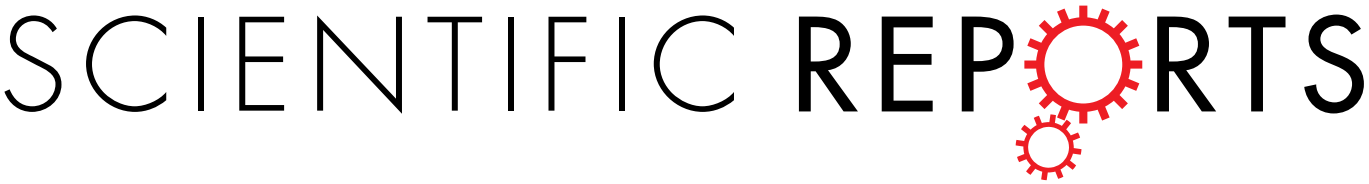

OPEN

The genetics of an early Neolithic pastoralist from the Zagros, Iran

M. Gallego-Llorente[1], S. Connell[2], E. R. Jones[1], D. C. Merrett[3], Y. Jeon[4,5], A. Eriksson[1,6], V. Siska[1], C. Gamba[2,7], C. Meiklejohn[8], R. Beyer[9], S. Jeon[4,5], Y. S. Cho[4,5], M. Hofreiter[10], J. Bhak[4], A. Manica[1,*] & R. Pinhasi[2,*]

Received: 28 May 2016
Accepted: 15 July 2016
Published: 09 August 2016

The agricultural transition profoundly changed human societies. We sequenced and analysed the first genome (1.39x) of an early Neolithic woman from Ganj Dareh, in the Zagros Mountains of Iran, a site with early evidence for an economy based on goat herding, ca. 10,000 BP. We show that Western Iran was inhabited by a population genetically most similar to hunter-gatherers from the Caucasus, but distinct from the Neolithic Anatolian people who later brought food production into Europe. The inhabitants of Ganj Dareh made little direct genetic contribution to modern European populations, suggesting those of the Central Zagros were somewhat isolated from other populations of the Fertile Crescent. Runs of homozygosity are of a similar length to those from Neolithic farmers, and shorter than those of Caucasus and Western Hunter-Gatherers, suggesting that the inhabitants of Ganj Dareh did not undergo the large population bottleneck suffered by their northern neighbours. While some degree of cultural diffusion between Anatolia, Western Iran and other neighbouring regions is possible, the genetic dissimilarity between early Anatolian farmers and the inhabitants of Ganj Dareh supports a model in which Neolithic societies in these areas were distinct.

The agricultural transition started in a region comprising the Ancient Near East and Anatolia ~12,000 years ago with the first Pre-Pottery Neolithic villages and the first domestication of cereals and legumes[1,2]. Archaeological evidence suggests a complex scenario of multiple domestications in a number of areas[3], coupled with examples of trade[4]. Ancient DNA (aDNA) has revealed that this cultural package was later brought into Europe by dispersing farmers from Anatolia (so called 'demic' diffusion, as opposed to non-demic cultural diffusion[5,6]) ~8,400 years ago. However a lack of aDNA from early Neolithic individuals from the Near East leaves a key question unanswered: was the agricultural transition developed by one major population group spanning the Near East, including Anatolia and the Central Zagros Mountains; or was the region inhabited by genetically diverse populations, as is suggested by the heterogeneous mode and timing of the appearance of early domesticates at different localities?

To answer this question, we sequenced the genome of an early Neolithic female from Ganj Dareh, GD13a, from the Central Zagros (Western Iran), dated to 10000-9700 cal BP[7], a region located at the eastern edge of the Near East. Ganj Dareh is well known for providing the earliest evidence of herd management of goats beginning at 9,900 BP[7–9]. It is a classic mound site at an altitude of ~1400 m in the Gamas-Ab Valley of the High Zagros zone in Kermanshah Province, Western Iran. It was discovered in the 1960s during survey work and excavated over four seasons between 1967 and 1974. The mound, ~40 m in diameter, shows 7 to 8 m of early Neolithic cultural deposits. Five major levels were found, labelled A through E from top to bottom. Extended evidence showed a

