

# **Age-related loss of the Y chromosome is associated with sex hormone binding globulin and selected drivers of clonal hematopoiesis**

Ahmed A.Z. Dawoud,<sup>1</sup> William J Tapper,<sup>1\*</sup> & Nicholas C.P. Cross<sup>1,2\*</sup>

<sup>1</sup> School of Medicine, University of Southampton, Southampton, Hampshire, SO17 1BJ, UK

<sup>2</sup> Wessex Regional Genetics Laboratory, Salisbury NHS Foundation Trust, Salisbury, Wiltshire, SP2 8BJ, UK

\* equal contributions

Correspondence to:

Professor N.C.P. Cross  
Wessex Regional Genetics Laboratory  
Salisbury NHS Foundation Trust  
Salisbury SP2 8BJ, UK

Tel: +(44) 1722 429080  
Fax: +(44) 1722 331531  
email: [ncpc@soton.ac.uk](mailto:ncpc@soton.ac.uk)

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

## Abstract

Mosaic loss of the Y-chromosome (LOY) is the most common somatic alteration in men. We aimed to assess the relationship between LOY and serum biomarkers in UK Biobank and explore the interaction with constitutional and somatic genetics. LOY was strongly associated with levels of sex hormone binding globulin (SHBG,  $\beta=0.12$ ,  $P_{FDR}=P=7.44 \times 10^{-36}$ , adjusted for age, age squared, gender, smoking status, smoking intensity and principal genetic components), a key regulator of testosterone bioavailability associated with diverse disorders including cancer and autoimmune diseases. Furthermore, LOY was associated with total testosterone (TT,  $\beta=0.09$ ,  $P_{FDR}=2.23 \times 10^{-20}$ ), but not bioavailable testosterone ( $P_{FDR}=0.46$ ) or free testosterone ( $P_{FDR}=0.75$ ). These relationships remained significant after sensitivity analysis that included comorbidities and body mass index (SHBG,  $\beta=0.08$ ,  $P_{FDR}=4.61 \times 10^{-21}$ ; TT,  $\beta=0.05$ ,  $P_{FDR}=4.13 \times 10^{-9}$ ). Mendelian randomisation suggested a causal effect of SHBG on LOY in BioBank Japan ( $P=6.58 \times 10^{-4}$ ) but there was no evidence for an effect of LOY on SHBG ( $P=0.46$ ). Assessment of cis-eQTLs for 13 genes associated with LOY identified two SNPs that were also associated with levels of SHBG, however only rs7141210 (imprinted *DLK1-MEG3* locus) modified the relationship between SHBG and LOY (rs7141210-T/T;  $P_{interaction}=0.04$ ) with low levels of SHBG seen specifically in men without LOY ( $\beta=-0.02$ ,  $P=0.001$ ), but not those with LOY ( $P=0.41$ ). Age-related clonal hematopoiesis (CH) defined by somatic driver mutations was not associated with sex hormone levels but was associated with LOY at clonal fractions  $>30\%$  (OR=1.52,  $P=2.92 \times 10^{-4}$ ). *TET2*, *TP53*, and *CBL* mutations were enriched in high level LOY cases, but not *DNMT3A* or *ASXL1*. Our findings thus identify independent relationships between LOY, sex hormones and CH.

## Introduction

Age-related somatic loss of the Y-chromosome (LOY) in peripheral blood leukocytes is the most prevalent chromosomal alteration in men.<sup>1</sup> LOY has been identified in as many as 20% of UKB male participants (median age = 58), but only 10% of these (2% of all men) had LOY involving >20% of leukocytes.<sup>2-4</sup> A series of genome wide association studies (GWAS) have characterized inherited genetic variation that predisposes to LOY,<sup>2,3,5</sup> with 156 independent loci explaining up to 34% of the heritability.<sup>2,6</sup> LOY has also been linked to smoking behaviour,<sup>1,3,7</sup> indicating that the environment as well as genetics is an important factor. LOY is associated with all-cause mortality, cancer mortality,<sup>1,8,9</sup> and a wide range of non-malignant conditions.<sup>1,10-13</sup> LOY is also associated with variation in blood cell counts<sup>4,14</sup> and has long been recognized as a recurrent clonal cytogenetic finding in hematological malignancies where, in the absence of other changes, it is associated with a good prognosis.<sup>15</sup> LOY with large clone size has been linked to the presence of somatic mutations associated with hematological malignancies and an elevated risk of developing myeloid neoplasia in two recent, small studies of selected cases.<sup>16,17</sup> It has been suggested that LOY might be a broad marker of genomic instability across different tissues and may exert its effects by altering immune cell function.<sup>2,18</sup>

Clonal hematopoiesis (CH) is a widespread phenomenon characterized by the presence of expanded clones of mutated blood cells,<sup>19</sup> predominantly in individuals over the age of 60. Large autosomal chromosomal alterations including gains, losses and copy number neutral loss of heterozygosity (CN-LOH) inferred from SNP array data have been used to identify clonality,<sup>20,21</sup> but CH is more commonly recognized by sequence analysis and the finding of pathogenic driver mutations associated with hematological malignancies, most commonly in the epigenetic regulators *DNMT3A*, *TET2*, and *ASXL1* but also a wide range of other genes.<sup>22-28</sup> Broad screens by whole exome sequencing (WES) or whole genome sequencing (WGS) have revealed that clonality in the absence of known driver mutations (unknown driver CH) is even more prevalent than CH with driver mutations.<sup>26</sup> Like LOY, CH has been linked to a wide range of malignant and non-malignant diseases.<sup>23,24,29-33</sup> Most prominently, CH defined by autosomal chromosomal alterations or driver mutations is associated with a significant risk (hazard ratios of 10-12.9) of subsequent hematological malignancies,<sup>20,21,24,25</sup> and recent studies showed a lineage-specific risk for mutations in genes associated with myeloid or lymphoid neoplasms.<sup>34,35</sup> The clinical consequences of unknown driver CH have not been defined, and the reason for clonality in these cases remains unclear. Furthermore, the extent to which mosaic LOY can be considered as CH has not been defined. Although age is the major risk factor for both CH and LOY, it has become clear that there is significant overlap between

genetic factors that predispose to both of these abnormalities. Broadly, inherited variants in cancer susceptibility genes and genes that are mutated in cancer feature prominently as risk factors for both CH and LOY.<sup>2,3,6,36-38</sup> As for external factors, prior chemotherapy is associated with CH characterized by *PPM1D* and *TP53* mutations,<sup>39</sup> whereas smoking is associated with *ASXL1* mutations as well as LOY.<sup>40</sup>

It is widely accepted that biochemical profiles change with age in a manner independent of specific disease states, for example ageing is associated with depletion of sex hormones.<sup>41</sup> Sex hormone binding globulin (SHBG) is a glycoprotein that binds to steroids with high affinity, with both 5 $\alpha$ -dihydrotestosterone and testosterone binding much more strongly than oestradiol.<sup>42</sup> In men, circulating testosterone levels are regulated by SHBG, with on average 58% bound to SHBG, 40% bound to albumin, and 2% as free testosterone (FT).<sup>43,44</sup> Binding to albumin is weak and so all non-SHBG-bound testosterone is considered as bioavailable testosterone (BAT).<sup>45</sup> Ageing is associated with a decline in total testosterone (TT), FT and BAT and an increase in SHBG<sup>41</sup> but the link, if any, between sex hormones and somatic genetic changes remains unexplored. In this study we aimed to determine if sex hormones and other serum markers are associated with LOY and to understand the interaction between biochemical factors with both constitutional and somatic genetics.

## Material and methods

### *Study cohort*

UK Biobank (UKB) is a large prospective cohort described in detail elsewhere<sup>46</sup> involving approximately 500,000 individuals aged between 40 and 69 years at recruitment. Genome wide SNP data derived from peripheral blood leucocytes was available for most participants, and WES data for 200,631 participants at the time of analysis.

### *Mosaic loss of the Y chromosome*

A previous study used SNP data to identify males with mosaic LOY (n=44,588; 20% of evaluable males) using a method which compared allelic intensities for statistically phased haplotypes of the pseudo-autosomal region 1 (PAR1).<sup>2</sup> This method for detecting LOY was considered to be less error prone than those based on the median genotyping intensity over the non-pseudoautosomal region of the Y chromosome and was able to detect mosaicism with a clonal fraction down to 1%.<sup>2,47</sup> The spectrum of LOY was categorised according to clonal fraction by considering the median change of B-allele frequency (BAF) in PAR1, specifically BAFs of 0.026, 0.056, and 0.088 were previously shown to correspond with clonal fractions of 10%, 20%, and 30% respectively.<sup>2</sup>

### *Biochemistry markers and sex hormones*

Measurements of 29 biochemistry markers were available from serum samples collected on recruitment to UKB.<sup>46</sup> Mass action equations were used to calculate FT and BAT from measurements of SHBG, TT and albumin as described.<sup>48</sup> Further details regarding the biochemical assay methods and external quality assurance are available at [https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/serum\\_biochemistry.pdf](https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/serum_biochemistry.pdf).

### *The relationship between biochemistry markers and LOY*

We focused on the subset of male UKB participants who were evaluable for LOY assessment (n=222,835).<sup>2</sup> The relationship between LOY and each of 31 biomarkers (29 measured and 2 calculated) was tested using multivariable linear regression in R. Continuous measures for each sex hormone were transformed into a normal distribution using inverse normal rank transformation and used as the dependant variable. For the basic model, the independent variables were LOY as a binary predictor, age, age squared, smoking status (never, previous, current), smoking intensity defined by average number of pack-years smoked by an individual over their adult lifetime<sup>49</sup> (Data-Field 20162) and the first 10 genetic principal components (10 PCA). For sensitivity analysis, we added more variables to the basic model, specifically binary variables derived from first occurrences of primary

hypertension (ICD10:I10), insulin-dependent diabetes (E10), and non-insulin-dependent diabetes (E11), plus body mass index (BMI) as a continuous variable. Effect sizes were reported as beta coefficients ( $\beta$ ) with 95% confidence intervals (95% CI). P-values were adjusted for 31 tests using the False Discovery Rate (FDR) method.<sup>50</sup> The average measure of serum biomarkers in participants with LOY were compared to those in LOY free controls using Mann–Whitney U tests.

