Clonal hematopoiesis and risk of prostate cancer in large samples of European ancestry men

Anqi Wang[1,20], Yili Xu[1,20], Yao Yu[2], Kevin T. Nead[2], TaeBeom Kim[2], Keren Xu[1], Tokhir Dadaev[3], Ed Saunders[3], Xin Sheng[1], Peggy Wan[1], Loreall Pooler[1], Lucy Y. Xia[1], Stephen Chanock[4], Sonja I. Berndt[4], Susan M. Gapstur[5], Victoria Stevens[5], Demetrius Albanes[4], Stephanie J. Weinstein[4], Vincent Gnanapragasam[6], Graham G. Giles[7-9], Tu Nguyen-Dumont[8,10], Roger L. Milne[7-9], Mark M. Pomerantz[11], Julie A. Schmidt[12,13], Konrad H. Stopsack[14], Lorelei A. Mucci[14], William J. Catalona[15], Kurt N. Hetrick[16], Kimberly F. Doheny[16], Robert J. MacInnis[7,9], Melissa C. Southey[7,8,10], Rosalind A. Eeles[4,17], Fredrik Wiklund[18], Zsofia Kote-Jarai[3], Adam J. de Smith[1], David V. Conti[1], Chad Huff[2], Christopher A. Haiman[1], Burcu F. Darst[1,19*]

[1] Center for Genetic Epidemiology, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California, US

[2] Department of Epidemiology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, US

[3] The Institute of Cancer Research, London, UK

[4] National Cancer Institute, National Institutes of Health, Bethesda, Maryland, US

[5] American Cancer Society, Atlanta, Georgia, US

[6] Division of Urology, Department of Surgery, University of Cambridge, Cambridge, UK

[7] Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, AU

[8] Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Melbourne, Victoria, AU

[9] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Victoria, AU

[10] Department of Clinical Pathology, The University of Melbourne, Victoria, AU

[11] Dana-Farber Cancer Institute, Boston, Massachusetts, US

[12] Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

[13] Department of Clinical Epidemiology, Department of Clinical Medicine, Aarhus University Hospital and Aarhus University, Aarhus N, Denmark

[14] Harvard T.H. Chan School of Public Health, Boston, Massachusetts, US

[15] Northwestern University Feinberg School of Medicine, Chicago, Illinois, US

[16] Center for Inherited Disease Research, Department of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD, US

[17] The Royal Marsden NHS Foundation Trust, London, UK

[18] Karolinska Institute, Solna, SE

[19] Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, US

[20] These authors contributed equally to this research

*Corresponding Author

Burcu F. Darst

1100 Fairview Ave N

Seattle, WA 98109 USA

1-206-667-1036

bdarst@fredhutch.org

Abstract

Little is known regarding the potential relationship between clonal hematopoiesis (CH) of indeterminate potential (CHIP), which is the expansion of hematopoietic stem cells with somatic mutations, and risk of prostate cancer, the fifth leading cause of cancer death of men worldwide. We evaluated the association of age-related CHIP with overall and aggressive prostate cancer risk in two large whole-exome sequencing studies of 75,047 European ancestry men, including 7,663 prostate cancer cases, 2,770 of which had aggressive disease, and 3,266 men carrying CHIP variants. We found that CHIP, defined by over 50 CHIP genes individually and in aggregate, was not significantly associated with overall (aggregate HR=0.93, 95% CI=0.76-1.13, P=0.46) or aggressive (aggregate OR=1.14, 95% CI=0.92-1.41, P=0.22) prostate cancer risk. CHIP was weakly associated with genetic risk of overall prostate cancer, measured using a polygenic risk score (OR=1.05 per unit increase, 95% CI=1.01-1.10, P=0.01). CHIP was not significantly associated with carrying pathogenic/likely pathogenic/deleterious variants in DNA repair genes, which have previously been found to be associated with aggressive prostate cancer. While findings from this study suggest that CHIP is likely not a risk factor for prostate cancer, it will be important to investigate other types of CH in association with prostate cancer risk.

Introduction

Age-related clonal hematopoiesis (CH), also referred to as clonal hematopoiesis of indeterminate potential (CHIP) in the absence of a hematologic malignancy, is the expansion of hematopoietic stem cells with somatic mutations and is increasingly common with older age. In addition to CH being a risk factor for myeloid malignancy development, it has also been associated with increased risk of all-cause mortality and cardiovascular disease(1-3). Individuals with solid tumors have been reported to be more likely to have clonal mosaicism than cancer-free participants(4). Age-related loss of chromosome Y (LOY) in circulating leukocytes has also been associated with increased risk of non-hematological cancer mortality(5). Further, in a two-sample Mendelian randomization analysis, genetically predicted LOY was reported to be associated with increased genetic risk of prostate cancer and other solid tumors(6). However, little is known regarding the potential impact of age-related CH on risk of prostate cancer, the fifth leading cause of cancer death of men worldwide(7). Pathogenic germline variants in many genes that are associated with CH, particularly DNA repair genes *ATM*, *CHEK2*, and *NBN*, are also associated with prostate and other non-hematologic cancers(8-11), suggesting a potential mechanistic link between these two conditions. In this investigation, we evaluated the association of age-related CHIP with overall and aggressive prostate cancer risk in two large whole-exome sequencing studies of European ancestry men.

