Runs of homozygosity: windows into population history and trait architecture

Francisco C. Ceballos[1,2], Peter K. Joshi[3], David W. Clark[3], Michèle Ramsay[1,4] and James F. Wilson[2,3]

Abstract | Long runs of homozygosity (ROH) arise when identical haplotypes are inherited from each parent and thus a long tract of genotypes is homozygous. Cousin marriage or inbreeding gives rise to such autozygosity; however, genome-wide data reveal that ROH are universally common in human genomes even among outbred individuals. The number and length of ROH reflect individual demographic history, while the homozygosity burden can be used to investigate the genetic architecture of complex disease. We discuss how to identify ROH in genome-wide microarray and sequence data, their distribution in human populations and their application to the understanding of inbreeding depression and disease risk.

Consanguinity
Mating among relatives, for example, first or second cousins. Literally 'of the same blood'.

[1]*Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Parktown 2193, Johannesburg, South Africa.* [2]*Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK.* [3]*Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK.* [4]*Division of Human Genetics, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Braamfontein 2000, Johannesburg, South Africa.*

Correspondence to J.F.W.
jim.wilson@ed.ac.uk

doi:10.1038/nrg.2017.109
Published online 15 Jan 2018

Mating between cousins, or inbreeding[1], is sometimes considered to be an unusual occurrence in humans; however, such consanguineous marriage is common across the planet. Surveys using genealogical data reveal that at least 10% of the global population (>700 million people) are the offspring of second cousins or closer[2]. Inbreeding is not distributed evenly around the globe, with higher incidences in areas where consanguinity is favoured culturally, such as parts of West and South Asia, but also occurs as a consequence of small population size and endogamy, even if there is random mating. Societal attitudes towards cousin marriage are greatly influenced by religious beliefs, with the Quran prohibiting marriages between close relatives, but permission is given for marriage between cousins, including double first-cousin unions. Even though many examples of consanguinity are cited in biblical texts, the Levitical code also forbids marriage between close kin.

Cousins share DNA that they have inherited from their common ancestors, and thus the offspring of cousin marriage may inherit identical chromosomal segments from both parents. The availability of denser genome-wide microsatellite scans in the mid-1990s led to the discovery of uninterrupted long runs of homozygous genotypes (known as runs of homozygosity (ROH)), the hallmark of these autozygous segments inherited from a recent common ancestor[3]. Members of two families recruited to construct the first human genetic maps — of Venezuelan and Old Order Amish ancestry — carried 4–16 ROH typically ~1.5–30 Mb in length, the most extreme individual having a total of ~195 Mb in ROH, consistent with close inbreeding[3]. More unexpected was the fact that, despite the relatively sparse and imperfect maps, 20% of the 100 individuals outside these two families (all Utah Mormons) carried at least one homozygous segment — ROH were thus likely to be common in human populations.

It becomes clear why ROH are common when we consider that an individual today is predicted to have half a billion (2^{29}) ancestors 29 generations ago (circa 1100, one generation after William the Conqueror), more than the estimated world population of ~310 million at that time[4,5]. This ancestor paradox is solved by the fact that many of the ancestors are the same people (known as pedigree collapse[6]). In most cases, given broad-scale and fine-scale human population genetic structure and a limited effective population size (N_e), ancestors will be shared more recently in time than the 12th century[7]: we are all inbred to some degree, and ROH capture this aspect of our individual demographic histories. To this end, they can even be analysed by free online utilities for genetic genealogists who have purchased direct-to-consumer genome scans. We do not inherit DNA from all our pedigree ancestors at this remove of generations[8]; however, we have to inherit DNA from some of them, and as the number of genealogical ancestors doubles every generation, eventually everyone has shared genetic ancestors between 300 and 1400 BC depending on assumptions about migration[9].

It has been known for over a century[10] that inbreeding increases the incidence of recessive disease, and the frequency of homozygotes is increased in relation to

the inbreeding level in the population. The long ROH in inbred individuals reveal the full, harmful effects of recessive deleterious variants present in the ROH, for example, to cause Mendelian diseases such as Tay–Sachs. Inbreeding usually leads to decreases in the vigour and reproductive fitness of offspring — known as inbreeding depression — as first noted by Charles Darwin in plants (BOX 1) and seen for numerous fitness-related traits in animals[11,12].

In this Review, we focus on the analyses of human ROH (rather than single-marker inbreeding coefficients) and their contributions to the understanding of human demographic history and to deciphering the genetic architecture of complex disease. We do not focus on Mendelian conditions or human knockouts. We discuss methodological considerations regarding the identification of ROH in microarray and sequence data sets, the distribution of ROH of different lengths across the genome and the globe and the relation of ROH to pedigree. We review the burgeoning literature on the influence of ROH on disease risk and quantitative traits and what has been learned about inbreeding depression in humans. We conclude with some recommendations for the assessment of ROH and highlight future research questions.

Origins of ROH and inbreeding depression

ROH arise when two copies of an ancestral haplotype are brought together in an individual: longer haplotypes inherited from recent common ancestors or shorter haplotypes from distant ones (background relatedness). Short ROH characterized by strong linkage disequilibrium (LD) among markers are not always considered autozygous but nevertheless are due to the mating of distantly related individuals. Different population histories give rise to divergent distributions of long and short ROH (FIG. 1). The ROH complement of outbred populations is related to their effective population size, with smaller populations tending to have more ROH and larger populations fewer ROH. Admixed populations, on account of their more distant shared ancestry across two or more ancestral populations, have fewer ROH than their respective parental populations. Consanguineous communities, on the other hand, have much longer ROH than those seen in outbred populations owing to very recent pedigree inbreeding loops, whereas populations that have undergone a population bottleneck carry a greater number of shorter ROH than cosmopolitan populations, reflecting deeper parental relatedness. Finally, populations with both reduced effective population size in the past and recent inbreeding have the greatest burden of ROH.

The causal mechanism for inbreeding depression is only partly understood, but empirical evidence in a number of species suggests that it is due mostly to increased homozygosity for (partially) recessive detrimental mutations maintained at low frequency in populations by mutation–selection balance, although the contribution of some loci with heterozygote advantage (overdominance) maintained at intermediate frequencies by balancing selection cannot be disregarded[13]. When dominant alleles at some loci decrease the trait value while others increase it, we do not expect any association with genome-wide homozygosity. However, if on average across all causal loci dominance is biased in one direction, for instance, to decrease the trait, we will see such an association. Such directional dominance arises owing to directional selection in evolutionary fitness-related traits.

Empirical studies[14] show that ROH are more enriched for homozygous deleterious variants than for non-deleterious variants. This emphasizes that ROH are important reservoirs of homozygous deleterious variation[15], although this is expected given the typically lower allele frequencies of deleterious variants compared with non-deleterious variants. Inbreeding increases the probability that a variant will be found in a homozygous state, so ROH are enriched for homozygotes at all allele frequencies. This enrichment is particularly strong for rare variants because a variant at frequency p is homozygous at frequency p^2 outside ROH and at frequency p inside ROH (where p is the population frequency of the allele). Homozygotes thus occur $(1/p)$ times more frequently inside ROH, so lower-frequency variants (including deleterious variants) are more strongly enriched. Theory also predicts that very strong inbreeding will in fact purge deleterious recessive alleles from the population as more copies are found in a homozygous state, and this has been observed in mountain and eastern lowland gorillas[16] but not in human genome data[17].

Methodological considerations

ROH calling requires high-density genome-wide scan data, now overwhelmingly from single nucleotide polymorphism (SNP) microarrays, but ROH analysis of short-read sequencing of the entire genome or exome will become more common as the price of these technologies decreases. A number of factors influence the quality of ROH calling, including the marker density, their distribution across the genome, the quality of the genotype calling (including error rates) and minor-allele frequencies. Microarray data are considered the gold standard with very low error rates (typically <0.1%); however, the content usually comprises ~1 million SNPs with allele frequencies >5%, chosen to best represent haplotype structure. Whole-genome sequencing (WGS), on the other hand, assays every variant, irrespective of allele frequency, although the low coverage often employed to maximize the number of individual genomes sequenced, and hence power for association, means that rare SNPs are called considerably less often, with higher error rates, than common SNPs. Hence, parameters of calling algorithms require tuning to the characteristics of the underlying data, and particular care must be taken with centromeres, duplications and other difficult regions. There are two major methods for identifying ROH: observational genotype-counting[18] and model-based[19].

Observational approaches. Algorithms, such as those implemented in PLINK[18], scan each chromosome by moving a window of fixed size along their length in

Endogamy
Marriage within the population or community.

Runs of homozygosity
(ROH). Contiguous regions of the genome where an individual is homozygous across all sites. This arises if the haplotypes transmitted from the mother and father are identical, having in turn been inherited from a common ancestor at some point in the past. It is important to note that this notion does not rely on a known pedigree and does not require an (arbitrary) baseline population (the first generation of ancestors or founders in a pedigree). However, ROH in practice are required to have an (arbitrary) minimum size, depending on the density of genotypes available, to distinguish identity-by-descent from chance.

Autozygous
Also known as homozygosity-by-descent; homozygosity arising at a locus owing to identity-by-descent.

Effective population size
(N_e). The size of an idealized population that would show the same amount of genetic drift or inbreeding, often thought of as the number of breeding individuals and usually lower than the census population size.

Demographic histories
The histories of the changes in population size; for example, populations may be large or small, of constant size, or expanding or contracting; may undergo bottlenecks (severe declines in population size) or founder events (establishment of populations by a limited number of ancestors); may be substructured geographically; or may admix with one another.

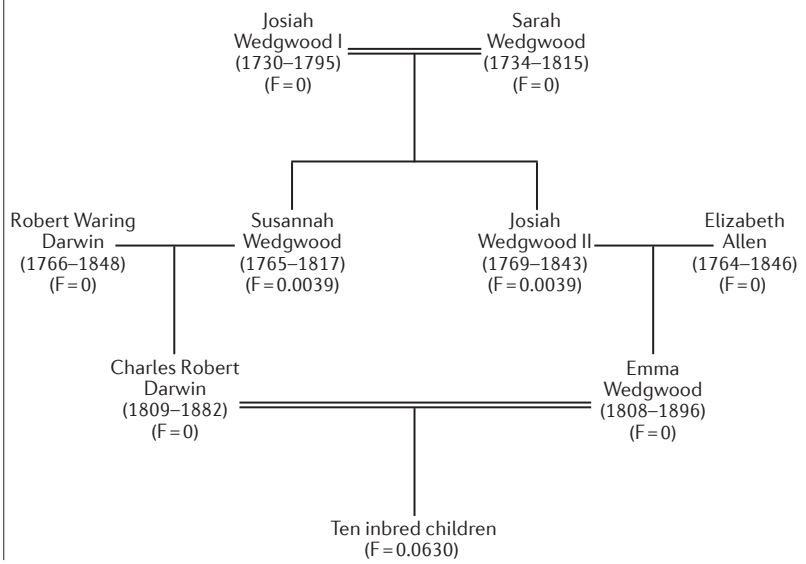
Inbreeding depression
The reduction in evolutionary fitness of a population or individual due to the presence of increased homozygosity arising from inbreeding. Values of traits related to fitness, such as fertility, are reduced.

Box 1 | **Inbreeding depression, Charles Darwin and royal dynasties**

The first research programme on the harmful effects of inbreeding, the mating of close relatives, was performed by Charles Darwin[81,82]. He carried out carefully controlled experiments in plants that involved self-fertilization and outcrossing between unrelated individuals in 57 species and showed that the offspring of self-fertilized plants were on average shorter, flowered later and produced fewer seeds than the progeny of cross-fertilized plants. He thus documented the phenomenon of inbreeding depression, the decline of traits that are closely related to fitness, now known to be caused by the increase in homozygosity in inbred individuals.

Darwin also had a personal interest in the adverse effects of inbreeding since he was married to his first cousin Emma Wedgwood (see the figure), and they had ten children who were often ill, three of whom died at an early age[83]. Charles and Emma would each have inherited large segments of the genomes of their grandparents Josiah Wedgwood I and Sarah Wedgwood, identical-by-descent, and transmitted some of these to their children, thus generating long runs of homozygosity (ROH) wherever the same segments were passed down each side of the pedigree. Darwin's concerns about the harmful effects of first-cousin marriage in his progeny have been considered exaggerated because they were based on the extrapolation from the ill effects of self-fertilization in plants to the outcomes of first-cousin marriage in humans. However, the possibility of inbreeding effects on the Darwin children is supported by the decrease in both childhood survival and male fertility detected in the progeny of a number of consanguineous marriages of the Darwin–Wedgwood dynasty[84,85].

