Autozygosity influences cardiometabolic disease-associated traits in the AWI-Gen sub-Saharan African study

Francisco C. Ceballos1, Scott Hazelhurst1,2, David W. Clark3, Godfred Agongo1,4, Gershim Asiki5, Palwende R. Boua1,6,7, F. Xavier Gómez-Olivé8, Felistas Mashinya9, Shane Norris1,6, James F. Wilson3,10 & Michèle Ramsay1,6✉

The analysis of the effects of autozygosity, measured as the change of the mean value of a trait among offspring of genetic relatives, reveals the existence of directional dominance or overdominance. In this study we detect evidence of the effect of autozygosity in 4 out of 13 cardiometabolic disease-associated traits using data from more than 10,000 sub-Saharan African individuals recruited from Ghana, Burkina Faso, Kenya and South Africa. The effect of autozygosity on these phenotypes is found to be sex-related, with inbreeding having a significant decreasing effect in men but a significant increasing effect in women for several traits (body mass index, subcutaneous adipose tissue, low-density lipoproteins and total cholesterol levels). Overall, the effect of inbreeding depression is more intense in men. Differential effects of inbreeding depression are also observed between study sites with different night-light intensity used as proxy for urban development. These results suggest a directional dominant genetic component mediated by environmental interactions and sex-specific differences in genetic architecture for these traits in the Africa Wits-INDEPTH partnership for Genomic Studies (AWI-Gen) cohort.
Obesity and its associated cardiometabolic diseases (CMDs) have been rapidly increasing in sub-Saharan Africa. The continent is in the turmoil of an epidemiological transition characterized by complex patterns of change in health and diseases interacting with demographic, dietary, economic and social determinants. However, this health and demographic transition is very heterogeneous across the African continent and varies dependant on the epidemiological transition at the site. The main purpose of the AWI-Gen (Africa Wits-INDEPTH partnership for Genomic studies) H3Africa Consortium study is to examine the prevalence of CMD-associated risk factors, regional burden, and to explore gene-gene and gene-environment interactions that contribute to disease risk. To achieve its goals, AWI-Gen collected genotype and phenotype data on more than 10,000 individuals from rural and urban sites across four sub-Saharan African (SSA) countries (Fig. 1). Biomarkers of CMD included six anthropometric traits (height, weight, body mass index (BMI), waist-to-hip ratio (WHR) and visceral and subcutaneous adipose tissue (VAT and SCAT)), four classical lipid traits (total-cholesterol (TC), high density lipoprotein cholesterol (HDL), low density lipoprotein cholesterol (LDL) and triglycerides (TG)) and three circulatory traits (pulse rate, diastolic and systolic blood pressure). The objective of this study is to learn more about the genetic architecture of CMD-associated traits in SSA by analyzing the contribution of inbreeding depression (ID) to these traits.

Dissecting the genetic architecture of complex traits provides a deeper understanding of disease etiology and insights that could contribute to screening, diagnosis, prognosis and therapy. ID, manifest by an effect of genomic homozygosity on phenotypic values, implies some degree of directional dominance or overdominance in the genetic architecture. ID is influenced by two main factors: the amount of inbreeding and thus autozygosity; and the degree to which the dominance of causal loci is biased in one direction (which can be mainly caused by selection pressure). For this study, the inbreeding coefficient is calculated from genomic data through the analysis of runs of homozygosity (ROH: genomic tracts where homozygous markers occur in an uninterrupted sequence). African genetic homozygosity is particularly understudied, e.g., 30,000 individuals of African ancestry were included within the largest (~1.4 M) published study, and of those, only 1000 were resident in SSA itself, all of whom were urban dwellers. Eleven traits in total are shared by this analysis and the one performed by Clark et al. 2019, with our study replicating results for height, weight, total cholesterol, LDL, BMI, and triglycerides levels and presenting a statistically-powered ROH analysis of Africans.

Results and discussion

Inbreeding in the AWI-Gen cohort. The mean genomic inbreeding coefficient, FROH, for the entire AWI-Gen cohort is 0.0093, equivalent to their parents being on average about second cousins, once removed, but likely generated in many cases from the sum of multiple more distant relationships. No differences in the FROH were found between men and women or between urban and rural sites (FROH men = 0.0045, women = 0.0049) (FROH urban sites = 0.0051, rural sites = 0.0049).