[1]Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, UK. [2]School of Archaeology and Earth Institute, University College Dublin, Belfield, Dublin 4, Ireland. [3]Department of Archaeology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. [4]The Genomics Institute, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea. [5]Department of Biomedical Engineering, School of Life Sciences, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea. [6]Integrative Systems Biology Laboratory, Division of Biological and Environmental Sciences & Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia. [7]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, Copenhagen 1350, Denmark. [8]Department of Anthropology, University of Winnipeg, Winnipeg, MB R3B 2E9, Canada. [9]McDonald Institute for Archaeological Research, University of Cambridge, Cambridge CB2 3ER, UK. [10]Evolutionary Adaptive Genomics, Institute for Biochemistry and Biology, Department of Mathematics and Natural Sciences, University of Potsdam, Karl-Liebknechtstraße 24-25, Potsdam, 14476, Germany. *These authors jointly supervised this work. Correspondence and requests for materials should be addressed to M.G.-L. (email: mg632@cam.ac.uk) or A.M. (email: am315@cam.ac.uk) or R.P. (email: ron.pinhasi@ucd.ie)

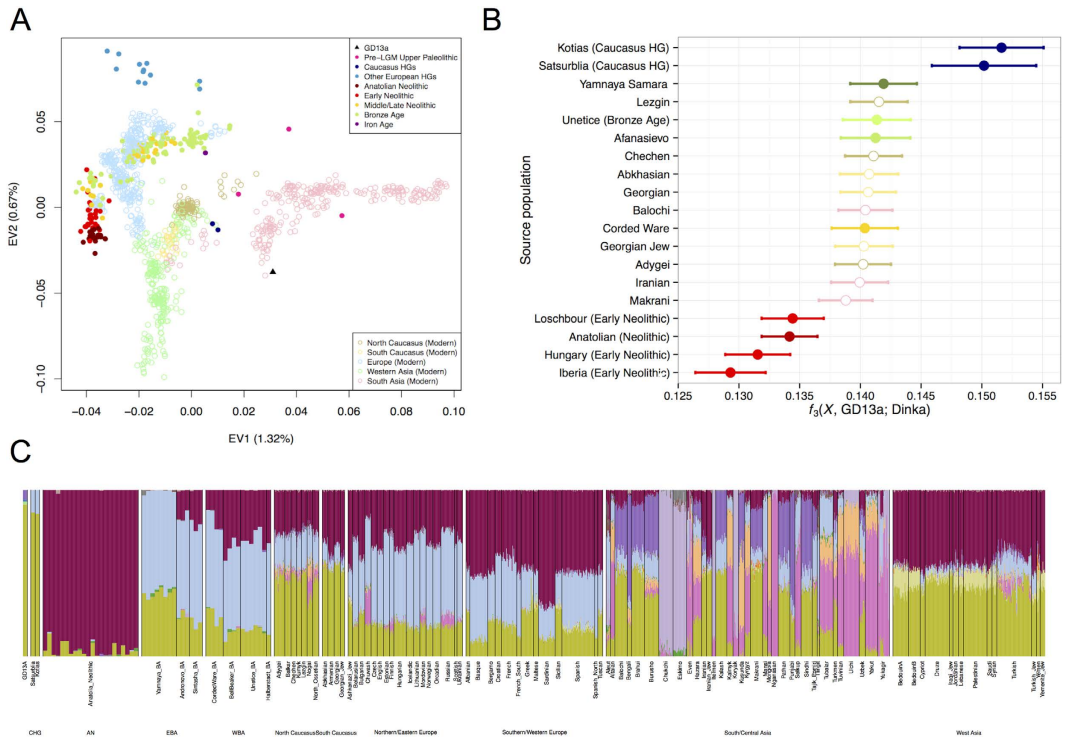

Figure 1. GD13a appears to be related to Caucasus Hunter Gatherers and to modern South Asian populations. (**A**) **PCA** loaded on modern populations (represented by open symbols). Ancient individuals (solid symbols) are projected onto these axes. **(B**) **Outgroup f_3(X, GD13a; Dinka)**, where Caucasus Hunter Gatherers (Kotias and Satsurblia) share the most drift with GD13a. Ancient samples have filled circles whereas modern populations are represented by empty symbols. **(C) ADMIXTURE** using K = 17, where GD13a appears very similar to Caucasus Hunter Gatherers, and to a lesser extent to modern south Asian populations.

warren of rooms with evidence of under-floor inhumations within what may be burial chambers and/or disused houses[10]. The current Minimum Number of Individuals is 116, with 56 catalogued as skeletons that had four or more bones recovered[11]. The individual analysed here was part of burial 13, which contained three individuals, and was recovered in level C in 1971 from the floor of a brick-walled structure. The individual sampled, 13A (referred to as GD13a throughout the text), was a 30–50 year old female; the other individuals in the burial unit were a second adult (13B) and an adolescent (13).

