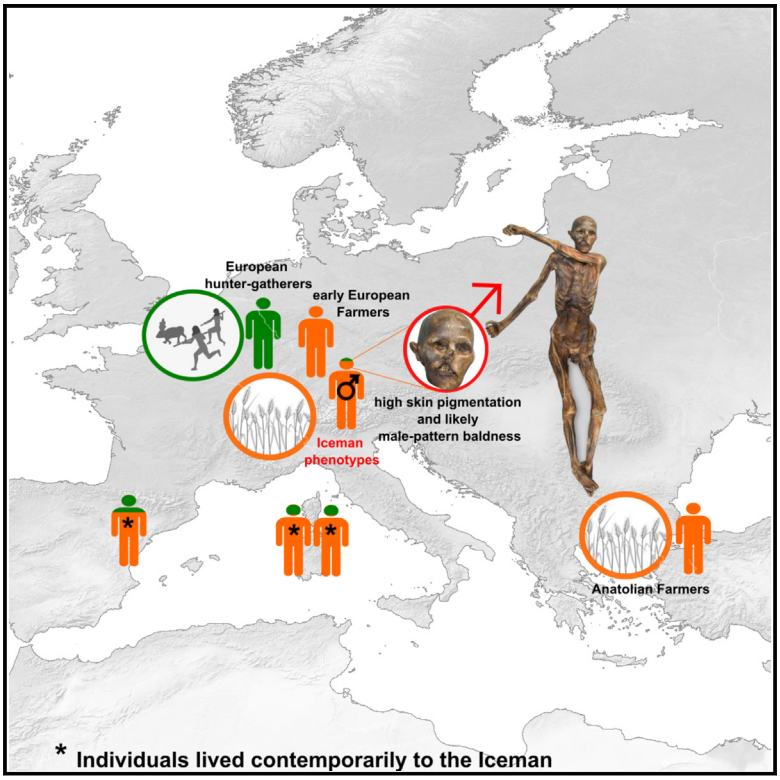
Cell Genomics

Short Article

High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry

Graphical abstract



Authors

Ke Wang, Kay Prüfer, Ben Krause-Kyora, ..., Albert Zink, Stephan Schiffels, Johannes Krause

Correspondence

albert.zink@eurac.edu (A.Z.), krause@eva.mpg.de (J.K.)

In brief

Wang et al. reported a newly generated high-coverage genome of the Tyrolean Iceman and revealed his unusually high Anatolian-farmer-related ancestry as well as his potential male-pattern baldness and high levels of skin pigmentation.

Highlights

- High-coverage genome of the Iceman
- Unusually high Anatolian-farmer-related ancestry
- Dark skin and likely bald

Wang et al., 2023, Cell Genomics *3*, 100377
September 13, 2023
https://doi.org/10.1016/j.xgen.2023.100377

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Short Article

High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry

Ke Wang,[1,2,3] Kay Prüfer,[2] Ben Krause-Kyora,[4] Ainash Childebayeva,[2] Verena J. Schuenemann,[5,6,7] Valentina Coia,[8] Frank Maixner,[8] Albert Zink,[8,*] Stephan Schiffels,[2] and Johannes Krause[2,9,*]

[1]MOE Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Science, Fudan University, Shanghai 200438, China
[2]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[3]Center of Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai 200438, China
[4]Institute of Clinical Molecular Biology, Kiel University, 24118 Kiel, Germany
[5]Department of Evolutionary Anthropology, University of Vienna, 1030 Vienna, Austria
[6]Institute of Evolutionary Medicine, University of Zurich, 8057 Zurich, Switzerland
[7]Human Evolution and Archaeological Sciences, University of Vienna, 1030 Vienna, Austria
[8]Eurac Research - Institute for Mummy Studies, Viale Druso 1, 39100 Bolzano, Italy
[9]Lead contact
*Correspondence: albert.zink@eurac.edu (A.Z.), krause@eva.mpg.de (J.K.)
https://doi.org/10.1016/j.xgen.2023.100377

SUMMARY

The Tyrolean Iceman is known as one of the oldest human glacier mummies, directly dated to 3350–3120 calibrated BCE. A previously published low-coverage genome provided novel insights into European prehistory, despite high present-day DNA contamination. Here, we generate a high-coverage genome with low contamination (15.3×) to gain further insights into the genetic history and phenotype of this individual. Contrary to previous studies, we found no detectable Steppe-related ancestry in the Iceman. Instead, he retained the highest Anatolian-farmer-related ancestry among contemporaneous European populations, indicating a rather isolated Alpine population with limited gene flow from hunter-gatherer-ancestry-related populations. Phenotypic analysis revealed that the Iceman likely had darker skin than present-day Europeans and carried risk alleles associated with male-pattern baldness, type 2 diabetes, and obesity-related metabolic syndrome. These results corroborate phenotypic observations of the preserved mummified body, such as high pigmentation of his skin and the absence of hair on his head.

INTRODUCTION

The Tyrolean Iceman (hereafter referred to as the Iceman), also known as "Ötzi," is the world's oldest glacier mummy. Radiocarbon dating and stable isotope analysis have revealed that the Iceman lived during the Chalcolithic (Copper Age) in the Southern slopes of the eastern Italian Alps.[1,2] His remains were found in the Italian part of the Ötztal Alps in 1991 and were directly dated to 3350–3120 calibrated BCE. In 2012, Keller et al. published the first whole-genome sequence of the Iceman.[3] Comparative analyses based on autosomal data reported a close genetic affinity between the Iceman and present-day Sardinians. These findings were, however, published before genomes from a larger number of ancient western Eurasian individuals became available. Genomic data from European ancient individuals from 3000 to 4000 BCE, who we consider contemporaneous populations to the Iceman, showed that the genetic similarity between the Iceman and present-day Sardinians is due to common genetic components that were geographically widespread across Europe during the Neolithic period.[4–8] The geographic region of the Alps, where the Iceman was discovered, remains, however, rather understudied.

The first Iceman genome from 2012 was generated using the ABI SOLiD sequencer platform, which requires complex computational infrastructure at high economical cost.[9] It had relatively low coverage (7.6×) compared with the high-coverage genome generated in this study and showed the presence of modern human DNA contamination. Therefore, thanks to the recent development of sequencing technologies (Illumina technology) with higher output and lower cost, which have become standard in the field of ancient DNA research,[9] we produced a new high-coverage genome for the Iceman (15.3× coverage) from the same left iliac bone sample used for the 2012 study, with minimal modern human contamination (0.5% ± 0.06%). We found that the Iceman shows unusually high early Neolithic-farmer-related ancestry among the analyzed European individuals from the 4th millennium BCE. Moreover, we show that the two ancestry components from European hunter-gatherer-related ancestry and early Neolithic-farmer-related ancestry present in the Iceman admixed rather recently before the Iceman's death (56 ± 21 generations ago, namely 4880 ± 635 BCE), suggesting the survival of

Please cite this article in press as: Wang et al., High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100377

hunter-gatherer groups south of the Alps as late as 5000 to 4000 BCE.

RESULTS

A new high-coverage genome

We obtained two samples of the left iliac bone and the surrounding tissue for a series of four extractions each in order to improve the amount of recovered ancient DNA and reduce modern human DNA contamination. In order to identify the extracts with the highest human DNA content for further processing, we compared the percentage of mapped human reads (i.e., endogenous DNA) after shotgun sequencing on an Illumina HiSeq platform, ranging from 1.58% to 51.02% endogenous human DNA (Table S1), for DNA libraries generated from these eight extracts (Table S1A). From the best two extracts (1412E2 and 1412E3), four double-stranded DNA libraries were generated with uracil DNA glycosylase (UDG), which reduces substitutions due to ancient DNA damage (i.e., deaminated cytosines) at the end of DNA fragments.[10] We then performed paired-end shotgun sequencing on a total of 36 Illumina HiSeq sequencing lanes for all four libraries (Tables S1B and S1C). We processed the raw sequencing data with EAGER 1.92.2[11] and obtained a combined alignment with average genomic coverage of 15.3× after removing duplicates. In the end, we obtained 45.4% endogenous human DNA content (Table S1), with more than 90.6% of the genome covered by at least five reads (STAR Methods). We estimated the contamination level in the high-coverage genome using ANGSD based on the heterozygosity on the haploid X chromosome.[12] The high-coverage Iceman genome has 0.5% ± 0.06% contamination, 10× less than the contamination level found in the previously published genome sequence[3] (7.5% ± 0.25%) (Table S2).

Genetic ancestry analysis

It has been shown that early Neolithic European farmers derived most of their ancestry from early Anatolian farmers, suggesting that farming spread with people from the Near East through Anatolia and the Balkan peninsula starting around 7000 BCE.[4,5,13] The arrival of farmers in Europe is followed by an increasing amount of admixture with local hunter-gatherers at variable levels during the initial expansion[14–16] and subsequently into the 4th millennium BCE. By the end of the 4th millennium BCE, admixture between early Neolithic-farmer-related ancestry originating from Anatolia and European hunter-gatherer-related ancestry had become prevalent in most parts of Europe.[5,15–19] Later, beginning from 2900 BCE, herders from the Pontic-Caspian steppe introduced substantial levels of so-called "Steppe-related ancestry" throughout Europe.[17] After the 3rd millennium BCE until present day, all three ancestry components are found in almost all European populations.[13]

We examined the Iceman's ancestry makeup in the context of those three ancestral components with corresponding representative proxies—western hunter-gatherers ("WHGs"), early Neolithic farmers from Anatolia ("Anatolia_N") or Germany ("Germany_EN_LBK"), and herders from the Samara region ("Russia _Samara_EBA_Yamnaya")—together with ancient populations from Germany, the northern Iberian Peninsula, Italy, and Sardinia also dated to the 4th millennium BCE (Figure 1; Table S3).