Results

Whole-exome sequence data was analyzed for 2,118 incident prostate cancer cases and 67,384 controls from the UK Biobank(12) and 2,770 aggressive and 2,775 non-aggressive prostate cancer cases from a cross-sectional case-only study from 12 international study sites,

referred to here as the Whole-Exome Sequencing Study in Prostate Cancer (WESP)(9). All men were of European ancestry.

Potential CHIP variants were called with Mutect2(13) and defined based on previous curations of variants within 74 established hematologic cancer genes(1, 14, 15). Specifically, CHIP variants in these genes were rare with minor allele frequencies (MAF) <0.1% in the study population, excluded variants with MAF >0.1% in the Genome Aggregation Database (gnomAD)(16) to reduce the potential of capturing germline variants, and were either deleterious (protein truncating or splice altering)(17) or specifically reported in Jaiswal et al.(1). Variant allelic fractions (VAFs) were >5% in the UK Biobank and >10% in WESP due to differences in exome sequencing coverage, thresholds previously suggested to increase the likelihood of a mutation being somatic(14, 18-20). VAFs were calculated with bcftools (fill-tags FORMAT/VAF) as the fraction of reads with the alternate allele(21). Several VAF thresholds were tested in sensitivity analyses and led to similar results and the same conclusions (see **Materials and Methods** for details). A total of 1,778 qualifying CHIP variants in 55 genes were identified in the UK Biobank, while 360 qualifying CHIP variants in 52 genes were identified in WESP (VAFs are described in **Table 1** and **Supplemental Tables 1-2**; **Materials and Methods**). Overall, 2,874 (4.1%) men in the UK Biobank and 392 (7.1%) men in WESP were found to carry a CHIP variant. The most commonly carried CHIP variants were in *DNMT3A* (33.6% of the 2,874 CHIP carriers in UK Biobank and 26.0% of the 392 CHIP carriers in WESP), *TET2* (25.2% of UK Biobank carriers and 22.7% of WESP carriers), and *ASXL1* (9.1% of UK Biobank carriers and 8.9% of WESP carriers; **Supplemental Figure 1**), consistent with previous studies(1, 22, 23). Given the overall CHIP carrier frequencies and study sample sizes,

the UK Biobank and WESP both had 80% power to detect an OR of 1.21 for the association of CHIP with overall and aggressive prostate cancer risk.

In the UK Biobank, the median age at blood draw was 57 years (interquartile range [IQR]=13) and the median time between blood draw and cancer diagnosis among cases was 4.1 years (IQR=3.6; **Supplemental Figure 2**). As expected, in the UK Biobank, CHIP status was significantly associated with age at blood draw, with the strongest associations observed with *DNMT3A* (+4.1 years, 95% CI=3.6-4.6, $P=4.4x10^{-55}$), *ASXL1* (+6.2 years, 95% CI=5.2-7.2, $P=1.3x10^{-35}$), *TET2* (+3.7 years, 95% CI=3.2-4.3, $P=2.6x10^{-35}$), *PPM1D* (+5.8 years, 95% CI=4.1-7.5, $P=3.8x10^{-11}$), and *SF3B1* (+8.8 years, 95% CI=5.5-12.0, $P=1.1x10^{-07}$) and across all 55 genes in aggregate (+3.2 years, 95% CI=2.9-3.5, $P=2.4x10^{-99}$; **Table 2**, **Figure 1**, and **Supplemental Figure 2-3**). However, no significant associations were observed between CHIP carrier status and age at prostate cancer diagnosis, with or without adjustment for age at blood draw, in the UK Biobank (**Table 2** and **Supplemental Figure 3**).

