1

Age-related loss of chromosome Y is associated with levels of sex hormone binding globulin and clonal hematopoiesis defined by *TET2*, *TP53*, and *CBL* mutations

Short title: Loss of Y chromosome and sex hormone binding globulin

Ahmed A.Z. Dawoud,[1,2] William J Tapper,[1]* & Nicholas C.P. Cross[1,2]*

[1] School of Medicine, University of Southampton, Southampton, Hampshire, SO17 1BJ, UK

[2] 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

29   **Abstract**

30   Mosaic loss of the Y-chromosome (LOY) in peripheral blood leukocytes is the most common somatic

31   alteration in men and linked to wide range of malignant and non-malignant conditions. LOY is

32   associated with age, smoking and constitutional genetics. Here, we aimed to assess the relationships

33   between LOY, serum biomarkers and clonal hematopoiesis (CH). LOY in UK Biobank was strongly

34   associated with levels of sex hormone binding globulin (SHBG), a key regulator of testosterone

35   bioavailability. Mendelian randomisation suggested a causal effect of SHBG on LOY but there was no

36   evidence for an effect of LOY on SHBG. In contrast, age-related CH defined by somatic driver mutations

37   was not associated with SHBG but was associated with LOY at clonal fractions above 30%. *TET2*, *TP53*,

38   and *CBL* mutations were enriched in LOY cases, but *JAK2* V617F was depleted. Our findings thus

39   identify independent relationships between LOY, sex hormone levels, and CH.

40

41   **Introduction**

42   Age-related somatic loss of the Y-chromosome (LOY) in peripheral blood leukocytes is the most

43   prevalent chromosomal alteration in men (*1*). Mosaic LOY, has been identified in as many as 20% of

44   male participants in large population-based cohort studies such as the UK Biobank (UKB; median age

45   = 58), but only 10% of affected individuals (2% of all men) had LOY involving >20% of leukocytes (*2-4*).

46   LOY is associated with all-cause mortality, cancer mortality (*1, 5, 6*), and a wide range of non-malignant

47   conditions (*1, 7-9*). LOY is also associated with variation in blood cell counts (*4, 10*), and has long been

48   recognized as a recurrent clonal cytogenetic finding in hematological malignancies where, in the

49   absence of other changes, it is associated with a good prognosis (*11*). Whether LOY is a direct driver

50   of clonality or merely a passenger event has not been established, but it has been suggested that LOY

51   might be a broad marker of genomic instability across different tissues, or that it may exert its effects

52   by altering immune cell function (*2, 12*).

53

54   Age-related clonal hematopoiesis (CH) is defined in most studies by the finding of specific somatically

55   acquired autosomal chromosomal alterations (*13, 14*), or a limited range of pathogenic driver

56   mutations (*15-20*) in the absence of a hematological neoplasm. The extent to which mosaic LOY, which

57   is much more common than mosaic autosomal alterations, overlaps with CH as defined above has

58   been explored (*18, 21*), but remains incompletely understood. Both LOY and CH are associated with a

59   wide range of malignant and non-malignant diseases (*15, 16, 22-26*). Most prominently, individuals

60   with CH have a significant risk (hazard ratios of 10-12.9) of developing a hematological malignancy

61   (*13, 14, 16, 17*), and recent studies showed a lineage-specific risk for mutations in genes associated

62   with myeloid or lymphoid neoplasms (*27, 28*). LOY with large clone size has been linked to the

63   presence of somatic mutations associated with hematological malignancies and an elevated risk of

64   developing myeloid neoplasia in two recent, small studies of selected cases (*21, 29*). Although age is

65   the major risk factor for the development of both CH and LOY, it has become clear that there is a

66   substantial overlap in constitutional genetic variation that predisposes to these abnormalities.