Traits affected by Inbreeding Depression. The substantial levels of genomic homozygosity exhibited by the AWI-Gen cohort allow us to explore their effect on 13 CMD-associated traits and their relationship to socio-economic status (Fig. 2a). We detected...
a negative association of inbreeding depression with socio-economic-status within each site; in general, having a higher F_{ROH} was associated with a lower socio-economic-status. When the entire cohort is considered, we detected significant negative ID for weight, BMI and SCAT. However, the intensity of the ID for these traits was not very strong and the offspring of the equivalent of a second cousin mating had on average, a reduction for each of the significantly associated traits in areas with less luminosity (0.70 kg in weight, 0.024 mmol/L in TG). Moreover, for TG levels that were not significant after multiple test correction in the overall analysis, there is a strongly significant ID effect in the study sites with less luminosity (0.024 mmol/L). As the sample size for the less developed sites is almost double that for more urban sites, we have more power to detect the effects of ID there. However, the effect sizes for ID on SCAT, LDL and TG in more urban sites were significantly different from less developed sites ($P = 2.2 \times 10^{-16}$, $2.2 \times 10^{-5}$ and $3.2 \times 10^{-9}$ respectively, using standard ANCOVA), interestingly showing that the effects in less developed sites were in the opposite direction for these traits (Fig. 2b). We suggest that environmental changes affecting lifestyle factors may be playing a role since food resources, comfort commodities and even primary healthcare are limiting factors in less developed sites. The results indicate that the differences in the effects of ID between more and less developed study sites are not due to unmeasured confounding variables, since we observed a negative effect of autozygosity on socio-economic-status (SES) in both groups (of similar strength), that is, increasing autozygosity was associated with decreased SES. Further, we accounted for the effect of SES by including it as a covariate in all ID analyses (see Methods). In order to ameliorate potential

### Table 1

| Trait       | Beta_{FROH} | P-val |
|-------------|-------------|-------|
| SES.Q       | -5.1111     | 0.00018* |
| Height      | -1.4794     | 0.17  |
| Weight      | -4.6106     | 0.00025* |
| BMI         | -4.1334     | 5e-04* |
| WHR         | 2.1297      | 0.14  |
| VAT         | -2.1673     | 0.094 |
| SCAT        | -5.8759     | 2e-06* |
| LDL         | -1.6556     | 0.23  |
| HDL         | -2.0608     | 0.14  |
| TG          | -3.9416     | 0.0032 |
| TC          | -1.2332     | 0.34  |
| Pulse       | 2.4121      | 0.088 |
| Systolic BP | 0.6185      | 0.64  |
| Diastolic BP| -0.4036     | 0.76  |

### Table 2

| Trait       | Beta_{FROH} | P-val |
|-------------|-------------|-------|
| SES.Q       | -1.4786     | 0.0069 |
| Height      | -1.1397     | 0.38  |
| Weight      | -2.2073     | 0.23  |
| BMI         | -5.7136     | 7.7e-05* |
| WHR         | -1.9983     | 0.4   |
| VAT         | -5.6484     | 2.8e-05* |
| SCAT        | -10.7616    | 2.1e-12* |
| LDL         | 4.4777      | 0.023 |
| HDL         | -3.7897     | 0.015 |
| TG          | -7.0931     | 4.1e-06* |
| TC          | 2.8058      | 0.27  |
| Pulse       | 3.2434      | 0.057 |
| Systolic BP | -0.9949     | 0.53  |
| Diastolic BP| 0.7254      | 0.63  |

**Fig. 2 Effect of inbreeding depression on cardiometabolic disease-associated traits in the AWI-Gen cohort.** To allow comparison between traits, effects estimates ($\beta_{FROH}$) and 95% confidence intervals are shown in units of within-group standard deviations. A regular Bonferroni correction for multiple testing was applied; significance is shown with an asterisk (0.05/14). a Normalized ($\beta_{FROH}$) and 95% confidence intervals on 14 quantitative traits known to be cardiometabolic disease-associated traits: socio-economic status (SES.Q); six anthropometric traits, height, weight, body mass index (BMI), waist-to-hip ratio (WHR), visceral adipose tissue (VAT) and subcutaneous adipose tissue (SCAT); four classical blood lipids, high and low density lipoprotein cholesterol (HDL, LDL), total cholesterol (TC), and triglycerides (TG); and three circulatory traits, pulse rate, systolic blood pressure (Systolic BP) and diastolic blood pressure (Diastolic BP). Sample sizes, effect estimates in measurement units, standard deviations and p-values are shown in Table S1. b Differences in the effects of normalised ($\beta_{FROH}$) and 95% confidence intervals between sites with an average night-light intensity below and above 5. Sample sizes, effect estimates in measurement units, standard deviations and p-values for dwellers living in sites with average night-light intensity below and above 5 are shown in Tables S2 and S3 respectively.
coefficient using the genomic relationship matrix ($F_{GRM}$) (see methods section) and $F_{outsideROH}$. We first fitted bivariate models with $F_{ROH}$ and $F_{GRM}$ as explanatory variables and then we explored the effect of $F_{outsideROH}$, which is not an estimator of the inbreeding coefficient, but describes the homozygosity of common SNPs outside ROH. As can be seen in Supplementary Table 4, for all the traits that were significant in the overall univariate analysis (SES.Q, weight, BMI, SCAT and TG) we find that $\beta_{F_{GRM}}$ is of a greater magnitude than $\beta_{F_{ROH}}$ in the conditional analysis. Furthermore, for all of the traits but SCAT, $\beta_{F_{GRM}}$ does not differ from zero, thus indicating that for these traits the variation of $F_{GRM}$ is not associated with any change in trait values. Like $F_{GRM}$, $F_{outsideROH}$ which captures common SNPs in strong linkage disequilibrium (LD), was not significant for any trait or any model (Supplementary Tables 1–3). These results suggest that autozygous rare recessive variants found in ROH, rather than homozygous common variants in strong LD, are causing the inbreeding depression and this is consistent with the dominance hypothesis $^4,13$. The fact that the bivariate models show that $F_{ROH}$ captures the signal better than $F_{GRM}$ and $F_{outsideROH}$ also suggests that in this example inbreeding depression is caused by directional dominance. Overdominance—the positive selection on heterozygotes bringing alleles to intermediate frequency—would predict that more common homozygous SNPs outside ROH would also have an effect.