Projecting both the high-coverage and the previously published[3] Iceman genome on modern western Eurasian genetic variation using principal-component analysis (PCA) (STAR Methods), the high-coverage genome is slightly shifted compared with the previously published one[3] (Figure 2). The high-coverage genome of the Iceman clusters between the two groups in the PCA formed by (1) Middle-Neolithic and Chalcolithic Europeans dated to the 4th millennium BCE and by (2) early Neolithic European farmers, which in turn fall closely together with earlier farmers from Anatolia. In the genetic affinity test, the high-coverage Iceman genome shows a similar pattern, presenting the closest genetic affinity to contemporaneous Europeans from the 4th millennium BCE and early Neolithic European farmers (Figure S1).

In the PCA (Figure 2), the high-coverage Iceman genome locates closer to the cluster of early Neolithic Farmers from Europe (farmers associated with Linear Pottery culture, in short "LBK" hereafter) and Anatolia_N than other Middle-Late Neolithic to Chalcolithic Europeans (Spain_MLN, Italy_Sardinia_N, Italy_Sardinia_C, Italy_N.SG, Germany_MN_Baalberge, Germany_MN_Salzmuende, Germany_MN_Esperstedt, Italy_Broion_CA.SG),[6,7,18,20] indicating that the Iceman may have more early Neolithic-farmer-related ancestry than other tested European individuals from the 4th millennium BCE.

To calculate the exact proportions of ancestral components in the Iceman and other contemporaneous ancient European groups, we applied *qpAdm* modeling to test three proxies for the early Neolithic-farmer-related ancestry—Germany_EN_LBK, Anatolia_N, and Germany_EN_LBK_Stuttgart.DG (a 7,000-year-old high-coverage shotgun genome from the LBK culture in Germany) (Tables S3 and S4). We found that the Iceman derives 90% ± 2.5% ancestry from early Neolithic farmer populations when using Anatolia_N as the proxy for the early Neolithic-farmer-related ancestry and WHGs as the other ancestral component (Figure 3; Table S4). When testing with a 3-way admixture model including Steppe-related ancestry as the third source for the previously published[3] and the high-coverage genome, we found that our high-coverage genome shows no Steppe-related ancestry (Table S5), in contrast to ancestry decomposition of the previously published Iceman genome.[3,17] We conclude that the 7.5% Steppe-related ancestry previously estimated for the previously published Iceman genome[3,17] is likely the result of modern human contamination.

Compared with the Iceman, the analyzed contemporaneous European populations from Spain and Sardinia (Italy_Sardinia_C, Italy_Sardinia_N, Spain_MLN) show less early Neolithic-farmer-related ancestry, ranging from 27.2% to 86.9% (Figure 3A; Table S4). Even ancient Sardinian populations,[7] who are located further south than the Iceman and are geographically separate from mainland Europe, derive no more than 85% ancestry from Anatolia_N (Figure 3; Table S4). The higher levels of hunter-gatherer ancestry in individuals from the 4th millennium BCE have been explained by an ongoing admixture between early farmers and hunter-gatherers in the Middle and Late Neolithic in various parts of Europe, including western Europe (Germany and France), central Europe, Iberia, and the

Please cite this article in press as: Wang et al., High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100377

CellPress
OPEN ACCESS

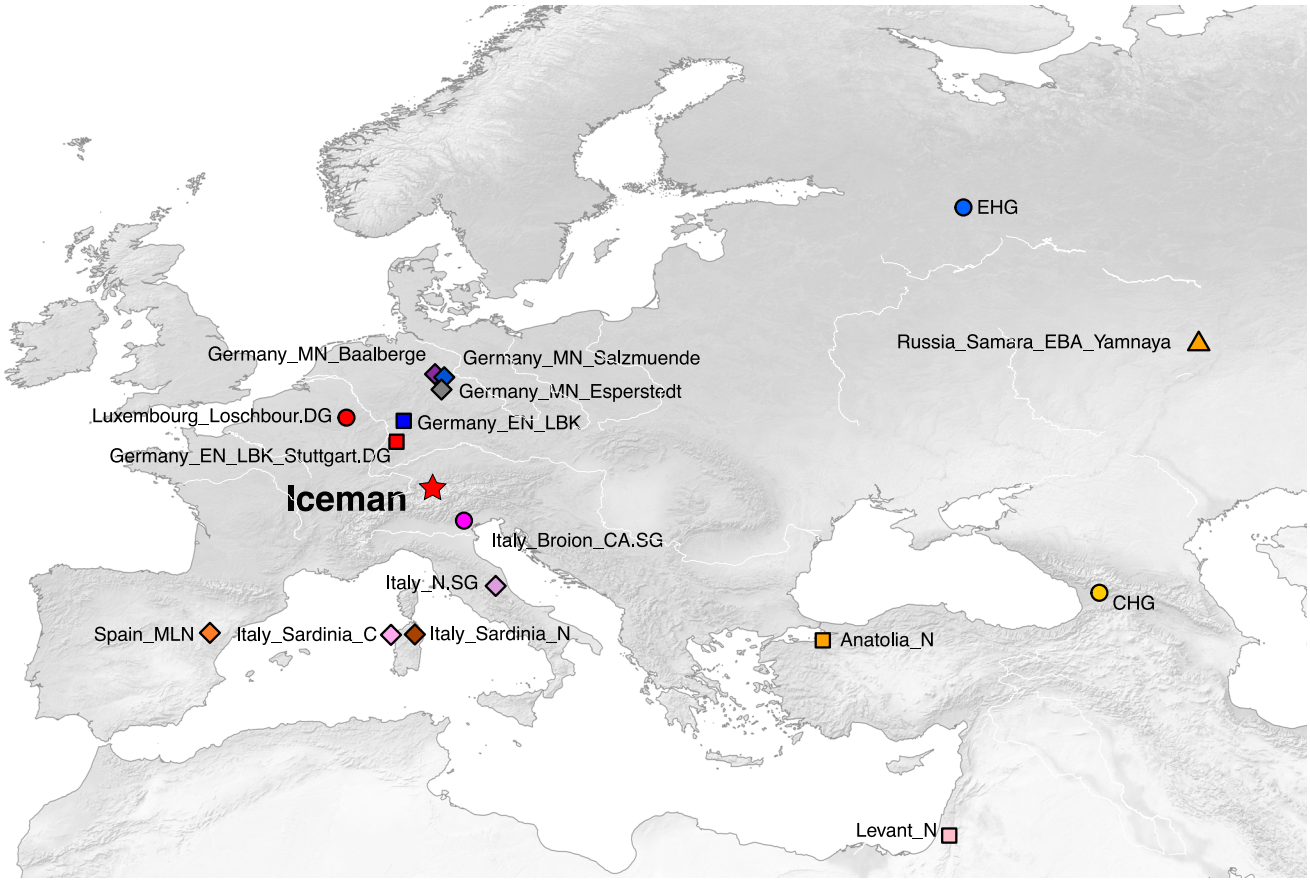


Figure 1. Geographic location of the Iceman and analyzed published ancient western Eurasian groups
See also Table S3.

Balkans.[5,15–19] Only individuals from Italy_Broion_CA.SG found to the south of the Alps present similarly low hunter-gatherer ancestry as seen in the Iceman.[21] We conclude that the Iceman and Italy_Broion_CA.SG might both be representatives of specific Chalcolithic groups carrying higher levels of early Neolithic-farmer-related ancestry than any other contemporaneous European group. This might indicate less gene flow from groups that are more admixed with hunter-gatherers or a smaller population size of hunter-gatherers in that region during the 5th and 4th millennium BCE.

Recent admixture between early farmers and hunter-gatherers in southern Europe

Given the high proportion of early Neolithic-farmer-related ancestry in the Iceman genome, we also tested if using the early Neolithic-farmer-related ancestry as a single source is sufficient. We found that *qpWave* results suggest neither Anatolia_N nor Germany_EN_LBK as an appropriate single source, confirming that the European hunter-gatherer-related ancestry is low but significantly present in the Iceman's genome ($p < 0.05$; Table S6).

We estimated the admixture date between the early Neolithic-farmer-related (using Anatolia_N as proxy) and WHG-related ancestry sources using *DATES*[22] to be 56 ± 21 generations before the Iceman's death, which corresponds to 4880 ± 635 calibrated BCE assuming 29 years per generation[23] (Figure 3B; Table S7) and considering the mean C14 date of this individual. Alternatively, using Germany_EN_LBK as the proxy for early Neolithic-farmer-related ancestry, we estimated the admixture date to be 40 ± 15 generations before his death (Table S7), or 4400 ± 432 calibrated BCE, overlapping with estimates from nearby Italy_Broion_CA.SG, who locate to the south of the Alps[7,18,20] (Figure 3B).

