleles between local and non-local surnames for that locality. In addition, for each of the groups defined in 1 to 3 above, F_{ST} was calculated using Weir and Cockerham's method[30], as implemented in the R package Geneland, Version 3.1.5[37]. All analyses were performed using R Version 2.9.1[38]. *HLA* haplotypes were inferred using PHASE Version 2.1[39, 40].

Admixture

In order to investigate further signals of fine scale population structure within the UK, point estimates of admixture were calculated using a maximum likelihood approach[31]. For these analyses, it was assumed that the Orkney population was a mixture of ancient British and Norse contributions, while the CN geographic populations of Oxfordshire and the Forest of Dean were assumed to be admixtures of ancient British and Anglo-Saxon[10]. The SW geographical populations were used as proxies for ancient British, and the E populations as a proxy for Anglo-Saxon ancestry. Norwegian or Swedish genetic data from the literature[24, 26, 36] were used to represent the Norse contribution.

SUPPLEMENTARY REFERENCES

32 Bøyum A: Separation of leucocytes from blood and bone marrow. *Scand J Clin Lab Invest* 1968; **21:** suppl. 97.

33 Singer VL, Jones LJ, Yue ST, Haugland RP: Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. *Anal Biochem* 1997; **249:** 228-238.

34 Tonks S, Marsh S, Bunce M, Bodmer JG: Molecular typing for HLA class I using ARMS-PCR: further development following the 12th International Histocompatibility Workshop. *Tissue Antigens* 1999; **53:** 175–183.

35 Bartlett S, Straub J, Tonks S, Wells RS, Bodmer JG, Bodmer WF: Alkaline-mediated differential interaction (AMDI): a simple automatable single-nucleotide polymorphism assay. *Proc Natl Acad Sci U S A* 2001; **98:** 2694–2697.

36 Middleton D, Menchaca L, Rood H, Komerofsky R: New allele frequency database: http://www.allelefrequencies.net. *Tissue Antigens* 2003; **61:** 403-407.

37 Guillot G, Mortier F, Estoup A: Geneland: A program for landscape genetics. *Molecular Ecology Notes* 2005; **5:** 712-715.

