Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome.

SI Appendix

Lara M. Cassidy[a,1], Rui Martiniano[a,1], Eileen M. Murphy[b], Matthew D. Teasdale[a], Jim Mallory[b], Barrie Hartwell[b], Daniel G Bradley[a,2]

[a]Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland
[b]School of Geography, Archaeology & Palaeoecology, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland
[1]LMC and RM contributed equally to this work
[2]To whom correspondence should be addressed. Email: dbradley@tcd.ie

Contents

S1: Archaeological Context

S1.1 Ballynahatty

The individual from Ballynahatty (A.64) was a female with an age-at-death of 18-35 years (1). Her remains were discovered in 1855 and comprised a near-complete skull recovered from the interior of a subterranean stone burial chamber discovered in a field to the north of the well-known henge monument, the Giant's Ring. The burial structure was sub-circular and its interior had been divided into a series of compartments – the skull was recovered from Compartment D and was associated with a collection of disarticulated bone and fragments of cremated bone. Radiocarbon dating of the cranium produced a result of 3343-3020 cal BC (UB-7059; 4465+/- 38 BP) at 95% probability (2), which places the individual into the later Middle Neolithic, or possibly Late Neolithic, periods (3). Her burial context, based on 19th century drawings and reports, is unusual in that its morphology does not readily fit within any of the recognised megalithic tomb traditions, although is most similar to passage grave structures.

S1.2 Rathlin Island

The three individuals from Rathlin Island were discovered during excavations in Church Bay in 2006 (4). The style of burial, the associated pottery and radiocarbon dates derived from the three individuals (Table S1.1) indicated that they were Early Bronze Age in date. C127 was a 40-60 year old male who had been buried in the interior of a stone cist (C124). He was lying on his left side in a crouched position along a north-south orientation, with the head to the north. He was a robust individual who would have had a height of approximately 5'11". His bones revealed signs of an active lifestyle and osteoarthritis, Schmorl's nodes and torsion of a number of lumbar spinous processes were apparent in his vertebrae, while os acromiale was visible in his left scapula. The tuber of his left calcaneus displayed healed lytic lesions which may have been caused by a soft tissue injury of his heel at some stage during his life (5). The remains of a complete tripartite bowl Food Vessel were recovered adjacent to his lower legs in the south-east corner of the cist (6). C122 (Sk 1) and C105 (Sk 2) were both young adult males whose disarticulated remains were recovered in association with a disturbed cist (C112) adjacent to the complete cist (C124). Disarticulated bone from an adult female was also recovered from the deposit. Signs of infection or inflammation were visible in the right parietal and mastoid region of the cranium of C122 (Sk 1) and in the right scapula of C105 (Sk 2) (5). The practice of using a single cist for the interment of

multiple individuals was also identified during excavations on Rathlin during the 1980s when Kenny Wiggins discovered a cist that contained the remains of at least five individuals (Grave 2). Interestingly, the radiocarbon dates of C122 (Sk 1) and C105 (Sk 2) revealed that it would have been impossible for them to have been buried at the same time. As such, it is possible that the cist had been re-opened at some stage to facilitate further burials or that one of the individuals was an 'ancestor' whose remains had been curated for a period of time prior to their burial in the cist (4).

Table S1.1 Radiocarbon dates and archaeological contexts for four ancient Irish individuals.

Site	Archaeological ID	Lab ID	Sample ID	C14 Lab Code	Radiocarbon Age BP	Calibrated date (95% probability) (Schulting et al. 2011, Reimer et al. 2004)	Context
Ballynahatty, Co. Down	A.64	BA64	Ballynahatty	UB-7059	4465 ± 38	3343-3020 BC	Middle to Late Neolithic
Glebe, Rathlin Island	C 127	RM127	Rathlin1	UBA-8707	3591 ± 29	2026-1885 BC	Early Bronze Age
Glebe, Rathlin Island	C 122 (Sk 1)	RSK1	Rathlin2	UBA-8705	3539 ± 54	2024-1741 BC	Early Bronze Age
Glebe, Rathlin Island	C 105 (Sk 2)	RSK2	Rathlin3	UBA-8706	3354 ± 28	1736-1534 BC	Early Bronze Age

S2: Sampling and Extraction

The petrous portion of the temporal bone from was obtained from the four ancient individuals. For two of these samples, Ballynahatty (A.64) and Rathlin1 (C127), the petrous was detached from the intact cranium using a dremel diamond wheel attached to a dental drill. This sampling took place in the School of Geography, Archaeology and Palaeoecology at Queen's University Belfast. The petrous bones were then transported to the Department of Genetics, TCD whereby all further sample preparation took place in facilities dedicated to ancient DNA analysis. Full body suits, face masks, gloves and hairnets were worn throughout the drilling, extraction and library creation processes. All tools and surfaces were consecutively cleaned with bleach, DNA-ExitusPlus, Ethanol and long exposures to UV light.

The samples were UV-irradiated for 20 minutes on each side. The surface of each bone was removed using a standard dentistry drill with a dremel engraving cutter attached. Fine powder was subsequently obtained from each sample with a Mixer Mill (MM 400, Retsch). DNA was extracted from approximately 150 mg of the sample bone powder produced, following the protocol outlined in (7) with modifications introduced by (8). An extraction control (1ml ddH_20) was carried throughout the procedure to keep account of possible modern exogenous DNA contamination and sample cross contamination.

Bone powder was placed in 1ml of lysis buffer (20 mM Tris HCL, pH 7.4; 47.5 mM EDTA, pH 8; 0.7% Sarkosyl NL30; 0.65 U/ml Proteinase K, recombinant, PCR Grade) and rotated at 300rpm for 24 hours at 55°C followed by 24 hours at 37°C. The samples were then spun at 10,000rpm for 4 mins and the supernatant was removed. A following 1ml of lysis buffer was added to the remaining pellets and the incubation step was repeated. After incubation the samples were centrifuged at 13,000rpm for 12 mins and the supernatant of each was transferred to an Amicon Ultra-4 Centrifugal Filter Unit 30kDA. Each sample was diluted in 3ml of 10mM Tris-EDTA Buffer and centrifuged at 2,500rpm until 250µl of the solution remained (12-20mins). The resulting flow through was discarded and another 3ml of 10mM Tris-EDTA Buffer was added to the filter. The samples were then spun at 2,500rpm until a final volume of 100µl was obtained (15-25mins).

Each extract was transferred from the filter to a silica column (MinElute PCR Purification Kit, Qiagen, Hilden, Germany) and purified following the manufacturer's instructions. Briefly, PB buffer was added to the silica column at approximately 5 times the volume of the extract. After the sample was spun at 13,000rpm for 1 minute and the flow through discarded, 750µl of PE buffer was added to the column. The sample was spun at 13,000rpm for a subsequent two minutes or until there was no visible liquid left.

Elution buffer (EB) at volume of 35µl was added to the column membrane and the sample DNA released from the filter by centrifugation at 13,000rpm for 1 minute into a fresh 1.5ml centrifuge tube.

S3: Library Creation and Sequencing

Libraries were prepared using 30µl of DNA extract following the method outlined in (9) with the following modifications as in (10): no DNA shearing by sonication was carried out due to the highly fragmented nature of ancient DNA molecules. All DNA purification steps were performed using the MinElute PCR Purification kit protocol as described above. Blunt end repair was done using NEBNext End Repair Module (New England BioLab Inc.), with all components scaled to 70% of the manufacturer's recommended volume, giving a final reaction volume of 70µl. This volume was then incubated for 15 min at 25°C followed by 5 min at 12°C. The final modification was the termination of Bst polymerase activity after the adapter fill-in reaction through the addition of a heat inactivation step (80°C for 20 min).

The library amplification reaction was set up for all samples and controls using Accuprime Pfx Supermix (Life Technology), primer IS4 (0.2µM), a specific indexing primer (0.2µM) and 3µl of sample library, with a total reaction volume of 25µl. Amplification took place under the following thermal cycling conditions: 5 min at 95°C; 12 cycles of 15 sec at 95°C, 30 sec at 60°C and 30 sec at 68°C; and a final extension step of 5 min at 68°C. The resulting PCR product was purified using the MinElute PCR Purification kit protocol described previously.

Quantification and quality assessment of the amplified libraries was performed on an Agilent 2100 Bioanalyzer, using a DNA-1000 chip and following manufacturer's protocol.

One amplified library from each sample was first screened on an Illumina MiSeq platform (Trinity Genome Sequencing Laboratory, Trinity College Dublin, Ireland) using 70 bp single-end sequencing and following the manufacturer's instructions for multiplex sequencing. The results of this sequencing were used to estimate endogenous DNA contents (Main Paper, Table 1; Table S3.1). Three additional amplified libraries were then created for each sample and these were subsequently sequenced on an Illumina HiSeq 2000 Platform using 100bp single-end sequencing at Beckman Coulter Genomics, Danvers, MA. Libraries from samples Ballynahatty and Rathlin1 were sequenced on 8 lanes in total between them, while samples Rathlin2 and Rathlin3 were sequenced on one shared lane. Sequencing results are displayed in Table S3.1.

Table S3.1. Summary of sequencing and alignment statistics for all sample libraries.

Sample ID	Library ID	Sequencing Platform	Raw Reads with length >34bp after adapter trimming	Total Mapped reads	Unique Reads with mapping quality >30	Duplication	Final Endogenous Content
Ballynahatty	LIB1-10	Illumina MiSeq	1932299	1610081	1311289	0.24%	67.86%
Ballynahatty	LIB1-23	Illumina HiSeq2000	284353240	236162337	166262242	14.58%	58.47%
Ballynahatty	LIB1-24	Illumina HiSeq2000	261340137	217496196	153125411	14.61%	58.59%
Ballynahatty	LIB1-25	Illumina HiSeq2000	297458417	247335459	163489349	19.86%	54.96%
Rathlin1	LIB1-12	Illumina MiSeq	1157817	847143	703264	0.18%	60.74%
Rathlin1	LIB1-26	Illumina HiSeq2000	275487235	200501983	148289824	14.12%	53.83%
Rathlin1	LIB1-27	Illumina HiSeq2000	243046349	176890003	132415709	13.01%	54.48%
Rathlin1	LIB1-28	Illumina HiSeq2000	241777482	175970753	131720012	13.06%	54.48%
Rathlin2	LIB1-08	Illumina MiSeq	1690602	800137	644959	0.14%	38.15%
Rathlin2	LIB1-75	Illumina HiSeq2000	42786666	19812965	16165004	2.17%	37.78%
Rathlin2	LIB1-76	Illumina HiSeq2000	73697710	34320081	27744257	3.04%	37.65%
Rathlin2	LIB1-77	Illumina HiSeq2000	58266248	27150183	22020894	2.68%	37.79%
Rathlin3	LIB1-09	Illumina MiSeq	2281602	1706599	1401348	0.20%	61.42%
Rathlin3	LIB1-90	Illumina HiSeq2000	15536366	11183917	9417979	1.23%	60.62%
Rathlin3	LIB1-91	Illumina HiSeq2000	15188194	10880384	9146209	1.26%	60.22%
Rathlin3	LIB1-92	Illumina HiSeq2000	18409063	13210282	11113844	1.29%	60.37%

Four amplified libraries were created for each sample. Final endogenous content refers to the percentage human DNA in the library after filtering by mapping quality (>30) and removing PCR duplicate reads.

S4: Data Processing and Read Mapping

S4.1 Irish Sequencing Data

Only reads with an exact match to the correct index sequence were retained for analysis to guard against possible contamination. The majority of aDNA molecules are short enough that single-end reads carry part of the P7 adapter at the 3' end. These residual adapter sequences were trimmed from reads using cutadapt v1.2.1 (11), with parameters enabled to discard reads under 34bp in length after trimming (-m 34) and to allow a minimum overlap of 1bp between the read and adapter (-O 1).

All remaining trimmed sequences were subsequently mapped to the human reference genome (hg19/GRCh37) with the mitochondrial genome replaced by the revised Cambridge reference sequence (Accession number NC_012920.1). BWA version 0.7.5 (12) was used for alignment with the seed disabled (-l 16500) as recommended in (13), the edit distance set to 0.02 to allow for more substitutions (-n 0.02) and the amount of gaps allowed raised (-o 2).

Read groups were added using Picard Tools v1.101 (http://broadinstitute.github.io/picard/). Reads were sorted using Samtools v0.1.19 (14) and reads with a mapping quality below 30 were at this point discarded from all further analysis. BAM files were merged to sample level using the MergeSamFiles Picard tool and PCR duplicates removed with Samtools (14). The RealignerTargetCreator and IndelRealigner tools from GenomeAnalysisTK v2.4-7 (15) were used to locally realign reads.

DNA damage patterns were assessed using mapDamage 2.0 (16) and the base quality scores of likely deaminated positions in reads were rescaled based on these patterns, as well as their initial qualities and their position in the read. As an additional safeguard to the high frequency of misincorporation sites found at read ends, each read had the base quality of two base pairs at both its 5' and 3' ends reduced to a PHRED score of 2.

The depth of genome coverage for each sample was estimated using Qualimap v2.1.1 (17).

S4.2 Published Ancient Data

Ancient data from twelve previous studies (10, 18–28) was processed alongside the four Irish genomes for use in downstream analyses. Information regarding the ancient samples included in this study can be found in Table S9.1.

S4.2.1 Skoglund et al. (2014), Lazaridis et al. (2014), Allentoft et al. (2015) & Günther et al. (2015)

BAM files were obtained from (19, 21, 26, 27) and realigned to the human reference genome (hg19/GRCh37) with the mitochondrial genome replaced by the revised Cambridge reference sequence (rCRS) with parameters as in S4.1. Following realignment the mapped reads were processed as described in S4.1, with modifications introduced for the data from (21) and (26). Firstly, in order to remain consistent with the data processing pipeline outlined in section S4.1, reads below 34bp in length were discarded. Secondly, for the data from (21), whose sample libraries were treated with Uracil-DNA-glycosylase (UDG) to remove deaminated cytosines and thus change typical aDNA damage patterns, mapDamage rescaling was not performed on this data. Iron Age individuals from (26) were not included in this study.

We only included in main analyses the genomes from these studies with a minimum of 200,000 secure genotype calls from the 354,212 SNPS targeted in (25) that overlap with the Affymetrix Human Origins array (21). For additional analyses with the Hellenthal et al, 2014 dataset (see Section S9) a minimum of 200,000 overlapping genotype calls with this dataset were required, increasing the number of eligible samples from (26) from 18 to 24.

S4.2.2 Gamba et al. (2014)

BAM files from (10) had their base qualities rescaled using mapDamage 2.0 and reads below 34bp in length were removed. It was not necessary to reduce the base quality of the first and last two base pairs on reads from these samples as these positions had previously been removed before mapping. Only samples above 0.81X coverage were included in further analysis. Iron Age individuals from this study were not included in analysis.

S4.2.3 Olalde et al. (2014) & Olalde et al. (2015)

Fastq files from (20, 28) were downloaded and aligned to the human reference genome (hg19/GRCh37) with the mitochondrial genome replaced by the revised Cambridge reference sequence (rCRS) and were subsequently processed as outlined in S4.1. In addition, reads below 34bp in length were removed. Only one individual, CB13, from (28) possessed sufficient genome coverage to be included in this study.

S4.2.4 Keller et al. (2012)

Fastq files representing both the forward read of each pair and singletons were downloaded from (18). The files were aligned to the human reference genome (hg19/GRCh37) with the mitochondrial genome replaced by the revised Cambridge Reference Sequence (rCRS) and mapped reads were processed as described in section S4.1, with reads below 34bp in length removed.

S4.2.5 Haak et al. (2015)

Ancient genotype data covering a subset of SNPs from the Human Origins array (21) was downloaded from (25). In addition to the genotypes for samples sequenced by these authors, the data downloaded included genotypes for three pleistocene hunter gatherers sequenced in separate studies (22–24). Only samples with a minimum of 200,000 genotype calls were considered.

S5: Data Authenticity

S5.1 Negative Control Analysis

As described previously in section S2, stringent anti-contamination measures were taken during bone drilling, DNA extraction and library creation, all of which were carried out in clean room facilities dedicated to ancient DNA analysis. Full body suits, face masks, gloves and hairnets were worn and all tools and surfaces were consecutively cleaned with bleach, DNA-ExitusPlus, Ethanol and long exposure to UV light.

To assess the effectiveness of these contamination safeguards negative controls were introduced during each stage of sample processing and sequenced alongside the four Irish samples on the Illumina MiSeq platform (Trinity Genome Sequencing Laboratory, Trinity College Dublin, Ireland). Two controls were introduced during the drilling stage, an air control and a water control. The air control consisted of an empty 2ml Eppendorf tube which was allowed to sit open for an hour in the cleaned work area prior to sample drilling. For the water control, 2ml of ddH_2O was shaken in the Mixer Mill (MM 400, Retsch), used to obtain fine powder from bone, after bleaching and UV exposure. An extraction control of 1ml ddH_20 was subsequently introduced during the relevant stage.

The sequencing results of these controls are displayed in Table S5.1. To estimate the approximate percentage contamination for each sample we compared the number of reads mapped to the human genome obtained from the controls with the number of human reads obtained from the four samples on the same sequencing run, following the methods of (10). The contamination fraction (C) for each sample was calculated as :

$$C = (R_b/V_b)/(R_s/V_s)$$

where R_b represents the number of total human reads in all controls, R_s the number of human reads for a specific sample, V_b the total volume of controls added to the pool for loading onto the sequencer and V_s the sample volume added to the same sequencing pool. The resulting values are shown in Table S5.2. All samples displayed extremely low contamination estimates (<0.16%).

S5.2 Patterns of molecular damage and sequence length distribution

Using mapDamage 2.0 (16) we assessed whether each ancient genome displayed the two main patterns typical of aDNA post-mortem degradation, short sequence length and C to T changes at the 5'-end of molecules (29, 30). Only reads with a minimum mapping quality of 30 were taken into account.

All samples displayed a similar sequence length distribution with a peak observed between 49-51 bp for each sample (Figure S5.1*C*). This fits with previous observations that ancient samples display average sequence lengths of less than 100bp (31). As samples were sequenced using both 70 bp and 100 bp single-end sequencing peaks at these points could be seen, representing all sequences greater than these lengths that were truncated in their respective sequencing runs.

An increase in C to T and G to A changes towards the 5' and 3' ends of reads was observed in all samples (Figure S5.1*A* and *B*). These changes are characteristic of ancient DNA molecules (29). C to T misincorporation frequencies at the 5' end of reads ranged from 14.53% in Rathlin1 to 21.04% in Rathlin2, while complementary G to A frequencies ranged from 11.47% in Rathlin1 to 18.70% in Rathlin2.

S5.3 Mitochondrial contamination estimates

In order to detect the presence of possible modern mitochondrial contamination, the proportion of secondary bases not matching the sample consensus sequence was calculated at both haplogroup defining positions and private sample mutations identified using HAPLOFIND (32) (section S7) as previously described in (10, 33). Bases with a quality below 30 were not considered. The contamination rate was defined as the percentage of the total number of secondary bases across all sample sites considered over the total base count across these positions (%C). Given the influence of sequencing errors, heteroplasmy, NUMTs and post-mortem molecular damage these estimates represent an upper limit of contamination. To address the issue of post-mortem damage this analysis was repeated removing sites where the defining SNP was a G or a C in the sample consensus sequence (%C-MD). Results are shown in Table S5.3. Conservative estimates of mitochondrial contamination ranged from 0.36-0.50%, while when possible molecular damage was also included somewhat higher estimates of 1.02-2.63% were observed, particularly so for Rathlin2 (2.63%), who displayed the highest misincorporation frequencies in the previous section. Due to this confounding effect of postmortem damage, it is the conservative estimate (%C-MD) that is presented in the main paper.