#### *Allelic SHBG score in UKB*

Thirteen genetic variants were associated with circulating SHBG levels by a previous two-stage GWAS, including 10 variants that achieved genome wide significance plus 3 independent cis variants that were identified by conditional analysis of the *SHBG* gene.<sup>51</sup> After excluding one SNP with heterogeneity towards females ( $P = 0.02$ , rs440837) and a second SNP on the X-chromosome (rs1573036), 11 SNPs were considered in relation to LOY. To consider both precision and power, two allelic scores were constructed. The first consisted of all 11 SNPs (rs12150660, rs1641537, rs1625895, rs6258, rs17496332, rs2411984, rs293428, rs780093, rs7910927, rs8023580 and rs4149056) while the second was based only on the 4 SNPs located in the vicinity of *SHBG* (rs12150660, rs1641537, rs1625895 and rs6258). Each score was calculated as the sum of number of risk factor-increasing alleles per SNP weighted by their corresponding genetic effect size. The allelic scores were calculated using Plink V1.9,<sup>52</sup> and transformed into a normal distribution using inverse normal rank transformation. A multivariable logistic regression model adjusted for age, sex, smoking status and 10 PCA was used to assess the relationship between LOY status (binary, dependent) and *SHBG* allelic scores (independent, continuous).

#### *Mendelian Randomisation*

Mendelian randomisation (MR) was used to assess the possibility of a causal relationship between SHBG and LOY using germline SNPs associated with circulating SHBG as instrumental variables, following the STROBE guidelines.<sup>53</sup> Eight SNPs were used as instrumental variables without considering estimated genetic effect sizes for LOY in 95,380 men recruited to BioBank Japan (BBJ).<sup>54</sup> Three of the eleven SNPs that were used to generate the allelic score (rs12150660, rs6258 and rs2411984) were excluded as genotypes were unavailable and/or non-informative in BBJ. We applied an inverse variance weighted model (IVW) with fixed effect using the TwoSamplesMR package.<sup>55</sup> To explore the effect of SNP selection, we repeated the MR analysis using a genome wide significance threshold ( $P < 5 \times 10^{-8}$ ) to select four SNPs (rs1641537, rs1625895, rs293428 and rs7910927) associated with SHBG.<sup>51</sup> We also performed leave-one-out analysis to evaluate the effect of each SNP on the analysis. To assess the effect of LOY on SHBG levels we examined SNPs ( $n=50$ )

previously associated with LOY<sup>54</sup> in BBJ. 10 SNPs were excluded as genotypes were unavailable and/or non-informative in UKB. The remaining 40 SNPs were used as instrumental variables without considering estimated genetic effect sizes for SHBG in UKB men.

### *Identification of CH*

A propensity score matching method in R<sup>56</sup> was used to select one control per case and to match for age in comparison to participants with LOY. Somatic mutations were called in LOY samples and matched controls using GATK (Version 4.1.9) and Mutect2<sup>57</sup> to process individual CRAM files in the tumour-only mode. Following best practice guidelines (<https://gatk.broadinstitute.org/hc/en-us/articles/360035531132>), a Panel Of Normal (PON; version 23<sup>rd</sup> August 2017) from the Broad Institute that were generated using Mutect2 on samples from the 1000 genomes project to identify recurrent artifacts, and germline variants from gnomAD were used to remove artefacts and germline variants. To identify putative somatic driver mutations, the analysis was restricted to rare variants with a minor allele frequency (MAF) <0.01 in gnomAD and a minimum number of reads supporting the mutated allele: 3 reads for point mutations and 6 reads for indels. Other methodological details for calling somatic variants are described in detail in our previous publication.<sup>31</sup> Variants that satisfied either of the following two criteria were selected: first; recurrent driver mutations as defined in our previous study;<sup>58</sup> second singleton variants that passed all Mutect2 filters with variant allele frequencies (VAF) between 0.1 and 0.2.<sup>26</sup> Driver mutations were classified into myeloid or lymphoid according to a published list of genes associated with myeloid neoplasms (n=76)<sup>23</sup> and genes associated with acute lymphoblastic leukemia or chronic lymphocytic leukemia in the Cancer Gene Census (CGC),<sup>59</sup> respectively (Table S1). Participants with variants in myeloid genes were considered as having myeloid CH (n=2,890), and those with variants in lymphoid genes as having lymphoid CH (n=532). Cases with mutations in genes involved in both myeloid and lymphoid neoplasms were considered as myeloid CH. To identify participants with evidence of clonality in the absence of pathogenic mutations in known driver genes, we also identified variants in all coding genes that were not defined as myeloid or lymphoid. Participants with singleton variants in any gene with VAF range between 0.1 and 0.2 not defined as myeloid or lymphoid were considered as having unknown driver CH (n=15,874).

### *The relationship between LOY and CH*

The association between LOY and all variants, driver variants, myeloid CH, lymphoid CH, and unknown driver CH were tested using logistic regression in R. We further assessed the relationship with LOY clone size categories ( $\leq 10\%$ , 10%-20%, 20%-30%, and  $>30\%$ ) using Fisher's exact tests. The

strength of the association was reported as odds ratio (OR) with 95% CI. P-values were adjusted for 20 tests using the FDR method. To test the relationship at the individual driver gene level, the analysis was focused on genes mutated in  $\geq 3$  males with LOY in  $>30\%$  of cells.

#### *Assessment of the co-existence of LOY and CH*

The co-existence of CH and LOY was tested by assessing the relationship between the BAF for LOY and VAF of driver mutations. For cases with two or more mutations the highest VAF was used. To avoid excess CH with VAF between 0.1-0.2, the analysis for myeloid and lymphoid CH was restricted to VAFs detected for recurrent driver mutations as defined in our previous study.<sup>31</sup> We assessed the relationship with LOY in  $\leq 10\%$  and  $>10\%$  of cells. The dependent variables were VAF, age and smoking status (never, previous, current). The strength of the association was reported as  $\beta$  coefficient with 95% CI. P-values were adjusted for 6 tests using the FDR method.

#### *The relationship between CH and sex hormones*

The association between CH and sex-hormone levels were tested using linear regression in R, with normally transformed sex hormone as the independent variable. The dependent variables were driver mutation state as a binary predictor, age, smoking status (never, previous, current), and 10 PCA. The association was reported as  $\beta$  coefficient with 95% CI. P-values were adjusted for 4 tests using the FDR method.

#### *Expression quantitative trait analyses*

The eQTLGen database incorporates expression quantitative trait locus (eQTL) data from blood samples from a total of 31,684 individuals.<sup>60</sup> To select a genetic proxy for gene expression, we filtered cis-eQTLs within a distance  $<1$  Mb, and  $FDR < 0.05$  and selected the SNP with the smallest FDR value, with no other genes showing a stronger association with the selected SNP to minimize horizontal pleiotropy. We restricted the analysis to directly genotyped SNPs with  $MAF > 0.05$  in UKB. 19 SNPs associated with LOY were associated with 27 genes by position, biological function, expression, or nonsynonymous variants in the gene.<sup>3</sup> Our analysis was restricted to 13 of these genes for which an expression proxy was identified.

#### *Assessment of the interaction between eQTL, LOY and SHBG levels*

The most significant eQTL SNP for each gene was encoded according to an additive model for the risk allele (0/1/2) in UKB and was factorized in each model with 0 as the reference (0/1 for heterozygous, and 0/2 if homozygous for the risk allele). The following statistical tests were applied,



and adjusted for age, smoking, 10 PCA, and multiple tests by the FDR method. First, the relationship between each eQTL and LOY was assessed using logistic regression in R where LOY (binary) was the dependent variable and eQTL was the independent variable. Second, the relationship between eQTLs and SHBG levels was assessed using linear regression in R where SHBG (continuous) was taken as the dependent variable and transformed to a normal distribution using rank transformation, and eQTLs were the independent variable. Finally, if an eQTL was significantly associated with both LOY and SHBG, the interaction effect of the eQTL and LOY on SHBG regression was assessed by linear regression in R. Inverse normal rank transformed SHBG was considered as a continuous dependent variable, and each of eQTL, LOY, and eQTL x LOY (interaction effect) as the independent variable. To visualize the interaction effect in our models, we used '*interactions*' in R, a tool that was developed to interpret statistical interactions in regression models.<sup>61</sup>

## Results

### *The relationship between LOY and biochemistry markers*

To investigate the relationship between LOY and serum biomarkers, we used previously published calls of LOY that were generated by utilizing long-range phasing information to analyse allele-specific genotyping intensities of 1,239 variants in the pseudo-autosomal region 1.<sup>2</sup> We restricted our analysis to the 222,835 males who passed QC, of whom 44,558 (20%) had LOY. Of these, the majority (n=31,952; 72%), had an estimated LOY clonal fraction of <10%. We compared the presence or absence of LOY with 29 biochemistry parameters that were directly measured by UKB, as well as estimated levels of FT (median = 0.21 nmol/L; range = 0.003 - 1.93) and BAT (median = 5.1 nmol/L; range = 0.09 - 45.68) that were derived from measurements of SHBG, TT and albumin.<sup>62</sup>

Univariate comparisons revealed that participants with LOY had higher median levels of alkaline phosphatase, apolipoprotein A, C-reactive protein, cystatin C, glucose, glycated hemoglobin (HbA1c), HDL cholesterol, TT, urea, SHBG and vitamin D. Lower medians levels in participants with LOY were seen for alanine aminotransferase, albumin, apolipoprotein B, aspartate aminotransferase, calcium, cholesterol, gamma glutamyltransferase, IGF-1, LDL direct, total bilirubin, total protein, triglycerides, urate and FT. (Table S2).

