**Supplemental Information**

# 40,000-Year-Old Individual from Asia Provides

# Insight into Early Population Structure in Eurasia

**Melinda A. Yang, Xing Gao, Christoph Theunert, Haowen Tong, Ayinuer Aximu-Petri, Birgit Nickel, Montgomery Slatkin, Matthias Meyer, Svante Pääbo, Janet Kelso, and Qiaomei Fu**

**A - Set 1**  **B - Set 2**  **C - Set 3**

**(i) m vs. ILSEI**

**(ii) m=0**

**(iii) m=1**

**(iv) m=2**

**(v) m=3**

**Figure S1. Treemix analysis for three sets of populations and individuals, related to Figure 2.** The three sets analyzed include (A) Set 1: the Tianyuan individual (TY), other ancient Eurasian individuals (UstIshim, Kostenki14, GoyetQ116-1, Vestonice16) and four present-day East and Southeast Asian populations (Han, Dai, Ami and Atayal); (B) Set 2: Set 1, the ancient Eurasian individual Malta1 and ancient Native American individual Anzick-1, and five present-day Native Americans (Mixe, Quechua, Surui, Karitiana and Chane), and (C) Set 3: Set 2 and a merged set of five ancient individuals from Nepal (AncTib). We ran the analysis for zero to ten migrations (m), and show the largest scaled residual magnitude for each set in row (i). The modeled tree and scaled residual plots for m=0 to m=3 are displayed in rows (ii) through (v). The African Mbuti population was used as an outgroup in all analyses. Set 1 used 559,802 SNPs, Set 2 used 406,040 SNPs and Set 3 used 231,462 SNPs. These results are for all sites. Bootstrap values are listed at the node if <100%, and the value indicates the fraction of replicates where the individuals to the right of the node group together. See also STAR Methods.