In the UK Biobank, we did not observe a significant difference in CHIP carrier frequencies between prostate cancer cases and controls (4.86% and 4.11%, respectively; HR=0.93, 95% CI=0.76-1.13, P=0.46; **Table 1**; see **Materials and Methods** for details). In WESP, CHIP carrier frequencies did not significantly differ when comparing cases with aggressive prostate cancer (7.15%; OR=1.14, 95% CI=0.92-1.41, P=0.22), prostate cancer death (6.38%; OR=1.02, 95% CI=0.81-1.30, P=0.84), or metastatic prostate cancer (7.07%; OR=1.22, 95% CI=0.81-1.83, P=0.34) to non-aggressive prostate cancer cases (6.99%; **Table 1**). Similarly, in gene-based tests, no significant associations were observed between CHIP carrier status and overall prostate cancer risk in the UK Biobank or with disease aggressiveness in WESP (**Table 1** and **Supplemental Figure 4**). Carrier frequencies for the aggregate of CHIP genes *PTEN*, *TP53*,

HLA-A, and *MAP2K1*, which have prior evidence of association with prostate cancer risk(11, 24-26), were not significantly associated with overall prostate cancer risk in the UK Biobank (HR=0.93, 95% CI=0.30-2.89, P=0.90) or with disease aggressiveness in WESP (OR=1.50, 95% CI=0.39-5.82, P=0.56).

We found weak evidence of association between genetic susceptibility to prostate cancer, measured by a polygenic risk score (PRS) constructed based on a previous publication(11), and CHIP carrier status across all CHIP genes in prostate cancer controls from the UK Biobank (OR=1.05 for each additional risk allele, 95% CI=1.01-1.10, P=0.01; **Materials and Methods**). CHIP carrier status for *DNMT3A* had the strongest association with the PRS in UK Biobank controls (OR=1.11, 95% CI=1.03-1.19, $P=5.6x10^{-3}$); however, this association was not significant after adjusting for multiple testing of all individual CHIP genes. We also observed a null association between carrier status for pathogenic/likely pathogenic/deleterious variants across 24 previously curated prostate cancer candidate DNA repair genes(9) and age-related CHIP carrier status across all CHIP genes in WESP (OR=0.99, 95% CI=0.72-1.36, P=0.96) and in the UK Biobank (OR=1.00, 95% CI=0.84-1.20, P=0.96; see **Materials and Methods** for details). DNA repair gene carrier status was also not significantly associated with age-related CHIP carrier status for individual CHIP genes in WESP or the UK Biobank, with the exception of *KDM6A* in the UK Biobank (OR=3.33, 95% CI=1.70-6.52, $P=4.4x10^{-4}$; all other P-values≥0.002). Likewise, we did not observe a significant association between the aggregate of DNA repair genes *BRCA2*, *ATM*, and *PALB2*, which we previously reported to be associated with aggressive prostate cancer(9), and age-related CHIP carrier status across all CHIP genes in WESP (OR=0.74, 95% CI=0.30-1.85, P=0.53) or the UK Biobank (OR=1.14, 95% CI=0.76-1.73, P=0.52).

Discussion

In two large European ancestry datasets, we found minimal evidence of an association between age-related CHIP and risk of overall or aggressive prostate cancer. These findings are supported by a previous investigation reporting that clonal mosaicism was associated with increased risk of non-hematological cancers, but not with prostate cancer(4).

While this study was based on two large datasets, if CHIP has a weak association with prostate cancer risk, a larger number of CHIP carrier prostate cancer cases would be needed to detect such an association. Further, our investigation was limited to men of European ancestry, and it is possible that the role of CHIP in prostate cancer risk could vary in non-European ancestry populations. For example, a germline sequencing study suggested that rare deleterious variants in *TET2* were in aggregate associated with prostate cancer risk in men of African ancestry(27). It is also possible that our approach to identifying CHIP variants may have introduced some non-differential exposure misclassification, although sensitivity analyses testing various VAF thresholds suggest that our findings are robust. In particular, based on previous estimates, the depth of our sequencing coverage in WESP provided ~50% sensitivity to detect CHIP variants in the VAF range of 5-10% and would need to be ~100x to capture variants in this range with closer to 100% sensitivity(14). Observed differences in CHIP carrier frequencies between the UK Biobank and WESP could be in part due to differences in age distributions and the PoN panels used to call CHIP variants.

Although our findings do not support an association between age-related CHIP and prostate cancer, a previous study found that therapy-related CH was associated with decreased survival in non-hematologic solid tumor cancer patients(22). As such, it may be relevant to

investigate the impact of CHIP on survival in post-therapy prostate cancer patients. Future studies of other types of CH may also provide important insights into prostate cancer risk, such as LOY, which is present in over 40% of men at age 70, is highly heritable, and has been previously associated with increased genetic risk of prostate cancer(6).

Materials and Methods

Participants and Genetic Sequencing

To investigate the association between CHIP variants and risk of overall prostate cancer, we analyzed 69,502 European ancestry men from the UK Biobank with whole exome-sequencing (WES) data, which was generated at the Regeneron Genetics Center(28, 29). On average, 95.6% of the targeted regions were sequenced with at least 20x coverage(28). Prostate cancer cases were identified through linkage to the NHS Central Register for the first diagnosis of prostate cancer. Quality control criteria applied to this sample of men from the UK Biobank have been previously described(12).