S5.4 X-chromosome contamination estimates

Sex determination (section S6) of the four Irish samples sequenced revealed the 3 Bronze Age samples to be male. Their haploid state for the X-chromosome allows us to use mismatches to inform about potential contamination. Contamination is expected to be more apparent at sites that are polymorphic within the population compared to neighbouring monomorphic sites due to the increased propensity for allelic diversity at these positions in contaminating individuals. With this in mind, we applied to our data the methods developed in (34) using ANGSD v.0.614 (35), which calculates the rate of heterozygous calls at known polymorphic sites and compares this to the rate at adjacent monomorphic sites. First, we generated a binary count file for the X-chromosome for each sample, taking only into account reads that have mapping quality ≥ 30 and bases with a quality ≥ 20. Next, we used the R script "contamination.R" which performs a Fisher's exact test and jackknife to estimate contamination, delimiting our analysis to unique regions of the X-chromosome ("RES/chrX.unique.gz") and using known HapMap polymorphic sites ("RES/HapMapChrX.gz") (http://popgen.dk/angsd/index.php/Contamination). ANGSD applies two different methods to estimate the contamination rate, described in reference (34): in Method 1, all major and minor bases are determined for each SNP site and adjacent positions, which are used to estimate background error rates for each sample. Method 1 is more sensitive and has greater precision than Method 2 as it takes all bases into account. However, it assumes that errors between reads and sites are independent. On the other hand, Method 2 is less precise but also less biased because it randomly samples a single read at each site, giving major and minor base counts of either 1 or 0. Major and minor base counts for both methods are displayed in Table S5.4, contingency tables based on these counts are shown in Table S5.5 and contamination estimates based on these results in Table S5.6. Low contamination estimates (below 1.65%) were observed in all three ancient males.

Figure S5.1. Misincorporation patterns and sequence length distributions for Ballynahatty, Rathlin1, Rathlin2 and Rathlin 3. (A) and (B) display mismatch frequency relative to the reference genome as a function of read position. (A) Shows the frequency of C to T transitions at the 5' ends of reads while (B) shows the frequency of G to A transitions at the 3' ends of reads. (C) shows the percentage of total reads found for any given read length for each sample.

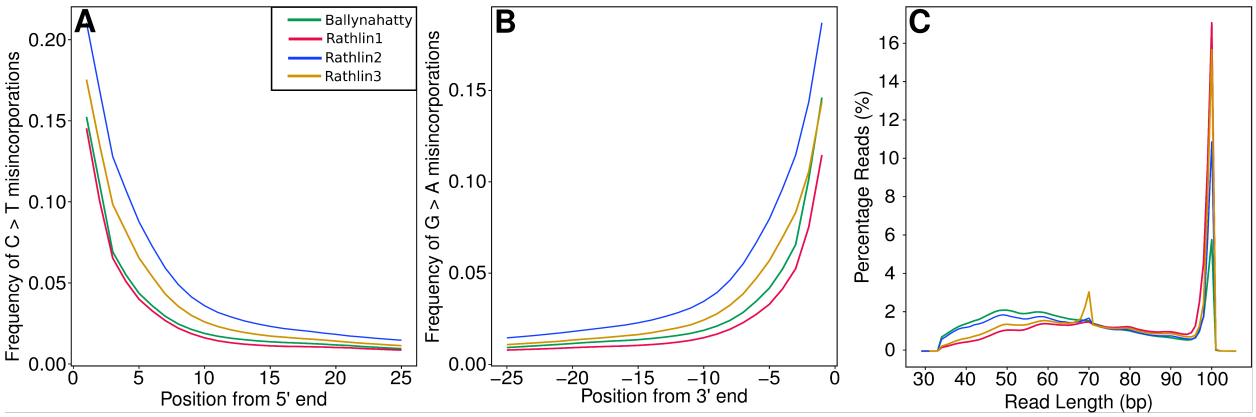


Table S5.1. Sequenced read numbers for negative controls.

Control ID	Total Reads	Reads mapped to hg19	Reads mapped (mapping quality >30)	Volume added to Sequencing pool (µl)
Air Control	4548	475	349	0.33
Water Control	1782	634	519	0.33
Extraction Control	1207	223	176	0.33
Total	7537	1332	1044	1

The three controls were sequenced alongside the four ancient samples on an Illumina MiSeq platform (Trinity Genome Sequencing Laboratory, Trinity College Dublin, Ireland) using 70 bp single-end sequencing.

Table S5.2. Sample contamination estimates based on negative control sequencing results.

Sample	Reads mapped to hg19	Volume added to Sequencing pool (µl)	Contamination (C)
Ballynahatty	1610081	1.76	0.15%
Rathlin1	847143	0.88	0.14%
Rathlin2	800137	0.96	0.16%
Rathlin3	1706599	0.90	0.07%

Sample mapped reads refer to those obtained from an initial sequencing run on an Illumina MiSeq platform (Trinity Genome Sequencing Laboratory, Trinity College Dublin, Ireland) using 70 bp single-end sequencing that also included the three controls. For each sample the contamination estimate (c) was calculated by comparing the total number of human reads across all controls to the total human reads of each sample, and adjusting for differences in pooling volumes as outlined in S5.1.

Table S5.3. Mitochondrial contamination estimates.

Sample	Markers used for contamination estimation	%C	%C-MD
Ballynahatty	T14766C, T72C, T16298C, 4769A, 6510A	1.54%	0.50%
Rathlin1	A11467G, A12308G, G12372A, C16192T, C16270T, T3197C, G9477A, T13617C, A14793G, C16256T, A15218G, A16399G, A15218G, A16399G , A9667G, C16291T, A12582G	1.02%	0.36%
Rathlin2	A11467G, A12308G, G12372A, C16270T, T3197C, G9477A, T13617C,C150T, T14182C, C1721T, A13637G, T16189C!, C3212T, G16398A	2.63%	0.39%
Rathlin3	C295T, T489C A10398G! A12612G, G13708A, C16069T, C150T, T152C!, G15257A, G15812A, C16193T, G10172A, C16278T!, T14819C	1.29%	0.38%

%C, represents the upper limit of contamination, while %C-MD is the estimated contamination rate when sites vulnerable to molecular damage are excluded from the calculation. Estimates are derived from the proportion of secondary bases at sample marker sites.

Table S5.4. Number of major and minor bases detected in SNP sites and adjacent positions in the X-chromosome for ancient male samples.

Method	Sample	Base	-4	-3	-2	-1	SNP site	1	2	3	4
Method 1	Rathlin1	minor base	557	539	601	637	1246	664	537	492	570
		major base	286890	286715	286414	286283	277489	286260	286656	287069	287175
	Rathlin2	minor base	107	94	113	92	235	125	84	91	96
		major base	27568	27588	27569	27586	27236	27605	27651	27621	27601
	Rathlin3	minor base	24	21	36	30	45	26	28	14	15
		major base	7992	8008	7995	8000	7956	7993	7988	8008	7998
Method 2	Rathlin1	minor base	92	110	91	102	217	106	83	84	99
		major base	46257	46239	46258	46247	46132	46243	46266	46265	46250
	Rathlin2	minor base	44	35	50	35	92	59	28	40	36
		major base	11274	11283	11268	11283	11226	11259	11290	11278	11282
	Rathlin3	minor base	11	6	19	18	18	13	13	4	7
		major base	3630	3635	3622	3623	3623	3628	3628	3637	3634

Method 1 and Method 2 were performed as in Rasmussen et al., 2011 (34).

Table S5.5. Contingency table for X-chromosome contamination estimates.

		Method 1		Method 2	
Sample	**Base**	**Polymorphic Sites**	**Average of Adjacent Sites**	**Polymorphic Sites**	**Average of Adjacent Sites**
Rathlin1	**major base**	277489	286682.75	46132	46253.125
	minor base	1246	574.625	217	95.875
	error	0.0045	0.0020	0.0047	0.0021
Rathlin2	**major base**	27236	27598.625	11226	11277.125
	minor base	235	100.25	92	40.875
	error	0.0086	0.0036	0.0082	0.0036
Rathlin3	**major base**	7956	7997.75	3623	3629.625
	minor base	45	24.25	18	11.375
	error	0.0057	0.003	0.005	0.0031

The number of major and minor bases found in known polymorphic sites in the three ancient male samples and the average number of major and minor bases at adjacent sites 4 bases upstream and downstream of the polymorphism and the observed probability of error (the ratio of minor bases to major bases).

Table S5.6. X-chromosome based contamination estimates in ancient male samples.

Sample	Contamination (%)	SE	P-value	Contamination (%)	SE	P-value
Rathlin1	0.688	0.041	2.20E-14	0.722	0.09747999	2.20E-14
Rathlin2	1.610	0.230	2.20E-16	1.646	0.2688079	2.17E-10
Rathlin3	1.014	0.230	2.20E-16	0.773	0.2688079	2.17E-10

Associated p-values for contingency tables used in this analysis were calculated using Fisher's exact test.

S6: Molecular Sex Determination

Osteoarchaeological analysis can provide accurate sex determination of ancient individuals based on differences in the metric dimensions of female and male skeletal anatomy (36, 37). Genetic sex determination can both confirm the findings of these studies and also take their place when an insufficient amount of skeletal material is available or the individual in question is a juvenile. Following the methods of (38) we identified the genetic sex of each ancient Irish human by calculating the ratio of reads aligning to the Y chromosome to reads aligning to both sex chromosomes (R_y). Only reads with a mapping quality over 30 were considered. Samples were assigned female if the upper bound of their confidence interval (CI) for R_y was lower than 0.016, while males required an R_y CI lower bound of above 0.075. The results (Table S6.1) identified Ballynahatty as female and the three Bronze Age individuals (Rathlin1, Rathlin2 and Rathlin3) as males, corroborating the osteoarchaeological findings.

Table S6.1. Molecular sex determination for Irish genomes.

Sample	Archaeological Sex	Ry	CI	Genetic Assignment
Ballynahatty	Female	0.004553	0.0045534766 - 0.0045534773	Female
Rathlin1	Male	0.092273	0.0922730811 - 0.092273109	Male
Rathlin2	Male	0.0923097	0.092309641 - 0.0923098134	Male
Rathlin3	Male	0.094700	0.0946998554 - 0.0947002383	Male

The ratio (Ry) denotes the ratio of reads aligning to the Y chromosome to the total number of reads aligning to both sex chromosomes. The 95% confidence interval (CI) was computed as $Ry \pm 1.96 \times Ry \times (1 - Ry)/(Nx + Ny)$, where Nx and Ny are the total number of reads aligning to the X chromosome and Y chromosome respectively. Only reads with a mapping quality above 30 were considered.

S7: Mitochondrial Genome Analysis

The mitochondrial coverage of each sample was obtained using the DepthOfCoverage tool in GenomeAnalysisTK v2.4-7 (15). To determine sample haplogroups we called consensus mitochondrial genome sequences using Samtools mpileup (with parameters -B and -Q 30) and vcfutils.pl (vcf2fq) (14) as in (19). HaploFind (32) was used to identify defining mutations and assign sample haplogroups. Only positions with a coverage of 6X or over were taken into account. Results are displayed in Table S7.1.

The Neolithic individual, Ballynahatty, was found to have all the defining mutations of haplogroup HV0 with the additional downstream T195C! mutation on the branch leading to haplogroups HV0b-HV0f. Haplogroup HV has been previously identified in ancient European individuals with highest frequencies seen in Early and Middle Neolithic groups from Germany and France (25, 39, 40). The haplogroup has also been described in hunter-gatherers from the Pitted Ware culture of southern Scandinavia (41, 42). The haplogroup has not to date been identified in any Mesolithic Europeans.

Two of the three Bronze age individuals, Rathlin1 and Rathlin2, belong the the haplogroup U5, which has been consistently observed as one of the dominant maternal lineages of European hunter-gatherer groups (21, 33, 43–46). Rathlin1 belongs to the haplogroup U5a1b1e, while Rathlin2 is a member of U5b2a2. The coalescence time estimate for the U5 haplogroup is approximately 25-30 thousand years as calculated by (47), with age estimates of 16-20 thousand years and 20-24 thousand years for the subhaplogroups U5a and U5b respectively. Haplogroup U5 is found at a frequency of 9% in modern day Europeans (48) with peaks seen in the Basque Country (49) and Scandinavia (50). The U5 haplogroup's age and distribution suggest its presence in Paleolithic Europe with a later expansion of its subclades from southern refugia after the Last Glacial Maximum (47). The U haplogroup shows an increase in frequency in ancient samples between the Early and Middle Neolithic, with another surge in frequency occurring in the Late Neolithic (25). These changes are thought to represent a resurgence of hunter gatherer lineages in the Middle Neolithic and an influx of steppe groups into Europe in the Late Neolithic and Bronze Ages (25).

The third Bronze Age individual, Rathlin3, is a member of haplogroup J2b1a. Haplogroup J is found at highest frequencies in the Middle East (13%), with a somewhat even frequency distribution across Europe (8.8%; refs. 48, 51). It is suggested that haplogroup J initially diversified in the Middle East where the majority of its subclades are distributed. Approximately 20% of J lineages belong to haplogroup J2, which

dates to approximately 37,000 years before present (51). While haplogroup J2 is predominantly Middle Eastern, the subclade J2b1 is found in Europe with highest frequencies seen in the Mediterranean-central and Atlantic regions. Our sample's haplogroup, J2b1a, is found almost exclusively in Europe and dates to approximately 11,000 years before present (51). As is the case with haplogroup HV, haplogroup J has been found in numerous ancient individuals, with highest frequencies seen in Early and Middle Neolithic groups (25, 39), suggesting its dispersal into Europe with agricultural populations.

Table S7.1. Mitochondrial haplogroup assignment and coverage for ancient Irish individuals.

Sample	Coverage (X)	Haplogroup	Defining haplogroup markers absent in sample haplotype	Sample mutations not associated with assigned haplogroup	Assignment Score	Haplo Score
Ballynahatty	246.09	HV0+195	**7521G**	4769A, 6510A, 16519T	0.95	1
Rathlin1	342.48	U5a1b1e	-	16519T	1	1
Rathlin2	103.68	U5b2a2	**4732G, 7521G, 7768G**, 16192T*	16519T, 16192insC	0.86	1
Rathlin3	64.88	J2b1a	**4216C, 5633T, 7146A, 7476T, 7521G**	14819C, 16519T	0.83	1

Absent haplogroup mutations marked in bold lacked sufficient sample coverage at these positions. The absent haplogroup mutation marked with an asterisk (*) was displaced by a private insertion in this sample.

S8: Y Chromosome Analysis

Three of the four ancient genomes were identified as genetically male (section S6). To determine the Y-chromosome haplogroup for each individual we assessed their allelic state at each SNP in the ISOGG database of Y chromosomal markers, omitting SNPs marked as under investigation. We also removed C/G and A/T SNPs lacking dbSNP information to avoid strand assignment errors.

The Pileup tool in GATK v2.4 (15) was used to report the base calls for all reads covering each site. Both a minimum base and mapping quality of 30 was required for consideration. Heterozygous sites were removed, possibly representing contamination, deamination or misalignment. The only exception to this were sites where the observed heterozygosity was explainable as deamination and the SNP in question was not a transition.

All samples could be securely assigned to haplogroup R1b1a2. This result was based on the presence of fourteen defining haplogroup mutations in Rathlin1, ten in Rathlin2 and four in Rathlin3 (Table S8.1), with an average coverage per informative SNP of 5.5X, 1.6X and 1.25X respectively. Moreover, all samples contained the derived allele at marker M529, the defining mutation of R1b1a2a1a2c, strongly suggesting their membership of this haplogroup. This marker was present at 4X coverage in Rathlin1, 1X coverage in Rathlin2 and 2X coverage in Rathlin3.

Rathlin1 could be further placed into the subhaplogroup R1b1a2a1a2c1g where the defining marker was present a 3X coverage. Rathlin2 also displayed the same C>T marker mutation, however the low sample coverage at this position (1X) prevented us from excluding deamination as a possible cause of the observation. Therefore, the final assigned haplogroup of Rathlin2 was R1b1a2a1a2c1 based on the presence of the A>C subhaplogroup mutation at 2X coverage.

All 3 Rathlin Bronze Age samples sequenced in the present study belong to the M529 lineage which is characterized by higher frequencies in the British Isles, reaching peak frequencies (> 94%) in the West of Ireland (52, 53).

One particular subclade of M529, defined by the mutation M222 (52), has a distribution which in the present day is considerably restricted to Ireland. Rathlin1, however, the only sample with reads covering SNP M222, presents the ancestral allele at this position.

Table S8.1. Defining Y chromosome haplogroup mutations for three Bronze Age male samples.

Haplogroup	Marker	SNP ID	Position in hg19/GRCh37	Ancestral Allele (fwd)	Derived Allele (fwd)	Rathlin1 Allele	Rathlin2 Allele	Rathlin3 Allele
R	M207	rs2032658	15581983	A	G	**G**	**G**	-
R	P224	rs17307398	17285993	C	T	**T**	-	**T***
R	P227	rs4481791	21409706	G	C	**C**	**C**	-
R	P232	rs9786261	23035132	G	A	**A**	**A***	**A***
R	P280	rs891407	21843090	C	G	**G**	**G**	-
R	P285	rs17249974	19267344	C	A	**A**	-	-
R	PF6014	rs1865680	6868118	G	A	**A**	-	-
R	M734	rs1348733	18066156	C	T	**T**	**T***	**T***
R1	P294	rs1005041	7570822	G	C	**C**	**C**	-
R1	P242	rs7067478	7647357	G	A	**A***	-	**A***
R1	P238	rs9785717	7771131	G	A	**A**	**A***	-
R1	P245	rs3853052	8633545	T	C	**C**	-	-
R1	P231	rs9786465	9989615	A	G	-	**G**	**G**
R1	M173	rs2032624	15026424	T	C	**C**	-	T*
R1	P225	rs17307070	15590342	G	T	**T**	-	-
R1	P286	rs1118473	17716251	C	T	**T**	**T***	**T***
R1	P236	rs9785959	17782178	C	G	**G**	-	**G**
R1	P234	rs9786197	21117888	T	C	**C**	**C**	-
R1	P233	rs9786232	21166358	T	G	**G**	-	-
R1	M306	rs1558843	22750583	C	A	**A**	-	**A**
R1b	M343	rs9786184	2887824	C	A	**A**	-	**A**
R1b1	M415	rs9786194	9170545	C	A	**A**	-	**A**
R1b1	L278	rs17249854	18914441	C	T	**T**	-	-
R1b1a	P297	rs9785702	18656508	G	C	**C**	**C**	**C**
R1b1a2	M269	rs9786153	22739367	T	C	**C**	**C**	-
R1b1a2	PF6399	rs2058276	2668456	C	T	**T**	-	-
R1b1a2	M520	rs9786486	4446430	T	A	-	**A**	**A**
R1b1a2	L265	rs9786882	8149348	A	G	**G**	**G**	-
R1b1a2	PF6475	rs66915044	17986687	C	A	**A**	-	-
R1b1a2	L1063	rs9786876	18167403	C	T	**T**	**T***	**T***