While compared with the admixture time between early Neolithic farmers and hunter-gatherers in other parts of southern Europe, for instance in Spain and southern Italy, we found that, particularly, the admixture with hunter-gatherers as seen in the Iceman and Italy_Broion_CA.SG is more recent (Figure 3B; Table S3), suggesting a potential longer survival of hunter-gatherer-related ancestry in this geographical region.

Effective population size and heterozygosity

The high-coverage genome allows for additional analyses, such as estimating effective population sizes through time and genome-wide heterozygosity, that are not possible with lower-coverage or SNP-captured genomes. Specifically, we estimated the population-size history of the population represented by the

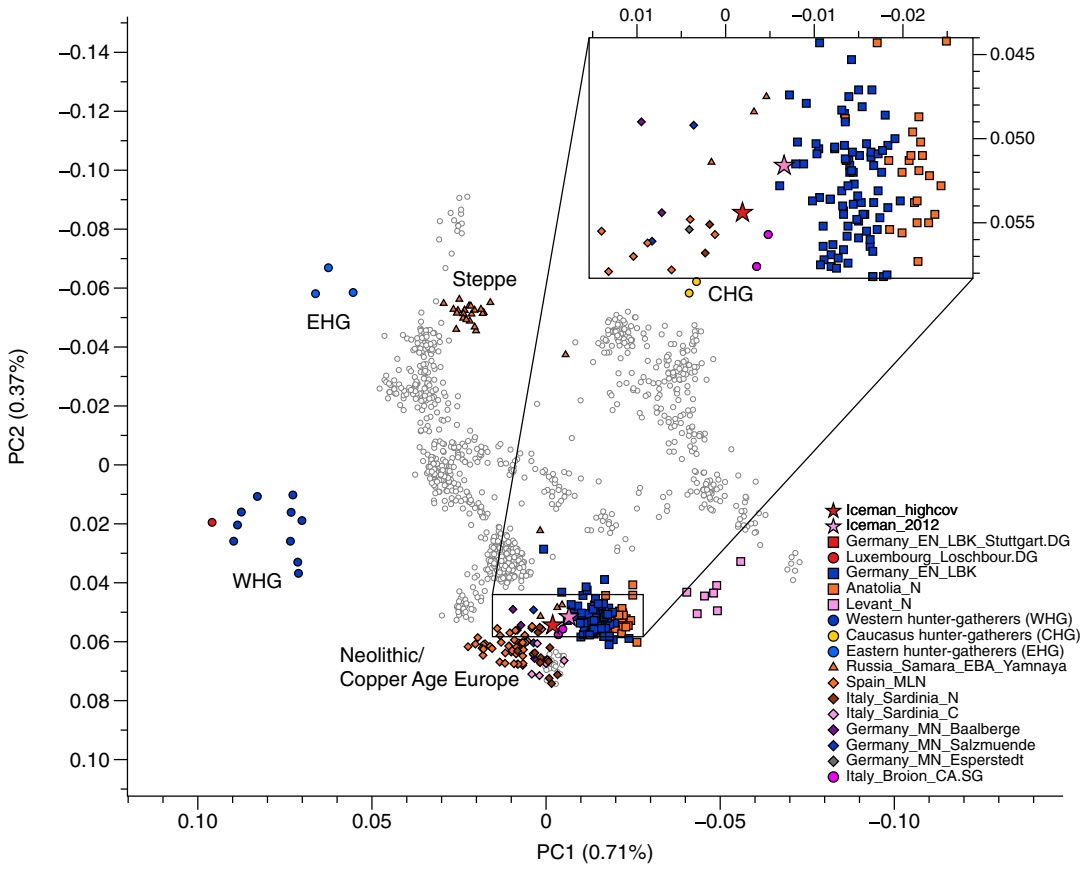


Figure 2. Principal-component analysis (PCA)

We project the high-coverage Iceman genome, the previously published Iceman genome, and related published ancient western Eurasian individuals onto present-day western Eurasian populations. See also Table S3.

Iceman and the two source populations represented by an early Neolithic farmer from Stuttgart in Germany ("Germany_EN_LBK_Stuttgart.DG") and by a Mesolithic hunter-gatherer from Loschbour in Luxembourg ("Luxembourg_Loschbour.DG")[13] using MSMC2.[24] The demographic histories estimated using the aforementioned three ancient high-coverage genomes share the same population bottleneck between 25,000 and 200,000 years ago, similar to that obtained from a present-day Sardinian individual, and they show a slight population size increase in a recent time epoch from 20,000–25,000 years ago (Figure S2). We observe a higher population size in recent times for the Iceman and the Germany_EN_LBK_Stuttgart.DG individual (both with high early Neolithic-farmer-related ancestry) compared with the Luxembourg_Loschbour.DG hunter-gatherer, which is possibly linked to the larger population size of early farming populations versus the hunter-gatherer populations in recent times.[25]

We estimated the rate of heterozygosity for the Iceman, Germany_EN_LBK_Stuttgart.DG, Luxembourg_Loschbour.DG, and a present-day Sardinian individual and plotted the per-chromosome estimate together with the standard error calculated from a weighted jackknife procedure in Figure S3 (Table S8). Both Germany_EN_LBK_Stuttgart.DG and the Iceman show higher heterozygosity levels than Luxembourg_Loschbour.DG, but the Iceman shows a relatively lower level than Germany_EN_LBK_Stuttgart.DG. This is consistent with the supposed relative isolation of the Iceman and the low WHG-derived ancestry seen in his genome.

New insights into the phenotypic traits and local ancestry assignments of the Iceman

The high-coverage genome allows us to investigate SNPs of phenotypic significance with sufficient coverage on individual allelic sites. We analyzed 147 SNPs of phenotypic interest, summarized in Table S9, including phenotypic sites examined in the previously published genome.[3] We newly reported alleles for phenotypic traits of the Iceman related to reduced hair curliness, black hair color, obesity-related metabolic disorders, reduced freckling, and male-pattern baldness (Tables S9 and S10), in addition to the previously reported phenotype of possibly light skin pigmentation, brown eyes, and blood type O from Keller et al.[3] (Table S9).

In particular, five SNPs (rs4988235, rs1050152, rs1495741, rs4751995, rs174546) assumed to be related to the adaptation to an agricultural lifestyle[26] suggest that the Iceman was a comparatively slow metabolizer with low concentrations of animal-oriented

Please cite this article in press as: Wang et al., High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100377

CellPress
OPEN ACCESS

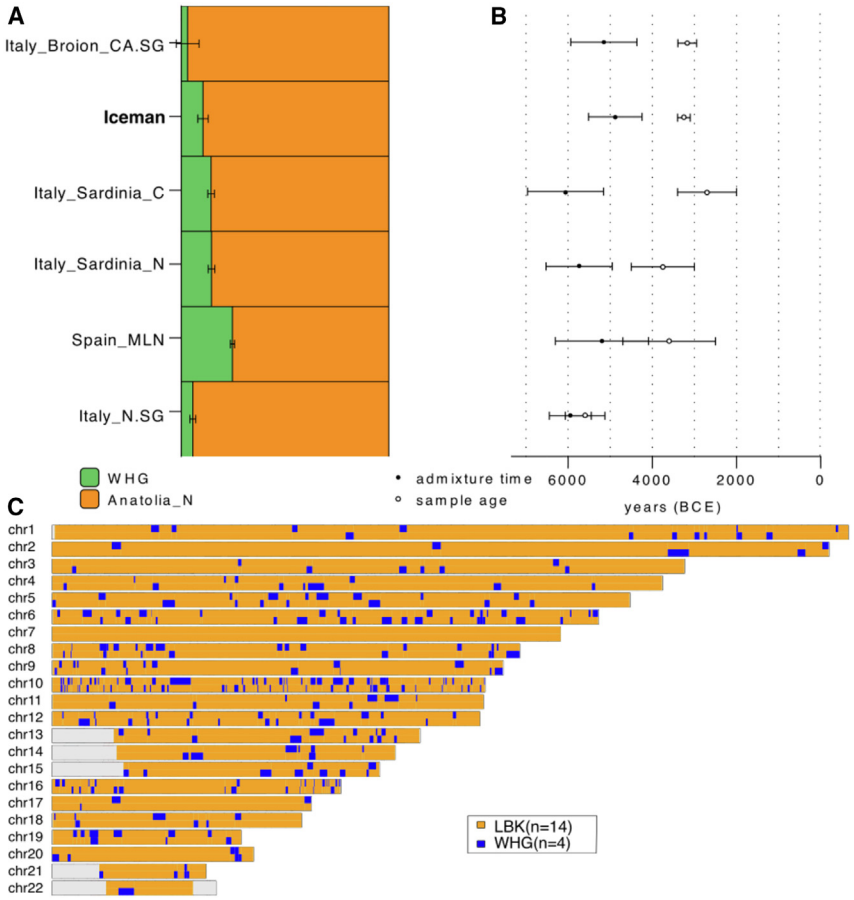


Figure 3. Global and local ancestral proportions and admixture time of ancient groups in the 4th millennium BCE

(A) *qpAdm* estimates on the proportion of western hunter-gatherer (WHG) and Anatolia_N ancestry in Iceman and contemporaneous ancient European groups.

(B) Dating the admixture time using *DATES* for target populations shown in panel (A).

