

# Inferring DNA methylation in non-skeletal tissues of ancient specimens

Received: 7 May 2023

Accepted: 1 October 2024

Published online: 20 November 2024

 Check for updates

Yoav Mathov<sup>1,2</sup>, Malka Nissim-Rafinia<sup>1</sup>, Chen Leibson<sup>1</sup>, Nir Galun<sup>1</sup>,  
Tomas Marques-Bonet<sup>3,4,5,6</sup>, Arye Kandel<sup>7</sup>, Meir Liebergal<sup>7</sup>,  
Eran Meshorer<sup>1,2</sup>✉ & Liran Carmel<sup>1</sup>✉

Genome-wide premortem DNA methylation patterns can be computationally reconstructed from high-coverage DNA sequences of ancient samples. Because DNA methylation is more conserved across species than across tissues, and ancient DNA is typically extracted from bones and teeth, previous works utilizing ancient DNA methylation maps focused on studying evolutionary changes in the skeletal system. Here we suggest that DNA methylation patterns in one tissue may, under certain conditions, be informative on DNA methylation patterns in other tissues of the same individual. Using the fact that tissue-specific DNA methylation builds up during embryonic development, we identified the conditions that allow for such cross-tissue inference and devised an algorithm that carries it out. We trained the algorithm on methylation data from extant species and reached high precisions of up to 0.92 for validation datasets. We then used the algorithm on archaic humans, and identified more than 1,850 positions for which we were able to observe differential DNA methylation in prefrontal cortex neurons. These positions are linked to hundreds of genes, many of which are involved in neural functions such as structural and developmental processes. Six positions are located in the neuroblastoma breaking point family (NBPF) gene family, which probably played a role in human brain evolution. The algorithm we present here allows for the examination of epigenetic changes in tissues and cell types that are absent from the palaeontological record, and therefore provides new ways to study the evolutionary impacts of epigenetic changes.

Changes in gene regulation often underlie phenotypic divergence<sup>1–4</sup>, making the identification of regulatory differences between archaic humans (Neanderthals and Denisovans) and anatomically modern ones a major goal of palaeogenetics<sup>5</sup>. Because RNA is rarely present in palaeontological remains<sup>6</sup>, changes in gene regulation must be inferred indirectly from the ancient DNA (aDNA) sequences themselves. Previous works showed that aDNA degradation signals can be harnessed to computationally reconstruct premortem genome-wide DNA methylation maps<sup>7–9</sup>. DNA methylation is a key epigenetic mark that strongly affects the activity level of regulatory regions such as promoters

and enhancers<sup>10</sup>. Hence, its reconstruction provides information on premortem gene activity patterns in ancient individuals. Indeed, the reconstruction of DNA methylation in Neanderthals, a Denisovan and anatomically modern humans allowed us to identify regulatory differences between these human groups and associate them with phenotypic changes, opening the field of palaeoepigenetics<sup>9,11,12</sup>.

However, DNA methylation is tissue-specific to the extent that methylation patterns in two different tissues of the same organism are often more diverged than methylation patterns in the same tissue in two different species<sup>13,14</sup>. Although occasionally aDNA is extracted from

A full list of affiliations appears at the end of the paper. ✉ e-mail: [eran.meshorer@mail.huji.ac.il](mailto:eran.meshorer@mail.huji.ac.il); [liran.carmel@huji.ac.il](mailto:liran.carmel@huji.ac.il)

soft tissues such as skin<sup>15</sup> and liver<sup>16</sup>, this is limited to a small number of highly conserved and relatively young samples. Generally, bones and teeth are the main sources of aDNA. In particular, these tissues are the source of all archaic human aDNA. Accordingly, our previous palaeoepigenetic studies focused on the evolution of the skeletal system<sup>9,12</sup>, but could provide only limited insights on other systems such as the nervous system<sup>8</sup>.

Yet, DNA methylation in one tissue still carries information on DNA methylation in other tissues of the same individual. The reason is that DNA methylation is erased almost completely in the zygote, and then re-established in a cell type-specific manner during development<sup>10</sup>. When a DNA methylation change that separates human groups is established during early developmental stages, it may propagate to all descendant cell types, and be simultaneously manifested in multiple tissues.

Here we show that for a certain class of DNA methylation changes, we are able to use parsimony considerations<sup>17</sup> to predict the timing of the change during development, rendering skeletal DNA methylation changes informative on the methylation state in other tissues. Testing our method on modern primate tissues, we show that we achieve precision of 0.7–0.92 in using DNA methylation changes in one tissue to predict corresponding changes in another tissue.

We applied our method to predict DNA methylation changes in prefrontal neurons between modern humans, archaic humans and chimpanzees. We found 71 genes that are associated with differential neuronal DNA methylation separating modern and archaic humans, as well as 870 genes that are associated with differential neuronal DNA methylation separating all humans from chimpanzees. We show that some of these genes are involved in neural functions. Moreover, we discovered six methylation changes in genes that carry the Olduvai domain, whose copy number is associated with an increase in brain volume and cognitive function, suggesting that these changes might be important in human brain evolution.

## Results

Given the almost complete reset of DNA methylation patterns in the zygote as part of the epigenetic reprogramming process<sup>18</sup>, evolutionary differences in DNA methylation patterns between species are manifested as changes that build up during embryonic development. A methylation change that occurs in cells that are not fully differentiated would propagate to all descendant cell lineages, unless reverted by a later change. For example, if a site goes through an evolutionary methylation change that builds up at a time that precedes the split of the mesodermal and the ectodermal germ layers, the difference in methylation level would show up both in bone and in neurons (Fig. 1a).

For any pair of tissues or cell types, we denote by ‘fundamental changes’ all those evolutionary changes to DNA methylation that were established before the developmental split between the two tissues or cell types. Evolutionary DNA methylation changes that occurred after this developmental time point would be called tissue-specific changes (Fig. 1a). Note that the terms tissue-specific and fundamental methylation changes should be always understood in the context of two specific tissues or cell types. For example, a methylation change affecting all endodermal tissues is still considered tissue-specific if we compare, for example, lung with bone. The basis of the algorithm we develop below is the fact that if we are able to determine, for two tissues or cell types, whether an evolutionary change in DNA methylation is either fundamental or tissue-specific, then the methylation level in one tissue or cell type would be informative on the methylation level in the other tissue or cell type.

### The inference procedure

We consider a three-species phylogeny, in which the species are marked as ‘reference’ (superscript r), ‘target’ (superscript t) and ‘outgroup’ (superscript o). For these three species, we examine five DNA methylation maps in two tissues (or cell types). In one tissue, denoted ‘all’

(superscript a), DNA methylation data are available for all three species. In the other tissue, denoted ‘partial’ (superscript p), DNA methylation data are available for the reference and the outgroup only (Fig. 1b). Our goal is to predict the methylation levels in the target in tissue ‘partial’.

A useful abstraction of the problem, which allows for convenient conceptualization of the approach, is to imagine that the DNA methylation in each CpG position is a binary variable that can be either high or low. We dub the combination of binary methylation levels across the five available DNA methylation maps in a specific CpG position a configuration. There are exactly three possible combinations of methylation states that describe CpG positions with differential methylation across the three species in tissue ‘all’ (Fig. 1c). An evolutionary methylation change in tissue ‘all’ can occur along the branch leading to the reference (reference-derived), along the branch leading to the target (target-derived) or along either of the two branches that separate the outgroup from the reference and target (outgroup-specific). Each of these three combinations may be associated with any of four different combinations of methylation states in the tissue ‘partial’, leading to a total of 12 possible configurations that should be examined (Fig. 1c).

For some configurations, it is possible to determine the developmental timing of the methylation change using parsimony, whereas for others it is not. In a previous publication<sup>17</sup> we outlined the procedure and demonstrated it in a particular configuration. Here we examine all possible configurations, identify those in which inference on the ‘target’ methylation can be made, develop a detailed algorithm to carry out the inference procedure and use it to predict DNA methylation levels in archaic human brains.

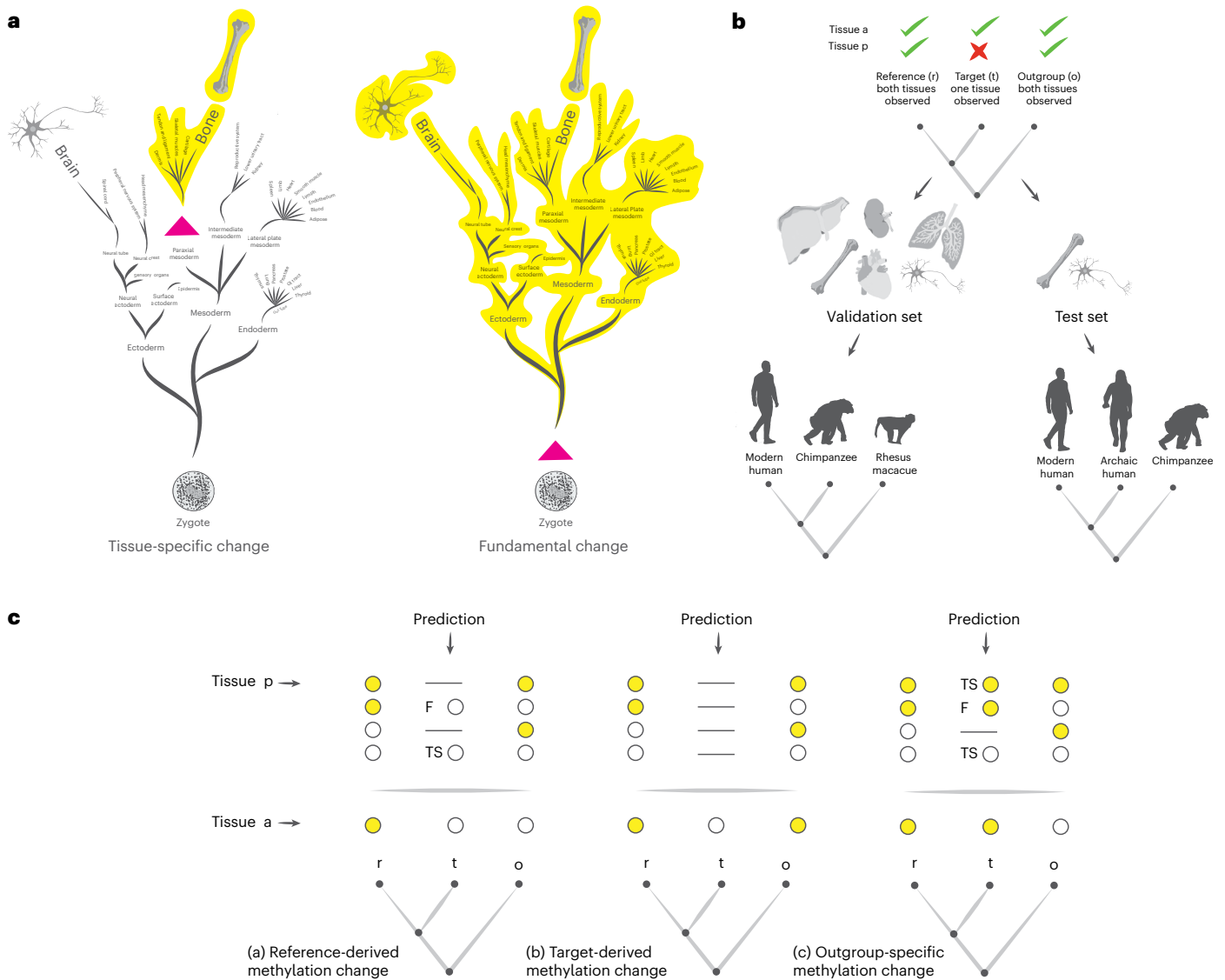
Of the 12 possible configurations, there are five for which parsimony considerations provide a prediction for the timing of the developmental change in methylation (Fig. 1c). In three configurations we predict that the change was tissue-specific, in two configurations the change was fundamental and in all others that it is impossible to determine the timing of the change. In tissue-specific configurations we end up predicting that the methylation change detected in tissue ‘all’ does not hold in tissue ‘partial’. These cases are less relevant for the current study, where our goal is to identify methylation changes in tissue ‘partial’. In the remainder of the article, therefore, we focus solely on the two configurations in which we conclude that the methylation change was fundamental. Note that because data are missing in tissue ‘partial’ of the ‘target’, we cannot make any predictions regarding developmental timing of the methylation change in target-derived changes.

