Quantitating and Dating Recent Gene Flow between European and East Asian Populations

Pengfei Qin¹, Ying Zhou¹, Haiyi Lou¹, Dongsheng Lu¹, Xiong Yang¹, Yuchen Wang¹, Li Jin², Yeun-Jun Chung³ & Shuhua Xu¹.⁴.⁵

¹Chinese Academy of Sciences (CAS) Key Laboratory of Computational Biology, Max-Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China, ²Ministry of Education (MOE) Key Laboratory of Contemporary Anthropology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China, ³Integrated Research Center for Genome Polymorphism, Department of Microbiology, The Catholic University Medical College, Socho-gu Seoul 137-701, Korea, ⁴School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China, ⁵Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China.

Historical records indicate that extensive cultural, commercial and technological interaction occurred between European and Asian populations. What have been the biological consequences of these contacts in terms of gene flow? We systematically estimated gene flow between Eurasian groups using genome-wide polymorphisms from 34 populations representing Europeans, East Asians, and Central/South Asians. We identified recent gene flow between Europeans and Asians in most populations we studied, including East Asians and Northwestern Europeans, which are normally considered to be non-admixed populations. In addition we quantitatively estimated the extent of this gene flow using two statistical approaches, and dated admixture events based on admixture linkage disequilibrium. Our results indicate that most genetic admixtures occurred between 2,400 and 310 years ago and show the admixture proportions to be highly correlated with geographic locations, with the highest admixture proportions observed in Central Asia and the lowest in East Asia and Northwestern Europe. Interestingly, we observed a North-to-South decline of European gene flow in East Asians, suggesting a northern path of European gene flow diffusing into East Asian populations. Our findings contribute to an improved understanding of the history of human migration and the evolutionary mechanisms that have shaped the genetic structure of populations in Eurasia.

The interaction between European and Asian populations has been historically influenced by many factors, including the political conditions created by great empires such as the Roman (31 BC–250 AD) and Mongolian empires (1207 AD–1360 AD) that fostered communication between European and Asian populations, and the famous Silk Road (206 BC) which acted historically as a commercial, religious, and cultural network interlinking the trade routes across the Eurasian landmass that connected East Asia with the Mediterranean and Europe. This raises the question of whether genetic admixture occurred during the interaction between people from different regions, particularly between Europe and East Asia. If there has been gene flow between populations, we can, in principle, estimate the time and magnitude of this gene flow by analyzing genomic data from modern human populations. Previous studies have revealed that the populations of Central Asia, such as Uygur, are typically admixed with genetic contributions from both European and East Asian populations¹–⁷. However, only a few cases of genetic admixture among other Eurasians, particularly between East Asians and Northwestern Europeans have been reported⁸,⁹. These studies reported the Asian or American contribution in European populations. But the European gene flow in Asians has not been well studied yet.

Here we present what is, to the best of our knowledge, a systematic investigation of genetic admixture between different European and Asian populations using genome-wide single nucleotide polymorphism (SNP) data from 34 populations, including seven European groups (EUR), 9 Central or South Asian groups (CSA), and 17 East Asian groups (EAS) and one Tibetan group from Tibet Plateau. Based on statistical analysis using 4 Population Test (f₄-test) and 3 Population Test (f₃-test)¹₀,¹¹, we detected gene flow between Europeans and Asians were
Results
Evidence of Gene Flow between EUR and EAS. We obtained genome-wide data of 34 Eurasian populations from 1000 Genomes Project (1 KG)\textsuperscript{13}, the Human Genome Diversity Panel (HGDP)\textsuperscript{14} and other studies\textsuperscript{1,15}. In total, 1,132 Eurasian samples with 186,506 genome-wide data of 34 Eurasian populations from 1000 Genomes Project (1 KG)\textsuperscript{13}, the Human Genome Diversity Panel (HGDP)\textsuperscript{14} and other studies\textsuperscript{1,15}. In total, 1,132 Eurasian samples with 186,506 SNPs integrated from different technical platforms were used for analysis (see Methods). To study the signal of gene flow between West and East Eurasian populations (European and Asian populations in Table 1), firstly we performed Principal Component Analysis (PCA)\textsuperscript{16} both with and without respectively African populations included. We quantitatively estimated the magnitude of these gene flows using two different \(f_4\) statistical approaches: \(f_4\) Ratio Estimation\textsuperscript{12}, and Regression Ancestry Estimation\textsuperscript{10}. And finally, we estimated the time of admixture events based on their admixture linkage disequilibrium (ALD) using ROLLOFF software\textsuperscript{15}.