38 R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2009 [Available from http://www.R-project.org].

39 Stephens M, Smith N, Donnelly P: A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68:** 978-989.

40 Stephens M, Scheet P: Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *Am J Hum Genet* 2005; **76:** 449-462.

	HLA-A																
	01	02	03	11	23	24	25	26	29	30	31	32	33	34	68	69	2n
SW	0.1687	0.3067	0.1564	0.0552	0.0184	0.0706	0.0245	0.0092	0.0153	0.0276	0.0307	0.0583	0.0092	0.0000	0.0307	0.0184	326
CN	0.1910	0.2697	0.1573	0.0712	0.0150	0.0412	0.0112	0.0449	0.0412	0.0225	0.0337	0.0375	0.0037	0.0000	0.0524	0.0075	267
E	0.1718	0.3037	0.1748	0.0521	0.0123	0.0828	0.0092	0.0245	0.0552	0.0215	0.0153	0.0429	0.0031	0.0031	0.0153	0.0123	326
N	0.1770	0.2708	0.1513	0.0756	0.0166	0.0938	0.0182	0.0121	0.0363	0.0348	0.0303	0.0318	0.0091	0.0015	0.0287	0.0121	661
Ork	0.1829	0.2914	0.0914	0.0800	0.0000	0.0971	0.0343	0.0286	0.0743	0.0229	0.0114	0.0457	0.0000	0.0000	0.0229	0.0171	175
	HLA-B																
	07	*08*	*13*	*14*	*15*	*18*	*27*	*35*	*37*	*38*	*39*	*40*	*41*	*42*	*44*	*45*	*47*
SW	0.1511	0.1704	0.0032	0.0675	0.0772	0.0354	0.0740	0.0707	0.0096	0.0032	0.0257	0.0322	0.0064	0.0000	0.1061	0.0000	0.0032
CN	0.1391	0.1466	0.0000	0.0376	0.0526	0.0451	0.0677	0.0902	0.0188	0.0150	0.0075	0.0714	0.0000	0.0038	0.1090	0.0038	0.0038
E	0.1335	0.1304	0.0155	0.0311	0.1025	0.0435	0.0559	0.1025	0.0093	0.0155	0.0031	0.0683	0.0031	0.0000	0.0901	0.0186	0.0000
N	0.1859	0.1045	0.0169	0.0276	0.0599	0.0522	0.0522	0.0814	0.0061	0.0077	0.0154	0.0353	0.0061	0.0215	0.1244	0.0046	0.0000
Ork	0.2216	0.0898	0.0299	0.0359	0.0838	0.0419	0.0240	0.0299	0.0060	0.0180	0.0240	0.0599	0.0000	0.0000	0.2036	0.0060	0.0000
	49	*50*	*51*	*52*	*53*	*54*	*55*	*57*	*58*	*59*	*78*	*2n*					
SW	0.0032	0.0129	0.0418	0.0225	0.0096	0.0064	0.0032	0.0450	0.0161	0.0032	0.0000	311					
CN	0.0038	0.0188	0.0564	0.0075	0.0188	0.0150	0.0113	0.0414	0.0075	0.0000	0.0075	266					
E	0.0124	0.0093	0.0683	0.0031	0.0186	0.0124	0.0155	0.0311	0.0000	0.0031	0.0031	322					
N	0.0230	0.0138	0.0538	0.0108	0.0154	0.0092	0.0123	0.0507	0.0015	0.0031	0.0046	651					
Ork	0.0120	0.0000	0.0000	0.0120	0.0000	0.0060	0.0240	0.0659	0.0000	0.0000	0.0060	167					
	HLA-C																
	01	*02*	*03*	*04*	*05*	*06*	*07*	*08*	*12*	*14*	*15*	*16*	*17*	*2n*			
SW	0.0184	0.0429	0.1227	0.0767	0.1411	0.0828	0.3681	0.0644	0.0245	0.0061	0.0215	0.0215	0.0092	326			
CN	0.0333	0.0259	0.1667	0.0815	0.1037	0.0889	0.3333	0.0444	0.0444	0.0111	0.0370	0.0296	0.0000	270			
E	0.0307	0.0583	0.1748	0.0583	0.0890	0.0920	0.3160	0.0307	0.0429	0.0153	0.0215	0.0583	0.0123	326			
N	0.0270	0.0375	0.1306	0.0691	0.1216	0.0871	0.3664	0.0330	0.0390	0.0120	0.0210	0.0405	0.0150	666			
Ork	0.0170	0.0170	0.1705	0.0341	0.1136	0.0909	0.3807	0.0511	0.0568	0.0114	0.0000	0.0568	0.0000	176			

	HLA-DRB1													
	01	*03*	*04*	*07*	*08*	*09*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*2n*
SW	0.1059	0.1558	0.2399	0.0997	0.0249	0.0156	0.0062	0.0654	0.0187	0.1090	0.0343	0.1215	0.0031	321
CN	0.1053	0.1654	0.1729	0.1015	0.0376	0.0188	0.0075	0.0940	0.0113	0.1090	0.0301	0.1316	0.0150	266
E	0.0786	0.1415	0.1950	0.1447	0.0409	0.0157	0.0000	0.0849	0.0220	0.1006	0.0346	0.1226	0.0189	318
N	0.1162	0.1437	0.1774	0.1468	0.0382	0.0153	0.0015	0.0550	0.0260	0.0902	0.0245	0.1636	0.0015	654
Ork	0.0613	0.1411	0.1595	0.1840	0.0368	0.0184	0.0000	0.0429	0.0184	0.0859	0.0307	0.2086	0.0123	163
	HLA-DQB1													
	02	*03*	*04*	*05*	*06*	*2n*								
SW	0.2250	0.3375	0.0281	0.1844	0.2250	320								
CN	0.2293	0.3421	0.0188	0.1504	0.2594	266								
E	0.2803	0.3408	0.0287	0.1369	0.2134	314								
N	0.2623	0.3083	0.0261	0.1503	0.2531	652								
Ork	0.2674	0.3198	0.0291	0.1105	0.2733	172								

Supplementary Table 1. Complete set of *HLA* allele frequencies. Populations are grouped into regions as defined in the main text.

