nature medicine

BRIEF COMMUNICATION

https://doi.org/10.1038/s41591-019-0459-6

CCR5-Δ32 is deleterious in the homozygous state in humans

Xinzhu Wei[1] and Rasmus Nielsen[1,2]

We use the genotyping and death register information of 409,693 individuals of British ancestry to investigate fitness effects of the *CCR5*-Δ32 mutation. We estimate a 21% increase in the all-cause mortality rate in individuals who are homozygous for the Δ32 allele. A deleterious effect of the Δ32/Δ32 mutation is also independently supported by a significant deviation from the Hardy–Weinberg equilibrium (HWE) due to a deficiency of Δ32/Δ32 individuals at the time of recruitment.

In late 2018, a scientist from the Southern University of Science and Technology in Shenzhen, Jiankui He, announced the birth of two babies whose genomes were edited using CRISPR[1]. No presentation of the experiment has appeared in the scientific literature, however online information[2] describes an introduction of mutations in the *CCR5* gene with the aim of mimicking the effect of the *CCR5*-Δ32 mutation, which provides protection against HIV in European individuals[3]. Although the mutations were not identical to *CCR5*-Δ32 (ref. [2]), and the consequences of the mutations are unknown, the stated purpose was nevertheless the prevention of HIV. The CRISPR experiment raises a number of obvious ethical issues. In addition, it is not clear whether the Δ32 mutation is beneficial. A mutation can be advantageous or disadvantageous depending on environmental conditions[4] and developmental stages[5]. In fact, despite the protection that Δ32 provides against HIV, and possibly other pathogens such as smallpox[6] and flavivirus[7], and although it facilitates recovery after stroke[8], it also appears to reduce protection against certain other infectious diseases such as influenza[9].

Direct fitness effects of individual segregating mutations are expected to be small, and are therefore very hard to measure directly. However, owing to the recent availability of large databases of genomic data, direct studies of fitness effects of individual mutations have now become feasible[10]. We might expect that the Δ32 mutation is deleterious in the homozygous state based on previous reports in smaller data sets, which show that individuals with the Δ32/Δ32 genotype have increased mortality when infected by influenza[9] and are four times more likely to develop certain infectious diseases[11]. Here we investigate this hypothesis using the genotyping and death register information of 409,693 individuals of British ancestry in the UK Biobank[12]. Δ32 has a frequency of 0.1159 in the British population and the UK Biobank contains data from thousands of individuals who are homozygous for the Δ32 allele, providing an opportunity to compare the mortality of these individuals to that of Δ32/+ and +/+ individuals.