On initial multivariate analysis adjusted for age, age squared, sex, 10 PCA, smoking status, smoking intensity and FDR corrected for multiple testing, we found that LOY as a binary predictor was most strongly associated with elevated levels of SHBG ( $\beta = 0.12$ , 95% CI: 0.11 - 0.13,  $P = 7.44 \times 10^{-36}$ ). SHBG binds steroids<sup>42</sup> and it is notable that the second strongest positive association was with TT ( $\beta = 0.09$ , 95% CI: 0.07 - 0.11,  $P = 2.23 \times 10^{-20}$ ). There was no association, however, between LOY and FT ( $P = 0.46$ ) or BAT ( $P = 0.75$ ) (Figure 1A and Table S3). Participants with LOY had higher levels of SHBG and TT (SHBG: median nmol/L = 41.54 vs 35.86,  $P < 0.001$ ; TT: median nmol/L = 11.74 vs 11.58,  $P < 0.001$ ; Mann-Whitney U tests) but lower levels of FT (median nmol/L = 0.19 vs 0.20,  $P < 0.001$ ), and BAT (median nmol/L = 4.78 vs 5.18,  $P < 0.001$ ). On sensitivity analysis corrected for hypertension, insulin dependent diabetes, non-insulin dependent diabetes, and BMI, LOY retained the positive association with SHBG ( $\beta = 0.08$ , 95% CI: 0.07 – 0.10,  $P = 4.61 \times 10^{-21}$ ) and TT ( $\beta = 0.05$ , 95% CI: 0.04 – 0.07,  $P = 4.13 \times 10^{-9}$ ; Figure 1B, and Table S3). Our observational results points to a direct relationship between levels of SHBG and LOY that cannot be explained by age, smoking, population stratification, common morbidities, BMI or free/bioavailable testosterone (Figure 2).

### *The relationship between genetically defined SHBG and LOY*

Published GWAS have identified multiple genetic determinants of SHBG levels in serum.<sup>51</sup> To understand the relationship between LOY and SHBG we generated allelic scores to summarize the genetic variation associated with SHBG and evaluated the scores as predictors of LOY. Since the allelic scores were derived from independent cohorts they represent unbiased instruments to assess the relationship with LOY in UKB.<sup>63</sup> We found that genetically predicted SHBG was significantly associated with the finding of LOY in UKB using the score estimated from 11 independent SNPs (OR = 1.02, 95% CI: 1.01 – 1.04,  $P = 5.59 \times 10^{-5}$ ).

To assess the possibility of a causal relationship between SHBG and LOY we performed MR analysis (Figure 3). In a liberal analysis, 8 independent SNPs were used to estimate the effect of SHBG on LOY (Table S4). Using an IVW model with fixed effect we identified a positive causal relationship ( $\beta = 0.15$ , 95% CI: 0.06 - 0.23,  $P = 6.58 \times 10^{-4}$ ). In a more conservative analysis restricted to 4 SNPs associated with SHBG with  $P < 5 \times 10^{-8}$ , the effect of SHBG on LOY was confirmed ( $\beta = 0.17$ , 95% CI: 0.07 - 0.26,  $P = 7.28 \times 10^{-4}$ ). Leave-one-out analysis (Table S5) found that significance was lost when rs7910927 at 10q21.3 within *JMJD1C* was excluded (liberal analysis:  $\beta = 0.08$ , 95% CI: -0.02 - 0.17,  $P = 0.13$ ; conservative analysis:  $\beta = 0.07$ , 95% CI: -0.04 - 0.19,  $P = 0.22$ ). To assess the possibility of bidirectional effect we utilized 40 SNPs with  $P < 5 \times 10^{-8}$  associated with LOY in Japanese men<sup>54</sup> to measure their effect on SHBG levels in UKB men<sup>64</sup> but no significant relationship was found ( $\beta = 0.02$ , 95% CI: -0.88 - 0.13,  $P = 0.46$ ).

### *The effect of gene expression on the relationship between SHBG and LOY*

To shed light on the possible mechanisms by which SHBG levels might promote LOY, we focused on 19 genomic regions robustly associated with LOY which had been confirmed in an independent GWAS.<sup>2,3</sup> Using the eQTLGene<sup>60</sup> database we identified eQTL SNPs to serve as valid proxies for 13 of the 19 regions (*ACAT1*, *BCL2*, *DLK1*, *HM13*, *MAD1L1*, *QKI*, *RBPM5*, *SEMA4A*, *SEN7*, *SEN8*, *SETBP1*, *SMPD2*, *TCL1A*, and *TSC22D2*). We found that the heterozygous state of 8/13 eQTLs and the homozygous state of 9/13 eQTLs were associated with LOY in UKB. The eQTL for *TCL1A* at 14q32.13 had the strongest association with LOY status (rs11849538\_G/C, OR = 0.83, 95% CI = 0.80 - 0.85,  $P = 6.32 \times 10^{-34}$ ; rs11849538\_G/G, OR = 0.63, 95% CI: 0.56 - 0.69,  $P = 1.48 \times 10^{-16}$ ), as shown in Figure 4 and Table S6. Only 2 of the 13 genes, however, were significantly associated with levels of SHBG. *MAD1L1* at 7p22.3 was positively associated with SHBG (rs10247428\_A/A,  $\beta = 0.02$ , 95% CI = 0.01 - 0.03,  $P = 0.003$ ), but this was in the opposite direction to its relationship with LOY. The *DLK1-MEG3* eQTL at 14q32.2 was negatively associated with SHBG (rs7141210-T/T,  $\beta = -0.02$ , 95% CI = -0.03 -

0.007,  $P = 0.02$ ) and also negatively associated with LOY (Table S6). Of these two signals, only rs7141210-T/T modified the relationship between LOY and SHBG ( $P_{\text{interaction}} = 0.04$ ) (Figure 5, Table S7). There was no influence of rs7141210-T on TT, FT or BAT (Table S8) indicating that the interaction was specific for SHBG.

### *The relationship between LOY and CH*

To assess the impact of somatic driver mutations and other markers of clonality on the relationship between SHBG, testosterone and LOY, we first assessed the relationship between somatic variants (driver and non-driver) and LOY. WES data was available for 17,759 participants with LOY, of whom 28% ( $n=4,981$ ) were estimated to have an LOY clone size  $\geq 10\%$  of leucocytes. For comparison, we randomly selected UKB age-matched controls ( $n=17,702$ ) that were negative for LOY. We identified recurrent somatic mutations in driver genes associated with myeloid CH and lymphoid CH, plus likely somatic variants in other genes that indicated clonality in the absence of known driver mutations, which we refer as unknown driver CH. Overall, the frequency of each CH subtype (myeloid, lymphoid, unknown) was similar between cases with LOY and controls. Striking differences emerged, however, when LOY was stratified by clone size. All CH (myeloid plus lymphoid plus unknown driver) was significantly associated with LOY in  $\geq 10\%$  of cells with a clear increase in the strength of the association with increasing LOY clone size (10-20% LOY, OR = 1.17,  $P = 1.81 \times 10^{-4}$ ; 20-30% LOY, OR = 2.20,  $P = 4.25 \times 10^{-27}$ ;  $\geq 30\%$  LOY, OR = 3.43,  $P = 2.42 \times 10^{-52}$ ). Both myeloid CH (OR = 1.42,  $P = 4.52 \times 10^{-3}$ ), and lymphoid CH (OR = 1.93,  $P = 0.01$ ) were significantly associated with LOY in  $\geq 30\%$  of cells but not LOY of smaller clone size. The relationship was most prominent, however, for unknown driver CH (OR = 2.46,  $P = 5.09 \times 10^{-34}$ ). Full results are presented in Figure 6 and Table S9. None of the three CH subtypes were associated with SHBG or the three measures of testosterone, indicating no effect of driver mutations on the relationship between LOY and sex hormones (Table S10).

To understand the relationship between CH and LOY in more detail, we assessed the association between somatic mutations in specific driver genes in participants with LOY in  $>30\%$  cells ( $n=823$ ) compared to LOY free controls ( $n=17,702$ ). *TET2* was the most significantly enriched mutated gene in LOY cases (4% versus 1.5% in controls, OR = 2.64,  $P = 9.58 \times 10^{-5}$ ) with *TP53* (OR = 6.96,  $P = 7.62 \times 10^{-3}$ ) and *CBL* (OR = 7.43,  $P = 0.04$ ) mutations also showing a significant enrichment (Table S11). Other genes, including *DNMT3A* and *ASXL1*, showed no enrichment in high level LOY cases.

The possibility that LOY and CH as defined by somatic mutations might co-exist in the same clone was assessed by analysing the relationship between LOY BAF and the VAFs of driver mutations.

Figure 7 shows a summary of the results at different ranges of LOY. Myeloid CH VAFs predicted BAF levels in samples with LOY >10% ( $\beta = 0.10$ , 95% CI = 0.05 - 0.15,  $P = 1.12 \times 10^{-4}$ ). Similar results were seen for lymphoid CH ( $\beta = 0.20$ , 95% CI = 0.06 - 0.35,  $P = 0.02$ ) and also unknown driver CH ( $\beta = 0.19$ , 95% CI = 0.14 - 0.24,  $P = 3.72 \times 10^{-12}$ ), which by definition was restricted to VAFs of 0.1 - 0.2 (Table S12).