Horizontal bars in (A) and (B) represent ±1 standard error (SE) estimated by *qpAdm* and *DATES* correspondingly.

(C) Local ancestry assignments across 22 autosomes of the Iceman genome. The majority of the Iceman genome is inferred to be of farmer origin (LBK, orange), and a small fraction is inferred to be of WHG origin (blue), consistent with the global ancestry proportion estimates from *qpAdm*. Gray area represents short arms on the chromosomes, which do not contain any unique genetic material. See also Table S4.

fatty acids but high concentrations of plant-oriented fatty acids and may have had an intermediate high-density lipoprotein-cholesterol concentration level in general (Tables S9 and S10).

While genetic information cannot yet be used to completely reconstruct the appearance of an individual, genetic models exist for specific phenotypic features. Among those, skin pigmentation is a relatively well-understood trait that can be inferred from genetic data. We examined 170 skin pigmentation-associated SNPs from the UK Biobank genome-wide association study (GWAS) for skin color[27,28] and retrieved diploid genotype information from 154 biallelic sites in Iceman. Each phenotype-informative SNP has a different effect size, i.e., the variance in pigmentation explained by individual SNPs is different. Thus, we combined the effect size of each pigmentation-informative SNP together with all examined effective alleles as an indication for the final phenotypic trait. To take effect size impact into consideration, we calculated a weighted genetic score given the individual SNP weight from the UK Biobank GWAS-estimated effect sizes,[27,28] which is the weighted proportion of dark pigmentation alleles used as an indicator of skin pigmentation. The weighted genetic score of dark pigmentation in the Iceman is estimated to be 0.591, higher than the score of present-day southern European populations taking Sardinians as an example (Table S11), which the Iceman shares closest genetic affinity to (Figure S1) and which represent the highest level of pigmentation among modern-day European groups,[29] although it is lower than the score of ancient LBK farmers and the Luxembourg_Loschbour.DG hunter-gatherer (Table S11).

The high-coverage genome also enables us to explore the ancestral origin of genomic regions along the genome given a set of phased genomes as ancestral references, making it possible to assign SNPs of phenotypic interests to a specific ancestral origin. We assigned local ancestry tracts across 22 autosomes of the Iceman using *RFMix*[30] (Figure 3C), employing an imputed and phased dataset as the reference for WHG and the early Neolithic-farmer-related ancestry (STAR Methods). The WHG ancestry tracts have an average length of 1.174cM on average across 22 autosomes, close to the expected tract length calculated from admixture time and the WHG admixture proportion (1.754cM). In total, genomic tracts of the early Neolithic-farmer-related ancestry account for 91.4% of the genome, with the remaining genomic chunks being assigned to WHG origin (in 8.6%), in line with the global ancestry estimation from *qpAdm* (90%/10%; Table S4). Based on the assigned local ancestral tract distribution, we inferred two farmer's diet-related SNPs (rs1495741, rs174546) mentioned above to be of LBK farmer origin, as expected.

DISCUSSION

The reconstruction of a high-quality genome of the Iceman using Illumina sequencing enabled reanalyses providing novel insights into the phenotypic traits and genetic history of this individual. Contamination estimates showed that the high-coverage genome is almost free of contamination, in contrast to the previous genome showing around 7% of human DNA contamination.

Unlike previous analyses performed on the previous Iceman genome,[17] we found no evidence for the presence of

Please cite this article in press as: Wang et al., High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100377

Steppe-related ancestry in the high-coverage one. Instead, his genome is best modeled as a genetic mixture between European hunter-gatherer-related ancestry and early Neolithic-farmer-related ancestry. The absence of Steppe-related ancestry is consistent with the dating of this individual preceding the arrival of Steppe-related ancestry in central and southern Europe.[17] We found that the Iceman, together with the contemporary Italy_Broion_CA.SG located to the south of the Alps, carries the largest proportion of early Neolithic-farmer-related ancestry among all contemporaneous European individuals analyzed so far, suggesting that these individuals were relatively isolated from other European individuals who were more genetically admixed with ancient European hunter-gatherers. The remote location of Alpine valleys might contribute to such an isolation. We did, however, not find lower levels of genetic diversity in the Iceman genome compared with other early European farmers (represented by LBK farmers in Germany), with no signs of inbreeding. Altogether, these observations add nuance to our understanding of the mixture processes underlying the rise of the European hunter-gatherer-related ancestry throughout Europe during the 5th and 4th millennium BCE.[5,15,17–19,31] While the general trend of rising European hunter-gatherer-related ancestry has been described throughout Europe, the presence of individuals with low European hunter-gatherer-related ancestry south of and within the eastern Italian Alps suggests regional heterogeneity in Late Neolithic Europe.

The high-coverage genome yielded further novel insights into the possible phenotypic traits of the Iceman, especially for complex traits regulated by multiple SNPs in the genome. We found that SNPs associated with an agricultural diet were present in the Iceman genome, two of which are assigned to local ancestry tracts of farmer origin, in line with the estimated high early Neolithic farmer ancestry. Our estimation of skin pigmentation for the Iceman based on over 100 regulatory SNPs related to that trait suggests that he displayed a rather dark skin, as also displayed by the actual mummy.[32] While this was discussed as a result of the mummification process itself,[33] our findings indicate a relatively dark skin complexion during his lifetime. Additional support for this assumption comes from a previous histological analysis of the Iceman's skin, where a small layer of brown melanin granules had been identified in the *stratum basale* of the epidermis.[32] The appearance of the baldness-related allele in the high-coverage Iceman genome may be related to the fact that almost no human hair was found with the otherwise well-preserved mummy.[34] Similar approaches like polygenic risk score and heritability calculation on other ancient high-coverage genomes might allow a more accurate reconstruction of complex phenotypic appearance such as skin pigmentation[35] of ancient human individuals.

Limitations of the study

This study produced a high-coverage genome for the Tyrolean Iceman that enabled the detection of high Anatolian-farmer-related genetic ancestry in his genome; a single individual has, however, limited resolution in representing the population history of his time and region. Nevertheless, another individual from Broion Italy, bordering the southern Alps, presents similarly high Anatolian-farmer-related ancestry, supporting the observation for the Iceman genome. Future studies with a denser sampling from the southern Alps will be needed to replicate our findings and show if the Iceman was an outlier or a representative of his population.

Moreover, this study makes exploratory analyses on cross-comparing genetic-predicting scores of phenotypes based on high-coverage ancient genomes. For instance, we estimate the possible phenotypes based on the presence of phenotype-related alleles and the prediction on skin pigmentation based on polygenetic risk sores. We caution that the actual phenotype is a combined effect from genetic mechanism and environment exposures through gene-by-environment interaction,[36] and multiple SNPs could be responsible for the heritability of complex traits like male-pattern baldness[37] and skin pigmentation.[27] Here, direct observation of the actual mummy allows us, however, to validate some of the findings such as pigmentation and baldness, corroborating the genetic prediction based on the genomic data. However, genomic data from ancient mummies is rather exceptional. In most ancient DNA studies, the predictive accuracy of ancient polygenetic scores for complex traits should be interpreted carefully together with various confounding factors, such as allelic turnover[38] or population stratification.[39]

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Genome generation
- METHOD DETAILS
  - Bioinformatic processing
  - Principal component analysis (PCA)
  - Outgroup f_3 statistics
  - qpAdm and qpWave analyses
  - Local ancestry decomposition
  - Effective population size estimates
  - Phenotypic SNP analyses
  - Estimates of heterozygosity

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2023.100377.

ACKNOWLEDGMENTS

We thank Alexander Peltzer for his help with the bioinformatic processing of shallow sequencing data. We thank the Multimedia department at the Max Planck Institute for Evolutionary Anthropology, Leipzig, and Michelle O'Reilly for help with the geographic map. We thank Dan Ju for discussion on the polygenetic risk score of skin pigmentation. V.J.S. was supported by the University of Zurich's University Research Priority Program "Evolution in Action: From Genomes to Ecosystems."