Table 1 | Testing for gene flow between EUR and EAS

| Population | Samples (after QC) | Region | Latitude | Longitude | \(f_{4,A}\) | \(f_{4,B}\) | \(f_3\) |
|------------|-------------------|--------|----------|-----------|-----------|-----------|--------|
| Dai        | 10 (10)           | East Asia | 21.4N | 101.0E   | -         | -         | -      |
| CHB        | 97 (97)           | East Asia | 39.9N | 116.5E   | 3.7       | -         | -      |
| CHS        | 100 (92)          | East Asia | 26.9N | 117.5E   | 2.3       | -         | -      |
| JPT        | 89 (89)           | East Asia | 35.9N | 138.6E   | 2.6       | -         | -      |
| Lahu       | 10 (10)           | East Asia | 24.3N | 99.4E    | 0.5       | -         | -      |
| Naxi       | 10 (9)            | East Asia | 27.7N | 98.3E    | 1.4       | -         | -      |
| She        | 10 (9)            | East Asia | 28.5N | 117.3E   | 0.9       | -         | -      |
| Yr         | 10 (10)           | East Asia | 28.6N | 100.5E   | 1.8       | -         | -      |
| Miao       | 10 (10)           | East Asia | 27.1N | 112.4E   | 1.9       | -         | -      |
| Tu         | 10 (10)           | East Asia | 34.1N | 102.3E   | 7.6       | -         | -      |
| Hezhen     | 10 (10)           | East Asia | 47.3N | 131.0E   | 6.2       | -         | -      |
| Mongolian  | 10 (10)           | East Asia | 44.6N | 114.9E   | 10.3      | -         | -      |
| Daur       | 10 (10)           | East Asia | 51.6N | 123.4E   | 7.3       | -         | -      |
| Oroqen     | 10 (10)           | East Asia | 35.9N | 116.5E   | 8.6       | -         | -      |
| Xibo       | 9 (9)             | East Asia | 45.0N | 121.9E   | 9.85      | -         | -      |
| Korean     | 100 (100)         | East Asia | 36.9N | 127.9E   | 2.55      | -         | -      |
| Tibetan    | 46 (46)           | Tibet Plateau | 31.1N | 88.7E    | 3.4       | -         | -      |
| Uygur      | 44 (42)           | Central Asia | 44.6N | 87.3E    | 31.1      | 50.7      | -60.0  |
| Hazara     | 25 (23)           | South Asia | 34.9N | 63.5E    | 33.5      | 57.0      | -56.6  |
| Pathan     | 25 (25)           | South Asia | 31.9N | 69.2E    | 33.8      | 18.5      | -21.8  |
| Burusha    | 25 (25)           | South Asia | 36.3N | 74.6E    | 30.9      | 29.8      | -23.8  |
| Makrani    | 25 (25)           | South Asia | 26.0N | 64.0E    | 19.6      | -12.7     | -1.8   |
| Balochi    | 24 (24)           | South Asia | 30.5N | 63.6E    | 27.9      | 2.1       | -6.9   |
| Brahui     | 25 (25)           | South Asia | 30.0N | 66.0E    | 26.3      | -0.6      | 0.4    |
| Kolash     | 23 (22)           | South Asia | 36.0N | 71.5E    | 33.2      | 14.2      | 39.1   |
| Sindhi     | 24 (24)           | South Asia | 25.5N | 59.0E    | 24.3      | 8.6       | -17.2  |
| French     | 29 (29)           | Europe   | 48.3N | 2.6E     | -         | -         | -      |
| CEU        | 82 (82)           | Europe   | 48.7N | 7.4E     | -         | 3.6       | -      |
| GBR        | 89 (85)           | Europe   | 51.8N | 0E       | -         | 5.0       | -      |
| FIN        | 93 (93)           | Europe   | 62.8N | 26.4E    | -         | -         | -      |
| Russian    | 25 (25)           | Europe   | 56.2N | 37.6E    | -         | 24.2      | -      |
| Adygei     | 17 (17)           | Europe   | 44.6N | 40E      | -         | 9.4       | -      |
| Orcadian   | 16 (15)           | Europe   | 59.0N | 3.1W     | -         | 4.1       | -      |

Note: \(f_{4,A}\) and \(f_{4,B}\) stand for Z-score of tests \(f_4(Yr,French,Dai,X)\) and \(f_4(Yr,French,Dai,X)\), respectively. \(f_3\) stands for Z-score of test \(f_3(K,French,Dai)\). Weighted Block Jackknife [block size of 5 cM] was used to correct LD among SNPs and estimate standard deviations. For these tests, we interpreted |Z-score| > 2 (bold) as significant evidence of admixture. A minus sign in table indicates that the test was not performed.

We assessed the gene flow between EUR and EAS using \(f_4\)-test and \(f_3\)-test for each Eurasian group. The \(f_4\)-test and \(f_3\)-test, which are model-based statistics, were designed to measure genetic drift along lineages quantitatively to detect gene flow between populations\textsuperscript{16,17,17}. In the current study, our samples did not cover all European regions. For example, Southern European populations were not included in the analysis because they may have been affected by a recent genetic contribution from populations with African ancestry. Several studies have revealed evidence of sub-Saharan African admixture among southern European populations but only rarely among northern European populations\textsuperscript{16,17,18}. Since ancestral populations predating admixture are unavailable, an alternative is to determine proper surrogates of ancestral EUR and EAS. We have employed an approach based on \(f\) statistics for each pair of EAS, EAS, and EUR, to compare the quantitative contribution they inherited from EUR (see Methods). We can therefore choose the least-admixed EAS as a surrogate of ancestral EAS. As a result, based on phylogeny as Supplementary Fig. S1, Dai from southern China could be considered to be a surrogate of ancestral EAS (Fig. 2a). Similarly, we applied our approach for each pair of EUR and found French holds the least genetic contribution from EAS than other EUR, and therefore can be