### Quantifying the performance

Based on the abstraction above, we developed an algorithm that accounts for the non-binary nature of DNA methylation and identifies CpG positions that are consistent with having gone through a fundamental methylation change (Methods). In these positions, we predict that the methylation levels in ‘target’ in tissue ‘partial’ are closer to those in the outgroup (for a reference-derived methylation change) or to those in the reference (for an outgroup-specific methylation change) (Fig. 1c).

Eventually, we would like to apply our algorithm to a triad of species made of modern human (reference), archaic human (target) and chimpanzee (outgroup). However, lacking DNA methylation in non-skeletal tissues of archaic humans, we optimized the parameters and quantified the performance based on a triad of extant species with available DNA methylation data in multiple tissues. To resemble the triad that we will eventually use, we used three modern primate species: modern humans as the ‘reference’, chimpanzees as the ‘target’ and rhesus macaques as the ‘outgroup’.

We trained the algorithm for the above triad of extant species, using heart as the tissue ‘all’ and kidney as the tissue ‘partial’ (Methods, Supplementary Tables 1 and 2 and Extended Data Figs. 1 and 2). For predicting reference-derived fundamental changes, we obtained a precision of



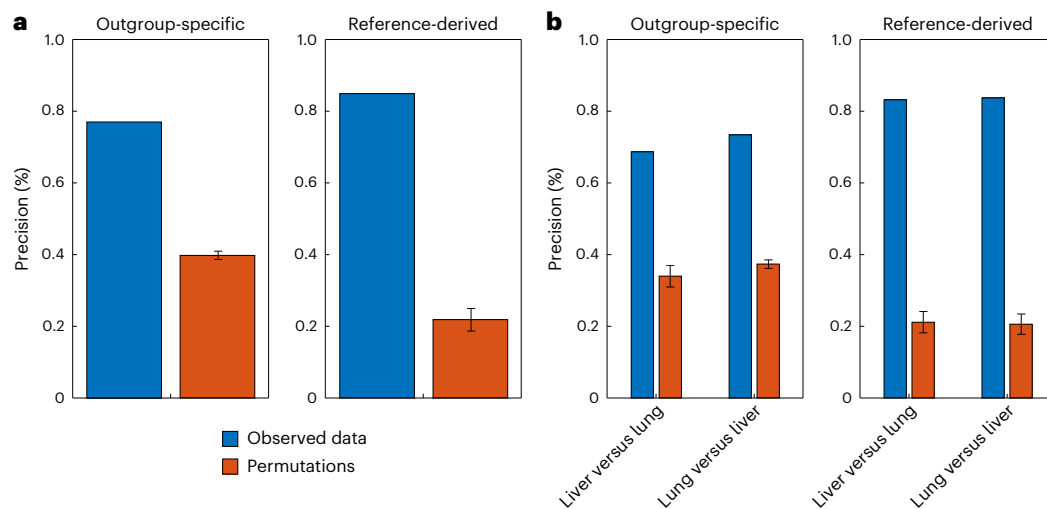
**Fig. 1 | Conceptual framework of the algorithm. a**, Schematic embryonic development trees. DNA methylation changes along the tree would propagate to all descendant lineages. Considering a pair of tissues, for example bone and brain, methylation changes that occurred earlier than their developmental split (right) are fundamental and will affect both tissues. Changes that occurred after the split (left) are tissue-specific and will affect only one of the tissues. **b**, Different triad phylogenies and paired tissues are considered by the algorithm

for validation and test sets. Methylation data are missing for the target 't' in tissue 'p', but available for tissue 'a' for all species. **c**, Identification of configurations, out of all 12 possible, where the methylation state in the target in tissue 'p' can be predicted. When prediction is possible, the type of methylation change, whether fundamental (F) or tissue-specific (TS), is denoted. Note that for clarity, methylation levels are considered as binary variables.

0.85, compared with an average of 0.22 in permuted data. For predicting outgroup-specific fundamental changes, we obtained a precision of 0.77, compared with an average of 0.40 in permuted data (Fig. 2a). Elevated precision for permuted data when predicting outgroup-specific fundamental changes is expected and, in fact, inherent to our algorithm. This stems from the fact that the evolutionary distance between the outgroup and the reference and/or target is, by definition, longer than the evolutionary distance between the target and the reference, leading to a greater similarity in the DNA methylation patterns of the reference and the target, and hence to an increased probability of finding by chance configurations that resemble outgroup-specific fundamental changes ('Discussion'). In any case, for both reference-derived changes and outgroup-specific changes, the algorithm shows training set precision which is significantly higher than random.

We validated our algorithm using DNA methylation data from liver and lung in these three species, which are tissues on which the algorithm was not trained. In each comparison, the precision achieved using

the real data significantly exceeded the precision that was achieved using permuted data. The observed precision in predicting fundamental changes was in the range 0.7–0.84, compared with 0.21–0.37 on permuted data (Fig. 2b and Supplementary Table 3). Flipping the roles of reference and target, namely picking modern human as the target and chimpanzees as the reference, yielded very similar performance estimations (Supplementary Table 3 and Extended Data Fig. 3). To evaluate the statistical significance of the precision levels we obtain, we performed 1,000 permutations, in which in each permutation we randomly shuffled the tags 'reference', 'target' and 'outgroup' across the samples, keeping the total number of reference, target and outgroup samples fixed. For reference-derived inference, we obtained that 0.004 of the permutations yielded higher precision than our algorithm, whereas this fraction was 0.021 for outgroup-specific inference. Combined, these results show that our algorithm achieves significantly higher precision than expected by chance, even when generalized to tissues and a species combination on which it was not trained.



**Fig. 2 | Precision in detecting fundamental DNA methylation changes using endodermal and mesodermal tissues. a**, Precision of the algorithm using the optimized parameters on observed data (blue) compared with permutations

(orange) in the detection of outgroup-specific (left) and reference-derived (right) fundamental changes. **b**, Validation of the algorithm on tissues that were not used for training. Error bars denote 1 s.d.

Because of the need to achieve sufficient statistical power, reconstructing DNA methylation maps of aDNA requires the use of moving averages across windows of consecutive CpG positions<sup>8</sup>. Although we showed in the past that this smoothing does not substantially affect the detection of differentially methylated regions (DMRs)<sup>9</sup>, we nevertheless wished to examine the effect of smoothing on the precision of our algorithm. We therefore applied the algorithm to a smoothed version of the above data (Methods) and found that the precision remains very similar to that obtained for non-smoothed data, and that it is always significantly higher than the precision achieved on permuted data (Supplementary Table 3).

### Identifying differential methylation in neurons

The method we developed can be used to infer on aDNA methylation in any tissue. Because the large volume and high complexity of the human brain stand out as hallmarks of modern humans, we decided to demonstrate the applicability of our method to this organ first. To this end, we applied the algorithm with bone serving as the tissue ‘all’ and neuronal cell lines from the prefrontal cortex as the tissue ‘partial’.

For bones, we considered published DNA methylation data (including data generated by us in previous works), data of four human femora measured by us for the current study and reconstructed DNA methylation maps in ancient samples, including maps we reconstructed in previous works<sup>10,17</sup> and a new map of Mesolithic anatomically modern human from Sweden<sup>19</sup> that we reconstructed for this study (Methods and Supplementary Table 4). DNA methylation in prefrontal cortex cell lines was taken from published studies<sup>20,21</sup>. In total, neuronal DNA methylation data comprise 26 modern humans, 11 chimpanzees and 15 macaques, and bone DNA methylation data comprise 11 modern humans (5 measured, 6 reconstructed), 2 archaic humans, 7 chimpanzees and 10 macaques.

We first estimated the precision of our algorithm on a triad of extant species, composed of modern human as reference, chimpanzee as target, and rhesus macaque as outgroup. This yielded a precision of 0.81 in predicting outgroup-specific fundamental changes, and 0.92 in predicting reference-derived fundamental changes, both being significantly higher than the values achieved for random permutations (Fig. 3a).

We then applied the algorithm to the triad composed of modern human as ‘reference’, archaic human as ‘target’ and chimpanzee as ‘outgroup’, and identified 1,750 CpG positions presenting putative outgroup-specific fundamental changes, and 122 CpG positions

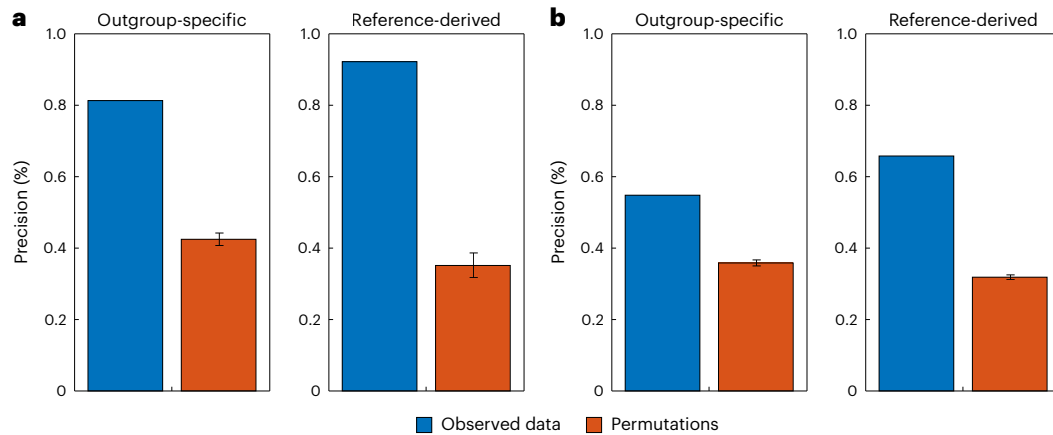
presenting putative reference-derived fundamental changes. Of these, 1,131 CpG positions (64.6%), presenting putative outgroup-specific changes, are located in the promoter or gene body of 870 genes, and 80 CpG positions (61.1%), presenting putative reference-derived changes, are located in the promoter or gene body of 71 genes (Supplementary Table 5).