## Discussion

Age-related mosaic LOY in peripheral blood leukocytes is known to be influenced by both genetic and environmental factors. We have found that LOY is also strongly with multiple serum biomarkers, most notably levels of SHBG, and that this association is independent of known confounders (age, age squared, smoking history, common comorbidities, BMI and the first 10 principal genetic components). Furthermore, we found genetic evidence to support the hypothesis that SHBG levels are causally linked to LOY, but no evidence that LOY has any effect on SHBG. SHBG regulates the level of circulating testosterone and, although we found that both FT and BAT were lower in men with LOY compared to those without LOY, there was no significant relationship between LOY and either FT or BAT on multivariate analysis. This finding is inconsistent with the free hormone hypothesis, which proposes that only the unbound fraction of testosterone is biologically active in target tissues,<sup>65</sup> and instead suggests that involvement of other pathways such as binding and internalization of SHBG-bound testosterone by specific cell types, including B-cells.<sup>66,67</sup> The mechanism by which SHBG promotes LOY is unclear, but from a genetic perspective the effect is not explained by variation at *SHBG* alone. Other loci are involved, particularly *JMJD1C* which encodes a histone demethylase previously linked to SHBG levels.<sup>51</sup>

To understand the influence of genetic factors on the relationship between SHBG and LOY in more detail, we focused on genetically predicted expression of genes linked to the development of LOY. We found that rs7141210-T, a marker for expression of genes in the *DLK1–MEG3* region at 14q32,<sup>60</sup> was associated with elevated LOY and SHBG, and that homozygosity for the T-allele of rs7141210 modified the relationship between LOY and SHBG. *DLK1–MEG3* is a large and complex imprinted cluster of genes and non-coding RNAs. The methylated paternally derived chromosome expresses the protein-coding genes *DLK1*, *RTL1* and *DIO3*, while the non-methylated maternally derived chromosome expresses the non-coding genes *MEG3*, *MEG8*, *asRTL1*, multiple miRNAs and lncRNAs.<sup>68</sup> Constitutional uniparental disomy (UPD) at 14q32 is associated with the developmental disorders Temple syndrome (maternal UPD) and Kagami–Ogata syndrome (paternal UPD) whereas somatically acquired paternal UPD is associated with CH and myeloid malignancies.<sup>69</sup> Genome-wide significant signals have been identified near *DLK1* in association with CH defined by acquired 14q UPD<sup>70</sup> and somatic driver mutations<sup>71</sup> as well as LOY.<sup>2,3,54</sup> Furthermore, a parent of origin specific effect of rs1555405-A linked to differential methylation has been defined in relation to platelet counts.<sup>72</sup> This SNP is in linkage disequilibrium with rs7141210 ( $D'=1$ ,  $R^2=0.7$ ), with rs7141210-T allele being correlated with rs1555405-A allele. Collectively these findings suggest the possibility of a parent of origin impact of rs7141210-T on the relationship between SHBG and LOY.

We have defined the relationship between CH and LOY in UKB. LOY in  $\geq 30\%$  of cells was associated with both myeloid and lymphoid CH, with 14% of affected individuals having one or more somatic driver mutation compared to 10% of controls ( $P = 2.92 \times 10^{-4}$ ; Table S9). At the level of individual genes, the most striking finding was that mutated *TET2* was associated with LOY, but not *DNMT3A* or *ASXL1*. Collectively, mutations in one of these genes accounts for 90% of cases of CH defined by sequence analysis, and our findings are consistent with the notion that CH with *TET2* mutations is different from CH with *DNMT3A* or *ASXL1* mutations. With larger studies, specific disease associations are emerging, for example CH with *TET2* mutations has been linked to chronic obstructive pulmonary diseases,<sup>73</sup> but not CH with *DNMT3A* mutations.

Most strikingly, however, unknown driver CH was seen in 65% of cases with high level ( $\geq 30\%$  of cells) LOY compared to just 43% of controls ( $P = 5.09 \times 10^{-34}$ ). Overall, 80% of cases with high level LOY had mutational evidence of clonality, with LOY in  $\geq 10\%$  of cells clearly associated with unknown driver CH. For the first time, therefore, our findings provide broad molecular confirmation that LOY  $\geq 10\%$  is indeed clonal, and we predict that comprehensive sequencing by WGS will confirm clonality in most cases. Importantly for our study, neither overall CH nor any CH subtype was associated with SHBG or measures of testosterone (Table S10). The driver of clonality in cases with LOY and unknown driver CH remains unclear but we found that the degree of LOY was strongly predicted by the VAF of the somatic variants used to define CH ( $P = 3.72 \times 10^{-12}$ ). This suggests that LOY might itself be a driver of clonality, as has been postulated recently from whole genome sequence data of single cell-derived hematopoietic cells colonies,<sup>74</sup> and that LOY therefore accounts for an appreciable proportion of unknown driver CH.

In summary, we conclude that (i) SHBG is associated with LOY, but this relationship cannot be explained by the free-hormone pathway; (ii) SHBG has a causal effect on LOY, but LOY has no effect on SHBG; (iii) expression of imprinted genes in the *DLK1-MEG3* locus alters the relationship between LOY and SHBG, but has no significant effect on other sex hormones; (iv) CH does not explain the relationship between LOY and SHBG, but confirms that LOY is clonal.

## Supplemental data

Supplemental Data includes 12 tables

## Declaration of interests

The authors declare no competing interests

## Acknowledgments

AAZD was supported by a Lady Tata International Award. The authors acknowledge the use of the IRIDIS High Performance Computing Facility and associated support services at the University of Southampton.

## Author contributions

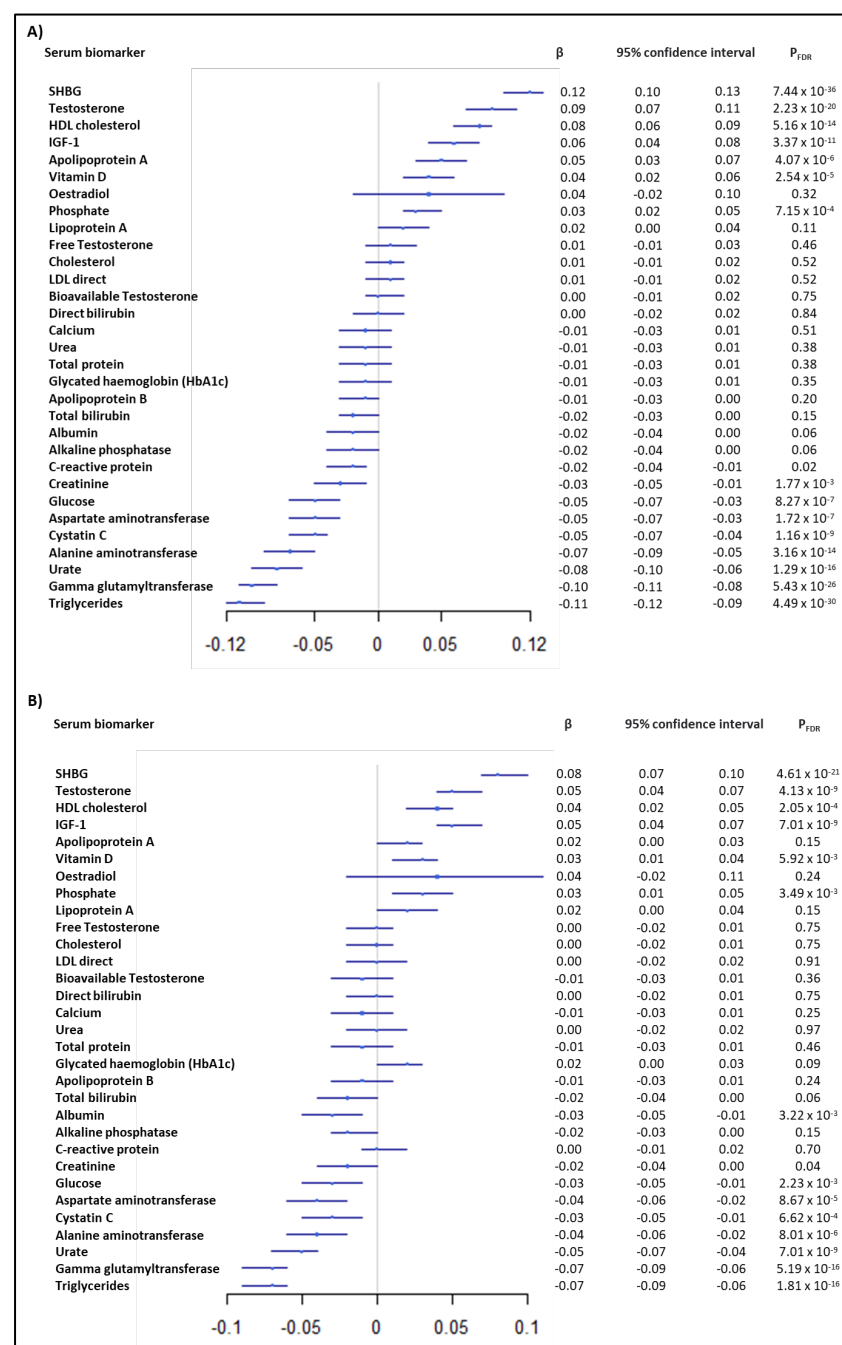
The study was designed by AAZD, WJT and NCPC. AAZD performed the data analysis in conjunction with WJT. All three authors refined the analysis and contributed to writing the paper.

## Web resources

Access to the UK Biobank Resource is available by application (<http://www.ukbiobank.ac.uk/>). eQTLGen database is available from (<https://www.eqtlgen.org>). PLINK (v1.9) is available from (<https://www.cog-genomics.org/plink/1.9/>). TwoSampleMR guide is available from (<https://mrcieu.github.io/TwoSampleMR/>). The code used for calculating FT and BAT is available from (<https://labrtorian.com/tag/free-testosterone/>). Further details regarding the biochemical assay methods and external quality assurance are available at [https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/serum\\_biochemistry.pdf](https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/serum_biochemistry.pdf).