R1b1a2	PF6485	rs877756	18719565	T	C	**C**	**C**	-
R1b1a2	L407	-	13887941	G	A	**A**	**A**	-
R1b1a2	L478	-	3274923	A	C	**C**	-	-
R1b1a2	L482	-	7863189	G	A	**A**	**A***	-
R1b1a2	L483	-	8233186	C	T	**T**	**T***	**T***
R1b1a2	L500	-	18180446	C	A	**A**	**A**	-
R1b1a2	L773	-	7220727	A	G	**G**	-	-
R1b1a2	L1353	-	19179540	G	A	**A**	-	**A***
R1b1a2	L150.1	rs9785831	10008791	C	T	**T**	**T**	-
R1b1a2a	L49.1	rs9786142	2842212	T	A	**A**	-	-
R1b1a2a	L23	rs9785971	6753511	G	A	**A**	**A***	**A***
R1b1a2a1	M412	rs9786140	8502236	G	A	**A**	-	**A***
R1b1a2a1a	L52	rs13304168	14641193	C	T	**T**	**T***	
R1b1a2a1a	L151	rs2082033	16492547	C	T	**T***	-	-
R1b1a2a1a	P311	rs9785659	18248698	A	G	**-**	-	-
R1b1a2a1a	P310	rs9786283	18907236	A	C	**C**	-	-
R1b1a2a1a2	P312	rs34276300	22157311	C	A	**A**	-	-
R1b1a2a1a2c	M529	rs11799226	15654428	C	G	**G**	**G**	**G**
R1b1a2a1a2c	L459	rs141984805	5275051	C	G	**G**	-	-
R1b1a2a1a2c	S461	rs146019383	28632468	G	C	**C**	-	-
R1b1a2a1a2c1	S521	-	2836431	A	C	**C**	**C**	-
R1b1a2a1a2c1	Z2542	rs138975640	17885577	C	T	**T***	-	-
R1b1a2a1a2c1a	S474	-	22735599	G	A	G	G	-
R1b1a2a1a2c1b	S5196	-	16250226	T	C	T	-	-
R1b1a2a1a2c1b	L513	-	7340450	C	T	C	-	C
R1b1a2a1a2c1b	S279	-	5818205	T	C	T*	-	-
R1b1a2a1a2c1c	FGC11134	-	21383466	T	A	T	-	T
R1b1a2a1a2c1d	S3666	rs11799049	14048611	A	G	A	-	-
R1b1a2a1a2c1e	S219	-	15702775	T	C	T	T*	-
R1b1a2a1a2c1f	S218	-	7193034	G	A	G	G	-
R1b1a2a1a2c1g	S192	rs138322855	9878915	C	T	**T**	**T***	-
R1b1a2a1a2c1g1	FGC3213	-	26722140	A	C	-	-	-
R1b1a2a1a2c1g1a	P314.2	-	2888672	T	C	T	T*	-

R1b1a2a1a2c1g1b	S3058	-	15683936	T	C	T	T*	-
R1b1a2a1a2c1g2a	s280	-	7061380	T	G	T	T	T
R1b1a2a1a2c1g3	S5488	-	19088630	C	T	C	T*	-
R1b1a2a1a2c1g3b	Z16294	-	21649408	G	A	G	-	G
R1b1a2a1a2c1g3c	L1336	-	19372904	T	C	T	-	T*
R1b1a2a1a2c1g5	L130	-	6753503	G	A	G	G	G
R1b1a2a1a2c1h	L371	-	18656470	T	G	T	T	T
R1b1a2a1a2c1i	S836	-	14331384	C	T	-	C	C
R1b1a2a1a2c1i	S524	-	16992602	T	C	T	-	-
R1b1a2a1a2c1j	S470	-	8736334	G	A	G	G	-
R1b1a2a1a2c1k	L1335	-	23154201	A	G	A	A*	-
R1b1a2a1a2c1l	S7200	-	21753900	C	A	C	C	C
R1b1a2a1a2c1l	FGC5494	-	3445114	A	G	-	-	-
R1b1a2a1a2c1m	S16264	-	14658766	A	G	A	A*	A*
R1b1a2a1a2c1m	S21225	-	18731825	A	G	A	-	-
R1b1a2a1a2c1n	Y5305	-	17681656	C	T	C	-	-
R1b1a2a1a2c1a1a1a1	M222	rs20321	14902414	G	A	G	-	-

Sample alleles were determined with the Pileup tool in GATK v2.4 (15). A minimum base and mapping quality of 30 and was required for consideration. Sample alleles in bold font match the derived allele of the given marker. Samples alleles marked with an asterisk (*) represent a possible deamination event, where the defining mutation is a C>T or G>A and sample coverage at this site is below 3X.

S9: Genotype calling and merging with modern datasets

S9.1 Human Origins Dataset

In order to analyse our whole genome data in combination with the ancient genotype data published in (25), the publicly available modern genotype dataset used in this study was downloaded, covering 354,212 SNPS overlapping with the Affymetrix Human Origins array and consisting of 1946 present-day humans from 167 populations initially published in (21).

The vast majority of published ancient genomes to date possess sequencing coverage too low for accurate diploid genotype calling. To circumvent this issue the pileup tool in GATK v2.4 (15) was used to report sample base calls for each position in the Human Origins dataset for 33 ancient samples from (10, 18–21, 26–28) (data processing detailed in section S4.2) and the four Irish individuals. Bases needed a minimum quality of 30 to be considered and sites with three or more different bases present were removed. A single base call was then picked at random from each site and duplicated to create a diploid homozygous genotype at that position for that individual. Creating genotype calls in this manner has the effect of increasing the appearance of drift artificially on the lineage specific to the individual but this drift should not occur in any particular direction and should not affect inferences is about relationships between individuals. This strategy was used for all ancient genomes regardless of coverage to avoid any possible bias it may introduce compared to other genotype calling procedures.

Using PLINK v1.90 (54), we merged our genotype calls from both the Irish and published whole genome data (10, 18–21, 26–28) with the modern Human Origins dataset and the genotypes of 38 ancients individuals published in (25) as well as genotypes for three Pleistocene hunter gatherers obtained from (25) but sequenced in separate studies (22–24). In cases where the allele of an ancient individual matched neither of the alleles given in the Human Origins dataset, we removed this position in the ancient sample. Approximately 250 of these mismatches occurred in each sample and are most likely the result of strand assignment errors, postmortem damage or sequencing error.

The final merged dataset was used for principal component analysis (section S10), ADMIXTURE analysis (section S11) as well as D- and f-statistics (section S12). Information on the above 82 ancient samples, including number of confident SNP calls for positions in the Human Origins array, is displayed in Table S9.1.

S9.2 Hellenthal et al. 2014 Dataset

The four Irish genomes and 43 whole ancient genomes (10, 18–21, 26–28) could also be studied in relation to a second modern SNP dataset based on samples from (55–58) genotyped using the Illumina 660W array in (59). This consists of a total of 474,491 analyzed autosomal SNPs in 1530 individuals from 95 different populations, including Ireland. SNP positions were mapped to build 37 (GRCh37) of the human genome and we excluded genotypes with a minor allele frequency below 0.5% and a genotyping rate below 99% across all samples.

Homozygous diploid genotypes were called for 43 previously published ancient genomes (10, 18–21, 26–28) as well as the four ancient Irish genomes for each position in the Hellenthal dataset as described in section S9.1. Included were six samples (not detailed in table S9.1) from (26), which did not have a sufficient number of genotype calls from the Human Origins dataset to be considered for the main analyses, but did meet the quota required for analysis with the Hellenthal dataset.

The genotype calls for all ancient individuals were subsequently merged with the modern dataset using PLINK v1.90 (54), again as described in S9.1. This final merged dataset was used for principal component analysis (Figure S10.2).

A second dataset used in CHROMOPAINTER analysis (Main Paper, Fig. 3) was created for analysis consisting of only high coverage European ancient genomes above 10X and the modern Hellenthal dataset. Diploid genotypes calls were generated for the six high coverage samples (Stuttgart, Loschbour, BR2, NE1, Ballynahatty and Rathlin1), using the UnifiedGenotyper tool in GATK v2.4-7 (15). Only bases with a quality above 30 were considered for calling (-mbq 30). Genotype calls were filtered for a depth of coverage of 10X or above and a genotype quality of 30 or above. The six samples were then merged individually with the modern dataset using PLINK v1.90 (54). In cases where either allele of an ancient individual matched neither of the alleles given in the Hellenthal 2014 dataset, this position was removed from the ancient sample. This problem only occurred at two SNPs in the dataset which are reported as triallelic in dbSNP. Number of confident diploid calls for each sample is reported in Table S14.1.

S9.3 1000 Genomes Phase 3 v5 and Simons Genome Diversity Project Datasets

To analyse runs of homozygosity (ROH) for the six high coverage ancient European samples (Stuttgart, Loschbour, BR2, NE1, Ballynahatty and Rathlin1) (section S13), diploid genotypes were called for autosomal transversion SNPs from the Phase 3 v5 1000 Genomes release (60) using the UnifiedGenotyper tool in GATK v2.4-7 (15) as described in the final paragraph of section S9.2. This variant set allows for the the high SNP density required for this type of analysis, while removing the complication of postmortem deamination. Sample genotype calls were subsequently merged with the same set of transversion SNPs from 2504 individuals of the Phase 3 v5 1000 Genomes release (60) and 23 individuals from the Simons Genome Diversity Project (61) using PLINK v1.90 (54), again as in S9.2. Only SNPs genotyped across all ancient individuals with a minor allele frequency above 0.5% were retained. This left 1,601,314 transversion SNPs for ROH analysis.

Table S9.1. Sample IDs, information and SNPS counts for all ancient samples used in study.

Reference	Original Sample ID	Current Study ID	Country of Origin	Population Grouping (D- and f- Statistics)	Context	Human Origins SNP calls
Seguin-Orlando et al. (2014)	K14	Kostenki14	Russia	N/A	Hunter Gatherer	328898
Raghavan et al. (2014)	MA-1	MA1	Russia	N/A	Hunter Gatherer	250517
Fu et al. (2014)	Ust'-Ishim	Ust_Ishim	Russia	N/A	Hunter Gatherer	268641
Haak et al. (2015)	I0061	Karelia_HG	Russia	Eastern_HG (EHG)	Hunter Gatherer	337770
Haak et al. (2015)	I0124	Samara_HG	Russia	Eastern_HG (EHG)	Hunter Gatherer	204440
Gamba et al. (2014)	KO1	KO1	Hungary	Western_HG (WHG)	Hunter Gatherer	238503
Olalde et al. (2014)	La Braña 1	LaBrana	Spain	Western_HG (WHG)	Hunter Gatherer	331277
Lazaridis et al. (2014)	Loschbour	Loschbour	Luxembourg	Western_HG (WHG)	Hunter Gatherer	353217
Haak et al. (2015)	I0011	Motala_HG1	Sweden	Scandinavian_HG (SHG)	Hunter Gatherer	225724
Haak et al. (2015)	I0012	Motala_HG2	Sweden	Scandinavian_HG (SHG)	Hunter Gatherer	289609
Haak et al. (2015)	I0013	Motala_HG3	Sweden	Scandinavian_HG (SHG)	Hunter Gatherer	248337
Haak et al. (2015)	I0014	Motala_HG4	Sweden	Scandinavian_HG (SHG)	Hunter Gatherer	307860
Haak et al. (2015)	I0015	Motala_HG5	Sweden	Scandinavian_HG (SHG)	Hunter Gatherer	282120
Haak et al. (2015)	I0016	Motala_HG6	Sweden	Scandinavian_HG (SHG)	Hunter Gatherer	272163
Lazaridis et al. (2014)	Motala12	Motala12	Sweden	Scandinavian_HG (SHG)	Hunter Gatherer	317396
Skoglund et al. (2014)	Ajvide58	Ajv58	Sweden	N/A	Hunter Gatherer	310980
Gamba et al. (2014)	NE1	Hungarian_EN1	Hungary	Hungarian_EN	Early Neolithic	352352
Gamba et al. (2014)	NE5	Hungarian_EN5	Hungary	Hungarian_EN	Early Neolithic	197914
Gamba et al. (2014)	NE6	Hungarian_EN6	Hungary	Hungarian_EN	Early Neolithic	227018
Gamba et al. (2014)	NE7	Hungarian_EN7	Hungary	Hungarian_EN	Early Neolithic	220144
Haak et al. (2015)	I0025	LBK_EN1	Germany	LBK_EN	Early Neolithic	304262
Haak et al. (2015)	I0026	LBK_EN2	Germany	LBK_EN	Early Neolithic	311990
Haak et al. (2015)	I0046	LBK_EN3	Germany	LBK_EN	Early Neolithic	263814
Haak et al. (2015)	I0054	LBK_EN4	Germany	LBK_EN	Early Neolithic	333922
Haak et al. (2015)	I0100	LBK_EN5	Germany	LBK_EN	Early Neolithic	338572
Lazaridis et al. (2014)	Stuttgart	Stuttgart	Germany	N/A	Early Neolithic	353142
Haak et al. (2015)	I0410	Spanish_EN1	Spain	Spanish_EN	Early Neolithic	294296
Haak et al. (2015)	I0412	Spanish_EN2	Spain	Spanish_EN	Early Neolithic	330260
Haak et al. (2015)	I0413	Spanish_EN3	Spain	Spanish_EN	Early Neolithic	292545
Olalde et al. (2015)	CB13	Cardial_EN	Spain	N/A	Early Neolithic	215014
Current Study	Ballynahatty	Ballynahatty	Ireland	N/A	Middle Neolithic	353009
Skoglund et al. (2014)	Gokhem2	Gok2	Sweden	N/A	Middle Neolithic	234743

Haak et al. (2015)	I0172	Esperstedt_MN	Germany	N/A	Middle Neolithic	276090
Haak et al. (2015)	I0406	Spanish_MN1	Spain	Spanish_MN	Middle Neolithic	320571
Haak et al. (2015)	I0407	Spanish_MN2	Spain	Spanish_MN	Middle Neolithic	233600
Haak et al. (2015)	I0408	Spanish_MN3	Spain	Spanish_MN	Middle Neolithic	318206
Haak et al. (2015)	I0108	Bell_Beaker_LN1	Germany	Bell_Beaker	Late Neolithic	257629
Haak et al. (2015)	I0111	Bell_Beaker_LN2	Germany	Bell_Beaker	Late Neolithic	205899
Haak et al. (2015)	I0112	Bell_Beaker_LN3	Germany	Bell_Beaker	Late Neolithic	337239
Allentoft et al. (2015)	RISE569	Bell_Beaker_Czech	Czech Republic	Bell_Beaker	Late Neolithic	205056
Haak et al. (2015)	I0058	BenzigerodeHeimburg_LN1	Germany	BenzigerodeHeimburg_LN	Late Neolithic	243960
Haak et al. (2015)	I0059	BenzigerodeHeimburg_LN2	Germany	BenzigerodeHeimburg_LN	Late Neolithic	238399
Haak et al. (2015)	I0103	Corded_Ware_LN1	Germany	Corded_Ware_LN	Late Neolithic	333213
Haak et al. (2015)	I0104	Corded_Ware_LN2	Germany	Corded_Ware_LN	Late Neolithic	332923
Haak et al. (2015)	I0118	Alberstedt_LN	Germany	N/A	Late Neolithic	346109
Allentoft et al. (2015)	RISE98	Nordic_LN	Sweden	N/A	Late Neolithic	337905
Günther et al.(2015)	ATP12-1420	Spanish_CA1	Spain	Spanish_CA	Copper Age	268393
Günther et al. (2015)	ATP16	Spanish_CA2	Spain	Spanish_CA	Copper Age	246480
Günther et al. (2015)	ATP2	Spanish_CA3	Spain	Spanish_CA	Copper Age	346793
Gamba et al. (2014)	CO1	CO1	Hungary	N/A	Copper Age	208581
Keller et al. (2012)	Iceman	Oetzi	Italy	N/A	Copper Age	345457
Haak et al. (2015)	I0231	Yamnaya1	Russia	Yamnaya	Copper Age	344303
Haak et al. (2015)	I0429	Yamnaya2	Russia	Yamnaya	Copper Age	215252
Haak et al. (2015)	I0438	Yamnaya3	Russia	Yamnaya	Copper Age	211116
Haak et al. (2015)	I0443	Yamnaya4	Russia	Yamnaya	Copper Age	340077
Allentoft et al. (2015)	RISE548	Yamnaya5	Russia	Yamnaya	Copper Age	220410
Allentoft et al. (2015)	RISE552	Yamnaya6	Russia	Yamnaya	Copper Age	290485
Allentoft et al. (2015)	RISE500	Andronovo1	Russia	Andronovo	Early Bronze	237751
Allentoft et al. (2015)	RISE505	Andronovo2	Russia	Andronovo	Early Bronze	341839
Gamba et al. (2014)	BR1	BR1	Hungary	Hungarian_Bronze	Early Bronze	171932
Allentoft et al. (2015)	RISE479	Vatya	Hungary	Hungarian_Bronze	Early Bronze	264115
Current Study	Rathlin1	Rathlin1	Ireland	Irish_Bronze	Early Bronze	353333
Current Study	Rathlin2	Rathlin2	Ireland	Irish_Bronze	Early Bronze	284047
Current Study	Rathlin3	Rathlin3	Ireland	Irish_Bronze	Early Bronze	202893
Allentoft et al. (2015)	RISE395	Sintashta1	Russia	N/A	Early Bronze	275743
Allentoft et al. (2015)	RISE511	Afanasievo1	Russia	N/A	Early Bronze	327954
Haak et al. (2015)	I0047	Unetice_EBA1	Germany	Unetice	Early Bronze	285142

Haak et al. (2015)	I0114	Unetice_EBA_relative	Germany	Unetice	Early Bronze	214641
Haak et al. (2015)	I0116	Unetice_EBA2	Germany	Unetice	Early Bronze	304713
Haak et al. (2015)	I0117	Unetice_EBA3	Germany	Unetice	Early Bronze	276893
Haak et al. (2015)	I0164	Unetice_EBA4	Germany	Unetice	Early Bronze	329147
Allentoft et al. (2015)	RISE150	Unetice_Poland	Poland	Unetice	Early Bronze	205990
Allentoft et al. (2015)	RISE577	Unetice_Czech	Czech Republic	Unetice	Early Bronze	212153
Gamba et al. (2014)	BR2	BR2	Hungary	Hungarian_Bronze	Late Bronze	352215
Allentoft et al. (2015)	RISE493	Karasuk1	Russia	Karasuk	Late Bronze	345106
Allentoft et al. (2015)	RISE495	Karasuk2	Russia	Karasuk	Late Bronze	341164
Allentoft et al. (2015)	RISE496	Karasuk3	Russia	Karasuk	Late Bronze	294150
Allentoft et al. (2015)	RISE497	Karasuk4	Russia	Karasuk	Late Bronze	350147
Allentoft et al. (2015)	RISE499	Karasuk5	Russia	Karasuk	Late Bronze	240623
Allentoft et al. (2015)	RISE502	Karasuk6	Russia	Karasuk	Late Bronze	240991
Haak et al. (2015)	I0099	Halberstadt_LBA	Germany	N/A	Late Bronze	333815
Allentoft et al. (2015)	RISE523	Mezhovskaya1	Russia	N/A	Late Bronze	306507

Samples are ordered by chronological context.

S10: Principal Component Analysis

To investigate the genetic affinities of our Irish genomes in the context of modern and ancient Eurasian populations we performed two principal component analyses (PCAs). The methods used to call sample genotypes and merge them with them with modern data is outlined in the previous section S9.

We used smartpca version 10210 from EIGENSOFT (62, 63) to carry out PCA on an extracted subset of 677 individuals from 54 West Eurasian populations (Table S10.1) from the modern Human Origins dataset (21) using a set of 354212 SNPs from the Human Origins array (21), with 82 ancient genomes, including the four Irish samples, projected onto the modern variation (lsqproject: YES option). The results are displayed in Fig.1*A* and *B* of the main paper. Figure S10.1 shows the same PCA with modern individuals alone, labeled with the first three letters of the population to which they belong.

PCA was also performed on 43 ancient individuals from (10, 18–21, 26–28) and the four Irish individuals, merged with the Hellenthal 2014 (59) dataset (Section S9.2). We carried out the PCA on a set of 472565 SNPs from a subset of 681 modern Eurasians from 41 populations (Table S10.1) and projected ancient individuals onto the modern variation as described above. The results are shown in Figure S10.2.

Comparing the results from both PCAs, based on genotypes from different SNP arrays and different individuals, we can determine that our findings are not the product of unseen bias in our dataset but in fact truly representative of patterns in Eurasian genetic variation, both modern and ancient.

Figure S10.1. Principal component plot of 677 West Eurasians from the modern Human Origins dataset. Individuals are labeled by the first three letters of their population grouping, listed in Table S10.1. Individuals are coloured by geographical region.