Archaic humans, which serve as ‘target’, comprise only two samples. To assess the possible reduction in power because of this small sample set, we repeated the analysis on the triad of extant species, but randomly downsampled the ‘target’ chimpanzee from 11 samples to only 2. We still detected 7 of the original 8 CpG positions associated with reference-derived changes, and 483 of the 531 original CpG positions associated with outgroup-specific changes. This suggests that the use of only two samples in the target set reduces the power of our analysis by approximately 10–12%.

We accounted for batch effects and evaluated the false discovery rate (FDR) for our actual triad by performing 1,000 permutations, shuffling the tags ‘reference’, ‘target’ and ‘outgroup’ of the bone samples and counting the number of detected reference-derived and outgroup-specific changes. On average, we observed 0.49 reference-derived CpG positions (FDR = 0.005) and 68.6 outgroup-specific CpG positions (FDR = 0.044).

A considerable number of the CpG positions associated with fundamental changes are grouped in clusters. This observation probably reflects the known correlation between adjacent CpG positions, and indicates regional changes in methylation in regulatory regions, where one often finds a close grouping of multiple differentially methylated CpGs<sup>22,23</sup>. We hypothesized that even isolated CpG positions showing fundamental changes would represent regional methylation changes. Namely, that the adjacent CpG positions would show changes in DNA methylation compatible with the same fundamental changes, despite not crossing the significance threshold. To test this, we computed the average methylation in such adjacent CpG positions, and compared it with both the nearby isolated CpG positions showing a fundamental change and the average methylation level in the genome. As expected, we found that in 89.1% of the cases the methylation level in adjacent CpGs clustered with the neighbouring isolated CpG position, suggesting that even loci featuring an isolated CpG position with a fundamental change might, in fact, represent a regional methylation change.

CpG positions with fundamental methylation changes are expected to preferentially fall inside regions we detected in a previous work as differentially methylated across the same triad, using a subset



**Fig. 3 | Precision in detecting fundamental DNA methylation changes using brain and bone.** **a, b**, Precision of the algorithm in detecting fundamental changes affecting both bone and prefrontal cortex neurons (**a**) or bulk prefrontal cortex tissue (**b**) on observed data (blue) and permuted data (orange). Error bars denote 1 s.d.

of the bone samples used in the current study<sup>9</sup>. Indeed, 172 of the 1,750 outgroup-specific positions (9.8%) fall in the list of 2,031 DMRs that separate chimpanzees and humans ( $P < 10^{-200}$  hypergeometric test), and 3 of the 122 reference-derived positions (2.5%) fall in the list of 873 anatomically modern human-derived DMRs ( $P = 2.8 \times 10^{-6}$ ). Likewise, 137 of the 870 (15.7%) genes associated with outgroup-specific changes overlap with the list of 1,185 differentially methylated genes separating chimpanzees and humans ( $P = 7.9 \times 10^{-40}$ ), and 5 of the 71 (7%) genes associated with reference-derived changes overlap with the list of 588 differentially methylated genes that have changed in modern humans ( $P = 4.8 \times 10^{-3}$ ). Notably, we do not expect a complete overlap, because we used an upgraded version of the reconstruction algorithm, including new filtering techniques and histogram matching<sup>24</sup>.

Testing for enrichment in biological functions of our gene lists yielded no significant terms. This is expected, given that a fundamental methylation change is, by definition, a change that is shared across multiple tissues. Specifically, fundamental methylation changes that are shared between bone and brain could be associated with brain, bone or any other tissue (Fig. 1a). Therefore, these changes could be related to a wide range of functions, making it unlikely that specific functional enrichment would be observed. However, given the importance of the brain in human evolution, and the lack of methylation data from brain in archaic human, we focus here on changes that might have a relation to the nervous system.

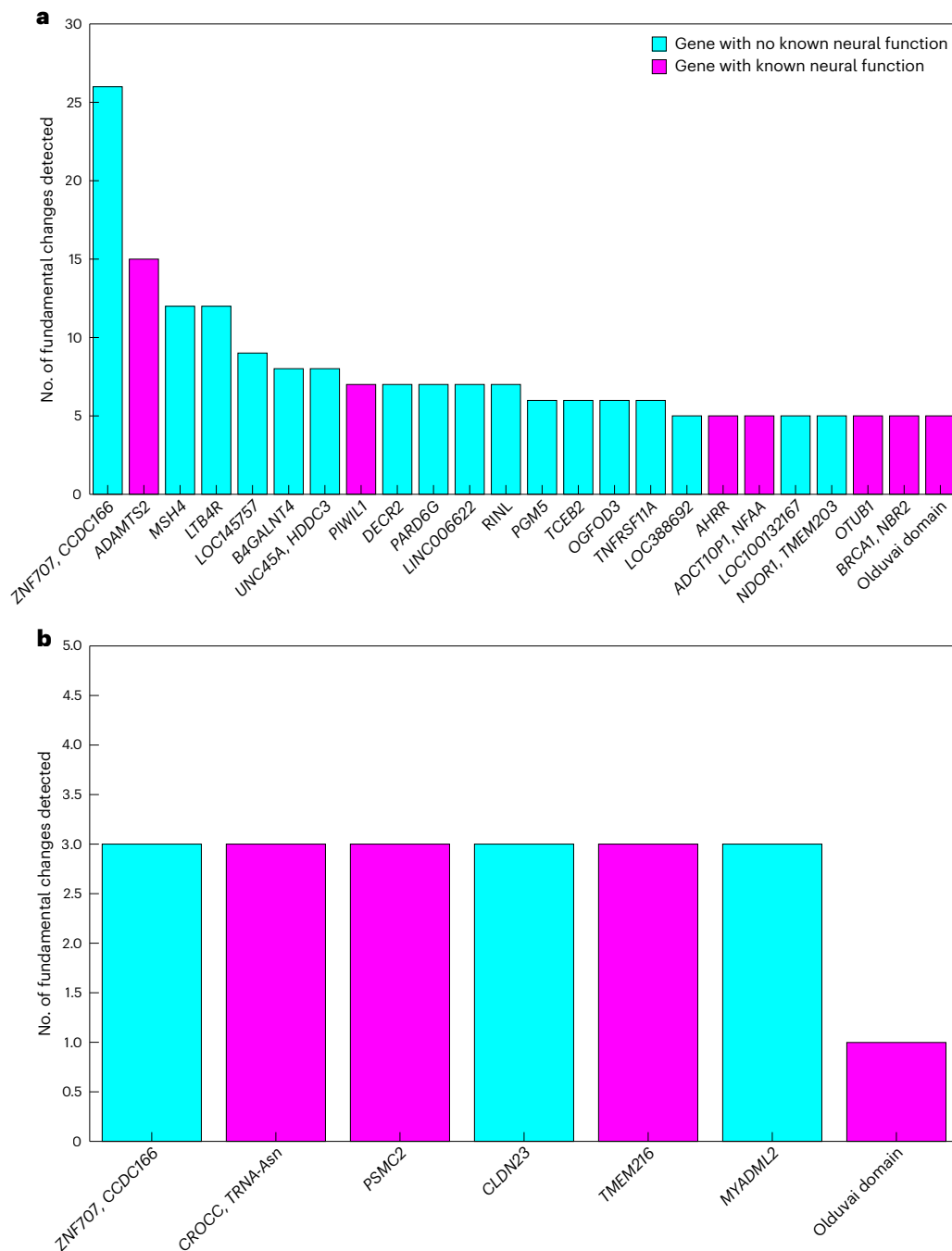
We focused on genes associated with a particularly large number of positions with fundamental changes (Fig. 4 and Supplementary Table 6). In this context, we identified 26 outgroup-specific changes and 3 reference-derived changes localized in the locus containing the genes *ZNF707* and *CCDC166*. Notably, our knowledge about the functions of these genes is limited, with scant evidence pointing at potential relevance to cancer<sup>25,26</sup>. Nonetheless, given its classification as a zinc-finger protein, it is reasonable to hypothesize that *ZNF707* probably possesses an as-yet undiscovered regulatory role.

Next in terms of the number of fundamental changes is the gene *ADAMTS2*, which exhibits 15 outgroup-specific changes. This gene product regulates neuronal migration by cleaving reelin<sup>27</sup>, and has been implicated in schizophrenia because it activates dopaminergic signaling and exhibits overexpression in the blood of people during episodes of psychosis (a phenomenon that can be reversed by antipsychotic medications)<sup>28,29</sup>. Moreover, in the Human Phenotype Ontology<sup>30</sup>, *ADAMTS2* shows associations with language impairment, intellectual disability, microcephaly, neurodevelopmental delay and other brain-related terms. The gene *NF-YA* harbours five outgroup-specific changes. *NF-YA* is a component of the NF-Y transcription factor, which protects neurons from cell death<sup>31</sup>. Inactivation of *NF-Y* has been linked

to various forms of neuronal pathologies<sup>32,33</sup>. *PIWIL1* exhibits seven outgroup-specific fundamental changes and regulates neuronal polarization and migration<sup>34</sup>. *BRCA1* harbours five outgroup-specific fundamental changes. Apart from being one of the most well-studied tumour suppressors, *BRCA1* plays a role in the survival of neural progenitors<sup>35</sup>. It is also expressed in the hippocampus, and has been observed to be under-expressed in patients with Alzheimer's disease. Knocking down this gene in the dentate gyrus leads to learning and memory deficits, impairs synaptic plasticity and neuronal shrinkage<sup>36</sup>. Notably, hypomethylation in a specific region in *BRCA1*, which does not overlap any of the positions with fundamental changes, has been associated with Alzheimer's disease<sup>37</sup>. Moreover, a variant of *BRCA1* has been associated with intellectual disability<sup>38</sup>. *AHRR*, which also harbours five changes, represses aryl hydrocarbon receptor (AHR), an important mediator for many cellular events. The main hallmarks of brain ageing, including oxidative stress, neuroinflammation and neurogenesis, are affected by AHR, suggesting its function as a regulator for ageing in the brain, together with its effect on the nervous system development<sup>39,40</sup>. The deubiquitylase *OTUB1* also exhibit five changes encoding the gene *OTUB1*. This gene mediates neuronal survival after intracerebral haemorrhage<sup>41,42</sup>. The effects of this protein on neuroimmune response contribute to pathologies associated with brain diseases such as Parkinson's disease and multiple sclerosis<sup>43,44</sup>.