67   Broadly, variants in cancer susceptibility genes and genes that are mutated in cancer feature

68   prominently as risk factors for both CH and LOY (*2, 3, 20, 30, 31*). As for external factors, smoking is

69   associated with both CH and LOY (*1, 3, 32, 33*) indicating that the environment as well as genetics is

70   an important factor.

71

72   Serum biochemical profiles are also known to change with age, for example ageing is associated with

73   depletion of sex hormones (*34*), and some of these changes have been linked to common age-related

74   disorders. However, any link between changes in serum markers and LOY or CH remains unexplored.

75   In this study we aimed therefore to assess the relationships between LOY, CH and serum biomarkers

76   and investigate the role of constitutional genetics in any associations.

77

78   **Results**

79   *The relationship between LOY and biochemistry markers*

80   To investigate the relationship between LOY and serum biomarkers, we used previously published

81   calls of LOY that were generated by utilizing long-range phasing information to analyse allele-specific

82   genotyping intensities of 1,239 variants in the pseudo-autosomal region 1 (*2*). We restricted our

83   analysis to the 222,835 males who passed QC, of whom 44,558 (20%) had LOY. Of these, the majority

84   (n=31,952; 72%), had an estimated LOY clonal fraction of <10%. We compared the presence or

85   absence of LOY with 29 biochemistry parameters that were directly measured by UKB, as well as

86   estimated levels of free testosterone (FT; median = 0.21 nmol/L; range = 0.003 - 1.93) and bioavailable

87   testosterone (BAT; median = 5.1 nmol/L; range = 0.09 - 45.68) derived from measurements of sex

88   hormone binding globulin (SHBG), total testosterone (TT) and albumin (*35*).

89

90   Univariate comparisons revealed that participants with LOY had higher median levels of alkaline

91   phosphatase, apolipoprotein A, C-reactive protein, cystatin C, glucose, glycated hemoglobin (HbA1c),

92   HDL cholesterol, TT, urea, SHBG and vitamin D. Lower median levels in participants with LOY were

93   seen for alanine aminotransferase, albumin, apolipoprotein B, aspartate aminotransferase, calcium,

94   cholesterol, gamma glutamyltransferase, IGF-1, LDL direct, total bilirubin, total protein, triglycerides,

95   urate and FT (Table S1).

96

97   On initial multivariate analysis adjusted for age, age squared, smoking status, smoking intensity, the

98   first 10 genetic principal components (10 PC) and using the false discovery rate (FDR) to correct for

99   multiple testing, we found that LOY as a binary predictor was most strongly associated with elevated

100   levels of SHBG (Figure 1; β = 0.12, 95% CI: 0.11 - 0.13, P = 7.44x10^{-36}). We therefore focused our

101   subsequent analysis on SHBG, although we note that there are several other potentially important

102   positive and negative associations (Figure 1).

103

104   SHBG is a glycoprotein that binds to steroids, particularly testosterone, with high affinity (*36*). It is

105   notable therefore that the second strongest positive association was with TT (β = 0.09, 95% CI: 0.07 -

106   0.11, P = 2.23 x 10^{-20}). In men, circulating testosterone levels are regulated by SHBG, with on average

107   58% of TT bound to SHBG, 40% bound to albumin, and 2% present as FT (*37, 38*). Binding to albumin

108   is weak and so all non-SHBG-bound testosterone is considered as BAT (*39*). Furthermore, ageing is

109   associated with a decline in FT and BAT, and an increase in SHBG (*40, 41*). There was no association,

110   however, between LOY and FT (P = 0.46) or BAT (P = 0.75) (Figure 1A). Participants with LOY had higher

111   levels of SHBG and TT (SHBG: median nmol/L = 41.54 vs 35.86, P < 0.001; TT: median nmol/L = 11.74

112   vs 11.58, P < 0.001; Mann-Whitney U tests) but lower levels of FT (median nmol/L = 0.19 vs 0.20, P <

113   0.001), and BAT (median nmol/L = 4.78 vs 5.18, P < 0.001). On sensitivity analysis with the addition of

114   hypertension, insulin dependent diabetes, non-insulin dependent diabetes, and body mass index

115   (BMI) to the co-variates above, LOY retained the positive association with SHBG (β = 0.08, 95% CI: 0.07

116   − 0.10, P = 4.61x10^{-21}) and TT (β = 0.05, 95% CI: 0.04 − 0.07, P = 4.13 x 10^{-9}; Figure 1B, and Table S2).