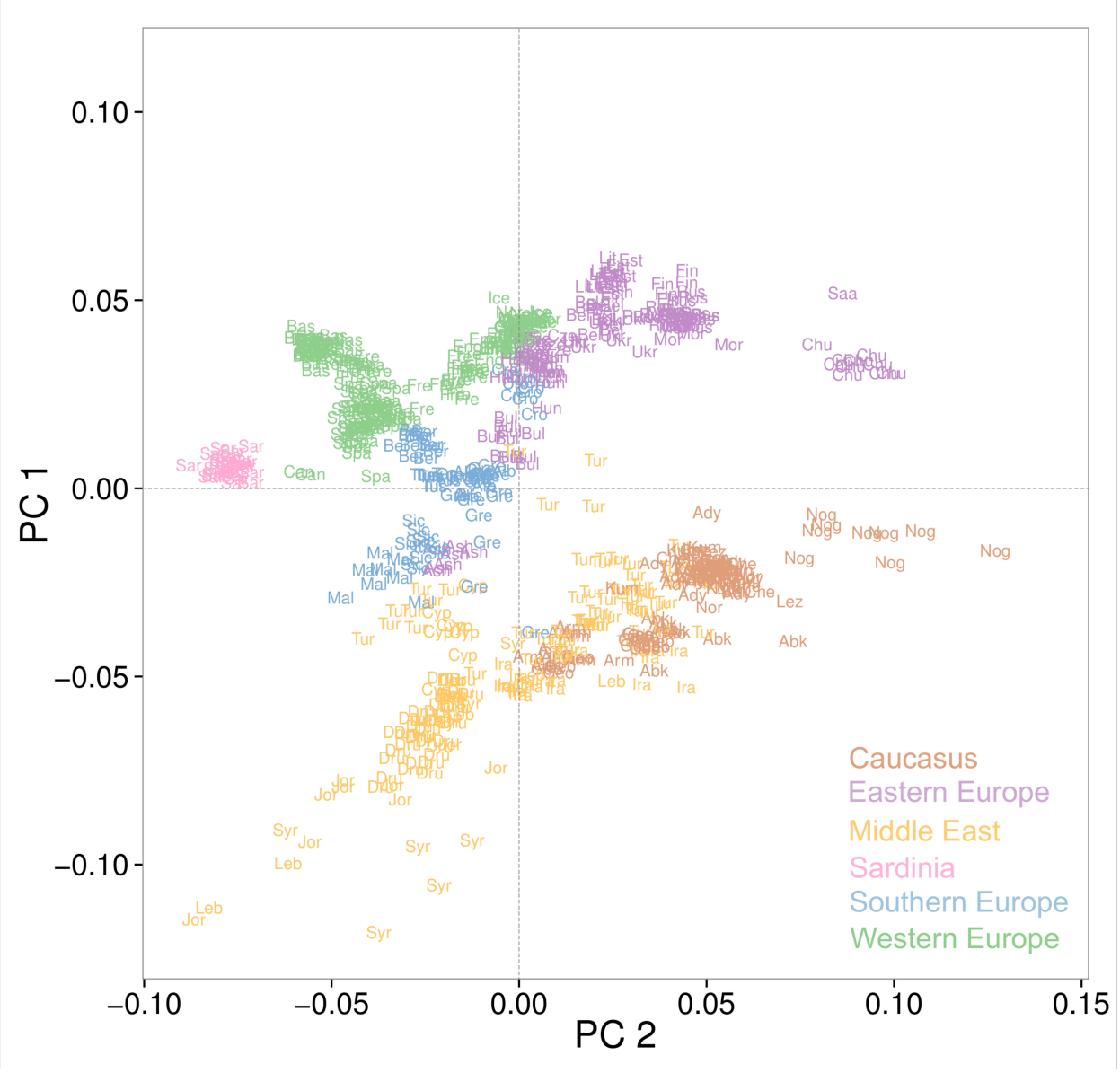


Figure S10.2. Principal component plot of 681 modern West Eurasians from the Hellenthal et al., 2014 dataset with 47 ancient individuals projected. Modern individuals are shown as transparent circles, coloured by geographical region. Ancient samples are displayed as opaque shapes with black outlines. Shape is determined by archaeological context, while colour is dependent on the region of origin of the ancient sample. We observe the same patterns and clustering as we see in Fig.1 *A* and *B* of the main paper.

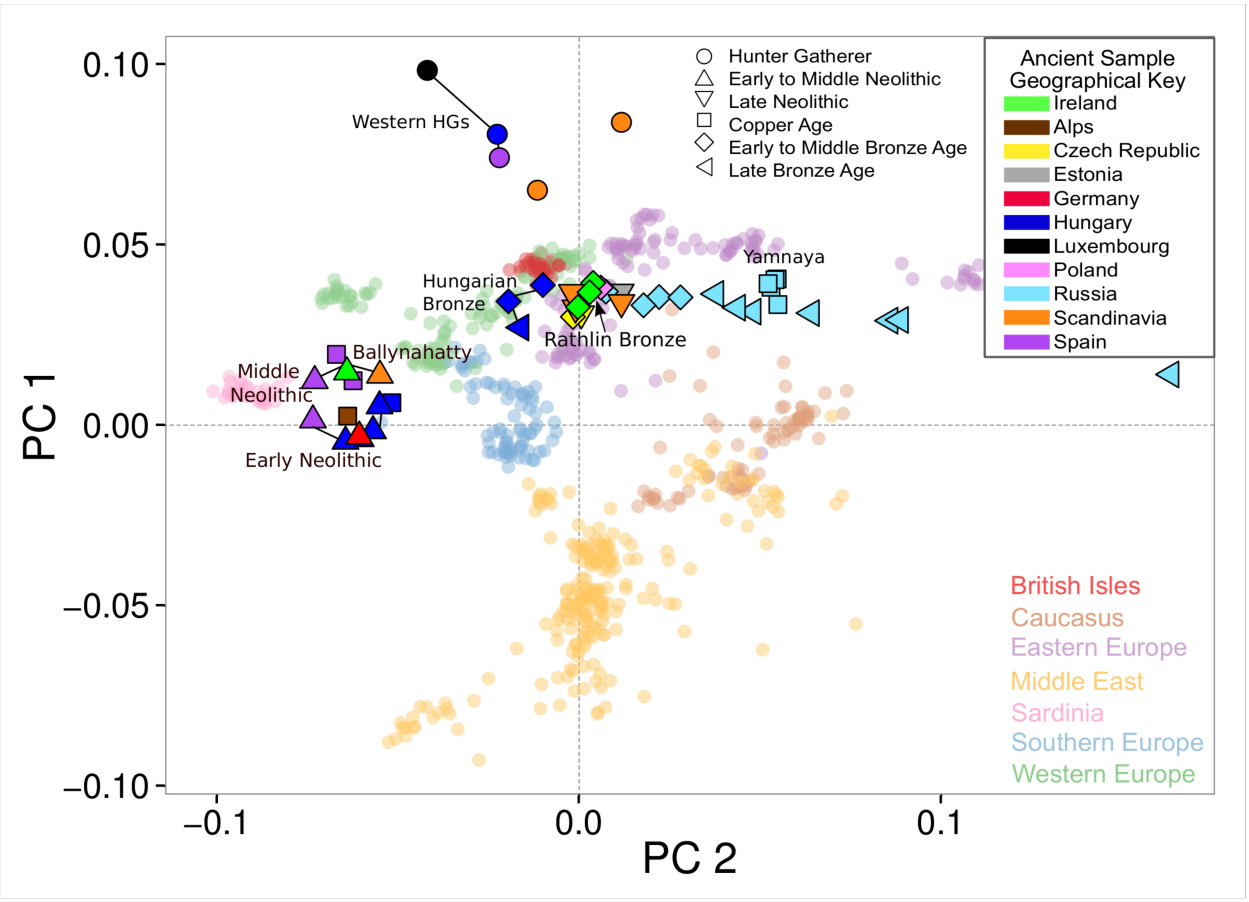


Table S10.1. List of populations used for principal component analyses.

Human Origins Dataset			Hellenthal 2014 Dataset		
Population Name	**Number of Individuals**	**Geographical Grouping for PCA**	**Population Name**	**Number of Individuals**	**Geographical Grouping for PCA**
Abkhasian	9	Caucasus	Adygei	17	Caucasus
Adygei	17	Caucasus	Armenian	16	Caucasus
Albanian	6	Southern Europe	Basque	24	Western Europe
Armenian	10	Caucasus	Bedouin	45	Middle East
Ashkenazi_Jew	7	Eastern Europe	Belorussian	8	Eastern Europe
Balkar	10	Caucasus	Bulgarian	18	Eastern Europe
Basque	29	Western Europe	Chuvash	17	Eastern Europe
Belarusian	10	Eastern Europe	Cypriot	12	Middle East
Bergamo	12	Southern Europe	Druze	42	Middle East
Bulgarian	10	Eastern Europe	EastSicilian	10	Southern Europe
Canary_Islanders	2	Western Europe	Egyptian	11	Middle East
Chechen	9	Caucasus	English	6	British Isles
Chuvash	10	Eastern Europe	Finnish	2	Eastern Europe
Croatian	10	Southern Europe	French	28	Western Europe
Cypriot	8	Middle East	Georgian	20	Caucasus
Czech	10	Eastern Europe	GermanyAustria	4	Western Europe
Druze	39	Middle East	Greek	20	Southern Europe
English	10	Western Europe	Hungarian	20	Eastern Europe
Estonian	10	Eastern Europe	Iranian	20	Middle East
Finnish	7	Eastern Europe	Ireland	7	British Isles
French	25	Western Europe	Jordanian	19	Middle East
French_South	7	Western Europe	Lezgin	18	Caucasus
Georgian	10	Caucasus	Lithuanian	10	Eastern Europe
Georgian_Jew	7	Caucasus	NorthItalian	12	Southern Europe
Greek	20	Southern Europe	Norwegian	18	Western Europe
Hungarian	20	Eastern Europe	Orcadian	15	British Isles
Icelandic	12	Western Europe	Palestinian	46	Middle East
Iranian	8	Middle East	Polish	16	Eastern Europe

Iranian_Jew	9	Middle East	Romanian	14	Eastern Europe
Iraqi_Jew	6	Middle East	Russian	25	Eastern Europe
Italian_South	1	Southern Europe	Sardinian	28	Sardinia
Jordanian	9	Middle East	Scottish	6	British Isles
Kumyk	8	Caucasus	SouthItalian	18	Southern Europe
Lebanese	8	Middle East	Spanish	34	Western Europe
Lezgin	9	Caucasus	Syrian	16	Middle East
Lithuanian	10	Eastern Europe	Turkish	17	Middle East
Maltese	8	Southern Europe	Tuscan	8	Southern Europe
Mordovian	10	Eastern Europe	Welsh	4	British Isles
Nogai	9	Caucasus	WestSicilian	10	Southern Europe
North_Ossetian	10	Caucasus			
Norwegian	11	Western Europe			
Orcadian	13	Western Europe			
Russian	22	Eastern Europe			
Saami_WGA	1	Eastern Europe			
Sardinian	27	Sardinia			
Scottish	4	Western Europe			
Sicilian	11	Southern Europe			
Spanish	53	Western Europe			
Spanish_North	5	Western Europe			
Syrian	8	Middle East			
Turkish	56	Middle East			
Turkish_Jew	8	Middle East			
Tuscan	8	Southern Europe			
Ukrainian	9	Eastern Europe			

S11: ADMIXTURE Analysis

To estimate the different components of ancestry in our samples in relation to both modern populations worldwide and published ancient data we used a model-based clustering approach performed by the program ADMIXTURE v.1.23 (64).

A preliminary analysis was carried out on 1941 present-day humans belonging to 162 populations from the Human Origins dataset (21) alongside 77 ancient individuals, including the four Irish genomes from this study (See S9.1 for information on genotype calling and dataset creation). This dataset had been filtered for related individuals and SNPs with a genotyping rate below 98.5%. The remaining 288,822 positions were subsequently pruned for SNPs in strong linkage disequilibrium using PLINK v1.90 (54), with the parameters (--indep-pairwise 200 25 0.4), leaving a final 179013 SNPs for analysis. ADMIXTURE was run with cross-validation enabled using the --cv flag for all ancestral population numbers from K=2 to K=20. This analysis was replicated 40 times over using random seeds chosen from the set 1 to 10000. The best of these replicates for each value of K, i.e. those with the maximum log likelihood (Figure S11.3), is presented in Figure S11.1.

We observed CV error to drop with increasing values of K until K=11 at which point error values began to slowly rise again (Fig S11.2). The lowest median CV error was attained for K=11.

Admixture analysis was then repeated at K=11 for 50 replicates using the same filtering, parameters and individuals as above with four additional genomes from Spain included (27, 28). As these genomes all were above 1X in coverage only 1023 SNPs were removed during the repeat filtering, leaving 177690 SNPs for analysis. The replicate possessing the highest log likelihood is displayed for all ancient individuals in Fig. 1*C*.

Figure S11.1. ADMIXTURE results for the best of 40 replicates for each value of K. Populations are arranged into geographical and temporal groupings. The lowest median CV error was attained for K=11.

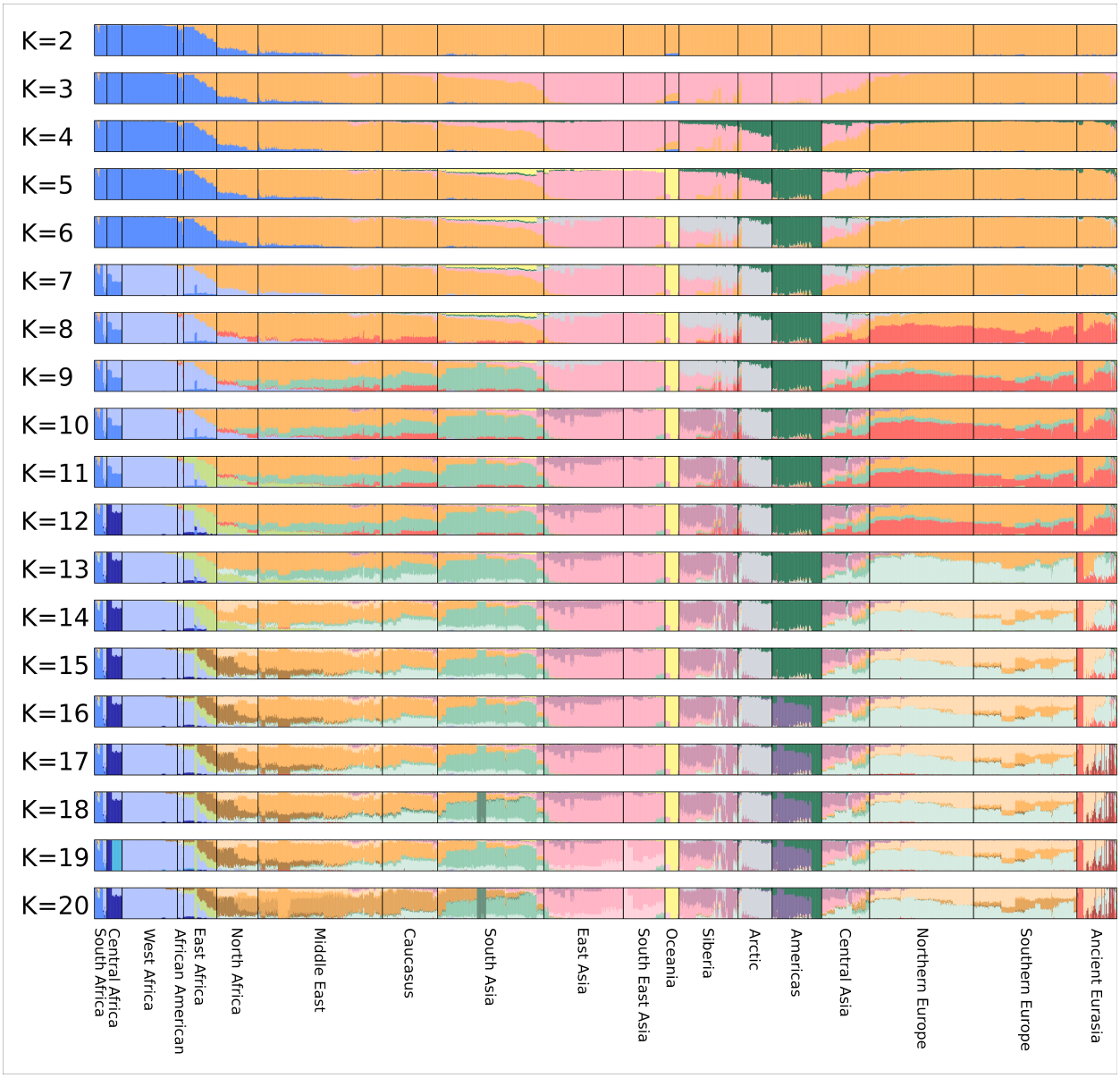


Figure S11.2. Cross validation errors obtained for each value of K across 40 replicates of ADMIXTURE analysis.