AUTHOR CONTRIBUTIONS

J.K. and A.Z. conceived the study. J.K., S.S., and A.Z. supervised the study. A.Z., F.M., and V.C. provided archaeological material and advised on the material background and interpretation. V.J.S. and B.K.-K. performed laboratory work. K.W., K.P., A.C., and S.S. analyzed data. K.W. and J.K. wrote the manuscript with input from all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 3, 2022
Revised: May 10, 2023
Accepted: July 13, 2023
Published: August 16, 2023

REFERENCES

1. Bonani, G., Ivy, S.D., Hajdas, I., Niklaus, T.R., and Suter, M. (1994). Ams 14C age determinations of tissue, bone and grass samples from the Ötztal Ice Man. Radiocarbon *36*, 247–250.
2. Müller, W., Fricke, H., Halliday, A.N., McCulloch, M.T., and Wartho, J.-A. (2003). Origin and migration of the Alpine Iceman. Science *302*, 862–866.
3. Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F., Leidinger, P., Backes, C., Khairat, R., Forster, M., et al. (2012). New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nat. Commun. *3*, 698.
4. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.
5. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.
6. Fernandes, D.M., Mittnik, A., Olalde, I., Lazaridis, I., Cheronet, O., Rohland, N., Mallick, S., Bernardos, R., Broomandkhoshbacht, N., Carlsson, J., et al. (2020). The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. Nat. Ecol. Evol. *4*, 334–345.
7. Marcus, J.H., Posth, C., Ringbauer, H., Lai, L., Skeates, R., Sidore, C., Beckett, J., Furtwängler, A., Olivieri, A., Chiang, C.W.K., et al. (2020). Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia. Nat. Commun. *11*, 939.
8. Raveane, A., Aneli, S., Montinaro, F., Athanasiadis, G., Barlera, S., Birolo, G., Boncoraglio, G., Di Blasio, A.M., Di Gaetano, C., Pagani, L., et al. (2019). Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. Sci. Adv. *5*, eaaw3492.
9. Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. *2012*, 251364.
10. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20130624.
11. Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biol. *17*, 60.
12. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinf. *15*, 356.
13. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.
14. Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-Del-Molino, D., van Dorp, L., López, S., Kousathanas, A., Link, V., et al. (2016). Early farmers from across Europe directly descended from Neolithic Aegeans. Proc. Natl. Acad. Sci. USA *113*, 6886–6891.
15. Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. Nature *555*, 197–203.
16. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature *551*, 368–372.
17. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.
18. Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C.J., Gandini, F., Pala, M., et al. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. Science *363*, 1230–1234.
19. Rivollat, M., Jeong, C., Schiffels, S., Küçükkalıpçı, İ., Pemonge, M.-H., Rohrlach, A.B., Alt, K.W., Binder, D., Friederich, S., Ghesquière, E., et al. (2020). Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. Sci. Adv. *6*, eaaz5344.
20. Antonio, M.L., Gao, Z., Moots, H.M., Lucci, M., Candilio, F., Sawyer, S., Oberreiter, V., Calderon, D., Devitofranceschi, K., Aikens, R.C., et al. (2019). Ancient Rome: A genetic crossroads of Europe and the Mediterranean. Science *366*, 708–714.
21. Saupe, T., Montinaro, F., Scaggion, C., Carrara, N., Kivisild, T., D'Atanasio, E., Hui, R., Solnik, A., Lebrasseur, O., Larson, G., et al. (2021). Ancient genomes reveal structural shifts after the arrival of Steppe-related ancestry in the Italian Peninsula. Curr. Biol. *31*, 2576–2591.e12.
22. Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., et al. (2019). The formation of human populations in South and Central Asia. Science *365*, eaat7487.
23. Fenner, J.N. (2005). Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am. J. Phys. Anthropol. *128*, 415–423.
24. Schiffels, S., and Wang, K. (2020). MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. In Statistical Population Genomics, J.Y. Dutheil, ed. (Springer US), pp. 147–166.
25. Marchi, N., Winkelbach, L., Schulz, I., Brami, M., Hofmanová, Z., Blöcher, J., Reyna-Blanco, C.S., Diekmann, Y., Thiéry, A., Kapopoulou, A., et al. (2022). The genomic origins of the world's first farmers. Cell *185*, 1842–1859.e18.
26. Mathieson, S., and Mathieson, I. (2018). FADS1 and the Timing of Human Adaptation to Agriculture. Mol. Biol. Evol. *35*, 2957–2970.
27. Ju, D., and Mathieson, I. (2021). The evolution of skin pigmentation-associated variation in West Eurasia. Proc. Natl. Acad. Sci. USA *118*, e2009227118.
28. Lab, N. (2018). UK Biobank GWAS. www.nealelab.is/uk-biobank.
29. Liu, F., Visser, M., Duffy, D.L., Hysi, P.G., Jacobs, L.C., Lao, O., Zhong, K., Walsh, S., Chaitanya, L., Wollstein, A., et al. (2015). Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. Hum. Genet. *134*, 823–835.
30. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet. *93*, 278–288.
31. Lipson, M., and Reich, D. (2017). A Working Model of the Deep Relationships of Diverse Modern Human Genetic Lineages Outside of Africa. Mol. Biol. Evol. *34*, 889–902.

32. Pabst, M.A., Letofsky-Papst, I., Bock, E., Moser, M., Dorfer, L., Egarter-Vigl, E., and Hofer, F. (2009). The tattoos of the Tyrolean Iceman: a light microscopical, ultrastructural and element analytical study. J. Archaeol. Sci. *36*, 2335–2341.

33. Spindler, K. (2000). Der Mann im Eis. Neue sensationelle Erkenntnisse über die Mumie in den Ötztaler Alpen (Goldmann).

34. Wilfing, H., Teschler-Nicola, M., Seidler, H., Weber, G., Schlagenhaufen, C., Irgolic, K.J., Gössler, W., Platzer, W., Spindler, K., Notdurfther, H., et al. (1993). Untersuchungen an Haarresten des Mannes vom Hauslabjoch. Naturwiss. Rundsch. *46*, 257–260.

35. Martin, A.R., Lin, M., Granka, J.M., Myrick, J.W., Liu, X., Sockell, A., Atkinson, E.G., Werely, C.J., Möller, M., Sandhu, M.S., et al. (2017). An Unexpectedly Complex Architecture for Skin Pigmentation in Africans. Cell *171*, 1340–1353.e14.

36. Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J.K., and Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. Elife *9*, e48376. https://doi.org/10.7554/eLife.48376.

37. Pirastu, N., Joshi, P.K., de Vries, P.S., Cornelis, M.C., McKeigue, P.M., Keum, N., Franceschini, N., Colombo, M., Giovannucci, E.L., Spiliopoulou, A., et al. (2017). GWAS for male-pattern baldness identifies 71 susceptibility loci explaining 38% of the risk. Nat. Commun. *8*, 1584.

38. Carlson, M.O., Rice, D.P., Berg, J.J., and Steinrücken, M. (2022). Polygenic score accuracy in ancient samples: Quantifying the effects of allelic turnover. PLoS Genet. *18*, e1010170.

39. Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., et al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. *43*, 519–525.

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

41. Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. *40*, e3.

42. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.

43. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224.

44. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

45. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

46. Prüfer, K. (2018). snpAD: an ancient DNA genotype caller. Bioinformatics *34*, 4165–4171.

47. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.

48. Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyzy, N., Flegontov, P., Zaporozhchenko, V., Immel, A., Wang, C.-C., Ixan, O., Khussainova, E., et al. (2019). The genetic history of admixture across inner Eurasia. Nat. Ecol. Evol. *3*, 966–976.

49. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.

50. Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. Curr. Biol. *23*, 553–559.

51. Biagini, S.A., Solé-Morata, N., Matisoo-Smith, E., Zalloua, P., Comas, D., and Calafell, F. (2019). People from Ibiza: an unexpected isolate in the Western Mediterranean. Eur. J. Hum. Genet. *27*, 941–951.

52. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.

53. Lazaridis, I., Alpaslan-Roodenberg, S., Acar, A., Açıkkol, A., Agelarakis, A., Aghikyan, L., Akyüz, U., Andreeva, D., Andrijašević, G., Antonović, D., et al. (2022). Ancient DNA from Mesopotamia suggests distinct Pre-Pottery and Pottery Neolithic migrations into Anatolia. Science *377*, 982–987.

54. Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. Nature *555*, 190–196.

55. Villalba-Mouco, V., van de Loosdrecht, M.S., Posth, C., Mora, R., Martínez-Moreno, J., Rojo-Guerra, M., Salazar-García, D.C., Royo-Guillén, J.I., Kunst, M., Rougier, H., et al. (2019). Survival of late Pleistocene Hunter-gatherer ancestry in the Iberian Peninsula. Curr. Biol. *29*, 1169–1177.e7.

56. Gallego Llorente, M., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W., Curtis, M.C., Stock, J.T., Coltorti, M., Pieruccini, P., et al. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. Science *350*, 820–822.

57. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.

58. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.

59. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.

60. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat. Commun. *6*, 8912.

61. Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J., and Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. Nat. Genet. *53*, 120–126.

62. Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., et al. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. Science *358*, 655–658.

63. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nat. Genet. *46*, 919–925.

64. Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.