117   Our observational results point to a direct relationship between levels of SHBG and LOY that cannot

118   be explained by age, age squared, smoking history or intensity, population stratification, common

119   morbidities, BMI or free/bioavailable testosterone (Figure 2).

120

121   To assess the possibility that the association between LOY and sex hormones might be due to

122   pleiotropic effects we included a polygenic risk score (PRS) consisting of 156 genome wide significant

123   SNPs (*2*) as a co-variate to account for genetic factors that predispose to LOY. When this score was

124   added as an independent variable in the sensitivity analysis (Table S3), LOY retained the positive

125   association with both SHBG (β = 0.08, 95% CI: 0.06 − 0.10, P = 8.09x10^{-18}) and TT (β = 0.06, 95% CI: 0.04

126   − 0.08, P = 1.34 x 10^{-9}) but the PRS was not significantly associated with SHBG (P = 0.63) or TT (P =

127   0.16). Thus, the association between LOY and SHBG and TT is not explained by established genetic risk

128   factors for LOY.

129

130

131   *The relationship between genetically defined SHBG and LOY*

132   Published GWAS have identified multiple genetic determinants of SHBG levels in serum of both men

133   and women (*42*). To understand the relationship between LOY and SHBG we used an 11 SNP PRS (of

134   which 8 SNPs were associated with SHBG at a level of genome wide significance and 3 were

135   identified by conditional analysis of the *SHBG* gene) to summarize the genetic variation associated

136   with SHBG and evaluated the score as a predictor of LOY. Since the score was derived from

137   independent cohorts it represents an unbiased instrument to assess the relationship with LOY in

138   UKB (*43*). We found that genetically predicted SHBG was significantly associated (OR = 1.02, 95% CI:

139   1.01 - 1.04, P = 5.59×10^{-5}) with the finding of LOY in UKB.

140

141   To assess the possibility of a causal relationship between SHBG and LOY we performed a Mendelian

142   randomisation (MR) analysis. Of the 11 SNPs used for the SHBG PRS, 3 were unavailable in Biobank

143   Japan (BBJ) and thus 8 were used to estimate the effect of SHBG on LOY in male BBJ participants (*44*)

144   (Figure 3, Tables S4 and S5). Using a standard inverse variance weighted (IVW) model with fixed

145   effects we identified a positive causal relationship (β = 0.15, 95% CI: 0.06 - 0.23, P = 6.58×10^{-4}), with

146   no evidence of heterogeneity (Q-test, P=0.26). Although not all alternative MR tests identified a

147   significant causal relationship (Table S5), effects were consistently seen in the same direction and

148   the non-significant EGGER intercept (P=0.8; Table S5) argues against the possibility of directional

149   horizontal pleiotropy. In a more conservative analysis restricted to the subset of four SNPs that were

150   strongly associated with SHBG in men (P < 5×10^{-8}), the effect of SHBG on LOY was confirmed (IVW, β

151   = 0.17, 95% CI: 0.07 - 0.26, P = 7.28×10^{-4}). Leave-one-out analysis (Table S6) found that significance

152   was lost when rs7910927 at 10q21.3 within *JMJD1C* was excluded (initial analysis, n=7 SNPs: β =

153   0.08, 95% CI: -0.02 - 0.17, P = 0.13; conservative analysis, n=3 SNPs: β = 0.07, 95% CI: -0.04 - 0.19, P =

154   0.22). rs7910927 was associated with LOY in UKB but did not reach genome wide significance (β =-

155   0.02, p=4.59×10^{-3}) and no association with LOY was seen in the vicinity of *JMJD1C* in BBJ, thus

156   arguing against the possibility of a pleiotropic effect of rs7910927 on both LOY and SHBG. To assess

157   the possibility of a bidirectional effect we utilized 40 SNPs with P < 5×10^{-8} associated with LOY in

158   Japanese men (*44*) to measure their effect on SHBG levels in UKB men (*45*), but no significant

159   relationship was found (β = 0.02, 95% CI: -0.01 - 0.05, P = 0.21).