We calculate the survival rate (1 − death rate) per year for each of the three Δ32 genotypes, from age 41 to age 78 (see Methods), which is the entire range allowed by the data available (Fig. 1a). Owing to the small sample size at ages 77 and 78, we primarily report the survival probability before age 76 (see Methods). The death rate from age 70 to 74 in the UK Biobank volunteers is 46–56% lower than that in the general UK population of the same age[13], probably owing to an ascertainment bias known as the 'healthy volunteer effect'[14]. Nevertheless, the relative death rates among different genotypes can still be compared to provide information about the fitness effects of specific mutations. The uncorrected survival probabilities to age 76 of individuals enrolled in the study is 0.8351 for Δ32/Δ32, 0.8654 for Δ32/+, and 0.8638 for +/+ (Fig. 1a), which implies that Δ32/Δ32 has an approximately 21% higher aggregated death rate before age 76 than the other genotypes. The average age of enrollment is 56.5 years, so the data largely reflect differences in mortality in individuals above this age. We can partially correct for the death registration delay and biased ascertainment using the general population's death rate per year. After correction, the individuals with the Δ32/Δ32 genotype are approximately 20% less likely to reach age 76 than individuals with the other genotypes (see Methods). To test the significance of the nominally lower survival rate of Δ32/Δ32, we first perform a log-rank test comparing the death rate of Δ32/Δ32 individuals to that of the other two genotypes (Z score = 2.37, one-tailed $P = 0.0089$). We also bootstrap the sample 1,000 times and find that Δ32/Δ32 individuals have a significantly higher death rate than the other two genotypes, whereas Δ32/+ and +/+ individuals have similar death rates (Supplementary Table 1). The increase in mortality of Δ32/Δ32 individuals is the highest at age 74, at which point it is 26.4% higher than the mortality of +/+ individuals (95% bootstrap confidence interval (3.0%,49.5%)). Similarly, a Cox model[15] for left truncated and right censored data also suggests that Δ32/Δ32 individuals have an average 21.4% elevated death rate across all ages (95% confidence interval 3.4% and 42.6%, one-tailed $P = 0.0089$). The fifth principal component is associated with Irish ancestry[12] and is also associated with a difference in mortality (two-sided $P = 2.5 \times 10^{-16}$) in the Cox model. However, when correcting for this effect using prinicipal component analysis (PCA) loadings as covariates, the increase in mortality of Δ32 is maintained (see Supplementary information). We note that despite the nominally large detected effect on survivorship, the P value of 0.0089 is only moderately small, owing to the low frequency of Δ32/Δ32 individuals and the generally low mortality in the cohort. The accuracy of the estimates will probably improve in future years as the mortality rate of the cohort increases.

Selection against homozygous individuals will lead to deviations from the HWE, which can be measured by the inbreeding coefficient (F). Deviations from the HWE at the time of enrollment, which is the time at which samples are obtained for genotyping, provides an assessment of the differential fitness of Δ32 genotypes that is independent from the previous analyses using death registry information obtained after enrollment. We test for deviations from the HWE consistent with a deleterious effect of Δ32 in homozygous individuals by calculating the allele-specific inbreeding coefficient

[1]Department of Integrative Biology and Statistics, University of California, Berkeley, Berkeley, CA, USA. [2]Centre for GeoGenetics, University of Copenhagen, Copenhagen, Denmark. e-mail: aprilwei@berkeley.edu; rasmus_nielsen@berkeley.edu

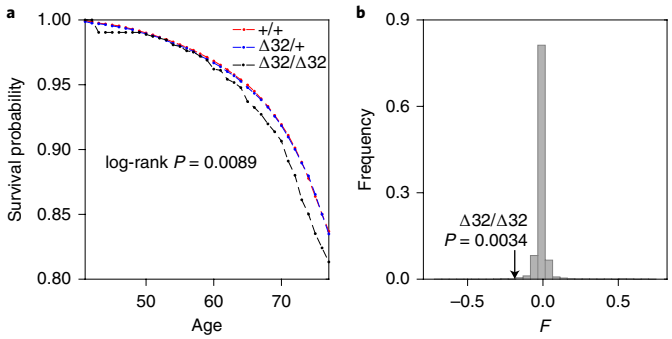


Fig. 1 | *CCR5*-Δ32 is deleterious in the homozygous state. a, Survival probabilities of the three Δ32 genotypes (+/+, Δ32/+ and Δ32/Δ32). The one-tailed *P* values from the log-rank tests up to age 76 are shown. The number of samples for which age information and genotype at Δ32 are both available is 395,704. **b**, The histogram of inbreeding coefficients, *F*, from 5,932 SNPs whose allele frequencies closely resemble that of Δ32. The black arrow points to the observed *F* of Δ32 ($F_{\Delta32/\Delta32} = -0.19$), calculated for the Δ32/Δ32 individuals. The sample size used in estimating *F* for each of the 5,932 SNPs varies from 7,896 to 409,607 with a mean of 405,428, and the sample size for Δ32 is 395,714.