Although studies in the Hutterites have shown a decrease in fecundity for inbred women (as well as evidence of reproductive compensation)[86], most information on inbreeding depression in humans relates to prereproductive survival. The mean decrease in survival to 10 years of age in the progeny of first cousins relative to the offspring of unrelated parents is estimated to be 3.5–4.4% across a large number of human populations[2,87]. The characterization of inbreeding depression for a wider range of inbreeding than that corresponding to first cousins (inbreeding coefficient (F)~0.0625) has been possible in royal dynasties: consanguineous lineages with well-recorded, deep pedigrees make very useful inbreeding laboratories[88–90]. In the House of Habsburg, strong inbreeding depression for both infant and child mortality was detected circa 1450–1800. A considerable reduction of this inbreeding effect on child survival in a fairly small number of generations was observed, potentially caused by the purging of deleterious alleles of a large effect, a mechanism previously observed for loss-of-function alleles in mountain gorillas[16].



search of stretches of consecutive homozygous SNPs. ROH are called by first calculating the proportion of completely homozygous windows that encompass that SNP. If this proportion is higher than a defined threshold, the SNP is designated as being in an ROH. A variable number of heterozygous or missing SNPs per window can be specified in order to tolerate genotyping errors and failures as well as rare new mutation events. Finally, an ROH is called if the number of consecutive SNPs in a homozygous segment exceeds a predefined threshold in terms of SNP number and/or covered chromosomal length. The simplicity of the PLINK approach allows distributed applications in large consortia[20], and SNP data may be pruned for LD if desired before ROH calling. Haplotype-matching algorithms (for example, GERMLINE[21]) for calculation of identity-by-descent (IBD) can also be used to identify ROH as a special case of IBD within an individual.

Model-based approaches. An alternative, computationally expensive approach implemented in the Beagle software program uses hidden Markov models (HMMs) to account for background levels of LD[22]. However, tests on simulated data showed that PLINK outperformed GERMLINE and Beagle in detecting ROH[23]. Further likelihood-based approaches use the log of the ratio of the probabilities of the genotype data under the hypotheses of autozygosity and non-autozygosity (incorporating population-specific allele frequency estimates) to infer the homozygosity status of sliding windows in each individual[19]. A population-specific threshold was defined from these log-odds scores, above which ROH are called. Gaussian kernel density estimates of the genome-wide log-odds scores revealed two modes in each population, and the local minimum was used as the threshold in each case. The distribution of ROH lengths was also modelled as a mixture of three Gaussian distributions, classifying ROH into size classes: very short ROHs (tens to hundreds of kb) reflecting LD patterns; intermediate ROH (hundreds of kb to 2 Mb) that result from background relatedness owing to genetic drift; and long ROH (over 1–2 Mb) arising from recent parental relatedness[19]. Despite providing increased sensitivity in the detection of shorter ROH, the need to estimate allele frequencies is a limitation of this approach — now implemented in the Garlic software[24]. In practice, the Gaussian mixture likelihood results are very highly correlated with those from PLINK[19]; however, the population-specific nature of ROH class boundaries will complicate meaningful meta-analysis.

Short-read sequence data. The increasing popularity of sequence data delivers the ultimate resolution, allowing even the shortest ROH to be identified; however, the genotype error rates are much higher than for microarray data. This is particularly true for low-coverage data (for example, fourfold depth), where there is a high probability that only one of the two chromosomes has been sampled at a specific site. Whole-exome sequences provide a further challenge, given the size of most exons and their sparsity across the genome[25–27]. Nevertheless, a number of HMM approaches have been implemented specifically for whole-genome or whole-exome sequencing. For example, H^3M^2 deploys a heterogeneous HMM taking into account distances between consecutive SNPs and outperforms GERMLINE and PLINK when applied to whole-exome sequences, especially for short and medium ROH[27]; however, analysis requires very

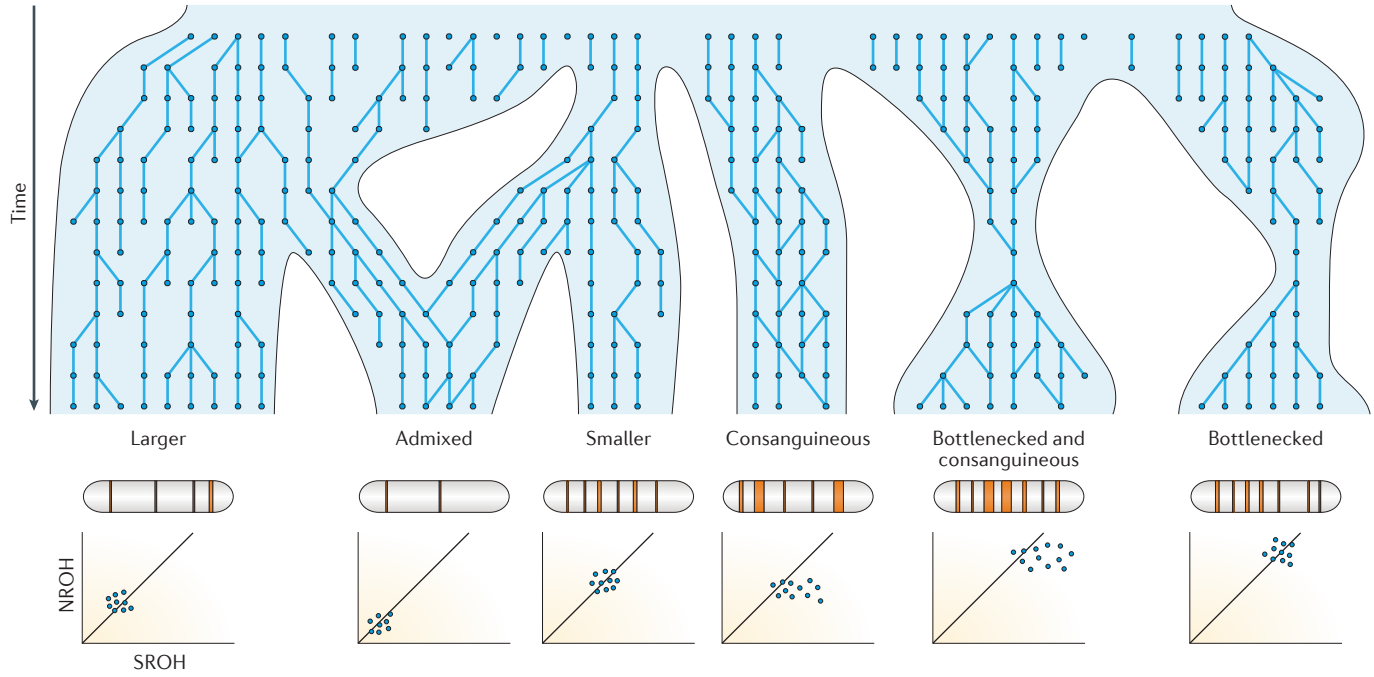


Figure 1 | **Demographic origins of ROH.** The demographic history of six diverse hypothetical populations is represented in the upper part of the plot. Representative pedigrees are indicated by dark blue lines connecting individuals (dots), loops show inbreeding and the population size is represented by the width of the light blue areas. Thus, bottlenecks are shown by a narrowing, which necessarily reduces the number of ancestral lineages that are present in the population; conversely, larger populations contain more ancestral lineages. Admixture is shown by a confluence of two hitherto separate populations and mating between the pedigree lineages therein. The consequences of each demographic scenario are illustrated below: schematic chromosomes showing the typical distribution of runs of homozygosity (ROH) in each and at the bottom a plot of the sum total length of ROH (SROH) versus the total number of ROH (NROH) expected in each scenario. As can be seen, the burden of ROH relates to the size of the population, with smaller populations having more and longer ROH than larger populations. Admixture brings together different haplotypes and typically reduces the number of ROH to very few short ROH, whereas bottlenecks increase the number of ROH, which are typically still relatively short. Consanguinity, on the other hand, adds a small number of very long ROH for those who are the offspring of cousin marriage, thus also increasing the variance in the sum of ROH, visible as a right shift in the NROH versus SROH plot. Some populations are both bottlenecked and practice consanguineous marriage, hence having many short and some long ROH, resulting in the highest burden of ROH.

Genetic architecture
The makeup of the genetic basis of a trait, in particular whether there are few or many causal loci, whether the causal variants are rare or common or have small or large effect sizes and the degree to which dominance plays a part.

Haplotype
A set of alleles on a chromosome or chromosomal segment inherited from one parent — often a series of alleles at neighbouring loci that are strongly statistically associated due to lack of recombination. Certain haplotypes may become common in the population owing to natural selection or drift until broken down over time by recombination.

Admixed
Genetic admixture occurs when mating begins between two previously separate populations and individuals within the new population have a mix of haplotypes from each parental population.

Inbreeding loops
Also known as pedigree loops; the connection in a pedigree between the maternal and paternal ancestors of an individual. The closed loops show how the same haplotypes could pass down both sides of families.

large mapped sequence read (bam) files. Further HMM methods for sequence data include BCFtools/ROH[28], which has similarly low error rates and can use much smaller variant call format (vcf) files, which contain only the SNP genotypes and quality scores. High-depth WGS holds the promise of the most accurate ROH detection and will allow assessment of the contribution of very short ROH to inbreeding depression.

Distribution of ROH

ROH are ubiquitous. A survey using ~700,000 SNP microarray genotypes for 209 HapMap individuals revealed for the first time how widespread megabase-scale ROH were, even among outbred individuals[29]. However, different continental populations have contrasting burdens: Africans generally have fewer ROH, reflecting their larger effective population size. Again, this survey identified cryptically inbred outliers: a Mormon from Utah and two Japanese individuals from Tokyo. Further studies verified these findings in European Americans[30,31] and East Asians[32]. Whereas hemizygous deletions could manifest as apparent ROH in genotype data, analysis of the fluorescent intensities showed a copy number of two in almost every case[31–35], and Mendelian transmission of haplotypes was observed in families[3,34]. Analysis of >3 million SNPs in the HapMap populations allowed identification of ROH down to 100 kb in length, which are dramatically more numerous: each individual carries hundreds to thousands of these, which in total comprise 400–500 Mb of the genomes of cosmopolitan Europeans and East Asians but only 160 Mb in Yoruba from Nigeria[33]. Thus, such short ROH account for more of the total sum of ROH than ROH >1 Mb, even for inbred individuals.

Correlation with pedigree inbreeding. The degree of individual inbreeding is measured using the inbreeding coefficient (F), the probability that an individual receives two alleles that are identical-by-descent at a given locus[36], which is also the expected proportion of the genome that is autozygous, for example, F = 0.0625 for the offspring of first cousins. The genomic inbreeding coefficient, F_{ROH}, measures the actual proportion of the autosomal genome that is autozygous — defined as the sum total length of ROH (SROH) over a specified minimum length threshold as a proportion of the total

genome length[35]. Another useful measure of ROH is the total number of ROH (NROH). F_{ROH} captures the total inbreeding coefficient of the individual, irrespective of pedigree accuracy or depth (or absence), within the resolution of the data set available (and hence the size of ROH that can be called). Early studies revealed that offspring of first cousins with autosomal recessive disease had a mean F_{ROH} of 11%, substantially higher than predicted, probably due to generations of consanguineous marriage[37]. Analysis of a broader spectrum of parental relatedness using accurate pedigrees from an isolated population demonstrated that F_{ROH} calculated using ROH >1.5 Mb in length correlated most strongly ($r = 0.86$) with inbreeding coefficients from six-generation pedigrees (F_{ped})[35]. Pedigrees provide only an expectation of the autozygosity, whereas ROH capture the realized autozygosity; in fact, siblings were shown to differ on average by 10 Mb in SROH. Demonstrably outbred individuals (with no inbreeding loops in at least the last five and probably ten generations) carried ROH up to 4 Mb in length but not longer, emphasizing that these shorter ROH are of considerable age[35]. In fact, across diverse samples, population mean F_{ped} also correlates well with F_{ROH} using ROH >5 Mb ($r = 0.87$) but not with F_{ROH} calculated from ROH <5 Mb[38].