prevalent for most populations including even those from western and other studies\textsuperscript{16,17,17}. In the current study, our samples did not cover all European regions. For example, Southern European populations were not included in the analysis because they may have been affected by a recent genetic contribution from populations with African ancestry. Several studies have revealed evidence of sub-Saharan African admixture among southern European populations but only rarely among northern European populations\textsuperscript{16,17,18}. Since ancestral populations predating admixture are unavailable, an alternative is to determine proper surrogates of ancestral EUR and EAS. We have employed an approach based on \(f\) statistics for each pair of EAS, EAS, and EUR, to compare the quantitative contribution they inherited from EUR (see Methods). We can therefore choose the least-admixed EAS as a surrogate of ancestral EAS. As a result, based on phylogeny as Supplementary Fig. S1, Dai from southern China could be considered to be a surrogate of ancestral EAS (Fig. 2a). Similarly, we applied our approach for each pair of EUR and found French holds the least genetic contribution from EAS than other EUR, and therefore can be
Figure 1 | PC plots indicate potential gene flow between EUR and EAS. (a) PC plot of 1,219 samples from 35 populations that were clustered into EAS, EUR, CSA, and African. (b) Fine resolution of the PC plot after removing YRI. Both plots were based on 96,538 pruned SNPs to reduce linkage disequilibrium relationships.

considered as a surrogate of ancestral EUR (Fig. 2b). We assumed a Hybrid Isolation (HI) model for admixed Eurasian populations with ancestries from both EUR and EAS, as shown in Supplementary Fig. S1. For each group of EAS, we assessed whether the statistic $f_4$(YRI, French; Dai, CSA) significantly deviates from zero (see Methods). If so, it should indicate gene flow from EUR. Otherwise, the population should have entire EAS ancestry. Accordingly, we found that most EAS significantly deviated from the (YRI(French, (Dai, EAS))) topology ($Z$-score > 2) (Table 1 and Supplementary Table S1), indicating gene flow from EUR. We similarly tested EUR gene flow in each group of EUR by test $f_4$(YRI, Dai; French, EUR) (see Methods). Most EUR significantly violated the (YRI, ((French, EUR),Dai) topology ($Z$-score > 2) (Table 1 and Supplementary Table S1), indicating that most EUR in our study inherited a genetic contribution from EAS ancestry.

For CSA, we used two $f_4$-tests, $f_4$(YRI, French, Dai; CSA) and $f_4$(YRI, Dai; French, CSA), and one $f_4$-test $f_4$(CSA, French, Dai). Most CSA showed an extremely significant violation of the (YRI, ((French, (Dai, CSA))) topology (Table 1 and Supplementary Table S1), which indicates that these CSA (such as Uygur, Haraza, Pathan, Burusho, Kalash and Sindhi) are typical admixed populations with a high level of gene flow from both EUR and EAS ancestries. Different from other CSA, Makrani showed significant signal of African ancestry. And there was no admixture signal for Brahui.

Quantitative Estimation of Gene Flow. To estimate the genetic contributions of EUR and EAS ancestries, we first performed a linear regression procedure10, which assumes that mixed populations considered as a surrogate of ancestral EUR (Fig. 2b). We assumed a Hybrid Isolation (HI) model for admixed Eurasian populations with ancestries from both EUR and EAS, as shown in Supplementary Fig. S1. For each group of EAS, we assessed whether the statistic $f_4$(YRI, French; Dai, CSA) significantly deviates from zero (see Methods). If so, it should indicate gene flow from EUR. Otherwise, the population should have entire EAS ancestry. Accordingly, we found that most EAS significantly deviated from the (YRI(French, (Dai, EAS))) topology ($Z$-score > 2) (Table 1 and Supplementary Table S1), indicating gene flow from EUR. We similarly tested EUR gene flow in each group of EUR by test $f_4$(YRI, Dai; French, EUR) (see Methods). Most EUR significantly violated the (YRI, ((French, EUR),Dai) topology ($Z$-score > 2) (Table 1 and Supplementary Table S1), indicating that most EUR in our study inherited a genetic contribution from EAS ancestry.

For CSA, we used two $f_4$-tests, $f_4$(YRI, French, Dai; CSA) and $f_4$(YRI, Dai; French, CSA), and one $f_4$-test $f_4$(CSA, French, Dai). Most CSA showed an extremely significant violation of the (YRI, ((French, (Dai, CSA))) topology (Table 1 and Supplementary Table S1), which indicates that these CSA (such as Uygur, Haraza, Pathan, Burusho, Kalash and Sindhi) are typical admixed populations with a high level of gene flow from both EUR and EAS ancestries. Different from other CSA, Makrani showed significant signal of African ancestry. And there was no admixture signal for Brahui.