$F_{\Delta32/\Delta32}$. However, there might be deviations from HWE in the data for multiple other reasons, including inbreeding and population structure. Therefore, we compare $F_{\Delta32/\Delta32}$ (see Methods) with the locus-specific value of *F* for other variants in the data with minor allele frequencies similar (± 0.0025) to that of Δ32. Only 20 out of 5,932 variants have a smaller *F* than $F_{\Delta32/\Delta32}$ (Fig. 1b; empirical one-tailed $P = 0.0034$). In addition, the deviation from the HWE for each age group also correlates with the deviation predicted by the survival probability (Spearman's $\rho = 0.67$, $P = 1.4 \times 10^{-4}$; see Supplementary information and Extended Data Fig. 1). These two independent analyses are largely consistent with each other and both indicate a substantial increase in mortality associated with the Δ32/Δ32 genotype.

Our results show that being homozygous for the Δ32 mutation is associated with reduced life expectancy in a modern cohort, despite the protective effect of the mutation against HIV[3]. This finding echoes the previous reports that Δ32 reduces resistance against influenza[9] and other infectious diseases[11]. We did not observe any difference in mortality between Δ32/+ and +/+ individuals (Supplementary Table 1), despite the fact that Δ32/+ also provides protection against HIV[3]. It could reflect the healthy volunteer effect in the UK Biobank cohort[13] if individuals affected by HIV, or suffering from higher mortality due to HIV infection, are less likely to be recruited. In that case, our estimates of death rates reflect individuals that have reduced exposure to HIV, and the conclusion regarding increased mortality of Δ32/Δ32 is then with reference to such individuals. If so, it would also imply that Δ32 is overdominant in the presence of HIV; that is, that individuals heterozygous for the mutation have the highest fitness. In the absence of HIV or other infectious agents for which the mutation provides protection, the mutation will be under negative directional selection. However, because only approximately 0.16% of the current British population is infected by HIV[16], the benefit from this protection is probably too small to have a detectable influence on survival probability in our study.

It is unclear exactly which factors are most important for the fitness effects of the Δ32 mutation. There are many phenotypic associations that are significant at 5% significance level after correction for multiple testing in the UK Biobank (see Supplementary information for the phenotypes), and the mutation is probably highly pleiotropic. Out of the 5,932 single-nucleotide polymorphisms (SNPs) with matching allele frequencies, only 76 have more phenotypic associations than Δ32 in terms of the UK Biobank phenotypes (empirical one-tail $P = 0.0128$, see Supplementary information).

It is perhaps not unexpected that homozygosity for a deletion in a functional gene is associated with reduced fitness. It underscores the idea that introduction of new or derived mutations in humans using CRISPR technology, or other methods for genetic engineering, comes with considerable risk even if the mutations provide a perceived advantage. In this case, the cost of resistance to HIV may be increased susceptibility to other, and perhaps more common, diseases.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41591-019-0459-6.

Received: 15 January 2019; Accepted: 16 April 2019;
Published online: 03 June 2019

References

1. Normile, D. *Science* **362**, 978–979 (2018).
2. Cyranoski, D. First CRISPR babies: six questions that remain *Nature* (30 November 2018).
3. Samson, M. et al. *Nature* **382**, 722 (1996).
4. Wei, X. & Zhang, J. *Genetics* **205**, 925–937 (2017).
5. Pavlicev, M. & Wagner, G. P. *Trends Ecol. Evol.* **27**, 316–322 (2012).
6. Galvani, A. P. & Slatkin, M. *Proc. Natl Acad. Sci.* **100**, 15276–15279 (2003).
7. Cahill, M. E., Conley, S., DeWan, A. T. & Montgomery, R. R. *BMC Infect. Dis.* **18**, 282 (2018).
8. Joy, M. T. et al. *Cell* **176**, 1143–1157 (2019).
9. Falcon, A. et al. *J. Gen. Virol.* **96**, 2074–2078 (2015).
10. Mostafavi, H. et al. *PLoS Biol.* **15**, e2002458 (2017).
11. Lim, J. K. & Murphy, P. M. *Exp. Cell Res.* **317**, 569–574 (2011).
12. Bycroft, C. et al. *Nature* **562**, 203 (2018).
13. Fry, A. et al. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
14. Delgado-Rodriguez, M. & Llorca, J. *J. Epidemiol. Community Health* **58**, 635–641 (2004).
15. Cox, D. R. & Oakes, D. *Analysis of Survival Data* (Chapman & Hall, 1984).
16. Nash, S et al. *Progress Towards Ending the HIV Epidemic in the United Kingdom: 2018 Report* (Public Health England, 2018).