Global distribution. The distribution of ROH across worldwide populations is structured at many scales from continental to tribal[19,38]. Analyses of longer and shorter ROH allow populations to be categorized into a number of broad classes that blend into one another (FIG. 2). The first class consists of consanguineous populations — many Muslim communities in Daghestan[39], Pakistan and West Asia (for example, Qataris[40], Balochis, Makrani, Bedouin and Druze), including Pakistanis in England[41] and also the Selkup of Siberia — that have an increased mean SROH and usually increased variance as well. As the relatively small number of very long ROH arising from the recent inbreeding loops influences the sum of ROH much more than the total number, these populations display a 'right shift' in the NROH versus SROH graph away from the trend line (FIG. 2b). Long tails in the distributions of SROH, or increased means, are also seen (FIG. 2a).

A second class includes numerous native populations from across the world with shared parental ancestry arising from isolation and endogamy over many generations, but comparatively little recent inbreeding. Such individuals carry few long ROH but have substantial enrichment for ROH in the 2–5 Mb length, leading to a relatively high NROH, and include Papua New Guinean Highlanders, the Koryak and Chukchi of Siberia, the Pulliyar, Kurumba and Piramalai Kallar castes of southern India, unadmixed Greenlanders and Athabaskans of North America, the click-speaking Hadza hunter–gatherers of Tanzania (FIG. 2a,b) and the Onge of the Andaman Islands in the Bay of Bengal[42,43]. Many other, mostly isolated groups display a less extreme profile of increased burden of shorter ROH, notably four Khoisan-speaking and two Pygmy populations, hunter–gatherers who stand out from the otherwise very low-ROH sub-Saharan Africans. Various European-heritage isolated populations are also known to carry many long ROH[20,44], for example, Amish (in whom ROH were first discovered[3]); Hutterites; populations of villages in Sardinia and Friuli-Venezia-Giulia, Italy[45], northern Sweden and Greece; Roma (gypsies); and Irish Travellers. Such an increased burden of ROH is not uncommon and was likely the default situation for much of human history before the farming revolution (BOX 2).

A third class shows signs of both ancient and recent inbreeding with an enrichment of both shorter and longer ROH, exemplified by Native American populations: the Karitiana and Surui of the Brazilian Amazon; Piapoco of the Colombian Amazon; and Pima of Mexico[19,38]. A fourth class — the most numerous globally — is from societies with much larger effective population sizes and thus much lower mean number and sum of ROH: East Asians typically have more ROH than Europeans, who in turn have more than South Asians, and sub-Saharan Africans have the least ROH. Indeed, shorter ROH in particular are correlated ($r = 0.82$, $P < 0.0001$) with overland distance from Addis Ababa in Ethiopia[19,38], reflecting serial bottlenecks during the dispersals across the globe. In these populations, there are also often different levels of cryptic inbreeding indicated by long tails on the densities of SROH >10 Mb (for example, for the Japanese population shown in FIG. 2a, ~5% of the sample show evidence of recent inbreeding). Finally, a class of mixed populations presents a heterogeneous picture, with admixed Native Americans and Hispanics/Latinos showing high variance in SROH, while African Americans[19] and Cape Coloureds have very little, and first-generation or second-generation mixed-race individuals have the fewest ROH of all populations. These differences arise from the specific histories of each admixed community in terms of the time depth of admixture and the burden of ROH in the parental populations. Native Americans have the highest mean SROH, but there is a wide distribution of Native American ancestry proportions in populations of Latin American descent[46]. The higher the Native American ancestry component, the greater the chance that these haplotypes will form ROH. The offspring of recent mixed-race partnerships, on the other hand, will have very few ROH, given the low chance of shared parental haplotypes, irrespective of the particular continental ancestry.

Sociodemographic factors that are not directly related to geography or principal components of ancestry can also influence ROH distributions. For example, in the Netherlands, which has a history of assortative mating by religious affiliation, F_{ROH} varies significantly between religious and non-religious groups[47]. Moreover, differential migration by educational status can induce systematic differences in F_{ROH} in highly educated, mobile versus less educated, less mobile population strata[48]. The effects of increased migration and urbanization through time also generate a secular trend in ROH such that younger European Americans have significantly lower burdens: NROH and SROH are predicted to have decreased by 14% and 24%, respectively, over the past century[49].

Population bottleneck
A severe decline in population size over a short time or a lesser reduction over a longer time, followed by a recovery.

Cosmopolitan populations
Populations that are not isolated; typical urban populations.

Overdominance
Also known as heterozygote advantage; overdominance occurs if the heterozygote trait value (phenotype) is outside the range of the trait values of the two homozygotes.

Balancing selection
When two or more alleles are favoured by natural selection rather than one, for example, when the heterozygote is fitter than either homozygote.

Dominance
Dominance is present at a genetic locus when the effect of one copy of an allele gives rise to a trait or phenotypic value that, rather than being halfway between the values for the two homozygotes, is nearer the trait value for a carrier of two copies of the allele. In this situation, the other allele is recessive.

Directional dominance
Directional dominance occurs when the dominance effect across all causal loci in the genome has a trend in one direction, that is, to raise or lower the trait, rather than the individual dominance effects at loci cancelling each other out.

Identity-by-descent
(IBD). The inheritance of an identical haplotype from both parents owing to it having been passed without recombination from a common ancestor in the baseline population.

Inbreeding coefficient
The probability, denoted F, of inheriting two alleles identical-by-descent at an autosomal locus in the presence of consanguinity. F is one-sixteenth for first-cousin offspring, one-sixty-fourth for second cousins and one-eighth for the progeny of avuncular or double first-cousin matings.

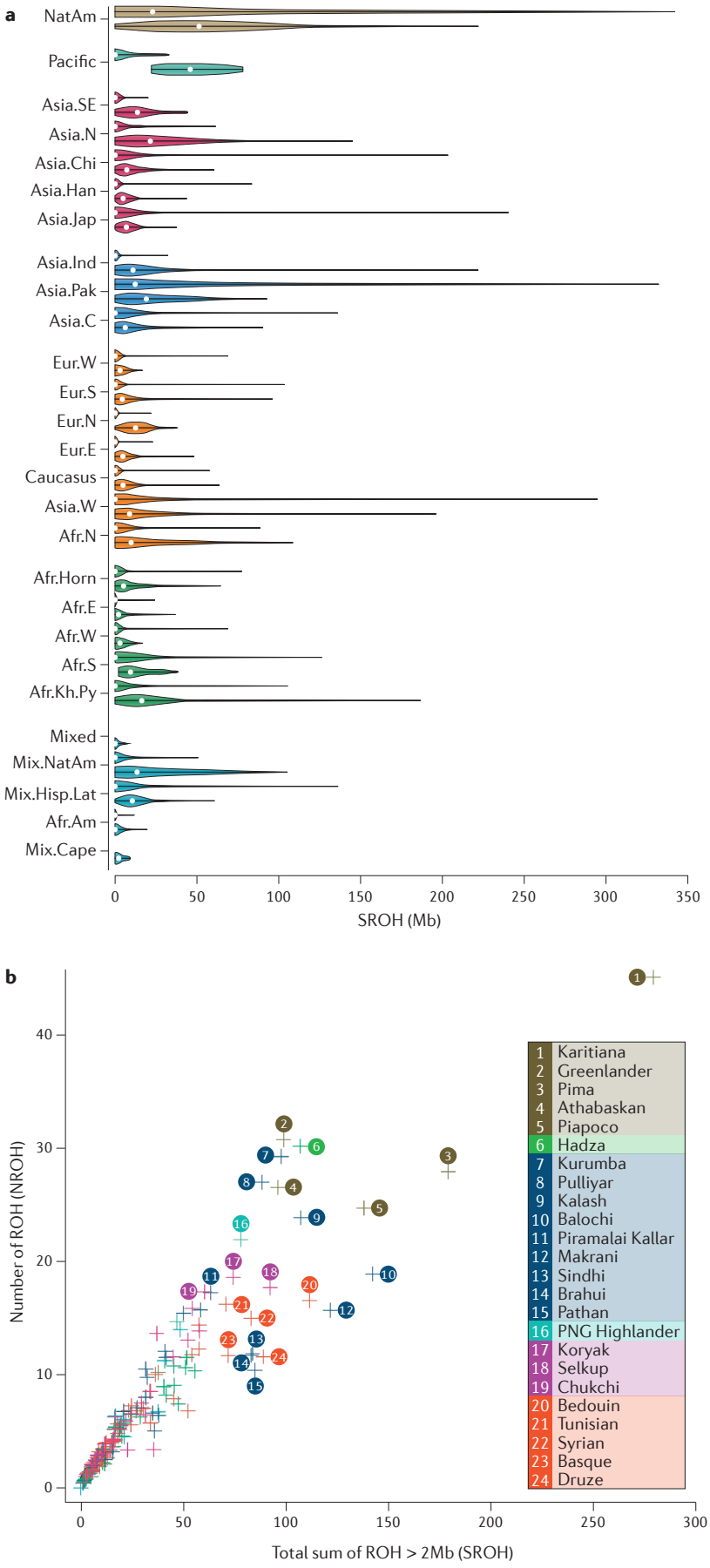


Figure 2 | **Global census of ROH. a** | Violin plot of sum length of runs of homozygosity (SROH) in 27 regional or demographic groups coloured by biogeographical continent (Americas in beige, Pacific in mint green, East Asia in pink, South Asia in midblue, West Eurasia in orange, sub-Saharan Africa in green and admixed in turquoise). The violin shows a coloured kernel density trace with the interquartile range as a black line and the median as a white circle. For each group, long runs of homozygosity (ROH) (>10 Mb) are plotted above and shorter ROH (2–5 Mb) are plotted below. Native Americans stand out with higher median SROH for both short and long ROH, whereas Pacific Islanders have a higher burden only of short ROH. Both West Asian and Pakistani populations have long tails in the distribution of long ROH, consistent with frequent close consanguinity. Mixed-race individuals have very few long ROH and the least short ROH. Northern Europeans are enriched for shorter ROH because the sample is mostly Finns. **b** | The mean SROH and number of ROH (NROH) are plotted for each of 160 populations with greater than three unrelated individuals sampled, coloured according to continent. Most populations have a complement of ROH similar to others from the same biogeographical continent; however, some stand out. For example, the Amazonian Karitiana have the highest SROH and NROH, the East African Hadza hunter–gatherers are similar to Native Greenlanders, and some North Asian groups, for example, the Selkup, are similar to Syrians and other Near Eastern populations. Populations with mean SROH >60 Mb are labelled. Published data[104–115] from the intersection of numerous microarrays were used (147,911 single nucleotide polymorphisms (SNPs) with minor-allele frequency >0.05); individuals not clustering with their population in principal components analysis (PCA) or showing high kinship were removed before plotting; admixed Native Americans were classified using PCA and admixture analyses. Minimum ROH length 2 Mb with ≥50 SNPs. South Asians include Pakistanis, Indians, Bangladeshis, Sri Lankans and Nepalese. East Asians include Chinese, Mongolians, Japanese and Koreans, together with Southeast Asians and indigenous Siberians. Western Eurasians comprise Europeans and West Asians, which in turn include North Africans. Afr, African; Am, American; C, central; Cape, Cape Coloured; Chi, Chinese minorities; E, east; Han, Han Chinese; Hisp, Hispanic; Ind, Indian; Jap, Japanese; Kh, Khoisan; Lat, Latino; Mix, mixed; N, north; NatAm, Native American; Pak, Pakistani; PNG, Papua New Guinea; Py, Pygmy; S, south; SE, southeast; W, west.