The site has been directly dated to 9650–9950 cal BP[7], and shows intense occupation over two to three centuries. The economy of the population was that of pastoralists with an emphasis on goat herding[7]. Archaeobotanical evidence is limited[12] but the evidence present is for two-row barley with no evidence for wheat, rye or other domesticates. This implies that the overall economy was at a much earlier stage in the development of cereal agriculture than that found in the Levant, Anatolia and Northern Mesopotamian basin.

Results

The petrous bone of GD13a yielded sequencing libraries comprising 18.57% alignable human reads that were used to generate 1.39-fold genome coverage. The sequence data showed read lengths and nucleotide misincorporation patterns indicative of post-mortem damage, and had low contamination estimates (<1%, see Supplementary Fig. S3). The mitochondrion of GD13a (91.74X) was assigned to haplogroup X, most likely to the subhaplogroup X2, which has been associated with an early expansion from the Near East[13,14] and has been found in early Neolithic samples from Anatolia[5], Hungary[15] and Germany[16].

We compared GD13a with a number of other ancient genomes and modern populations[6,15–27], using principal component analysis (PCA)[28], ADMIXTURE[29] and outgroup f_3 statistics[30] (Fig. 1). GD13a did not cluster with any other early Neolithic individual from Eurasia in any of the analyses. ADMIXTURE and outgroup f_3 statistics identified Caucasus Hunter-Gatherers of Western Georgia, just north of the Zagros mountains, as the group genetically most similar to GD13a (Fig. 1B,C), whilst PCA also revealed some affinity with modern Central South Asian populations such as Balochi, Makrani and Brahui (Fig. 1A and Fig. S4). Also genetically close to GD13a were ancient samples from Steppe populations (Yamanya & Afanasievo) that were part of one or more Bronze age migrations into Europe, as well as early Bronze age cultures in that continent (Corded Ware)[16,21], in line with previous relationships observed for the Caucasus Hunter-Gatherers[24].

We further investigated the relationship between GD13a and Caucasus Hunter-Gatherers using *D*-statistics[30,31] to test whether they formed a clade to the exclusion of other ancient and modern samples (Table S4). A large number of Western Eurasian samples (both modern and ancient) showed significant excess genetic affinity to