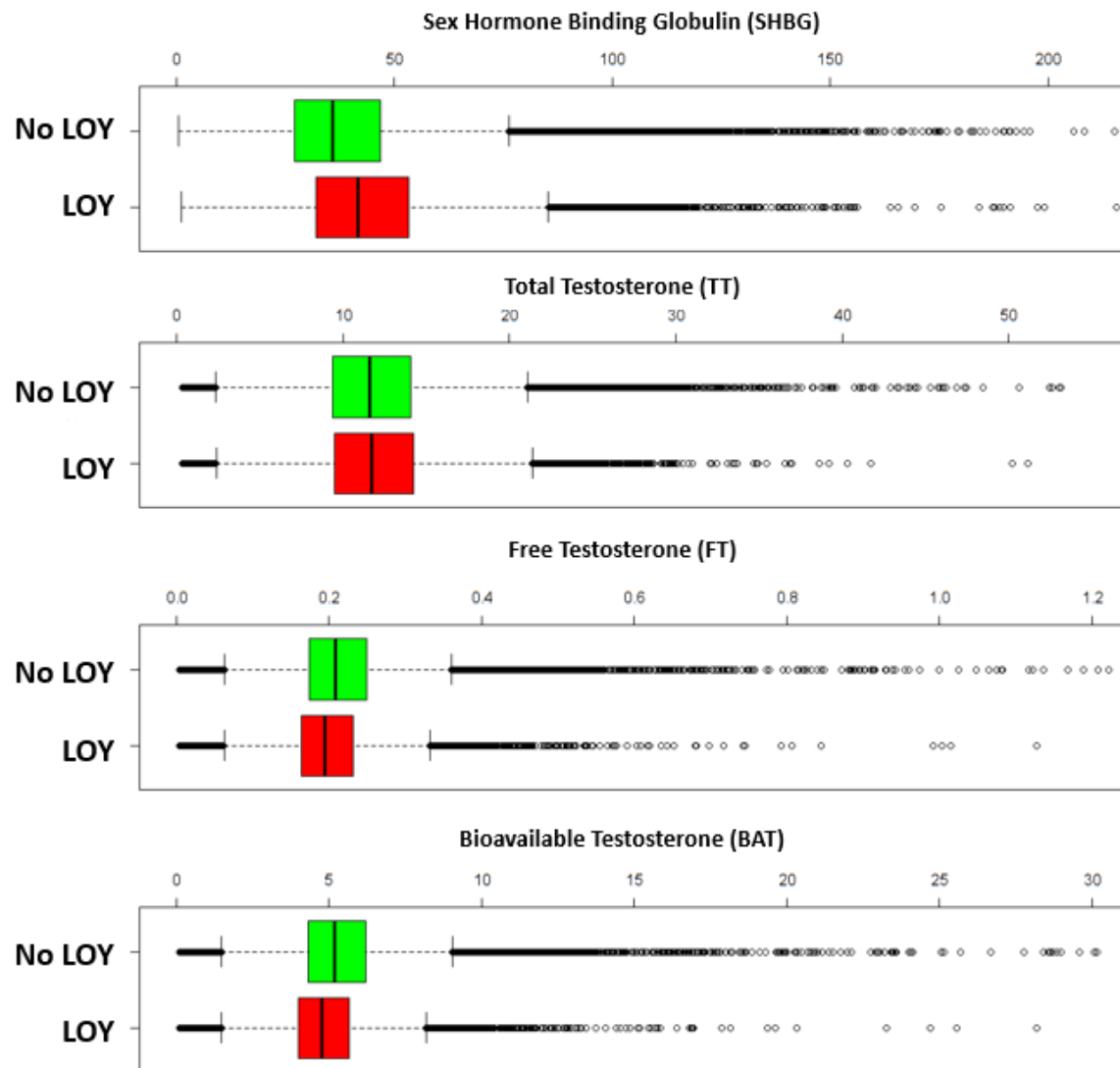


## Figures and legends



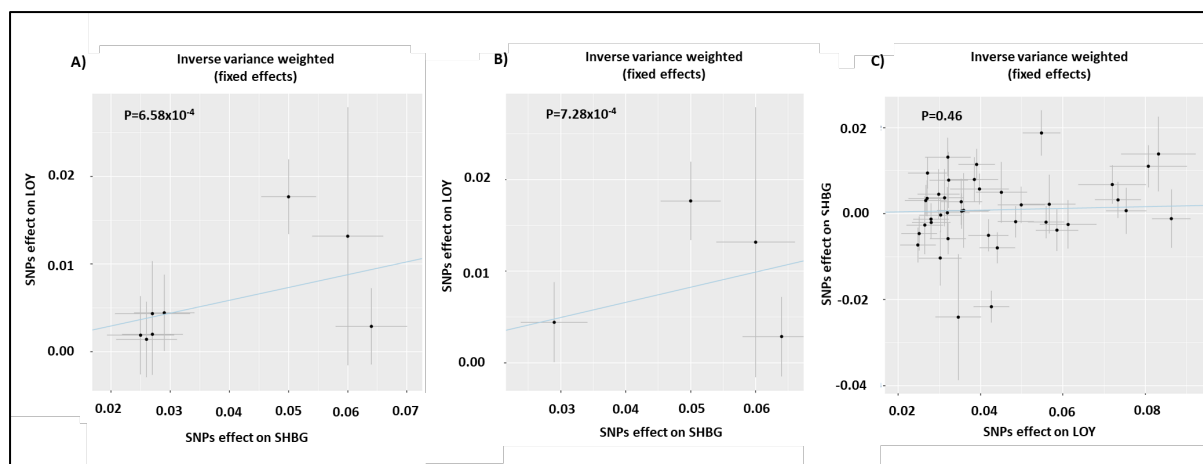
**Figure 1: The relationship between LOY and biochemistry markers**

The relationship between LOY and each of 31 biomarkers (29 measured and 2 calculated) was tested using multivariable linear regression in R in 222,835 UKB males. A) basic models: each linear model considered age, age squared, smoking status, smoking intensity 10PCA and multiple testing; B) sensitivity model: hypertension, insulin dependent diabetes, non-insulin dependent diabetes, and BMI were added to each basic model.



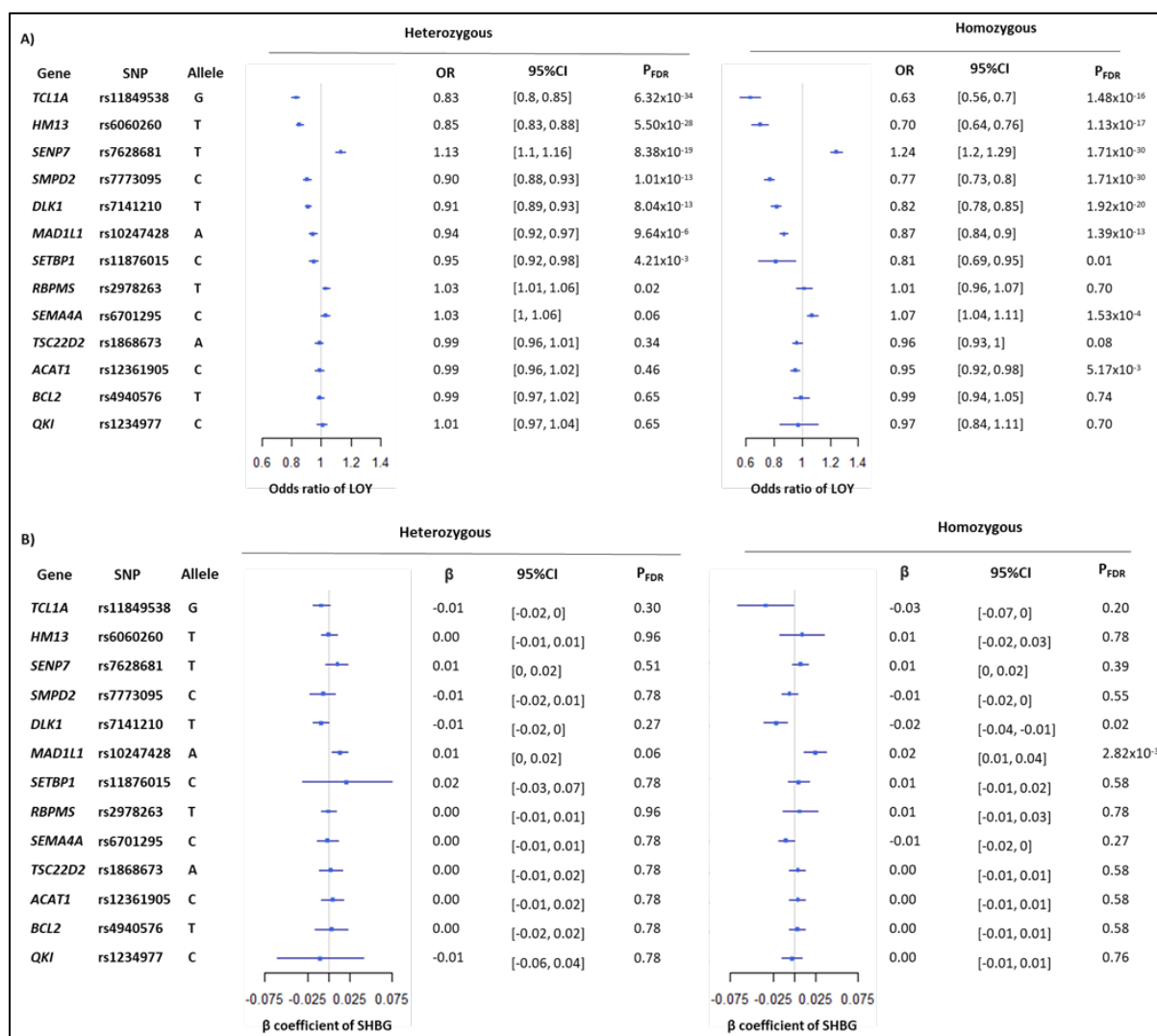
**Figure 2. The relationship between LOY and levels of sex hormones**

The box plots summarize serum sex hormone measurements in participants without LOY ( $n=178,277$ ) and with LOY ( $n=4,458$ ). SHBG: median nmol/L = 41.54 vs 35.86,  $P < 0.001$ ; TT: median nmol/L = 11.74 vs 11.58,  $P < 0.001$ ; FT: median nmol/L = 0.19 vs 0.20,  $P < 0.001$ ; BAT: median nmol/L = 4.78 vs 5.18,  $P < 0.001$ .