Two recent studies found evidence for the same conclusion. Clark et al. $^{10}$, analyzing a large sample size of 1.4 M found that for all the traits affected by inbreeding depression $\beta_{F_{GRM}}$ was of greater magnitude and in many cases $\beta_{F_{GRM}}$ was also non-significant. Our results therefore confirm what Clark et al. $^{10}$ and Johnson et al.$^{14}$ found.

**Genomic regional $F_{ROH}$ effects.** We tested whether the genome-wide ID effects we observed came from a small number of major loci with large effects or resulted from the polygenic effect of many loci with small effects. We thus divided the genome into 1,000 non-overlapping 3 Mb-wide windows and assessed whether homozygosity of these regions was associated with the CMD-associated traits (see Methods). Only two traits were suggestively significantly influenced by individual regions: BMI in men (Fig. 4a, Supplementary Table 5) and VAT in women (Fig. 4b, Supplementary Table 6). For all other traits we have no evidence of major loci that may be exerting large effects, rather the ID appears to be polygenic in origin. For BMI in men, a significant window on chromosome 14 was detected, including 23 protein-coding genes (Supplementary Table 5). Published genome-wide association studies (GWAS) have reported associations with BMI in two of these genes ($KCNH5^{15}$ and $FUT8^{16}$), which are known to be pleiotropic. $KCNH5$ (potassium voltage-gated channel subfamily H member 5) has been found to be associated not just with BMI but also with SCAT$^{17}$, intelligence$^{18}$, and sleep duration$^{19}$. $FUT8$ (fucosyltransferase 8) has also been associated with IgG glycosylation patterns$^{20}$, age at menarche$^{21}$, schizophrenia$^{22}$, plasma N-glycans$^{23}$, head circumference$^{24}$, the plasma proteome$^{25}$, multiple sclerosis$^{26}$ and systolic blood pressure$^{16}$. Thus, the strongest directionally dominant effects on BMI are shared with loci also imparting additive effects discovered by GWAS. Two windows on chromosomes 2 and 4 were found to be associated with VAT in women (Supplementary Table 6). No GWAS hits were previously found for this trait in these regions; however both regions include genes with multiple associations with other traits.

In this study we analysed the effect of inbreeding depression on 13 cardiometabolic disease-associated traits in the AWI-Gen
cohort with the intention of better understanding the genetic architecture of these traits in sub-Saharan Africa. We found significant evidence for inbreeding depression for four phenotypes, which was strongly enhanced in less developed rural regions and also stronger in men. Our results suggest a complex genetic architecture for these traits with interactions by rurality. Larger studies of African populations will further illuminate the factors contributing to complex disease risk in sub-Saharan African populations.

Methods

Population genetics overview. Our objective was to explore the directionally dominant component of the genetic architecture of 13 cardiometabolic disease (CMD) associated traits measured in 10,776 sub-Saharan African (SSA) individuals recruited by the AWI-Gen Initiative. In order to achieve this, we measured the effect of inbreeding depression on these risk factors. In classical population genetics, inbreeding depression (ID) is defined as the reduction of the mean fitness in a population because of inbreeding4. Currently this definition has been generalized for any complex trait as the change in the mean phenotypic value in a population because of inbreeding. Considering the combined effect of all the loci that affect a character, as far as the genotypic value of the loci combined additively is concerned, the mean character value of a population with inbreeding coefficient (F) is given by:

$$M_F = M_0 - 2F \sum d \bar{p} \bar{q},$$

(1)

where $$M_0$$ is the population mean before inbreeding and $$p$$ and $$q$$ are allele frequencies. The change of the mean on inbreeding is therefore3

$$-2F \sum dpq$$

(2)