Acknowledgements

The authors thank D. Feehan, M. Slatkin, and P. Wilton for discussions about death rate estimation, and R. Durbin, C. Freeman, and G. McVean for discussions about UK Biobank markers. This work is supported by US National Institutes of Health (NIH) grant R01GM116044 to R.N.

Author contributions

X.W. and R.N. designed the study and wrote the manuscript. X.W. analyzed the data.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-019-0459-6.

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-019-0459-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to X.W. or R.N.

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

Methods

The study population. This study uses the UK Biobank data under application number 33672 and basket IDs 10997 and 2000429. It complies with ethical regulations of the University of California (UC) Berkeley and the data are accessed under the Material Transfer Agreement between the UK Biobank and UC Berkeley.

In the UK Biobank, 409,693 volunteers have self-reported British ancestry confirmed by PCA[12], which constitutes roughly 0.62% of the entire British population. Our main analysis is performed on these volunteers, unless otherwise stated. There are 75,970 volunteers in the UK Biobank whose data are labeled as of non-British ancestry, which are used to investigate the effect of Δ32 in populations other than the British. The UK Biobank volunteers were recruited during 2006–2010 and 2.9% of the volunteers (13,831) have a recorded age at death (all cause).

Marker selection and validation. SNP rs62625034 (coordinate 3:46414975 in GRCh37) is a directly genotyped SNP that is used to identify Δ32 (rs333) based on the following validations. First, the Affymetrix probe used for this SNP is CCATACAGTCAGTATCAATTCTGGAAGAATTTCCA[G/T] ACATTAAAGATAGTCATCTTGGGGCTGGTCCTGCC, based on annotation files 'Axiom_UKBiLEVE.na34.annot.csv' and 'Axiom_UKB_WCSG.na34.annot.csv'. The targeted region of this probe fully includes the 32-bp deletion in rs333, given rs333 (Δ32) has coordinate 3: 46414947-46414978 in GRCh37. Second, rs62625034 is not called as a SNP in the 1000 Genome database and a recent study on variants in *CCR5* (ref. [17]) also confirmed that it could be detected only in one of the Denisovian samples. However, the detected allele frequency by the probe of rs62625034 in the UK Biobank is 0.1159 among the British ancestry genomes, which does not resemble the frequency of rs62625034 but closely resembles the frequency of rs333 (0.1237) in the European and the British population (CEU and GBR) in the 1000 Genomes data. Third, SNP rs113010081, a directly genotyped SNP in the UK Biobank data, is in strong linkage disequilibrium with rs333 in the 1000 Genomes data, with a r^2 of 0.93 combining CEU and GBR in 1000 Genomes data (https://ldlink.nci.nih.gov/?var1=rs333&var2=rs113010081&pop=CEU%2BGBR&tab=ldpair). We calculate the Pearson correlation between rs113010081 and the probe of rs62625034 using the UK Biboank British ancestry genotypes and obtain $r^2 = 0.94$, which again resembles the correct linkage disequilibrium between rs113010081 and rs333. In addition, there is no other SNP that is in as strong a linkage disequilibrium with rs113010081 in the targeted region of this probe (https://ldlink.nci.nih.gov/?var=rs113010081&pop=CEU%2BGBR&r2_d=r2&tab=ldproxy). Last, we also estimate the survival probability for rs113010081, and the results are similar to that obtained for rs62625034 (not shown).

