The origin of Aboriginal Australians as inferred from the genomic sequence of a hundred-year-old lock of hair

Morten Rasmussen*[1,2], Xiaosen Guo*[2,3], Yong Wang*[4], Simon Rasmussen[5], Anders Albrechtsen[6], Line Skotte[6], Stinus Lindgreen[6], Kirk E. Lohmueller[4], Mait Metspalu[7], Thibaut Jombart[8], Toomas Kivisild[9], Weiwei Zhai[10], Ludovic Orlando[1], Ene Metspalu[7], Kasper Nielsen[5], María C. Ávila-Arcos[1], J. Víctor Moreno-Mayar[1,11], M. Thomas P. Gilbert[1,2], Ole Lund[5], Agata Wesolowska[5], Monika Karmin[7] , Lucy A. Weinert[8], , Bo Wang[3], Jun Li[3], Shuaishuai Tai[3], Fei Xiao[3], Tsunehiko Hanihara[12], Andrea Manica[13], George van Driem[14], Aashish R. Jha[15], François-Xavier Ricaut[16], Peter de Knijff[17], Andrea B. Migliano[18], Irene Gallego-Romero[15], Karsten Kristiansen[2,3,6], Søren Brunak[5,23], Peter Forster[19], Bernd Brinkmann[20], Olaf Nehlich[21], Michael Richards[21,22], Ramneek Gupta[5], Anders Krogh[6], Robert Foley[9], Marta M. Lahr[9], Francois Balloux[8], Thomas Sicheritz-Pontén[5,23], Richard Villems[7,24], Rasmus Nielsen[4,6], Wang Jun[2,3,6], Eske Willerslev[1,2]

Correspondence to: Eske Willerslev[1,2], Wang Jun[2,3,6] Correspondence and requests for materials should be addressed to E.W. (Email: ewillerslev@snm.ku.dk) or J.W. (Email: wangj@genomics.org.cn).

1. Centre for GeoGenetics, Natural History Museum of Denmark & Department of Biology, University of Copenhagen, Øster Voldgade 5-7, DK-1350, Denmark.
2. Sino-Danish Genomics Center, BGI-Shenzhen, Shenzhen, 518083, China & University of Copenhagen, Denmark.
3. BGI-Shenzhen, Shenzhen, 518083, China

4. Departments of Integrative Biology and Statistics, UC-Berkeley, 4098 VLSB, Berkeley, CA 94720, USA
5. Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark
6. Department of Biology, University of Copenhagen, Ole Maaloes Vej 5, DK-2200, Denmark.
7. Department of Evolutionary Biology, Tartu University and Estonian Biocentre, 23 Riia Street, 510101 Tartu, Estonia
8. MRC Centre for Outbreaks, Department of Infectious Disease Epidemiology, Imperial College Faculty of Medicine, St Mary's Campus, Norfolk Place, London W2 1PG, UK
9. Leverhulme Centre for Human Evolutionary Studies, Department of Biological Anthropology, Henry Wellcome Building, Fitzwilliam Street, University of Cambridge, Cambridge, CB2 1QH, UK
10. Beijing Institute of Genomics, Chinese Academy of Sciences, No.7 Beitucheng West Road, Chaoyang District, Beijing, 100029, China
11. Undergraduate Program on Genomic Sciences. National Autonomous University of Mexico. Avenida Universidad s/n Chamilpa 62210, Cuernavaca, Morelos, Mexico.
12. Department of Anatomy, Kitasato University School of Medicine, 1-15-1 Kitasato, Minami-ku, Sagamihara, 252-0374, Japan
13. Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK
14. Institut für Sprachwissenschaft, Universität Bern, Länggassstrasse 49, 3000 Bern 9, Switzerland
15. Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA

16. Laboratoire d´Anthropologie Moléculaire et Imagerie de Synthèse (AMIS) Université de Toulouse (Paul Sabatier) - CNRS UMR 5288, 37 allées Jules Guesdes, 31073 Toulouse cedex 3, France
17. Department of Human and Clinical Genetics, Postzone S5-P, Leiden University Medical Centre, Albinusdreef 2, Leiden, The Netherlands
18. Department of Anthropology, University College London, Gower Street - London - WC1E 6BT, UK
19. Murray Edwards College, ¼University of Cambridge,¼ CB3 0DF, UK
20. FG (Institute for Forensic Genetics), D-48161 Münster, Germany
21. Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig, Germany
22. Department of Anthropology, University of British Columbia, 6303 NW Marine Drive, Vancouver, BC, Canada, V6T 1Z1
23. Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3A, DK-2200 Copenhagen, Denmark
24. Estonian Academy of Sciences, 6 Kohtu Street, 10130 Tallinn, Estonia

We present the first Aboriginal Australian genomic sequence, obtained from a lock of hair donated by an Aboriginal man living in South-Western Australia in the early 20th century. The genome was sequenced to an average depth of 11x, showing no evidence of European admixture, and an estimated contamination level of < 0.5%. It represents the first high-depth ancient genome sequence from outside the permafrost regions. Using population genetic approaches based on whole genome data, we find evidence that: i) Eurasians and Aboriginal Australians diverged from Africans at the same time; ii) Aboriginal Australians were the first to separate from this ancestral non-African gene pool some 50-60 thousand years ago, 10-25 thousand years prior to the split between Europeans and Asians; iii) Aboriginal Australians received substantial gene flow from Asians at a later date, suggesting that the Neanderthal genetic signal in the Aboriginal Australian genome, similar to that found in Eurasians, may be secondarily derived. Our findings imply that current views on the population history of Aboriginal Australian based on uniparental markers and SNP chip data are over simplistic. As such our understanding of modern human evolutionary history will likely benefit from being revisited using ancient and modern human genomics.