Turning our attention to genes associated with reference-derived changes, of eight such genes with three altered positions (the highest number of such alternations in a gene), three show a relationship to the brain. Mutations in *TMEM216* have been linked to syndromes that involve brain abnormalities<sup>45,46</sup>. *TMEM216* regulates ciliogenesis, a process with substantial importance in brain development, and disruptions in *TMEM216* function lead to impairments in various neurodevelopmental processes<sup>47</sup>. Another gene linked to ciliary function is *CROCC*. *CROCC* encodes rootletin, a protein involved in the formation of ciliary rootlets. Mutations in this gene in *Drosophila* have been associated with sensory deficits<sup>48</sup>. *PSMC2* is a component of the 26S proteasome complex that plays a role in various neural processes, including synaptic plasticity<sup>49,50</sup> and brain autophagy<sup>51</sup>. Notably, impairments in the 26S proteasome complex have been strongly associated with the presence of brain inclusions such as alpha-synuclein and tau tangles<sup>52</sup>.

Five CpG positions associated with outgroup-specific fundamental changes, and another one associated with reference-derived changes, are located in genes from the neuroblastoma breaking point family (NBPF) gene family, inside the LOC100288142 locus (Fig. 4). NBPF proteins are enriched with a protein domain called the Olduvai domain that is key to human brain evolution. This domain, which is found almost exclusively in NBPF proteins, went through a human-specific increase in copy number,



**Fig. 4 | Genes featuring multiple fundamental DNA methylation changes.**

**a**, Genes featuring more than four outgroup-specific fundamental changes. Instances in which changes overlap multiple genes are consolidated in the same column. Methylation alterations overlapping genes associated with the Olduvai

domain (*NBPF8*, *NBPF9* and *NBPF10*) are collectively assembled in the column furthest right. **b**, Genes featuring more than two reference-derived fundamental changes, along with the reference-derived change that intersects with *NBPF9*, which harbours the Olduvai domain.

reaching approximately 300 copies. Chimpanzees show less than half that number of copies, and the number drops to several dozens of copies in other Old World monkeys, and to around a single copy in non-primate mammals. The domain is absent in non-mammals<sup>53</sup>. Olduvai domain copy number variation is correlated with brain volume<sup>54,55</sup>, cognitive functions<sup>56</sup> and autistic symptoms<sup>57–59</sup>. It is also negatively correlated with schizophrenia<sup>60</sup>. *NBPF* genes in general have been found to enhance neural stem cell proliferation<sup>61</sup>. Four of the five outgroup-specific fundamental changes are located in *NBPF10*, one of the *NBPF* genes with the highest number of human-specific Olduvai domain copies. *NBPF10* is adjacent to *NOTCH2NL*, an additional gene linked to human brain volume<sup>62,63</sup>, and it has been suggested that the two evolved in tandem<sup>64</sup>.

To further explore potential regulatory effects of the differential methylation, we examined whether CpG positions with fundamental methylation changes tend to overlap enhancers associated with bones and neurons more than would be expected by chance (Methods). In bones, we observed a significant overlap for both outgroup-specific (298 intersecting positions;  $P = 9.04 \times 10^{-10}$ ) and reference-derived CpG positions (28 intersecting positions;  $P = 2.66 \times 10^{-4}$ ). In neurons, significant overlap was only evident for outgroup-specific CpG positions (81 intersecting positions;  $P = 5.37 \times 10^{-7}$ ), whereas no significant overlap was observed for reference-derived CpG positions (4 intersecting positions). Next, we examined genes that are associated with enhancers that overlap CpG positions with differential methylation. Although

there are no such genes for reference-derived CpG positions, there are 330 and 480 genes for outgroup-specific CpG positions in bones and neurons, respectively. Enrichment analysis of these genes revealed similar patterns, despite representing two distinct tissues, highlighting organelle functions over tissue-specific ones (Supplementary Table 7). This aligns with expectations, considering that fundamental changes affect multiple tissues.

We next searched for motifs that overlap these enhancers and transcription factors that bind these motifs (Supplementary Table 8). In outgroup-specific changes we found enrichment of transcription factors that are related to brain development (GLI1, GLI2, ZIC1, NHLH2, NR2F2, ZIC2 and ZNF148;  $FDR = 2.3 \times 10^{-3}$ ) and specifically forebrain development (GLI1, GLI2, ZIC1, NHLH2 and NR2F2;  $FDR = 0.01$ ).

### Application to bulk prefrontal cortex

In addition to prefrontal neurons, DNA methylation is also available from bulk prefrontal cortex tissue from humans, chimpanzees and rhesus macaque (Methods and Extended Data Fig. 4)<sup>65,66</sup>. To examine whether the methylation changes we identified in neurons can be detected in whole tissues, we repeated the analysis with these bulk DNA methylation measurements.

We first evaluated the performance of our algorithm using the same triad of extant species as we used before, setting modern human as 'reference', chimpanzee as 'target' and rhesus macaque as 'outgroup', with bone serving as the tissue 'all' and bulk prefrontal cortex as the tissue 'partial'. Unfortunately, there is a small number of samples of DNA methylation in bulk prefrontal cortex (three humans, three chimpanzees and two rhesus macaques) with relatively lower coverage (Supplementary Table 4), compromising the power of our approach to detect differentially methylated positions. Indeed, although the precision we achieved is significantly higher than for permuted data ( $P < 10^{-4}$ ), we observed lower values than in previous analyses. We obtained a precision of 0.66 in predicting reference-derived changes (compared with a mean of 0.32 for permuted data) and a precision of 0.55 in predicting outgroup-specific changes (compared with a mean of 0.36 for permuted data) (Fig. 3b).

We then considered archaic humans as 'target' and chimpanzees as 'outgroup' and applied our algorithm to predict differential methylation in bulk brain tissue. We found 396 CpG positions associated with putative outgroup-specific changes and 53 CpG positions associated with putative reference-derived changes. The outgroup-specific CpG positions are associated with 248 genes and the reference-derived CpG positions are associated with 35 genes (Supplementary Table 9). Reassuringly, we found high consistency between these CpG positions and those identified using prefrontal neurons. In total, 156 of the 396 of the outgroup-specific CpG positions appear among the 1,750 positions identified using prefrontal neurons ( $P < 10^{-300}$ , hypergeometric test) and 8 of the 53 of the reference-derived CpG positions appear among the 122 positions identified using prefrontal neurons ( $P = 4.7 \times 10^{-44}$ ). Similarly, of the 248 genes associated with outgroup-specific changes in bulk prefrontal cortex, 116 were found in the corresponding list of 870 genes identified using prefrontal neurons ( $P < 10^{-100}$ ) and 8 of the 35 genes associated with reference-derived changes were also identified using neurons ( $P = 3.54 \times 10^{-16}$ ).

### Discussion

Based on the embryonic developmental timing of an evolutionary methylation change, we defined two families of changes; fundamental changes are those that occurred before the developmental split between a pair of tissues, whereas tissue-specific changes are those that occurred after the split. In this work, we developed an algorithm to detect CpG positions that underwent fundamental changes, because they result in DNA methylation differences across the examined species that are also replicated in the tissue 'partial' (Fig. 1c). Similar methodology can be used to detect tissue-specific changes as well, although these

result in no change in DNA methylation across the species in the tissue 'partial' (Fig. 1c), hence potentially presenting more limited evolutionary interest. The main algorithmic modification that is required would be to use statistical tests for equivalence, such as the two one-sided t-tests (TOST) procedure<sup>67</sup>.

The algorithm can be used to determine the methylation state in non-skeletal tissues of archaic humans, as long as methylation data are available for this tissue in both modern humans and chimpanzees (or any other non-human ape, in the case that bone methylation data are available for this species as well). This ability to obtain information on DNA methylation in archaic tissues that are not accessible in the palaeontological record opens new ways to examine changes in gene regulation and their potential effect on evolutionary adaptations in humans.

To quantify performance, we tested the algorithm on three extant species, comprising modern humans, chimpanzees and rhesus macaques (Fig. 2a). The longer branches on this phylogenetic tree (median divergence time between modern humans and rhesus macaques is 28.9 million years ago (Ma), compared with the median divergence time between modern humans and chimpanzees of 6.4 Ma (ref. 68)) result in an underestimation of the performance of the algorithm, because the likelihood of independent reversal of the methylation state along each branch increases. We therefore estimate that our algorithm works with even higher precisions than those reported here. We used the algorithm to find genomic positions in which the methylation state can be determined in the brains of archaic humans. We focused on the brain because of its central role in human evolution and the potential implications of brain-related differences between human groups, and between humans and chimpanzees. However, the algorithm we present is general and can be used to determine the methylation state in other archaic tissues. In fact, brain is probably among the most challenging choices, given its large developmental distance from bone (Fig. 1a). A tissue that is developmentally closer to bone, such as muscle or heart, is likely to result in a larger number of positions whose methylation state can be determined.

We showed that we are able to detect two types of fundamental DNA methylation changes. Reference-derived changes are those in which DNA methylation levels in modern humans are different from those in both archaic humans and chimpanzees, representing methylation patterns that are unique to modern humans. Outgroup-specific changes are those in which DNA methylation levels in both modern and archaic humans are different from those in chimpanzees, representing methylation patterns that are shared across modern and archaic humans. In all analyses, we detect fewer reference-derived positions than outgroup-specific ones, although with higher precision. This is expected considering the much shorter evolutionary time span in which the methylation change could have occurred for reference-derived changes compared with outgroup-specific ones (Fig. 1b). Shorter distances mean fewer methylation change events, which result in lower numbers of detected events, but also in lower probability for reversal of methylation changes, and therefore higher precision of parsimony-based inference.

We present here evidence for a strong connection between genes associated with fundamental changes and neuronal functions. Among the most interesting findings is the identification of six fundamental changes in NBPF genes carrying Olduvai domains, which seem to be associated with human brain evolution. Whereas the association of Olduvai domains to brain evolution has focused thus far on their copy number, our findings suggest that methylation changes might also have a role in this; in particular, *NBPF10*, has four fundamental methylation changes and is also one of the genes with the largest number of Olduvai domains. With that, future research is imperative to establish causality and validate the functional importance of these methylation changes on phenotypes.

The number of CpG positions that can be tested for differential methylation is constrained by the limited availability of DNA

methylation data from primate bone, because part of the data was produced using Illumina Infinium MethylationEPIC BeadChip methylation array (Supplementary Table 4), which represent approximately 3% of (but enriched for functional) CpG positions in the genome. As a result, our set of CpG positions associated with fundamental changes represents only a subset of the positions that show such an association. In addition, restrictions come from the fact that we considered a CpG position only if data were available in at least two samples from each group. This criterion resulted in a restricted set of CpGs available for analysis, ranging from 1% to 17% of autosomal CpGs, depending on the tissues selected for each analysis. These limitations underscore the need for an expanded dataset of bone methylation.

For reference-derived changes, we require that the mean methylation in the 'target' will be closer to the mean methylation of the 'outgroup' than to that of the 'reference', whereas for outgroup-specific changes we require the opposite. Addition of this requirement was important for reducing the level of false predictions, which result from the small number of samples in the 'target' compared with the other groups. However, when this additional requirement is used, the algorithm is unable to detect CpG positions that have gone through gradual methylation change (for example, that the 'outgroup' is unmethylated, the 'target' is partially methylated and the 'reference' is highly methylated). Such positions could potentially be detected as being simultaneously reference-derived and outgroup-specific methylation changes. However, mean methylation in the 'target' group will be closer to either the 'outgroup' or the 'reference'. Therefore, the additional requirement will lead the algorithm to detect only one of the two types of fundamental changes, and information on gradual changes will be lost. We hope that in the future more methylation maps of archaic humans will be published, making use of the additional requirement unnecessary, and allowing for the detection of gradual fundamental methylation changes.