		All samples			Pilot samples		
Pop. size	*Distance away (km)*	*Rural G'parents*	*4 G'parents*	*3 G'parents*	*Rural G'parents*	*4 G'parents*	*3 G'parents*
20000	2	0.541	0.332	0.137	0.536	0.317	0.145
30000	2	0.614	0.420	0.139	0.606	0.400	0.153
40000	2	0.646	0.455	0.140	0.624	0.426	0.146
50000	2	0.668	0.484	0.138	0.642	0.452	0.139
75000	2	0.719	0.551	0.130	0.670	0.489	0.140
100000	2	0.741	0.580	0.129	0.698	0.525	0.138
125000	2	0.757	0.599	0.128	0.719	0.555	0.130
150000	2	0.805	0.662	0.123	0.814	0.664	0.135
200000	2	0.837	0.711	0.113	0.855	0.728	0.119
250000	2	0.847	0.724	0.111	0.872	0.757	0.112
300000	2	0.859	0.743	0.108	0.880	0.772	0.104
20000	5	0.473	0.272	0.126	0.469	0.245	0.147
30000	5	0.571	0.376	0.139	0.565	0.347	0.161
40000	5	0.610	0.414	0.144	0.587	0.376	0.156
50000	5	0.640	0.447	0.146	0.613	0.408	0.153
75000	5	0.699	0.526	0.136	0.646	0.452	0.151
100000	5	0.724	0.553	0.137	0.676	0.487	0.149
125000	5	0.743	0.577	0.136	0.703	0.525	0.144
150000	5	0.795	0.647	0.128	0.805	0.648	0.137
200000	5	0.831	0.699	0.118	0.849	0.717	0.118
250000	5	0.841	0.715	0.115	0.868	0.751	0.109
300000	5	0.855	0.735	0.111	0.876	0.766	0.102
20000	10	0.375	0.198	0.100	0.369	0.168	0.134
30000	10	0.514	0.322	0.127	0.508	0.290	0.160
40000	10	0.564	0.365	0.140	0.538	0.325	0.159
50000	10	0.603	0.405	0.143	0.572	0.364	0.154
75000	10	0.675	0.498	0.138	0.612	0.414	0.154
100000	10	0.703	0.528	0.141	0.642	0.447	0.158
125000	10	0.726	0.555	0.140	0.676	0.490	0.156
150000	10	0.785	0.635	0.129	0.790	0.625	0.145
200000	10	0.824	0.690	0.120	0.837	0.698	0.128
250000	10	0.837	0.708	0.118	0.861	0.737	0.119
300000	10	0.851	0.729	0.113	0.870	0.753	0.113

Supplementary Table 2. Proportion of all grandparents classed as rural according to their distance (2km, 5km or 10km) from an urban area of a given population size. Estimates given separately for all the 3,865 geocoded samples (All samples) and those that were genotyped (Pilot samples). Estimates were made separately for all grandparents, as well as those who had either 3 or 4 rural grandparents.

Population	*LQ>19, Dist<83km*	*LQ>19, dist<120km*	*LQ>19*	*LQ>45, dist<83km*	*Proportion LQ>45, dist<120km*	*LQ>45*	*LQ>120, dist<83km*	*LQ>120, dist<120km*	*LQ>120*	*LQ>200*	*LQ>300*
Cornwall	0.083	0.100	0.133	0.000	0.017	0.050	0.000	0.000	0.000	0.000	0.000
Cumbria	0.086	0.103	0.121	0.069	0.086	0.086	0.000	0.000	0.000	0.000	0.000
Devon	0.063	0.076	0.101	0.025	0.025	0.051	0.000	0.000	0.025	0.000	0.000
Forest of Dean	0.075	0.164	0.269	0.060	0.090	0.104	0.015	0.030	0.030	0.030	0.000
Kent/Sussex	0.082	0.122	0.163	0.082	0.102	0.102	0.020	0.061	0.061	0.061	0.020
Lincolnshire	0.033	0.100	0.133	0.033	0.067	0.100	0.000	0.033	0.067	0.033	0.000
North East	0.088	0.132	0.140	0.051	0.059	0.066	0.000	0.007	0.007	0.000	0.000
Norfolk	0.050	0.060	0.060	0.040	0.050	0.050	0.030	0.030	0.030	0.010	0.000
Pembrokeshire	0.231	0.333	0.385	0.026	0.026	0.026	0.026	0.026	0.026	0.000	0.000
Oxfordshire	0.051	0.076	0.076	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Yorkshire	0.055	0.076	0.097	0.014	0.028	0.048	0.000	0.000	0.000	0.000	0.000
All populations	0.075	0.109	0.134	0.034	0.046	0.058	0.007	0.013	0.017	0.008	0.001

Supplementary Table 3. Proportion of surnames excluded because of multiple observed peaks or broad distributions. The numbers are incorporated in the proportions classified as local in Table 3.