The evolutionary history of Aboriginal Australians remains debated. It is generally believed that a single out-of-Africa dispersal gave rise to all contemporary non-Africans sometime before 50 thousand years ago (50 ka BP)[1]. Genetic data have so far supported this view, pointing to Aboriginal Australians having split from Eurasian populations approximately 40-70 ka BP[2-4]. However, an alternative hypothesis proposes an additional earlier dispersal of anatomically moderns humans (AMH) from Africa some 60-75 ka BP[5, 6]. According to this model, the descendants of the earlier expansion became admixed or were replaced by later dispersing populations with the exception of a few groups, among whom would be Aboriginal Australians[6, 7].

Although critical to understanding the early dispersal of AMH outside Africa, only a few genetic studies have been conducted on Aboriginal Australians to date. These are restricted to uniparental markers and nuclear SNPs, and point to a shared ancestry with other non-Africans[2-4, 8]. They further reveal extensive recent admixture with immigrants of European descent[8, 9]. Interest in Aboriginal Australian genetics is not just a matter of history, as present-day communities often suffer from poorer health and higher rates of chronic disease than other Australians[10]. A non-admixed Aboriginal Australian genomic sequence should offer not only new insights into the evolution of human diversity outside Africa, but also provide a reference for future genome-wide disease association studies that might improve our understanding of the increased disease susceptibility and risk in this population.

Such a genome could be reconstructed from contemporary samples collected in isolated populations from, for example, northern Australia[4]; alternatively, it could be obtained from ancient samples collected closer to first contact with European immigrants. However, genomic sequencing of ancient human remains is difficult due to contamination by recent human DNA and massive microbial loads[11-13]. The use of ancient human hair overcomes most of these problems as it can be efficiently de-contaminated[14]. The disadvantage is that hair generally contains far less DNA than bone or teeth. Consequently, high-depth ancient genomic data have only been reported so far from human hair preserved in ideal frozen conditions[15]. Nevertheless, hair constitutes an important resource in many museum collections and could hold the key to obtaining sequence data from historic populations. Together with the inherent importance of describing the genome of an individual from a traditional, hunter-gatherer population[16], the challenge of reconstructing an ancient human genome from non-permafrost areas was a further motivation for this study.

DNA damage, contamination, and admixture estimates

We used 0.6 grams of human hair from the Duckworth Collection, Cambridge, UK, given by a young Aboriginal Australian male, and entered into the collection by the anthropologist Dr. Alfred C. Haddon in the early 20th century. The records show that the individual lived in Golden Ridge (Golden Station) in South-Western Australia, an area where two Aboriginal languages were spoken until recently - KALAAKO and MADUWONGGA; both languages are reported as extinct today[17].

We applied the same experimental procedure to the historical ancient hair as we used for the first ancient human genome of an extinct Palaeo-Eskimo[15], with minor differences (Supplementary Information, section 2; SI2), and sequenced the genome to an average depth of 11x. Despite its relatively young age, the DNA in the hair showed a high degree of fragmentation, with an average length of 69-bp. Cytosine to guanine and guanine to adenine mis-incorporation levels typical of ancient DNA[18] are low (maximum 3% per base), and cluster within five nucleotides at both read ends (Table 1). Such regions were trimmed in order to improve SNP call quality (SI3, 9). Isotopic dietary and mobility analyses of the hair further suggest good preservation, and point to a diet with a large intake of terrestrial protein and little seasonal variation (SI1).

The genome was assembled and genotyped using BWA[19] and the SAM tools suite[20]. This gave us a total of 2,782,401 SNPs of which 449,115 are considered high confidence and were used in further analyses (SI5). Of these, 28,395 (6.3%) have not been previously reported (Table 1). Despite extensive handling of the hair by people of European ancestry, contamination levels are estimated to be < 0.5% based on analysis of the heterozygosity on the X chromosome. Similarly, no European input could be detected at the genotype level (SI10).

Uniparental markers

The Australian individual's mitochondrial genome (mtDNA) was sequenced to an average depth of 338x, and we found it to belong to a new sub-clade of haplogroup O (hg O) that we term hg O1a (SI11.1). Haplogroup O is one of the four major lineage groups specific to Australia, and has been reported from the Northern Territories (15%) and the Warlpiri Desert (16%)[2-4]. Hg O derives from Eurasian founder hg N, one of the three founder haplogroups associated with the dispersal from Africa 50-65 ka BP[21]. Similarly, based on 188 high-quality Y-chromosome SNPs, we assigned the aboriginal Y-chromosome to the MNOPS macro-haplogroup (SI11.2). While the O and P members of this para-clade account for the majority of East and West Eurasian Y chromosomes, the unresolved MNOPS* lineages are common (>5%) only among contemporary populations of Australasia (Island Southeast Asia, Australia, New Guinea and Island Melanesia)[4, 22]. Both uniparental markers fall within the known pattern found among contemporary Aboriginal Australians[4], and are compatible with a common African ancestral population for all Eurasians and Aboriginal Australians[23]. However, uniparental markers do not allow for probabilistic testing of the different hypotheses of origins. This is not the case with genomic variation, since genomic polymorphisms can be treated as independent replicates, which, on average, should reflect past demography.

Comparison to worldwide genetic diversity

Among the high-confidence SNPs, 84,304 overlap with genotyping data from the Illumina 650/660k chip. We compared these to SNP chip data from 1,225 genotyped individuals belonging to 80 populations, including 49 previously unpublished individuals (SI12.1). The SNP data comprise four contemporary Aboriginal individuals from Northern Australia who were selected because their mtDNA and Y-chromosome DNA showed no evidence of European admixture. Individuals from the Kusunda and Aeta populations – two populations of hunter-gatherers from Nepal and the Philippines respectively - were also included. Both groups have been hypothesised to be possible descendants from an early wave of dispersal that also included the initial colonisation of Australasia[24, 25].