Please cite this article in press as: Wang et al., High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100377

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human archaeological skeletal material	This study	Iceman
Chemicals, peptides, and recombinant proteins		
USER enzyme	New England Biolabs	Cat# M5505
Uracil Glycosylase inhibitor (UGI)	New England Biolabs	Cat# M0281
AccuPrime Pfx DNA polymerase	Invitrogen	Cat#12344024
Phusion High Fidelity DNA polymerase	Thermo Scientific	Cat# F530S
1x Tris-EDTA pH 8.0	AppliChem	Cat# A85690500
0.5 M EDTA pH 8.0	Life Technologies	Cat# AM9261
10x Buffer Tango	Life Technologies	Cat# BY5
Isopropanol	Merck	Cat# 1070222511
Ethanol	Merck	Cat# 1009832511
Proteinase K	Sigma Aldrich	Cat# P2308
Guanidine hydrochloride	Sigma Aldrich	Cat# G3272
3M Sodium Acetate pH 5.2	Sigma Aldrich	Cat# S7899
Tween-20	Sigma Aldrich	Cat# P9416
5M NaCl	Sigma Aldrich	Cat# S5150
ATP 100 mM	Thermo Fisher Scientific	Cat# R0441
1 M Tris-HCl pH 8.0	Thermo Fisher Scientific	Cat# 15568025
GeneAmp 10x PCR Gold Buffer	Thermo Fisher Scientific	Cat# 4379874
dNTP Mix	Thermo Fisher Scientific	Cat# R1121
T4 DNA Polymerase	New England Biolabs	Cat# M0203
T4 Polynucleotide Kinase	New England Biolabs	Cat# M0201
Bst 2.0 DNA Polymerase	New England Biolabs	Cat# M0537
Quick Ligation Kit	New England Biolabs	Cat# M2200L
Critical commercial assays		
MinElute PCR Purification Kit	QIAGEN	Cat# 28006
Deposited data		
Raw and analyzed data	This study	ENA: PRJEB56570
Software and algorithms		
EAGER v1.92.2	Peltzer et al.[11]	https://github.com/apeltzer/EAGER-GUI
AdapterRemoval v2.2.0	Schubert et al.[40]	https://github.com/MikkelSchubert/adapterremoval
BWA v0.7.12	Li and Durbin[41]	http://bio-bwa.sourceforge.net
dedup v0.12.2	Peltzer et al.[11]	https://github.com/apeltzer/DeDup
snpAD	Pruefer[42]	https://bioinf.eva.mpg.de/snpAD/
Eigensoft v7.2.1	Patterson et al.[43]	https://github.com/DReichLab/EIG
DATES v753	Narasimhan et al.[22]	https://github.com/priyamoorjani/DATES
admixtools v5.1	Patterson et al.[44]	https://github.com/DReichLab/AdmixTools
Schmutzi	Renaud et al.[45]	https://bioinf.eva.mpg.de/schmutzi
ANGSD v0.910	Korneliussen et al.[12]	http://www.popgen.dk/angsd/index.php/ANGSD
contamMix 1.0-10	Fu et al.[46]	https://github.com/plfjohnson/contamMix
RFMix v2.03-r0	Maples et al.[29]	http://med.stanford.edu/bustamantelab/
MSMC2	Schiffels and Wang[47]	https://github.com/stschiff/msmc-tools/
GLIMPSE	Rubinacci et al.[48]	https://odelaneau.github.io/GLIMPSE

Please cite this article in press as: Wang et al., High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100377

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed and will be fulfilled by the lead contact, Johannes Krause (krause@eva.mpg.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Sequencing data are available at the European Nucleotide Archive (ENA) under ENA: PRJEB56570.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Genome generation

A bone biopsy was taken from the Iceman's left ilium (0.1 g) under sterile conditions in the Iceman's preservation cell at the South Tyrol Archaeological Museum in Bolzano, Italy. DNA was extracted at the Institute for Archaeological Sciences, University of Tübingen, Germany, by using two powdered samples (1411 and 1412) of bone material and surrounding tissue from pelvic bone for a series of four sequential extractions each to reduce contamination and maximise the recovery of endogenous DNA. We followed the extraction protocol by Dabney et al.,[49] and varied incubation times and temperatures in four test extractions. In the first extraction (E1) the samples were incubated for 10 min with the extraction buffer at 37°C. The samples were then centrifuged and the extraction buffer removed. These steps were then repeated for a second extraction (E2). In the third extraction (E3) the incubation time was elongated to an overnight incubation at 37°C; and for the last extraction (E4) one hour at 56°C was used. From each of the extractions, we took an aliquot of 10μl to build a double-stranded, dual-indexed DNA library,[40,41] for which shallow shotgun sequencing was performed on an Illumina HiSeq platform in order to identify the optimal extracts for deeper sequencing. The resulting reads were processed with the EAGER[11] pipeline, revealing the highest percentage of mapped reads to the human reference genome for the extracts 1412E2 and 1412E3 (Table S1). Of these two extracts, four additional double-stranded DNA libraries were prepared with uracil DNA glycosylase (UDG) treatment.[40,41] To generate a high coverage genome, we then subjected the libraries to paired-end shotgun sequencing on a total of 36 Illumina HiSeq sequencing lanes.

METHOD DETAILS

Bioinformatic processing

We processed raw sequencing data per sequencing lane following the pipeline in *EAGER 1.92.2*,[11] with adaptors removed by *AdapterRemoval v2*[42] reads mapped to the human reference genome by *BWA v0.7.12*[45] and polymerase chain reaction (PCR) duplicates removed by *Dedup v0.12.2*.[11] We merged sequencing data for each sequencing library and performed duplication removal again for each individual-library-based BAM (Table S1). Across four sequencing libraries, there are 2,221,240,494 sequencing reads in total, out of which 1,008,074,503 reads mapped to the human reference genome hs37d5. Then, we merged four individual-library-based BAM files (after removing duplicates), and ended up with 707,832,853 reads (Table S1), based on which we ran *Dedup v0.12.2*[11] again, and finally obtained 678,475,568 reads for genomic coverage calculation and downstream analyses.

We called diploid genotypes with *snpAD*[46] using a minimum base quality (Phred-scaled) of 30 and a minimum mapping quality (Phred-scaled) of 30. It resulted in 1,062,059 SNPs overlapping with the 1240k SNP panel. The diploid genotype was later merged with a publicly released dataset of ancient and present-day individuals for the same set of ~1.24million SNPs (https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadrdownloadable-genotypes-present-day-and-ancient-dna-data, v44.3) and individuals from Saupe et al. 2021.[21] The details of analyzed genetic groups in this study are summarised in Table S3. We estimated nuclear contamination using *ANGSD v0.910*.[12] We estimated mitochondria contamination using *schmutzi*[43] and *contamMix 1.0-10*.[50]

Principal component analysis (PCA)

We carried out a principal components analysis using *smartpca* v16000 from the *eigensoft v7.2.1* package[44] with "lsproject: YES" and no shrink mode. The PCA is based on 68 western Eurasian populations in the Human Origins datasets (Table S3).[4,13,48,51] We projected our new high coverage Iceman genome and the previously published Iceman shotgun genome data from Keller et al. 2012[3] onto the PC space calculated from modern western Eurasians. We used DataGraph to visualize the PCA results.

Outgroup f_3 statistics

We examined the genetic affinity of the high coverage Iceman genome to previously published relevant ancient genetic groups and worldwide modern populations utilizing *outgroup* f_3(Iceman, Ancient populations; Mbuti.DG) *and outgroup* f_3(Iceman, Modern population X, Mbuti.DG) statistics. We used the *qp3Pop v435* programs in the *AdmixTools v5.1* package[47] for the f_3 statistics calculations. Modern population X and Mbuti.DG are from Simons Diversity Genome Project (SGDP).[52]

Please cite this article in press as: Wang et al., High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry, Cell Genomics (2023), https://doi.org/10.1016/j.xgen.2023.100377

qpAdm and qpWave analyses

We modeled the ancestry of the high coverage Iceman genome using *qpAdm v810* and *qpWave v410* in the *AdmixTools-5.1* package.[47] We tested various two-way combinations of WHG/Luxembourg_Loschbour.DG and Antolia_N/Germany_EN_LBK/Germany_EN_LBK_Stuttgart.DG[5,13,16,19,22,53,54] for Iceman and five contemporaneous ancient groups around the 4th millennium BCE from southern Europe (Italy_N.SG,[20] Italy_Broion_CA.SG,[21] Italy_Sardinia_C,[6,7] Italy_Sardinia_N,[6,7] Spain_MLN[5,16,18,55]). We used a list of 13 ancient populations as our default outgroup list, including *Ethiopia_4500BP_published.SG*[56], *Russia_Ust_Ishim.DG*[57], *Russia_Kostenki14*[58], *Belgium_UP_GoyetQ116_1_published_all*[58], *Czech_Vestonice16*[58], *Russia_MA1_HG.SG*[59], *Spain_ElMiron*[58], *Italy_North_Villabruna_HG*[58], *Iran_GanjDareh_N*[4,53], *Israel_Natufian_published*[4,53], *CHG*[60], *EHG*[5,58], *Levant_N*.[4] To replicate the modeling results of the previously published Iceman genome[3] in Haak et al. 2015,[17] we also applied the same 'O9' and 'O9N' outgroup list used in Haak et al. 2015,[17] and summarized corresponding modeling results in Table S5.

Local ancestry decomposition

To examine the ancestral tracts of specific ancestry, we decomposed the high-coverage Iceman genome using *RFMix*,[30] which identifies genomic chunks of contiguous ancestry of a given phased genome using a reference panel of phased haplotypes from two or three source populations. As *Rfmix* requires more than one phased genome for each source population as reference, we used four ancient genomes[13,19] as the WHG source, and 14 ancient LBK genomes[5,13,16,19] as the early Neolithic farmer source. Low coverage WHG and LBK genomes were imputed and phased with GLIMPSE[61] using 1000 Genomes as a reference panel. The high coverage genome Luxembourg_Loschbour.DG from Luxembourg (22x, 6221-5986 calBCE[62]) and Germany_EN_LBK_Stuttgart.DG from Germany (19x, 5307-5071 calBCE)[13] were called and phased following the same strategy as described for the high-coverage Iceman genome. We applied DATES[22] to estimate the admixture date in the Iceman using the same set of reference individuals, and estimated the expected average length of WHG ancestry tract by 1/admixture date * (1 – WHG ancestry proportion).