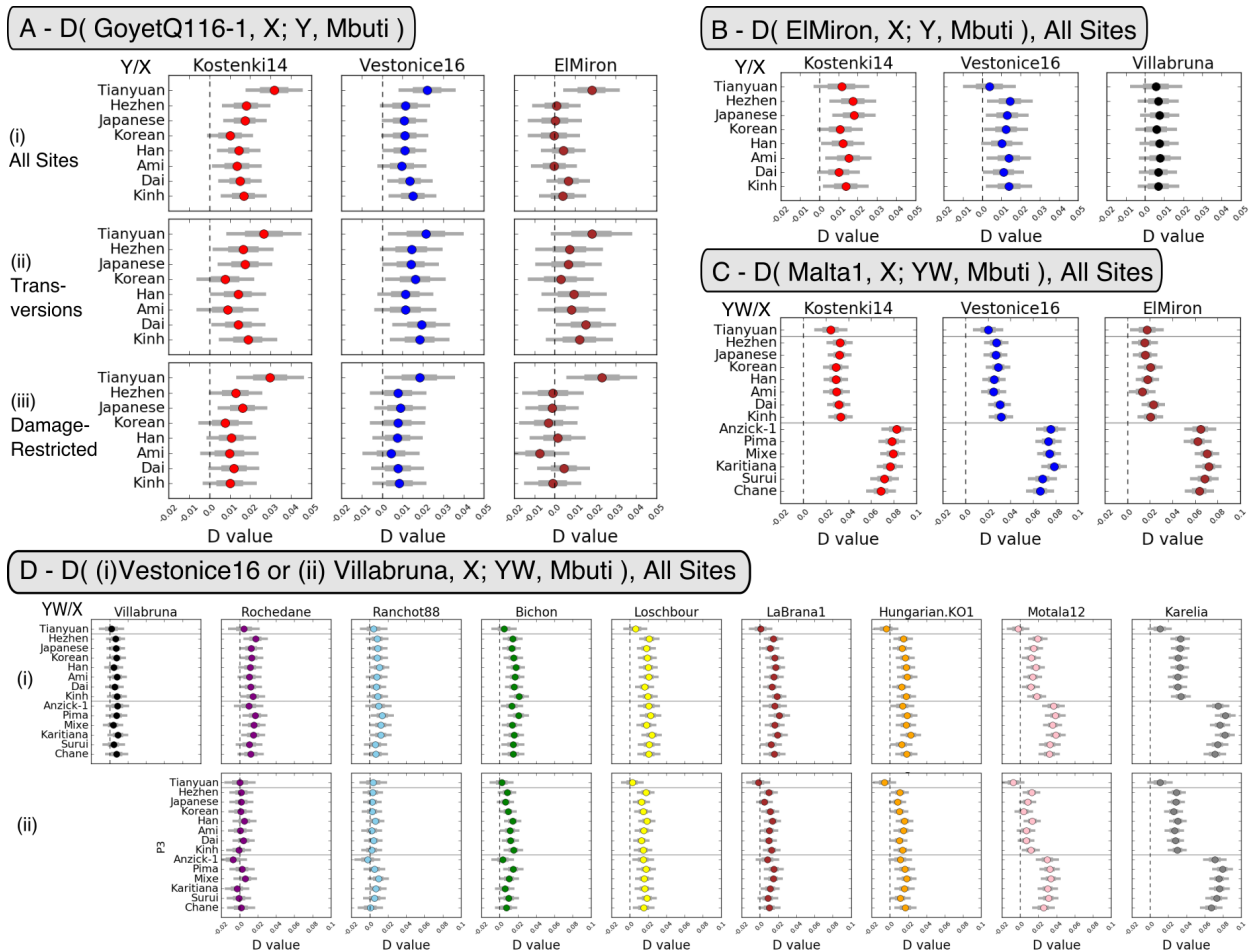
**Figure S2**. **D values and their 95% confidence intervals (thin whiskers) from D***(Ancient Eurasian, Ancient Eurasian; Tianyuan/East and Southeast Asian/Native American, Mbuti)***, related to Figure 2.** (A) focuses on comparisons for GoyetQ116-1 for (i) all sites, (ii) transversions only, and (iii) damage-restricted data for GoyetQ116-1. (B) focuses on ElMiron, (C) focuses on Malta1, and (D) highlights results for ancient West Eurasians <15 kya, as compared to the ancient individuals (i) Vestonice16 or (ii) Villabruna. Thick bars are within one standard error of the estimate, and thin bars are within 1.96 standard error of the estimate. See also Table S3 and STAR Methods.
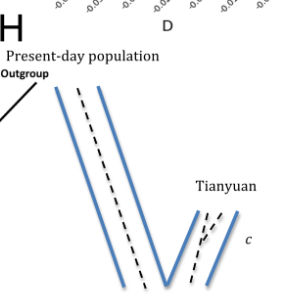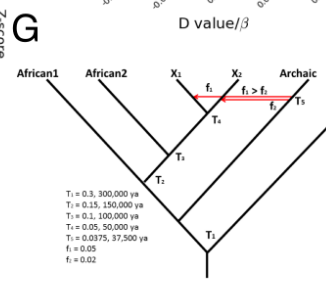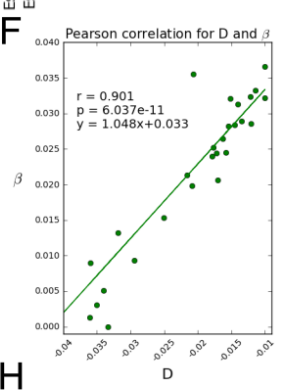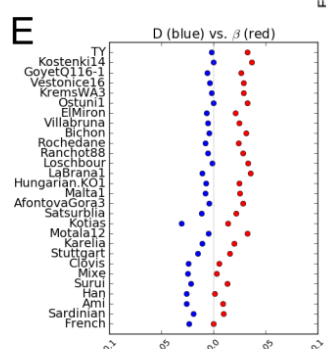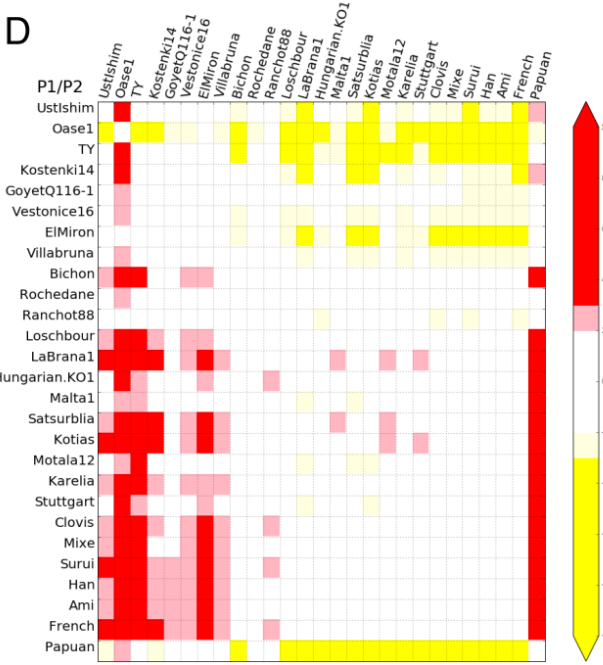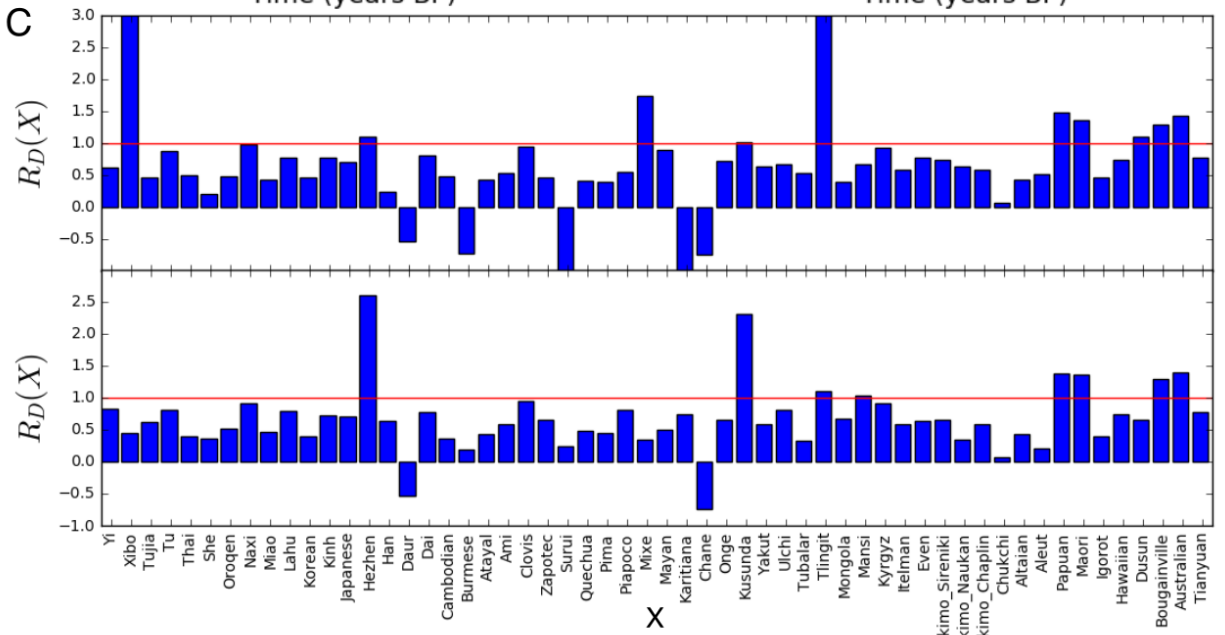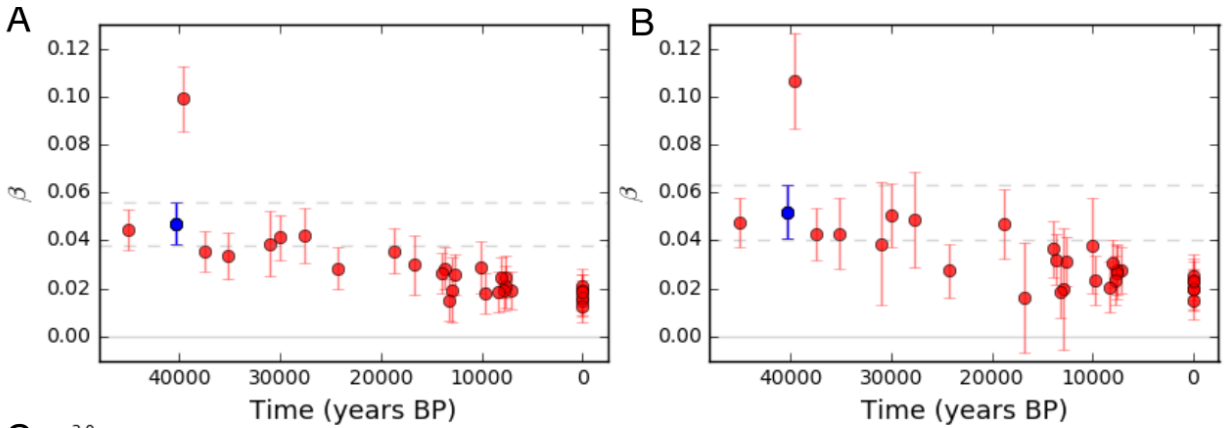
**Figure S3. Exploring the possibility of archaic admixture into the Tianyuan individual (A-G) and model for a test of direct ancestry (H), related to the STAR Methods and Table 1.