Quantitative Estimation of Gene Flow. To estimate the genetic contributions of EUR and EAS ancestries, we first performed a linear regression procedure10, which assumes that mixed populations share similar demographical histories. One Oceania population Papuan, which were not considered experiencing any recent admixture with EUR or EAS, was added into our model as an out-group11. For EAS with phylogeny displayed in Supplementary Fig. S2, we plotted $f_4$(YRI, French, Dai, EAS) against $f_4$(YRI, Papuan, French, EAS), and each group should fall along a line with a negative slope (equation (3)) if most groups shared same degree of drift $T$ and $W$. By carrying out the least-squares that best fit all EAS data, we could estimate the parameters of the linear model, allowing us to calculate the admixture proportion for each group. As expected, we observed a well-fit linear model for all EAS in our study, with a correlation coefficient $r^2 = 0.8$ (Supplementary Fig. S3a), which indicates that most EAS are likely to have inherited alleles from the same ancestral populations and shared similar demographical histories (same $T$ and $W$ in Supplementary Fig. S2). We subsequently calculated the admixture proportion given the values of drift $T$ and $W$ that could be estimated based on parameters of the linear model. An estimation of admixture proportions for each EAS is provided in Table 2. We detected $2.8 \pm 0.2\%$ EUR ancestry in Northern Han Chinese (CHB), which was more than that in Southern Han Chinese (CHS; $1.7 \pm 0.1\%$), Japanese ($2.2 \pm 0.2\%$) and Korean ($1.6 \pm 0.2\%$) populations. Northeast Asians such as Oroqen, Mongolian, Hezhen, and Daur (nomads who historically lived alongside Russians and Caucasians) inherited significantly more alleles from EUR: Mongolian $10.9 \pm 0.1\%$, Oroqen $9.6 \pm 0.2\%$, Daur $8.0 \pm 0.2\%$, and Hezhen $6.8 \pm 0.2\%$.

Similarly, we plotted $f_4$(YRI, Dai, French, EUR) against $f_4$(YRI, Papuan; EUR, Dai) for EUR, in which a linear model provided a good fit to the data ($r^2 = 0.9$) (Supplementary Fig. S3b), indicating most groups underwent the same amount of drift as measured by $W$ and $M$ (Supplementary Fig. S2). The proportions of admixture from ancestral EUR and EAS were estimated, and are shown in Table 2. CEU populations mostly originating from France and Germany had a small fraction ($0.7 \pm 0.8\%$) of genetic material from EAS. People from Great Britain such as British (GBR) and Orcadian inherited 2.5%–3.8% from ancestral EAS. Finnish (FIN) and Russians inherited significantly more genetic material (>12%) from ancestral EAS, which is consistent with their historical record of admixture with Mongolian populations. Besides, Adygei from Caucasus inherited $3.2 \pm 1.0\%$ from ancestral EAS.

This regression method was unsuitable for studying CSA, because there was no correlation between $f_4$(YRI, Dai; French, CSA) against $f_4$(YRI, Papuan; Dai, CSA) which is required in this approach. We therefore performed $F_2$ Ratio Estimation12 to calculate the admixture proportion for CSA. For each CSA (CSA), the admixture proportion could be estimated directly from the ratio of these two $f_4$-tests $f_4$(YRI, Papuan; CSA, Dai) and $f_4$(YRI, Papuan, French, Dai) (equation (4)). The results (Table 2) revealed that CSA, located in the middle of the Eurasian continent, mostly are typical admixed populations with a high level of gene flow from EUR and EAS ancestries. Uygur from Northwest China exhibit 52.4% EUR ancestry. For those populations from Pakistan, Pathan exhibits 78.4% EUR ancestry. Burusho and Hazara have 67.9% and 50.2% EUR ancestry, respectively. Admixture
proportions for some CSA are relatively small, for which Kalash has 20.7% EUR ancestry, Sindhi 11%, Balochi 2.1%.

We also performed F4 Ratio Estimation to estimate admixture proportion for EAS and EUR. Using this method, we obtained comparable results to the regression method (Supplementary Table S2). For example, we detected 9.7% EUR ancestry in the Mongolian population. EUR ancestry is more pronounced among Northern Han Chinese than southern Han Chinese. In addition, Russian and Finnish populations show 13% Asian ancestry.