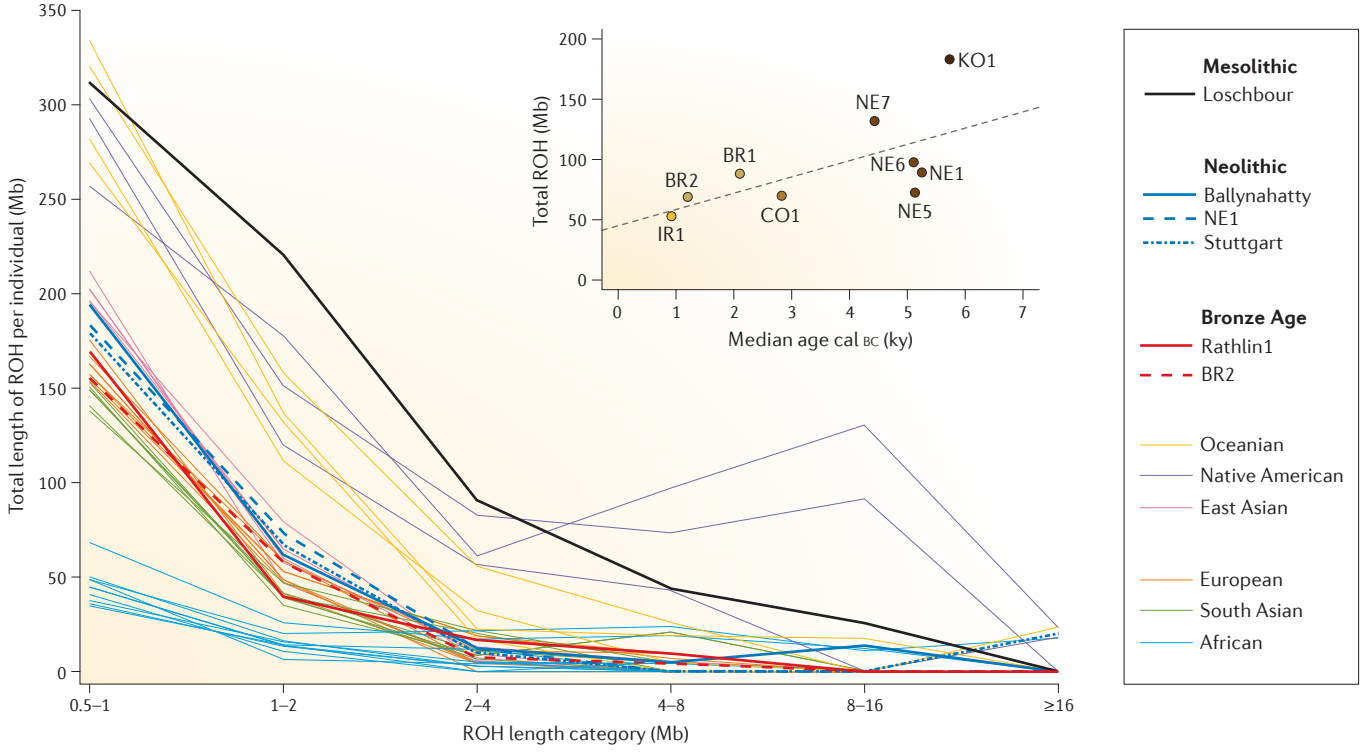
Limits of homozygosity. Complete hydatidiform moles are a very rare form of non-viable pregnancy wherein the oocyte is enucleated and fertilized by a sperm. Thus, the mole contains only sperm-derived DNA and is homozygous across the entire genome; they have been used to provide accurate haplotype maps[50]. Uniparental disomy (UPD) occurs when both copies of a chromosome, or segment of a chromosome, are inherited from one parent and therefore also generates ROH if two copies of one parental chromosome are present. However, the observation of Mendelian transmission of haplotypes giving rise to ROH demonstrates that most ROH are not due to UPD or other cytogenetic abnormalities[34]. Indeed, analysis of a large series of children with developmental delay or autism revealed UPD to be rare and to manifest with very long ROH, with the shortest

Box 2 | **ROH in ancient humans, Neanderthals and great apes**

The ability to generate genome-wide genotypes or whole-genome sequences from ancient DNA has ushered in a new era in understanding human population history, including via runs of homozygosity (ROH). It is striking that both Upper Palaeolithic and Mesolithic hunter–gatherers, from Luxembourg, Switzerland and Georgia (~6,000–11,000 BC), carried very high levels of ROH, comparable to those of modern Oceanians and certain Siberian, Indian and Greenlander populations. By contrast, Neolithic skeletons from Northern Ireland, Hungary, southwest Germany and Anatolia (~3,200–6,000 BC) showed much-reduced levels, comparable to those of modern East Asians, with Bronze Age samples from Northern Ireland and Hungary (~2,000–1,200 BC) even lower, similar to the levels of modern Europeans[42,91–93] (see the figure). Plotted in the main figure[91] is the sum of ROH in different megabase (Mb) length categories for one Mesolithic (bold black line), three Neolithic (bold blue lines) and two Bronze Age (bold red lines) samples, along with representative individuals from modern populations (thin lines with colours indicating continent of origin). The results imply that the Western and Caucasus hunter–gatherers lived in relatively isolated, endogamous societies, unlike the Neolithic farmers, who appear to have arrived as part of a folk migration with a large effective population size. There is in fact a relationship (see the figure inset) between the median calibrated carbon-14 date (in thousands of years (ky) before the common era, BC) of nine ancient Hungarian skeletons and the sum total length of ROH (SROH) ($r^2 = 0.4$, $P = 0.06$)[92]. Additionally, the very early (~8,000 BC) Neolithic sample from Boncuklu in Anatolia, who was probably an indigenous forager who adopted cultivation, was intermediate in ROH distribution to the Mesolithic and later Neolithic samples[93].

A considerably more extreme burden of ROH was discovered in a Neanderthal woman from the Altai mountains of Siberia. She carried 20 ROH longer than ~8 Mb, consistent with an inbreeding coefficient of 0.125, and was therefore the product of an avuncular, half-sibling or double first-cousin relationship over 50,000 years ago[94]. Analysis of shorter ROH (2–8 Mb) revealed evidence of background inbreeding over and above the recent consanguinity, such that the Altai Neanderthal carried more ROH of this length than the Karitiana, who are known to be among the most homozygous modern human populations (FIG. 2b), with increased burden of ROH in all length categories[38]. Remarkably, this was also the case for the Denisovan sample — derived from another 50,000-year-old archaic hominin from Siberia — implying that mating between relatives was not uncommon for either species at this time. Short ROH identified on chromosome 21 in Neanderthals from Spain and Croatia also resembled that of the Denisovan[95].

The distribution of ROH in hominins can be put into perspective by comparison with the other great apes. The endangered mountain gorillas have exceptionally high levels of homozygosity, with an average of 34% of their genomes in ROH[16]. Nineteen per cent of their genomes consist of ROH between 2.5 and 10 Mb, easily more than the most homozygous reported human and the Altai Neanderthal. The homozygosity implies several generations of recent as well as ancient inbreeding in the ancestry of the seven individuals sampled. Eastern lowland gorillas also show exceedingly high levels of ROH, about double the sum and number of ROH that are typical among the Karitiana[96]. By contrast, Western lowland gorillas, chimpanzee, bonobo and orangutan subspecies are much less homozygous, even if often considerably higher than most outbred human populations, averaging Oceanian levels for bonobos, for instance. Thus, the great majority of humans are at the lower end of the hominid homozygosity spectrum, and only very isolated populations reflect the pattern that is prevalent in most great apes, where habitat fragmentation has reduced breeding population sizes dramatically. Eastern lowland and mountain gorillas are considerably more homozygous than any human or other great ape population. Figure adapted with permission from REF. 91, Proceedings of the National Academy of Sciences; and REF. 92, Macmillan Publishers Limited.



being 13.5 Mb (REF. 51). Thus, UPD is unlikely to confound analyses of ROH to any great degree, particularly as subjects with ROH on multiple chromosomes can be excluded as UPD cases. Incest — mating between first-degree relatives — will generate an extreme burden of ROH, with ~25% of the genome expected to be in ROH. Several such cases have been found through clinical screening of children with intellectual disabilities

or congenital abnormalities using microarrays[52–54], and incest was common among the Pharaohs, for example, Tutankhamun[55] and the Ptolemies.

In the data presented in FIG. 2, the most extreme individual has an SROH >500 Mb, including 342 Mb in ROH >10 Mb in length, probably the result of avuncular union[56], also observed in a Neanderthal sample (BOX 2). Of 3,851 individuals, 112 have SROH >160 Mb, which is a conservative lower boundary for the equivalent of offspring of first cousins or closer[35]; only five of these had no ROH >10 Mb. Outside of populations with high mean SROH, first-cousin offspring are seen in two Japanese individuals, one Uzbek, one South African Bantu-speaker, four Colombians and one Mexican (mestizos — recalling the Maracaibo Venezuelan family where ROH were first discovered[3]). Assessing F_{ROH} not only reveals interesting demographic historical information but also allows prediction of the increased risk of rare recessive diseases[57].

Distribution across the genome. ROH are somewhat more common in regions of high LD and low recombination[29] and are particularly prevalent on the X chromosome[58] and regions of low genetic diversity[59]. These observations are linked by low recombination: the X chromosome spends one-third of its time in the male germline, where (with the exception of the small pseudo-autosomal regions) it cannot recombine, and low-recombination regions have low SNP diversity. Recombination breaks up chromosomal segments over generations, and thus low-recombination regions allow greater persistence of long ancestral haplotypes and an increased chance that they come together to form ROH. Over and above this, there is a very uneven distribution along the genome, with a number of comparatively short regions with significant excesses of ROH — known as ROH islands — on each chromosome[19,31,35,58,59], as well as coldspots[19]. These ROH islands dominate the population of ROH in typical outbred individuals, and while they are present in all populations, they are overshadowed by much longer ROH arising from recent pedigree loops in inbred individuals[29,35]. The common ancestors are recent enough that recombination has had little opportunity to break up the segments, and so these ROH are more randomly distributed across the genome. This difference is illustrated by the distribution of ROH >1 Mb in length on chromosome 1 (which reflects the genome-wide pattern[58]) in the relatively outbred Tuscans from the 1000 Genomes Project (FIG. 3a) versus that in the consanguineous Punjabis (FIG. 3b). There is a distinct tendency for Tuscans to carry ROH in the same places — ROH islands — where commonly >10% of the population carries an ROH (FIG. 3c). More randomly sited ROH are also observed. Fine-scale investigation reveals remarkably consistent sharing of ROH boundaries from person to person, probably due to ancestral recombination events[59] (and once more highlighting the pervasive influence of recombination on ROH distributions). Whereas ROH islands are also present in the Punjabis, in some cases at the same loci as in Europeans, a significant minority of the population carries much longer ROH scattered across the genome, elevating the baseline proportion of individuals who are autozygous (FIG. 3d).

In some cases, ROH islands are due to homozygosity of one common haplotype, but in other cases, multiple different haplotypes contribute to the ROH[58]. The origin of ROH islands is subject to debate, but it appears that there are extended haplotypes segregating at high frequencies in the population in these regions. In some cases, this can be explained by selection; for example, there is an ROH island around the lactase (*LCT*) gene on chromosome 2q21 in Europeans, the site of very strong selection for the ability to metabolize lactose as an adult[58], and numerous other islands are probable targets of recent positive selection[19]. Another potential explanation is that ROH islands include small inversions that suppress recombination[58] — or some may not be truly autozygous. Whole-genome sequence data will facilitate an assessment of whether hitherto ungenotyped variants in ROH islands are also homozygous and thus the potential contribution of rare variants in these regions to disease risk.

ROH and complex disease

Although homozygosity mapping[60] has successfully identified the loci underlying many hundreds of rare recessive disorders, mostly in high-homozygosity populations, attention has only recently turned to the relationship between ROH and complex diseases[61]. The now familiar challenges of small effect sizes at many loci have been explored in real and simulated data, showing that sample sizes of 12,000–65,000 individuals would be required to detect effects in populations with cosmopolitan effective population sizes[7]. Even in small effective population sizes of ~1,000, conservative but realistic effect size estimates imply that ~5,000 samples are required for 80% power.