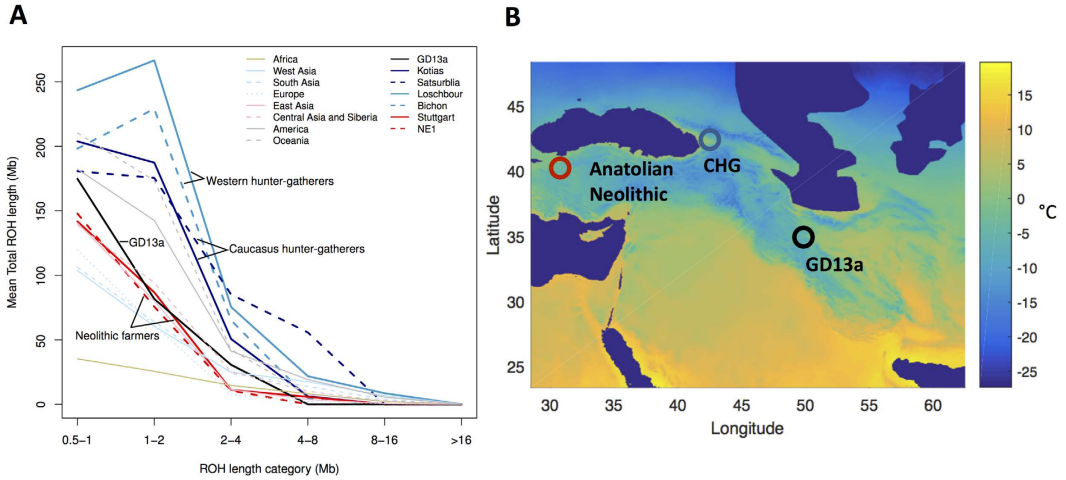

Figure 2. GD13a did not undergo a recent large population bottleneck. (**A**) GD13a has similar runs of homozygosity (ROH) lengths to Neolithic individuals, while Caucasus Hunter Gatherers (Kotias and Satsurblia), like European Hunter Gatherers (Loschbour and Bichon), underwent recent large population bottlenecks potentially associated withthe LGM. (**B**) Map showing geographical location of Anatolian Neolithic samples, Caucasus Hunter Gatherers (CHG) and GD13a. Background colours indicate mean temperature (°C) of coldest quarter during the LGM (data from the worldclim database60 generated by the CCSM4 model)[60], with LGM sea levels. Map of populations was generated with MATLAB R2015b (Mathworks, http://www.mathworks.com/)[61].

the Caucasus Hunter-Gatherers, whilst none did with GD13a. Overall, these results point to GD13a having little direct genetic input into later European populations compared to its northern neighbours.

To better understand the history of the population to which GD13a belonged, we investigated the distribution of lengths of runs of homozygosity (ROH) (Fig. 2A). A bias towards a high frequency of both long and short ROH is indicative of past strong bottlenecks followed by population re-expansion. GD13a has a distribution with few long ROH (>2 Mb), similar to that of the descendants of Anatolian early farmers (represented by the European farmers NE1[15] and Stuttgart[17]). In contrast, both Western[17] and Caucasus Hunter-Gatherers[24] have relatively more long as well as short ROH. Thus, GD13a is the descendant of a group that had relatively stable demography and did not suffer the bottlenecks that affected more northern populations.

The phenotypic attributes of GD13a are similar to the neighbouring Anatolian early farmers and Caucasus Hunter-Gatherers. Based on diagnostic SNPs, she had dark, black hair and brown eyes (see Supplementary). She lacked the derived variant (rs16891982) of the *SLC45A2* gene associated with light skin pigmentation but likely had at least one copy of the derived *SLC24A5* allele (rs1426654) associated with the same trait. The derived *SLC24A5* variant has been found in both Neolithic farmer and Caucasus Hunter-Gatherer groups[5,15,24] suggesting that it was already at appreciable frequency before these populations diverged. Finally, she did not have the most common European variant of the *LCT* gene (rs4988235) associated with the ability to digest raw milk, consistent with the later emergence of this adaptation[5,15,21].

It is possible that farmers related to GD13a contributed to the eastern diffusion of agriculture from the Near East that reached Turkmenistan[32] by the 6th millennium BP, and continued further east to the Indus Valley[33]. However, detecting such a contribution is complicated by a later influx from Steppe populations with Caucasus Hunter-Gatherer ancestry during the Bronze Age. We tested whether the Western Eurasian component found in Indian populations can be better attributed to either of these two sources, GD13a and Kotias (a Caucasus Hunter Gatherer), using *D*-statistics to detect gene flow into an ancestral Indian component (represented by the Onge). Overall, for all tests where a difference could be detected, Kotias and GD13a were equally likely sources (Fig. S9 and Table S6). Whilst the attribution of part of the Western Eurasia component seen in India to Bronze Age migrations is supported by dating of last contact based on patterns of Linkage Disequilibrium[34], our analysis highlights the possibility that part of that component might derive from earlier contact during the eastern diffusion of agriculture.

Discussion

GD13a had little direct genetic input into later European populations compared to the Caucasus Hunter-Gatherers (its northern neighbours) as demonstrated using *D*-statistics. This lack of connectivity with neighbouring regions might have arisen early on, since we also find that hunter-gatherers from the Caucasus show higher affinity to Western Hunter-Gatherers and early Anatolian farmers; this result suggests the possibility of gene flow between the former and these two latter groups to the exclusion of GD13a. An alternative, but not mutually exclusive, explanation for this pattern is that GD13a might have received genetic input from a source equally distant from all other European populations, and thus basal to them.