This shows that inbreeding will change the mean value of a character in a population when the sum of the genotypic value of the heterozygotes ($$d$$) is different from 0 (i.e., the character needs to exhibit directionally dominant (or overdominant) genetic architecture).4 Further, when loci are combined additively, the change of the mean on inbreeding should be directly proportional to the coefficient of inbreeding.4 This allows us to detect inbreeding depression in complex characteristics exhibiting directional dominance using regression analysis as long as the population under study practices some inbreeding. It is important to realize that the genetic architecture of a character, including the effects of ID, is not necessarily constant among different populations. The intensity of the ID and the genetic architecture of a trait depend on the selection pressure, environmental factors and population structure inasmuch as genetic frequencies change between populations. The considerable burden of long ROH (>1.5 Mb) present in some sub-Saharan populations28 provides the opportunity to test whether CMD-associated traits gathered by the AWI-Gen initiative exhibit inbreeding depression.

AWI-Gen Cohort. The Africa Wits-INDEPTH Partnership for Genomic Studies (AWI-Gen) is an NIH-funded Collaborative Centre of the Human Heredity and Health in Africa (H3Africa) Consortium. It is a partnership between the University of the Witwatersrand and the International Network for Demographic Evaluation of Populations and Their Health (INDEPTH) that includes Health and Demographic Surveillance System (HDSS) Centres and the Developmental Pathways for Health Research Unit (DPhRUs), that have longitudinal cohorts in Navrongo, Ghana; Nanoro, Burkina Faso; Nairobi, Kenya; Agincourt, South Africa; Dikgale, South Africa and Soweto, South Africa. In total, over 12,500 participants were recruited between August 2013 and August 2016, of which 10,776 were aged 40–60 years. Pregnant women, first-degree relatives, recent immigrants and individuals with physical impairments preventing the measurement of different characteristics were excluded. The objectives of the AWI-Gen study can be summarized as follows: (1) To build capability for genomic research in these centers and countries by providing opportunities to develop skills including biostatistics, genomics, and bioinformatics. (2) To understand population structure and genetic architecture among recruiting sites in order to inform analysis strategies and evaluate impact across different ethnicities. (3) To understand the prevalence and genetic basis of cardiometabolic diseases (CMDs). The study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (Wits) (protocol numbers M121029 and M170880), and each contributing Centre obtained additional local ethics approval as required. Data and samples were collected following community engagement and individual informed consent.

Biomarkers of CMD gathered by the AWI-Gen study include anthropometric variables including height, weight, body mass index (BMI), waist-to-hip ratio (WHR) and fat distribution (visceral adipose tissue (VAT)) and subcutaneous adipose tissue (SCAT); lipid composition variables including low density lipoprotein cholesterol (LDL), high density lipoprotein cholesterol (HDL), total cholesterol (TC) and triglycerides (TG); circulatory traits including pulse rate and blood pressure (systolic and diastolic). Besides the above biomarkers a socioeconomic-status variable was also calculated for each individual. This variable (SES) was obtained by adding the items a person has in their household from among a list defined items, determined as appropriate for their country. The SES quintiles for each study site were calculated and individuals assigned to a quintile. This was used in the analysis to remove a potential cultural component related to socioeconomic-status as is shown in29. See Trait Definition section.

AWI-Gen individuals were genotyped using the H3Africa Custom Genotyping Array, which is a high-density genotyping array (2.267 million SNVs) designed by an H3Africa project team to maximize capturing common variation in African population. Manufactured by Illumina, the array has been used by several H3Africa projects. Details of the array and the SNVs captured can be found at https://www.h3abionet.org/h3afirica-chip. A total of 11,076 samples were genotyped. We removed duplicate and non-autosomal SNVs; SNVs with genotype missing rate > 0.01; minor allele frequency < 0.01 and SNVs that deviated from Hardy-Weinberg equilibrium test.
(p-value < 5.0 × 10^{-4}). Similarly, at the sample level, individuals with genotype missing rate > 0.02, that failed the sex check and one of the individuals from each related (PHIAT > 0.8—potential duplicates) were removed. Also, 159 first degree relatives were removed from the Agincourt site. These QC steps were performed using the H3Africa GWAS pipeline and resulted in the final dataset containing 1,733,121 SNVs and 10,617 individuals.