Discriminant analysis of principal components (DAPC)[26], principal component analysis (PCA)[27] and weighted PCA[28] were used to investigate between-population discrimination (SI12.3). The results show a cline of genetic differentiation from Africa to Asia, and then Australasia, reproducing the widely accepted scenario of a single human dispersal out of Africa. They are also consistent with the Australasian population deriving from one of the serial founders shared with Asians, but not with Europeans, followed by isolation of the Australo-Melanesian populations (Fig. 1a). This pattern is confirmed using only the 542 individuals from 43 Asian and Australasian populations (SI.12.3). Additionally, DAPC group assignment of individuals indicates that Australo-Melanesian populations form a well-defined genetic cluster (a-score=96%, SI12.3). There is no genetic signal to indicate that the Kusunda might be remnants of an earlier wave of colonists to Australia. However, the Aeta's clustering towards the Australo-Melanesian sample is in agreement with the possibility of wider regional spread of genes carried by an early wave of dispersal to Australasia.

We further explored the genetic differentiation of the ancient and modern Aboriginal Australians using the model-based structure-like approach ADMIXTURE[29]. The ancient Aboriginal genome is virtually indistinguishable from three of the four modern Aboriginal Australians, suggesting a common genetic history between the ancient South-Western and modern Northern Aboriginal Australians - the outlier being an individual with probable recent Asian admixture (also evident in the PCA/DAPC plots (SI12)). Although, we observe a close genetic relationship between all Australasian populations, a minor South Asian (Indian) signal is also visible in Aboriginal Australians which is not present in Melanesian populations. This may suggest that the peopling of Sahul (the Pleistocene continent of Australia, New Guinea and Tasmania) was not a single event[30], or that Aboriginal Australians have experienced more recent gene flow from South Asia[31], or that the signal has been erased from Melanesians due to higher levels of recent gene flow.

Population genetic estimates of divergence times

While the above comparisons to SNP chip data allow for detailed analyses of genetic similarity, they are less suitable for estimating population genetic parameters. This is because of the inherently biased ascertainment process used when selecting SNPs for chip genotyping platforms[32-34]. Therefore, to examine if the Aboriginal Australian sequence shows evidence of an early divergence from the Eurasian lineage, we developed a new population genetic analysis method for estimating demographic parameters from diploid whole genome data. In brief, the method characterises pairs of genomes in terms of their inferred transition probability matrix of pairwise allelic patterns. The method, therefore, gains information from both haplotype and joint allele frequency patterns. The observed matrix of genomic transitions between allelic patterns is then fitted to the data using a composite likelihood approach. The method allows joint estimation of divergence times and migration rates between pairs of populations, and further takes sequencing errors into account (SI13).

Using this method, we estimate a divergence time between the Aboriginal Australian sequence and representative European or East Asian sequences of ~ 2,500 generations, corresponding to ~ 50-60 thousand years assuming a generation time of 20-25 years. In contrast, we find the divergence time between representative European and East Asian sequences to be ~ 1,500 generations. All three populations, however, have a similar divergence time to the representative African sequence. This indicates that the Aboriginal Australians diverged from Asians and Europeans prior to the split between these populations. The high degree of genetic similarity between Aboriginal Australians and Asians, as measured by eigenvalues in the PCA/ DAPC analyses or by Identity-By-Descent (IBD) (Table 2), can be explained in our model by gene-flow, subsequent to the divergence between Europeans and East Asians.

Beyond demographic modelling, we also employed a test that compares the patterns of similarity between Asian or Aboriginal Australian individuals to African and European individuals (SI14). This test is closely related to the *D* test used in Green et al.[13], but is far more robust to errors because it only includes sites in which all alleles have been observed at least twice in modern humans. Furthermore, by only considering those sites where the

Aboriginal Australian differs from an Asian genome, this test is able to detect subtle demographic signals in the data that may be masked by the large amounts of gene-flow between Australians and Asians. Taking those sites where Australians (AUS) differ from Asians (ASN), and comparing AUS and ASN with the Ceph European Sample (CEU) representing Europe and the Yoruba representing Africa (YRI), we predict an equal number of sites suggesting Grouping I ((YRI, ASN), (CEU, AUS)) and Grouping II ((YRI, AUS), (CEU, ASN)). However, we find a statistically significant excess of sites (51.3%) grouping the Yoruba and Australian genomes together (Grouping II) relative to the Yoruba and Asian genomes together (Grouping I, 48.7%). The observed excess of sites showing Grouping II is expected under our best-fitting demographic model (Table 2). We estimate that an approximately 16% increased error rate in the Aboriginal Australian compared to the Asian sequences would be required to explain this result as the outcome of sequencing errors (SI14.3). However, when estimating the relative increase in error in the Aboriginal in relation to the Asian individuals, only a 0.2% increase is observed (SI17). Therefore, the greater affinity between the Aboriginal Australian and African observed here cannot be explained by errors in the Aboriginal Australian sequence. This suggests that a model where Aboriginal Australians were recently derived from Asian populations is not compatible with the data, unless there was gene flow between modern European and Asian populations, but not between Europeans and Aboriginal Australians. Instead, these results support a model where Aboriginal Australians split from the Eurasian population before Asian and European populations split from each other (SI14), which is consistent with the other lines of evidence presented above.

Archaic hominin admixture

Recent ancient DNA analyses have suggested that human populations outside Africa show evidence for admixture with Neanderthals[13]. Also, Melanesians possibly show evidence of increased admixture with a newly inferred hominin group, the Denisovans[35]. We used two approaches to test for such admixture in the sequence of the Aboriginal Australian. First, we examine if previously identified high-confidence Neanderthal admixture segments in Europeans and Asians can also be found in the Aboriginal Australian. As shown in the SI15, we find that the proportion of such segments in the Aboriginal Australian closely matches that observed in European and Asian sequences. Second, we use the *D* test (also known as the ABBA/BABA test) from Green et al.[13] and Reich et al.[35] to look for shared ancestry with these two hominin groups (SI16). The test is based on comparing the proportion of shared derived alleles between an outgroup sequence (in this case the Denisovan or Neanderthal) and two ingroup sequences. We find that the *D* test is highly sensitive to errors in the ingroup sequences being compared (SI16.3) and shared errors are of particular concern when the comparisons involve both an ingroup and outgroup ancient DNA sequence. We cannot exclude these results being influenced by such errors, but we nonetheless find it interesting that the test shows a relative increase in allele sharing between the Denisovan and the Aboriginal Australian, in accordance with the hypothesis of increased admixture between Denisovans and modern humans in Australasia[35].