Estimation of survival probability. The UK Biobank death records are updated quarterly with the UK National Health Service (NHS) Information Centre for participants from England and Wales, and by NHS Central Register, Scotland for participants from Scotland. However, the death records are not made available immediately to researchers. The latest date of death among all registered deaths in the downloaded data is 16 February 2016, and we use this date to approximate the time of last death entry, and assume that after this date we have no mortality or viability information for the volunteers. We use five entries from the UK Biobank data—the age at recruitment, the date of recruitment, the year of birth, month of birth, and the age at death—to calculate the number of individuals (N_i) who are ascertained from age i to age $i+1$, and the occurrence of death observed from these N_i individuals during the interval of age i to age $i+1$ is O_i. Using this information, we calculate the ascertained age for each individual. We ignore the partially ascertained age to avoid biases from censoring. For example, an individual recruited at age 45.2, and reaching age 52.3 on 16 February 2016, who does not have a reported death in our data, is treated as being observed from age 46 to age 52, thus this volunteer contributes to N_{46}, N_{47}, N_{48}, N_{49}, N_{50}, N_{51}. As another example, a person who is recruited at age 65.7, and who could have reached age 72.6 by 16 February 2016 but has a reported death at age 69.7 will contribute to N_{66}, N_{67}, N_{68}, N_{69}, and this volunteer will also contribute to O_{69}. This volunteer does not contribute to N_{70}, because death has already occurred before age 70. The death rate per year is then calculated as $h_i = O_i/N_i$, and the probability of surviving to age $i+1$ is $S_i = \prod_{n=1}^{n=i} h_n$. The UK Biobank data allow estimation of death rates from h_{41} to h_{77}, but because N_{77} is smaller than 800, we have to assume that $h_{76} = h_{77}$ and combined these two ages in our estimation. We estimate h_i separately for the three different Δ32 genotypes. We mainly report the survival probability before age 76, as there is sufficient data to obtain accurate estimates, but the estimated survival probabilities to age 77 and 78 are also shown in Fig. 1.

As the exact birth dates of the volunteers are considered sensitive, we do not have access to these. The age at recruitment in the UK Biobank is rounded down to nearest integer age, and we approximate the exact age using the date of recruitment, the year of birth, and month of birth, assuming that everyone is born on the 15th of their birth month. In rare cases, when the date of recruitment is very close to a person's birthday, the approximated age could be smaller than the age at recruitment provided by the UK Biobank and in these rare cases we instead round up the estimated age. After applying this rounding scheme, if there are no errors in the data, under no scenario should the estimated age be smaller than the integer age at recruitment. However, there are 17 individuals whose estimated age is smaller than the age at recruitment, and we exclude these individuals in the death rate calculation. Among them, 15 are of British ancestry.

Although the UK Biobank routinely imports death records from the national databases, the healthy volunteer effect[13] can still lead to a substantial underestimation of the death rate per year h_i compared to the general population. The delay of the death records may be affected by many factors, including time of recruitment, age of death, cause of death, and various socioeconomic factors[18]. However, if we assume that these biases are independent of the Δ32 genotype, we can then estimate the death rate correction factor C_i for each age i, and estimate the death rate per year and the survival probability for the three different Δ32 genotypes in the general population. To do this, we download the national life tables in the UK (nltuk1517reg.xls) from the Office of National Statistics (https://www.ons.gov.uk), which contain the death rate per year for the entire British population each year from 1980 to 2017, estimated for males and females separately. We average the death rate per year from 2006 to 2016 to represent the death rate H_i of the general population. We then use h_i/H_i to estimate C_i. We then calculate a corrected death rate for each Δ32 genotype. For example, the corrected death rate for +/+ is $h_{i,+/+}/C_i$. We use the corrected death rates to estimate the corrected survival probability (S_C). The inferred survival probability after correction (S_C) to age 76 are 0.7565, 0.7589, and 0.7111 for genotypes +/+, Δ32/+, and Δ32/Δ32, respectively. With this crude correction, the probability of death before age 76 in the general population is $(1 - S_{C,\Delta32/\Delta32})/(1 - S_{C,\Delta32/+}) - 1$, approximately 20% higher for Δ32/Δ32 individuals than for heterozygous individuals. We note that although the calculations of death rates could be more accurate, for example by using exact birthdays (which we did not have access to), the significant difference in death rates between genotypes is unlikely to be explained by this effect. However, our survival analyses may underestimate the beneficial effects of Δ32 in some age groups owing to ascertainment biases caused by the healthy volunteer effect[13].