**Trait definition.** 14 different traits and different control variables like age, sex, site of sampling, education and occupation were also considered in the analysis. All traits are defined below under headings in the format (short name –) full name-units. Further details can be found in Ali et al.1

**Socio-economic-demographic variables**
- **age—years.** Age at data collection (calculated using the self-declared date of birth and the date of the interview with the individual).
- **sex.** Sex was self-reported and validated by genomic data.
- **site—Site of sampling.** Recruitment site where the sample was collected: Nanoro (Burkina Faso), Navronvo (Ghana), Nairobi (Kenya), Agincourt (South Africa), Dikgale (South Africa), Soweto (South Africa).
- **night-light—Night-time luminosity.** The average value of light pixels were obtained using night-light geo-tiffs from NOAA’s NGDC Earth Observation Group (EOG): https://ngdc.noaa.gov/eog/ and spatial boundaries are taken from the Global Administrative Boundaries Database: http://gadm.org/.
- **edu—Education.** The highest level of education was self-reported according to four categories: (1) No formal education, (2) Primary, (3) Secondary, (4) Tertiary. Tertiary qualifications such as certificates, diplomas, or degrees.
- **occu—Occupation.** Self-declared employment according to five categories. (1) Self-employed. (2) Formal full-time employment by someone else. (3) Part-time employment by someone else. (4) Informal employment (dependent on the availability of work). (5) Unemployed.

**SES.Q—Socio-economic-status quintiles.** The socio-economic status of the individual was calculated using a list of household goods appropriate for each site (not all sites included all variables as some items were not relevant in specific settings): electricity, solar energy, power generator, alternative power source, television, radio, motor vehicle, bicycle, refrigerator, washing machine, sewing machine, telephone, cell phone, microwave, DVD player, torch, gas lamp, kerosene lamp with a glass, wall sockets, water tap, toilet, window, table, sofa, set, wall clock, bed, mattress, blankets, cable, other livestock, poultry, tractor, plough, etc. In order to exclude units with qualifications such as certificates, diplomas, or degrees, different sites, the SES quintiles were calculated, and individuals assigned to a quintile. This approach was used to remove a potential cultural component related to socio-economic-status2.

**Anthropometry variables**
- **Height—meters.** Standing height measured in meters.
- **Weight—kg.** Weight measured in kilograms.
- **BMI—Body mass index—kg/m².** Weight in kilograms divided by height in meters squared.
- **WHR—Waist: Hip ratio—no units.** Calculated by dividing the individual waist circumference in centimeters by the hip circumference also in centimeters.
- **VAT—Visceral adiposity tissue—cm.** Visceral (medial) fat measured using ultrasound.
- **SCAT—Subcutaneous adiposity tissue—cm.** Subcutaneous (transverse) fat measured using ultrasound.
- **LDL—Low-density lipoprotein cholesterol—mmol/L.** Lipid composition (in fasting serum).
- **HDL—High-density lipoprotein cholesterol—mmol/L.**
- **TC—Total cholesterol—mmol/L.**
- **TG—Total triglycerides—mmol/L.**

**Circulatory**
- **Systolic BP—Systolic blood pressure—mmHg.** Three readings were taken during a single session, two minutes apart, the first was discarded and the second two readings were averaged.
- **Diastolic BP—Diastolic blood pressure—mmHg.** Three readings were taken during a single session, the first was discarded and the second two readings were averaged.
- **Pulse—Heart rate—beats per minute.** Three readings taken during a single session, the first was discarded and the second two readings were averaged.

**Assessing site economic and urban development.** As a proxy for the urban and economic development of the different AWI-Gen sites we used night-light intensity (luminosity). Luminosity has been widely used as a proxy in countries where GDP data are either not available or of poor quality13,14. It was found to be highly correlated with GDP per capita and other measures of prosperity, like electricity provision15, and can therefore be considered as a valid proxy13,14. More recently, using Demographic and Health Surveys (DHS) from 29 African countries, night-time light intensity has been found to correlate strongly with indicators of household wealth, education, and health. Also, the variation in night-time light explained a substantial share in the variation of these indicators16.

Night light data are made available by the National Geophysical Data Center (NGDC) of the National Oceanic and Atmospheric Administration of the US, and originate from images taken by satellites of the Defense Meteorological Satellite Program (DMSP) of the U.S. Department of Defense between 1992 and 2013. Night light intensity data are available on pixel (grid cell) level, with each pixel corresponding to 30 × 30 arc s, i.e., one value represents the average night light intensity of an area of 0.86 square kilometers (on the equator). Night light intensity is measured by an integer ranging from 0 (unlit) to 63. We use the latest version (4.0) of the data.

**ROH calling.** ROH longer than 300 Kb were called using PLINK software with the following parameters:
- **homozyg-snp 30.** Minimum number of SNPs that a ROH is required to have.
- **homozyg-kb 300.** Length in Kb of the sliding window.
- **homozyg-density 30.** Required minimum density to consider a ROH (1 SNP in 30 Kb).
- **homogyz-window-snp 30.** Number of SNPs that the sliding window must have.
- **homogyz-gap 1000.** Length in Kb between two SNPs in order to be considered in two different segments.

No linkage disequilibrium pruning was performed. These conditions were previously used and validated by different published studies and were shown to call ROH that correspond to autozygous segments in which all SNPs (including those not present on the array) are homozygous-by-descent28,37.