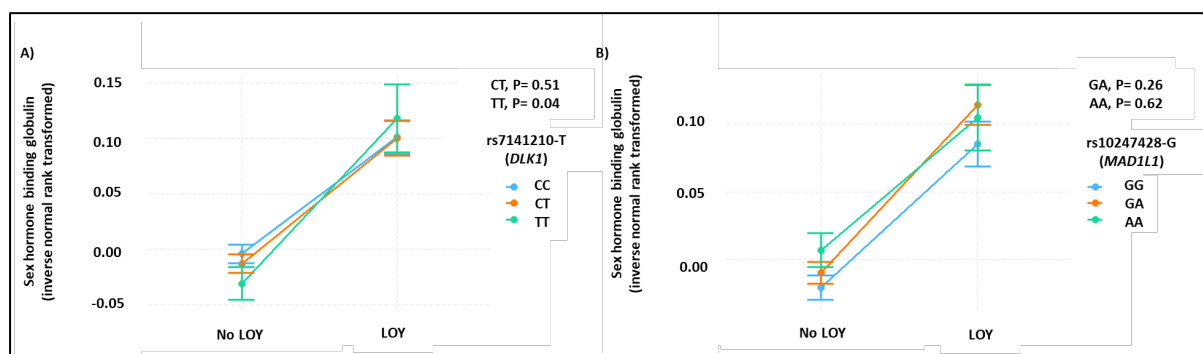
**Figure 3. Mendelian randomisation using an inverse variance weighted model to estimate the causal relationship between SHBG and LOY**

(A) Liberal analysis using 8 independent SNPs associated with SHBG. The IVW test estimated a significant positive effect of SHBG on LOY ( $P = 6.58 \times 10^{-4}$ ). (B) Conservative analysis using 4 SNPs ( $P = 7.28 \times 10^{-4}$ ). (C) analysis using 40 independent SNPs associated with LOY. The IVW test estimated no effect of LOY on SHBG ( $P = 0.46$ ). The line of regression is indicated in blue, and the axes show  $\beta$  coefficients for SNP effects on SHBG and LOY.



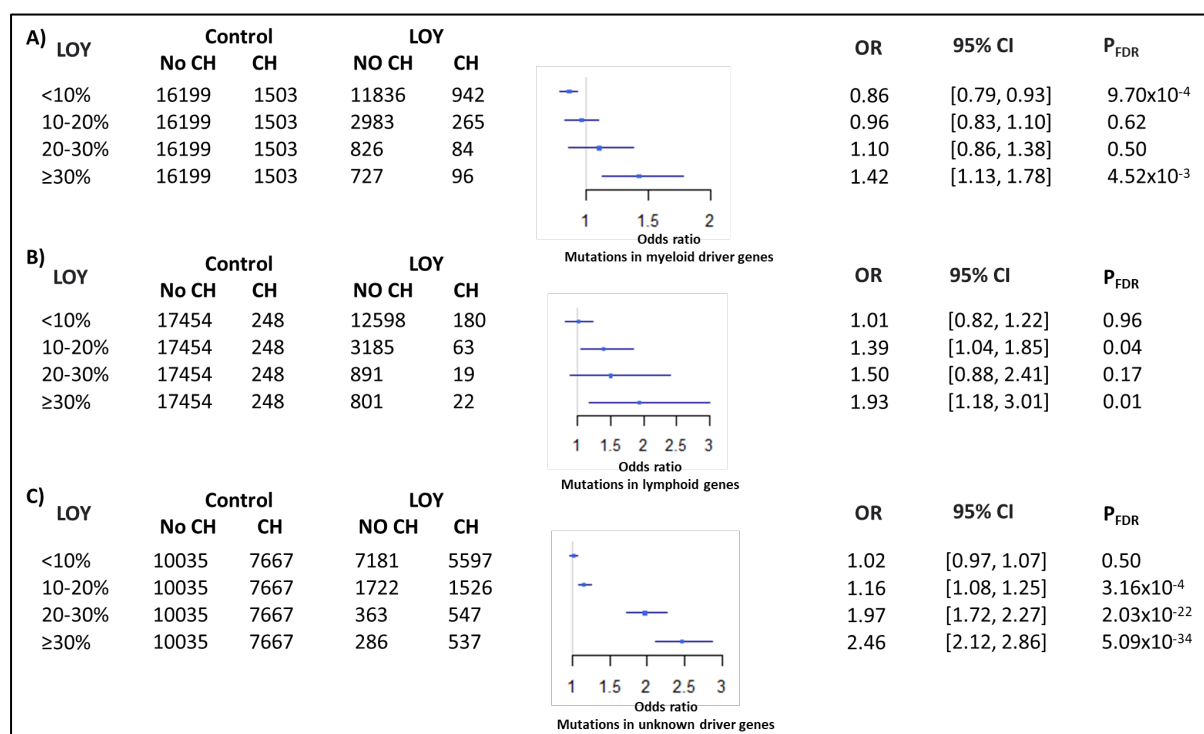
**Figure 4. The relationship between the predicted expression of 13 genes and each of LOY and SHBG.**

eQTL SNPs were used as proxies for gene expression and as assessed as predictors for LOY (panel A) and SHBG levels (panel B) incorporating age, smoking status, and 10 PCA as covariates.



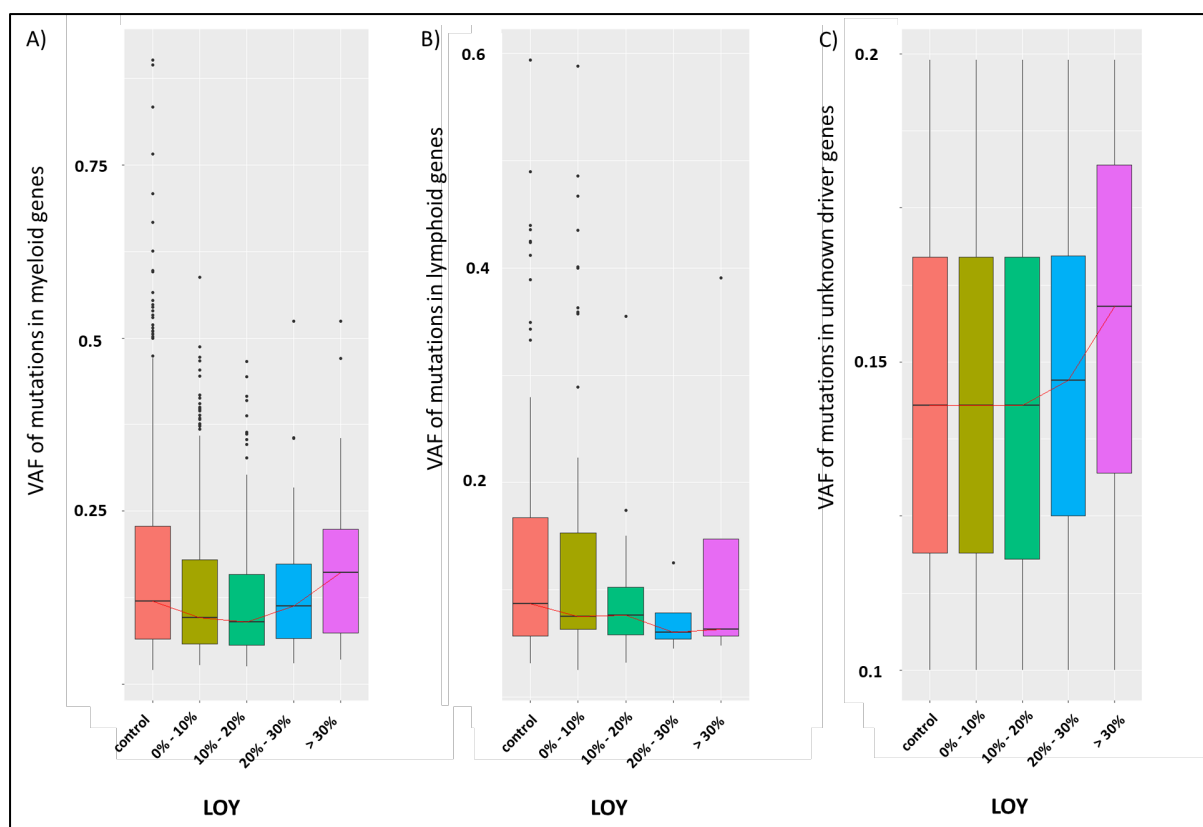
**Figure 5. The interaction between genetically predicted expression of *DLK1-MEG3* and *MAD1L1*, according to LOY status and SHBG levels**

Summary of SHBG values for men with and without LOY for (A) rs7141210 as a proxy for *DLK1-MEG3* expression and (B) rs10247428 as a proxy for *MAD1L1* expression. The interaction effect of each eQTL and LOY on SHBG was assessed by linear regression and was significant for rs7141210-T/T (green; left panel) in comparison to (C/C, blue). No significant difference was seen between rs10247428 genotypes.



**Figure 6. The relationship between LOY and CH**

LOY was stratified according to the clonal size (<10%, 10%-20%, 20%-30%, and ≥30%) and the proportion of participants with CH within each group was compared with controls. (A) myeloid CH, (B) lymphoid CH, (C) unknown driver CH.



**Figure 7. The relationship between LOY clonal size and CH VAFs**

Boxplots summarizing the distribution of VAFs of somatic mutations in controls and cases with LOY broken down by clone size. (A) myeloid CH, (B) lymphoid CH, (C) unknown driver CH. The red lines connect median values.