Estimation of Admixture Times. To provide a more detailed description of demographic history, we attempted to estimate the time that admixture events occurred. Given the low level of gene flow for some Eurasian groups, we needed an accurate and sensitive method to estimate admixture times. We used ROLLOFF, which examines pairs of SNPs and assesses how ALD decreases with genetic distance in admixed populations. We used each admixed EAS using ROLLOFF with EAS (Dai) and EUR (French) ancestries. By fitting an exponential distribution to each run of ROLLOFF with least-squares, we obtained the number of generations since admixture. As shown in Table 2, Han Chinese (CHB) received recent genetic input from EUR approximately 48 ± 1.2 generations (1383 years) ago assuming a generation time of 29 years. Similarly, Korean populations received recent gene flow from EUR approximately 45 ± 3.9 generations ago. EUR gene flow reached Japanese populations 31.2 generations (1,763 years) ago, which probably followed the continental migration of Yayoi that began ~2300 years ago and continued for the next 1000 years. The admixture of EUR in Xibo population could be dated back to 10.7 ± 0.7 generations, or ~310 years ago. In the past, Xibo populations lived in Siberia and North-Eastern China. In the mid-18th century, part of the Xibo population migrated to North-Western China due to the policy of guarding the frontier by the government of the Qing Dynasty. This estimation of admixture time correlates accurately with this historical event. Interestingly, according to our estimation, Mongolians received their European ancestry 32.9 ± 0.9 generations, or 954 years ago. This is around the time of the expansion of Mongolians in military and politics in the 12th century, who in turn built the largest contiguous land empire in human history. However, we did not always observe the extent of ALD with genetic distance due to the extremely low level of gene flow for some groups, such as CEU and CHS.

Similar approaches were applied for both EUR and CSA. Two populations from Great Britain (GBR and Orcadian) shared a similar recent admixture time with EUR (82.8 ± 39.6 and 78.1 ± 8.6 generations, respectively), which were older than other EUR. Adyghe received their EAS ancestry 24.1 ± 1.2 generations, or 699 years ago. The gene flow from EAS to Finnish and Russian populations could be dated back 64.2 ± 1.1 and 45.2 ± 1.3 generations, or 1,862 and 1,311 years, respectively. Our estimation of admixture between EAS and EUR in CSA populations was 26.4 ± 0.5 generations for

![Figure 2](https://www.nature.com/scientificreports/)

**Figure 2** | Determining surrogates of ancestral populations. $f_4(YRI, EUR; EAS, EAS_i)$ and $f_4(YRI, EAS; EUR, EUR_i)$ were used to identify least-admixed populations as ancestral surrogates of (a) EAS and (b) EUR, respectively. Positive values in (a) indicate there is less EUR ancestry in EAS than in EAS_i, and positive values in (b) indicate there is less EAS ancestry in EUR than in EUR_i.
generations ago. It continuous gene flows, the estimation of occurred at different times. Based on the decay of the extended ALD graphic history of CSA, and multiple waves of admixture may have Balochi, respectively. However, we were aware of the complex demo-
frac{EUR ancestry in Kalash was derived from 61.8
2.3 generations for Burusho and Pathan, respectively. The small
waves of admixture happened.

Admixture in the Uygur. According to the historical records, ancestors of Uygur can be traced to ancient Chidi and Dingling populations living in this region in the 3rd century B.C. The Silk Road crossed Xinjiang which improved the communication between EUR and EAS can be traced back to 206 B.C. Admixture is likely to have taken place since then, or even earlier. However, our study and one previous study11 gave a very short estimation of about 26
ancestors of Uygur can be traced to ancient Chidi and Dingling
populations did not receive as much gene flow from EUR (1.6% for
ancestry than populations in the South, such as She, Tujia, Yi, Miao,
Northern Han Chinese, and Tu inherited much more European
ancestry than populations in the North such as Mongolian were likely to contact with people from Caucasus region or other European-like populations around them in history and intermediate the EUR gene flow into other EAS22. Our data reveal that populations in the North such as Mongolian, Xibo, Oroqen, Hezhen, Northern Han Chinese, and Tu inherited much more European ancestry than populations in the South, such as She, Tujia, Yi, Miao, and Southern Han Chinese. Exceptionally, Japanese and Korean populations did not receive as much gene flow from EUR (1.6% for Korean, 2.2% for Japanese) as other Northern EAS did, which is likely due to their isolated location in Northeast Asia.

Discussion
In this study, by analyzing genome-wide SNP data, we revealed that recent genetic admixture did occur and have been prevalent in Eurasia continent, notably, gene flows have been detected even between northern European and East Asian populations which are geographically far away from each other and generally considered as well-differentiated populations.