Craniometric analyses

To address the origins of aboriginal Australians, we also analysed an extensive dataset consisting of 37 cranial traits measured on 6,245 individuals distributed into 107 indigenous human populations worldwide[36]. Craniometric features are heritable, highly variable, and largely unaffected by natural selection, and have therefore been used like genetic markers for inferring human demography and migrations. After correcting for the effects of size and sex, the data were submitted to DAPC and weighted PCA as for the genetic data. Morphometric data contained less signal than genome-wide SNP data, and only DAPC was able to reveal morphological differences between African, European, and Arctic populations (SI12.4). Interestingly, Australo-Melanesians and African populations exhibit strikingly similar cranial features (Fig. 2). This similarity could be indicative of common ancestry in these populations and could be consistent with two waves of dispersal out-of-Africa. Alternatively, it points to Africans and Australo-Melanesians sharing plesiomorphic traits (i.e. retention of ancestral characters), suggesting an evolutionary trajectory among the latter that is distinct from that of Eurasians.

Inference of disease risk

The disease burden of Australians of Aboriginal ancestry is estimated to be two-and-a-half times greater than that of other Australians[10]. A single Aboriginal genome does not offer the statistical rigour of genome-wide disease association studies. However, because this genome shows no evidence of recent European admixture, it can still pinpoint areas of future research interest. For example, we observe rare variation in a microRNA hsa-mir-196a2, potentially involved in the cleavage of mRNAs of HOX gene clusters that has been associated with several types of cancers[37]. We also find the genotype rs2294008(T;T) affecting the Prostate stem cell antigen[38] that has been shown to increase certain cancer risks and a variant in the VEGF gene encoding the vascular endothelial growth factor A associated with age-related loss of vision[39]. For additional observations see (SI18).

Discussion and conclusion

Our results show that it is possible to obtain a high-depth human genome sequence from historically ancient hair samples stored under conditions that were far from those in ideal permafrost regions[15]. Having such high coverage across the genome allows us to estimate proper genotypes, improving population genetic analyses. We also reveal that an uncontaminated genome sequence can be obtained despite handling by researchers of European decent, further demonstrating that historical ancient genomes can provide a means to avoid the issue of recent admixture that is often found in many modern-day indigenous populations. As such, our results provide a benchmark for conducting genomic-scale studies on the many historical hair samples in museums worldwide. Such ancient genomes also provide an important means to obtain basic genetic information to be used in tandem to larger genome-wide disease association studies in modern heavily admixed populations.

The genome of a single individual may or may not be representative of the evolutionary history of all Aboriginal Australians. The SNP chip data show an extraordinary degree of similarity between the modern Aboriginals from Northern Australia and the historic individual from the Southwest, pointing to a common history between at least these groups. Furthermore, based on our genome sequence data, we may conclude that the ancestors of the Aboriginal Australian man, and possibly all Aboriginal Australians, are as distant to Africans as are other Eurasians. However, the Aboriginal ancestors were the earliest to split from this broad ancestral non-African gene pool, 50-60 ka BP. In contrast, the ancestors of contemporary Europeans and Asians diverged from each other 30-50 ka BP i.e. 10-25 thousand years later. Genomic data further suggest that descendants from that first split received significant gene-flow from ancestral Asians after the latter had diverged from Europeans, reducing the primary dispersal genetic signal, and so increasing the similarity with

Asians. We can also conclude that the Aboriginal Australian genome, like other non-Africans, has a set of genes observed in Neanderthals, which may have been acquired by the gene flow event between Australians and Asians, rather than direct contact between Australian ancestors and Neanderthals.

An important lesson from the different analyses conducted in this study is that simple inferences of similarity may not tell the whole story. Real demographic histories are more complex than what can be revealed by analyses of similarity or shared genetic identity. For example, the mtDNA and Y-chromosome data show derived features shared by Australians and Eurasians but cannot easily discriminate between models of initial divergence and isolation, from models of later discrete gene flows in deep past. Simple analyses based on genetic similarity presented here, including PCA, DAPC, and admixture analyses, similarly point to a close genetic relationship between Australians and Asians, but are agnostic regarding demographic models. Only the craniometric analyses point to a different history, suggesting a unique relationship between Australian and Africans. Analyses of whole genome-data, using population genetics models, yield different results. They indicate that the closer affiliation of the Aboriginal Australians to Asians than to Europeans (evident from the PCA, DAPC, and admixture plots), is the likely result of significant secondary Asian gene-flow rather than shared history. They further suggest that the craniometric similarity between Africans and Aboriginal Australians does not reflect a private relationship between these groups, but rather retention of plesiomorphic traits consistent with a different out-of-Africa history of Aboriginal Australians and Eurasians. The implication is that reconstructing human population history requires sampling very substantial parts of the genome. The final important lesson to be learned is that signatures of admixture between modern humans outside Africa and ancient hominins may have spread among non-African populations through significant levels of more recent or secondary gene-flow rather than through direct contact between non-African populations and Neanderthals. Such signatures should be interpreted with great care.