Effective population size estimates

We estimated effective population sizes using MSMC2[63] for the high coverage Iceman and a modern Sardinian genome,[64] as well as for Luxembourg_Loschbour.DG and Germany_EN_LBK_Stuttgart.DG that represent ancient European hunter-gatherer and early Neolithic farmer populations respectively. As MSMC2 requires high coverage genomes, we randomly chose a high coverage modern Sardinian genome (SS6004474) published in Pruefer et al. 2014,[64] together with the high-coverage ancient genomes Luxembourg_Loschbour.DG and Germany_EN_LBK_Stuttgart.DG. We followed the MSMC2 tutorial (https://github.com/stschiff/msmc-tools/blob/master/msmc-tutorial) for preparing individual masks and used VCF files with reliable diploid genotype calls from *snpAD*. For Sardinian SS6004474, we used masks as described in Supplementary section 5b of ref. [62] instead of individual masks. Together with negative universal masks (https://github.com/wangke16/MSMC-IM/tree/master/masks), we generated the required format of input files for MSMC2 using *generate_multihetsep.py*. We run MSMC2 using command line *msmc2 –I 0,1 –o outfile.msmc2 input.chr*.multihetsep.txt*.

Phenotypic SNP analyses

We first examined a list of 147 SNPs encoded for traits related to diabetes, male-pattern baldness, metabolic disorders, skin/hair/eye pigmentation and agriculturalist's diet etc. In particular, we examined the list of SNPs across the whole genome for Type 2 diabetes and male-pattern baldness phenotypic trait based on previous findings from GWAS studies. (https://www.ebi.ac.uk/gwas/api/search/downloads/studies_alternative). For each locus, we report the risk allele and called genotype in the Iceman genome, as summarized in Tables S9 and S10. We also examined five SNPs (FADS1/rs174546, SLC22A4/rs1050152, MCM6 (LCT)/rs4988235, NAT2/rs1495741, PRLP2/rs4751995), which are related to putative agricultural adaptations as summarized in Mathieson et al. 2018.[26] Meanwhile, we examined 170 skin-pigmentation related SNPs analyzed in Ju and Mathieson 2020,[27] and calculated the polygenic genetic score by a weighted sum for dark pigmentation phenotypic trait in the Iceman, Germany_EN_LBK_Stuttgart.DG, Luxembourg_Loschbour.DG and Sardinian SS6004474.

Estimates of heterozygosity

We estimate average heterozygosity levels for the Iceman, Germany_EN_LBK_Stuttgart.DG, Luxembourg_Loschbour.DG and a modern Sardinian individual by taking the fraction of the number of heterozygous sites over the total number of sites across 22 autosomes. We also estimated heterozygosity by applying a set of filters for ancient genomes, which filtered out regions of low mappability and sequence complexity, indels, and unusually high or low coverage after correcting for local GC content as described in Supplementary section 5b of ref. [62] We calculated error bars for our heterozygosity estimates via weighted jackknife following formulas provided here (https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/wjack.pdf).

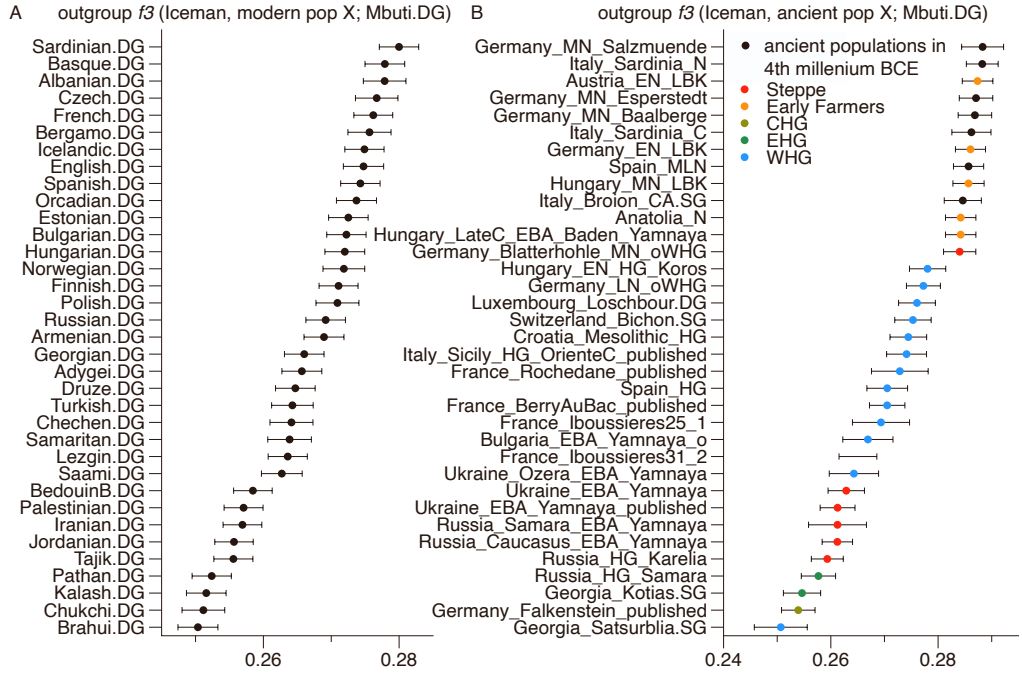
Cell Genomics, Volume *3*

Supplemental information

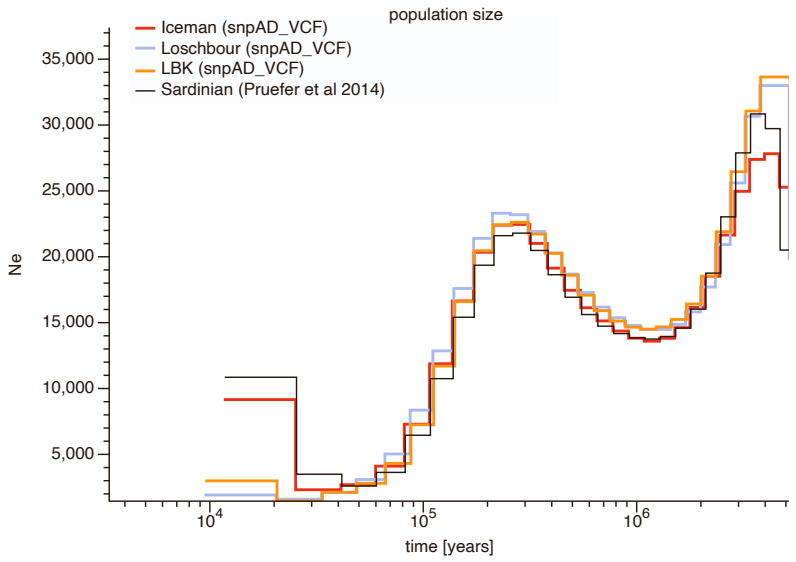
High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry

Ke Wang, Kay Prüfer, Ben Krause-Kyora, Ainash Childebayeva, Verena J. Schuenemann, Valentina Coia, Frank Maixner, Albert Zink, Stephan Schiffels, and Johannes Krause

- **Figure S1**. **Genetic affinity tests, related to STAR Methods**. Iceman shows close genetic affinity to modern Sardinians and ancient populations with high early Neolithic-Farmer-related ancestry in Europe. A. outgroup *f3*(Iceman, modern population X; Mbuti.DG). B. outgroup *f3*(Iceman, ancient population X; Mbuti.DG). In (A), we tested high coverage modern populations from SGDP dataset and reported the top 35 f_3-statistics with more 50,000 SNPs used in the tests. In (B), we tested ancient European groups dated to the 4th millennium BCE and populations related to three key ancient ancestry components and report the top 35 f_3-statistics here. The grouping details of ancient populations, as well as their related publications[65–68], are reported in Extended Table S2. We plot one standard error bar in *f*-statistics tests.



- **Figure S2**. **Effective population size estimates, related to STAR Methods**. We estimated effective population sizes for Iceman, Luxembourg_Loschbour.DG, Germany_EN_LBK_Stuttgart.DG, Sardinian using MSMC2. The estimate of effective population size of Iceman is very similar to Sardinian SS6004474. We find a slightly higher population size for Iceman and Germany_EN_LBK_Stuttgart.DG (both with high early European farmer ancestry), in comparison to the Luxembourg_Loschbour.DG hunter-gather in the recent time epoch, which is however difficult to confirm due to the limited resolution in recent times of MSMC2 method.



- **Figure S3. Heterozygosity distributions, related to STAR Methods.** Red square shows the heterozygosity calculated across the whole genome, together with red error bars calculated from weighted jacknife. Black dots represent the heterozygosity distribution calculated per autosome. One standard error bar is plotted here. The chromosome 21 of the Sardinian genome shows unusual high heterozygosity likely attributing to the uncertainty caused by the small chromosome size in which case limited data is available for heterozygosity calculation.