Table 2 | Estimation of admixture proportion and admixture time

| Population | Region  | Gene flow (EAS%) | Date of admixture (Generation) | Date of admixture (Year) |
|------------|---------|------------------|-------------------------------|--------------------------|
| CHB        | East Asia | 97.2 ± 0.2 | 47.7 ± 1.2 | 1383.3 ± 34.8 |
| CHS        | East Asia | 98.3 ± 0.1 | NULL | NULL |
| JPT        | East Asia | 97.8 ± 0.2 | 60.8 ± 31.2 | 1763.2 ± 904.8 |
| Tu         | East Asia | 92.5 ± 0.2 | 43.4 ± 1.3 | 1258.6 ± 37.7 |
| Hezhen     | East Asia | 93.2 ± 0.2 | 37.8 ± 7.1 | 1096.2 ± 205.9 |
| Mongolian  | East Asia | 89.1 ± 0.1 | 32.9 ± 0.9 | 954.1 ± 26.1 |
| Daur       | East Asia | 92 ± 0.2 | 54.5 ± 3.6 | 1580.5 ± 104.4 |
| Oроqen     | East Asia | 90.4 ± 0.2 | 41.8 ± 1.6 | 1212.2 ± 46.4 |
| Xibo       | East Asia | 89.7 ± 0.1 | 10.7 ± 0.7 | 310.3 ± 20.3 |
| Korean     | East Asia | 98.4 ± 0.2 | 45 ± 3.9 | 1305 ± 113.1 |
| Tibetan    | Tibet Plateau | 94.5 ± 0.1 | 47.8 ± 1.8 | 1386.2 ± 52.2 |
| Uygur      | Central Asia | 47.6 ± 0.4 | 26.4 ± 0.5 | 765.6 ± 14.5 |
| Hazara     | Central Asia | 49.8 ± 0.3 | 24.4 ± 0.2 | 707.6 ± 5.8 |
| Burusho    | South Asia | 32.1 ± 0.3 | 49.2 ± 1.1 | 1426.8 ± 31.9 |
| Pathan     | South Asia | 21.6 ± 0.3 | 51.2 ± 2.3 | 1484.8 ± 66.7 |
| Balochi    | South Asia | 97.9 ± 0.4 | 82.0 ± 4.1 | 2376.9 ± 118.9 |
| Kalash     | South Asia | 79.3 ± 0.3 | 61.8 ± 5.3 | 1792.8 ± 152.3 |
| Sindhi     | South Asia | 89 ± 0.4 | NULL | 2049.7 ± 89.1 |
| CEU        | Europe | 0.7 ± 0.8 | NULL | NULL |
| GBR        | Europe | 2.5 ± 1 | 82.8 ± 39.6 | 2401.2 ± 1148.4 |
| FIN        | Europe | 12.5 ± 0.9 | 64.2 ± 1.1 | 1861.8 ± 31.9 |
| Russian    | Europe | 12.4 ± 1 | 45.2 ± 1.3 | 1310.8 ± 37.7 |
| Adygei     | Europe | 3.2 ± 1 | 24.1 ± 1.2 | 698.9 ± 34.8 |
| Orcadian   | Europe | 3.8 ± 1 | 78.1 ± 8.6 | 2264.9 ± 249.4 |

Note: Admixture time was estimated with ROLLOFF assuming the putative ancestral populations were Dai and French. Standard errors were computed using Weighted Block Jackknife by removing one chromosome each of 22 times. We assume 29 years for each generation. NULL indicates that no obvious ALD decay was observed.
ancestors in previous study11. Our approach and simplified model in this study has also been shown to be robust using surrogates of real ancestral populations (of all populations, EUR only and EAS only) as the measure of ascertainment bias by using markers with minor allele frequency (MAF) >5% (Supplementary Fig. S5 and S6). The previous application of ROLLOFF in time estimations has also been shown to be robust using surrogates of real ancestors in previous studies11. Our approach and simplified model in admixture estimations will still be robust when the admixture history is complicated, such as multiple admixture events or pre-mixed ancestors. The estimation of admixture proportion would not be remarkable affected (simulations in Supplementary Fig. S7 and S8), while dating admixture mainly reflects the recent admixture events (simulations in Supplementary Fig. S4 and S9, Supplementary Table S3).

In addition, our results are consistent with some recent studies8,9, while we identified and estimated EUR admixture in EAS which was not well described in previous studies. We found signals of gene flow in many populations with reasonable interpretations, especially with respect to EUR admixture in EAS.

Our estimation of admixture time was based on the assumption of an HI model, which is a simplified model for the complex untraceable admixture reality. In realistic scenarios, most admixtures could be continuous or multiple waved. The real parameters of admixture such as proportions and time are usually incomprehensible. Even though some estimations here might be indecipherable to real history, we could estimate, by modeling admixture based on genetic diversity, LD and ancestral segmental distribution, many useful parameters such as the effective admixture proportion and effective admixture time which are helpful to further evolutionary medical studies. The algorithm (ROLLOFF) we applied to date admixture in our study is LD-based, one tricky issue for such algorithm is that we could not precisely distinguish admixture LD from the background LD, especially when the admixture level is extremely low. That’s why ROLLOFF lose power for dating admixture in some populations with low level of gene flow.

Taken together, although fine-scale dissection of the demographic history of human populations in Eurasia and precise estimation of evolutionary parameters need improved methods and data, our current study provided an overall picture of subsequent genetic interaction among well-differentiated populations. Our results advanced our understanding of the history of human migration and the evolutionary mechanisms that have shaped the genetic structure of populations in Eurasia.

**Methods**

**Population Samples and Data.** Samples of 1,256 individuals in total from 34 Eurasian populations, one African population Yoruban (YRI) and one Oceanian population Papuan were obtained in this study. Genome-wide SNP data were acquired from both public datasets (1000 Genomes Project (1 KG)13, the Human Genome Diversity Panel (HGDP)14) and other studies in which there are 100 Korean samples (South Korea), 46 Tibetans (Tibet15) and 44 Uygurs (Xinjiang, China1), genotyped by Affy6.0. Details of samples were listed in Table 1. The combined data was further filtered to exclude individuals with >10% missing genotypes, and SNPs with missing rate >10% as well as those exhibiting Hardy-Weinberg disequilibrium (p <0.001). Pairwise kinship coefficients were estimated in each population24. According to the relationship inference criteria, individuals in the first (e.g. full-sibs), second (e.g. half-sibs) and third (e.g. first cousins) degrees of relationship were removed from our study. At last, we obtained 1,132 Eurasian samples with 186,506 SNPs after integration and quality control.