Using our findings to distinguish between the different models on the early spread out-of-Africa of AMH is more enigmatic. In order to distinguish between a single *versus* multiple out-of-Africa dispersal scenarios to give rise to contemporary non-Africans, we need to find the geographical location of the 50-60 ka BP split of the Aboriginal ancestors from the Eurasian branch – something we cannot do in any reliable way at the moment. If the Neanderthal genetic signature is of primary origin, it may suggest that the split happened outside Africa, possibly in the Middle East, shortly after the Neanderthal admixture event. However, if the Neanderthal genetic signature derives from secondary Asian gene-flow, the divergence event could have taken place either within or outside Africa (Fig. 3). Likewise, if we take for granted a recent genome based divergence estimate of ~140 ka BP between contemporary Africans and non-Africans[40], our data suggest that the Aboriginal lineage diverged from other Eurasians some 80-90 thousand years after the Eurasian lineage was formed. This makes the divergence split less likely, but not impossible, to have happened in Africa. For the event to have occurred within Africa, the African population would have had to have been characterised by severe, long-term population structuring. There is, however, with the exception of the MIS5 modern humans in the Levant (Skhul and Qafzeh), which appear to become extinct, no convincing palaeontological evidence for modern humans outside Africa prior to ~45 ka BP. From an archaeological standpoint, this makes a within Africa Aboriginal and Eurasian split, followed by multiple dispersals, more likely. More evidence, based on ancient as well as modern genomes, is needed to distinguish fully between the different out-of-Africa dispersal models.

Method summary

The genomic sequence was generated from a historical ancient hair; data was aligned and genotyped using standard tools[19, 20]. A number of comparative analyses were run against a large sample size of genotyped individuals, these include DAPC[26], PCA[27, 28], and ADMIXTURE[29]. Several novel and modified population genetic analyses that compare contemporary and ancient genomes were used to infer the split time between Aboriginal Australians and Eurasians- detailed descriptions of all methods are provided in the Supplementary Information.

Author statement

The ancient aboriginal Australian hair sample was donated by a young man living in Western Australia, native of Golden Ridge (today known as Golden Station), 15 miles E of Kalgoorlie in the early 20th century. The sample was accessioned by Dr Alfred C. Haddon (Cambridge) into the Archaeology & Anthropology Museum collections in 1923 with the number Oc.1.1. It later became part of the Duckworth Collection, the University of Cambridge's main collection of human and non-human primate historical and prehistorical biological samples. The collection is kept solely for the purpose of scientific research and learning. Permission for sampling for destructive analyses is given by the Director. Hair samples donated by living individuals are not subject to the British Government's Human Tissue Act 2004.

Appropriate permissions were obtained for the genetic study of all modern human samples. For the modern aboriginal Australian samples the current study is an upgrade of previously published work[4], and has been approved by the local ethic committees. The Kusunda genome campaign was conducted by informed consent with the logistic assistance and active participation in the field of the *Ādivāsī Janjāti Utthān Rāṣṭrīya Pratiṣṭhān* (National Foundation for the Development of Indigenous Nationalities, NFDIN) under the auspices of the *Sthānīya Vikās Mantrālaya* (Ministry of Local Development) of the Government of Nepal.

Acknowledgements

The project was supported by the Lundbeck Foundation, the Danish National Research Foundation, the Novo Nordisk Foundation, Clare College, Newnham College (Cambridge), Wellcome Trust, Leverhulme Trust, the Biotechnology, Biological Sciences Research Council (BBSRC) grant BB/H008802/1, Estonian SF grants 7858, 8973 and 0270177, EU RDF Centre of Excellence in Genomics, and a Miller Research Fellowship from UC Berkeley (KEL). We thank K. Magnussen, T. B. Brand, P. S. Olsen, NCIP in the Philippines, Aeta and Kusunda people for help and support.

Author contributions

E.W. initially conceived and headed the project (J.W. headed research at BGI). M.R. and E.W. designed the experimental research project setup. M. M. L. and R.F. provided the hair sample along with detailed information and archaeological context. F.B., A.M., G.vD., A.R.J., F.X.R., P.dK., I.G.R., and T. H. provided modern Kusunda samples and cranial data. P.F., B.B., provided access to contemporary Aboriginal Australian samples. A. M. provided Aeta samples. O.N. and Mi. R. conducted isotopic analyses. M.R. and M.T.P.G. did the ancient DNA extractions. M.R. did the library builds, amplifications and purifications prior to sequencing. X.G., B.W., J.L., S.T., F.X. and W.J. did the Illumina sequencing on the prepared libraries and provided access to three Asian genomes of high quality and depth. R.V. ran Illumina genotypings and provided all genotyping data. S.L. and A.K. did the initial raw read processing, and subsampled the genomes. S.R. and T.S.P. mapped and genotyped ancient and modern data. S.R., O.L., A.W., K.N., R.G., S.B. and T.S.P. did de novo assembly, metagenomics analyses, disease/phenotype mapping and HLA typing. M.R., M.A.C.C., and J.V.M.M. did various bioinformatic analyses. L.O. did the damage estimates and data filtering. A.A., L.S. and R.N. did the contamination estimates and abbababa analyses. T.J. and F.B. did DAPC, weighted PCA and all analyses on cranial data. M.M., E.M., M.K. and R.V. did genotype data prep, PCA plots, admixture analyses. T.K. did mtDNA and Y-chr analyses. Y.W., K.E.L. and R.N. did the demographic modeling and populations genomic analyses. W.Z. and R.N. did the archaic hominin admixture analyses. M.M. did the figures. E.W. and M.R. wrote the majority of the manuscript, with critical input from F.B., M.L., R.F., R.N., R.V., S.R., K.E.L., T.K., M.M., T.S.P., A.A., L.S., L.O., Y.W., S.L.. W.Z., R.G., M.T.P.G. and the remaining authors.

References

1. Stringer, C. B. & Andrews, P. Genetic and fossil evidence for the origin of modern humans. *Science* **239,** 1263-1268 (1988).
2. Ingman, M & Gyllensten, U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res.* **13,** 1600-1606 (2003).
3. van Holst Pellekaan *et al.* Mitochondrial genomics identifies major haplogroups in Aboriginal Australians. *Am. J. Phys. Anthropol.* **131,** 282-294 (2006).
4. Hudjashov, G. *et al.* Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc. Natl. Acad. Sci. U S A.* **104,** 8726-8730 (2007).
5. Cavalli-Sforza, P & Menozzi, A. Piazza, The History and Geography of Human Genes (Princeton Univ. Press, Princeton, NJ, 1994).
6. Lahr, M. M. & Foley, R. Multiple Dispersals and Modern Human Origins. *Evol. Anthropol.* **3,** 48-60 (1994).