ABO (rs7853989)

	SW	*N*	*CN*	*E*	*OR*
SW	0	-0.0013	0.0062	0.0028	0.0590
N	-0.0013	0	0.0031	0.0006	0.0532
CN	0.0062	0.0031	0	-0.0022	0.0191
E	0.0028	0.0006	-0.0022	0	0.0269
OR	0.0590	0.0532	0.0191	0.0269	0

HLA-B

	SW	*N*	*CN*	*E*	*OR*
SW	0	0.0032	-0.0001	0.0006	0.0137
N	0.0032	0	0.0027	0.0023	0.0059
CN	-0.0001	0.0027	0	-0.0006	0.0154
E	0.0006	0.0023	-0.0006	0	0.0135
OR	0.0137	0.0059	0.0154	0.0135	0

NRY

	SW	*N*	*CN*	*E*	*OR*
SW	0	0.0128	0.0095	0.0204	0.1441
N	0.0128	0	-0.0072	-0.0017	0.0976
CN	0.0095	-0.0072	0	-0.0072	0.0829
E	0.0204	-0.0017	-0.0072	0	0.0820
OR	0.1441	0.0976	0.0829	0.0820	0

Supplementary Table 4. Pairwise F_{ST} estimates for the 4 regions defined in the main text along with the Orcadian samples (OR). F_{ST} estimates have been calculated for each individual locus and the three loci shown here are those that gave some consistently high estimates.

Autosomes

No stratification

	SW	N	CN	E
SW	0	0.0018	0.0003	0.0010
N	0.0018	0	0.0011	0.0007
CN	0.0003	0.0011	0	-0.0004
E	0.0010	0.0007	-0.0004	0

Local Surnames *Non-local Surnames*

LQ=19,distance>83km

	SW	*N*	*CN*	*E*
SW	0	0.0004	-0.0006	0.0035
N	0.0004	0	0.0026	0.0010
CN	-0.0006	0.0026	0	-0.0014
E	0.0035	0.0010	-0.0014	0

	SW	*N*	*CN*	*E*
SW	0	0.0031	0.0020	0.0004
N	0.0031	0	0.0007	0.0017
CN	0.0020	0.0007	0	-0.0004
E	0.0004	0.0017	-0.0004	0

LQ=45, distance>83km

	SW	*N*	*CN*	*E*
SW	0	-0.0002	-0.0019	0.0021
N	-0.0002	0	0.0018	0.0000
CN	-0.0019	0.0018	0	-0.0022
E	0.0021	0.0000	-0.0022	0

	SW	*N*	*CN*	*E*
SW	0	0.0024	0.0021	0.0002
N	0.0024	0	0.0003	0.0016
CN	0.0021	0.0003	0	-0.0001
E	0.0002	0.0016	-0.0001	0

LQ=120, distance>83km

	SW	*N*	*CN*	*E*
SW	0	-0.0023	-0.0029	0.0033
N	-0.0023	0	-0.0010	-0.0019
CN	-0.0029	-0.0010	0	-0.0075
E	0.0033	-0.0019	-0.0075	0

	SW	*N*	*CN*	*E*
SW	0	0.0020	0.0012	0.0002
N	0.0020	0	0.0004	0.0014
CN	0.0012	0.0004	0	-0.0003
E	0.0002	0.0014	-0.0003	0

LQ=19, distance>120km

	SW	*N*	*CN*	*E*
SW	0	-0.0003	0.0006	0.0027
N	-0.0003	0	0.0026	0.0017
CN	0.0006	0.0026	0	-0.0011
E	0.0027	0.0017	-0.0011	0

	SW	*N*	*CN*	*E*
SW	0	0.0028	0.0016	0.0004
N	0.0028	0	0.0004	0.0009
CN	0.0016	0.0004	0	-0.0008
E	0.0004	0.0009	-0.0008	0

NRY

No stratification

	SW	N	CN	E
SW	0	0.0128	0.0095	**0.0204**
N	0.0128	0	-0.0072	-0.0017
CN	0.0095	-0.0072	0	-0.0072
E	**0.0204**	-0.0017	-0.0072	0

Local Surnames *Non-local Surnames*

LQ=19, distance>83km

	SW	*N*	*CN*	*E*
SW	0	-0.0096	**0.0253**	0.0162
N	-0.0096	0	0.0078	-0.0043
CN	**0.0253**	0.0078	0	-0.0314
E	0.0162	-0.0043	-0.0314	0

	SW	*N*	*CN*	*E*
SW	0	-0.0032	-0.0022	0.0071
N	-0.0032	0	-0.0056	-0.0087
CN	-0.0022	-0.0056	0	-0.0104
E	0.0071	-0.0087	-0.0104	0