160

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

162   To explore the hypothesis that germline variants associated with LOY might regulate genes that

163   could modify the association between LOY and SHBG, we focused on 19 genomic regions associated

164   with LOY in UKB (*3*) and replicated a larger UKB dataset as well as three independent cohorts (*2*).

165   Using the eQTLGene database (*46*), we identified eQTL SNPs to serve as valid proxies for 13 of the 19

166   regions (*ACAT1, BCL2, DLK1, HM13, MAD1L1, QKI, RBPMS, SEMA4A, SENP7, SENP8, SETBP1, SMPD2,*

167   *TCL1A,* and *TSC22D2*). We found that 8/13 eQTLs were associated with LOY in UKB but only 2 were

168   associated with levels of SHBG, albeit at borderline levels of significance (Figure 4; Table S7). The

169   eQTL allele associated with increased expression of *MAD1L1* at 7p22.3 was positively associated with

170   SHBG (rs10247428_A, β = 0.02, 95% CI = 0.00 - 0.03, P = 0.04, but this was in the opposite direction

171   to its relationship with LOY. The eQTL allele associated with reduced expression of *DLK1*, part of

172   *DLK1-MEG3* imprinted locus at 14q32.2 was negatively associated with SHBG (rs7141210-T, β = -

173   0.01, 95% CI = -0.03 – 0.00, P=0.05; Figure 4, Table S8) and also negatively associated with LOY.

174   rs7141210 and LOY showed no interactive effects on sex hormones levels, although for SHBG this

175   approached nominal significance (Table S9; P=0.07).

176

177   *The relationship between LOY, SHBG and other forms of CH*

178   To understand the impact of somatic mutations on the relationship between SHBG, testosterone

179   and LOY, we assessed the relationship between somatically acquired single nucleotide variants

180   (SNVs), sex hormones and LOY. Whole exome sequence (WES) data was available for 17,759

181   participants with LOY, of whom 28% (n=4,981) were estimated to have an LOY clone size ≥ 10% of

182   leukocytes. For comparison, we randomly selected UKB age matched male controls (n=17,702) who

183   were negative for LOY. We identified recurrent somatic SNVs in driver genes associated with myeloid

184   CH and lymphoid CH (Table S10), plus likely somatic SNVs in other genes. This latter group are

185   expected to be mostly passenger mutations that do not confer a selective advantage but indicate

186   clonality in the absence of known driver mutations, a situation we refer to as 'unknown driver CH'.

187   Overall, the frequency of each CH subtype (myeloid, lymphoid, unknown driver) was similar between

188   cases with LOY and controls. Striking differences emerged, however, when LOY was stratified by

189   clone size. All SNV-associated CH (myeloid plus lymphoid plus unknown driver) was significantly

190   associated with LOY in ≥ 10% of cells with a clear increase in the strength of the association with

191   increasing LOY clone size (Table S11; 10-20% LOY, OR = 1.17, P = 1.81×10^{-4}; 20-30% LOY, OR = 2.20, P

192   = 4.25×10^{-27}; ≥30% LOY, OR = 3.43, P = 2.42×10^{-52}). Similar results were seen for unknown driver CH

193   considered alone (10-20% LOY, OR=1.16, P= 3.16×10^{-4}; 20-30% LOY, OR=1.97, 2.03×10^{-22}; ≥30% LOY,