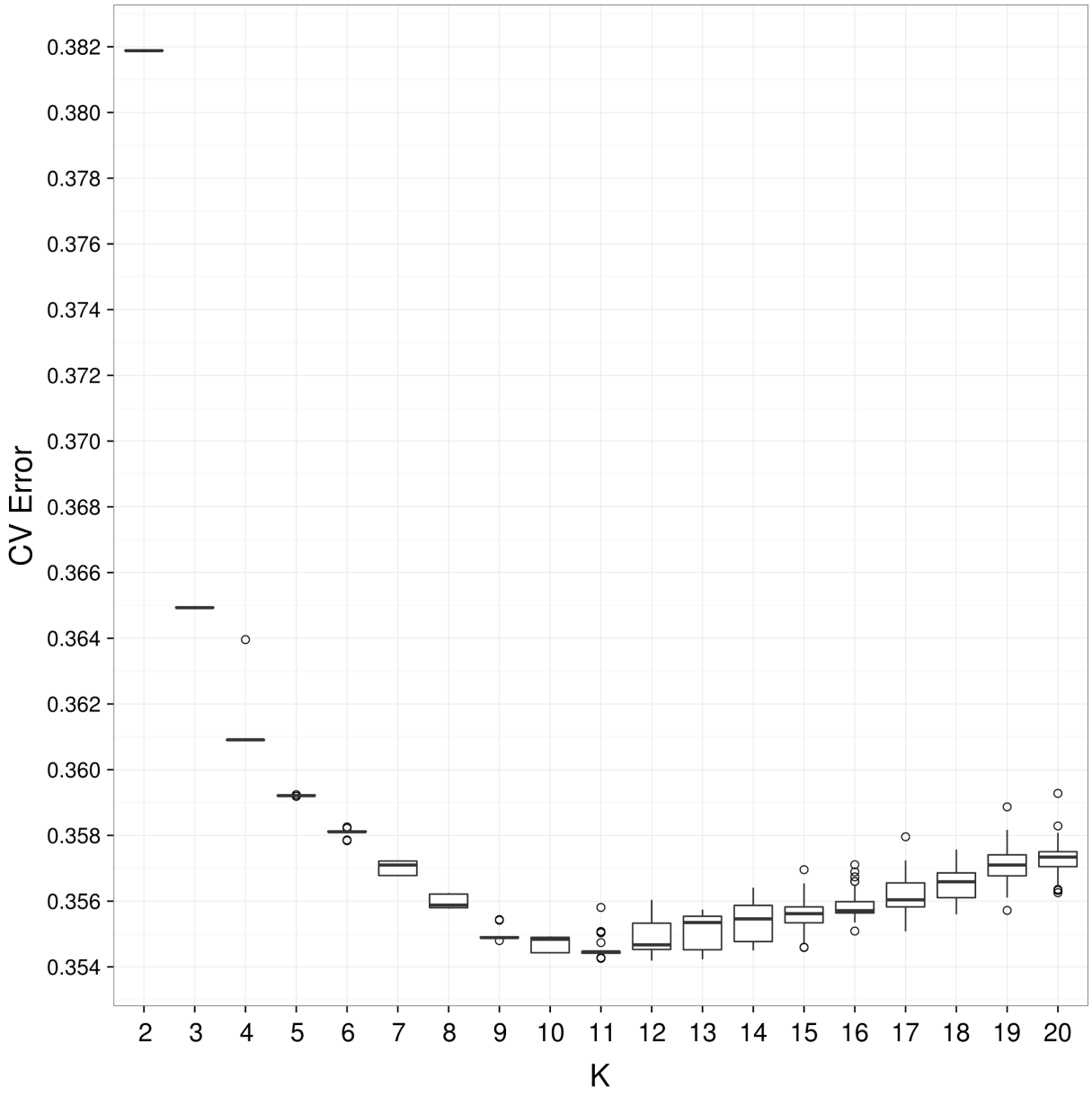
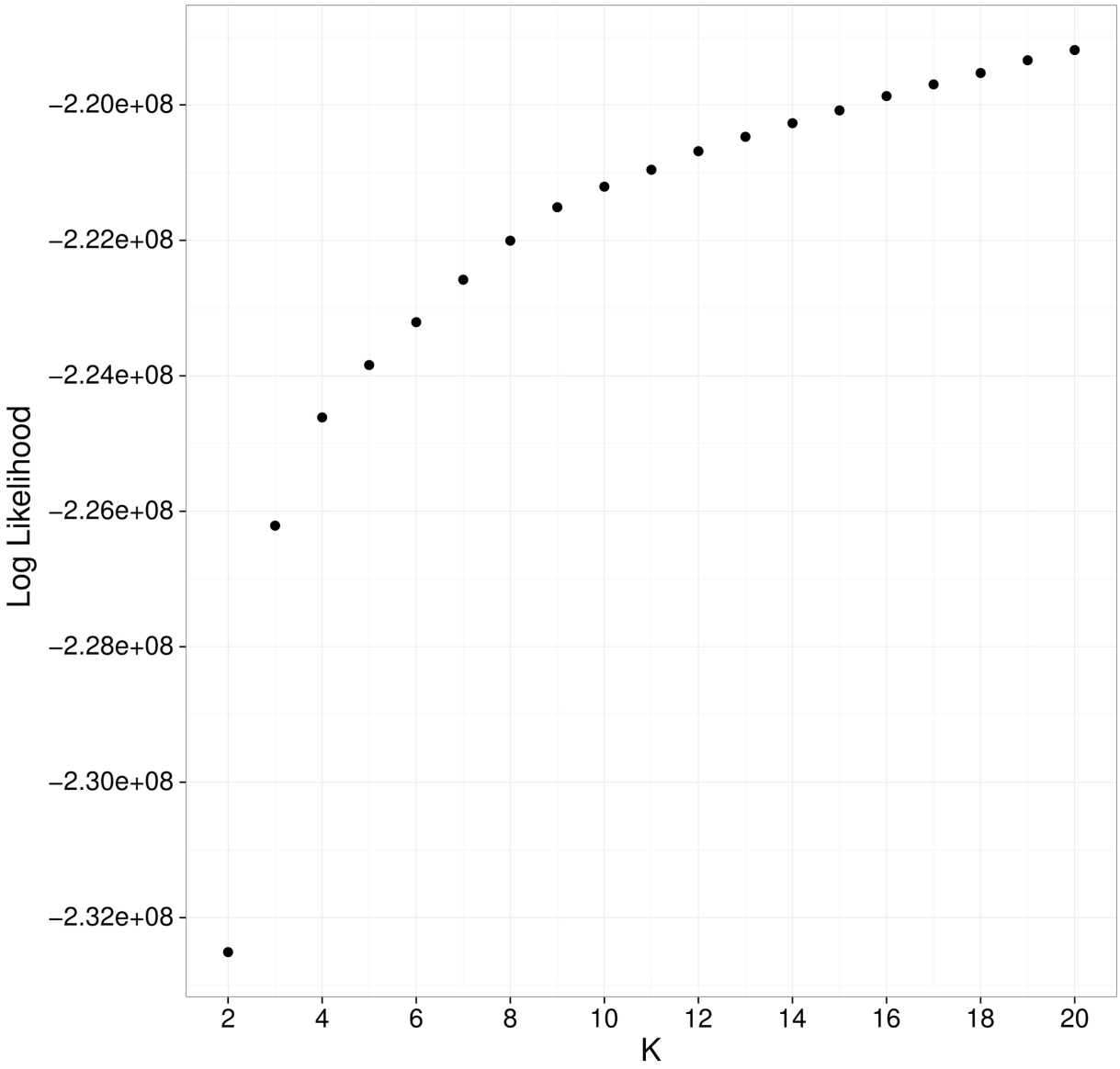


Figure S11.3 Log Likelihood for the best of 40 replicates of ADMIXTURE analysis for each value of K.



S12: Formal Tests of Admixture implemented with D- and f-Statistics

In order to confirm the relationships between ancient samples that were visualised in PCA and ADMIXTURE analysis we carried out formal tests of admixture using D-statistics (65) and f-statistics (66, 67). These were employed using the ADMIXTOOLS package (67) with significance assessed using a block jackknife of 5cM in size. Statistics were considered significant if they possessed a Z-score greater than 3 (66, 67), which corresponds to a p-value of less than 0.001. All tests were carried out on the same dataset used in PCA and ADMIXTURE analysis (described in S9.1). For any tests that required an outgroup the Mbuti population of Central Africa was used, given their undetectable levels of Eurasian admixture (68)**.** For information on ancient individuals used and population keys see Table S9.1

S12.1 Outgroup f_3-statistics

We used outgroup f_3-statistics to explore relationships between the ancient Irish genomes and other ancient populations and individuals. These were constructed in the form f_3(Mbuti; IA, X), where IA was each Irish ancient genome in turn and X was any other ancient individual or population. The magnitude of the resulting f_3-statistic is a measure of the amount of genetic drift shared between IA and OA since their split from Mbuti. Results are displayed in Figure S12.1.

Ballynahatty was seen to share the most drift with other Neolithics followed by western hunter gatherers (WHG), with the maximum scores observed for Spanish Cardial (Cardial_EN), Epicaridial (Spanish_EN) and Middle Neolithics (Spanish_MN), as well as the Scandinavian Middle Neolithic Gok2. It is interesting to note that both the Scandinavian and Spanish Middle Neolithic individuals come from Megalithic contexts as does Ballynahattty.

The Irish Bronze Age individuals all shared high levels of drift with two western hunter gatherers, the Hungarian KO1 and Loschbour. Rathlin2, who falls temporally between the earlier Rathlin1 and later Rathlin3, also displayed high levels of drift with both these individuals. However, there does not appear to be an especially high amount of drift between Rathlin1 and Rathlin3. All three Rathlin Islanders also had relatively high values of sharing with the Corded Ware and other European Late Neolithic and Bronze Age populations, with a Sintashta individual giving a high value for Rathlin2 and Rathlin3 but not Rathlin1. All three also show lower affinities to the early Neolithic LBK culture, with magnitude of drift

increasing with Middle Neolithics. Of all Neolithic samples tested, Rathlin2 shares it's highest affinity with Cardial_EN, Rathlin3 with Spanish_MN, and Rathlin1 with the Scandinavian Middle Neolithic Gok2.

S12.2 D-Statistics

All D-Statistic test results are reported in the supporting Dataset S1 file under Supporting Information.

S12.2.1 Neolithic Ballynahatty

It is now clear that Neolithic genomes sampled in Hungary, Germany, Spain and Scandinavia all have majority contributions from early farmers migrating from the south (10, 18, 19, 21, 25, 43, 69). In PCA Ballynahatty is seen clustering with these genomes and shows a similar admixture profile to other Middle Neolithics, suggesting she too shares this large component of early farmer (EF) ancestry.

To confirm this, we first investigated Ballynahatty's relationship to the three Mesolithic hunter gather groupings from (25), eastern hunter gatherer (EHG), Scandinavian hunter gather (SHG) and western hunter gatherer (WHG) using the statistic D(Mbuti, Ballynahatty; HG Group1, HG Group2). Of the three, Ballynahatty shared the most alleles with WHG.

We compared Ballynahatty's hunter gatherer affinities to six Middle Neolithic and European Copper Age individuals/groups, Gok2 (Scandinavian), CO1 (Hungarian), Esperstedt_MN (German), Oetzi (Alpine), Spanish_CA and Spanish_MN all of which clustered close to Ballynahatty on the PCA and shared similar admixture profiles. The statistic D(Mbuti, MN/CA; HG Group1, HG Group2) was used, where MN/CA refers to all of the aforementioned samples. We observed Ballynahatty and Spanish_CA to have the strongest preferences for WHG over SHG and EHG, while Esperstedt_MN displayed the least.

The possibility of Ballynahatty being a WHG herself was ruled out with the statistic D(Mbuti, Ballynahatty; X, Y) where X and Y formed the three combinations of the Spanish, Luxembourg and Hungarian WHG genomes. In all cases the WHG genomes belong in a clade excluding Ballynahatty. Of the three WHG individuals Ballynahatty appeared to share the least amount of affinity with LaBrana and the largest with Loschbour, but not significantly so.

The degree of WHG admixture in Ballynahatty was compared to three Early Neolithic groups, LBK_EN, Hungarian_EN and Spanish_EN, as well as two Early Neolithic individuals, Stuttgart and Cardial_EN (See table S9.1) using the statistic D(Mbuti, WHG; Ballynahatty, EN). This gave a significant indication of WHG introgression in Ballynahatty compared to early Spanish, Hungarian and LBK Neolithics (|Z|= 1.342 - 7.063), akin to that seen when Ballynahatty is replaced with other Middle Neolithics. Middle Neolithic samples showed the lowest affinities to WHG when Cardial_EN was the Early Neolithic used, indicating possible increased WHG in Cardial_EN compared to other Early Neolithics.

The test D(Mbuti, WHG; Ballynahatty, MN/CA) gives no significant statistic of gene flow suggesting Ballynahatty forms a clade with other Middle Neolithics separate to WHG. The only exception to this is when MN/CA is Spanish_MN which appears to have experienced significantly more WHG gene flow than Ballynahatty.

To measure the rough proportion of Ballynahatty's ancestry derived from WHG we used the f_4 ratio f_4(Mbuti, Loschbour; Ballynahatty, Dai)/f_4(Mbuti, Loschbour; KO1, Dai) (70). From this we estimated a WHG component of 42 ± 2% within a predominantly early farmer genome, a result similar to that seen in ADMIXTURE analysis.

Ballynahatty's affinity to different Early Neolithic populations was assessed using the statistic D(Mbuti, Ballynahatty; EN1, EN2). Ballynahatty appeared to share the most alleles with Cardial_EN and Spanish_EN. In order to confirm that this observation was not simply the result of increased WHG ancestry in Cardial_EN and Spanish_EN relative to other Early Neolithics, the test was repeated with other MN/CA members in the place of Ballynahatty to see if patterns of allele sharing differed in these samples. This test revealed Early Neolithic affinity to be somewhat correlated to geographical location. The Hungarian Copper Age sample, CO1, showed the highest affinity with Hungarian_EN, Spanish_MN and Spanish_CA with Spanish_EN and Cardial_EN, and the German Esperstedt_MN with LBK_EN.

The affinity of Ballynahatty to different Middle Neolithic and Chalcolithic (MN/CA) samples was also explored as above (D(Mbuti, Ballynahatty; MN/CA1, MN/CA2)), and the Irish farmer was seen to have the most allele sharing with the Scandinavian individual Gok2, followed by Spanish_MN. Both this and the above result agree with those from the outgroup f_3-statistics, identifying other Atlantic Neolithic individuals as the samples with whom Ballynahatty shares the most genetic drift.

S12.2.2 Bronze Age Rathlin

Prior studies (10, 25, 26) convincingly demonstrate that Central European genomes from the Late Neolithic and Early Bronze Age differ from the preceding Middle Neolithic due to a massive immigration of Steppe herders linked to cultures such as the Yamnaya. This is initially and most strongly apparent at the Corded Ware horizon. Accordingly, we employed a series of tests to gauge whether the ancestries of the Rathlin Early Bronze Age genomes were subject to this influence and if so to what extent.

First, we confirmed that the five Middle Neolithic and Southern Central European Copper Age individuals, Ballynahatty, Gok2 (Scandinavian), CO1 (Hungarian), Oetzi (Alpine) and Esperstedt_MN (German), and two populations Spanish_MN and Spanish_CA, the whole group referred to as MN/CA, always form clades with each other to the exclusion of the Irish Bronze Age individuals. For this we used the statistic D(Mbuti, Irish_Bronze; MN/CA1, MN/CA2). No significant gene flow ($|Z| > 3$) was observed. However, slightly higher correlation with Gok2 compared to other MN/CA members was seen for Rathlin1 and Rathlin3. This affinity is also noted in the previous outgroup f_3-statistics. Overall, the results suggests that the Irish Bronze Age was not simply a continuation of the Irish Middle Neolithic, or indeed any European Middle Neolithic group so far sampled.

We then demonstrated significant evidence for both Yamnaya ($|Z|= 2.341 - 9.596$) and EHG ($|Z|= 1.618 - 8.864$) introgression into each Irish Bronze Age sample when placed in a clade with the different members of MN/CA in turn using the statistic D(Mbuti, Yamnaya or EHG; Irish Bronze Age, MN/CA), with the strongest signals occurring with Ballynahatty and Spanish_CA as the MN/CA sample. Ballynahatty, of all Middle Neolithics, has the lowest affinity to EHG, while Esperstedt_MN has the highest, which may account for the variance in the level of Yamnaya/EHG introgression detected in the Irish Bronze depending on the MN/CA member used. Interestingly, no significant WHG gene flow into the Irish Bronze age was seen to the exclusion of MN/CA groups, with the reverse appearing to be the case for a majority combinations. This demonstrates that the increased affinity to Eastern Hunter Gatherer and Yamnaya ancestry in the Irish Bronze Age is not likely the result of more local hunter gatherer admixture.

We also showed that the Irish Bronze Age samples display significant Middle Neolithic introgression (|Z| = 3.485 - 9.911) when placed in a clade with Yamnaya (D(Mbuti, MN/CA; Irish Bronze, Yamnaya)). We observed different levels of introgression into the Irish Bronze Age depending on the MN/CA group or individual in question, with the highest Z scores obtained for Spanish_CA, Ballynahatty and the Scandinavian Individual, Gok2. The variance in detectable gene flow depending on the MN/CA member used, while possibly again the result of differing MN/CA affinities with Yamnaya/EHG as noted above, changed when the Irish Bronze Age samples were replaced by other Late Neolithic and Bronze Age (LN/BA) populations from Germany. These groups showed the highest levels of gene flow from Spanish_MN, Spanish_CA and Esperstedt_MN, with the exception of the Bell Beaker group who had the highest affinity with the Spanish_MN, Spanish_CA and Oetzi. The Nordic LN showed greatest introgression from the two Chalcolithics individuals, Oetzi and CO1.

The above result hinted at possible increased correlation between Ballynahatty and Irish Bronze Age compared to continental samples. However, the Irish Bronze Age showed no significant introgression from Ballynahatty when placed in a clade with continental LN/BA populations (D(Mbuti, Ballynahatty; Irish Bronze Age, Continental LN/BA)), with results actually suggesting higher correlation between Ballynahatty and certain continental samples, most significantly the Hungarian Bronze Age (|Z| = 1.909 - 2.881). A similar result was obtained when Ballynahatty was replaced with the Scandinavian Middle Neolithic Gok2. However, when Spanish_MN or Esperstedt_MN took Ballynahatty's place significant introgression into continental LN/BA populations to the exclusion of the Irish Bronze Individuals was seen, with the notable exception of the Nordic LN. This indicates that while the Irish Bronze Age contains a substantial amount of Middle Neolithic ancestry we find no strong evidence to suggest that the Irish population from which Ballynahatty came was its source. It is also not likely to be identical to the component of Middle Neolithic ancestry found in German LN/BA samples, or indeed any continental LN/BA so far sampled.

We also observed that for most cases the Irish Bronze Age samples form a clade with each other to the exclusion continental LN/BA samples (D(Mbuti, Continental LN/BA; X, Y), where X and Y are a pair of Irish Bronze Age individuals), with Rathlin3 showing slightly higher affinities to continental samples compared to the earlier Rathlin1 and Rathlin2. The exception to this is the strong signal of introgression seen into Rathlin1 from BenzigerodeHeimburg_LN when placed in clade with Rathlin2 or Rathlin3 (|Z| = 2.206 - 3.594). This preference is also seen in outgroup f_3-statistics.

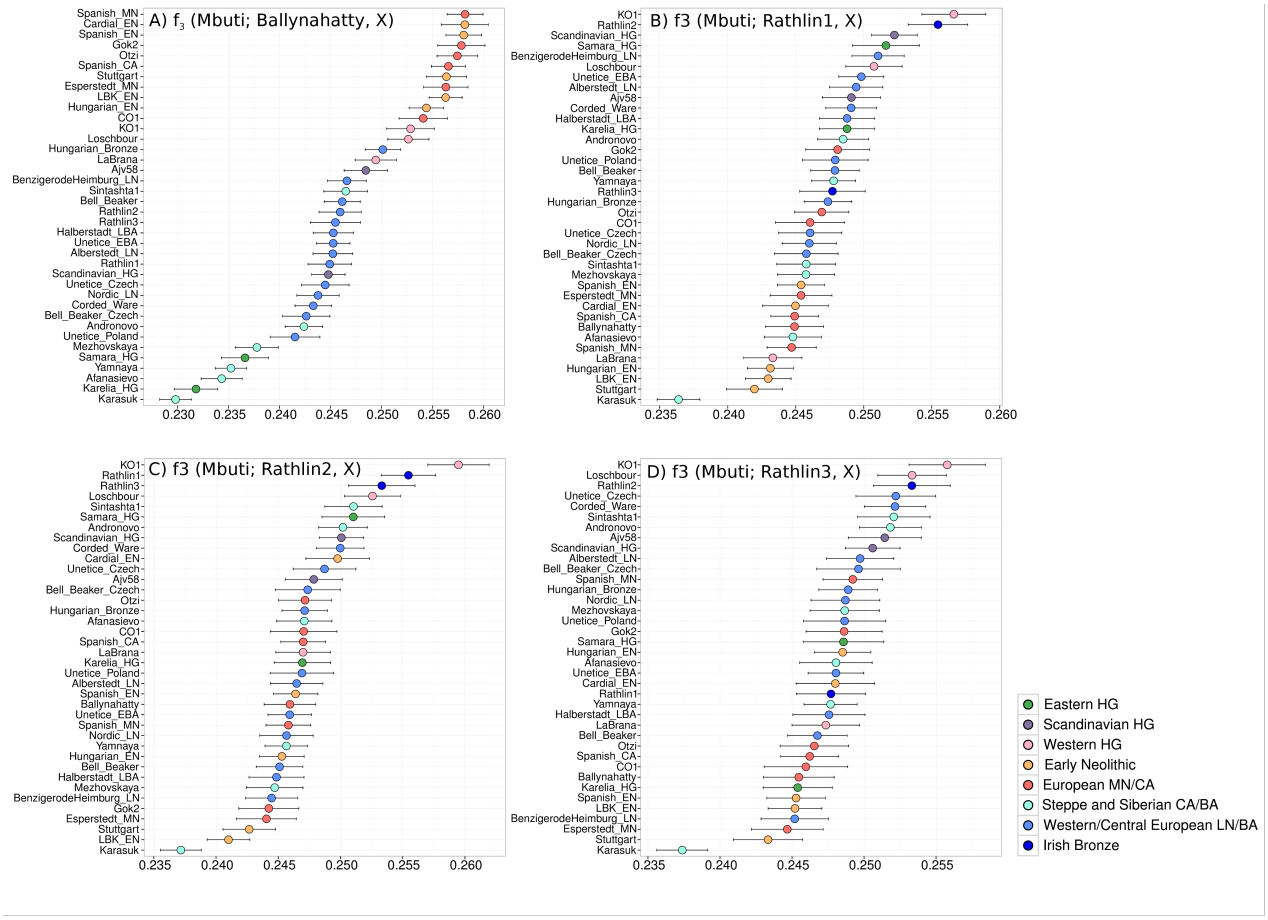
The above results, along with PCA and ADMIXTURE analysis, indicates that the Irish Bronze Age is composed of a mixture of Middle Neolithic and Steppe Ancestry as previously shown to be the case with other continental LN/BA samples (25, 26). To envisage the proportion of Yamnaya/EHG to Middle Neolithic ancestry in each Irish Bronze Age sample we implemented three methods. The first comes from ADMIXTURE analysis (Figure 1), where we make use of the fact that the green component of West and Central European ancestry does not appear until the Bronze Age. We presume an ultimate source of this component as the Yamnaya where it is present at an average ratio of 40:60 to the red 'hunter gatherer' component. This red component of ancestry is also in the European Bronze Age and most likely the result of contributions from both Middle Neolithic and Yamnaya populations. In our three Irish Bronze Age samples we see the green component at a level of 6-13%, which, taking into account the 40:60 ratio, implies a total of 14-33% Yamnaya ancestry and 67-86% Middle Neolithic in the Irish Bronze Age.