**Calculating genomic inbreeding coefficients.** The inbreeding coefficient or F<sub>E</sub> is defined as the probability that an individual receives two alleles identical-by-descent38. Traditionally, F<sub>E</sub> was measured using deep genealogies: currently, we can obtain an estimate of this parameter without having any genealogy by using genomic approaches. Different genomic inbreeding coefficients were calculated.

**F<sub>ROH</sub>** measures the actual proportion of the autosomal genome that is autozygous over and above a specific minimum length ROH threshold39. When analyzing ROH > 1.5 Mb, F<sub>ROH</sub> correlates most strongly (r = 0.86) with the inbreeding coefficient obtained from an accurate six-generation pedigree38. Using extended pedigrees of the royal European dynasties, with complex inbreeding loops, it has been found that above the 10th generation the change in the inbreeding coefficient is less than 1%40. Also, it has been found that individuals with no inbreeding loops in at least 5 generations (and probably 10) carried ROH up to 4 Mb in length but not longer41. F<sub>ROH</sub> using a genomic approach, captures the total inbreeding coefficient (F<sub>E</sub>) within the resolution of the data available and the size of ROH that can be called9.

**F<sub>GRM** An alternative genomic inbreeding coefficient was obtained using PLINK’s parameter —eb (Flag3). This coefficient described by Yang et al. 2011 (Eq1)41 is defined as:

\[
F_{GRM} = \frac{\sum_{i=1}^{N} \text{ROH}_i \times 1.5 \text{Mb}}{3 \text{Gb}}
\]

Where:
- \(\text{ROH}_i\) is the number of SNPs in the reference allele frequency of the ith SNP and \(N\) is the number of copies of the reference allele. The reference allele frequencies were site-specific and included only loci with MAF > 0.05.

Clark et al. 2020 showed that F<sub>GRM</sub> is downwardly biased in real data and that this is proportional to the ratio of the inbreeding coefficient (F<sub>E</sub>) and F<sub>GRM** as expected when the difference between F<sub>GRM</sub> and F<sub>ROH</sub> can be considered as estimation error. In the same work, they compared the relative abilities of F<sub>ROH</sub> and F<sub>GRM** to capture inbreeding using the pedigree information of 47927 Icelanders with a 10 generation pedigree. The correlation was highest for F<sub>ROH** (r = 0.779) in comparison to F<sub>GRM** (r = 0.682).

**F<sub>inbreed</sub>** An additional genomic inbreeding measure, F<sub>inbreed</sub>, was calculated as the genomic fraction of homozygous SNPs outside ROH.

\[
F_{inbreed} = \frac{\text{O(HOM)} - \text{E(HOM)}}{\text{N} - \text{E(HOM)}}
\]

Where:
- \(\text{O(HOM)} = \text{O(HOM)} - \text{NSNP}_{ROH}\)
- \(\text{E(HOM)} = \frac{\text{N} - \text{NSNP}_{ROH}}{\text{N}} \times \text{E(HOM)}\)

\[
N = N - \text{NSNP}_{ROH}
\]

Where: O(HOM) is the observed number of homozygous SNPs, E(HOM) is the expected number of homozygous SNP according to H-W proportions, N is the
measured trait values closer to the univariate effect of FROH. FROH is a known vector of individual \( \mathbf{F}_{\text{ROH}} \) as described above, \( \mathbf{b} \) is a vector of unknown covariates effects including the mean (\( \mu \)). \( \mathbf{X} \) is a known design matrix for the fixed effects, and \( \mathbf{e} \) is an unknown vector of residuals. The traits used in this study have been the subject of genome-wide association meta-analysis (GWAMA), phenotype modeling, such that inclusion of covariates were chosen with reference to leading consortia: The Genetic Investigation of Anthropometric Traits (GIANT) for anthropometry, Global Lipids Genetics Consortium (GLGC) for lipids composition International Consortium for Blood Pressure (ICBP) for blood pressures.

### Measuring genome-specific local Inbreeding Depression

To assess whether inbreeding depression is caused by common or rare variants, \( \mathbf{F}_{\text{ROH}} \) is an estimate of autozygosity which increases the homozygosity of all variants, both common and rare. In contrast, \( \mathbf{F}_{\text{GRM}} \) is calculated from common SNPs (\( >5\% \text{MAF} \)) and correlates well with the homozygosity of common SNPs but less with rare SNP which may be in weak LD. Following previous studies\(^{10}\) and in order to assess if ID is caused by common or rare variants, we performed bivariate models of all traits (\( \text{Trait} - F_{\text{ROH}} + F_{\text{GRM}} \)) to establish whether the observed inbreeding effects associate more strongly with \( F_{\text{ROH}} \) or \( F_{\text{GRM}} \). By analyzing the excess homozygosities of SNPs, extracted from the UK Biobank, at seven minor allele frequencies, Clark et al.\(^{10}\) showed that homozygosity of common SNPs is better predicted by \( F_{\text{GRM}} \), but rare variant homozygosity is better predicted by \( F_{\text{ROH}} \).