Overall, we present an algorithm to infer DNA methylation in non-skeletal archaic tissues. The algorithm can be applied to any triad of species or groups and could be useful whenever DNA methylation in a certain tissue is unavailable or difficult to obtain.

## Methods

### DNA methylation data

We collected and generated DNA methylation data from seven tissues and cell types (neurons, bone, kidney, heart, liver, lung and bulk brain) across three species (modern human, chimpanzee and rhesus macaque). In addition, we used reconstructed DNA methylation data in bones of modern and archaic humans. Data were collected only from autosomes, to avoid the unique methylation patterns that characterize sex chromosomes.

### Bone DNA methylation of present-day individuals published in this study.

DNA methylation maps from four femur head bones from present-day humans were generated using whole-genome bisulfite sequencing (WGBS). Patients were females with osteoarthritis. Samples were extracted during total hip replacement surgery and were taken from the healthy part of the bone. All four patients signed a form according to Helsinki approval 0178-13-HMO. DNA extraction and the WGBS protocol are described in a previous publication<sup>9</sup> (Supplementary Table 4). In brief, DNA was extracted from bones using a QIAamp DNA Investigator kit (Qiagen, catalogue no. 56504) and DNA libraries were built with the Illumina TruSeq Sample Preparation kit. Bisulfite treatment was applied in two rounds using the EpiTect Bisulfite kit (Qiagen) and paired-end sequencing was performed on an Illumina Hi-Seq 2000 instrument. We aligned the reads to hg19 genome assembly using Bismark v.0.23.0 with the following parameters: `--bowtie2--non_bs_mm-p 4 --multicore 4`. We then used `bismark_methylation_extractor` to extract methylation calls with the following parameters: `-p --parallel 4 --bedGraph`. We did not perform filtration of CpG positions by coverage,

as we used the default Bismark setting of including positions with a minimum coverage of 1×.

**Previously published modern DNA methylation.** Neuronal data include WGBS DNA methylation data measured in neurons isolated from prefrontal cortices of humans, chimpanzees and macaques. Data for humans were downloaded from Gene Expression Omnibus (GEO) accession number [GSE107638](#) (ref. 20), taking only the 25 healthy controls. Data for chimpanzees and macaques were downloaded from GEO accession number [GSE151768](#) (ref. 21). WGBS DNA methylation data for kidneys, hearts, livers and lungs of four specimens for each of these species were downloaded from GEO accession number [GSE112356](#) (refs. 69,70). Bulk WGBS brain data that were generated from prefrontal cortices of three modern humans and three chimpanzees was downloaded from GEO accession number [GSE37202](#) (ref. 66). Bulk WGBS data of two macaques were downloaded from GEO accession number [GSE77124](#) (ref. 65). For bone, we used ten Illumina Infinium MethylationEPIC BeadChip (850K) methylation arrays from rhesus<sup>71</sup> and data acquired by Gokhman et al.<sup>9</sup> including one modern human and one chimpanzee WGBS map, one chimpanzee reduced representation bisulfite sequencing map and four chimpanzee 850K methylation arrays (Supplementary Table 4).

Most non-human DNA methylation data were already mapped to the human hg19 reference genome. However, this was not done for rhesus bulk brain samples, and we therefore mapped these data to hg19 using `liftOver`<sup>72</sup>. This mapped ~30% of the rhesus CpG positions to a corresponding CpG position in human. However, in a significantly enriched number of cases, the rhesus CpG positions were mapped to a position preceding a human CpG position by a single base (Extended Data Fig. 4). We considered these as valid mappings, ending up with ~40% of the rhesus CpG positions being successfully mapped to the human hg19 genome reference.

**Ancient DNA methylation maps.** We used our newest version of the RoAM software tool<sup>24</sup> to reconstruct the methylome of a Mesolithic anatomically modern human individual from Stora Karlsö, Sweden (~9,000 years ago) that had been previously sequenced<sup>19</sup> (Supplementary Table 4). In addition, we used our previously published methylation maps of five ancient anatomically modern humans<sup>9</sup>, one Neanderthal and one Denisovan<sup>8</sup>, which at the time of writing are the only uracil DNA glycosylase-treated high-coverage archaic samples, two prerequisites for computational reconstruction of aDNA methylation<sup>17</sup>.

**Batch effects.** To minimize batch effects stemming from combining data from different sources, we took several measures. First, all data used for optimization and cross-validation were taken from the same laboratory<sup>69,70</sup>. Similarly, all the brain data, including both prefrontal cortex neurons and bulk brain tissue, were taken from the same laboratory<sup>20,21,65,66</sup>.

Most of the bone data were produced by us, either specifically for this paper or in previous work<sup>9</sup>. To address consistency between the modern bone samples and the ancient ones, we used histogram matching during the DNA methylation reconstruction, fitting the reconstructed histogram to that of previously published modern bone data<sup>9,73</sup>.

Finally, remaining batch effects are accounted for in the estimated false discovery rates of our method.

### Detecting CpGs positions with differential DNA methylation

The algorithm receives methylation data from three species and two tissues (Fig. 1b). Only CpG positions where, for each tissue and species, data are available from at least two samples were considered. In addition, only positions where the variance in methylation levels in each tissue and species was <0.08 were considered. This threshold was selected to guarantee that a minimum set of two samples will have at most a 40% difference in methylation level.



**Identifying reference-derived fundamental changes.** Let us have methylation data for  $N$  CpG positions and let  $0 \leq m_{i,j}^{s,u} \leq 1$  be the measured methylation level of sample  $j$  at CpG position  $i$  ( $i = 1, \dots, N$ ), in species  $s$  and tissue  $u$ . We say that the methylation level at CpG position  $i$  is reference-derived in tissue ‘all’ if

$$|t_i^{r,a}| \geq t^{r,a} + k_a^{rd} \sigma^{r,a}.$$

Here,  $t_i^{r,a}$  is the statistic of an equal-variance  $t$ -test at CpG position  $i$  in tissue ‘all’, contrasting the ‘reference’ with the ‘target’ and ‘outgroup’, and  $t^{r,a}$  and  $\sigma^{r,a}$  are the estimators of the mean and the standard deviation of the statistic across the  $N$  positions.  $k_a^{rd}$  is a parameter.

Similarly, we say that CpG position  $i$  shows differential methylation between the ‘reference’ and the ‘outgroup’ in tissue ‘partial’ if

$$|t_i^{r,o,p}| \geq t^{r,o,p} + k_p^{rd} \sigma^{r,o,p}.$$

Here,  $t_i^{r,o,p}$  is the statistic of an equal-variance  $t$ -test at CpG position  $i$  in tissue ‘partial’, contrasting the ‘reference’ with the ‘outgroup’, and  $t^{r,o,p}$  and  $\sigma^{r,o,p}$  are the estimators of the mean and the standard deviation of the statistic across the  $N$  positions.  $k_p^{rd}$  is a parameter.

To identify reference-derived fundamental changes (left-hand column in Fig. 1c), we look for CpG positions  $i$  that satisfy:

$$|t_i^{r,a}| \geq t^{r,a} + k_a^{rd} \sigma^{r,a} \tag{1a}$$

$$|t_i^{r,o,p}| \geq t^{r,o,p} + k_p^{rd} \sigma^{r,o,p}, \tag{1b}$$

$$t_i^{r,a} \cdot t_i^{r,o,p} > 0. \tag{1c}$$

Inequality (equation (1c)) guarantees that the methylation difference between the ‘reference’ and the ‘outgroup’ in both tissues is in the same direction.

We found it useful to add a fourth condition to equation (1a–c), verifying that the mean methylation in ‘target’ is closer to the mean methylation of the ‘outgroup’ than to the mean methylation of the reference. If  $\overline{m_i^{r,a}}$ ,  $\overline{m_i^{t,a}}$  and  $\overline{m_i^{o,a}}$  are the mean methylation across samples in tissue ‘all’ in position  $i$  of the ‘reference’, ‘target’ and ‘outgroup’, respectively, then

$$\left| \overline{m_i^{t,a}} - \overline{m_i^{r,a}} \right| > \left| \overline{m_i^{o,a}} - \overline{m_i^{r,a}} \right|. \tag{1d}$$

In the implementation, we kept this additional condition optional because the algorithm detects methylation changes in high precision even without it. However, it is advisable to use it when the number of samples in ‘target’ is small compared with the other groups, because it prevents mispredictions that arise as a result of the lower weight of the ‘target’ samples in the  $t$ -test. Because only two archaic human samples serve as ‘target’ in the current study, we used this additional condition throughout the paper. Positions that satisfy equation (1a–d) are those for which we predict that the methylation in ‘target’ in tissue ‘partial’ clusters with that in ‘outgroup’, and different from that in ‘reference’ (Fig. 1c).

**Identifying outgroup-specific fundamental changes.** We follow a similar approach to identify fundamental changes in which the methylation change in tissue ‘all’ is outgroup-specific (right-hand column in Fig. 1c). CpG positions that we predict have gone through outgroup-specific fundamental changes should satisfy:

$$|t_i^{o,a}| \geq t^{o,a} + k_a^{os} \sigma^{o,a} \tag{2a}$$

$$|t_i^{o,p}| \geq t^{o,p} + k_p^{os} \sigma^{o,p}, \tag{2b}$$

$$t_i^{o,a} \cdot t_i^{r,o,p} > 0. \tag{2c}$$

Here,  $t_i^{o,a}$  is the statistic of an equal-variance  $t$ -test at CpG position  $i$  in tissue ‘all’ contrasting the ‘outgroup’ with the ‘target’ and ‘reference’,  $t^{o,a}$  and  $\sigma^{o,a}$  are the estimators of the mean and the standard deviation of the statistic across the  $N$  positions, and  $k_a^{os}$  and  $k_p^{os}$  are parameters. Note that the  $t$ -test for tissue ‘partial’, equation (2b), is identical to the one for the reference-derived fundamental changes, equation (1b), because in both cases we wish to find a significant difference in tissue ‘partial’ between the ‘outgroup’ and the ‘reference’. Yet, we allow the parameters  $k_p^{rd}$  and  $k_p^{os}$  to be potentially different.

Here, too, we added a fourth condition that guarantees that the mean methylation in ‘target’ is closer to ‘reference’ than to ‘outgroup’,

$$\left| \overline{m_i^{t,a}} - \overline{m_i^{o,a}} \right| > \left| \overline{m_i^{r,a}} - \overline{m_i^{o,a}} \right|. \tag{2d}$$

In the implementation, this condition is optional, but we used it throughout this work.