To investigate the association between CHIP variants and risk of aggressive prostate cancer, 5,545 European ancestry men were included from WESP, consisting of 12 large European and US studies(9). WES was performed at the Center for Inherited Diseases Research with average targeted exon coverage of 56x and 95.7% of targeted regions sequenced with at least 10x coverage. Aggressive prostate cancer was defined as men who either died due to prostate cancer, had metastatic disease, had stage T4 disease, or had stage T3 disease with a Gleason score ≥8 tumor. Non-aggressive prostate cancer was defined as men who had stage T1/T2 disease and a Gleason score ≤6 tumor, with 71.3% also having ≥10 years of follow-up to

indicate that they were alive and without recurrence. Details of the study design, sequencing procedures, and quality control were previously described(9). Informed consents were obtained from all participants, and study protocols were approved by respective institutional review boards.

Identification of Clonal Hematopoiesis Variants

Somatic CHIP variants were identified based on a list of 74 genes with previously reported mutations in human hematologic cancers(1, 14, 15). Somatic short variants (SNVs and Indels) were identified using the GATK toolkit following the GATK's best practices workflows. In short, somatic variant calling was carried out using the GATK Mutect2(13) in tumor-only mode. A panel of normals (PoN) was incorporated to filter out commonly seen sequencing artefacts. A subset of 100 randomly selected UK Biobank individuals under 40 years of age were used as the PoN for UK Biobank, while WESP used the PoN provided in the GATK resource bundle consisting of several hundred normals. Population allele frequencies of common and rare variants from gnomAD were provided in the GATK resource bundle as an external reference of germline variants and were utilized to filter out possible germline variants. Somatic short variants were further filtered through the GATK FilterMutectCalls. Variants included among these genes were those with deleterious (protein truncating or splice altering) functional consequences(17) or those specifically reported in Jaiswal et al.(1) with minor allele frequencies (MAF) <0.1% (in the respective UK Biobank or WESP data) and a variant allelic fraction (VAF) >5% in the UK Biobank and >10% in WESP given differences in sequence coverage, as previously suggested to increase the likelihood of a mutation being somatic(14, 18, 19). This lower bound VAF in WESP was selected given the coverage of our exome sequencing, as 30-

50x coverage has been reported to be able to robustly call CHIP variants with VAF >10%(14). Based on previous literature(19, 20), we also conducted sensitivity analyses using VAFs between 10-40%, 10-60%, and 10-40% or 60-90%. These sensitivity analyses all led to similar results and the same conclusions as our primary analysis. VAFs were calculated with bcftools (fill-tags FORMAT/VAF) as the fraction of reads with the alternate allele(21). We excluded variants with MAF >0.1% in the Genome Aggregation Database (gnomAD)(16) and those in simple tandem repeat regions(30, 31). This led to a total of 1,778 variants in 55 CHIP genes identified in the UK Biobank and 360 variants in 52 CHIP genes identified in WESP (**Supplemental Tables 1-2**).

Identification of DNA Repair Gene Variants

A total of 24 previously curated prostate cancer candidate DNA repair genes(9) were identified, as DNA repair genes have been shown to predispose to prostate cancer(9, 10, 32-34) and clonal hematopoiesis(8). Within these genes, pathogenic/likely pathogenic/deleterious (P/LP/D) variants were considered and defined as rare variants (MAF<0.01) in the study population that had either a Variant Effect Predictor (VEP) Impact score of "high"(17) and/or a ClinVar classification of Pathogenic or Likely Pathogenic(35). We excluded the known low/moderate prostate cancer risk variant c.9976A>T (rs11571833) in *BRCA2*(36).

Polygenic Risk Score Construction

In the UK Biobank, where GWAS data was available, a PRS was constructed based on a multi-ancestry prostate cancer genome-wide association study (GWAS) meta-analysis of >230,000 men, where we developed a PRS using 269 variants and corresponding multi-ancestry weights and found that the PRS was highly predictive of prostate cancer risk across

populations(11). Of these 269 variants, 267 were present in the UK Biobank data and had an imputation info score >0.50 (median info score=0.99). The PRS was calculated as a weighted sum of the number of risk alleles among the 267 variants, using the variant-specific multi-ancestry weights we previously reported.