LQ=45, distance>83km

	SW	*N*	*CN*	*E*
SW	0	-0.0192	0.0058	-0.0143
N	-0.0192	0	**0.0754**	0.0077
CN	0.0058	**0.0754**	0	-0.0110
E	-0.0143	0.0077	-0.0110	0

	SW	*N*	*CN*	*E*
SW	0	0.0097	0.0003	0.0197
N	0.0097	0	-0.0020	-0.0102
CN	0.0003	-0.0020	0	-0.0073
E	0.0197	-0.0102	-0.0073	0

LQ=120, distance>83km

	SW	*N*	*CN*	*E*
SW	0	**0.0450**	**0.0382**	-0.0098
N	**0.0450**	0	**0.2815**	**0.1051**
CN	**0.0382**	**0.2815**	0	-0.0078
E	-0.0098	**0.1051**	-0.0078	0

	SW	*N*	*CN*	*E*
SW	0	0.0099	-0.0028	0.0159
N	0.0099	0	0.0021	-0.0094
CN	-0.0028	0.0021	0	-0.0010
E	0.0159	-0.0094	-0.0010	0

LQ=19, distance>120km

	SW	*N*	*CN*	*E*
SW	0	-0.0161	-0.0072	0.0138
N	-0.0161	0	-0.0088	0.0037
CN	-0.0072	-0.0088	0	-0.0302
E	0.0138	0.0037	-0.0302	0

	SW	*N*	*CN*	*E*
SW	0	0.0009	-0.0059	0.0076
N	0.0009	0	-0.0055	-0.0098
CN	-0.0059	-0.0055	0	-0.0143
E	0.0076	-0.0098	-0.0143	0

LQ=45, distance>120km

	SW	N	CN	E
SW	0	-0.0008	-0.0014	0.0021
N	-0.0008	0	0.0010	-0.0003
CN	-0.0014	0.0010	0	-0.0021
E	0.0021	-0.0003	-0.0021	0

	SW	N	CN	E
SW	0	0.0021	0.0012	0.0001
N	0.0021	0	0.0004	0.0011
CN	0.0012	0.0004	0	-0.0009
E	0.0001	0.0011	-0.0009	0

LQ=45, distance>120km

	SW	N	CN	E
SW	0	-0.0165	0.0137	-0.0144
N	-0.0165	0	**0.0878**	0.0120
CN	0.0137	**0.0878**	0	-0.0141
E	-0.0144	0.0120	-0.0141	0

	SW	N	CN	E
SW	0	0.0144	-0.0046	**0.0243**
N	0.0144	0	0.0005	-0.0108
CN	-0.0046	0.0005	0	-0.0037
E	**0.0243**	-0.0108	-0.0037	0

LQ=120, distance>120km

	SW	N	CN	E
SW	0	-0.0028	-0.0007	0.0031
N	-0.0028	0	0.0000	-0.0020
CN	-0.0007	0.0000	0	-0.0028
E	0.0031	-0.0020	-0.0028	0

	SW	N	CN	E
SW	0	0.0020	0.0007	0.0003
N	0.0020	0	0.0004	0.0014
CN	0.0007	0.0004	0	-0.0007
E	0.0003	0.0014	-0.0007	0

LQ=120, distance>120km

	SW	N	CN	E
SW	0	**0.0421**	-0.0142	-0.0154
N	**0.0421**	0	**0.1375**	**0.0766**
CN	-0.0142	**0.1375**	0	-0.0374
E	-0.01540449	**0.0766**	-0.0374	0

	SW	N	CN	E
SW	0	0.0149	-0.0029	**0.0208**
N	0.0149	0	0.0006	-0.0102
CN	-0.0029	0.0006	0	-0.0063
E	**0.0208**	-0.0102	-0.0063	0

LQ=19

	SW	N	CN	E
SW	0	0.0010	0.0031	0.0025
N	0.0010	0	0.0036	0.0024
CN	0.0031	0.0036	0	0.0009
E	0.0025	0.0024	0.0009	0

	SW	N	CN	E
SW	0	0.0017	0.0009	0.0007
N	0.0017	0	0.0000	-0.0019
CN	0.0009	0.0000	0	-0.0018
E	0.0007	-0.0019	-0.0018	0

LQ=19

	SW	N	CN	E
SW	0	-0.0032	0.0010	0.0119
N	-0.0032	0	0.0057	-0.0035
CN	0.0010	0.0057	0	-0.0045
E	0.0119	-0.0035	-0.0045	0

	SW	N	CN	E
SW	0	-0.0055	-0.0069	0.0089
N	-0.0055	0	-0.0066	-0.0121
CN	-0.0069	-0.0066	0	-0.0181
E	0.0089	-0.0121	-0.0181	0