## References

1. Forsberg, L.A., Rasi, C., Malmqvist, N., Davies, H., Pasupulati, S., Pakalapati, G., Sandgren, J., de Ståhl, T.D., 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. *Nature Genetics* 46, 624-628. 10.1038/ng.2966.
2. Thompson, D.J., Genovese, G., Halvardson, J., Ulirsch, J.C., Wright, D.J., Terao, C., Davidsson, O.B., Day, F.R., Sulem, P., and Jiang, Y. (2019). Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* 575, 652-657.
3. Wright, D.J., Day, F.R., Kerrison, N.D., Zink, F., Cardona, A., Sulem, P., Thompson, D.J., Sigurjonsdottir, S., Gudbjartsson, D.F., Helgason, A., et al. (2017). Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility. *Nature Genetics* 49, 674-679. 10.1038/ng.3821.
4. Lin, S.H., Loftfield, E., Sampson, J.N., Zhou, W., Yeager, M., Freedman, N.D., Chanock, S.J., and Machiela, M.J. (2020). Mosaic chromosome Y loss is associated with alterations in blood cell counts in UK Biobank men. *Sci Rep* 10, 3655. 10.1038/s41598-020-59963-8.
5. Zhou, W., Machiela, M.J., Freedman, N.D., Rothman, N., Malats, N., Dagnall, C., Caporaso, N., Teras, L.T., Gaudet, M.M., Gapstur, S.M., et al. (2016). Mosaic loss of chromosome Y is associated with common variation near TCL1A. *Nature Genetics* 48, 563-568. 10.1038/ng.3545.
6. Silver, A.J., Bick, A.G., and Savona, M.R. (2021). Germline risk of clonal haematopoiesis. *Nature Reviews Genetics* 22, 603-617. 10.1038/s41576-021-00356-6.
7. Dumanski, J.P., Rasi, C., Lönn, M., Davies, H., Ingelsson, M., Giedraitis, V., Lannfelt, L., Magnusson, P.K., Lindgren, C.M., Morris, A.P., et al. (2015). Mutagenesis. Smoking is associated with mosaic loss of chromosome Y. *Science* 347, 81-83. 10.1126/science.1262092.
8. Noveski, P., Madjunkova, S., Sukarova Stefanovska, E., Matevska Geshkovska, N., Kuzmanovska, M., Dimovski, A., and Plaseska-Karanfilska, D. (2016). Loss of Y Chromosome in Peripheral Blood of Colorectal and Prostate Cancer Patients. *PLoS One* 11, e0146264. 10.1371/journal.pone.0146264.
9. Machiela, M.J., Dagnall, C.L., Pathak, A., Loud, J.T., Chanock, S.J., Greene, M.H., McGlynn, K.A., and Stewart, D.R. (2017). Mosaic chromosome Y loss and testicular germ cell tumor risk. *Journal of Human Genetics* 62, 637-640. 10.1038/jhg.2017.20.
10. Baliakas, P., and Forsberg, L.A. (2021). Chromosome Y loss and drivers of clonal hematopoiesis in myelodysplastic syndrome. *Haematologica* 106, 329-331. 10.3324/haematol.2020.266601.
11. Pérez-Jurado, L.A., Cáceres, A., Esko, T., López de Heredia, M., Quintela, I., Cruz, R., Lapunzina, P., Carracedo, Á., Group, S.C., and González, J.R. (2022). Clonal chromosomal mosaicism and loss of chromosome Y in men are risk factors for SARS-CoV-2 vulnerability in the elderly. *MedRxiv*, 2020.2004.2019.20071357. 10.1101/2020.04.19.20071357.
12. Dumanski, J.P., Lambert, J.C., Rasi, C., Giedraitis, V., Davies, H., Grenier-Boley, B., Lindgren, C.M., Campion, D., Dufouil, C., Pasquier, F., et al. (2016). Mosaic Loss of Chromosome Y in Blood Is Associated with Alzheimer Disease. *Am J Hum Genet* 98, 1208-1219. 10.1016/j.ajhg.2016.05.014.
13. Haitjema, S., Kofink, D., van Setten, J., van der Laan, S.W., Schoneveld, A.H., Eales, J., Tomaszewski, M., de Jager, S.C.A., Pasterkamp, G., Asselbergs, F.W., and den Ruijter, H.M. (2017). Loss of Y Chromosome in Blood Is Associated With Major Cardiovascular Events During Follow-Up in Men After Carotid Endarterectomy. *Circ Cardiovasc Genet* 10, e001544. 10.1161/circgenetics.116.001544.



14. 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. 10.1038/s41586-020-2426-2.
15. Greenberg, P.L., Tuechler, H., Schanz, J., Sanz, G., Garcia-Manero, G., Solé, F., Bennett, J.M., Bowen, D., Fenaux, P., Dreyfus, F., et al. (2012). Revised International Prognostic Scoring System for Myelodysplastic Syndromes. *Blood* 120, 2454-2465. 10.1182/blood-2012-03-420489.
16. Ouseph, M.M., Hasserjian, R.P., Dal Cin, P., Lovitch, S.B., Steensma, D.P., Nardi, V., and Weinberg, O.K. (2021). Genomic alterations in patients with somatic loss of the Y chromosome as the sole cytogenetic finding in bone marrow cells. *Haematologica* 106, 555-564. 10.3324/haematol.2019.240689.
17. Ljungström, V., Mattisson, J., Halvardson, J., Pandzic, T., Davies, H., Rychlicka-Buniowska, E., Danielsson, M., Lacaze, P., Cavelier, L., Dumanski, J.P., et al. (2022). Loss of Y and clonal hematopoiesis in blood—two sides of the same coin? *Leukemia* 36, 889-891. 10.1038/s41375-021-01456-2.
18. Dumanski, J.P., Halvardson, J., Davies, H., Rychlicka-Buniowska, E., Mattisson, J., Moghadam, B.T., Nagy, N., Weglarczyk, K., Bukowska-Strakova, K., Danielsson, M., et al. (2021). Immune cells lacking Y chromosome show dysregulation of autosomal gene expression. *Cell Mol Life Sci* 78, 4019-4033. 10.1007/s00018-021-03822-w.
19. Steensma, D.P., Bejar, R., Jaiswal, S., Lindsley, R.C., Sekeres, M.A., Hasserjian, R.P., and Ebert, B.L. (2015). Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood, The Journal of the American Society of Hematology* 126, 9-16.
20. Jacobs, K.B., Yeager, M., Zhou, W., Wacholder, S., Wang, Z., Rodriguez-Santiago, B., Hutchinson, A., Deng, X., Liu, C., and Horner, M.-J. (2012). Detectable clonal mosaicism and its relationship to aging and cancer. *Nature Genetics* 44, 651–658.
21. Laurie, C.C., Laurie, C.A., Rice, K., Doherty, K.F., Zelnick, L.R., McHugh, C.P., Ling, H., Hetrick, K.N., Pugh, E.W., and Amos, C. (2012). Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nature Genetics* 44, 642.
22. Miller, P.G., Qiao, D., Rojas-Quintero, J., Honigberg, M.C., Sperling, A.S., Gibson, C.J., Bick, A.G., Niroula, A., McConkey, M.E., Sandoval, B., et al. (2022). Association of clonal hematopoiesis with chronic obstructive pulmonary disease. *Blood* 139, 357-368. 10.1182/blood.2021013531.
23. Jaiswal, S., Natarajan, P., Silver, A.J., Gibson, C.J., Bick, A.G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., and Ardissino, D. (2017). Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *New England Journal of Medicine* 377, 111-121.
24. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burt, N., and Chavez, A. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. *New England Journal of Medicine* 371, 2488-2498.
25. Genovese, G., Köhler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., and Fromer, M. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *New England Journal of Medicine* 371, 2477-2487.
26. Zink, F., Stacey, S.N., Norddahl, G.L., Frigge, M.L., Magnusson, O.T., Jonsdottir, I., Thorgerisson, T.E., Sigurdsson, A., Gudjonsson, S.A., and Gudmundsson, J. (2017). Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* 130, 742-752.
27. Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., and Miller, C.A. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature Medicine* 20, 1472.

28. 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*. 10.1038/s41586-020-2819-2.
29. Bick, A.G., Popadin, K., Thorball, C.W., Uddin, M.M., Zanni, M.V., Yu, B., Cavassini, M., Rauch, A., Tarr, P., Schmid, P., et al. (2022). Increased prevalence of clonal hematopoiesis of indeterminate potential amongst people living with HIV. *Scientific Reports* 12, 577. 10.1038/s41598-021-04308-2.
30. Zekavat, S.M., Lin, S.-H., Bick, A.G., Liu, A., Paruchuri, K., Wang, C., Uddin, M.M., Ye, Y., Yu, Z., Liu, X., et al. (2021). Hematopoietic mosaic chromosomal alterations increase the risk for diverse types of infection. *Nature Medicine* 27, 1012-1024. 10.1038/s41591-021-01371-0.
31. Dawoud, A.A.Z., Gilbert, R.D., Tapper, W.J., and Cross, N.C.P. (2022). Clonal myelopoiesis promotes adverse outcomes in chronic kidney disease. *Leukemia* 36, 507-515. 10.1038/s41375-021-01382-3.
32. Buscarlet, M., Provost, S., Zada, Y.F., Barhdadi, A., Bourgoin, V., Lépine, G., Mollica, L., Szuber, N., Dubé, M.-P., and Busque, L. (2017). DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood* 130, 753-762.
33. Jaiswal, S. (2020). Clonal hematopoiesis and nonhematologic disorders. *Blood* 136, 1606-1614. 10.1182/blood.2019000989.
34. Niroula, A., Sekar, A., Murakami, M.A., Trinder, M., Agrawal, M., Wong, W.J., Bick, A.G., Uddin, M.M., Gibson, C.J., Griffin, G.K., et al. (2021). Distinction of lymphoid and myeloid clonal hematopoiesis. *Nature Medicine* 27, 1921-1927. 10.1038/s41591-021-01521-4.
35. Saiki, R., Momozawa, Y., Nannya, Y., Nakagawa, M.M., Ochi, Y., Yoshizato, T., Terao, C., Kuroda, Y., Shiraishi, Y., Chiba, K., et al. (2021). Combined landscape of single-nucleotide variants and copy number alterations in clonal hematopoiesis. *Nature Medicine* 27, 1239-1249. 10.1038/s41591-021-01411-9.
36. Zhao, Y., Stankovic, S., Koprulu, M., Wheeler, E., Day, F.R., Lango Allen, H., Kerrison, N.D., Pietzner, M., Loh, P.-R., Wareham, N.J., et al. (2021). GIGYF1 loss of function is associated with clonal mosaicism and adverse metabolic health. *Nature Communications* 12, 4178. 10.1038/s41467-021-24504-y.
37. Jones, A.V., Chase, A., Silver, R.T., Oscier, D., Zoi, K., Wang, Y.L., Cario, H., Pahl, H.L., Collins, A., and Reiter, A. (2009). JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nature Genetics* 41, 446-449.
38. 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. 10.1038/s41586-020-2819-2.
39. Coombs, C.C., Zehir, A., Devlin, S.M., Kishtagari, A., Syed, A., Jonsson, P., Hyman, D.M., Solit, D.B., Robson, M.E., and Baselga, J. (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.
40. Dawoud, A.A.Z., Tapper, W.J., and Cross, N.C.P. (2020). Clonal myelopoiesis in the UK Biobank cohort: ASXL1 mutations are strongly associated with smoking. *Leukemia* 34, 2660-2672. 10.1038/s41375-020-0896-8.
41. Gray, A., Feldman, H.A., McKinlay, J.B., and Longcope, C. (1991). Age, Disease, and Changing Sex Hormone Levels in Middle-Aged Men: Results of the Massachusetts Male Aging Study\*. *The Journal of Clinical Endocrinology & Metabolism* 73, 1016-1025. 10.1210/jcem-73-5-1016.
42. Dunn, J.F., Nisula, B.C., and Rodbard, D. (1981). Transport of Steroid Hormones: Binding of 21 Endogenous Steroids to Both Testosterone-Binding Globulin and Corticosteroid-Binding