**
(A-B) show the archaic admixture fraction with a 95% confidence interval for each ancient individual and present-day population plotted against the average date (yBP) of the individual or population using all sites (A) or transversions only (B). The blue point indicates the Tianyuan (TY) individual and dashed gray lines indicate the 95% confidence interval for the TY individual. (C) shows $R_D(X)$ using all sites, using the Mbuti as the African population and French as the West Eurasian population, and X as listed on the x-axis. Crossing the red line (y=1) indicates Denisovan admixture. For present-day populations, a randomly selected individual from the population is included (top) or all individuals per population are included (bottom). (D) shows Z-scores for *D(P1, P2; Mota, Chimp)* for transversions only, where Mota is a 4,000-year-old Ethiopian sample and P1 and P2 are non-Africans listed on the x- and y-axes. Red indicates *D>0*, such that P1 and Mota share alleles relative to P2, white indicates *D=0*, such that P1 and P2 share similar number of alleles to Africans, and yellow indicates *D<0*, such that P2 and Mota share alleles relative to P1. Similar results are found replacing Mota with present-day African populations and using all sites. (E) compares *D(Ust'-Ishim, X; Mbuti, Chimp)* (blue) and the admixture fraction β (red) from archaic hominins to non-Africans using transversions only and (F) shows the correlation from (E) between D and β. (G) describes