7. Lahr, M.M. & Foley. R.A. Towards a theory of modern human origins: Geography, demography and diversity in recent human evolution. *Yearbook of Physical Anthropology* **41,** 137-176 (1998).

8. McEvoy, B. P. *et al.* Whole-genome genetic diversity in a sample of Australians with deep aboriginal ancestry. *Am. J. Hum. Gen.* **87,** 297-305 (2010).

9. Taylor, D. A. *et al.* Knowing your DNA database: Issues with determining ancestral Y haplotypes in a Y-Filer database. *Forensic Science International: Genetics Supplement Series* **2**, 411–412 (2009).

10. Begg, S. *et al.* Burden of disease and injury in Australia: Australian Institute of Health and Welfare AIHW 2007-05-25 (2003).

11. Green, R. E. *et al.* Analysis of one million base pairs of Neanderthal DNA. *Nature* **444,** 330-336 (2006).

12. Poinar H. N. *et al.* Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311,** 392-394 (2006).

13. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328,** 710-722 (2010).

14. Gilbert, M. T. P. *et al.* Paleo-Eskimo mtDNA Genome Reveals Matrilineal Discontinuity in Greenland. *Science* **320,** 1787-1789 (2008).

15. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463,** 757-762 (2010).

16. Schuster, S. C. *et al.* Complete Khoisan and Bantu genomes from southern Africa. *Nature* **463,** 943-947 (2010).

17. Lewis, M. Paul (ed.), 2009. Ethnologue: Languages of the World, Sixteenth edition. Dallas, Tex.

18. Binladen, J. *et al.* Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics* **172,** 733-741 (2006).

19. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754-1760 (2009).

20. Li H, *et al.* The sequence alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078-2079 (2009).

21. Macaulay, V. *et al.* Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308,** 1034-1036 (2005).

22. Karafet, T. M. *et al.* Major east-west division underlies Y chromosome stratification across Indonesia. *Mol. Biol. Evol.* **27,** 1833-44 (2010).

23. Mellars, P. Going east: New genetic and archaeological perspectives on the modern human colonization of Eurasia? *Science* **313,**796-800 (2006).

24. Lahr, M.M. The Evolution of Modern Human Cranial Diversity: A Study in Cranial Variation. Cambridge, Cambridge University Press (1996).

25. Whitehouse, P. *et al.* Kusunda: an Indo-Pacific language in Nepal. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 5692-5695 (2004).

26. Jombart, T. *et al.* Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics **11,** 94 (2010).

27. Patterson, N. *et al.* Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).

28. Dray, S. & Dufour, A.B. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software. **22,** 1–20 (2007).

29. Alexander, D. H. *et al.* Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).

30. Birdsell J. B. The recalibration of a paradigm for the first peopling of Greater Australia. In: Allen J. *et al.* editors. Sunda and Sahul prehistoric studies in Southeast Asia, Melanesia and Australia. London: Academic Press. p 113–167 (1977).

31. Redd, J. A. *et al.* Gene flow from the Indian subcontinent to Australia. *Curr. Biol.* **12,** 673 – 677 (2002).

32. Nielsen R. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* **154,** 931-942 (2000).

33. Clark, A. G. *et al.* Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res.* **15,**1496-1502 (2005).

34. Albrechtsen, A. *et al.* Ascertainment biases in SNP chips affect measures of population divergence. *Mol. Biol. Evol.* **27,** 2534-2547 (2010).

35. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468,** 1053-1060 (2010).

36. Manica, A. *et al.* The effect of ancient population bottlenecks on human phenotypic variation. *Nature* **448,** 346-348 (2007).

37. Zhibin, H. *et al.* Common genetic variants in pre-microRNAs were associated with increased risk of breast cancer in Chinese women. *Hum. Mutat.* **30,** 79-84 (2009).

38. Xifeng, W. *et al.* Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. *Nat. Genet.* **41,** 991-995 (2009).

39. Churchill, A. J. et al. VEGF polymorphisms are associated with neovascular age-related macular degeneration. *Hum. Mol. Genet.* **15,** 2955-2961 (2006).

40. Gutenkunst R. N. *et al.* Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genet.* **10,** e1000695 (2009).

Table 1. Ancient genome summary statistics.

Mapping statistics	Giga-bases
Total raw bases	209.3
Total mapped bases[a]	66.9
Bases covered[a]	1.79
Bases covered[a] ≥ 10x depth	0.76

Variant	Number
SNPs	2,782,401
High confidence SNPs	449,115
Indels	215,189
High confidence indels	22,576

[a] Using hg19 as reference

Table 2. Number of sites showing Groupings 1 and 2 in the aboriginal Australian genome.

	Grouping 1[a]	Grouping 2[a]
YRI	1	1
AUS	0	1
CEU	0	0
ASN	1	0
Observed Number[b]	14001	14758
Observed Proportion	48.7%	51.3%[c]
(95% CI)	(48.1%-49.3%)	(50.7%-51.9%)
Expected proportion under best-fitting model[d]	48.3%	51.7%
Expected proportion under alternative model[e]	49.9%	50.1%

The results are from using NA19239 (for YRI), NA12891 (for CEU), HG00421 (for ASN), and the aboriginal Australian genome.

[a] The patterns shown here do not reflect ancestral vs. derived states (i.e. they have not been polarized). Rather these patterns represent the two ways in which the eligible SNPs can partition the four genomes.

[b] The average number of eligible SNPs showing Groupings 1 and 2 from 100 different replicates of sampling alleles at heterozygous sites.

[c] We reject the null hypothesis that this value is equal to 50%. The median P-value from a binomial test is $<2 \times 10^{-5}$. Note that the 95% CI calculated here assumes that all the SNPs are independent of each other, and as such, is likely too narrow.