The Last Glacial Maximum (LGM) made entire regions in northern Eurasia uninhabitable, and therefore a number of hunter-gatherer populations likely moved to the south. For Europe there may be a separation between

Western and Eastern populations with minimal occupation of the Central European plains[22]. For Eastern Europe, Central Asia and the northern Near East, glaciation itself was less a factor. In these areas, our understanding of how populations weathered the LGM is still vague at best. It has previously been suggested that differences in the frequency of long and short runs of homozygosity in ancient samples may be associated with different demographic experiences during the LGM[15,24]. Neolithic farmers, with their lower frequency of short ROH, have been argued to have been relatively little affected by the LGM compared to Western and Caucasus Hunter-Gatherers[15,24] which are characterised by more long ROH (>2 Mb). GD13a has a profile similar to that of the descendants of Anatolian farmers (i.e. early European farmers), suggesting that her ancestors also faced more benign conditions compared to populations further north. Superimposing the sampling locations of these groups onto climatic reconstructions from the LGM (Fig. 2B), however, does not reveal clear climatic differences among the regions. It is possible that the ancestors of the Anatolian and Ganj Dareh farmers spent the LGM in areas further south or east, which experienced milder climate. But it is also possible that they exploited local pockets of favourable climate (refugia). Whilst high elevation sites in the Zagros were abandoned during the LGM[35], there are a number of sites in the valleys that were occupied during that period and might have experienced more favourable conditions[36].

The archaeological record indicates an eastward Neolithic expansion from the eastern regions of the Near East into Central and South Asia[32,37]. Our analysis shows that both the Caucasus Hunter Gatherer Kotias and GD13a are plausible sources for the Eurasian Ancestry found in that part of Asia. Even though part of the Western Eurasian component found in India can be linked to Bronze Age migrations by dating the last contact using Linkage Disequilibrium (thus coming from the Kotias lineage), our results highlight the possibility of an older contribution from a source genetically close to GD13a (which would be hard to disentangle from the later gene flow from the Steppe). Eventually, ancient DNA from the Indus Valley will be needed to detect conclusively whether any genetic traces were left by the eastward Neolithic expansion from the Near East, or whether this process was mostly cultural.

The presence of two distinct lineages (Anatolian-like agriculturalists and Zagros mountain herders) in the Near East at the beginning of the Neolithic transition raises an interesting question regarding the independence of innovations arising at different locations. Even within the Central Zagros, economies vary greatly in their rate and pathway towards Neolithisation[35]. Ganj Dareh, in the high Zagros, has the earliest known evidence for goat domestication[7–9], and the foothills of the Zagros mountains have also been argued to have been the site of early farming[3]. In addition, early sites such as Sheikh-e Abad (11.650-9,600 cal BP) provide evidence of early stages of barley cultivation[38]. Were these innovations independent of similar achievements that made up the Neolithic package that North West Anatolians brought into Europe? Or were they exchanged culturally? If the latter, it would imply a cultural diffusion in the absence of much genetic interchange.

Methods

DNA extraction and library preparation. Sample preparation, DNA extraction and library preparation were carried out in dedicated ancient DNA facilities at University College Dublin. The dense part of the petrous bone was isolated, cleaned and sequenced following experimental procedures outlined in Gamba *et al.*[15]. DNA was extracted from 310 mg of ground bone powder using a double-digestion and silica column method as described in Gamba *et al.*[39]. Indexed Illumina sequencing libraries were constructed with a protocol based on Meyer *et al.*[40] with modifications including blunt end repair using NEBNext End Repair Module (New England BioLabs Inc) and heat inactivation of Bst DNA polymerase[15].

Sequence processing and alignment. Libraries were sequenced over a flow cell on a HiSeq 2000 at the TheragenEtex (South Korea) using 100 bp single-end sequencing. Adapter sequences were trimmed from the 3′ ends of sequences using cutadapt version 1.3[41], conservatively requiring an overlap of 1 base pair (bp) between the adapter and the read. Reads were aligned using BWA[42], with the seed region disabled, to the GRCh37 build of the human genome with the mitochondrial sequence replaced by the revised Cambridge reference sequence (NCBI accession number NC_012920.1). Data from separate lanes were merged using Picard MergeSamFiles (http://picard.sourceforge.net/) and duplicate reads from the same library amplification were filtered using SAMtools rmdup[43]. Sequences were further filtered to remove those with mapping quality <30 and length <30 bp. Indels were realigned using RealignerTargetCreator and IndelRealigner from the Genome Analysis Toolkit[44]. The first and last 2 bp of each read were soft-clipped to a base quality of 2. The average genome-wide depth of coverage was calculated using the *genomecov* function of bedtools[45]. A summary of alignment statistics can be found in Table S1.