LQ=45

	SW	N	CN	E
SW	0	0.0026	0.0051	0.0032
N	0.0026	0	0.0021	0.0002
CN	0.0051	0.0021	0	0.0007
E	0.0032	0.0002	0.0007	0

	SW	N	CN	E
SW	0	0.0021	0.0008	0.0003
N	0.0021	0	0.0000	0.0012
CN	0.0008	0.0000	0	-0.0006
E	0.0003	0.0012	-0.0006	0

LQ=45

	SW	N	CN	E
SW	0	-0.0072	**0.0322**	0.0074
N	-0.0072	0	**0.0467**	0.0022
CN	**0.0322**	**0.0467**	0	-0.0018
E	0.0074	0.0022	-0.0018	0

	SW	N	CN	E
SW	0	-0.0002	-0.0062	0.0120
N	-0.0002	0	-0.0040	-0.0097
CN	-0.0062	-0.0040	0	-0.0021
E	0.0120	-0.0097	-0.0021	0

LQ=120

	SW	N	CN	E
SW	0	-0.0006	0.0054	0.0047
N	-0.0006	0	0.0007	-0.0013
CN	0.0054	0.0007	0	0.0009
E	0.0047	-0.0013	0.0009	0

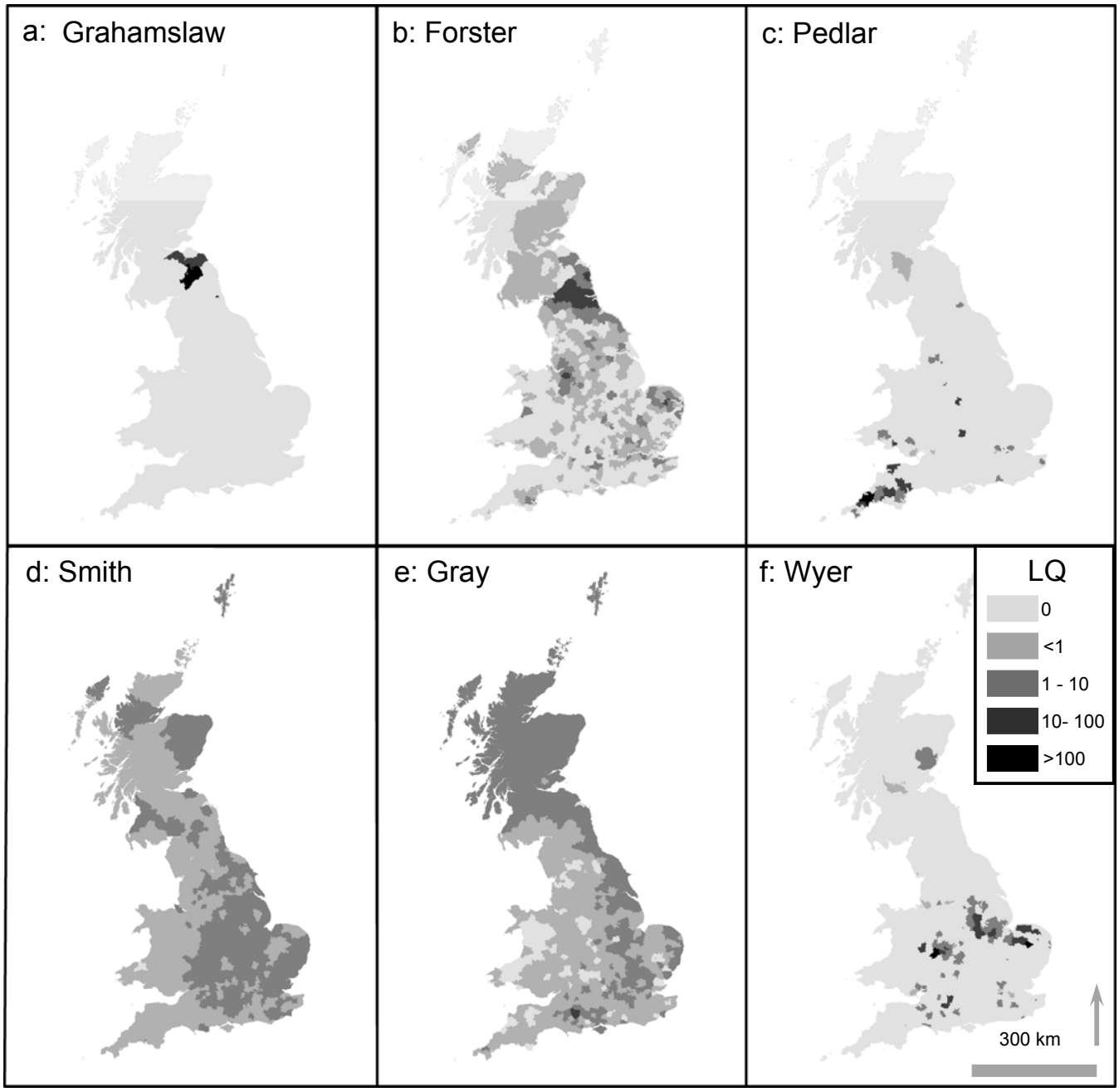
	SW	N	CN	E
SW	0	0.0018	0.0004	0.0003
N	0.0018	0	0.0006	0.0014
CN	0.0004	0.0006	0	-0.0007
E	0.0003	0.0014	-0.0007	0