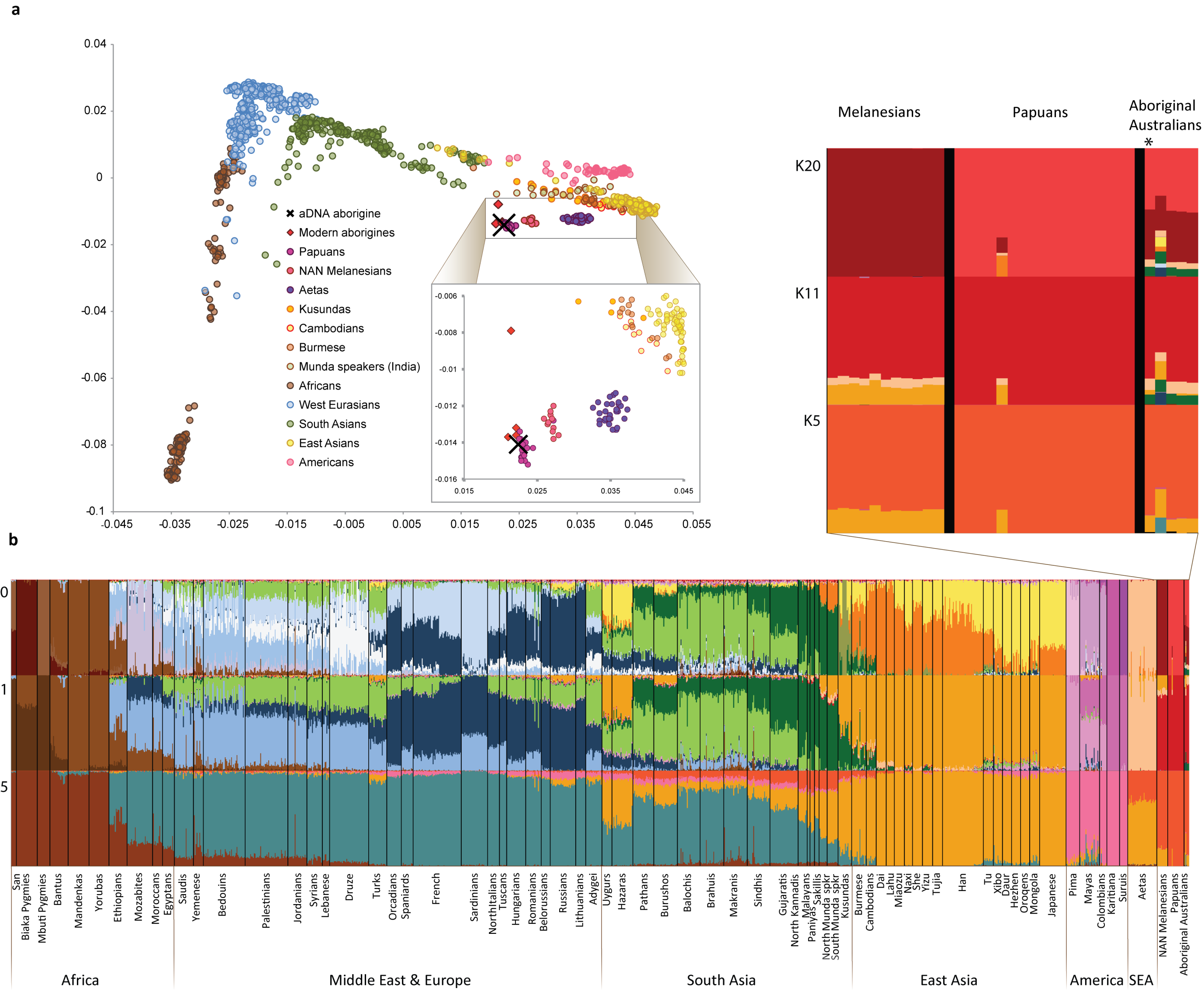
[d] The expected proportion obtained from our best-fitting demographic model where aboriginal Australians split from Eurasian populations 2,500 generation ago, prior to the split of European and Asian populations (SI14).

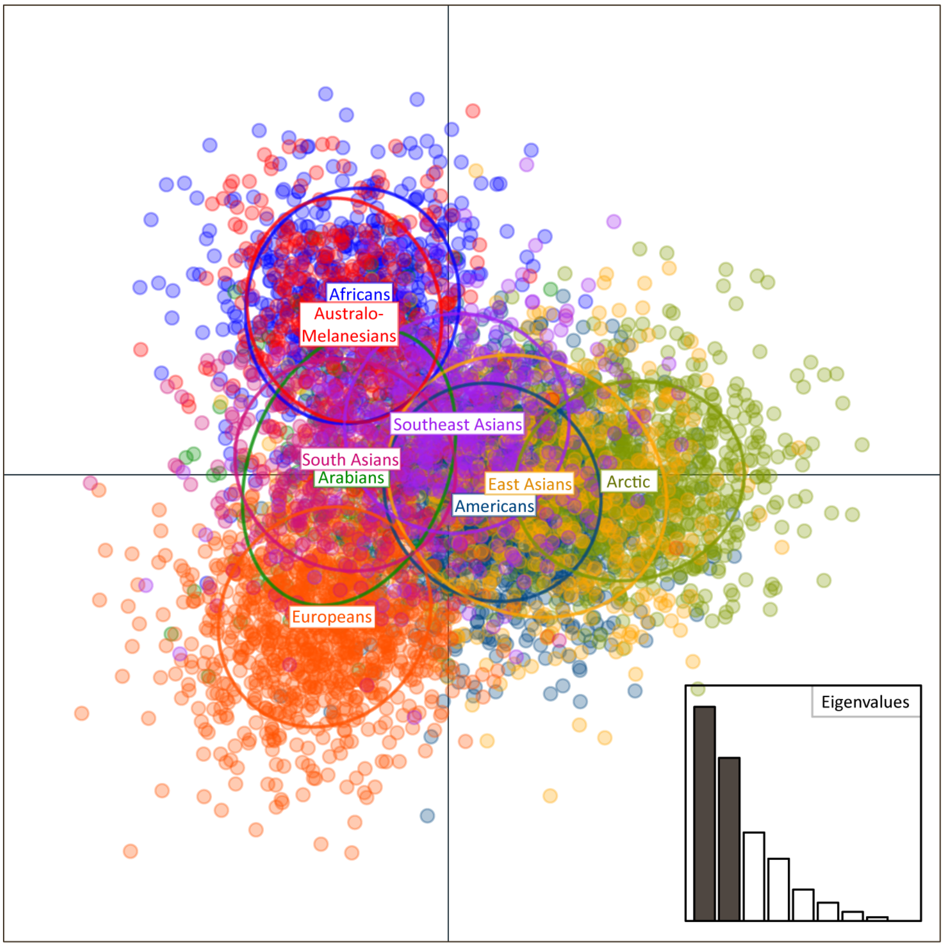
[e] The expected proportion obtained from coalescent simulations under a demographic model where aboriginal Australians split from Asian populations 1,500 generations ago (SI13).