Authenticity of results. The data were assessed for the presence of typical signatures of post-mortem DNA damage[46,47]. The sequence length distribution of molecules was examined as outlined in Gallego Llorente *et al.*[48] (Fig. S2) while the prevalence of nucleotide misincorporation sites at the ends of reads was evaluated using mapDamage 2.0 and a random subsample of 1 million reads[49] (Fig. S3).

The mitochondrial contamination rate was assessed by evaluating the proportion of non-consensus bases at haplogroup defining positions in the mitochondrial genome[15,50]. Only bases with quality ≥20 were used. The X chromosome contamination rate could not be evaluated as the sample was determined to be female, using the script described in ref. 51.

Mitochondrial Haplogroup Determination. To determine to which haplogroup the mitochondrion of GD13a belonged, a consensus sequence was generated using ANGSD[52]. Called positions were required to have a depth of coverage ≥3 and only bases with quality ≥20 were considered. The resulting FASTA files were uploaded to HAPLOFIND[53] for haplogroup determination. Coverage was calculated using GATK DepthOfCoverage[44].

Dataset preparation for population genetic analyses. Genotypes were called in GD13a at sites which overlapped those in the Human Origins dataset (Lazaridis *et al.*[17], filtered as described in Jones *et al.*[24]) using GATK Pileup[44]. Triallelic SNPs were discarded and bases were required to have quality ≥30. For positions with more than one base call, one allele was randomly chosen with a probability equal to the frequency of the base at that position. This allele was duplicated to form a homozygous diploid genotype for each position called in GD13a. This method of SNP calling was also used for selected ancient samples described in Jones *et al.*[24] Cassidy *et al.*[25], Gunther *et al.*[26], Omrak *et al.*[6] and Olalde *et al.*[27]. Genotype calls for these ancient samples were merged with calls from modern samples found in the Human Origins dataset and ancient samples provided in the Mathieson *et al.*[5] dataset which also included genotype calls for previously published ancient samples[15–17,19–23,27]. To avoid biases caused by post-mortem DNA damage, only transversion sites were used for PCA, ADMIXTURE, f_3-statistics and *D*-statistics.

Principal component analysis. To explore GD13a and other ancient samples in the context of modern variation in Eurasia, we performed PCA with a panel of contemporary populations (196 contemporary populations, 145,004 transversion SNPs). The analysis was carried out using SmartPCA[28]; the components were loaded on the contemporary populations, and the ancient individuals were projected onto these dimensions (Fig. 1 and Fig. S4).

ADMIXTURE. A clustering analysis was performed using ADMIXTURE version 1.23[29], using the full panel of modern and ancient samples described above. SNPs in linkage disequilibrium were thinned using PLINK (v1.07)[54] with parameters –indep-pairwise 200 25 0.5[16], resulting in a set of 116,834 SNPs for analysis. Clusters (K) (2–20) were explored using 3 runs with fivefold cross-validation at each K with different random seeds. The minimal cross-validation error was found at K = 17, but the error already starts plateauing from roughly K = 10, implying little improvement from this point onwards (Fig. S5). The ADMIXTURE proportions are shown in Fig. S6 for all samples and in Fig. 1C for GD13a and selected modern and ancient populations harbouring the component dominant in GD13a.

Outgroup f_3-statistics and *D*-statistics. Outgroup f_3-statistics and *D*-statistics were performed using the qp3Pop and qpDstat programs from the ADMIXTOOLS package[30].

Neighbour-joining tree. We used a custom Matlab script to calculate pairwise pi from genome-wide genotype data using a panel of 22 individuals (from the dataset described above), including GD13a, representative ancient samples, and different modern populations from the same geographic area as GD13a, and generated an unweighted pair group method with arithmetic mean (UPGMA) tree using the seqlinkage function in Matlab's Bioinformatics Toolbox[55].