The second method estimated the amount of Middle Neolithic ancestry in the Irish Bronze Age by using the ratio f_4(Mbuti, Ballynahatty ; X, Dai)/f_4(Mbuti, Ballynahatty; Gok2, Dai). These two Middle Neolithics were chosen due to their lower affinity with Yamnaya and EHG compared to Esperstedt_MN and Spanish_MN. This gave us an estimated range of 72 ± 4% to 74 ± 5% for the Middle Neolithic component in the Irish Bronze Age genomes.

Finally, we followed the methods of (25), using the qpWave and qpAdm programs in the ADMIXTOOLS package (67), to estimate the mixture proportions of three different source populations, LBK (Early Neolithic), Loschbour (WHG) and Yamnaya, in the Irish Bronze Age group (the target population). A set of 15 outgroup populations (Ami, Biaka, Bougainville, Chukchi, Eskimo, Han, Hadza, Ju_hoan_North, Karitiana, Mbuti, Nganasan, Papuan, She, Ulchi and Yoruba), devoid of any recent gene flow with the source or target populations, was used to estimate mixture proportions, based on the idea that these outgroups are not identically related to the source populations. The results received indicated a WHG proportion at 26 ± 12%, Early Neolithic at 35 ± 6% and Yamnaya at 39 ± 8%

The overlap of these three estimates falls at approximately 32%.

Figure S12.1. Outgroup f_3-Statistics for each ancient Irish Individual. Tests in the form f_3(Mbuti; IA, X), where IA is an Irish ancient genome and X is any other ancient individual or population. Data points are coloured by archaeological context.



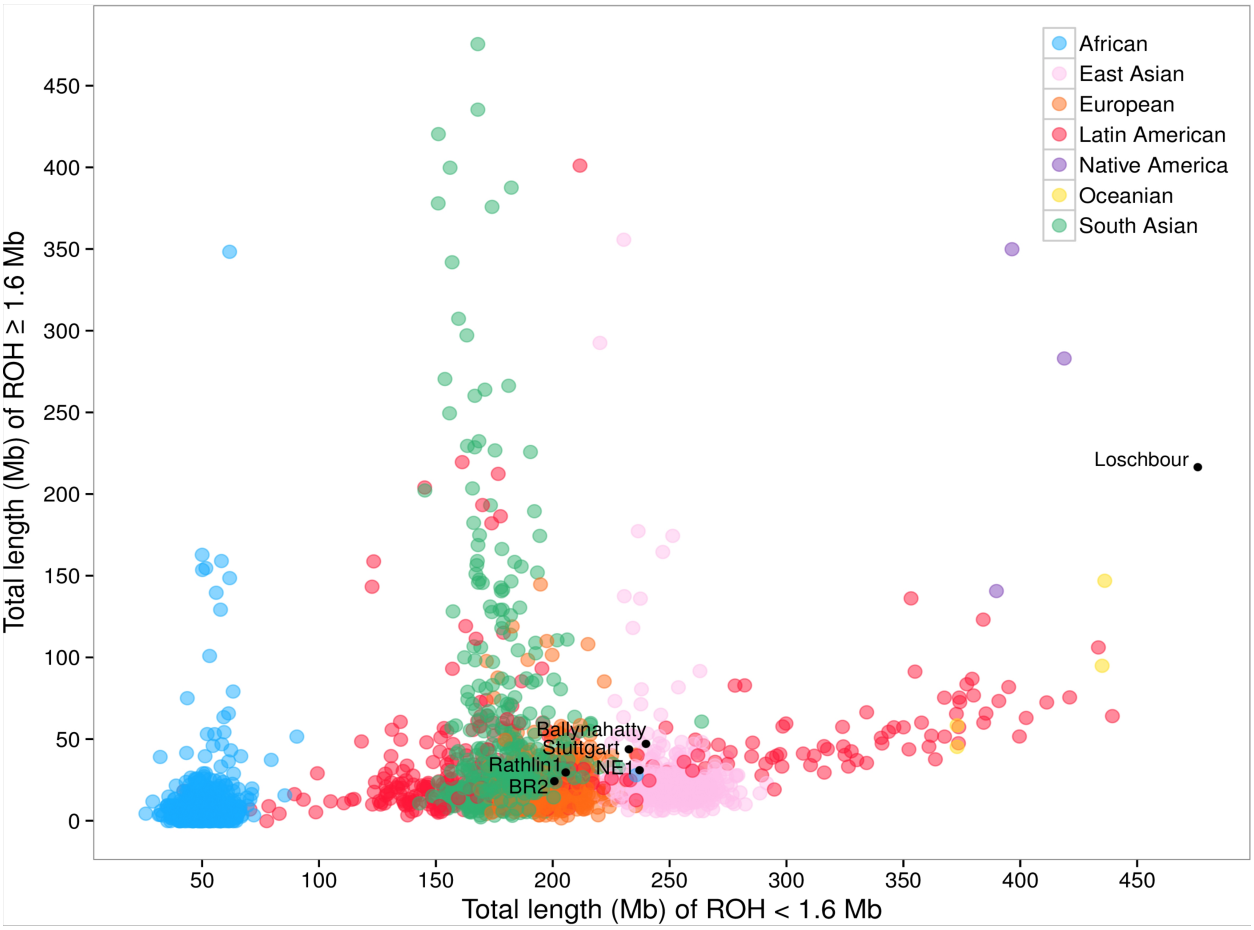
S13: Runs of Homozygosity

Runs of homozygosity (ROH) analysis was carried out on Ballynahatty and Rathlin1 and the four other high coverage ancient European samples available, one Mesolithic from Luxembourg (21), Loschbour, two Early Neolithics from Hungary (10) and Germany (21), NE1 and Stuttgart, and one Bronze Age sample from Hungary (10), BR2, alongside 2527 modern individuals using 1,601,314 transversion SNPs called securely across all samples (detailed dataset preparation is described in SI 9.3). ROH were called using PLINK v1.90 (54) following the specifications used in (10) (--homozyg, --homozyg-density 50, --homozyg-gap 100, --homozyg-kb 500, --homozyg-snp 50, --homozyg-window-het 1, --homozyg-window-snp 50 and --homozyg-window-threshold 0.05).

Individual ROH were divided first into two classes, long ROH (≥ 1.6 Mb), indicative of recent endogamy, and short to intermediate (<1.6 Mb), a result of more ancient population bottlenecks (71). For both classes, the summed total length of ROH was calculated for every individual. These values were then plotted against each other (Figure S13.1) giving an overview of ROH length distribution for all genomes, modern and ancient.

To investigate the length distribution of ROH in more detail we placed our ROH into size bins (72) and calculated the total length of ROH in each bin for each individual. For this analysis we selected single individuals from a subset of populations from 1000 Genomes Phase 3 v5, avoiding those known to have experienced substantial historical admixture. Each individual selected displayed the median value of that population for total fraction of the genome under ROH. In cases where two median individuals existed one was selected at random. We also filtered the 23 modern individuals from the Simons Genome Diversity Project (61) removing those belonging to a population already represented in the 1000 Genomes set. The remaining individuals exist in pairs belonging to various populations. We selected one individual at random from each pair for analysis, with the exception of those from Native American and Oceanian populations where both members of the pair were retained. This left us with a total of 32 modern individuals alongside our 6 ancient genomes for ROH size bin analysis, plotted in the Main Paper Fig. 2.

Figure S13.1. Total lengths of short to intermediate ROH (<1.6Mb) plotted against the total length of long ROH (≥1.6 Mb). ROH calls for 2527 modern and 6 ancient individuals were based on 1.6M autosomal transversions with minor allele frequencies above 0.5%



S14: Haplotype-based Analysis with ChromoPainter

We followed a similar approach to (21) and used FineStructure v.2 (www.paintmychromosomes.com; ref. 73) to investigate population structure in our data. The analysis was performed on the same set of high coverage ancient samples in the above S13. Diploid genotype calling is described in the final paragraph of S9.2. In order to maximize the number of SNPs for analysis, each high coverage sample was analysed separately with a set of modern Eurasian genotypes from the Hellenthal 2014 dataset (59) (Modern individuals are the same as those used in PCA; see Table S10.1), keeping only SNPs confidently called across all individuals within each dataset (238,762-356,240 SNPs, Table S14.1). Modern populations were divided into larger geographical groupings based on (59).

For each dataset, we split the genotype data by chromosome and phased genotypes with SHAPEIT v2.r778 (74). Haplotype files were converted to ChromoPainter format using "impute2chromopainter.pl" and a recombination map was created using "makeuniformrecfile.pl" both available at http://www.paintmychromosomes.com/. We then ran the ChromoPainter analysis, linked mode, first by estimating "mu" and "Ne", followed by estimating "c" and using "Chromocombine" to create a genome wide ChromoPainter output for all individuals. We calculated median chunk donation from each ancient individual to populations in the modern dataset (Table S14.2) in order to ascertain patterns of haplotypic affinity and performed linear regression analysis using the "visreg" package (75) of the R programming language (76) (Figure S14.1). The correlogram shown in Figure S14.2. was produced using the "corrgram" package (http://CRAN.R-project.org/package=corrgram).

To allow for a visual comparison of the geographical distribution of haplotype sharing between ancient samples and modern populations (Main paper, Fig. 3), we first calculated the median of the chunks donated by each ancient individual to each separate population (Table S14.2). These values were then were then normalised by scaling between 0 and 1 and plotted as interpolated maps using the *Geostatistical Analyst* tool of the ArcMap component of the ArcGIS (v10.1) software suite (ESRI).

Figure S14.1 Selected pairwise regression plots of haplotype donation by ancient genomes to a set of modern genomes. Colour keys are West Asia (brown), South Middle East (orange), East Europe (purple) South Europe (blue) and Northwest Europe (green). Specific samples are labelled: En - English; Ir - Ireland; Sa - Sardinian; Sc - Scottish; Sp - Spanish; We - Welsh.

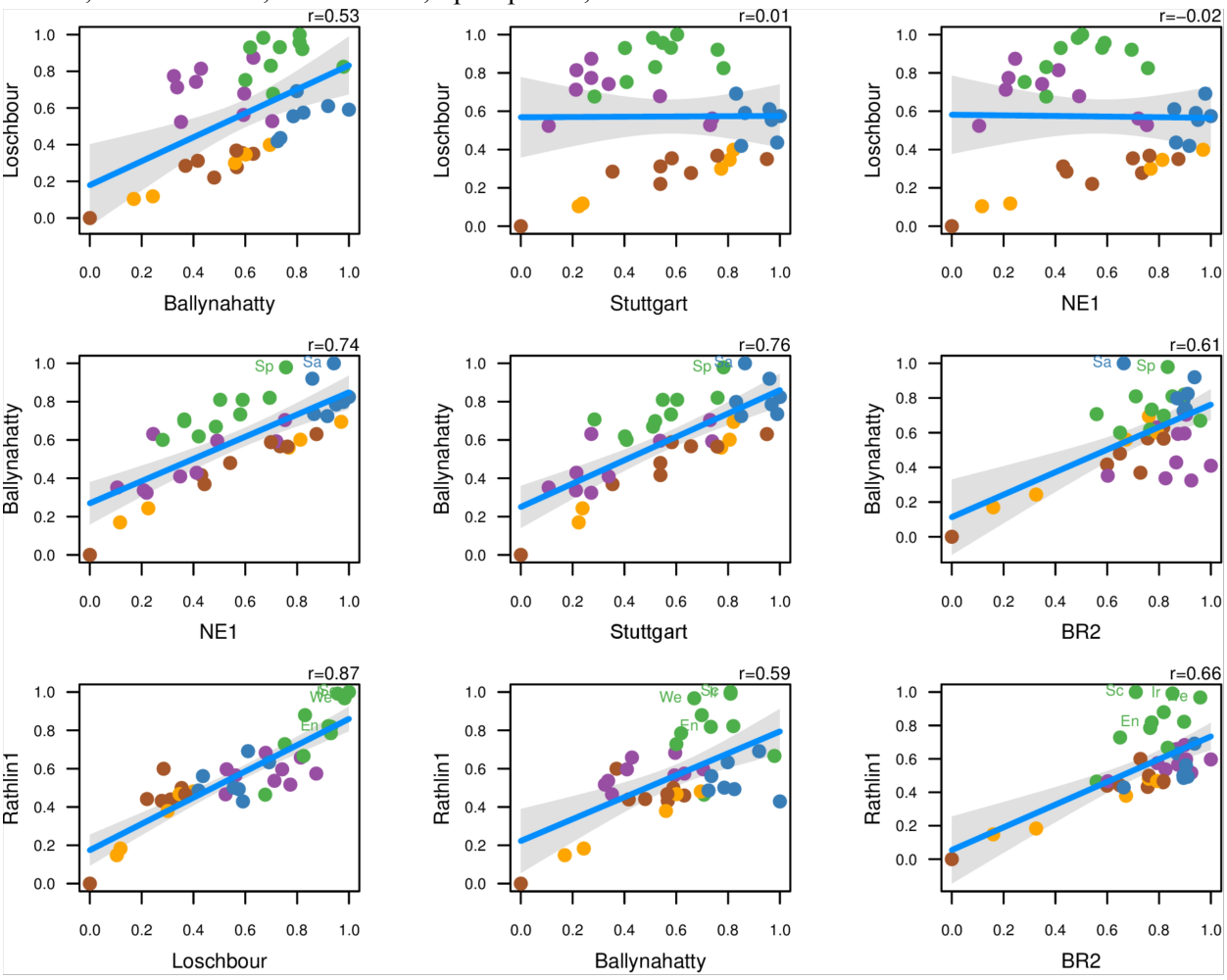


Figure S14.2. Correlation matrix plot of median chunk donation from ancient samples to modern populations. Left-hand side numbers represent correlation coefficients, confidence intervals and p-values for the pairwise correlation tests.

Table S14.1. Number of SNPs used in each separate ChromoPainter run.

Ancient Sample	Context	SNPs
Ballynahatty	Irish Middle Neolithic	238762
Rathlin1	Irish Bronze Age	257516
BR2	Hungarian Bronze Age	356240
LBK	German Early Neolithic	334345
Loschbour	Luxembourg Mesolithic	347914
NE1	Hungarian Early Neolithic	347671

Each ancient sample was merged separately with the modern dataset. Only genotypes called at high quality in the ancient sample were retained for analysis.

Table S14.2. Median haplotype donation from high coverage Ancients to modern populations.

Population	Hellenthal et al. 2014 Grouping	Ballynahatty	NE1	Rathlin1	BR2	Loschbour	Stuttgart
Adygei	West Asia	20.68	24.661	23.240	23.488	20.572	26.263
Armenian	West Asia	23.376	29.392	23.060	25.908	19.577	28.188
Basque	Northwest Europe	25.878	23.644	23.824	22.865	31.063	22.107
Bedouin	South Middle East	16.267	19.79	16.341	16.707	14.589	21.112
Belorussian	East Europe	20.903	24.381	28.418	27.603	35.017	20.967
Bulgarian	East Europe	23.838	29.17	26.210	27.713	27.767	29.526
Chuvash	East Europe	19.525	19.618	23.865	23.528	26.663	19.217
Cypriot	West Asia	24.509	31.578	23.720	26.83	21.676	32.972
Druze	West Asia	13.231	17.977	12.825	14.239	11.596	17.465
EastSicilian	South Europe	27.28	32.751	24.689	28.033	27.559	33.268
Egyptian	South Middle East	25.657	33.06	24.220	25.977	23.075	30.873
English	Northwest Europe	26.346	27.001	32.213	26.143	38.388	26.927
Finnish	East Europe	24.521	21.779	26.431	26.49	36.735	21.907
French	Northwest Europe	27.901	28.77	32.299	28.072	38.089	29.863
Georgian	West Asia	21.815	26.399	23.284	24.252	17.939	26.256
GermanyAustria	Northwest Europe	25.721	23.648	33.658	26.862	35.486	25.931
Greek	South Europe	26.191	32.232	24.341	28.043	23.651	31.351
Hungarian	East Europe	23.889	25.62	28.995	28.098	31.116	26.236
Iranian	West Asia	23.758	28.846	24.648	25.984	21.784	26.979
Ireland	Northwest Europe	27.712	27.146	36.313	27.383	39.116	26.412
Jordanian	South Middle East	23.25	29.913	21.811	24.617	20.226	30.092
Lezgin	West Asia	19.837	24.858	27.028	25.472	19.786	23.243
Lithuanian	East Europe	19.034	21.383	25.067	28.503	33.855	21.901
NorthItalian	South Europe	27.514	33.191	27.829	27.658	31.498	31.03
Norwegian	Northwest Europe	24.298	24.512	31.425	26.05	38.35	24.028
Orcadian	Northwest Europe	23.969	22.349	30.072	24.263	33.238	24.132
Palestinian	South Middle East	17.58	21.484	17.164	19.261	14.999	21.354
Polish	East Europe	20.559	23.398	26.957	29.67	32.947	23
Romanian	East Europe	25.823	29.687	26.962	28.193	26.781	29.392
Russian	East Europe	19.257	21.215	25.537	26.969	32.093	20.937
Sardinian	South Europe	31.118	32.621	22.989	24.481	28.592	31.59
Scottish	Northwest Europe	27.715	25.798	36.512	25.197	40.362	27.316
SouthItalian	South Europe	26.385	31.434	26.127	28.182	24.148	33.621
Spanish	Northwest Europe	30.743	29.748	28.613	27.097	35.322	30.236
Syrian	South Middle East	23.988	30.613	23.896	26.466	21.554	30.624
Turkish	West Asia	23.346	29.84	23.886	26.855	22.182	29.845
Tuscan	South Europe	29.681	31.328	29.202	28.696	29.158	33.133
Welsh	Northwest Europe	25.202	25.539	35.745	29.037	39.871	25.798
WestSicilian	South Europe	27.977	33.526	24.509	28.289	28.124	33.792

S15: Phenotypic Analysis

A panel of SNPs, small indels and multi-allelic markers were chosen for examination in the four ancient Irish humans. This was composed of variants shown to have well-established phenotypic associations with traits such as pigmentation, immunity, disease susceptibility, diet and blood group. Also included were markers linked to genetic disorders relatively common in Ireland today.

For the two genomes with coverage over 10X (Ballynahatty and Rathlin1) genotype calls were generated using UnifiedGenotyper in GATK v2.4-7 (15). Only bases with a quality above 30 were considered for calling (-mbq 30). All genotype calls are reported in the accompanying tables, however genotypes with a quality below 30 are not considered to be reliable.

We could not call diploid genotypes to a satisfactory level of accuracy in the two genomes under 2X coverage (Rathlin2 and Rathlin3) using the UnifiedGenotyper tool. Instead we use the Pileup tool in GATK v2.4 (15) to report the base calls for all reads covering each site with a base quality over 30.

S15.1 Pigmentation

To investigate the pigmentation phenotype of each ancient Irish human two approaches were taken. For the two higher coverage samples the Hirisplex SNP panel for forensic hair and eye colour inference was utilized (77). These SNPs are also reported for the low coverage genomes, however lack of data prevented accurate results from the prediction models (Table S15.1). The second approach, based on (21), saw both low and high coverage samples being assessed at SNPs marking several haplotypes in the HERC2/OCA2 region associated with eye colour (78, 79) (Tables S15.3 and S15.4), as well as 16 SNPs comprising a 78-kb SLC24A5 haplotype (C11) that is strongly associated with the derived light skin colour allele A111T of rs1426654 (80) (Table S15.5).