Genome-wide effects in case–control studies. Many different diseases and risk factors, from cancer to cognition, have been tested for association with either the burden of ROH (SROH) or their number (NROH) or for association of individual ROH with the phenotype (TABLE 1). Whereas 12 studies found no evidence of association, 14 reported an association with genome-wide NROH and/or SROH. However, only four of these positive associations have sample sizes above the minimum (~12,000 individuals) estimated to have power to detect the effect sizes expected for complex traits[7]. Power also depends on the variance in SROH, which is highly correlated with mean SROH, such that more homozygous populations are more powerful. An interesting example is provided by serial analyses of schizophrenia risk and ROH, whereby an initial meta-analysis of >9,000 cases and >12,000 controls provided evidence that a 1% increase in F_{ROH} conferred a 17% increase in risk of schizophrenia[62]. However, analysis of a much larger sample set, totalling nearly 40,000 subjects, found a much-attenuated signal and concluded that there was no reliable association between burden of ROH and case status[63]. Confounding, publication bias and other biases may therefore also be at play[64].

Genomic inbreeding coefficient
F_{ROH}: the proportion of the genome that is in ROH. F and F_{ROH} have been shown to be highly correlated.

Avuncular union
Marriage or mating between an uncle and niece or aunt and nephew.

Confounding
Literally, confusion. Statistical confounding arises when the association between a proposed explanatory variable and an outcome is distorted by the presence of a third variable associating with both. Unless all confounding can be excluded, causal inferences cannot be made from observational associations.

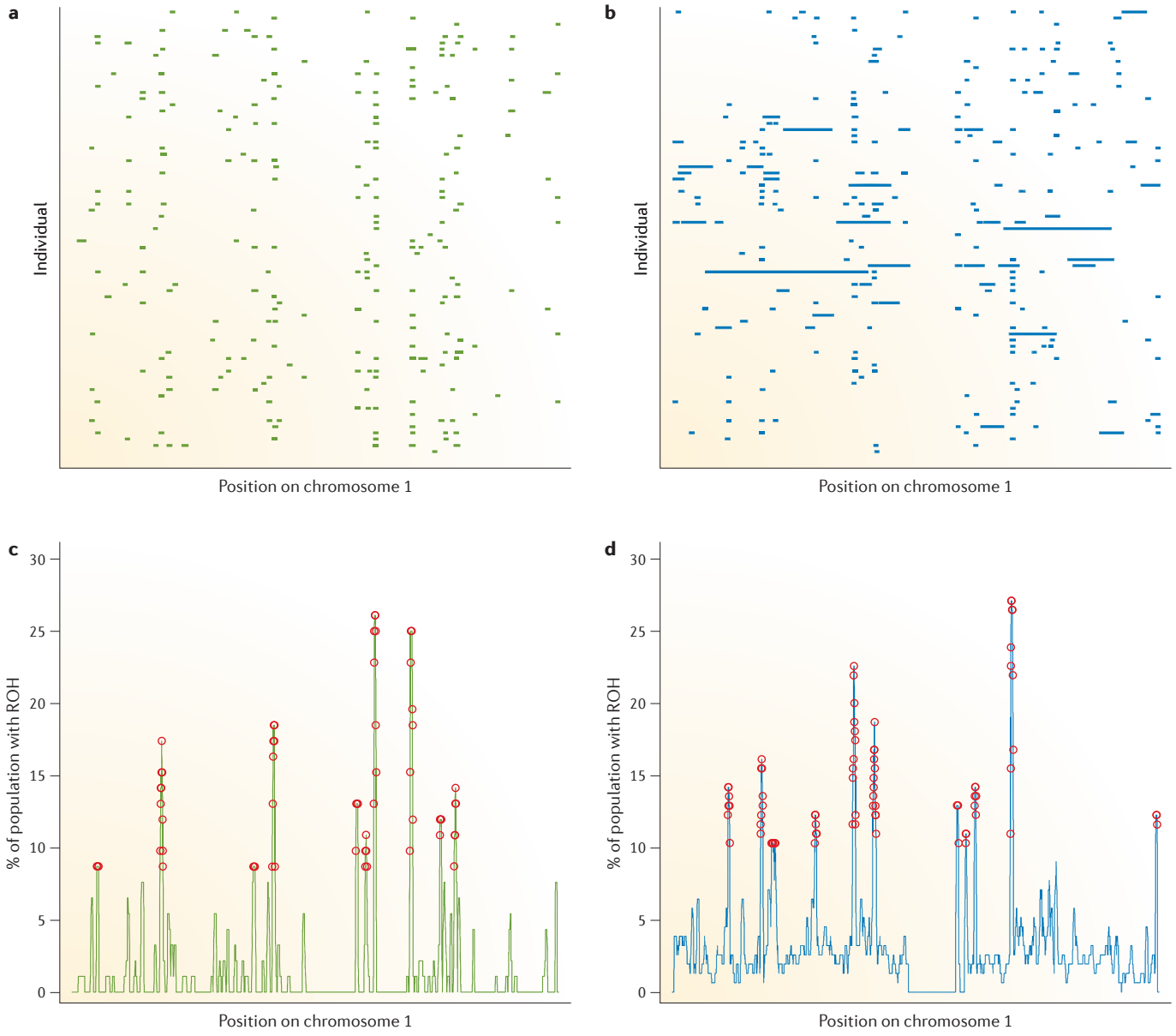


Figure 3 | **Genomic distributions reveal common ROH islands and random patterning of long ROH.** The size and location of runs of homozygosity (ROH) over 1 Mb in length across the genome are represented by an analysis of chromosome 1 for the first 70 individuals in each of two populations from the 1000 Genomes Project[104], together with the proportion of each population that carries ROH in each genomic location. **a** | Genomic distribution for Tuscans from Central Italy (Toscani in Italia; TSI). A uniform pattern of short, shared ROH in islands dominates the picture, although a few short ROH are found outside islands owing to distant pedigree loops. **b** | Distribution for Punjabis from Lahore, Pakistan (PJL). Again, the ROH islands stand out as concentrations of short ROH; however, in some individuals, very long ROH are also present due to the frequency of consanguineous marriage in this population. In contrast to the ROH islands, the longer ROH are distributed relatively randomly across the chromosome. **c** | Percentage of 92 TSI with ROH across chromosome 1. There are ~10 ROH islands in Tuscans, where typically 10% to >20% of the population carries an ROH, against a background of 0.4% outside ROH islands. **d** | Percentage of 155 PJL with ROH. The Punjabis provide a stark contrast: the proportion of the population carrying ROH along the genome is greatly increased, averaging 2.5% outside ROH islands. The proportion of individuals in an ROH was assessed while sliding a 100 kb window across chromosome 1. Windows with a red circle have a significant enrichment of ROH by a binomial test ($P < 2 \times 10^{-5}$ with Bonferroni correction for 2,500 windows).

Indeed, this inconsistency, particularly for case–control analyses, is a common feature of ROH studies and may be due to confounding by factors such as education, socio-economic status, rurality and cultural influences, which might be associated with both inbreeding and the end points considered[63]. In genome-wide association studies (GWAS), after accounting for population genetic structure by, for example, using principal components of ancestry, quantitative geneticists can usually rely on the random

Table 1 | **Studies of ROH and quantitative or disease phenotypes**

Phenotype	Design	Number	Population	Results	Refs
Schizophrenia	Case–control	178 cases; 144 controls	Jewish ancestry, USA	Individual ROH over-represented in cases	31
Schizophrenia	Case–control	9,288 cases; 12,456 controls	Multiple countries and ancestries	Association with SROH	62
Schizophrenia	Case–control	18,562 cases; 21,268 controls	Multiple countries and ancestries	No association with NROH, SROH or individual ROH	63
Major depression	Case–control	9,238 cases; 9,521 controls	9 European populations	No association with NROH, SROH or individual ROH	65
Major depression	Case–control	1,834 cases; 2,131 controls	Dutch	No association with NROH, SROH or individual ROH	47
Bipolar disorder	Case–control	506 cases; 510 controls	European ancestry, UK	No association with NROH, SROH or individual ROH	116
Alzheimer	Case–control	837 cases; 550 controls	North European ancestry	Association with NROH and SROH	117
Alzheimer	Case–control	1,917 cases; 3,858 controls	African-American	Association with NROH and SROH	118
Autism	Extreme homozygous cases	5,431	Multiple ancestries, USA	Individual ROH over-represented in cases	119
Autism	Family-based	2,584 trios	Multiple ancestries	Individual ROH over-represented in cases	120
Autism, speech delay	Case–control	315 cases; 1,115 controls	Han Chinese	Individual ROH over-represented in cases	121
Intellectual disabilities	Family-based	2,108 families	Multiple ancestries, USA	Individual ROH over-represented in cases	122
Intellectual disabilities	Cases	668	Italian	Association of SROH with degree of disability	123
Intellectual disabilities	Cases	267	Russian	Individual ROH detected in cases; no controls	124
Psychosis	Case–control	203 cases; 125 controls	Pacific Islanders	Association with NROH and SROH	125
Colorectal cancer	Case–control	74 cases; 264 controls	Jewish cases, European ancestry controls, USA	Association with NROH and SROH	126
Colorectal cancer	Case–control	921 cases; 626 controls	European ancestry, UK	No association with NROH, SROH or individual ROH	127
Colorectal cancer	Case–control	48 cases; 100 controls	Saudi Arabian	No association with NROH; individual ROH over-represented in cases	128
Childhood acute leukaemia	Case–control	824 cases; 2,398 controls	European ancestry, UK	No association with NROH, SROH or individual ROH	129
Breast and prostate cancer	Case–control	1,183 cases; 1,185 controls	15 worldwide populations	No association with NROH, SROH or individual ROH	130
Lung cancer	Case–control	788 cases; 830 controls	European ancestry, USA	Individual ROH over-represented and under-represented in cases	131
Breast cancer	Case–control	906 cases; 1,217 controls	German	Association with NROH and SROH	132
Thyroid cancer	Case–control	659 cases; 431 controls	Italian	Association with NROH and SROH; individual ROH over-represented in cases	133
Rheumatoid arthritis	Case–control	2,000 cases; 3,000 controls	European ancestry, UK	Individual ROH over-represented in cases	71
Amyotrophic lateral sclerosis	Case–control	605 cases; 1,179 controls	Irish	Association with NROH and SROH; individual ROH over-represented in cases	134
Multiple sclerosis	Case–control	88 cases; 178 controls	Orkney Islanders	No association with SROH	135
Multiple sclerosis	Case–control	29 cases; 28 controls	Faroe Islanders	No association with NROH	136
Coronary artery disease	Case–control	12,123 cases; 12,197 controls	11 European populations	Association with NROH and SROH	137

Table 1 (cont.) | **Studies of ROH and quantitative or disease phenotypes**

Phenotype	Design	Number	Population	Results	Refs
Longevity	Population-based	5,974	Dutch	No association with NROH, SROH or individual ROH	138
Bone mineral density	Population-based	8,647	European and Chinese	Individual ROH associated with trait	72
Height	Population-based	9,383	European ancestry, USA	Individual ROH associated with trait	139
Height	Population-based	35,808	35 European populations	Association with NROH and SROH	66
Cognitive ability	Population-based	2,329	Twins from 32 countries	No association with NROH, SROH or individual ROH	140
Cognitive ability	Population-based	4,854	European ancestry	Association with NROH and SROH	67
Education	Population-based	2,089	Dutch	Association with NROH and SROH	48
Height, income, education, anhedonia, lifetime health	Population-based	5,368	Finnish	Association with SROH for 5 traits but not for 12 others tested	141
Height, lung function, cognitive ability, education	Population-based	53,300–354,224	102 cohorts from 5 continents	Association with SROH for 4 traits but not for 12 others tested; β_{FROH} between −4.7 and −2.9	20

In this table, phenotypes are ordered by neuropsychiatric disorders, cancer, autoimmune and cardiovascular diseases, followed by quantitative traits. β_{FROH}, the effect size estimate of $F_{ROH} = 1$, expressed in units of intrasex phenotypic standard deviations; NROH, number of runs of homozygosity; ROH, runs of homozygosity; SROH, sum of runs of homozygosity.

distribution of alleles, obviating the need for carefully matched cases and controls. However, concerns over confounding are much less easily dismissed for ROH, the burden of which can change greatly in one generation[65]. Social class, genetic isolation and many other potentially confounding variables can and often do associate with parental relatedness (consider, for example, European royal families (BOX 1), although the point is much broader).