**Principle component analysis.** Principle component analysis (PCA) was performed with EIGENSOFT16 version 5.0.1 based on 96,538 pruned SNPs, which were randomly selected from 186,506 autosomal SNPs of all merged samples with interval distance larger than 10 kb to avoid high linkage disequilibrium.

**Modeling Admixture between EUR and EAS.** We assumed that admixture events between EUR and EAS occurred following the HI model (Supplementary Fig. S1). Since ancestral populations predating admixture are unavailable, an alternative is to determine proper surrogates of ancestral EUR and EAS. An approach was employed to select the least-admixed EAS and EUR exhibiting minimal gene flow. To choose proper surrogates for ancestral EUR, we applied a set of f4-tests to the proposed relationship (YRI, (EUR, (EAS, EAS))), for each pair of EUR, EAS and EUR(x) under consideration.

\[
f_{4}(YRI, EUR_{anc}; EAS_{x}, EAS) = p_{x} \times (1 - p_{x}) \times T + p_{x} \times (1 - p_{x}) \times T = (p_{x} - p_{y}) \times T \tag{1}
\]

where EUR_{anc} represents the ancestral populations of EUR, p_{x} and p_{y} is the EUR contribution to EAS_{x} and EAS, respectively. T is the quantity of drift depicted in Supplementary Fig. S1. Equation (1), (2), (3) and (4) were derived according to the
The expectation of $f$ statistics is mathematically appropriate to the drift paths in the admixture graph. A positive value means that EUR contribution to EAS is smaller than to EAS. We can therefore easily choose the least-admixed EAS as a surrogate of ancestral EAS. For EUR, we applied similar $f_4$-tests to the proposed relationship ($YRI, (EUR_i, EUR_j), EAS_i)$ for each pair of EUR (EUR_i, and EUR_j).

\[
f_4(YRI, EUR_{anc}; EUR_i, EUR_j) = p_i \times (1 - p_i) \times W - p_j \times (1 - p_j) \times W
\]

where $EAS_{anc}$ represents the ancestral populations of EAS, $p_i$ and $p_j$ is the EUR contribution to EUR, and $W$ is the quantity of drift depicted in Supplementary Fig. S1. A population consistently exhibiting positive values of the $f_4$-test was treated as the surrogate of the ancestral EUR.

**Statistical Analysis to Detect Gene Flow between EUR and EAS.** We applied two model-based statistics (the $f_3$-test and the $f_4$-test) to detect gene flow between Eurasian populations. Weighted Block Jackknife\(^2\),\(^3\), which drops 5 centimorgan (cM) blocks\(^6\) of the genome in each run, was used to compute a standard error of the statistic. We tested for admixture by assessing whether the statistic is more than 2 standard deviations from zero.

For EAS, we assessed whether the statistics $f_4(YRI, EUR_{anc}; EUR_i, EAS)$, whose expected value should be proportional to $p_i \times T$, were significantly different from zero (Supplementary Fig. S1). A Z-score deviating from zero ($\geq 2$) could be considered as evidence of gene flow from EUR. Similarly, we examined the value of $f_4(YRI, EUR_{anc}; EUR_{anc}, EUR_i)$ for EUR, which is expected to equal the product of the EUR contributions and their shared drift ($1 - p_i \times W$) for any EUR (Supplementary Fig. S1). In this case, a Z-score $\geq 2$ could be considered as evidence of gene flow from EUR. In contrast to EAS and EUR, for each CSA (CSA), we used two $f_4$-tests, i.e. $f_4(YRI, EUR_{anc}; EUR_{anc}, CSA)$ and $f_4(YRI, EUR_{anc}; EUR_{anc}, CSA)$, and one $f_4$-test, i.e. $f_4(CSA, EUR_{anc}; EUR_{anc})$. Z-scores deviating from zero provided evidence of admixture.

**Quantitative Estimation of the Gene Flow between EUR and EAS.** We used two different $f_3$ based approaches, Regression Ancestry Estimation\(^7\) and $F_4$ Ratio Estimation\(^7\),\(^8\), to estimate the extent of gene flow between EUR and EAS.

**Regression Ancestry Estimation.** We investigated a linear regression model based on two sets of $f_4$-test values. Taking EUR as an example, the expected value of $f_4(YRI, EUR_{anc}; EUR_{anc}, EAS)$ is proportional to $p_i \times T$. We designed an additional $f_3$-test, i.e. $f_3(YRI, EUR_{anc}; EUR_{anc}, EAS)$, and the expected value is proportional to $(1 - p_i) \times W$, where $W$ is a measurement of genetic drift, as described in Supplementary Fig. S2. Notably, if all EAS were derived from the same admixture event, genetic drift as measured by $T$ and $W$ should be:

\[
S = 4 \times p_i \times (1 - p_i) \times L \times T
\]

Where $S$ is the number of segments with one allele from ancestry A and the other allele from ancestry B (as shown in Supplementary Fig. S10), $p_i$ is contribution from ancestry A, $L$ is chromosome length in Morgans and $T$ is time in generations since admixture.