- Globulin in Human Plasma. *The Journal of Clinical Endocrinology & Metabolism* 53, 58-68. 10.1210/jcem-53-1-58.
43. Wheeler, M.J. (1995). The Determination of Bio-Available Testosterone. *Annals of Clinical Biochemistry* 32, 345-357. 10.1177/000456329503200401.
44. Vermeulen, A., StoïCa, T., and Verdonck, L. (1971). The Apparent Free Testosterone Concentration, An Index of Androgenicity. *The Journal of Clinical Endocrinology & Metabolism* 33, 759-767. 10.1210/jcem-33-5-759.
45. Manni, A., Pardridge, W.M., Cefalu, W., Nisula, B.C., Bardin, C.W., Santner, S.J., and Santen, R.J. (1985). Bioavailability of albumin-bound testosterone. *J Clin Endocrinol Metab* 61, 705-710. 10.1210/jcem-61-4-705.
46. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., and O'Connell, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203-209.
47. Loh, P.-R., Genovese, G., Handsaker, R.E., Finucane, H.K., Reshef, Y.A., Palamara, P.F., Birmann, B.M., Talkowski, M.E., Bakhoum, S.F., and McCarroll, S.A. (2018). Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* 559, 350.
48. Vermeulen, A., Verdonck, L., and Kaufman, J.M. (1999). A Critical Evaluation of Simple Methods for the Estimation of Free Testosterone in Serum. *The Journal of Clinical Endocrinology & Metabolism* 84, 3666-3672. 10.1210/jcem.84.10.6079.
49. Kim, W., Moll, M., Qiao, D., Hobbs, B.D., Shrine, N., Sakornsakolpat, P., Tobin, M.D., Dudbridge, F., Wain, L.V., Ladd-Acosta, C., et al. (2021). Interaction of Cigarette Smoking and Polygenic Risk Score on Reduced Lung Function. *JAMA Network Open* 4, e2139525-e2139525. 10.1001/jamanetworkopen.2021.39525.
50. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
51. Coviello, A.D., Haring, R., Wellons, M., Vaidya, D., Lehtimäki, T., Keildson, S., Lunetta, K.L., He, C., Fornage, M., Lagou, V., et al. (2012). A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation. *PLoS Genet* 8, e1002805. 10.1371/journal.pgen.1002805.
52. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81, 559-575.
53. Smith, G.D., Davies, N.M., Dimou, N., Egger, M., Gallo, V., Golub, R., Higgins, J.P., Langenberg, C., Loder, E.W., and Richards, J.B. (2019). STROBE-MR: guidelines for strengthening the reporting of Mendelian randomization studies. *PeerJ Preprints*. 2167-9843.
54. Terao, C., Momozawa, Y., Ishigaki, K., Kawakami, E., Akiyama, M., Loh, P.-R., Genovese, G., Sugishita, H., Ohta, T., Hirata, M., et al. (2019). GWAS of mosaic loss of chromosome Y highlights genetic effects on blood cell differentiation. *Nature Communications* 10, 4719. 10.1038/s41467-019-12705-5.
55. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., and Langdon, R. (2018). The MR-Base platform supports systematic causal inference across the human phenotype. *Elife* 7, e34408.
56. Zhang, Z. (2017). Propensity score method: a non-parametric technique to reduce model dependence. *Ann Transl Med* 5, 7. 10.21037/atm.2016.08.57.
57. 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. *Nature Biotechnology* 31, 213-219. 10.1038/nbt.2514.

58. Dawoud, A.A.Z., Gilbert, R.D., Tapper, W.J., and Cross, N.C.P. (2021). Clonal myelopoiesis promotes adverse outcomes in chronic kidney disease. *Leukemia*. 10.1038/s41375-021-01382-3.
59. Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nature Reviews Cancer* 18, 696-705. 10.1038/s41568-018-0060-1.
60. Vösa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nature Genetics* 53, 1300-1310. 10.1038/s41588-021-00913-z.
61. Long, J.A. (2019). Interactions: Comprehensive, user-friendly toolkit for probing interactions. R package version 1.1. 0.
62. Emadi-Konjin, P., Bain, J., and Bromberg, I.L. (2003). Evaluation of an algorithm for calculation of serum "Bioavailable" testosterone (BAT). *Clinical Biochemistry* 36, 591-596. [https://doi.org/10.1016/S0009-9120\(03\)00076-6](https://doi.org/10.1016/S0009-9120(03)00076-6).
63. Burgess, S., and Thompson, S.G. (2013). Use of allele scores as instrumental variables for Mendelian randomization. *International Journal of Epidemiology* 42, 1134-1144. 10.1093/ije/dyt093.
64. Neale, R. (2018). UK Biobank GWAS round 2 [<http://www.nealelab.is/uk-biobank/>]. 01/08/2018 ed. Neale Lab.
65. Mendel, C.M. (1989). The Free Hormone Hypothesis: A Physiologically Based Mathematical Model\*. *Endocrine Reviews* 10, 232-274. 10.1210/edrv-10-3-232.
66. Hammes, A., Andreassen, T.K., Spoelgen, R., Raila, J., Hubner, N., Schulz, H., Metzger, J., Schweigert, F.J., Luppa, P.B., Nykjaer, A., and Willnow, T.E. (2005). Role of Endocytosis in Cellular Uptake of Sex Steroids. *Cell* 122, 751-762. <https://doi.org/10.1016/j.cell.2005.06.032>.
67. Balogh, A., Karpai, E., Schneider, A.E., Hetey, S., Szilagyi, A., Juhasz, K., Laszlo, G., Hupuczi, P., Zavodszky, P., Papp, Z., et al. (2019). Sex hormone-binding globulin provides a novel entry pathway for estradiol and influences subsequent signaling in lymphocytes via membrane receptor. *Scientific Reports* 9, 4. 10.1038/s41598-018-36882-3.
68. Rocha, S.T.d., Edwards, C.A., Ito, M., Ogata, T., and Ferguson-Smith, A.C. (2008). Genomic imprinting at the mammalian Dlk1-Dio3 domain. *Trends in Genetics* 24, 306-316. <https://doi.org/10.1016/j.tig.2008.03.011>.
69. Chase, A., Leung, W., Tapper, W., Jones, A.V., Knoop, L., Rasi, C., Forsberg, L.A., Guglielmelli, P., Zoi, K., and Hall, V. (2015). Profound parental bias associated with chromosome 14 acquired uniparental disomy indicates targeting of an imprinted locus. *Leukemia* 29, 2069-2074.
70. Loh, P.-R., Genovese, G., and McCarroll, S.A. (2020). Monogenic and polygenic inheritance become instruments for clonal selection. *Nature* 584, 136-141. 10.1038/s41586-020-2430-6.
71. Kessler, M.D., Damask, A., O'Keefe, S., Van Meter, M., Banerjee, N., Semrau, S., Li, D., Watanabe, K., Horowitz, J., Houvras, Y., et al. (2022). Exome sequencing of 628,388 individuals identifies common and rare variant associations with clonal hematopoiesis phenotypes. *MedRxiv*, 2021.2012.2029.21268342. 10.1101/2021.12.29.21268342.
72. Zink, F., Magnusdottir, D.N., Magnusson, O.T., Walker, N.J., Morris, T.J., Sigurdsson, A., Halldorsson, G.H., Gudjonsson, S.A., Melsted, P., Ingimundardottir, H., et al. (2018). Insights into imprinting from parent-of-origin phased methylomes and transcriptomes. *Nature Genetics* 50, 1542-1552. 10.1038/s41588-018-0232-7.
73. Buscarlet, M., Provost, S., Zada, Y.F., Barhdadi, A., Bourgoin, V., Lépine, G., Mollica, L., Szuber, N., Dubé, M.-P., and Busque, L. (2017). DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood, The Journal of the American Society of Hematology* 130, 753-762.

74. Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K.J., Mende, N., Calderbank, E.F., Jung, H., Mitchell, T., Coorens, T.H.H., Spencer, D.H., et al. (2022). Clonal dynamics of haematopoiesis across the human lifespan. *Nature* *606*, 343-350. 10.1038/s41586-022-04786-y.