Association Testing

We evaluated the association between CHIP variants and prostate cancer risk using gene-based analyses, considering carrier status for qualifying CHIP variants within each gene individually. We also evaluated the association between CHIP variants and prostate cancer risk by aggregating across all genes, considering carrier status for any qualifying variants. Men carrying ≥1 allele among the identified CHIP genes (individually or in aggregate, depending on the assessment) were considered carriers. In the UK Biobank, we examined associations between carrier status and overall incident prostate cancer, age at enrollment, and age at diagnosis (for cases only), adjusting for age at enrollment (for overall prostate cancer and age at diagnosis) and the first 10 genetic principal components of ancestry to account for potential population stratification. In WESP, we examined associations between carrier status and aggressive versus non-aggressive prostate cancer, death due to prostate cancer versus non-aggressive prostate cancer, and metastatic versus non-aggressive prostate cancer. Age at diagnosis, study, country, and the first three principal components of ancestry were adjusted for as covariates. Logistic regression models were used for binary prostate cancer outcomes, while linear regression models were used for continuous age outcomes. Cox proportional hazards models were used to evaluate incident prostate cancer status in the UK Biobank with age in years as the time metric, using age at blood draw as the entry time and age at prostate cancer diagnosis as the exit time.

We evaluated the association between the prostate cancer PRS and age-related CHIP carrier status across all CHIP genes and for individual CHIP genes in prostate cancer controls from the UK Biobank. This analysis was performed using logistic regression models with CHIP status as the outcome and the continuous PRS as the predictor, adjusting for age at blood draw and the first 10 principal components of ancestry. We also evaluated the association between carrier status for P/LP/D variants in DNA repair genes and age-related CHIP carrier status using logistic regression models with CHIP status as the outcome and DNA repair gene carrier status as the predictor. In the UK Biobank, analyses were adjusted for age at blood draw and the first 10 principal components of ancestry, and in WESP, analyses were adjusted for age at prostate cancer diagnosis, study, country, and the first three principal components of ancestry. Analyses were performed aggregating across all 24 DNA repair genes and separately aggregating across three DNA repair genes: *BRCA2, ATM*, and *PALB2*, which we previously reported to be associated with aggressive prostate cancer(9).

A Bonferroni-corrected P-value <0.05 was considered statistically significant (P-values presented in the "Results" section are unadjusted). R 3.6.0 was used for all analyses.

Acknowledgements

This work was supported by the National Cancer Institute at the National Institutes of Health grant (K99 CA246063, BFD and R01 CA196931, CAH), an award from the Achievement Rewards for College Scientists Foundation Los Angeles Founder Chapter (BFD), and a Prostate Cancer Foundation Challenge Award (KHS and LAM). KTN is a Cancer Prevention Research Institute of Texas (CPRIT) Scholars in Cancer Research and supported by CPRIT RR190077.

This research has been conducted using the UK Biobank Resource under application numbers 42195 and 70505. Sequencing services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268201200008I.

Conflict of Interest Statement

The authors have no conflicts of interest to report.

Data Availability

Whole-exome sequencing data along with the clinical status of each WESP participant in this investigation is available through the database of genotypes and phenotypes (dbGaP, accession number: phs001524.v1.p1).

References

1 Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A. *et al.* (2014) Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*, **371**, 2488-2498.

2 Terao, C., Suzuki, A., Momozawa, Y., Akiyama, M., Ishigaki, K., Yamamoto, K., Matsuda, K., Murakami, Y., McCarroll, S.A., Kubo, M. *et al.* (2020) Chromosomal alterations among age-related haematopoietic clones in Japan. *Nature*, **584**, 130-135.

3 Loh, P.R., Genovese, G. and McCarroll, S.A. (2020) Monogenic and polygenic inheritance become instruments for clonal selection. *Nature*, **584**, 136-141.

4 Jacobs, K.B., Yeager, M., Zhou, W., Wacholder, S., Wang, Z., Rodriguez-Santiago, B., Hutchinson, A., Deng, X., Liu, C., Horner, M.J. *et al.* (2012) Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet*, **44**, 651-658.

5 Forsberg, L.A., Rasi, C., Malmqvist, N., Davies, H., Pasupulati, S., Pakalapati, G., Sandgren, J., Diaz de Stahl, T., Zaghlool, A., Giedraitis, V. *et al.* (2014) Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat Genet*, **46**, 624-628.

6 Thompson, D.J., Genovese, G., Halvardson, J., Ulirsch, J.C., Wright, D.J., Terao, C., Davidsson, O.B., Day, F.R., Sulem, P., Jiang, Y. *et al.* (2019) Genetic predisposition to mosaic Y chromosome loss in blood. *Nature*, **575**, 652-657.

7 Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*, **71**, 209-249.

8 Silver, A.J., Bick, A.G. and Savona, M.R. (2021) Germline risk of clonal haematopoiesis. *Nat Rev Genet*, in press.

9 Darst, B.F., Dadaev, T., Saunders, E., Sheng, X., Wan, P., Pooler, L., Xia, L.Y., Chanock, S., Berndt, S.I., Gapstur, S.M. *et al.* (2021) Germline Sequencing DNA Repair Genes in 5545 Men With Aggressive and Nonaggressive Prostate Cancer. *J Natl Cancer Inst*, **113**, 616-625.