LQ=120

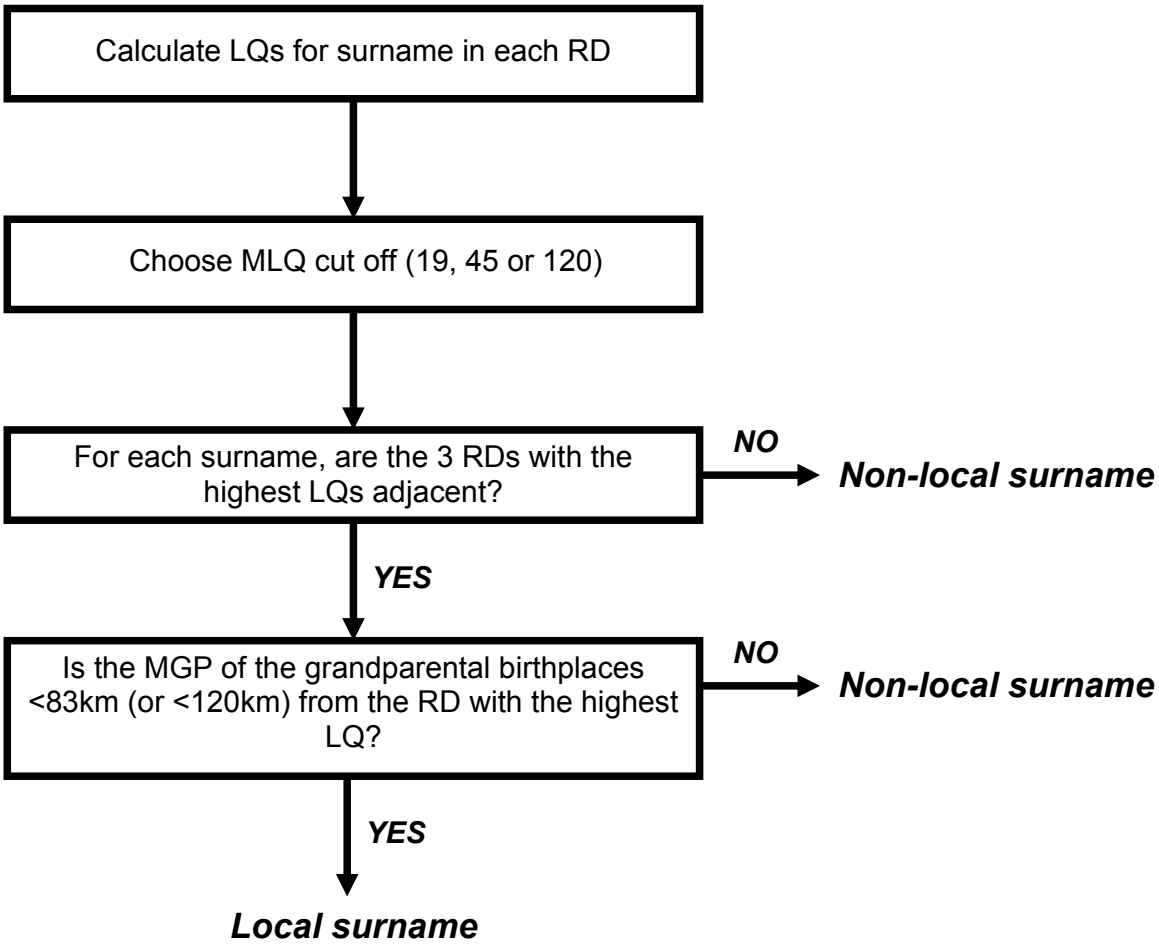
	SW	N	CN	E
SW	0	**0.0421**	-0.0142	-0.0154
N	**0.0421**	0	**0.1375**	**0.0766**
CN	-0.0142	**0.1375**	0	-0.0374
E	-0.0154	**0.0766**	-0.0374	0