**Simulations.** Both backward time and forward time simulations were conducted in this study for different purposes. To evaluate the potential influence on our results of ascertainment bias, using surrogates instead of real ancestors and pre-mixed ancestors in the estimations of admixture, we performed coalescent simulations implemented in ms\(^{25}\). We used the same parameters of demographic history as those in a previous study\(^8\). Simulation study of each scenario was repeated one hundred times. Command lines were listed in Supplementary Fig. S5 and Supplementary Fig. S7.

The influence of ascertainment bias was evaluated using SNPs from three different ascertainment sets, which are based on minor allele frequency (MAF) of all simulated populations, EUR only and EAS only. We compared the estimations using all alleles and common alleles only (MAF$>5\%$).

Forward time simulations were conducted to generate Eurasian admixed (Uygur-like) haplotypes with ancestries from ancestral surrogates (e.g. CEU and CHB), with genotype data downloaded from HapMap 3\(^{26}\) and phased by Beagle v3.3.2\(^{27}\). Ancestral segments with recombination breakpoints information in Uygyor genomes were recorded during simulations. Simulated admixture scenarios were listed in Supplementary Table S3. Forward time simulations were also conducted to evaluate the influence of pre-mixed ancestries on estimation of admixture time (Supplementary Fig. S9).

1. Xu, S., Huang, W., Qian, J. & Jin, L. Analysis of genomic admixture in Uygur and its implication in mapping strategy. *Am. J. Hum. Genet.* 84, 883–894 (2009).
2. Xu, S. & Jin, L. A genome-wide analysis of admixture in Uyghurs and a high-density admixture map for disease-gene discovery. *Am. J. Hum. Genet.* 83, 322–336 (2008).
3. Xu, S., Jin, W. & Jin, L. Haplotype-sharing analysis showing uygurs are unlikely genetic donors. *Mol. Biol. Evol.* 26, 2197–2206 (2009).
4. Comas, D. et al. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am. J. Hum. Genet.* 63, 1824–38 (1998).
5. Tai, S. P. et al. [HLA-DRB alleles polymorphism in Han, Hui, Uyghur and Tibetan populations in northwestern China]. Yi Chu Xue Bao 26, 447–57 (1999).
6. Su, Y. et al. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am. J. Hum. Genet.* 65, 1718–24 (1999).
7. Yao, Y.-G., Kong, Q.-P., Wang, C.-Y., Zhu, C.-L. & Zhang, Y.-P. Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in china. *Mol. Biol. Evol.* 21, 2265–80 (2004).
8. Hellenthal, G. et al. A genetic atlas of human admixture history. *Science* 343, 747–751 (2014).
9. Lipson, M. et al. Efficient moment-based inference of admixture parameters and sources of gene flow. *Mol. Biol. Evol.* 30, 1788–802 (2013).
10. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* 461, 489–494 (2009).
11. Patterson, N. et al. Ancient admixture in human history. *Genetics* 192, 1065–93 (2012).
12. Moorjani, P., Patterson, N. & Hirschhorn, J. The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet.* 7, e1001373, doi:10.1371/ journal.pgen.1001373 (2011).
13. Abecasis, G. R. et al. A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–73 (2010).
14. Li, J. Z. et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319, 1100–4 (2008).
15. Xu, S. et al. A genome-wide search for signals of high-altitude adaptation in tibetans research article. 28, 1093–1101 (2011).
16. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* 2, e190, doi:10.1371/journal.pgen.0020190 (2006).
17. Auton, A. et al. Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res.* 19, 795–803 (2009).
18. Ruvolo, G. M. et al. Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *Am. J. Hum. Genet.* 72, 1058–64 (2003).
19. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* 128, 415–23 (2005).
Acknowledgments

We thank LetPub and Science Craft for their linguistic assistance during the preparation of this manuscript and Dr. Wenfei Jin for his discussion on the data analysis. These studies were supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (CAS) (XDB13040100), by the National Science Foundation of China (NSFC) grants (91331204, 31171218). S.X. is Max-Planck Independent Research Group Leader and member of CAS Youth Innovation Promotion Association. S.X. also gratefully acknowledges the support of the National Program for Top-notch Young Innovative Talents of The “Wanren Jihua” Project.

Author contributions

S.X. conceived and designed the study. L.J. and Y.C. contributed to population samples and genotyping data. P.Q. performed data analysis, with contribution from Y.Z., H.L., D.L., X.Y. and Y.W. S.X. and P.Q. wrote the paper.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports
Competing financial interests: The authors declare no competing financial interests.
How to cite this article: Qin, P. et al. Quantitating and Dating Recent Gene Flow between European and East Asian Populations. Sci. Rep. 5, 9500; DOI:10.1038/srep09500 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/