**Measuring performance.** Whenever the true value of the methylation in the ‘target’ at tissue ‘partial’ is known, we can use it to estimate the precision of our algorithm. To this end, we define CpG positions for which there is truly a reference-derived methylation change in tissue ‘partial’ as those that satisfy

$$|t_i^{r,p}| \geq t^{r,p} + k_t^{rd} \sigma^{r,p}. \tag{3}$$

Here,  $t_i^{r,p}$  is the statistic of an equal-variance  $t$ -test at CpG position  $i$  in tissue ‘partial’, contrasting the ‘reference’ with the ‘target’ and ‘outgroup’,  $t^{r,p}$  and  $\sigma^{r,p}$  are the estimators of the mean and the standard deviation of the statistic across the  $N$  positions, and  $k_t^{rd}$  is a parameter. Similarly, we define CpG positions for which there is truly an outgroup-specific methylation change in tissue ‘partial’ as those that satisfy

$$|t_i^{o,p}| \geq t^{o,p} + k_t^{os} \sigma^{o,p}, \tag{4}$$

where  $t_i^{o,p}$  is the statistic of an equal-variance  $t$ -test at CpG position  $i$  in tissue ‘partial’, contrasting the ‘outgroup’ with the ‘target’ and ‘reference’,  $t^{o,p}$  and  $\sigma^{o,p}$  are the estimators of the mean and the standard deviation of the statistic across the  $N$  positions, and  $k_t^{os}$  is a parameter.

Let  $S_p$  be the set of CpG positions for which we predict a reference-derived fundamental change, namely, those positions that satisfy equation (1a–d). Let  $S_e$  be the set of reference-derived fundamental changes that are considered ‘true’ changes, defined as the collection of positions that satisfy

$$|t_i^{r,a}| \geq t^{r,a} + k_a^{rd} \sigma^{r,a} \tag{5a}$$

$$|t_i^{r,p}| \geq t^{r,p} + k_t^{rd} \sigma^{r,p}, \tag{5b}$$

$$t_i^{r,a} \cdot t_i^{r,p} > 0, \tag{5c}$$

where equation (5a) is identical to equation (1a) and (5b) is identical to equation (3). Then, the precision of our algorithm that predicts reference-derived fundamental changes is

$$\text{Pr} = \frac{|S_p \cap S_e|}{|S_p|}, \tag{6}$$

where  $|S|$  is the size of set  $S$ .

Analogously, for outgroup-specific fundamental changes we define  $S_p$  as the set of positions that satisfy equation (2a–d), and  $S_e$  as

the set of positions that are taken as ‘true’ outgroup-specific fundamental change, namely the set of positions that satisfy

$$|\ell_i^{o,a}| \geq \ell^{o,a} + k_a^{os} \sigma^{o,a}, \quad (7a)$$

$$|\ell_i^{o,p}| \geq \ell^{o,p} + k_p^{os} \sigma^{o,p}, \quad (7b)$$

$$\ell_i^{o,a} \cdot \ell_i^{o,p} > 0, \quad (7c)$$

where equation (7a) is identical to equation (2a) and (7b) is identical to equation (4). Precision is defined as in equation (6).

**Parameter estimation.** Overall, the algorithms that predict reference-derived changes and outgroup-specific changes are independent of each other, and each respective set of parameters can be optimized separately. Each algorithm uses two parameters,  $k_a^{rd}$  and  $k_p^{rd}$ , for predicting reference-derived changes (equation (1a–d)), and  $k_a^{os}$  and  $k_p^{os}$  for predicting outgroup-specific changes (equation (2a–d)). Moreover, the performance of each algorithm is determined using an additional parameter, either  $k_i^{rd}$  for the reference-derived changes or  $k_i^{os}$  for the outgroup-specific changes. All three parameters for each algorithm were optimized together. Permuted data were generated by pairing a random CpG position in tissue ‘partial’ to each position in tissue ‘all’. For each analysis, 10,000 permuted datasets were generated. The optimized parameters were chosen as the values that gave the largest difference between the precision of the algorithm on the real data and its precision on permuted data.

We performed the optimization process by using chimpanzee as ‘target’, modern humans as ‘reference’ and rhesus macaque as ‘outgroup’ and choosing heart as tissue ‘all’ and kidney as tissue ‘partial’. We scanned all values of the parameters in a grid of values from 2 to 4 using linear spacing of 0.2 for each parameter. Notably, for each value of the parameters, both in the reference-derived and the outgroup-specific comparisons, the precision of the algorithm when applied to the real data was higher than the average precision of the corresponding permuted data (Extended Data Fig. 1). Moreover, not a single permutation in any set of parameters achieved precision that exceeded that of the real data. The optimized values of the parameters are given in Supplementary Table 1.

Some of the optimized parameters take their value at the edge of the grid, pointing at the possibility of achieving even higher precision using values that are outside the current grid. However, we noticed that in these cases precision reaches a plateau, so that the improvement is minor when changing the value of the parameter. Combined with the fact that values of the parameters outside the grid decreased the number of detections and hence reduced the power of the analysis, we decided to bound the value of the parameters and not extend the grid (Extended Data Fig. 2). The set of optimized parameters is remarkably robust to the selection of the training dataset and changed only slightly when using other tissue combinations (Supplementary Table 2).

### Smoothing data

Smoothing of methylation data in modern samples was performed in the cross-validation part only. It was done using a moving average with a fixed window size over CpG positions. We used a window size of 31 CpG positions, to match typical window sizes used for reconstructing ancient methylation<sup>8,9</sup>.

### Gene enrichment analysis

A CpG position with a predicted fundamental change is associated with a gene if it resides in the gene body or in the promoter region, defined as 5,000 bp upstream of the transcription start site to 1,000 bp downstream of it. Gene body coordinates and gene names were determined using the hg19 genome reference downloaded from UCSC genome assembly<sup>74</sup>.

We used DAVID<sup>75,76</sup> to test for enrichment of our gene lists in biological functions. We used all default datasets.

### Enhancer enrichment analysis

Enhancer locations were downloaded from EnhancerAtlas<sup>77</sup>. For bone enhancers we used the osteoblasts datasets. For neuron enhancers we combined data of embryonic stem cell neurons and cerebellum neurons. Enrichment was evaluated using a hypergeometric test followed by Benjamini–Hochberg FDR procedure. EnhancerAtlas includes a comprehensive list of genes associated with each enhancer. We used DAVID to investigate functional enrichments for genes linked to enhancers that intersect with the detected positions. Further, we used MEME-ChIP<sup>78</sup> to examine which of the positions that overlap enhancers also overlap binding motifs, and then used TOMTOM<sup>79</sup> to identify transcription factors that bind these motifs.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

All sequencing and methylation data generated for this study was deposited in the National Center for Biotechnology Information’s Gene Expression Omnibus under GEO access number [GSE276666](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE276666).

### Code availability

The MATLAB code can be downloaded from <http://carmelab.huji.ac.il/software.html>.

### References

- King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Romero, I. G., Ruvinsky, I. & Gilad, Y. Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genet.* **13**, 505–516 (2012).
- Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
- Carroll, S. B. Evo-Devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
- Colbran, L. L. et al. Inferred divergent gene regulation in archaic hominins reveals potential phenotypic differences. *Nat. Ecol. Evol.* **3**, 1598–1606 (2019).
- Fordyce, S. L. et al. Deep sequencing of RNA from ancient maize kernels. *PLoS ONE* **8**, e50961 (2013).
- Pedersen, J. S. et al. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* **24**, 454–466 (2014).
- Gokhman, D. et al. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* **344**, 523–527 (2014).
- Gokhman, D. et al. Differential DNA methylation of vocal and facial anatomy genes in modern humans. *Nat. Commun.* **11**, 1189 (2020).
- Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).
- Mathov, Y., Batyrev, D., Meshorer, E. & Carmel, L. Harnessing epigenetics to study human evolution. *Curr. Opin. Genet. Dev.* **62**, 23–29 (2020).
- Gokhman, D. et al. Reconstructing Denisovan anatomy using DNA methylation maps. *Cell* **179**, 180–192.e10 (2019).
- Pai, A. A., Bell, J. T., Marioni, J. C., Pritchard, J. K. & Gilad, Y. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet.* **7**, e1001316 (2011).