10 Matejcic, M., Patel, Y., Lilyquist, J., Hu, C., Lee, K.Y., Gnanaolivu, R.D., Hart, S.N., Polley, E.C., Yadav, S., Boddicker, N.J. *et al.* (2020) Pathogenic Variants in Cancer Predisposition Genes and Prostate Cancer Risk in Men of African Ancestry. *JCO Precis Oncol*, **4**, 32-43.

11 Conti, D.V., Darst, B.F., Moss, L.C., Saunders, E.J., Sheng, X., Chou, A., Schumacher, F.R., Olama, A.A.A., Benlloch, S., Dadaev, T. *et al.* (2021) Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat Genet*, **53**, 65-75.

12 Darst, B.F., Sheng, X., Eeles, R.A., Kote-Jarai, Z., Conti, D.V. and Haiman, C.A. (2021) Combined Effect of a Polygenic Risk Score and Rare Genetic Variants on Prostate Cancer Risk. *Eur Urol*, in press.

13 Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S. and Getz, G. (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*, **31**, 213-219.

14 Bick, A.G., Weinstock, J.S., Nandakumar, S.K., Fulco, C.P., Bao, E.L., Zekavat, S.M., Szeto, M.D., Liao, X., Leventhal, M.J., Nasser, J. *et al.* (2020) Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature*, **586**, 763-768.

15 Jaiswal, S., Natarajan, P., Silver, A.J., Gibson, C.J., Bick, A.G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D. *et al.* (2017) Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med*, **377**, 111-121.

16 Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434-443.

17 McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P. and Cunningham, F. (2016) The Ensembl Variant Effect Predictor. *Genome Biol*, **17**, 122.

18 Zhu, B., Xiao, Y., Yeager, M., Clifford, G., Wentzensen, N., Cullen, M., Boland, J.F., Bass, S., Steinberg, M.K., Raine-Bennett, T. *et al.* (2020) Mutations in the HPV16 genome induced by APOBEC3 are associated with viral clearance. *Nat Commun*, **11**, 886.

19 Shin, H.T., Choi, Y.L., Yun, J.W., Kim, N.K.D., Kim, S.Y., Jeon, H.J., Nam, J.Y., Lee, C., Ryu, D., Kim, S.C. *et al.* (2017) Prevalence and detection of low-allele-fraction variants in clinical cancer samples. *Nat Commun*, **8**, 1377.

20 Desai, P., Mencia-Trinchant, N., Savenkov, O., Simon, M.S., Cheang, G., Lee, S., Samuel, M., Ritchie, E.K., Guzman, M.L., Ballman, K.V. *et al.* (2018) Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat Med*, **24**, 1015-1023.

21 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

22 Coombs, C.C., Zehir, A., Devlin, S.M., Kishtagari, A., Syed, A., Jonsson, P., Hyman, D.M., Solit, D.B., Robson, M.E., Baselga, J. *et al.* (2017) Therapy-Related Clonal Hematopoiesis

in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell*, **21**, 374-382 e374.

23 Genovese, G., Kahler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M. *et al.* (2014) Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*, **371**, 2477-2487.

24 Cancer Genome Atlas Research, N. (2015) The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, **163**, 1011-1025.

25 Wedge, D.C., Gundem, G., Mitchell, T., Woodcock, D.J., Martincorena, I., Ghori, M., Zamora, J., Butler, A., Whitaker, H., Kote-Jarai, Z. *et al.* (2018) Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. *Nat Genet*, **50**, 682-692.

26 Mancuso, N., Gayther, S., Gusev, A., Zheng, W., Penney, K.L., Kote-Jarai, Z., Eeles, R., Freedman, M., Haiman, C., Pasaniuc, B. *et al.* (2018) Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat Commun*, **9**, 4079.

27 Koboldt, D.C., Kanchi, K.L., Gui, B., Larson, D.E., Fulton, R.S., Isaacs, W.B., Kraja, A., Borecki, I.B., Jia, L., Wilson, R.K. *et al.* (2016) Rare Variation in TET2 Is Associated with Clinically Relevant Prostate Carcinoma in African Americans. *Cancer Epidemiol Biomarkers Prev*, **25**, 1456-1463.

28 Szustakowski, J.D., Balasubramanian, S., Kvikstad, E., Khalid, S., Bronson, P.G., Sasson, A., Wong, E., Liu, D., Wade Davis, J., Haefliger, C. *et al.* (2021) Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat Genet*, **53**, 942-948.

29 Van Hout, C.V., Tachmazidou, I., Backman, J.D., Hoffman, J.D., Liu, D., Pandey, A.K., Gonzaga-Jauregui, C., Khalid, S., Ye, B., Banerjee, N. *et al.* (2020) Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature*, **586**, 749-756.

30 Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res*, **12**, 996-1006.