194   OR = 2.46, P = 5.09×10^{-34}). Of note, LOY clone size was also related to age. The mean age of cases

195   with LOY <10%, 10-20% and ≥30% was 61.7, 64.2 and 65.4 respectively (<10% vs 10-20%, $P=1.03 \times 10^{-10}$

196    $^{-10}$; 10-20 vs ≥30%, $P=7.4 \times 10^{-8}$) but the associations between CH and LOY clone size remained

197   significant when adjusting for age, age squared, smoking history and intensity, and 10 PC (Table

198   S12). By contrast, both myeloid CH (OR = 1.42, P = 4.52×10^{-3}), and lymphoid CH (OR = 1.93, P = 0.01)

199   were significantly associated with LOY in ≥30% of cells but not LOY of smaller clone size (Figure 5).

200   Our findings thus confirm and extend a previously reported association between LOY and unknown

201   driver CH (*17*). Notably, however, none of the three CH SNV-associated subtypes were associated

202   with SHBG or the three measures of testosterone, indicating no effect of driver mutations on the

203   relationship between LOY and sex hormones (Table S13).

204

205   To understand the relationship between CH and LOY in more detail, we assessed the association

206   between somatic mutations in specific driver genes in participants with LOY in ≥30% cells (high level

207   LOY; n=823) compared to LOY free controls (n=17,702). Focusing on genes (n=9) that had driver

208   mutations in ≥3 cases (Table S14), *TET2* was the most significantly enriched mutated gene in LOY

209   cases (4% versus 1.5% in controls, OR = 2.64, P = 9.58×10^{-5}) with *TP53* (OR = 6.96, P = 7.62×10^{-3}) and

210   *CBL* (OR = 7.43, P = 0.04) mutations also showing a significant enrichment (Table 1). These

211   relationships remained significant after excluding cases with a diagnosis of cancer at or prior to study

212   recruitment (n=2998; Table S15). We found no significant association between these three genes

213   and SHBG or TT, but lower levels of FT and BAT were noted in men with *TP53* mutations (Table S16).

214   The other 6 genes, including *DNMT3A* and *ASXL1*, showed no enrichment in high level LOY cases and

215   in fact the prevalence of *DNMT3A* mutations was lower in cases with high level LOY compared to

216   controls, although the difference was not significant (3.8% vs 2.4%; P=0.11). To explore the

217   possibility that some somatic abnormalities might be negatively associated with LOY, and thus

218   potentially missed by focusing on high level LOY and driver mutations in ≥3 cases, we compared

219   mutational status with LOY at any level. As shown on Table S17, *JAK2* V617F but no other gene/SNV

220   was negatively associated with LOY (OR=0.39, 95%CI=0.22-0.66, $P_{FDR}=6.78 \times 10^{-3}$).

221

222   The possibility that LOY and CH as defined by SNVs might co-exist in the same clone was assessed by

223   analysing the relationship between LOY B-allele frequencies (BAF) and the variant allele frequencies

224   (VAF) of driver mutations. Figure 6 shows a summary of the results at different ranges of LOY.

225   Myeloid CH VAFs were associated with BAF levels in samples with LOY ≥10% (β = 0.13, 95% CI = 0.05

226   - 0.20, P = 4.89×10^{-3}). Similar results were seen for unknown driver CH (β = 0.18, 95% CI = 0.10 - 0.26,

227   P = 4.67×10^{-5}), which by definition was restricted to VAFs of 0.1 - 0.2 (Table S18).

228

229   **Discussion**

230   Age-related mosaic LOY in peripheral blood leukocytes is known to be influenced by both genetic

231   and environmental factors. We have found that LOY is also strongly associated with multiple serum

232   biomarkers, most notably levels of SHBG, and that this association is independent of known

233 confounders (age, age squared, smoking history, common comorbidities, BMI and 10 PC).