14. Böck, J. et al. Cell type and species-specific patterns in neuronal and non-neuronal methylomes of human and chimpanzee cortices. *Cereb. Cortex* **28**, 3724–3739 (2018).
15. Palkopoulou, E. et al. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* **25**, 1395–1400 (2015).
16. Seguin-Orlando, A. et al. Pros and cons of methylation-based enrichment methods for ancient DNA. *Sci. Rep.* **5**, 11826 (2015).
17. Gokhman, D., Meshorer, E. & Carmel, L. Epigenetics: it's getting old. Past meets future in paleoepigenetics. *Trends Ecol. Evol.* **31**, 290–300 (2016).
18. Morgan, H. D., Santos, F., Green, K., Dean, W. & Reik, W. Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* **14**, R47–R58 (2005).
19. Günther, T. et al. Population genomics of Mesolithic Scandinavia: investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol.* **16**, e2003703 (2018).
20. Berto, S. et al. Accelerated evolution of oligodendrocytes in the human brain. *Proc. Natl Acad. Sci. USA* **116**, 24334–24342 (2019).
21. Jeong, H. et al. Evolution of DNA methylation in the human brain. *Nat. Commun.* **12**, 2021 (2021).
22. Zhang, W., Spector, T. D., Deloukas, P., Bell, J. T. & Engelhardt, B. E. Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements. *Genome Biol.* **16**, 14 (2015).
23. Affinito, O. et al. Nucleotide distance influences co-methylation between nearby CpG sites. *Genomics* **112**, 144–150 (2020).
24. Mathov, Y. et al. RoAM: computational reconstruction of ancient methylomes and identification of differentially methylated regions. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.08.08.607143> (2024).
25. Chen, S. et al. A novel DNA methylation signature associated with lymph node metastasis status in early gastric cancer. *Clin. Epigenetics* **14**, 18 (2022).
26. Kim, J. et al. Identification of candidate genes associated with susceptibility to ovarian clear cell adenocarcinoma using *cis*-eQTL analysis. *J. Clin. Med.* **9**, 1137 (2020).
27. Yamakage, Y. et al. A disintegrin and metalloproteinase with thrombospondin motifs 2 cleaves and inactivates Reelin in the postnatal cerebral cortex and hippocampus, but not in the cerebellum. *Mol. Cell. Neurosci.* **100**, 103401 (2019).
28. Crespo-Facorro, B., Prieto, C. & Sainz, J. Schizophrenia gene expression profile reverted to normal levels by antipsychotics. *Int. J. Neuropsychopharmacol.* **18**, pyu066 (2015).
29. Ruso-Julve, F. et al. Dopaminergic control of ADAMTS2 expression through cAMP/CREB and ERK: molecular effects of antipsychotics. *Transl. Psychiatry* **9**, 306 (2019).
30. Gargano, M. A. et al. The human phenotype ontology in 2024: phenotypes around the world. *Nucleic Acids Res.* **52**, D1333–D1346 (2024).
31. Yamanaka, T. et al. NF-Y inactivation causes atypical neurodegeneration characterized by ubiquitin and p62 accumulation and endoplasmic reticulum disorganization. *Nat. Commun.* **5**, 3354 (2014).
32. Yamanaka, T. et al. Differential roles of NF-Y transcription factor in ER chaperone expression and neuronal maintenance in the CNS. *Sci. Rep.* **6**, 34575 (2016).
33. Yamanaka, T. et al. Mutant Huntingtin reduces HSP70 expression through the sequestration of NF-Y transcription factor. *EMBO J.* **27**, 827–839 (2008).
34. Zhao, P. P. et al. Novel function of PIWIL1 in neuronal polarization and migration via regulation of microtubule-associated proteins. *Mol. Brain* **8**, 39 (2015).
35. Pao, G. M. et al. Role of BRCA1 in brain development. *Proc. Natl Acad. Sci. USA* **111**, E1240–E1248 (2014).
36. Suberbielle, E. et al. DNA repair factor BRCA1 depletion occurs in Alzheimer brains and impairs cognitive function in mice. *Nat. Commun.* **6**, 8897 (2015).
37. Mano, T. et al. Neuron-specific methylome analysis reveals epigenetic regulation and tau-related dysfunction of BRCA1 in Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **114**, E9645–E9654 (2017).
38. Aljouneh, A. et al. Dual BRCA1 and BRCA2 pathogenic variants in an adolescent with syndromic intellectual disability. *Clin. Case Rep.* **10**, e6202 (2022).
39. Gassmann, K. et al. Species-specific differential AhR expression protects human neural progenitor cells against developmental neurotoxicity of PAHs. *Environ. Health Perspect.* **118**, 1571–1577 (2010).
40. Ojo, E. S. & Tischkau, S. A. The role of AhR in the hallmarks of brain aging: friend and foe. *Cells* **10**, 2729 (2021).
41. Xie, L. et al. OTUB1 attenuates neuronal apoptosis after intracerebral hemorrhage. *Mol. Cell. Biochem.* **422**, 171–180 (2016).
42. Shen, J. et al. Deubiquitylating enzyme OTUB1 facilitates neuronal survival after intracerebral hemorrhage via inhibiting NF- $\kappa$ B-triggered apoptotic cascades. *Mol. Neurobiol.* **61**, 1726–1736 (2024).
43. Wang, X. et al. OTUB1 inhibits CNS autoimmunity by preventing IFN- $\gamma$ -induced hyperactivation of astrocytes. *EMBO J.* **38**, e100947 (2019).
44. Xing, X., Xu, F., Wang, Y. & Liu, H. Role of the OTUB1/IRF7/NOX4 axis in oxidative stress injury and inflammatory responses in mice with Parkinson's disease. *Psychogeriatrics* **23**, 32–44 (2023).
45. Edvardson, S. et al. Joubert Syndrome 2 (JBTS2) in Ashkenazi Jews is associated with a *TMEM216* mutation. *Am. J. Hum. Genet.* **86**, 93–97 (2010).
46. Valente, E. M. et al. Mutations in *TMEM216* perturb ciliogenesis and cause Joubert, Meckel and related syndromes. *Nat. Genet.* **42**, 619–625 (2010).
47. Guo, J. et al. Developmental disruptions underlying brain abnormalities in ciliopathies. *Nat. Commun.* **6**, 7857 (2015).
48. Chen, J. V. et al. Rootletin organizes the ciliary rootlet to achieve neuron sensory function in *Drosophila*. *J. Cell Biol.* **211**, 435–453 (2015).
49. Tai, H. C., Besche, H., Goldberg, A. L. & Schuman, E. M. Characterization of the brain 26S proteasome and its interacting proteins. *Front. Mol. Neurosci.* **3**, 1236 (2010).
50. Yi, J. J. & Ehlers, M. D. Emerging roles for ubiquitin and protein degradation in neuronal function. *Pharm. Rev.* **59**, 14–39 (2007).
51. Ugun-Klusek, A. et al. Continued 26S proteasome dysfunction in mouse brain cortical neurons impairs autophagy and the Keap1-Nrf2 oxidative defence pathway. *Cell Death Dis.* **8**, e2531–e2531 (2017).
52. Myeku, N. et al. Tau-driven 26S proteasome impairment and cognitive dysfunction can be prevented early in disease by activating cAMP-PKA signaling. *Nat. Med.* **22**, 46–53 (2015).
53. Popesco, M. C. et al. Human lineage-specific amplification, selection, and neuronal expression of DUF1220 domains. *Science* **313**, 1304–1307 (2006).
54. Zimmer, F. & Montgomery, S. H. Phylogenetic analysis supports a link between DUF1220 domain number and primate brain expansion. *Genome Biol. Evol.* **7**, 2083–2088 (2015).
55. Dumas, L. J. et al. DUF1220-domain copy number implicated in human brain-size pathology and evolution. *Am. J. Hum. Genet.* **91**, 444–454 (2012).
56. Davis, J. M. et al. DUF1220 copy number is linearly associated with increased cognitive function as measured by total IQ and mathematical aptitude scores. *Hum. Genet.* **134**, 67 (2015).

57. Davis, J. M. et al. DUF1220 dosage is linearly associated with increasing severity of the three primary symptoms of autism. *PLoS Genet.* **10**, 1004241 (2014).
58. Davis, J. M., Searles Quick, V. B. & Sikela, J. M. Replicated linear association between DUF1220 copy number and severity of social impairment in autism. *Hum. Genet.* **134**, 569 (2015).
59. Davis, J. M., Heft, I., Scherer, S. W. & Sikela, J. M. A third linear association between Olduvai (DUF1220) copy number and severity of the classic symptoms of inherited autism. *Am. J. Psychiatry* **176**, 643–650 (2019).
60. Searles Quick, V. B., Davis, J. M., Olincy, A. & Sikela, J. M. DUF1220 copy number is associated with schizophrenia risk and severity: implications for understanding autism and schizophrenia as related diseases. *Transl. Psychiatry* **5**, e697 (2015).
61. Keeney, J. G. et al. DUF1220 protein domains drive proliferation in human neural stem cells and are associated with increased cortical volume in anthropoid primates. *Brain Struct. Funct.* **220**, 3053–3060 (2014).
62. Fiddes, I. T. et al. Human-specific *NOTCH2NL* genes affect Notch signaling and cortical neurogenesis. *Cell* **173**, 1356–1369.e22 (2018).
63. Florio, M. et al. Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. *eLife* **7**, e32332 (2018).
64. Fiddes, I. T., Pollen, A. A., Davis, J. M. & Sikela, J. M. Paired involvement of human-specific Olduvai domains and *NOTCH2NL* genes in human brain evolution. *Hum. Genet.* **138**, 715–721 (2019).
65. Mendizabal, I. et al. Comparative methylome analyses identify epigenetic regulatory loci of human brain evolution. *Mol. Biol. Evol.* **33**, 2947–2959 (2016).
66. Zeng, J. et al. Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *Am. J. Hum. Genet.* **91**, 455–465 (2012).
67. Schuirman, D. J. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biopharm.* **15**, 657–680 (1987).
68. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
69. Pavlovic, B. J., Blake, L. E., Roux, J., Chavarria, C. & Gilad, Y. A comparative assessment of human and chimpanzee iPSC-derived cardiomyocytes with primary heart tissues. *Sci. Rep.* **8**, 15312 (2018).
70. Blake, L. E. et al. A comparison of gene expression and DNA methylation patterns across tissues and species. *Genome Res.* **30**, 250–262 (2020).
71. Housman, G., Quillen, E. E. & Stone, A. C. Intraspecific and interspecific investigations of skeletal DNA methylation and femur morphology in primates. *Am. J. Phys. Anthropol.* **173**, 34–49 (2020).
72. Hinrichs, A. S. et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
73. Barouch, A., Mathov, Y., Meshorer, E., Yakir, B. & Carmel, L. Reconstructing DNA methylation maps of ancient populations. *Nucleic Acids Res.* **52**, 1602–1612 (2024).
74. Karolchik, D. et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–D496 (2004).
75. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
76. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
77. Gao, T. & Qian, J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res.* **48**, D58–D64 (2020).
78. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
79. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. *Genome Biol.* **8**, R24 (2007).

## Acknowledgements

This publication was made possible through the support of a grant from the John Templeton Foundation (grant ID 61739 to L.C. and E.M.). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. This study was also funded by the Israel Science Foundation (grant no. 2436/22 to L.C.) and by the Ministry of Innovation, Science & Technology (grant no. grant 1001584586 to L.C. and E.M.). T.M.-B. is supported by funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 864203), PID2021-126004NB-I00 (MICIIN/FEDER, UE), 'Unidad de Excelencia María de Maeztu', funded by the AEI (CEX2018-000792-M), NIH 1R01HG010898-01A1 and Secretaria d'Universitats i Recerca and CERCA Programme del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2021 SGR 00177). L.C. is the Snyder Granadar chair in Genetics. E.M. is the Arthur Gutterman Family Chair for Stem Cell Research. We wish to thank Avigal Be'er for the design of Fig. 1.

## Author contributions

Y.M., E.M. and L.C. designed the research. Y.M. conducted the analysis. M.N.-R., C.L., T.M.-B., A.K. and M.L. contributed the production of data. N.G. helped collecting data. E.M. and L.C. supervised the study. Y.M., E.M. and L.C. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41559-024-02571-w>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41559-024-02571-w>.

**Correspondence and requests for materials** should be addressed to Eran Meshorer or Liran Carmel.

**Peer review information** *Nature Ecology & Evolution* thanks Alison Barton and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share

adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory

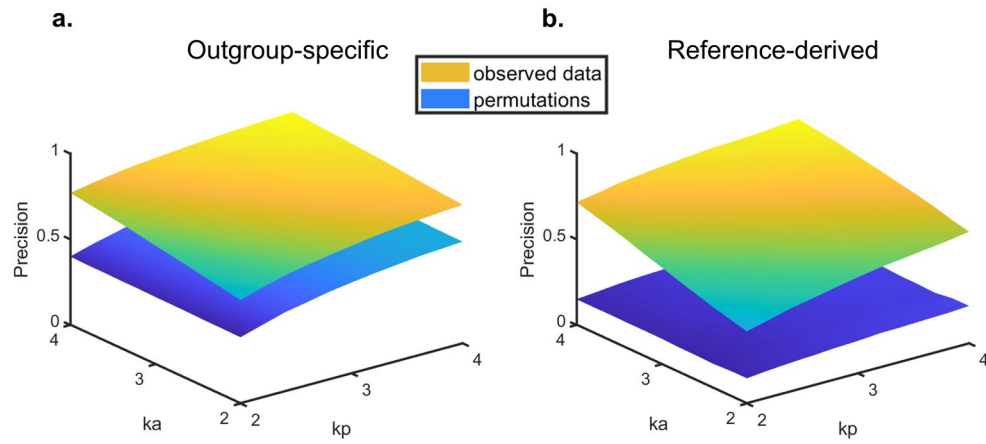
regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

---

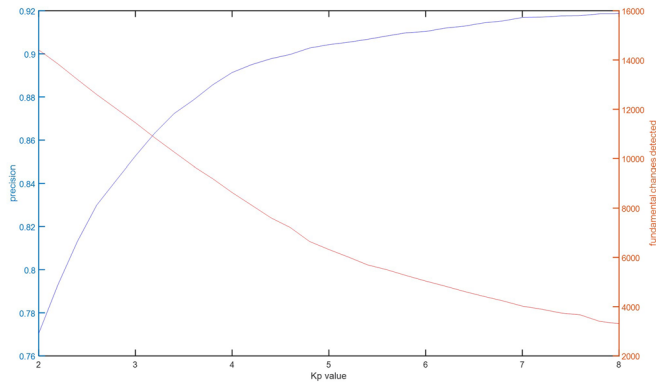
<sup>1</sup>Department of Genetics, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel. <sup>2</sup>Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Jerusalem, Israel. <sup>3</sup>Institute of Evolutionary Biology (UPF-CSIC), PRBB, Barcelona, Spain. <sup>4</sup>Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain. <sup>5</sup>CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. <sup>6</sup>Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Barcelona, Spain. <sup>7</sup>Orthopedic Department, Hadassah – Hebrew University Medical Center, Jerusalem, Israel.