### Total number of non-missing genotyped SNPs and \( N_{\text{SNP,ROH}} \)

The total number of non-missing genotyped SNPs and \( N_{\text{SNP,ROH}} \) is the number of homozygous SNPs found in ROH.

### Testing inbreeding depression

Inbreeding depression (ID) was modeled by a multiple regression:

\[
y = \beta_0 + P_{\text{ROH}} \times \mathbf{F}_{\text{ROH}} + \mathbf{X} \times \mathbf{e} + \mathbf{r}
\]

Where \( y \) is a vector of measured trait values, \( \beta_0 \) is the univariate effect of \( F_{\text{ROH}} \), \( F_{\text{ROH}} \) is a known vector of individual \( \mathbf{F}_{\text{ROH}} \) as described above, \( \mathbf{b} \) is a vector of unknown covariates including the mean (\( \mu \)). \( \mathbf{X} \) is a known design matrix for the fixed effects, and \( \mathbf{e} \) is an unknown vector of residuals. The traits used in this study have been the subject of genome-wide association meta-analysis (GWAMA), phenotype modeling, such that inclusion of covariates were chosen with reference to leading consortia: The Genetic Investigation of Anthropometric Traits (GIANT) for anthropometry, Global Lipids Genetics Consortium (GLGC) for lipids composition International Consortium for Blood Pressure (ICBP) for blood pressures.

### Reporting summary

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

### Data availability

AWI-Gen phenotype dataset is available at database accession id: EGA000010002482. AWI-Gen genotype dataset accession id: EGA000010001996.

### Night-time luminosity

The average values of light pixels were obtained using night-light geo-tifs from NOAA’s NGDC Earth Observation Group (EOD): [https://ngdc.noaa.gov/eog/](https://ngdc.noaa.gov/eog/) and spatial boundaries are taken from the Global Administrative Boundaries Dataset [http://gadm.org/](http://gadm.org/). Data used correspond to the 2013 version, last accessed on February 10, 2020.

Received: 6 May 2020; Accepted: 12 October 2020; Published online: 13 November 2020