the model simulated in *ms* [S1] and loosely adapted from SI16a.3 Model 3 in [S2]. Rather than directly simulating more ancient and more recent samples and including purifying selection on admixed regions, we construct a simpler model that achieves a similar difference in D-statistics, where $X_1$ receives higher amounts of Neanderthal admixture than $X_2$. We assume a recombination rate of $1.3 \times 10^{-8}$ per bp per generation, a mutation rate of $2.5 \times 10^{-8}$ per generation, 25 years per generation, and times of divergence and admixture as listed in the bottom left. We use θ=10 such that every replicate would generate approximately 150,000 SNPs. We simulate two (haploid) simulated samples per population to generate five diploid individuals. (H) describes the population tree model assumed for the test of direct ancestry described in the Methods. The dashed lines indicate the coalescence history at one SNP for an allele in the present-day population and the two alleles from the Tianyuan individual. *c* is the probability that the two alleles from the Tianyuan individual coalesce prior to the most recent common ancestral population (shown here) or they do not coalesce. See also Table S4 and STAR Methods.
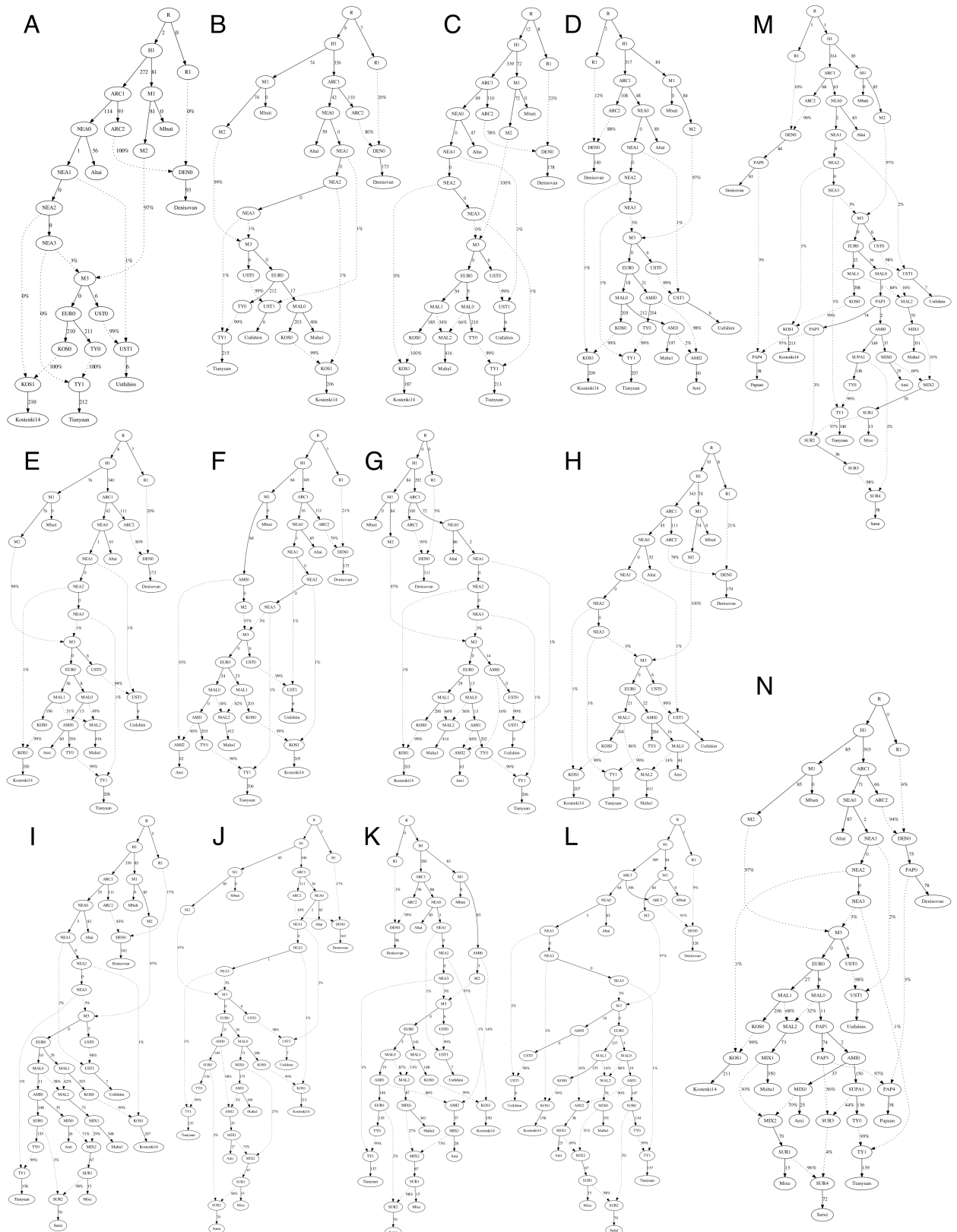
**Figure S4. Admixture graph depicting a potential model of relationships between the Tianyuan individual and other ancient Eurasians, and present-day non-African**