The Neolithic individual, Ballynahatty, was assigned a dark, black hair colour and brown eyes by the Hiriplex prediction model at high probability scores (Table S15.2). The predicted eye colour was in agreement with the observed genotypes of Ballynahatty in SNPs marking several haplotypes in the HERC2/OCA2 region where the individual was either homozygous for the allele not associated with blue

eyes or a heterozygote. In regards to skin colour, Ballynahatty was found to be homozygous for the SLC24A5 haplotype C11 associated with lighter pigmentation.

For the high coverage Bronze Age Individual, Rathlin1, the Hirisplex probability scores, although not high enough for confident predictions, suggest a light hair shade, possibly blonde in colour, and brown eyes. Examining the various haplotypes in the HERC2/OCA2 region Rathlin1 was found to be a heterozygote in each case for the blue eye colour associated haplotype. The individual was also found to be homozygous for the SLC24A5 pigmentation-lightening haplotype C11.

In the two lower coverage Bronze Age individuals, Rathlin2 and Rathlin3, we found both displayed the blue eye associated allele in the vast majority base calls at sites marking several haplotypes in the HERC2/OCA2 region. Both samples also appear to carry the C11 haplotype of the SLC24A5 region.

S15.2 ABO blood groupings

To ascertain each sample's ABO blood group genotypes a panel of six diagnostic SNPs in the ABO gene were assessed (Table S15.6). Three of these, rs8176719 (81), rs505922 and rs687289 (82), were used as markers for the O allele, two for the B allele, rs8176746 and rs8176746, and one, rs8176704, for the A2 allele (82). Ballynahatty and Rathlin1 both presented as heterozygotes for the ancestral A1 allele and the O allele. Rathlin2 is also a possible A1/O or B/O heterozygote, however low coverage depth at the diagnostic sites and the risk of deamination obscuring results does not allow us to confirm this result. Finally, Rathlin3 appears to contain at least one A1 allele.

S15.3 Genetic Disorders at high frequency in present day Ireland

There are a number of inherited conditions common in Europe, for which some of their causative mutations reach peak frequencies either on the island of Ireland or more generally in Northwestern Europe. We chose to examine five of these disorders, namely haemochromatosis, phenylketonuria (PKU), α_1-antitrypsin deficiency, classic galactosemia (GALT) and cystic fibrosis (CF).

Haemochromatosis is an autosomal recessive condition that occurs in European populations, typified by iron accumulation. Three point mutations in the HFE gene are responsible for the majority of cases, C282Y, H63D and S65C, with C282Y being by far the most penetrant of the three. Ballynahatty appeared

to be a heterozygote for H63D (GATK quality 99; SNP coverage 10X; Allele depths 6X and 4X) and Rathlin1 a heterozygote for C282Y (GATK quality 95; SNP coverage 6X; Allele depths 3X and 3X). As C282Y is a G to A mutation we examined the SNP visually in IGV (83), to check for any higher risk of deamination at these derived A alleles due to read position. However, all A alleles were found to be low risk, situated away from the 3' prime end of reads, with high base quality scores after mapDamage rescaling.

PKU, characterised by an increase in blood phenylalanine concentrations sometimes resulting in hyperphenylalaninaemia, is caused by mutations in phenylalanine hydroxylase gene (PAH), over 400 of which have been identified . The incidence of PKU in Europe is about one case per 10 000 livebirths. This rises to a peak of one in every 4000 in Ireland (84, 85). We examined 29 of the most common PKU mutations in Europe (86) in our samples, including R408W, which is found at highest frequency in the west of Ireland (86), however none of our ancient individuals were found to carry a causative allele. α_1-antitrypsin deficiency, an autosomal co-dominant disorder, clinically characterised by liver disease and early-onset emphysema, is most frequently caused by one of two mutations in the α_1-antitrypsin deficiency gene (AAT), the S and the Z alleles (87). The S allele reaches its highest frequencies in Iberia, while the Z allele is most prevalent in the UK, coastal France and the Baltic countries (88). Neither of these alleles were identified in our ancient Irish individuals. Ireland has the world's highest prevalence of cystic fibrosis, an autosomal recessive disorder that affects the lungs and digestive system (89). While over 1700 mutations have been identified in the CFTR gene less than 20 occur at a frequency of over 0.1%, while a single mutation ΔF508 accounts for approximately 70% of CF cases (90). None of these mutations were identified in our ancient samples. Classic galactosemia (GALT) is an autosomal recessive metabolic disorder than can result in liver disease, and occurs relatively frequently in Ireland. We explored three mutations commonly associated with the disease, however none of the ancient samples carried a causative allele.

S15.4 Classical HLA typing

Numerous studies have reported a significant association with the C282Y mutation and the HLA-A*03-B*07 haplotype (91–95). To investigate this we used HLA*IMP:02 (96) to infer classical HLA types in the two high coverage genomes. Diploid genotypes for a set of SNPs from HapMap 3 covering the xMHC region of chromosome 6 were called in Ballynahatty and Rathlin1 using the UnifiedGenotyper tool in GATK v2.4-7 (15). Only bases with a quality above 30 were considered for calling (-mbq 30). Genotype

calls were filtered for a depth of coverage of 15X or above and a genotype quality of 30 or above. Data conversion to HLA*IMP format, quality control, and alignment to the HapMap reference strand was performed by the HLA Imputation front-end program. This resulted in 5262 genotypes in Rathlin1 and 4540 genotypes in Ballynahatty covering the xMHC region used for imputation of HLA types, which was carried out by the HLA*IMP web service. The results are displayed in Table S15.8. Note that these do not contain inter-locus phasing information.

Rathlin1 was found to be a carrier of both the HLA-A*0301 and HLA-B*0702 alleles, associated with the C282Y mutation, further supporting the sample's heterozygote status. Furthermore, Rathlin1, also carries the HLA alleles C*0702, DRB1*1501, and DQB1*0602, which together with A*0301 and B*0702 compose the PGF haplotype, one of the most frequent long-range MHC in Northern Europeans (~10%). This haplotype has been shown to offer protection against type 1 diabetes, while predisposing to other diseases such as multiple sclerosis and systemic lupus erythematosus (97–99).

S15.5 SNPs with evidence of recent positive selection in humans

Sample genotypes were also assessed at panel of 19 SNPs that display evidence of recent positive selection, the majority of which have well validated phenotypic associations to diet and immunity. These SNPs were chosen based on a similar SNP panel in the previous high coverage ancient genome study by (21). Sample genotypes and base calls are shown in Table S15.9

The most well documented genomic region associated with dietary change is the LCT lactase persistence locus. The high coverage Neolithic individual, Ballynahatty, was homozygous for the ancestral alleles at both the LCTa and LCTb loci, while the high coverage Bronze Age male was a heterozygote at both positions. Base calls for the two lower coverage Bronze Age samples indicated the presence of the ancestral allele in these individuals.

All four samples, where accurate genotyping was possible, appeared to carry the ancestral alleles at ALDH2, ADH1Ba and ADH1Bb. The derived versions of these alleles are linked to slower alcohol metabolism.

We also evaluated sample genotypes at 6 SNPs in the NAT2 gene used by the NAT2pred tool (http://nat2pred.rit.albany.edu/) to determine acetylation phenotype, which can be rapid, intermediate or

slow. The slow phenotype is thought to be under positive selection, possibly related to the advent of farming (100). Phenotype prediction was only possible for our two high coverage genomes, both of which were discovered to be intermediate acetylators.

The final two loci linked to diet examined were FADS1 and NADSYN1, both of which were shown to carry a signal of selection (101), with the most significant SNPs identified as rs174546 and rs7940244 respectively. FADS1 is involved in fatty acid metabolism with the derived allele at rs174546 associated with a decrease in triglyceride levels (102). The Neolithic individual, Ballynahatty, was found to be a heterozygote at this position, while Rathlin1 is most likely a homozygote for the ancestral allele. Variation in NADSYN1 has been linked to circulating vitamin D levels, with high variability in derived allele frequency at rs7940244 across Northern European populations (103, 104). Both high coverage individuals were found to be homozygous for the ancestral allele at this position.

Ancestral alleles were observed at rs17822931 (ABCC1), which determines earwax type (105), and rs3827760 (EDAR), involved in tooth morphology and hair thickness (106). The derived allele at both these loci is found at high frequencies in East Asians and Native Americans today.

Table S15.1. Ballynahatty, Rathlin1, Rathlin2 and Rathlin3 genotype and base calls for the Hirisplex 24 SNP panel for pigmentation phenotype prediction.

Gene	SNP ID	Ancestral Allele (fwd)	Derived Allele (fwd)	Ballynahatty (GQ)	Rathlin1 (GQ)	Rathlin2	Rathlin3
MC1R	N29insA	C	insA	C/C (21)	C/C (21)	-	-
MC1R	rs11547464	G	A	G/G (57)	G/G (60)	G	-
MC1R	rs885479	G	A	G/G (33)	G/G (39)	GG	-
MC1R	rs1805008	C	T	C/C (21)	C/C (33)	C	-
MC1R	rs1805005	G	T	G/G (30)	G/G (39)	GG	-
MC1R	rs1805006	C	A	C/C (30)	C/C (39)	CCC	-
MC1R	rs1805007	C	T	C/C (42)	C/C (48)	C	C
MC1R	rs1805009	G	C	G/G (30)	G/G (30)	-	-
MC1R	rs201326893 (Y152OCH)	C	A	C/C (36)	C/C (60)	C	C
MC1R	rs2228479	G	A	G/G (24)	G/G (51)	GG	-
MC1R	rs1110400	T	C	T/T (33)	T/T (48)	T	T
SCL45A2	rs28777	C	A	A/A (45)	A/A (21)	AA	-
SLC45A2 (MATP)	rs16891982	C	G	C/G (99)	G/G (39)	GGGG	G
KITLG	rs12821256	T	C	T/T (39)	T/T (33)	CCCCT	-
EXOC2	rs4959270	C	A	A/A (30)	C/A (91)	-	A
IRF4	rs12203592	C	T	T/T (27)	C/C (42)	T	-
TYR	rs1042602	C	A	C/C (21)	C/A (99)	C	A
OCA2	rs1800407	C	T	C/C (24)	C/C (30)	CCCCC	CC
SLC24A4	rs2402130	G	A	A/A (30)	A/A (27)	AA	AA
HERC2	rs12913832	A	G	A/A (30)	A/G (99)	-	-
PIGU	rs2378249	A	G	A/A (27)	A/A (27)	AA	-
SLC24A4	rs12896399	G	T	G/T (99)	T/T (36)	TTT	G
TYR	rs1393350	G	A	A/G (99)	G/G (57)	GG	G
TYRP1	rs683	A	C	A/C (99)	C/C (42)	A	A

GQ refers to the Genotype Quality outputted from GATK UnifiedGenotyper for the genotype calls of the two genomes with coverage over 10X (Ballynahatty and Rathlin1). Genotype calls are not shown for the two genomes under 2X coverage (Rathlin2 and Rathlin3). Instead all bases calls with a quality above 30 are reported.

Table S15.2. Hirisplex probabilities for hair and eye colour in Ballynahatty and Rathlin1.

		Ballynahatty	Rathlin1
Hair Colour	Brown	0.137	0.276
	Red	0	0.001
	Black	**0.861**	0.282
	Blond	0.002	**0.441**
Hair Shade	Light	0.005	**0.614**
	Dark	**0.995**	0.386
Eye Colour	Blue	0.002	0.229
	Intermediate	0.062	0.128
	Brown	**0.973**	**0.643**

The highest probability values for hair shade, hair colour and eye colour are marked in bold.

Table S15.3. Ballynahatty, Rathlin1, Rathlin2 and Rathlin3 genotype and base calls for three blue-eye associated haplotypes in the OCA2-HERC2 region reported in Donnelly et al., 2012 (79)

Haplotype	Gene	SNP ID	Ancestral Allele (fwd)	Blue Eye Associated Allele (fwd)	Ballynahatty (GQ)	Rathlin1 (GQ)	Rathlin2	Rathlin3
BEH1	OCA2	rs4778138	G	A	A/G (99)	A/G (99)	AAA	A
BEH1	OCA2	rs4778241	A	C	A/A (45)	A/C (99)	A	CC
BEH1	OCA2	rs7495174	G	A	A/G (55)	A/G (51)	AA	A
BEH2	HERC2	rs1129038	C	T	C/C (42)	C/T (99)	TT	-
BEH2	HERC2	rs12913832	A	G	A/A (30)	A/G (99)	-	-
BEH3	HERC2	rs916977	T	C	T/T (33)	T/C (99)	C	TC
BEH3	HERC2	rs1667394	C	T	C/C (51)	C/T (99)	T	-

GQ refers to the Genotype Quality outputted from GATK UnifiedGenotyper for the genotype calls of the two genomes with coverage over 10X (Ballynahatty and Rathlin1). Genotype calls are not shown for the two genomes under 2X coverage (Rathlin2 and Rathlin3). Instead all bases calls with a quality above 30 are reported.

Table S15.4. Ballynahatty, Rathlin1, Rathlin2 and Rathlin3 genotype and base calls for a 13 SNP haplotype in the OCA2-HERC2 region identified in Eiberg et al., 2008 (78) and found in 97% of people with blue eyes.

Gene	SNP ID	Other Allele (fwd)	Blue Eye Associated Allele (fwd)	Ballynahatty (GQ)	Rathlin1 (GQ)	Rathlin2	Rathlin3
OCA2	rs4778241	A	C	A/A (45)	A/C (99)	A	CC
HERC2	rs1129038	C	T	C/C (42)	C/T (99)	TT	-
HERC2	rs12593929	G	A	G/A (59)	G/A (53)	-	A
HERC2	rs12913832	A	G	A/A (30)	A/G (99)	-	-
HERC2	rs7183877	A	C	C/A (50)	C/C (30)	-	C
HERC2	rs3935591	T	C	T/T (45)	T/C (99)	CCC	CCC
HERC2	rs7170852	T	A	T/T (48)	T/A (99)	AAA	A
HERC2	rs2238289	G	A	G/G (24)	A/G (99)	A	A
HERC2	rs3940272	T	G	T/T (6)	T/T (3)	-	-
HERC2	rs8028689	C	T	T/C (99)	T/C (87)	T	TT
HERC2	rs2240203	C	T	T/C (99)	T/C (99)	T	-
HERC2	rs11631797	A	G	A/A (15)	A/G (98)	G	-
HERC2	rs916977	T	C	T/T (33)	T/C (99)	C	TC

GQ refers to the Genotype Quality outputted from GATK UnifiedGenotyper for the genotype calls of the two genomes with coverage over 10X (Ballynahatty and Rathlin1). Genotype calls are not shown for the two genomes under 2X coverage (Rathlin2 and Rathlin3). Instead all bases calls with a quality above 30 are reported.

Table S15.5. Ballynahatty, Rathlin1, Rathlin2 and Rathlin3 genotype and base calls for 16 SNPs marking the SLC24A5 haplotype (C11) found in virtually all chromosomes carrying the derived light skin colour allele (A111T).

SNP ID	Ancestral Allele (fwd)	Derived allele (fwd)	Ballynahatty (GQ)	Rathlin1 (GQ)	Rathlin2	Rathlin3
rs1834640	G	**A**	A/A (24)	A/A (33)	A	-
rs2675345	G	**A**	A/A (18)	A/A (33)	-	-
rs2469592	G	**A**	A/A (21)	A/A (30)	A	AAAA
rs2470101	C	**T**	T/T (45)	T/T (27)	TT	-
rs938505	**C**	T	C/C (30)	C/C (33)	CC	-
rs2433354	T	**C**	C/T (16)	C/C (30)	CC	C
rs2459391	G	**A**	A/A (42)	A/A (60)	A	A
rs2433356	A	**G**	G/G (48)	G/G (51)	GGG	G
rs2675347	G	**A**	A/A (30)	A/A (18)	AA	AA
rs2675348	G	**A**	A/A (30)	A/A (33)	AAA	-
rs1426654 (A111T)	G	**A**	A/A (42)	A/A (57)	-	-
rs2470102	G	**A**	A/A (39)	A/A (33)	A	-
rs16960631	**A**	G	A/A (21)	A/A (54)	AA	-
rs2675349	G	**A**	A/A (33)	A/A (42)	A	-
rs3817315	T	**C**	C/C (51)	C/C (39)	CCC	-
rs7163587	T	**C**	C/C (42)	C/C (36)	-	C

The allelic state associated with C11 is marked in bold. GQ refers to the Genotype Quality outputted from GATK UnifiedGenotyper for the genotype calls of the two genomes with coverage over 10X (Ballynahatty and Rathlin1). Genotype calls are not shown for the two genomes under 2X coverage (Rathlin2 and Rathlin3). Instead all bases calls with a quality above 30 are reported.

Table S15.6. Ballynahatty, Rathlin1, Rathlin2 and Rathlin3 genotype and base calls for a panel of 6 SNPs used to diagnose ABO blood type.

SNP ID	Ancestral A1 Allele (fwd)	Derived Allele (fwd)	Allele Diagnosed	Ballynahatty (GQ)	Rathlin1 (GQ)	Rathlin2	Rathlin3
rs8176719	C	del	O	C/del (99)	C/del (99)	-	CC
rs505922	C	T	O	C/T (59)	C/T (99)	TC	-
rs687289	A	G	O	A/G (19)	A/G (96)	AG	-
rs8176746	G	T	B	G/G (3)	G/G (18)	-	G
rs8176747	C	G	B	C/C (3)	C/C (30)	-	C
rs8176704	G	A	A2	G/G (54)	G/G (48)	G	G
	Predicted Blood Group Genotype:			A1/O	A1/O	-	-

GQ refers to the Genotype Quality outputted from GATK UnifiedGenotyper for the genotype calls of the two genomes with coverage over 10X (Ballynahatty and Rathlin1). Genotype calls are not shown for the two genomes under 2X coverage (Rathlin2 and Rathlin3). Instead all bases calls with a quality above 30 are reported.

Table S15.7. Ballynahatty, Rathlin1, Rathlin2 and Rathlin3 genotype and base calls for SNPs associated with autosomal recessive disorders common in Ireland today.