Estimation of *F*. $F_{\Delta32/\Delta32}$ is estimated from the equation $P_{\Delta32/\Delta32} = (1 + F_{\Delta32/\Delta32}) P_{\Delta32} P_{\Delta32}$, where $P_{\Delta32}$ and $P_{\Delta32/\Delta32}$ are the observed frequencies of Δ32 and Δ32/Δ32, respectively. When $F_{\Delta32/\Delta32}$ is significantly lower than 0, it implies that the observed fraction of Δ32/Δ32 individuals is lower than expected under the HWE, consistent with increased mortality of Δ32/Δ32 individuals. The *F* of other SNPs are estimated similarly.

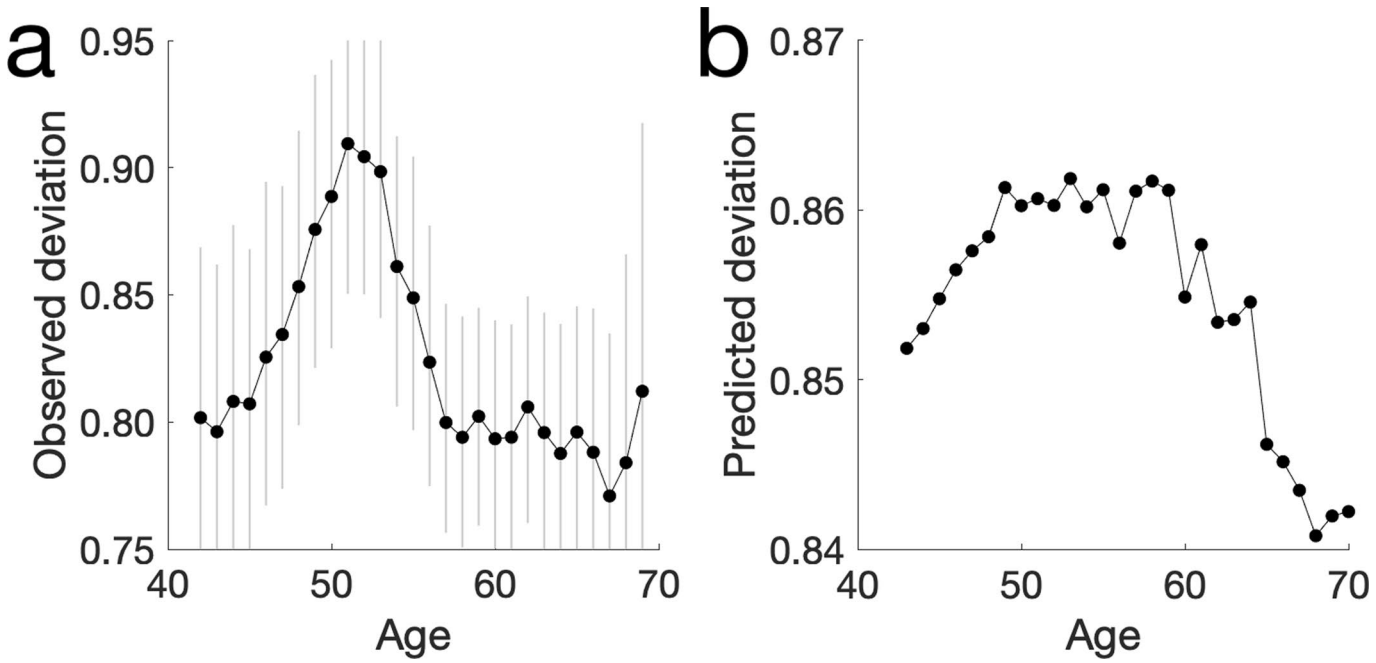
Statistical analysis. One-tailed *P* values from log-rank tests are used in Fig. 1a and Supplementary Table 1. In Fig. 1b, the empirical one-tailed *P* value from the *F* of 5,932 SNPs is used. Bootstrap 95% confidence intervals are shown as error bars in Extended Data Fig. 1a, and are used in Supplementary Table 1. Spearman's correlation is used in Extended Data Fig. 1. In addition, the details of the statistical tests are given where they are mentioned.