31 Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D. and Kent, W.J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res*, **32**, D493-496.

32 Pritchard, C.C., Mateo, J., Walsh, M.F., De Sarkar, N., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R. *et al.* (2016) Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N Engl J Med*, **375**, 443-453.

33 Wu, Y., Yu, H., Li, S., Wiley, K., Zheng, S.L., LaDuca, H., Gielzak, M., Na, R., Sarver, B.A.J., Helfand, B.T. *et al.* (2020) Rare Germline Pathogenic Mutations of DNA Repair Genes Are Most Strongly Associated with Grade Group 5 Prostate Cancer. *Eur Urol Oncol*, **3**, 224-230.

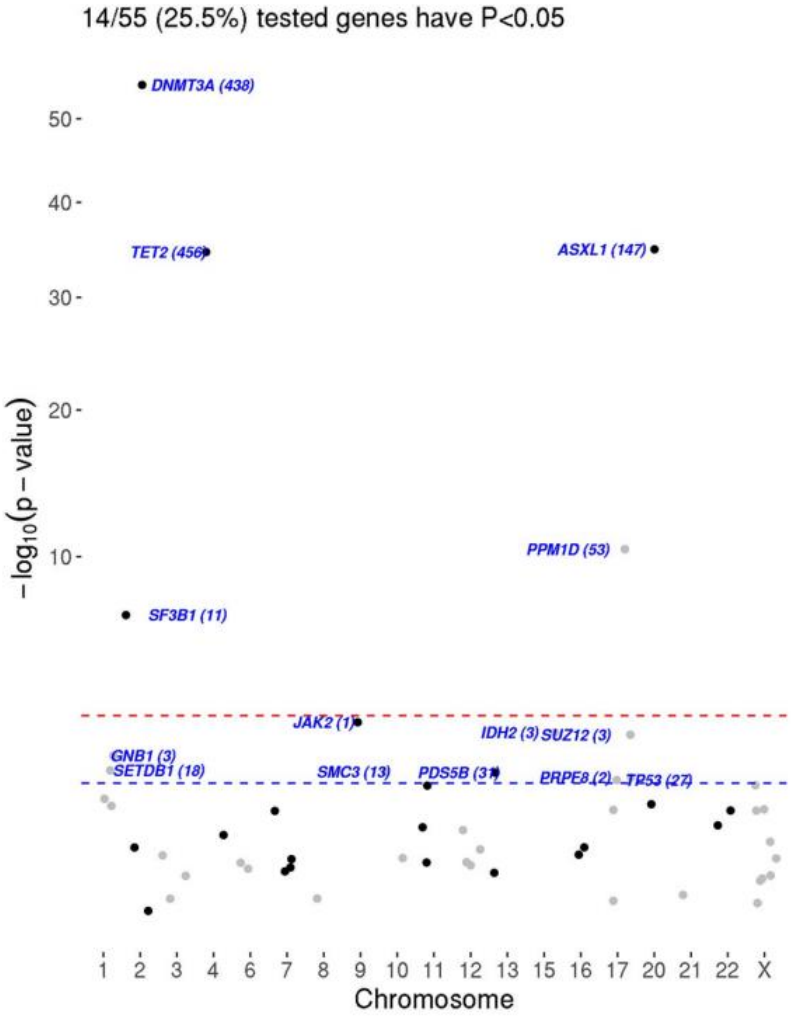
34 Leongamornlert, D.A., Saunders, E.J., Wakerell, S., Whitmore, I., Dadaev, T., Cieza-Borrella, C., Benafif, S., Brook, M.N., Donovan, J.L., Hamdy, F.C. *et al.* (2019) Germline DNA Repair Gene Mutations in Young-onset Prostate Cancer Cases in the UK: Evidence for a More Extensive Genetic Panel. *Eur Urol*, **76**, 329-337.

35 Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M. and Maglott, D.R. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*, **42**, D980-985.

36 Meeks, H.D., Song, H., Michailidou, K., Bolla, M.K., Dennis, J., Wang, Q., Barrowdale, D., Frost, D., Embrace, McGuffog, L. *et al.* (2016) BRCA2 Polymorphic Stop Codon K3326X and the Risk of Breast, Prostate, and Ovarian Cancers. *J Natl Cancer Inst*, **108**.

Legends to Figures

Figure 1. Manhattan plots of associations between CHIP genes and age at blood draw in the UK Biobank. CHIP: Clonal hematopoiesis of indeterminate potential.