234 Furthermore, we found genetic evidence to support the hypothesis that SHBG levels are causally

235 linked to LOY, but no evidence that LOY has any effect on SHBG. SHBG regulates the level of

236 circulating testosterone and, although we found that both FT and BAT were lower in men with LOY

237 compared to those without LOY, there was no significant relationship between LOY and either FT or

238 BAT on multivariate analysis. This finding is inconsistent with the free hormone hypothesis, which

239 proposes that only the unbound fraction of testosterone is biologically active in target tissues (*47*),

240 and instead suggests that other pathways may be involved. The mechanism by which SHBG

241 promotes LOY is unclear but from a genetic perspective the effect is not explained by variation at

242 *SHBG* alone. Other loci are involved, particularly *JMJD1C* which encodes a histone demethylase

243 previously linked to SHBG levels (*42*).

244

245 To understand the influence of genetic factors on the relationship between SHBG and LOY in more

246 detail, we focused on genetically predicted expression of genes linked to the development of LOY. We

247 identified one eQTL (rs7141210-T, associated with reduced expression of *DLK1*) that was negatively

248 associated with both LOY and SHBG. *DLK1* encodes a delta-like non-canonical notch ligand but this

249 gene is part of the large and complex *DLK1-MEG3* imprinted cluster of genes and non-coding RNAs.

250 The methylated paternally derived chromosome expresses the protein-coding genes *DLK1*, *RTL1* and

251 *DIO3*, while the non-methylated maternally derived chromosome expresses the non-coding genes

252 *MEG3*, *MEG8*, *asRTL1*, multiple miRNAs and lncRNAs (*48*). Constitutional uniparental disomy (UPD) at

253 14q32 is associated with the developmental disorders Temple syndrome (maternal UPD) and Kagami–

254 Ogata syndrome (paternal UPD) whereas somatically acquired paternal UPD is associated with CH and

255 myeloid malignancies (*49*). Genome-wide significant signals have been identified near *DLK1* in

256 association with CH defined by acquired 14q UPD (*50*), and somatic driver mutations (*51*), as well as

257 LOY (*2, 3, 44*). Collectively, these findings suggest that the impact of rs7141210-T on the relationship

258 between SHBG, CH and LOY merits further investigation.

259

260 CH is typically defined by pathogenic driver mutations associated with hematological malignancies,

261 most commonly in the epigenetic regulators *DNMT3A*, *TET2*, and *ASXL1*, which collectively account

262 for 90% of CH cases (*15-20*). Broad screens by WES or whole genome sequencing (WGS) have revealed

263 that clonality in the absence of known driver mutations (unknown driver CH) is even more prevalent

264 than CH with driver mutations (*18*). In this study we have defined the relationship between CH and

265 LOY. LOY in ≥30% of cells was associated with both myeloid and lymphoid CH, with 14% of affected

266 individuals having one or more somatic driver mutations compared to 10% of controls (P = 2.92 x 10⁻

267 [4]; Table S11). At the level of individual genes, the most striking finding was that mutated *TET2* was

268 associated with LOY, but not *DNMT3A* or *ASXL1*. Our findings are consistent with the notion that CH

269 with *TET2* mutations is different from CH with *DNMT3A* or *ASXL1* mutations as indicated by some

270 clinical studies, for example CH with *TET2* mutations has been linked to chronic obstructive pulmonary

271 diseases (*52*), but not CH with *DNMT3A* mutations.

272

273 With regard to the relationship between LOY and CH, our data confirm and extend previous

274 observations (*17, 18*). As detailed in Table S11, unknown driver CH was seen in 65% (537/823) of

275 cases with high level (≥30% of cells) LOY compared to just 43% (7667/17702) of controls (P = 5.09 x

276 10^{-34}). Overall, 80% (655/823) of cases with high level LOY had mutational evidence of clonality (i.e.

277 CH defined as myeloid, lymphoid or unknown), with LOY in ≥10% of cells clearly associated with

278 unknown driver CH. Our findings thus provide statistical evidence that LOY ≥10% co-exists with CH.