In the Netherlands, for example, apparent association was seen between F_{ROH} and major depressive disorder; however, further investigation revealed that this was confounded by religion, and no association remained after accounting for religiosity[47]. Further complications may arise owing to ascertainment biases between cases and controls or experimental biases, for example, genotyping and ROH calling not being performed jointly for cases and controls. For ROH, this may be particularly sensitive as the precise length and evidence for the existence of an ROH may depend on the density (and of course accuracy) of the genotyping microarray used[20].

Quantitative traits. More consistency has been observed in studies of the association between SROH and quantitative traits, perhaps owing to the larger sample sizes, the use of common microarrays within study populations and the avoidance of unmatched controls. An exceptionally large study of up to 354,224 subjects found the regression coefficient between trait and the proportion of the genome in ROH to be −2.9 (0.2), −3.5 (0.7), −4.7 (0.6) and −4.6 (0.7) phenotypic standard deviations (standard errors in brackets) for height, forced expiratory lung volume in 1 second, cognitive ability and educational attainment[20], respectively (FIG. 4). The offspring of first cousins are thus predicted to be 1.2 cm shorter, have 140 ml lower lung expiratory capacity, have 0.29 standard deviations less generalized cognitive function (~4.3 IQ points) and attain 9.7 fewer months of education. Although the association with height was already known[66], the relationship with general cognitive ability was replicated in a study of 4,854 subjects of European ancestry[67], and effect sizes were consistent with pedigree-based estimates for stature and IQ[68,69]. However, no effects of homozygosity were observed for a series of cardiometabolic risk factors, despite very large sample sizes[20].

Despite consistency across these studies, it remains important to consider confounding. One study of ~2,000 subjects of Dutch ancestry showed that parental educational attainment is very strongly related to offspring SROH, a relationship fully explained by the distance between the parents' birthplaces; that is, more educated individuals moved further before finding a spouse, and the two parents were thus, on average, less genomically related[48]. Social behaviours can thus confound associations between SROH and complex traits. Nonetheless, and reassuringly for causal inference of the effect of SROH on traits, an international multicohort study[20] also showed that the effect sizes observed were similar in populations with higher and lower mean SROH, a consistency expected if the effect is from SROH but not expected under a model of confounding by socioeconomic status. The signals of association were also robust to stratification by geography or demographic history and inclusion of educational attainment as a covariate[20]. As expected under the directional dominance model, associations were also found in populations for which there is very limited

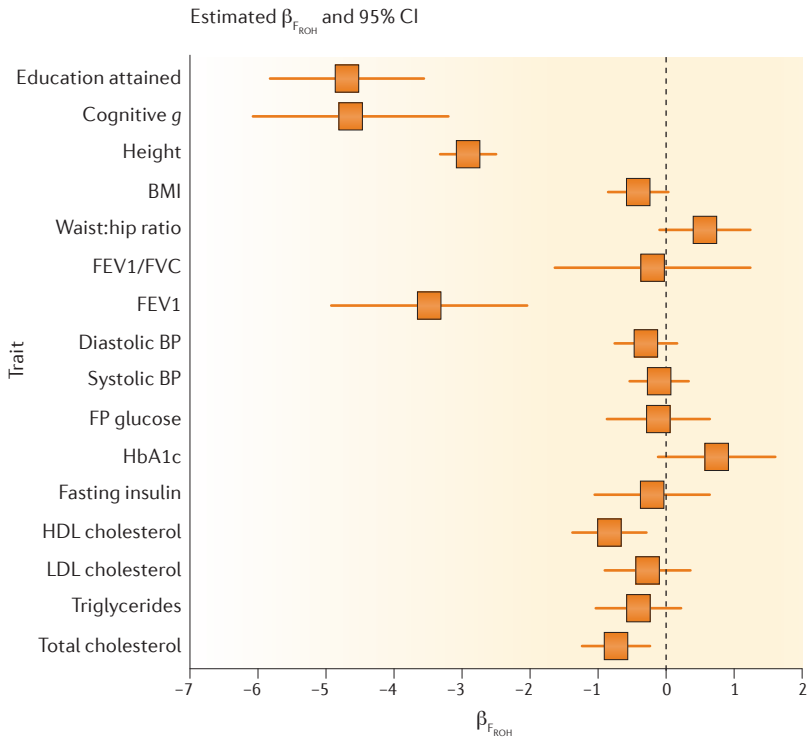


Figure 4 | **Effect of genome-wide homozygosity on 16 complex traits.** Significant effects are observed for two trait groups: stature-related (height and forced expiratory volume in 1 second (FEV1)) and cognition-related (Spearman's *g*, which is a measure of generalized cognitive ability, and educational attainment)[20]. The other, mainly cardiometabolic risk factors show no effect. β_{FROH}, the effect size estimate of $F_{ROH} = 1$, in standard deviation units; BMI, body mass index; BP, blood pressure; FP, fasting plasma; FVC, forced vital capacity; HbA1c, haemoglobin A1c; HDL, high-density lipoprotein; LDL, low-density lipoprotein. Figure from REF. 20, Macmillan Publishers Limited.

migration and sometimes communal living, such as the Hutterites, Amish and populations of Italian hill towns and mountain villages.

Fitness-related traits are thought to show inbreeding depression and/or directional dominance owing to the limited effectiveness of purifying selection on rare, deleterious recessive alleles. The associations between SROH and stature and cognition thus suggest that these traits — or perhaps more likely some underlying trait or traits — are components of Darwinian fitness and have a rare, recessive element to their genetic architecture[13]. The contribution of this component to genetic variance may be low, should the individual alleles be very rare and fully recessive, even if there are many such alleles. However, the decline in ROH due to increasing panmixia and urbanization will be beneficial in terms of reducing the burden of recessive Mendelian disease and risk factors for complex traits showing directional dominance[70].

Individual ROH associations. The postulated model that homozygosity at rare, deleterious recessive alleles gives rise to directional dominance implies that there are specific loci within the genome giving rise to these effects. In principle, such loci should be discernible through a GWAS, although recessive models have been much less used than additive ones (and there is less power to detect rare variants than common ones). A slightly different approach identifies regions in which the occurrence of ROH is tested for association with the trait, aiming to detect a different class of variant from that found in GWAS. However, caution must be used in such exercises: multiple testing of large numbers of independent regions requires proper adjustment, and confounding by genome-wide homozygosity needs to be accounted for. Twelve studies observed associations of individual ROH with diverse phenotypes (TABLE 1); however, all sample sizes were smaller than that judged to be effective for a single whole-genome statistical test[7], let alone the multiple tests performed in a genome-wide ROH regional association study. Furthermore, with the exception of an ROH association across the human leukocyte antigen (HLA) region for rheumatoid arthritis[71] and possibly one at 1q31.3 with bone mineral density[72], these associations are not robust and have not been replicated. Thus, although homozygosity mapping in inbred populations has been exceptionally successful for monogenic recessive disorders, ROH mapping studies have had less success for complex traits. Nevertheless, when performed in well-powered samples such as the UK Biobank and other very large studies, individual ROH associations have the advantage of potentially being able to identify rare recessive variants in a cost-efficient manner without large-scale WGS.

Final considerations

ROH studies are proliferating, with applications to diverse topics, from improved discovery of human knockouts to estimation of the human mutation rate (BOX 3). Isolated populations yield a higher number of homozygous loss-of-function mutations per sequence than cosmopolitan groups[73,74], with consanguineous populations providing a yet greater harvest[41,75], in accordance with their differing burdens of homozygosity (FIGS 2b,3d).

ROH calling. We have already noted the different kinds of genotype data, assays, algorithms and parameterizations that can be used in ROH calling, and such heterogeneity is also prevalent in clinical genetics laboratories[54]. In some circumstances, it is likely that these differences could influence not only the measured ROH but also the apparent strength of association with traits. To help aid comparability of results, we suggest that studies of ROH using microarray data should adopt the following criteria. The microarray should be genome-wide and have at least 300,000 SNPs, and the presented results should include (accompanied by standard error) the use of the same protocol as the ROHgen consortium[20]: ROH, called by PLINK, should consider SNPs >5% allele frequency, have at least 50 SNPs and be 1.5 Mb long, with allowance for missing and heterozygous calls. This is not to preclude presentation of central results on another basis, for example, also including

Darwinian fitness
The expected relative contribution of an individual or allele to the next generation of the population. It is the ability of an organism of a particular genotype to survive and leave viable offspring in its particular environment, captured in the phrase 'the survival of the fittest', although reproduction of the fittest might be more apt.

Panmixia
Random mating rather than mating structured by geography, ethnicity, socioeconomic status or other factors.

Box 3 | **Estimating the human mutation rate from ROH**

The mutation rate is a central parameter in biology — being the key to timing of the molecular clock as well as informing our expectations of mutations in cancer and understanding the incidence of genetic diseases — yet it has proved remarkably difficult to ascertain accurately[97]. There are two main methods of measuring the rate, the first being a phylogenetic approach, such as comparing sequences of a well-dated ancient sample with a modern sample. More commonly, multiple modern samples are sequenced, and the time separating them is estimated by calibration to an external reference, such as fossil evidence for the split between humans and chimpanzees; numerous such studies estimated the rate to be about 10^{-9} per site per year. The large number of generations separating the samples means that many mutations are observed; however, there are a number of downsides, including the impact of ancestral polymorphism and the fact that calibration is only as good as the accuracy of the fossil date and assumed phylogenetic tree. The second approach is to use direct observation of mutations in pedigrees, such as parent–offspring trios, using whole-genome or exome sequencing. However, the number of *de novo* mutations is low because of the small number of generations being assayed. Much debate has ensued as pedigree estimates of mutation rates are consistently about half of those using fossil calibrations ($0.4–0.6 \times 10^{-9}$ per nucleotide per year, with generation times between 20 and 30 years[98]).

Runs of homozygosity (ROH) provide a third method[99,100] that circumvents some of these caveats: heterozygous mutations found within autozygous segments must have arisen since the most recent common ancestor (MRCA); hence, they provide another direct estimate of the mutation rate[3]. The number of generations to the MRCA can be inferred from the ROH length distribution and verified from pedigree information. Thus, no external calibration is required, and because this approach makes use of mutations that have occurred over many generations, there is good statistical power. Care must be taken to correct for both the effects of gene conversion and errors arising from the inaccurate calling of ROH ends. Analysis of 4,353 exome sequences from consanguineous British South Asians[41] revealed >10 gigabases of autozygous sequences containing 932 *de novo* mutations, with an average of 6.6 generations separating the two copies (a little more than first cousins), providing an estimate equivalent to 0.5×10^{-9} mutations per exonic nucleotide per year, assuming 30-year generations[101]. Approaches using any identical-by-descent haplotype sharing (including one copy) across deeper time depths provide similar estimates[102,103].

shorter ROH, but presentation somewhere against a common baseline would enable readers to compare and contrast results and see how changes in method influence results. The inbreeding coefficient F_{ROH} can be calculated as the SROH divided by the length of the autosomal genome. The continued evolution of sequencing technologies means more research is required before recommendations can be drafted for ROH calling with these data. Good practices in quantitative genetics, such as large sample sizes, control of population structure, replication studies and accurate matching of cases and controls, or better, using case–cohort studies, will increase the likelihood of robust findings.

Future directions. The very large genomic data sets now becoming available[76–79] offer researchers a unique opportunity to better understand the influence of ROH on complex disease architecture. Such data sets will allow well-powered, broad surveys of phenotypes — including omic analytes that are mechanistically proximal to the gene and potentially fitness-related immune traits — to delineate the scale of inbreeding depression and to identify genomic regions with recessive effects on both complex traits and rare Mendelian diseases. A number of future research questions are suggested here. Does the burden of homozygosity caused by reduced population size have the same effect as recent consanguinity, and can this be used to infer whether the variants responsible for inbreeding depression are rare or common, recessive or overdominant? Recent consanguinity gives rise to long ROH that bring almost all variants, from common to very rare, into a homozygous state. By contrast, more distant inbreeding causes shorter ROH that make homozygous only the variants present in the shared ancestor, which are by definition common variants. If (unobserved) mutations have occurred in either the maternal or paternal line since the time of a common ancestor, these will not be homozygous in the inbred individual. Thus, analysing the effect of different lengths of ROH may reveal the relative contributions of rare and common variants: greater effects per megabase for larger ROH imply that rarer variants are causing the inbreeding depression.