Runs of homozygosity. In order to examine runs of homozygosity (ROH) we used imputation to infer diploid genotypes in our sample following the method described in Gamba *et al.*[15]. We used GATK Unified genotyper[44] to call genotype likelihoods at SNP sites in Phase 3 of 1,000 genomes project[56] (version 5a downloaded from the BEAGLE website, https://faculty.washington.edu/browning/beagle/beagle.html). Genotype likelihoods were called for alleles observed in the 1,000 Genomes Project and equal likelihoods were set for positions with no spanning sequence data as well as positions where the observed genotype could be explained by deamination. Genotypes were imputed using Beagle 4.0 with default parameters in intervals of 1 Mb[57]. We imposed a genotype probability threshold of 0.99 (any SNP without a genotype exceeding this threshold had a missing genotype assigned) while converting to PLINK-format genotype data. These data were merged with the dataset used in Jones *et al.*[24] and ROH analysis was carried out as outlined in Gamba *et al.*[15] and Jones *et al.*[24].

Phenotypes of interest. Genes associated with a particular phenotype in modern populations were explored in GD13a. Observed genotypes were called using GATK Unified genotyper[44], calling alleles present in Phase 1 of 1,000 genomes dataset[58] with base quality ≥20. As many diagnostic markers had 1-fold coverage or less, we also used imputation to infer genotypes at these positions. Imputation was performed as described in section S11, imputing at least 1 Mb upstream and downstream of the SNP position (this interval was reduced for those variants within the first 1 Mb of the chromosome). The Hirisplex prediction model[59] was used to explore hair and eye colour (Table S5). For the observed data, if the sample had 1x coverage, the variant was called as homozygous for that allele. Hair and eye colour predictions were confirmed using imputed data.

References

1. Blockley, S. P. E. & Pinhasi, R. A revised chronology for the adoption of agriculture in the Southern Levant and the role of Lateglacial climatic change. *Quaternary Science Reviews* **30,** 98–108 (2011).
2. Goring-Morris, A. N. & Belfer-Cohen, A. Neolithization Processes in the Levant: The Outer Envelope. *Current Anthropology* **52,** S195–S208 (2011).
3. Riehl, S., Zeidi, M. & Conard, N. J. Emergence of Agriculture in the Foothills of the Zagros Mountains of Iran. *Science* **341,** 65–67 (2013).
4. Aurenche, O. & Kozlowski, S. K. *La naissance du néolithique au proche Orient ou Le paradis perdu* (Errance, 1999).
5. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528,** 499–503 (2015).
6. Omrak, A. *et al.* Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Current Biology* **26,** 270–275 (2016).
7. Zeder, M. A. & Hesse, B. The Initial Domestication of Goats (Capra hircus) in the Zagros Mountains 10,000 Years Ago. *Science* **287,** 2254–2257 (2000).
8. Zeder, M. A. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *PNAS*, doi: 10.1073/pnas.0801317105 (2008).
9. Zeder, M. A. The Origins of Agriculture in the Near East. *Curr Anthrop* **52,** 221–235 (2011).
10. Smith, P. E. L. Architectural Innovation and Experimentation at Ganj Dareh, Iran. *World Archaeology* **21,** 323–335 (1990).