α1-Antitrypsin deficiency (AAT gene)							
Mutation	**SNP ID**	**Ancestral Allele (fwd)**	**Derived Allele (fwd)**	**Ballynahatty (GQ)**	**Rathlin1 (GQ)**	**Rathlin2**	**Rathlin3**
E342K (PI*Z)	rs28929474	C	T	C/C (57)	C/C (60)	CCC	C
E264V (PI*S)	rs17580	T	A	T/T (48)	T/T (27)	-	TTT
Haemochromatosis (HFE gene)							
Mutation	**SNP ID**	**Ancestral Allele (fwd)**	**Derived Allele (fwd)**	**Ballynahatty (GQ)**	**Rathlin1 (GQ)**	**Rathlin2**	**Rathlin3**
H63D	rs1799945	C	G	C/G (99)	C/C (42)	C	CCC
S65C	rs1800730	A	T	A/A (33)	A/A (48)	A	AAA
C282Y	rs1800562	G	A	G/A (24)	G/A (95)	GGG	-
Cystic Fibrosis (CFTR gene)							
Mutation	**SNP ID**	**Ancestral Allele (fwd)**	**Derived Allele (fwd)**	**Ballynahatty (GQ)**	**Rathlin1 (GQ)**	**Rathlin2**	**Rathlin3**
394delTT	rs121908769	TT	del	TT/TT (30)	TT/TT (36)	T	T
R117H	rs78655421	G	A	G/G (27)	G/G (42)	GGG	-
1078delT	rs121908744	T	del	T/T (27)	T/T (42)	-	T
A455E	rs74551128	C	A	C/C (21)	C/C (6)	-	C
Delta I507	rs121908745	TCT	del	TCT/TCT (36)	TCT/TCT(36)	T	-
Delta F508	rs113993960	CTT	del	CTT/CTT (42)	CTT/CTT (36)	CC	-
IVS10, G-A, -1	rs76713772	G	A	G/G (42)	G/G (36)	G	GG
G542X	rs113993959	G	T	G/G (33)	G/G (36)	G	G
G551D	rs75527207	G	A	G/G (21)	G/G (27)	G	-
R553X	rs74597325	C	G	C/C (21)	C/C (21)	C	-
A561E	rs121909047	C	A	C/C (42)	C/C (24)	CC	-
W846X	rs267606722	G	A	G/G (24)	G/G (18)	-	G
R1066C	rs78194216	C	T	C/C (24)	C/T (7)	CC	C
R1162X	rs74767530	C	T	C/C (51)	C/C (27)	-	-
3659delC	rs121908811	C	del	C/C (51)	C/C (33)	CC	CCCC
W1282X	rs77010898	G	A	G/G (36)	G/G (36)	GGG	-

N1303K	rs80034486	C	G	C/C (21)	C/C (15)	CCCC	C

Phenylketonuria (PAH gene)

Mutation	SNP ID	Ancestral Allele (fwd)	Derived Allele (fwd)	Ballynahatty (GQ)	Rathlin1 (GQ)	Rathlin2	Rathlin3
IVS12DS, G-A, +1	rs5030861	C	T	C/C (33)	C/C (30)	CCC	C
Y414C	rs5030860	T	C	T/T (24)	T/T (36)	TTTT	TT
R408Q	rs5030859	C	T	C/C (24)	C/C (33)	CCCC	C
R408W-H1	rs5030858	G	A	G/A (16)	G/G (36)	GGGG	G
A403V	rs5030857	G	A	G/G (33)	G/G (45)	GGG	-
E390G	rs5030856	T	C	T/T (54)	T/T (54)	TTTTTT	TT
V388M	rs62516101	C	G,T	C/C (45)	C/C (51)	CCCCCC	CC
IVS10AS, G-A, -11	rs5030855	C	T	C/C (39)	C/C (54)	C	C
S349P	rs62508646	A	C,G	A/A (42)	A/A (33)	-	AA
L348V	rs62516092	G	C	G/G (42)	G/G (33)	-	GG
A300S	rs5030853	C	A	C/C (42)	C/C (36)	C	C
F299C	rs62642933	A	C	A/A (39)	A/A (36)	A	A
IVS7DS, G-A, +1	rs5030852	C	A,T	C/C (27)	C/C (45)	-	-
P281L	rs5030851	G	A	G/G (30)	G/G (42)	-	-
E280K	rs62508698	C	G,T	C/C (33)	C/C (36)	C	-
G272X	rs62514952	C	A	C/C (33)	C/C (33)	CC	-
R261Q	rs5030849	C	G,T	C/C (27)	C/C (27)	C	-
R261X	rs5030850	G	C,A	G/G (27)	G/G (27)	GG	-
R252W	rs5030847	G	C,A	G/G (24)	G/G (27)	GG	-
R243X	rs62508588	C	T,A	C/C (39)	C/C (24)	C	-
R243X	rs5030846	G	A	G/G (45)	G/G (27)	G	-
W187X	rs62507336	C	T,G	C/C (24)	C/C (21)	C	-
W187X	rs62507272	A	G	A/A (21)	A/A (24)	A	-
R158Q	rs5030843	C	T,G	C/T (4)	C/C (51)	CC	C
R111X	rs76296470	G	A	G/G (45)	G/G (51)	G	G
A104D	rs62642929	G	T	G/G (39)	G/G (54)	G	-
I65T	rs75193786	A	T,C,G	A/A (63)	A/A (33)	AA	A
L48S	rs5030841	A	G	A/A (42)	A/A (24)	AAAA	A
G46S	rs74603784	C	T	C/C (42)	C/C (21)	CCCC	C
F39L	rs62642926	G	C	G/G (33)	G/G (15)	GGG	G

Classic Galactosemia (GALT)

Mutation	SNP ID	Ancestral Allele (fwd)	Derived Allele (fwd)	Ballynahatty (GQ)	Rathlin1 (GQ)	Rathlin2	Rathlin3
Q1888R	rs75391579	A	G	A/A (33)	A/A (39)	AAAA	A
R333G	rs111033800	C	T,G	C/T (1)	C/C (45)	CCC	CC
F194L	rs111033726	T	C	T/T (45)	T/T (57)	TTT	T

GQ refers to the Genotype Quality outputted from GATK UnifiedGenotyper for the genotype calls of the two genomes with coverage over 10X (Ballynahatty and Rathlin1). Genotype calls are not shown for the two genomes under 2X coverage (Rathlin2 and Rathlin3). Instead all bases calls with a quality above 30 are reported.

Table S15.8. HLA*IMP2 results for high coverage Irish genomes.

Sample	HLA Locus	A	B	C	DPB1	DQA1	DQB1	DRB1
Ballynahatty	**Allele 1**	0201	5101	1402	0401	0301	0402	0801
	Posterior Probability	1	0.993	1	0.933	1	1	1
	Allele 2	2402	1302	0501	0201	0401	0302	0401
	Posterior Probability	0.973	0.242	0.987	0.84	0.993	0.993	0.873
Rathlin1	**Allele 1**	0301	0702	0702	0201	0102	0602	1501
	Posterior Probability	1	1	0.993	0.580	1	0.993	1
	Allele 2	3101	2705	0202	1301	0103	0603	1301
	Posterior Probability	0.653	0.973	1	0.320	0.953	1	0.980

Posterior probabilities for the most likely HLA alleles are shown for each locus.

Table S15.9. Ballynahatty, Rathlin1, Rathlin2 and Rathlin3 genotype and base calls for SNPs that have undergone recent positive selection in modern humans.

Gene	Associated Trait/Function	SNP ID	Ancestral Allele (fwd)	Derived Allele (fwd)	Ballynahatty (GQ)	Rathlin1 (GQ)	Rathlin2	Rathlin3
MCM6/LCTa	Lactase Persistence	rs4988235	G	A	G/G (21)	G/A (75)	AGGGGG	G
LCTb	Lactase Persistence	rs182549	C	T	C/C (45)	C/T (99)	CC	-
FADS1	Fatty Acid Metabolism	rs174546	T	C	C/T (99)	T/T (24)	-	-
NADSYN1	Circulating Vitamin D levels	rs7940244	C	T	C/C (48)	C/C (51)	TC	C
ALDH2	Alcohol Metabolism	rs671	G	A	G/G (24)	G/G (30)	-	G
ADH1Ba	Alcohol Metabolism	rs3811801	G	A	G/G (72)	G/G (45)	GGG	GGGG
ADH1Bb	Alcohol Metabolism	rs1229984	C	T	C/C (10)	C/C (30)	CC	-
NAT2	Acetylation Status	rs1041983	C	T	C/C (27)	C/T (99)	CCT	-
NAT2	Acetylation Status	rs1801280	T	C	T/C (95)	T/T (51)	TT	-
NAT2	Acetylation Status	rs1799929	C	T	C/T (57)	C/C (24)	CC	T
NAT2	Acetylation Status	rs1799930	G	A	G/G (36)	G/A (96)	-	G
NAT2	Acetylation Status	rs1208	A	G	A/G (93)	A/A (15)	GAA	-
NAT2	Acetylation Status	rs1799931	G	A	G/G (30)	G/G (45)	-	-
EDAR	Tooth Morphology and Hair Thickness	rs3827760	A	G	A/A (27)	A/A (45)	AAA	-
ABCC1	Earwax type	rs17822931	C	T	C/C (24)	C/C (24)	-	C
AGT	Hypertension	rs699	G	A	A/A (24)	G/G (27)	-	A
KITLG	Melanin pathway	rs4590952	G	A	A/A (18)	G/A (56)	G	-
CYP3A4	Protection from some forms of cancer	rs2740574	C	T	T/T (36)	T/T (42)	T	-
CYP3A5	Drug Metabolism	rs776746	T	C	C/C (42)	C/C (33)	-	-

GQ refers to the Genotype Quality outputted from GATK UnifiedGenotyper for the genotype calls of the two genomes with coverage over 10X (Ballynahatty and Rathlin1). Genotype calls are not shown for the two genomes under 2X coverage (Rathlin2 and Rathlin3). Instead all bases calls with a quality above 30 are reported.

References

1. Beck LA (1995) Standards for data collection from human skeletal remains. Edited by Jane E. Buikstra and Douglas H. Ubelaker. 272 pp. Fayetteville: Arkansas Archeological Survey Research Series No. 44, 1994. $25.00 (paper). *Am J Hum Biol* 7(5):672–672.

2. Schulting RJ, Murphy E, Carleton Jones, Warren G (2012) New dates from the north and a proposed chronology for Irish court tombs. *Proc R Ir Acad C Archaeol Celt Stud Hist Linguist Lit* 112C:1–60.

3. Whitehouse NJ, et al. (2014) Neolithic agriculture on the European western frontier: the boom and bust of early farming in Ireland. *J Archaeol Sci* 51(0):181–205.

4. Sloan B, Carver N, Murphy E, McGranaghan C (2008) A Bronze Age Cist Burial at Glebe, Rathlin Island, County Antrim. *Ulster Journal of Archaeology* 67:60–83.

5. Murphy E, McGranaghan C (2008) Osteological report on human remains from Glebe, Rathlin Island, County Antrim, pp. 72-83 in Sloan, B., A Bronze Age cist burial at Glebe, Rathlin Island, County Antrim. *Ulster Journal of Archaeology* 67:72–83.

6. Carver N (2008) The Food Vessel, pp. 66-68 in Sloan, B., A Bronze Age Cist Burial at Glebe, Rathlin Island, County Antrim. *Ulster Journal of Archaeology* 67:60–83.

7. Yang DY, Eng B, Waye JS, Dudar JC, Saunders SR (1998) Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am J Phys Anthropol* 105(4):539–543.

8. MacHugh DE, Edwards CJ, Bailey JF, Bancroft DR, Bradley DG (2000) The extraction and analysis of ancient DNA from bone and teeth: a survey of current methodologies. *Anc Biomol* 3(2):81–103.

9. Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010(6):db.prot5448.

10. Gamba C, et al. (2014) Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun* 5:5257.

11. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1):10–12.

12. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.

13. Schubert M, et al. (2012) Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* 13:178.

14. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

15. McKenna A, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.

16. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684.

17. Okonechnikov K, Conesa A, García-Alcalde F (2015) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. doi:10.1093/bioinformatics/btv566.

18. Keller A, et al. (2012) New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat Commun* 3:698.

19. Skoglund P, et al. (2014) Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* 344(6185):747–750.

20. Olalde I, et al. (2014) Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* 507(7491):225–228.

21. Lazaridis I, et al. (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513(7518):409–413.

22. Seguin-Orlando A, et al. (2014) Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* 346(6213):1113–1118.

23. Fu Q, et al. (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514(7523):445–449.

24. Raghavan M, et al. (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87–91.

25. Haak W, et al. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522(7555):207–211.

26. Allentoft ME, et al. (2015) Population genomics of Bronze Age Eurasia. *Nature* 522(7555):167–172.

27. Günther T, et al. (2015) Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1509851112.

28. Olalde I, et al. (2015) A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Mol Biol Evol*. doi:10.1093/molbev/msv181.

29. Brotherton P, et al. (2007) Novel high-resolution characterization of ancient DNA reveals C> U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35(17):5717–5728.

30. Briggs AW, et al. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* 104(37):14616–14621.

31. Shapiro B, Hofreiter M (2014) A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. *Science* 343(6169):1236573.

32. Vianello D, et al. (2013) HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. *Hum Mutat* 34(9):1189–1194.

33. Sánchez-Quinto F, et al. (2012) Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr Biol* 22(16):1494–1499.

34. Rasmussen M, et al. (2011) An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* 334 (6052):94–98.

35. Korneliussen T, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15(1):356.

36. Seidemann RM, Stojanowski CM, Doran GH (1998) The use of the supero-inferior femoral neck diameter as a sex assessor. *Am J Phys Anthropol* 107(3):305–313.

37. Haas J, Buikstra JE (1994) *Standards for Data Collection from Human Skeletal Remains: Proceedings of a Seminar at the Field Museum of Natural History, Organized by Jonathan Haas* (Arkansas Archeological Survey).

38. Skoglund P, Storå J, Götherström A, Jakobsson M (2013) Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J Archaeol Sci* 40(12):4477–4482.

39. Brandt G, et al. (2013) Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* 342(6155):257–261.

40. Lacan M, et al. (2011) Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc Natl Acad Sci U S A* 108(24):9788–9791.

41. Skoglund P, et al. (2012) Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336(6080):466–469.

42. Malmström H, et al. (2009) Ancient DNA Reveals Lack of Continuity between Neolithic Hunter-Gatherers and Contemporary Scandinavians. *Curr Biol* 19(20):1758–1762.

43. Bramanti B, et al. (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326(5949):137–140.

44. Fu Q, et al. (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* 23(7):553–559.

45. Bollongino R, et al. (2013) 2000 years of parallel societies in Stone Age Central Europe. *Science* 342(6157):479–481.

46. Hervella M, et al. (2012) Ancient DNA from hunter-gatherer and farmer groups from Northern Spain supports a random dispersion model for the Neolithic expansion into Europe. *PLoS One* 7(4):e34417.

47. Malyarchuk B, et al. (2010) The peopling of Europe from the mitochondrial haplogroup U5 perspective. *PLoS One* 5(4):e10285.

48. Richards M, et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67(5):1251–1276.

49. García O, et al. (2011) Using mitochondrial DNA to test the hypothesis of a European post-glacial human recolonization from the Franco-Cantabrian refuge. *Heredity* 106(1):37–45.

50. Tambets K, et al. (2004) The Western and Eastern Roots of the Saami—the Story of Genetic "Outliers" Told by Mitochondrial DNA and Y Chromosomes. *Am J Hum Genet* 74(4):661–682.

51. Pala M, et al. (2012) Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am J Hum Genet* 90(5):915–924.

52. Myres NM, et al. (2011) A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur J Hum Genet* 19(1):95–101.

53. Moore LT, McEvoy B, Cape E, Simms K, Bradley DG (2006) A Y-chromosome signature of hegemony in Gaelic Ireland. *Am J Hum Genet* 78(2):334–338.

54. Chang CC, et al. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.

55. Busby GBJ, et al. (2012) The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. *Proc Biol Sci* 279(1730):884–892.

56. Li JZ, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319(5866):1100–1104.

57. Behar DM, et al. (2010) The genome-wide structure of the Jewish people. *Nature* 466(7303):238–242.

58. Henn BM, et al. (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci U S A* 108(13):5154–5162.

59. Hellenthal G, et al. (2014) A genetic atlas of human admixture history. *Science* 343(6172):747–751.

60. 1000 Genomes Project Consortium, et al. (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74.

61. Prüfer K, et al. (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43–49.

62. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2(12):e190.

63. Price AL, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8):904–909.

64. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.

65. Green RE, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328(5979):710–722.

66. Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* 461(7263):489–494.

67. Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192(3):1065–1093.

68. Gurdasani D, et al. (2015) The African Genome Variation Project shapes medical genetics in Africa. *Nature* 517(7534):327–332.

69. Haak W, et al. (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310(5750):1016–1018.

70. Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192(3):1065–1093.

71. Pemberton TJ, et al. (2012) Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 91(2):275–292.

72. Kirin M, et al. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5(11):e13996.

73. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genet* 8(1):e1002453.

74. Delaneau O, Marchini J, Zagury J-F (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* 9(2):179–181.

75. Breheny P, Burchett W (2012) visreg: Visualization of regression models. *R package version*:2–0.

76. Team RC (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012.

77. Walsh S, et al. (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet* 7(1):98–115.

78. Eiberg H, et al. (2008) Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum Genet* 123(2):177–187.

79. Donnelly MP, et al. (2012) A global view of the OCA2-HERC2 region and pigmentation. *Hum Genet* 131(5):683–696.

80. Canfield VA, et al. (2013) Molecular phylogeography of a human autosomal skin color locus under natural selection. *G3* 3(11):2059–2067.

81. Paré G, et al. (2008) Novel association of ABO histo-blood group antigen with soluble ICAM-1: results of a genome-wide association study of 6,578 women. *PLoS Genet* 4(7):e1000118.

82. Wolpin BM, et al. (2009) ABO blood group and the risk of pancreatic cancer. *J Natl Cancer Inst* 101(6):424–431.

83. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26.

84. Blau N, van Spronsen FJ, Levy HL (2010) Phenylketonuria. *Lancet* 376(9750):1417–1427.

85. Zschocke J, Graham CA, Carson DJ, Nevin NC (1995) Phenylketonuria mutation analysis in Northern Ireland: a rapid stepwise approach. *Am J Hum Genet* 57(6):1311–1317.

86. Zschocke J (2003) Phenylketonuria mutations in Europe. *Hum Mutat* 21(4):345–356.

87. Stoller JK, Lacbawan FL, Aboussouan LS (2006) Alpha-1 Antitrypsin Deficiency. *GeneReviews(®)*, eds Pagon RA, et al. (University of Washington, Seattle, Seattle (WA)).

88. Blanco I, Fernandez E, Bustillo EF (2001) Alpha-1-antitrypsin PI phenotypes S and Z in Europe: an analysis of the published surveys. *Clin Genet* 60(1):31–41.

89. Farrell P, et al. (2007) Diagnosis of cystic fibrosis in the Republic of Ireland: epidemiology and costs. *Ir Med J* 100(8):557–560.

90. Castellani C, et al. (2008) Consensus on the use and interpretation of cystic fibrosis mutation analysis in clinical practice. *J Cyst Fibros* 7(3):179–196.

91. Barton JC, Acton RT (2002) HLA-A and -B alleles and haplotypes in hemochromatosis probands with HFE C282Y homozygosity in central Alabama. *BMC Med Genet* 3:9.

92. Distante S, et al. (2004) The origin and spread of the HFE-C282Y haemochromatosis mutation. *Hum Genet* 115(4):269–279.

93. Pacho A, et al. (2004) HLA haplotypes associated with hemochromatosis mutations in the Spanish population. *BMC Med Genet* 5:25.

94. Barton JC, Wiener HW, Acton RT, Go RC-P (2005) HLA haplotype A* 03-B* 07 in hemochromatosis probands with HFE C282Y homozygosity: frequency disparity in men and women and lack of association with severity of iron overload. *Blood Cells Mol Dis* 34(1):38–47.

95. Rodriguez LM, et al. (2015) Ancestral association between HLA and HFE H63D and C282Y gene mutations from northwest Colombia. *Genet Mol Biol* 38(1):8–13.

96. Dilthey AT, Moutsianas L, Leslie S, McVean G (2011) HLA* IMP—an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics* 27(7):968–972.

97. Traherne JA, et al. (2006) Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLoS Genet* 2(1):e9.

98. Larsen CE, Alper CA (2004) The genetics of HLA-associated disease. *Curr Opin Immunol* 16(5):660–667.

99. Barcellos LF, et al. (2003) HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course. *Am J Hum Genet* 72(3):710–716.

100. Patin E, et al. (2006) Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes. *Am J Hum Genet* 78(3):423–436.

101. Mathieson I, et al. (2015) *Eight thousand years of natural selection in Europe*.

102. Bokor S, et al. (2010) Single nucleotide polymorphisms in the FADS gene cluster are associated with delta-5 and delta-6 desaturase activities estimated by serum fatty acid ratios. *J Lipid Res* 51(8):2325–2333.

103. Ahn J, et al. (2010) Genome-wide association study of circulating vitamin D levels. *Hum Mol Genet* 19(13):2739–2745.

104. Price AL, et al. (2009) The impact of divergence time on the nature of population structure: an example from Iceland. *PLoS Genet* 5(6):e1000505.

105. Yoshiura K-I, et al. (2006) A SNP in the ABCC11 gene is the determinant of human earwax type. *Nat Genet* 38(3):324–330.

106. Fujimoto A, et al. (2008) A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum Mol Genet* 17(6):835–843.