### References

1. Ali, S. A. et al. Genomic and environmental risk factors for cardiometabolic disorders in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. *Global Health Action* 11, 1507133 (2018).
2. Ramsay, M. et al. H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Glob. Health, Action, Med., Genomics* 1, e20 (2016).
3. Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J. & Richards, J. B. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.* 19, 110–124 (2018).
4. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* 10, 783–796 (2009).
5. Falconer, D. S. & Mackay, T. F. C. *Quantitative genetics* (Pearson, 1996).
6. Crnokrak, P. & Roff, D. A. Inbreeding depression in the wild. *Heredity* 83, 260–270 (1999).
7. Ballatton, T. & Kirkpatrick, M. Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. *Genetica* 75, 71–80 (2000).
8. Kirkpatrick, M. & Jorde, P. The effects of a Bottleneck on inbreeding depression and the genetic load. *Am. Naturalist* 155, 154–167 (2000).
9. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.* 19, 220–234 (2018).
10. Clark, D. W. et al. Association of autozygosity with a broad range of human phenotypes. *Nat. Commun.* 10, 4957 (2019).
11. Joshi, P. K. et al. Directional dominance on stature and cognition in diverse human populations. *Nature* 523, 459–462 (2015).
12. McQuillan, R. et al. Evidence of inbreeding depression on human height. *PLoS Genet.* 8, e1002655 (2012).
13. Nakatsuka, N. et al. The promise of discovering population-specific disease-associated genes in South Asia. *Nat. Genet.* 49, 1403–1407 (2017).
14. Johnson, E. C., Evans, L. M. & Keller, M. C. Relationships between estimated autozygosity and complex traits in the UK Biobank. *PLoS Genet.* 14, e1007556 (2018).
15. Namjou, B. et al. EMR-linked GWAS study: investigation of variation landscape of loci for body mass index in children. *Front. Genet.* 4, 268 (2013).
16. Kichaev, G. et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am. J. Hum. Genet.* 104, 65–75 (2019).
17. Fox, C. S. et al. Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet.* 8, e1002695 (2012).
18. Hill, W. D. et al. A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Mol. Psychiatry* 24, 169–181 (2019).
19. Doherty, A. et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nat. Commun.* 9, 5257 (2018).
20. Wahl, A. et al. Genome-Wide Association Study on Immunoglobulin G Glycosylation Patterns. *Front. Immunol.* 9, 277 (2018).
21. Perry, I. R. et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* 514, 92–97 (2014).
22. Goes, F. S. et al. Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am. J. Med. Genet. Part B: Neuropsychiatr. Genet.* 168, 549–659 (2015).
23. Huffman, J. E. et al. Polymorphisms in B3GAT1, SLC9A9 and MAGT5 are associated with variation within the human plasma N-glycome of 3533 European adults. *Hum. Mol. Genet.* 20, 5000–5011 (2011).
24. Comuzzie, A. G. et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One* **7**, e51954 (2012).
25. Sun, B. R. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
26. Baranzini, S. E. et al. Genetic variation influences glutamate concentrations in brains of patients with multiple sclerosis. *Brain* **133**, 2603–2611 (2010).
27. Crow, J. F. & Kimura, M. An introduction to population genetics theory. (Harper & Row, New York, 1970).
28. Ceballos, F. C., Hazelhurst, S. & Ramsay, M. Runs of homozygosity in sub-Saharan African populations provide insights into complex demographic histories. *Hum. Genet.* **138**, 1123–1142 (2019).
29. Kabudula, C. W. et al. Assessing changes In household socioeconomic status In rural South Africa, 2001–2013: a distributional analysis using household asset indicators. *Soc. Indic. Res.* **133**, 1047–1073 (2017).
30. Baichoo, S. et al. Developing reproducible bioinformatics analysis workflows for heterogeneous computing environments to support African genomics. *BMC Bioinformatics* **19**, 457 (2018).
31. Anthony, M. Night lights and regional income inequality in Africa. *WIDER Working Paper Series* **085** (2015).
32. Elliott, R. J. R., Strobl, E. & Sun, P. The local impact of typhoons on economic activity in China: a view from outer space. *J. Urban Econ.* **88**, 50–66 (2015).
33. Baskaran, T., Min, B. & Uppal, Y. Election cycles and electricity provision: evidence from a quasi-experiment with Indian special elections. *J. Public Econ.* **126**, 64–73 (2015).
34. Chen, X. & Nordhaus, W. D. Using luminosity data as a proxy for economic statistics. *Proc. Natl Acad. Sci. USA* **108**, 8589–8594 (2011).
35. Henderson, J. V., Storeygard, A. & Weil, D. N. Measuring economic growth in rural South Africa, 2001–2013: a distributional analysis using household asset indicators. *Soc. Indic. Res.* **133**, 1047–1073 (2017).
36. Bruenderle, A. & Hodler, R. Nighttime lights as a proxy for human development at the local level. *PLoS One* **13**, e0202231 (2018).
37. Ceballos, F. C., Hazelhurst, S. & Ramsay, M. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics* **19**, 106 (2018).
38. Templeton, A. R. & Read, B. Inbreeding, one word, several meanings, much confusion. *Biol. Conserv.* **75**, 91–105 (1996).
39. McQuillan, R. et al. Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **83**, 359–372 (2008).
40. Alvarez, G., Ceballos, F. C. & Quintero, C. The role of inbreeding in the extinction of a European royal dynasty. *PLoS ONE* **4**, e5174 (2009).
41. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
42. Amin, N., van Duijn, C. M. & Aulchenko, Y. S. A genomic background based method for association analysis in related individuals. *PLoS One* **2**, e1274 (2007).
43. Karsen, L. C., van Duijn, C. M. & Aulchenko, Y. S. The GenABEL Project for statistical genomics. *F1000Research* 5, 914 (2016).

**Acknowledgments**

F.C.C. received a postdoctoral fellowship from the South African National Research Foundation. The AWI-Gen Collaborative Centre is funded by the National Human Genome Research Institute (NHGRI), Office of the Director (OD), Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD), the National Institute of Environmental Health Sciences (NIEHS), the Office of AIDS research (OAR) and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), of the National Institutes of Health (NIH) under award number U54HG06938 and its supplements, as part of the H3Africa Consortium. Additional funding came from the Department of Science and Technology, South Africa, award number DST/CON 0056/2014. M.R. is 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 National Research Foundation. J.W. acknowledges support from the UK MRC Human Genetics Unit quinquennial programme grant (MC_UU_00007/10).

**Author contributions**

F.C.C.: research plan development, data analysis, figures, and manuscript preparation; S.H.: advisor on genomic analysis, data analysis, and manuscript preparation; D.C.: data analysis. GoAg, GeAs, P.B., X.G.O., F.M., and S.N.: AWI-Gen dataset collection and preparation; M.R.: principal advisor for research plan development, data analysis, and manuscript preparation; J.W.: advisor for research plan development, data analysis, and manuscript preparation. All authors read and approved the final manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-19595-y.

**Correspondence and requests for materials should be addressed to M.R.**

**Peer review information Nature Communications thanks Daniel Howrigan and other, anonymous, reviewers for their contributions to the peer review of this works. Peer review reports are available.**

**Reprints and permission information is available at http://www.nature.com/reprints**

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

---

© The Author(s) 2020