11. Merrett, D. C. Bioarchaeology in Early Neolithic Iran: assessment of health status and subsistence strategy. Unpublished Ph.D. Dissertation. (University of Manitoba, 2004).
12. van Zeist, W., Smith, P. E. L., Palfenier-Vegter, R. M., Suwijn, M. & Casparie, W. A. An archaeobotanical study of Ganj Dareh Tepe, Iran. *Palaeohistoria* **26,** 201–224 (1984).
13. Reidla, M. *et al.* Origin and Diffusion of mtDNA Haplogroup X. *Am J Hum Genet* **73,** 1178–1190 (2003).
14. Richards, M. *et al.* Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67,** 1251–1276 (2000).
15. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5** (2014).
16. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522,** 207–211 (2015).
17. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).
18. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507,** 225–228 (2014).
19. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
20. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat Commun* **3,** 698 (2012).
21. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522,** 167–172 (2015).
22. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346,** 1113–1118 (2014).
23. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524,** 216–219 (2015).
24. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun* **6,** 8912 (2015).
25. Cassidy, L. M. *et al.* Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *PNAS* **113,** 368–373 (2016).
26. Günther, T. *et al.* Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *PNAS* **112,** 11917–11922 (2015).
27. Olalde, I. *et al.* A common genetic origin for early farmers from Mediterranean Cardial and Central European LBK cultures. *Mol Biol Evol* msv181, doi: 10.1093/molbev/msv181 (2015).
28. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
29. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).
30. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192,** 1065–1093 (2012).
31. Green, R. E. *et al.* A Draft Sequence of the Neandertal Genome. *Science* **328,** 710–722 (2010).
32. Harris, D. R. *et al. Origins of Agriculture in Western Central Asia: An Environmental-Archaeological Study*. (University of Pennsylvania Press, 2010).
33. Gangal, K., Sarson, G. R. & Shukurov, A. The Near-Eastern Roots of the Neolithic in South Asia. *PLoS ONE* **9,** e95714 (2014).
34. Kirin, M. *et al.* Genomic Runs of Homozygosity Record Population History and Consanguinity. *PLOS ONE* **5,** e13996 (2010).
35. Matthews, R. & Nashli, H. F. *The Neolithisation of Iran*. 3, (Oxbow Books, 2013).
36. Tsuneki, A. In: R. Matthews & H. Fazeli Nashili (Eds) *The Neolithisation of Iran: The Formation of New Societies* 94–96 (Oxbow Books, 2013).
37. Weeks, L. *et al.* The Neolithic Settlement of Highland SW Iran: New Evidence from the Mamasani District. *Iran* **44,** 1–31 (2006).
38. Whitlam, J., Ilkhani, H., Bogaard, A. & Charles, C. In *The Earliest Neolithic of Iran: 2008 Excavations at Sheikh-e Abad and Jani* (eds Matthews, R., Matthews, W. & Mohammadifar, Y.) Ch. 15, 175–184 (2013).
39. Gamba, C. *et al.* Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol Ecol Resour* **16,** 459–469 (2016).
40. Meyer, M. & Kircher, M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc.* **2010,** pdb.prot5448 (2010).
41. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17,** pp. 10–12 (2011).
42. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
43. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
44. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).
45. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
46. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA* **104,** 14616–14621 (2007).
47. Brotherton, P. *et al.* 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,** 5717–5728 (2007).
48. Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350,** 820–822 (2015).
49. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29,** 1682–1684 (2013).
50. Sánchez-Quinto, F. *et al.* Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers. *Curr. Biol.* **22,** 1494–1499 (2012).
51. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40,** 4477–4482 (2013).
52. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinform.* **15,** 356 (2014).
53. Vianello, D. *et al.* HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. *Hum. Mutat.* **34,** 1189–1194 (2013).
54. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
55. The MathWorks, Inc. MATLAB and Bioinformatics Release 2016b (2016).
56. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
57. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81,** 1084–1097 (2007).
58. Consortium, T. 1000 G. P. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
59. Walsh, S. *et al.* The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.* **7,** 98–115 (2013).
60. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25,** 1965–1978 (2005).
61. The MathWorks Inc. MATLAB version 8.6.0. Natick, Massachusetts (2015).

Acknowledgements

A.M. was supported by ERC Consolidator Grant 647787 'LocalAdaptation'; R.P. by ERC Starting Grant: ERC-2010-StG 26344 ("ADNABIOARC"); M.H. by ERC Consolidator Grant 310763 'GeneFlow'; C.G. was supported

by the Irish Research Council for Humanities and Social Sciences (IRCHSS) ERC Support Programme and the Marie-Curie Intra-European Fellowships (FP7-IEF-328024); S.C. was supported by the Irish Research Council for Humanities and Social Sciences (IRCHSS) ERC Support Programme; J.B. was supported by the 2014 Research fund (1.140077.01) of Ulsan National Institute of Science & Technology (UNIST) and Geromics internal research funding; J.B. and Y.S.C. were supported by the Research Fund (14-BR-SS-03) of Civil-Military Technology Cooperation Program; R.B. was supported by ERC Consolidator Grant 617627 "ADaPt"; and M.G. by a BBSRC DTP studentship.

Author Contributions

D.C.M., C.M. and R.P. obtained the sample and provided expertise in the archaeology. C.G. and S.C. extracted and prepared the genetic material, Y.J., S.J., Y.S.C. and J.B. sequenced the sample, M.G.-L., E.R.J., V.S., A.E., R.B. and A.M. did the genetic analysis. M.H. helped with interpretation.

Additional Information

Accession codes: Raw reads from Ganj Dareh 13a are available for download through the EBI European Nucleotide Archive (ENA) accession number PRJEB13189.

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Gallego-Llorente, M. *et al.* The genetics of an early Neolithic pastoralist from the Zagros, Iran. *Sci. Rep.* **6**, 31326; doi: 10.1038/srep31326 (2016).