Is there any evidence that mixed-race individuals differ in fitness-related traits from their peers, owing to either heterosis or outbreeding depression? Is the effect of inbreeding sex-specific in humans, as has been observed in other species[80]? Is it possible to identify specific genomic regions where ROH influence complex traits, and if so, do these loci correspond to known GWAS hits, or does regional ROH mapping offer a complementary method for identifying novel biology? One of the strengths of ROH analyses is that long homozygous segments can be reliably identified even from relatively modest marker densities. However, the increasing availability of WGS will soon allow shorter ROH to be more reliably called, in larger data sets, than is currently possible. This should permit the effect of very short ROH on disease risk to be quantified as well as potentially shedding further light on the demographic history of human populations. ROH studies will further illuminate the scope and mechanism of inbreeding depression in humans.

Gene conversion
A mechanism of recombination where one DNA sequence is replaced by a highly homologous one, leaving the sequences identical. In mammals, gene conversion tracts are usually short, between 200 bp and 1 kb.

Heterosis
Also called hybrid vigour; the propensity when inbred lines of, for example, maize or domesticated animals are crossed to result in hybrids that are fitter than either parent. The trait values that were reduced by inbreeding depression increase after outbreeding.

Outbreeding depression
When the offspring of distantly related mates are less fit than the parents; for example, if one homozygote has the highest fitness, outbreeding will usually increase the number of heterozygotes and thus reduce fitness.

1. Cavalli-Sforza, L. L. & Bodmer, W. *The Genetics of Human Populations* (W. H. Freeman & Co Ltd, 1978).
2. Bittles, A. H. & Black, M. L. Consanguinity, human evolution, and complex diseases. *Proc. Natl Acad. Sci. USA* **107**, 1779–1786 (2010).
3. Broman, K. W. & Weber, J. L. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am. J. Hum. Genet.* **65**, 1493–1500 (1999).
This seminal study is the first to identify long ROH, showing they are common in humans.
4. Jones, C. M. R. M. *Atlas of World Population History*. (Facts On File, 1978).
5. Biraben, J.-N. An essay concerning mankind's demographic evolution. *J. Hum. Evol.* **9**, 655–663 (1980).
6. Gunderson, R. C. *Connecting Your Pedigree Into Royal, Noble and Medieval Families*. (Genealogical Society of Utah, 1980).
7. Keller, M. C., Visscher, P. M. & Goddard, M. E. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* **189**, 237–249 (2011).
This paper shows that F_{ROH} is the preferred genomic inbreeding measure and that sample sizes in the tens of thousands will be needed to detect inbreeding depression in humans.
8. Donnelly, K. P. The probability that related individuals share some section of genome identical by descent. *Theor. Popul. Biol.* **23**, 34–63 (1983).
9. Rohde, D. L., Olson, S. & Chang, J. T. Modelling the recent common ancestry of all living humans. *Nature* **431**, 562–566 (2004).
10. Garrod, A. E. The incidence of alkaptonuria: a study in chemical individuality. *Lancet Infect. Dis.* **2**, 1616–1620 (1902).
11. Hoffman, J. I. *et al.* High-throughput sequencing reveals inbreeding depression in a natural population. *Proc. Natl Acad. Sci. USA* **111**, 3775–3780 (2014).
12. Huisman, J., Kruuk, L. E., Ellis, P. A., Clutton-Brock, T. & Pemberton, J. M. Inbreeding depression across the lifespan in a wild mammal population. *Proc. Natl Acad. Sci. USA* **113**, 3585–3590 (2016).
Using a well-studied wild deer population from Scotland with genomic data, this paper finds effects of homozygosity on offspring survival, birth weight, juvenile survival and other components of fitness.
13. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796 (2009).
14. Szpiech, Z. A. *et al.* Long runs of homozygosity are enriched for deleterious variation. *Am. J. Hum. Genet.* **93**, 90–102 (2013).
15. Alsalem, A. B., Halees, A. S., Anazi, S., Alshamekh, S. & Alkuraya, F. S. Autozygome sequencing expands the horizon of human knockout research and provides novel insights into human phenotypic variation. *PLoS Genet.* **9**, e1004030 (2013).
16. Xue, Y. *et al.* Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**, 242–245 (2015).
17. Scott, E. M. *et al.* Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat. Genet.* **48**, 1071–1076 (2016).
18. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
19. Pemberton, T. J. *et al.* Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* **91**, 275–292 (2012).
20. Joshi, P. K. *et al.* Directional dominance on stature and cognition in diverse human populations. *Nature* **523**, 459–462 (2015).
This paper is the largest study of ROH to date and found robust evidence for inbreeding effects on cognition and height-related traits in many populations across the world.
21. Gusev, A. *et al.* Whole population, genome-wide mapping of hidden relatedness. *Genom. Res.* **19**, 318–326 (2009).
22. Browning, S. R. & Browning, B. L. High-resolution detection of identity by descent in unrelated individuals. *Am. J. Hum. Genet.* **86**, 526–539 (2010).
23. Howrigan, D. P., Simonson, M. A. & Keller, M. C. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* **12**, 460 (2011).
This paper uses both simulated and real data to show that PLINK outperformed other software for the detection of ROH.
24. Szpiech, Z. A., Blant, A. & Pemberton, T. J. GARLIC: Genomic Autozygosity Regions Likelihood-based Inference and Classification. *Bioinformatics* **33**, 2059–2062 (2017).
25. Zhuang, Z., Gusev, A., Cho, J. & Pe'er, I. Detecting identity by descent and homozygosity mapping in whole-exome sequencing data. *PLoS ONE* **7**, e47618 (2012).
26. Pippucci, T., Magi, A., Gialluisi, A. & Romeo, G. Detection of runs of homozygosity from whole exome sequencing data: state of the art and perspectives for clinical, population and epidemiological studies. *Hum. Hered.* **77**, 63–72 (2014).
27. Magi, A. *et al.* H3M2: detection of runs of homozygosity from whole-exome sequencing data. *Bioinformatics* **30**, 2852–2859 (2014).
28. Narasimhan, V. *et al.* BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**, 1749–1751 (2016).
29. Gibson, J., Morton, N. E. & Collins, A. Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* **15**, 789–795 (2006).
This seminal paper demonstrates that ROH are ubiquitous in human populations.
30. Simon-Sanchez, J. *et al.* Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum. Mol. Genet.* **16**, 1–14 (2007).
31. Lencz, T. *et al.* Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc. Natl Acad. Sci. USA* **104**, 19942–19947 (2007).
32. Li, L. H. *et al.* Long contiguous stretches of homozygosity in the human genome. *Hum. Mutat.* **27**, 1115–1121 (2006).
33. International HapMap, C. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
34. Curtis, D. Extended homozygosity is not usually due to cytogenetic abnormality. *BMC Genet.* **8**, 67 (2007).
35. McQuillan, R. *et al.* Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **83**, 359–372 (2008).
Using well-studied isolate populations, this paper shows a strong correlation of genomic and pedigree inbreeding coefficients and that outbred individuals could harbour ROH up to 4 Mb in length.
36. Wright, S. Coefficients of Inbreeding and relationship. *Amer. Naturalist* **56**, 330–338 (1922).
37. Woods, C. G. *et al.* Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am. J. Hum. Genet.* **78**, 889–896 (2006).
38. Kirin, M. *et al.* Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **5**, e13996 (2010).
This survey of ROH across different populations, continents and demographic histories allows classification of populations into four major groups in terms of their ROH burden.
39. Karafet, T. M. *et al.* Extensive genome-wide autozygosity in the population isolates of Daghestan. *Europ. J. Hum. Genet.* **23**, 1405–1412 (2015).
40. Mezzavilla, M. *et al.* Increased rate of deleterious variants in long runs of homozygosity of an inbred population from Qatar. *Hum. Hered.* **79**, 14–19 (2015).
41. Narasimhan, V. M. *et al.* Health and population effects of rare gene knockouts in adult humans with related parents. *Science* **352**, 474–477 (2016).
This first large survey of gene knockouts in a consanguineous population describes homozygous loss of function for hundreds of genes.
42. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
43. Waldman, Y. Y. *et al.* The genetic history of Cochin Jews from India. *Hum. Genet.* **135**, 1127–1143 (2016).
44. Gilbert, E., Carmi, S., Ennis, S., Wilson, J. F. & Cavalleri, G. L. Genomic insights into the population structure and history of the Irish Travellers. *Sci. Rep.* **7**, 42187 (2017).
45. Esko, T. *et al.* Genetic characterization of northeastern Italian population isolates in the context of broader European genetic diversity. *Europ. J. Hum. Genet.* **21**, 659–665 (2013).
46. Bryc, K. *et al.* Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc. Natl Acad. Sci. USA* **107** (Suppl. 2), 8954–8961 (2010).
47. Abdellaoui, A. *et al.* Association between autozygosity and major depression: stratification due to religious assortment. *Behav. Genet.* **43**, 455–467 (2013).
48. Abdellaoui, A. *et al.* Educational attainment influences levels of homozygosity through migration and assortative mating. *PLoS ONE* **10**, e0118935 (2015).
This study is a great example of how confounding effects, such as assortative mating, can influence ROH analyses.
49. Nalls, M. A. *et al.* Measures of autozygosity in decline: globalization, urbanization, and its implications for medical genetics. *PLoS Genet.* **5**, e1000415 (2009).
50. Higasa, K. *et al.* Evaluation of haplotype inference using definitive haplotype data obtained from complete hydatidiform moles, and its significance for the analyses of positively selected regions. *PLoS Genet.* **5**, e1000468 (2009).
51. Papenhausen, P. *et al.* UPD detection using homozygosity profiling with a SNP genotyping microarray. *Am. J. Med. Genet. A* **155A**, 757–768 (2011).
52. Schaaf, C. P. *et al.* Identification of incestuous parental relationships by SNP-based DNA microarrays. *Lancet* **377**, 555–556 (2011).
53. Sund, K. L. *et al.* Regions of homozygosity identified by SNP microarray analysis aid in the diagnosis of autosomal recessive disease and incidentally detect parental blood relationships. *Genet. Med.* **15**, 70–78 (2013).
54. Grote, L. *et al.* Variability in laboratory reporting practices for regions of homozygosity indicating parental relatedness as identified by SNP microarray testing. *Genet. Med.* **14**, 971–976 (2012).
55. Hawass, Z. *et al.* Ancestry and pathology in King Tutankhamun's family. *J. Am. Med. Assoc.* **303**, 638–647 (2010).
56. Leutenegger, A. L., Sahbatou, M., Gazal, S., Cann, H. & Genin, E. Consanguinity around the world: what do the genomic data of the HGDP-CEPH diversity panel tell us? *Europ. J. Hum. Genet.* **19**, 583–587 (2011).
57. Jalkh, N. *et al.* Genome-wide inbreeding estimation within Lebanese communities using SNP arrays. *Europ. J. Hum. Genet.* **23**, 1364–1369 (2015).
58. Curtis, D., Vine, A. E. & Knight, J. Study of regions of extended homozygosity provides a powerful method to explore haplotype structure of human populations. *Ann. Hum. Genet.* **72**, 261–278 (2008).
59. Nothnagel, M. *et al.* Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Hum. Mol. Genet.* **19**, 2927–2935 (2010).
This is the first study to perform in-depth analysis of ROH islands, regions of the genome where a high proportion of people are homozygous.
60. Lander, E. S. & Botstein, D. Homozygosity mapping — a way to map human recessive traits with the DNA of inbred children. *Science* **236**, 1567–1570 (1987).
61. Rudan, I., Campbell, H., Carothers, A. D., Hastie, N. D. & Wright, A. F. Contribution of consanguinuity to polygenic and multifactorial diseases. *Nat. Genet.* **38**, 1224–1225 (2006).
62. Keller, M. C. *et al.* Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. *PLoS Genet.* **8**, e1002656 (2012).
63. Johnson, E. C. *et al.* No reliable association between runs of homozygosity and schizophrenia in a well-powered replication study. *PLoS Genet.* **12**, e1006343 (2016).
64. Panagiotou, O. A., Willer, C. J., Hirschhorn, J. N. & Ioannidis, J. P. The power of meta-analysis in genome-wide association studies. *Annu. Rev. Genom. Hum. Genet.* **14**, 441–465 (2013).
65. Power, R. A. *et al.* A recessive genetic model and runs of homozygosity in major depressive disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **165B**, 157–166 (2014).
66. McQuillan, R. *et al.* Evidence of inbreeding depression on human height. *PLoS Genet.* **8**, e1002655 (2012).
67. Howrigan, D. P. *et al.* Genome-wide autozygosity is associated with lower general cognitive ability. *Mol. Psychiatry* **21**, 837–843 (2016).
68. Morton, N. E. Effects of inbreeding on IQ and mental retardation. *Proc. Natl Acad. Sci. USA* **75**, 3906–3908 (1978).
69. Schull, W. J. Inbreeding and maternal effects in the Japanese. *Eugen. Q.* **9**, 14–22 (1962).