**populations, related to Figure 3.** The graph begins with (A) the base graph model adapted from Figure 3 of [S3], with UstIshim and the Tianyuan individual added, and subsequently adding Malta1 (B-C), the Ami (D-H), the Mixe and Surui (I-L) and the Papuan (M-N). The number of SNPs in each analysis is (A) 751,351; (B-C) 539,047; (D-H) 539,025; (I-L) 538,978, and (M-N) 538,964. |maxZ| is (A) 0.817, (B) 1.846, (C) 1.207, (D) 1.989, (E) 1.839, (F) 1.843, (G) 1.632, (H) 1.836, (I) 2.896, (J) 2.896, (K) 2.851, (L) 2.787, and (M-N) 2.896. See also Table S2i and STAR Methods.

## Supplemental References

S1.    Hudson, R.R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics *18*, 337-338.

S2.    Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science *338*, 222-226.

S3.    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.

S4.    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.

S5.    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.

S6.    Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature *524*, 216-219.

S7.    Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Paeaebo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proceedings of the National Academy of Sciences of the United States of America *110*, 2223-2227.

S8.    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*, 202-205.

S9.    Seguin-Orlando, A., Korneliussen, T.S., Sikora, M., Malaspinas, A.S., Manica, A., Moltke, I., Albrechtsen, A., Ko, A., Margaryan, A., Moiseyev, V., et al. (2014). Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. Science *346*, 1113-1118.

S10.   Raghavan, M., Steinruecken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Avila-Arcos, M.C., Malaspinas, A.-S., et al. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. Science *349*.

S11.   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. Nature Communications *6*, 8912.

S12. Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., DeGiorgio, M., Stafford, T.W., Jr., Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S.M., et al. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature *506*, 225-229.

S13. 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.

S14. 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.

S15. Olalde, I., Allentoft, M.E., Sánchez-Quinto, F., Santpere, G., Chiang, C.W., DeGiorgio, M., Prado-Martinez, J., Rodríguez, J.A., Rasmussen, S., Quilez, J., et al. (2014). Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature *507*, 225-228.

S16. 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.

S17. Siska, V., Jones, E.R., Jeon, S., Bhak, Y., Kim, H.M., Cho, Y.S., Kim, H., Lee, K., Veselovskaya, E., Balueva, T., et al. (2017). Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. Sci Adv *3*, e1601877.

S18. 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 throughout the African continent. Science *350*, 820-822.

S19. Jeong, C., Ozga, A.T., Witonsky, D.B., Malmström, H., Edlund, H., Hofman, C.A., Hagan, R.W., Jakobsson, M., Lewis, C.M., Aldenderfer, M.S., et al. (2016). Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. Proc Natl Acad Sci U S A *113*, 7485-7490.

S20. Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature *468*, 1053-1060.