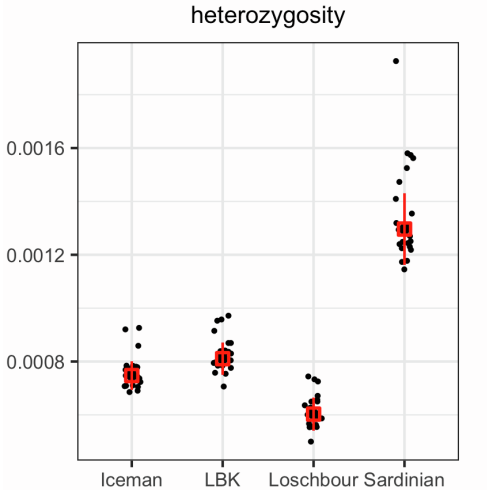


Table S2. Contamination estimates of the newly generated high-coverage Iceman genome and the previously published genome[3], related to STAR Methods. We employed *ANGSD xContam*[12], *schmutzi*[46] and *contamMix*[47] for testing contamination levels in both genomes. *ANGSD xContam* provides estimates on standard errors, while *schmutzi* reports 95% confidence interval and *contamMix* reports 97.5% confidence interval which are shown in brackets.

Method	Iceman		Iceman old genome (Keller et al 2012)	
	Contamination	Standard Error	Contamination	Standard Error
ANGSD (Method1: old_llh)	0.00449	2.27e-4	0.0745	1.35e-3
ANGSD (Method1: new_llh)	0.00449	2.28e-4	0.0743	1.35e-3
ANGSD (Method2: old_llh)	0.00524	5.96e-4	0.0756	2.48e-3
ANGSD (Method2: new_llh)	**0.00524**	**5.97e-4**	**0.0754**	**2.49e-3**

Schmutzi (5-95%)	0.00 (0.00-0.01)	0.02(0.01-0.03)
contamMix (2.5-97.5%)	0.003 (0.003-0.004)	0.035 (0.027-0.048)

- **Table S6**. **Cladality test using qpWave, related to STAR Methods**. We examined if a single source with early Neolithic-farmer-related ancestry is enough for explaining the genetic makeup of Iceman, using Anatolia_N/Germany_EN_LBK/ Germany_EN_LBK_Stuttgart.DG as the ancestral proxy.

pleft pop 1	pleft pop 2	p-value
Iceman	Anatolia_N	0.004825056
Iceman	Germany_EN_LBK	0.045266206
Iceman	Germany_EN_LBK_Stuttgart.DG	0.450214154

- **Table S7**. **Admixture time estimates for Iceman and relevant ancient European groups using *DATES*, related to STAR Methods**. Admixture time and standard error estimates are reported in generations here.

Target	Source1	Source2	Admixture Time	Str. Err.
Iceman	WHG	Anatolia_N	56.226	21.281
Iceman	WHG	Germany_EN_LBK	39.734	14.887
Iceman	WHG	Germany_EN_LBK_Stuttgart.DG	62.02	28.846
Spain_MLN	WHG	Anatolia_N	55.054	4.178
Spain_MLN	WHG	Germany_EN_LBK	50.345	3.925
Spain_MLN	WHG	Germany_EN_LBK_Stuttgart.DG	75.995	14.961
Italy_Sardinia_C	WHG	Anatolia_N	115.895	19.672
Italy_Sardinia_C	WHG	Germany_EN_LBK	109.342	32.863
Italy_Sardinia_C	WHG	Germany_EN_LBK_Stuttgart.DG	72.564	40.615
Italy_Sardinia_N	WHG	Anatolia_N	68.557	8.17
Italy_Sardinia_N	WHG	Germany_EN_LBK	55.069	7.487
Italy_Sardinia_N	WHG	Germany_EN_LBK_Stuttgart.DG	60.772	24.626
Italy_Broion_CA.SG	WHG	Anatolia_N	68.325	25.987
Italy_Broion_CA.SG	WHG	Germany_EN_LBK	108.376	78.349
Italy_Broion_CA.SG	WHG	Germany_EN_LBK_Stuttgart.DG	75.773	32.924
Germany_MN_Baalberge	WHG	Anatolia_N	33.223	8.844
Germany_MN_Baalberge	WHG	Germany_EN_LBK	19.425	17.408
Germany_MN_Baalberge	WHG	Germany_EN_LBK_Stuttgart.DG	115.315	20.247
Germany_MN_Salzmuende	WHG	Anatolia_N	81.619	47.86
Germany_MN_Salzmuende	WHG	Germany_EN_LBK	106.426	48.84
Germany_MN_Salzmuende	WHG	Germany_EN_LBK_Stuttgart.DG	117.324	45.268
Germany_MN_Esperstedt	WHG	Anatolia_N	26.189	22.908
Germany_MN_Esperstedt	WHG	Germany_EN_LBK	16.034	11.32
Germany_MN_Esperstedt	WHG	Germany_EN_LBK_Stuttgart.DG	92.13	75.32
Germany_LN_oWHG	WHG	Anatolia_N	17.253	8.514
Germany_LN_oWHG	WHG	Germany_EN_LBK	4.366	8.759
Germany_LN_oWHG	WHG	Germany_EN_LBK_Stuttgart.DG	-57.713	42.963

- **Table S8. Heterozygosity estimates, related to STAR Methods**. We estimate the heterozygosity using all 22 autosomes, and estimate standard errors using weighted jacknife calculation across 22 autosomes (see Methods) and examined heterozygosity with a set of filters (see Methods). We use individual SS6004474 as the representative for present-day Sardinians.

Individual ID	Heterozygosity	Heterozygosity after a set of filters
Iceman	0.000747346± 5.342197e-05	0.000745729±5.173587e-05
Germany_EN_LBK_Stuttgart.DG	0.000810571± 6.047406e-05	0.000807305±5.887978e-05
Luxembourg_Loschbour.DG	0.000602824±6.196018e-05	0.000589333±6.074216e-05
present-day Sardinian	0.001296318± 0.0001341288	0.000763694±4.234146e-05

- **Table S10. Summary of phenotype-informative SNPs in Iceman, related to STAR Methods**. We examined 147 SNPs related to phenotypic traits and summarized the phenotypic traits mentioned in the main text that are related to male-pattern baldness, obesity, type 2 diabetes, metabolic disorders, and pigmentation here. The full list of all examined SNPs is reported in Table S9.

rs ID	chr_pos	gene	ref_snp ADvcf	alt_snp ADvcf	effective allele	phenotypic trait associated with effective allele	Iceman genotype from snpAD
rs4988235	2_136608646	MCM6	G	.	A	lactase persistent	G/G
rs1050152	5_131676320	SLC22A4	C	T	T	Crohn' disease	missing
rs1495741	8_18272881	NAT2	G	A	G	NAT2 rapid metabolizer	A/A
rs4751995	10_118397884	PRLP2	A	G	G	digestion on plant membrane lipids	A/A
rs174546	11_61569830	FADS1	C	T	T	higher concentration of alpha-linolenic acids, and lower concentrations of eicosapentaenoic acids	C/T
rs17646946	1_152062767	TCHH	G	A	A	reduced hair curliness	G/A
rs3791783	2_190924163	MSTN	T	C	T	obesity-related	T/C
rs8192678	4_23815662	PPARGC1A	C	T	T	increased blood pressure, type 2 diabetes	T/T
rs35395	5_33948589	SLC45A2	T	C	C	light skin	C/C
rs16891982	5_33951693	SLC45A2	C	G	G	light skin	G/G
rs1042713	5_148206440	ADRB2	G	A	A	high total cholesterol, LDL cholesterol levels	G/A
rs4959270	6_457748	EXOC2	C	A	A	black hair color	C/A
rs1800795	7_22766645	LOC541472	C	.	C	obesity-related metabolic disorders	C/C
rs2153271	9_16864521	BNC2	C	T	C	decreased frecking	C/T
rs1815739	11_66328095	ACTN3	T	C	T	sprinter-type muscle performance	T/C
rs12913832	15_28365618	HERC2	A	.	A	brown eyes	A/A
rs1426654	15_48426484	SLC24A5	A	.	A	light skin	A/A
rs9939609	16_53820527	FTO	T	A	A	obesity risk and type 2 diabetes risk	A/A
rs2497938	X_66563018	SNPXBR102	T	.	T	male-pattern baldness	T/T

- **Table S11**. **Polygenic risk score of skin pigmentation, related to STAR Methods.** We estimated polygenetic risk scores for four ancient individuals calculated based on 170 UK BioBank SNPs (See also Table S9).

Individual ID	Polygenic Risk Score
Iceman	0.591405
Luxembourg_Loschbour.DG	0.675488
Germany_EN_LBK_Stuttgart.DG	0.620631
present-day Sardinian	0.589335

65. Nikitin, A.G., Stadler, P., Kotova, N., Teschler-Nicola, M., Price, T.D., Hoover, J., Kennett, D.J., Lazaridis, I., Rohland, N., Lipson, M., et al. (2019). Interactions between earliest Linearbandkeramik farmers and central European hunter gatherers at the dawn of European Neolithization. Sci. Rep. *9*, 19544.

66. Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. Nat. Commun. *5*, 5257.

67. Wang, C.-C., Reinhold, S., Kalmykov, A., Wissgott, A., Brandt, G., Jeong, C., Cheronet, O., Ferry, M., Harney, E., Keating, D., et al. (2019). Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. Nat. Commun. *10*, 590.

68. González-Fortes, G., Jones, E.R., Lightfoot, E., Bonsall, C., Lazar, C., Grandal-d'Anglade, A., Garralda, M.D., Drak, L., Siska, V., Simalcsik, A., et al. (2017). Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin. Curr. Biol. *27*, 1801-1810.e10.