279 We found that the degree of LOY was associated with the VAF of the somatic variants used to define

280 CH (P = $3.72x10^{-12}$), however this finding does not clarify whether LOY might be a direct driver of

281 clonality, as has been postulated recently (*53*), or whether LOY is simply a passenger event

282 associated with as yet uncharacterized factors that promote clonality. Importantly for our study,

283 however, neither overall CH nor any CH subtype was associated with SHBG or measures of

284 testosterone (Table S13).

285

286 In summary, we conclude that (i) SHBG is associated with LOY, but this relationship cannot be

287 explained by the free-hormone hypothesis; (ii) SHBG has a likely causal effect on LOY, but LOY has no

288 effect on SHBG; (iii) CH does not explain the relationship between LOY and SHBG.

289

290 **Material and Methods**

291 *Study cohort*

292 UKB is a large prospective cohort described in detail elsewhere (*54*), involving approximately 500,000

293 individuals aged between 40 and 69 years at recruitment. Genome wide single nucleotide

294 polymorphism (SNP) data derived from peripheral blood leucocytes was available for most

295 participants, and WES data for 200,631 participants at the time of analysis. All participants provided

296 informed consent according to the Declaration of Helsinki, and UK Biobank received ethical approval

297 from the Northwest multi-centre Research Ethics Committee (REC reference 11/NW/0382). The

298 current study was conducted under approved UK Biobank application number 35273.

299

300

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) (*2*). 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% (*2*), but the majority of events are detected at cell fractions >5% (Figure 7). The spectrum of LOY was categorised according to clonal fraction by considering the median change of 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 (*2*).

Biochemistry markers and sex hormones

Measurements of 29 biochemistry markers were available from serum samples collected on recruitment to UKB (*54*). Mass action equations were used to calculate FT and BAT from measurements of SHBG, TT and albumin as described (*55*). 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.

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) (*2*). WES data was available for 40% of these individuals (n=89,811). 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 history (never, previous, current), smoking intensity defined by average number of pack-years smoked by an individual over their adult lifetime (Data-Field 20162) (*56*), and 10 PC. 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 BMI as a continuous variable. Effect sizes were reported as beta coefficients (β) with 95% confidence intervals (95% CI). P-values were adjusted for 31 tests using the False Discovery Rate (FDR) method. The average measure of serum biomarkers in participants with LOY were compared to those in LOY free controls using Mann–Whitney U tests.

PRS for LOY in UKB

156 genome wide significant variants were associated with LOY by a previous GWAS (*2*). A PRS was calculated as the sum of number of risk alleles per SNP weighted by their corresponding genetic effect size using the score function in Plink V1.9, and transformed into a normal distribution using inverse normal rank transformation (*57*). We added the PRS to the sensitivity analysis of each serum biomarker. Effect sizes for both LOY, and LOY score were reported as beta coefficients (β) with 95% CI (Table S3).

PRS for SHBG in UKB

Thirteen genetic variants were associated with circulating SHBG levels by a previous two-stage GWAS (*42*) of both men and women from multiple cohorts that did not include UKB or BBJ. The 13 variants included 10 that achieved genome wide significance plus 3 independent cis variants that were identified by conditional analysis of the *SHBG* gene. After excluding one SNP with heterogeneity towards females (P = 0.02, rs440837) and a second SNP on the X-chromosome (rs1573036), 11 SNPs (rs12150660, rs1641537, rs1625895, rs6258, rs17496332, rs2411984, rs293428, rs780093, rs7910927, rs8023580 and rs4149056) were used to calculate an SHBG PRS as described above. A multivariable logistic regression model adjusted for age, age squared, smoking history and intensity and 10 PC was used to assess the relationship between LOY status (binary, dependent) and the SHBG PRS in UKB.