✉ e-mail: [eran.meshorer@mail.huji.ac.il](mailto:eran.meshorer@mail.huji.ac.il); [liran.carmel@huji.ac.il](mailto:liran.carmel@huji.ac.il)



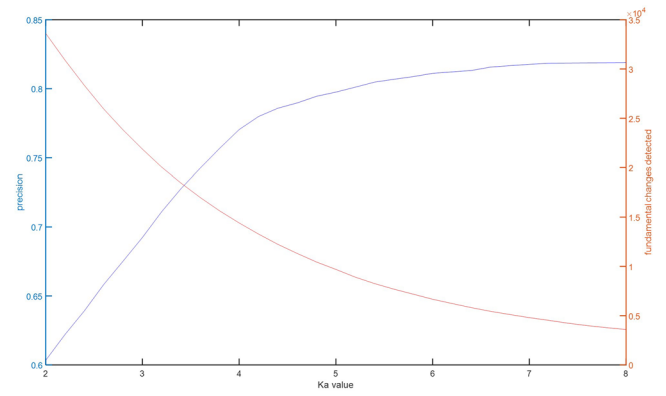
**Extended Data Fig. 1 | Precision on training set as a function of the parameters  $k_a$  and  $k_p$ .** Precision of the algorithm in detection of **a** outgroup-specific fundamental changes and **b** reference-derived fundamental changes, for each value of the parameters  $k_a$  and  $k_p$ . In yellow: the precision when running the algorithm on real data, in blue: the precision when running on permutations.

a.

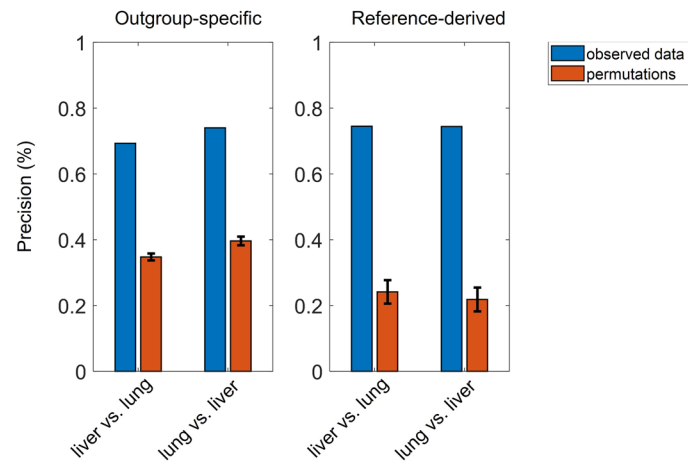


**Extended Data Fig. 2 | Precision versus number of detections.** Precision reaches plateau in higher values of the parameters, whereas power is decreasing. **a**, Precision and the number of outgroup-specific changes detected as a function

b.

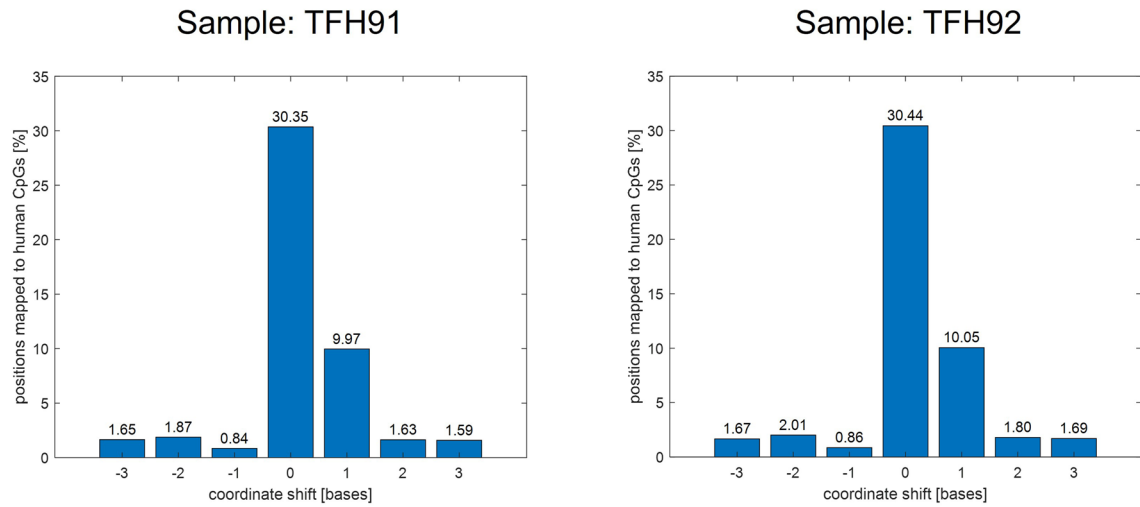


of the parameter  $K_p$  when the other parameters are fixed to the optimized value. **b**, A similar comparison for the parameter  $K_a$  while the rest of the parameters are fixed to the optimized value.



**Extended Data Fig. 3 | Precision on an independent dataset.** Validating the algorithm by measuring precision on independent dataset. To complement the results presented in the main text, here chimpanzees were selected as *reference* and modern humans as *target*.





**Extended Data Fig. 4 | Mapping of CpG positions in the rhesus genome to the human genome.** Percentage of CpG positions successfully mapped to a human CpG position from rhesus bulk brain samples. If a rhesus CpG position was

originally mapped by liftOver to coordinate  $c$ , the figure shows the percentage of positions where the human reference genome has a CpG at coordinate  $c + shift$ .

Corresponding author(s): \_\_\_\_\_

Last updated by author(s): YYYY-MM-DD \_\_\_\_\_

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Reported
Reporting on race, ethnicity, or other socially relevant groupings	irrelevant
Population characteristics	irrelevant
Recruitment	irrelevant
Ethics oversight	irrelevant

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences     Behavioural & social sciences     Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<i>Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.</i>
Data exclusions	<i>Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Replication	<i>Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.</i>
Randomization	<i>Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.</i>
Blinding	<i>Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.</i>

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<i>Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).</i>
Research sample	<i>State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.</i>
Sampling strategy	<i>Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.</i>
Data collection	<i>Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.</i>
Timing	<i>Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.</i>

Data exclusions	<i>If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Non-participation	<i>State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.</i>
Randomization	<i>If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.</i>

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Prediction of DNA methylation level in non skeletal samples of ancient DNA using parsimony.
Research sample	DNA methylation from rhesus macaques, chimpanzees,, archaic and modern human from 7 different tissues or cell types
Sampling strategy	All publically available DNA methylation maps
Data collection	All data are publically available. Modern human bones were collected under Helsinki approval 0178-13-HMO
Timing and spatial scale	Up to 28.9 million years
Data exclusions	No data were excluded
Reproducibility	Irrelevant
Randomization	Permutation tests and other tools described in Methods
Blinding	Irrelevant

Did the study involve field work?  Yes  No

## Field work, collection and transport

Field conditions	<i>Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).</i>
Location	<i>State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).</i>
Access & import/export	<i>Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).</i>
Disturbance	<i>Describe any disturbance caused by the study and how it was minimized.</i>

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.
Validation	Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.
Authentication	Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.
Mycoplasma contamination	Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

## Palaeontology and Archaeology

Specimen provenance	Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.
Specimen deposition	Indicate where the specimens have been deposited to permit free access by other researchers.
Dating methods	If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.
<input type="checkbox"/> Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.	
Ethics oversight	Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Animals and other research organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.
Wild animals	Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.
Reporting on sex	Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.
Field-collected samples	For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.
Ethics oversight	Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	<i>Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.</i>
Study protocol	<i>Note where the full trial protocol can be accessed OR if not available, explain why.</i>
Data collection	<i>Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.</i>
Outcomes	<i>Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.</i>

## Dual use research of concern

Policy information about [dual use research of concern](#)

### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
<input type="checkbox"/>	<input type="checkbox"/>	Public health
<input type="checkbox"/>	<input type="checkbox"/>	National security
<input type="checkbox"/>	<input type="checkbox"/>	Crops and/or livestock
<input type="checkbox"/>	<input type="checkbox"/>	Ecosystems
<input type="checkbox"/>	<input type="checkbox"/>	Any other significant area

### Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
<input type="checkbox"/>	<input type="checkbox"/>	Demonstrate how to render a vaccine ineffective
<input type="checkbox"/>	<input type="checkbox"/>	Confer resistance to therapeutically useful antibiotics or antiviral agents
<input type="checkbox"/>	<input type="checkbox"/>	Enhance the virulence of a pathogen or render a nonpathogen virulent
<input type="checkbox"/>	<input type="checkbox"/>	Increase transmissibility of a pathogen
<input type="checkbox"/>	<input type="checkbox"/>	Alter the host range of a pathogen
<input type="checkbox"/>	<input type="checkbox"/>	Enable evasion of diagnostic/detection modalities
<input type="checkbox"/>	<input type="checkbox"/>	Enable the weaponization of a biological agent or toxin
<input type="checkbox"/>	<input type="checkbox"/>	Any other potentially harmful combination of experiments and agents

## Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>

## ChIP-seq

### Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

#### Data access links

May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

#### Files in database submission

Provide a list of all files available in the database submission.

#### Genome browser session (e.g. [UCSC](#))

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

### Methodology

#### Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

#### Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

#### Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

#### Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

#### Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

#### Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

#### Sample preparation

Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

#### Instrument

Identify the instrument used for data collection, specifying make and model number.

#### Software

Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.

#### Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

#### Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

## Magnetic resonance imaging

### Experimental design

#### Design type

Indicate task or resting state; event-related or block design.

Design specifications *Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*

Behavioral performance measures *State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

## Acquisition

Imaging type(s) *Specify: functional, structural, diffusion, perfusion.*

Field strength *Specify in Tesla*

Sequence & imaging parameters *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI  Used  Not used

## Preprocessing

Preprocessing software *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

## Statistical modeling & inference

Model type and settings *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis:  Whole brain  ROI-based  Both

Statistic type for inference *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

(See [Eklund et al. 2016](#))

Correction *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

## Models & analysis

n/a | Involved in the study  
  Functional and/or effective connectivity  
  Graph analysis  
  Multivariate modeling or predictive analysis

Functional and/or effective connectivity *Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*

Graph analysis *Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis *Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*