Tables

			#Carriers (Carrier Frequencies)						UK Biobank		WESP					
	Variant Allelic Fraction, median (min-max)		UK Biobank		WESP				Case vs controls		Aggressive vs non-aggressive		Prostate cancer death vs non-aggressive		M1 vs. non-aggressive	
Gene (#Variants in UKB/WESP)	UK Biobank	WESP	Control (N=67,384)	Case (N=2,118)	Non-Agg (N=2,775)	Agg (N=2,770)	Death (N=2,052)	M1 (N=467)	HR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P
DNMT3A (438/79)	0.10 (0.05-0.56)	0.18 (0.10-0.48)	929 (1.38%)	36 (1.70%)	53 (1.91%)	49 (1.77%)	37 (1.80%)	8 (1.71%)	1.04 (0.75-1.45)	0.80	1.05 (0.70-1.59)	0.80	1.09 (0.70-1.70)	0.70	1.10 (0.50-2.46)	0.81
ASXL1 (147/30)	0.11 (0.05-0.48)	0.16 (0.10-0.40)	255 (0.38%)	7 (0.33%)	14 (0.50%)	20 (0.72%)	13 (0.63%)	3 (0.64%)	0.77 (0.37-1.62)	0.50	1.61 (0.79-3.28)	0.19	1.61 (0.73-3.54)	0.24	1.95 (0.52-7.38)	0.33
TET2 (456/88)	0.14 (0.05-0.75)	0.21 (0.10-0.60)	690 (1.02%)	34 (1.61%)	41 (1.48%)	48 (1.73%)	30 (1.46%)	8 (1.71%)	0.97 (0.69-1.36)	0.86	1.33 (0.86-2.05)	0.20	1.14 (0.70-1.87)	0.60	1.69 (0.75-3.81)	0.21
JAK2 (1/1)	0.17 (0.07-0.90)	0.13 (0.11-0.19)	44 (0.07%)	1 (0.05%)	4 (0.14%)	2 (0.07%)	2 (0.10%)	0 (0%)	0.55 (0.08-3.89)	0.55	0.65 (0.10-4.24)	0.65	0.85 (0.14-5.25)	0.86	--	--
All 55 UKB/52 WESP CHIP Genes (1,778/360)	0.11 (0.05-0.90)	0.19 (0.10-1.0)	2,771 (4.11%)	103 (4.86%)	194 (6.99%)	198 (7.15%)	131 (6.38%)	33 (7.07%)	0.93 (0.76-1.13)	0.46	1.14 (0.92-1.41)	0.22	1.02 (0.81-1.30)	0.84	1.22 (0.81-1.83)	0.34

Table 1. Associations of four common CHIP genes and the aggregate of all identified CHIP genes with overall and aggressive prostate cancer risk.

-- Estimate could not be reliably calculated due to lack of carriers.

CHIP: Clonal hematopoiesis of indeterminate potential; WESP: Whole-Exome Sequencing Study in Prostate Cancer; M1: Metastatic prostate cancer; UKB: UK Biobank

	Age at blood draw[1]		Age at prostate cancer diagnosis[2]		Age at prostate cancer diagnosis adjusted for age at blood draw[2]	
Gene (#Variants)	Age difference in years (95% CI)	P	Age difference in years (95% CI)	P	Age difference in years (95% CI)	P
DNMT3A (438)	4.08 (3.57-4.60)	$4.4x10^{-55}$	1.35 (-0.39-3.09)	0.13	-0.67 (-1.39-0.04)	0.07
ASXL1 (147)	6.22 (5.24-7.19)	$1.3x10^{-35}$	0.78 (-3.15-4.70)	0.70	-1.32 (-2.94-0.30)	0.11
TET2 (457)	3.74 (3.15-4.33)	$2.6x10^{-35}$	1.16 (-0.63-2.95)	0.20	0.52 (-0.22-1.26)	0.17
JAK2 (1)	3.83 (1.47-6.19)	$1.5x10^{-03}$	1.91 (-8.44-12.25)	0.72	2.39 (-1.87-6.66)	0.27
All 55 CHIP Genes (1,778)	3.25 (2.95-3.55)	$2.4x10^{-99}$	0.81 (-0.23-1.86)	0.13	-0.21 (-0.64-0.22)	0.34

Table 2. Associations of four common CHIP genes and the aggregate of all identified CHIP genes with age at blood draw and prostate cancer diagnosis in the UK Biobank.

[1] Analysis performed in all participants

[2] Analysis performed in incident prostate cancer cases

CHIP: Clonal hematopoiesis of indeterminate potential

Abbreviations

Clonal hematopoiesis (CH); clonal hematopoiesis of indeterminate potential (CHIP); loss of chromosome Y (LOY); Whole-Exome Sequencing Study in Prostate Cancer (WESP); polygenic risk score (PRS); whole-exome sequencing (WES); minor allele frequency (MAF); Genome Aggregation Database (gnomAD); pathogenic/likely pathogenic/deleterious (P/LP/D); Variant Effect Predictor (VEP); genome-wide association study (GWAS); Variant allelic fraction (VAF)