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

Data, code, and research notebook availability. The genotype and death registry information are available with the permission of the UK Biobank. Analytical results and scripts are accessible from https://github.com/AprilWei001/CCR5-delta32. In addition, a detailed experimental notebook covering the entire development of this project is available at the following depository: https://xinzhuaprilwei.weebly.com/download/ccr5-delta32.

References

17. Hoover, K. C. *PloS ONE* **13**, e0204989 (2018).
18. Patel, V. *Impact of Registration Delays on Mortality Statistics: 2016* (Office for National Statistics, 2016); https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/impactofregistrationdelaysonmortalitystatistics2016



Extended Data Fig. 1 | The deviation from HWE with age. **a**, The observed deviation using age at recruitment estimated. Each dot represents one age group. The grey error bars show the 95% confidence intervals estimated from bootstrap the genotypes of individuals recruited at each age 1000 times. The sample size used for each error bar ranges from 15191 to 100117 with a mean of 65479. **b**, The predicted deviation from HWE using the corrected survival probability. A total of 395704 samples are used. The observed and predicted values are significantly correlated (Spearman's correlation coefficient $\rho = 0.67$, $P = 1.4 \times 10^{-4}$).

nature research

Corresponding author(s): Xinzhu Wei, Rasmus Nielsen

Last updated by author(s): Apr 10, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The genotype and death registry information are available through the UK Biobank application 33672. Software used: Plink2
Data analysis	Analytical results and scripts are available on https://github.com/AprilWei001/CCR5-delta32

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The genotype and death registry information are available with the permission of the UK Biobank. Analytical results and scripts are accessible through (https://github.com/AprilWei001/CCR5-delta32). In addition, a detailed experimental notebook covering the entire development of this project is available at depository (https://xinzhuaprilwei.weebly.com/download/ccr5-delta32).

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	409,693. This is the number of British ancestry volunteer that are genotyped in the UK Biobank. The sample size is sufficient because the delta32 has a relatively high MAF (0.1159).
Data exclusions	Exclude non-British ancestry volunteers from the UK Biobank to control for the genetic background and for the purpose of calculating Hardy-Weinberg proportion. Exclude samples whose estimated age from year/month of birth do not agree with self-reported age.
Replication	Each reported result is confirmed by several statistical approaches, in addition to two lines of independent evidences confirming each other.
Randomization	Randomization were not employed.
Blinding	Yes, data is collected and de-identified by UK Biobank.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	British ancestry. Age 40-69 at recruitment. Genotype information from blood-derived DNA.
Recruitment	The UK Biobank recruited volunteers by sending out invitation letters to homes of people aged 40-69. Volunteers then signed up at assessment centers. There can be "healthy volunteer effect" such that people who volunteer are likely healthier than the general population.
Ethics oversight	UC Berkeley, UK Biobank

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