70. Rudan, I. *et al.* Quantifying the increase in average human heterozygosity due to urbanisation. *Europ. J. Hum. Genet.* **16**, 1097–1102 (2008).
71. Yang, H. C., Chang, L. C., Liang, Y. J., Lin, C. H. & Wang, P. L. A genome-wide homozygosity association study identifies runs of homozygosity associated with rheumatoid arthritis in the human major histocompatibility complex. *PLoS ONE* **7**, e34840 (2012).
72. Yang, T. L. *et al.* Genome-wide survey of runs of homozygosity identifies recessive loci for bone mineral density in Caucasian and Chinese populations. *J. Bone Miner. Res.* **30**, 2119–2126 (2015).
73. Kaiser, V. B. *et al.* Homozygous loss-of-function variants in European cosmopolitan and isolate populations. *Hum. Mol. Genet.* **24**, 5464–5474 (2015).
74. Lim, E. T. *et al.* Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* **10**, e1004494 (2014).
75. Saleheen, D. *et al.* Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* **544**, 235–239 (2017).
76. Sudlow, C. *et al.* UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
77. Gaziano, J. M. *et al.* Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
78. Nagai, A. *et al.* Overview of the BioBank Japan project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
79. Chen, Z. *et al.* China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int. J. Epidemiol.* **40**, 1652–1666 (2011).
80. Ebel, E. R. & Phillips, P. C. Intrinsic differences between males and females determine sex-specific consequences of inbreeding. *BMC Evol. Biol.* **16**, 36 (2016).
81. Darwin, C. R. *The Variation of Animals and Plants Under Domestication.* (John Murray, 1868).
82. Darwin, C. R. *The Effects of Cross and Self Fertilisation in the Vegetable Kingdom.* (John Murray, 1876).
83. Berra, T. M. *Darwin & His Children: His Other Legacy.* (Oxford Univ. Press, 2013).
84. Berra, T. M., Alvarez, G. & Ceballos, F. C. Was the Darwin/Wedgwood dynasty adversely affected by consanguinity? *Bioscience* **60**, 376–383 (2010).
85. Alvarez, G., Ceballos, F. C. & Berra, T. M. Darwin was right: inbreeding depression on male fertility in the Darwin family. *Biol. J. Linn. Soc.* **114**, 474–483 (2015).
86. Ober, C., Hyslop, T. & Hauck, W. W. Inbreeding effects on fertility in humans: evidence for reproductive compensation. *Am. J. Hum. Genet.* **64**, 225–231 (1999).
87. Bittles, A. H. & Neel, J. V. The costs of human inbreeding and their implications for variations at the DNA level. *Nat. Genet.* **8**, 117–121 (1994).
88. Ceballos, F. C. & Alvarez, G. Royal dynasties as human inbreeding laboratories: the Habsburgs. *Heredity* **111**, 114–121 (2013).
89. Alvarez, G. & Ceballos, F. C. Royal inbreeding and the extinction of lineages of the Habsburg dynasty. *Hum. Hered.* **80**, 62–68 (2015).
90. Alvarez, G., Ceballos, F. C. & Quinteiro, C. The role of inbreeding in the extinction of a European royal dynasty. *PLoS ONE* **4**, e5174 (2009).
91. Cassidy, L. M. *et al.* Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proc. Natl Acad. Sci. USA* **113**, 368–373 (2016).
92. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
93. Kılınç, G. M. *et al.* The demographic development of the first farmers in Anatolia. *Curr. Biol.* **26**, 2659–2666 (2016).
94. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
95. Kuhlwilm, M. *et al.* Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* **530**, 429–433 (2016).
96. Prado-Martinez, J. *et al.* Great ape genetic diversity and population history. *Nature* **499**, 471–475 (2013).
97. Callaway, E. DNA mutation clock proves tough to set. *Nature* **519**, 139–140 (2015).
98. Kong, A. *et al.* Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
99. Campbell, C. D. *et al.* Estimating the human mutation rate using autozygosity in a founder population. *Nat. Genet.* **44**, 1277–1281 (2012).
100. Alkuraya, F. S. Autozygome decoded. *Genet. Med.* **12**, 765–771 (2010).
101. Narasimhan, V. M. *et al.* in *65th Annual Meeting of The American Society of Human Genetics* PgmNr 353 (Baltimore, MD, 2015).
102. Lipson, M. *et al.* Calibrating the human mutation rate via ancestral recombination density in diploid genomes. *PLoS Genet.* **11**, e1005550 (2015).
103. Palamara, P. F., Lencz, T., Darvasi, A. & Pe'er, I. Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.* **91**, 809–822 (2012).
104. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
105. Behar, D. M. *et al.* The genome-wide structure of the Jewish people. *Nature* **466**, 238–U112 (2010).
106. Busby, G. B. J. *et al.* The role of recent admixture in forming the contemporary west Eurasian genomic landscape. *Curr. Biol.* **25**, 2518–2526 (2015).
107. Henn, B. M. *et al.* Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc. Natl Acad. Sci. USA* **108**, 5154–5162 (2011).
108. Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
109. Hellenthal, G. *et al.* A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
110. Metspalu, M. *et al.* Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am. J. Hum. Genet.* **89**, 731–744 (2011).
111. Pagani, L. *et al.* Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.* **91**, 83–96 (2012).
112. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
113. Schlebusch, C. M. *et al.* Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* **338**, 374–379 (2012).
114. Hodoglugil, U. & Mahley, R. W. Turkish population structure and genetic ancestry reveal relatedness among Eurasian populations. *Ann. Hum. Genet.* **76**, 128–141 (2012).
115. Yunusbayev, B. *et al.* The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol. Biol. Evol.* **29**, 359–365 (2012).
116. Vine, A. E. *et al.* No evidence for excess runs of homozygosity in bipolar disorder. *Psychiatr. Genet.* **19**, 165–170 (2009).
117. Nalls, M. A. *et al.* Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics* **10**, 183–190 (2009).
118. Ghani, M. *et al.* Association of long runs of homozygosity with Alzheimer disease among african american individuals. *JAMA Neurol.* **72**, 1313–1323 (2015).
119. Chahrour, M. H. *et al.* Whole-Exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genet.* **8**, 236–244 (2012).
120. Casey, J. P. *et al.* A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum. Genet.* **131**, 565–579 (2012).
121. Lin, P. I. *et al.* Runs of homozygosity associated with speech delay in autism in a taiwanese han population: evidence for the recessive model. *PLoS ONE* **8**, e72056 (2013).
122. Gamsiz, E. D. *et al.* Intellectual disability is associated with increased runs of homozygosity in simplex autism. *Am. J. Hum. Genet.* **93**, 103–109 (2013).
123. Gandin, I. *et al.* Excess of runs of homozygosity is associated with severe cognitive impairment in intellectual disability. *Genet. Med.* **17**, 396–399 (2015).
124. Iourov, I. Y., Vorsanova, S. G., Korostelev, S. A., Zelenova, M. A. & Yurov, Y. B. Long contiguous stretches of homozygosity spanning shortly the imprinted loci are associated with intellectual disability, autism and/or epilepsy. *Mol. Cytogenet.* **8**, 77 (2015).
125. Melhem, N. M. *et al.* Characterizing runs of homozygosity and their impact on risk for psychosis in a population isolate. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **165B**, 521–530 (2014).
126. Bacolod, M. D. *et al.* The signatures of autozygosity among patients with colorectal cancer. *Cancer Res.* **68**, 2610–2621 (2008).
127. Spain, S. L. *et al.* Colorectal cancer risk is not associated with increased levels of homozygosity in a population from the United Kingdom. *Cancer Res.* **69**, 7422–7429 (2009).
128. Siraj, A. K. *et al.* Colorectal cancer risk is not associated with increased levels of homozygosity in Saudi Arabia. *Genet. Med.* **14**, 720–728 (2012).
129. Hosking, F. J. *et al.* Genome-wide homozygosity signatures and childhood acute lymphoblastic leukemia risk. *Blood* **115**, 4472–4477 (2010).
130. Enciso-Mora, V., Hosking, F. J. & Houlston, R. S. Risk of breast and prostate cancer is not associated with increased homozygosity in outbred populations. *Europ. J. Hum. Genet.* **18**, 909–914 (2010).
131. Orloff, M. S., Zhang, L., Bebek, G. & Eng, C. Integrative genomic analysis reveals extended germline homozygosity with lung cancer risk in the PLCO cohort. *PLoS ONE* **7**, e31975 (2012).
132. Thomsen, H. *et al.* Inbreeding and homozygosity in breast cancer survival. *Sci. Rep.* **5**, 16467 (2015).
133. Thomsen, H. *et al.* Runs of homozygosity and inbreeding in thyroid cancer. *BMC Cancer* **16**, 227 (2016).
134. McLaughlin, R. L. *et al.* Homozygosity mapping in an Irish ALS case-control cohort describes local demographic phenomena and points towards potential recessive risk loci. *Genomics* **105**, 237–241 (2015).
135. McWhirter, R. E., McQuillan, R., Visser, E., Counsell, C. & Wilson, J. F. Genome-wide homozygosity and multiple sclerosis in Orkney and Shetland Islanders. *Europ. J. Hum. Genet.* **20**, 198–202 (2012).
136. Binzer, S. *et al.* High inbreeding in the Faroe Islands does not appear to constitute a risk factor for multiple sclerosis. *Mult. Scler.* **21**, 996–1002 (2015).
137. Christofidou, P. *et al.* Runs of homozygosity: association with coronary artery disease and gene expression in monocytes and macrophages. *Am. J. Hum. Genet.* **97**, 228–237 (2015).
138. Kuningas, M. *et al.* Runs of homozygosity do not influence survival to old age. *PLoS ONE* **6**, e22580 (2011).
139. Yang, T. L. *et al.* Runs of homozygosity identify a recessive locus 12q21.31 for human adult height. *J. Clin. Endocrinol. Metab.* **95**, 3777–3782 (2010).
140. Power, R. A., Nagoshi, C., DeFries, J. C., Wellcome Trust Case Control Consortium 2 & Plomin, R. Genome-wide estimates of inbreeding in unrelated individuals and their association with cognitive ability. *Europ. J. Hum. Genet.* **22**, 386–390 (2014).
141. Verweij, K. J. *et al.* The association of genotype-based inbreeding coefficient with a range of physical and psychological human traits. *PLoS ONE* **9**, e103102 (2014).

Acknowledgements
This work was supported by the Medical Research Council Human Genetics Unit quinquennial programme grant 'QTL in Health and Disease.' F.C.C. is supported by the South African National Research Foundation (NRF), and M.R. holds a South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Technology and administered by the NRF. The authors thank T. Gonzalez for help with figures and G. Alvarez, R. Vilas, O. Polašek, T. Esko, A. Wright, H. Campbell and C. Haley for helpful discussions and comments on the manuscript.

Author contributions
F.C.C. and J.F.W. researched data for the article. F.C.C., P.K.J., D.W.C. and J.F.W. wrote the manuscript. All authors contributed to reviewing and editing the manuscript before submission.

Competing interests statement
The authors declare no competing financial interests.

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

FURTHER INFORMATION
Jim Wilson's homepage: http://www.ed.ac.uk/mrc-human-genetics-unit/research/wilson-group

ALL LINKS ARE ACTIVE IN THE ONLINE PDF