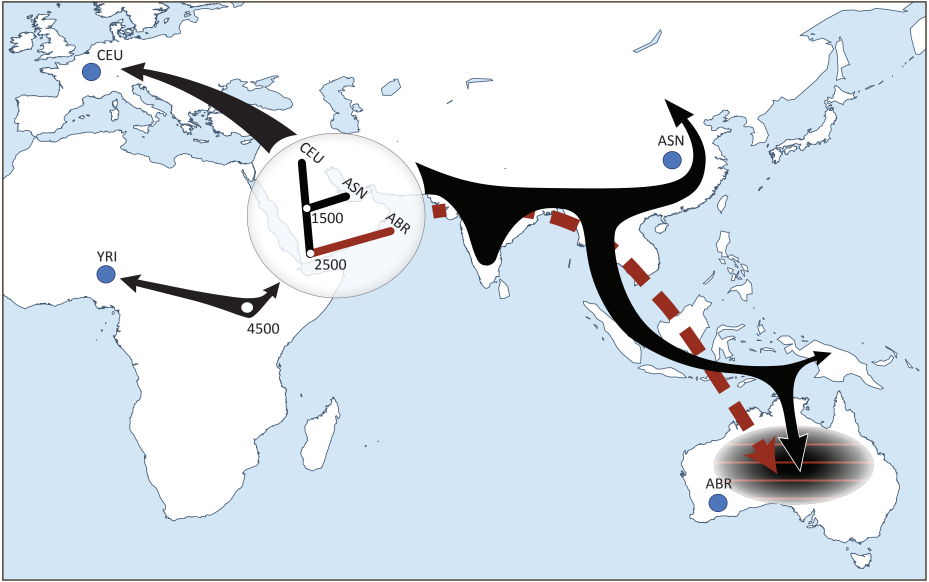
Fig. 1. a, PCA plot (PC1 versus PC2) of the studied populations and the ancient genome of the aboriginal Australian (marked with a cross). Insert shows the Australasian populations (SI12.3). **b**, Ancestry proportions of the studied 1,225 individuals from 80 populations and the ancient Aboriginal Australian as revealed by the ADMIXTURE program[29] with $K = 5$, K=11 and K=20. Each individual is represented by a stacked column of the K proportions, with fractions indicated on the y axis. The Australasian populations are shown in detail in the insert (SI12.2).

Fig. 2. First and second principal components of the DAPC of 6,245 skulls measured for 37 cranial traits from 107 native human populations distributed worldwide. Individuals are represented by dots. Populations are indicated by colours and ellipses which model 95% of the corresponding variability. The inset displays the eigenvalues of the analysis, with represented axes in black and retained axes in grey. The genetic differentiation amongst groups is proportional to the magnitude of the corresponding eigenvalues (SI12.4).

Fig. 3. Reconstruction of early spread of modern humans, the divergence of the Aboriginal Australian and Eurasian lineages (in generations), and the secondary Asian gene-flow to the Aboriginal Australian lineage.


a
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.045
-0.035
-0.025
-0.015
-0.005
0.005
0.015
0.025
0.035
0.045
0.055
✖ aDNA aborigine
◆ Modern aborigines
● Papuans
● NAN Melanesians
● Aetas
● Kusundas
○ Cambodians
● Burmese
● Munda speakers (India)
● Africans
○ West Eurasians
● South Asians
● East Asians
● Americans
-0.006
-0.008
-0.01
-0.012
-0.014
-0.016
0.015
0.025
0.035
0.045
Melanesians
Papuans
Aboriginal Australians
*
K20
K11
K5
b
K20
K11
K5
San
Biaka Pygmies
Mbuti Pygmies
Bantus
Mandenkas
Yorubas
Ethiopians
Mozabites
Moroccans
Egyptans
Saudis
Yemenese
Bedouins
Palestinians
Jordanians
Syrians
Lebanese
Druze
Turks
Orcadians
Spaniards
French
Sardinians
NorthItalians
Tuscans
Hungarians
Romanians
Belorussians
Russians
Lithuanians
Adygei
Uygurs
Hazaras
Pathans
Burushos
Balochis
Brahuis
Makranis
Sindhis
Gujaratis
North Kannadis
Malayans
Paniyas
Sakillis
North Munda spkr
South Munda spkr
Kusundas
Burmese
Cambodians
Dai
Lahu
Miaozu
Naxi
She
Yizu
Tujia
Han
Tu
Xibo
Daur
Hezhen
Oroqens
Mongola
Japanese
Pima
Mayas
Colombians
Karitiana
Suruis
Aetas
NAN Melanesians
Papuans
Aboriginal Australians
Africa
Middle East & Europe
South Asia
East Asia
America
SEA


Africans
Australo-Melanesians
Southeast Asians
South Asians
Arabians
East Asians
Arctic
Americans
Europeans
Eigenvalues


CEU
YRI
4500
CEU
ASN
1500
ABR
2500
ASN
ABR