	SW	N	CN	E
SW	0	0.0149	-0.0029	**0.0208**
N	0.0149	0	0.0006	-0.0102
CN	-0.0029	0.0006	0	-0.0063
E	**0.0208**	-0.0102	-0.0063	0

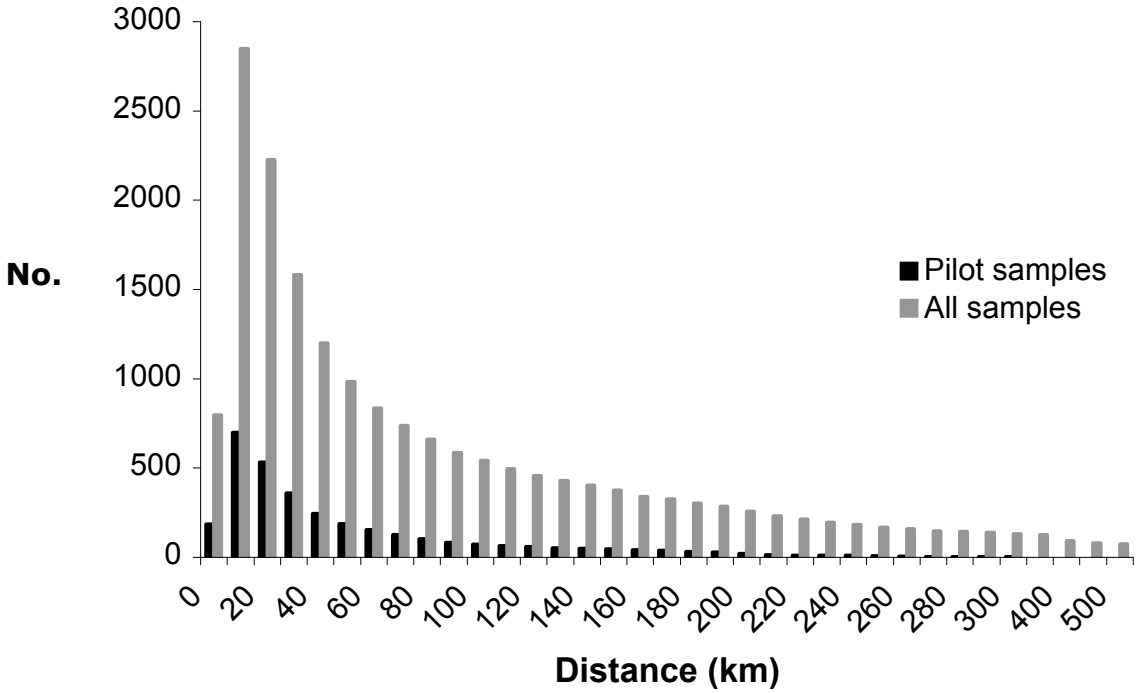
Supplementary Table 5. F_{ST} estimated pairwise between regions for autosomal and NRY data. F_{ST}s have been calculated with no stratification as well as stratification by local or non-local surnames. Surnames were classified as local depending on different exclusion criteria. The two main criteria were a minimum LQ of the district with the highest LQ and maximum distance of the MGP from that district for each sample. Values highlighted in bold are >0.02 and are only found in the NRY data.



Supplementary Figure 1. Location Quotient (LQ) distributions for a selection of surnames. Shade represents LQ level in the district. Surnames a) - c) are examples of localised surnames, d) and e) are widespread. f) is an example of a surname that has a Registration District (RD) with a very high LQ but also has two further non-adjacent RDs with similarly high LQs.



Supplementary Figure 2. Flow diagram for method used to classify a surname as either local or non-local.



Supplementary Figure 3. Graph of the distribution of the MD between grandparental birthplaces for each sample. Distributions are given for the complete set of geocoded samples (All samples) and those genotyped (Pilot samples).

Supplementary Figure 4. Geographic frequency of allele and haplotype distributions. Frequencies are shown for the populations genotyped (Figure 2). Shading goes from darker to lighter as the frequency of the haplotype (*NRY*, *HLA*) or minor allele (*MC1R* and *ABO* SNPs) decreases. Frequencies are from Table 2.