Mendelian Randomisation

MR was used to assess the possibility of a causal relationship between SHBG and LOY using germline SNPs associated with circulating SHBG (*42*) as instrumental variables for LOY in BBJ, following the STROBE guidelines (*58*). Of the 11 SNPs described above for the SHBG PRS, three (rs12150660, rs6258 and rs2411984) were unavailable in BBJ. The remaining 8 SNPs (of which 6 were genome wide significant for SHBG and two were *cis* variants identified by conditional analysis of the *SHBG* gene) were thus used as instrumental variables without considering estimated genetic effect sizes for LOY in 95,380 BBJ men (*44*). The analysis was performed using the TwoSamplesMR package (*59*), with the IVW (fixed effects) model considered as the standard method. Other models were used for sensitivity analysis, and the MR-EGGER model as a method to account for horizontal pleiotropy. To explore the effect of SNP selection, we repeated the MR analysis using a genome wide significance threshold specifically in men (P < 5x10^{-8}) to select a subset of four SNPs (rs1641537, rs1625895, rs293428 and rs7910927) associated with SHBG (*42*). We also performed a leave-one-out analysis to evaluate the effect of each SNP on the analysis. To assess the effect of LOY on SHBG

369   levels we examined SNPs (n=50) previously associated with LOY (*44*) in BBJ. 10 SNPs were excluded

370   as genotypes were unavailable and/or non-informative in UKB. The remaining 40 SNPs were used as

371   instrumental variables without considering estimated genetic effect sizes for SHBG in UKB men.

372

373   *Identification of CH*

374   A propensity score matching method in R (*60*), was used to select one control per case and to match

375   for age in comparison to participants with LOY. Somatic mutations were called in LOY samples and

376   matched controls using GATK (Version 4.1.9) and Mutect2 (*61*), to process individual CRAM files in

377   the tumour-only mode. Following best practice guidelines (https://gatk.broadinstitute.org/hc/en-

378   us/articles/360035531132), a Panel Of Normal (PON; version 23rd August 2017) from the Broad

379   Institute that were generated using Mutect2 on samples from the 1000 genomes project to identify

380   recurrent artifacts, and germline variants from gnomAD were used to remove artefacts and germline

381   variants. To identify putative somatic driver mutations, the analysis was restricted to rare variants

382   with a minor allele frequency (MAF) <0.01 in gnomAD and a minimum number of reads supporting

383   the mutated allele: 3 reads for point mutations and 6 reads for indels. Other methodological details

384   for calling somatic variants are described in detail in our previous publication (*24*). Variants that

385   satisfied either of the following two criteria were selected: first; recurrent driver mutations as

386   defined in our previous study (*24*); second singleton variants that passed all Mutect2 filters with VAF

387   between 0.1 and 0.2 (*18*). Driver mutations were classified into myeloid or lymphoid according to a

388   published list of genes associated with myeloid neoplasms (n=76) (*15*) and genes associated with

389   acute lymphoblastic leukemia or chronic lymphocytic leukemia in the Cancer Gene Census (CGC)

390   (*62*), respectively (Table S1). Participants with variants in myeloid genes were considered as having

391   myeloid CH (n=2,890), and those with variants in lymphoid genes as having lymphoid CH (n=532).

392   Cases with mutations in genes involved in both myeloid and lymphoid neoplasms were considered

393   as myeloid CH. To identify participants with evidence of clonality in the absence of pathogenic

394   mutations in known driver genes, we also identified variants in all coding genes that were not

395   defined as myeloid or lymphoid. Participants with singleton variants in any gene with VAF range

396   between 0.1 and 0.2 not defined as myeloid or lymphoid were considered as having unknown driver

397   CH (n=15,874).

398

399   *The relationship between LOY and CH*

400   The association between LOY and all variants, driver variants, myeloid CH, lymphoid CH, and

401   unknown driver CH were tested using logistic regression in R. We further assessed the relationship

402   with LOY clone size categories (<10%, 10%-20%, 20%-30%, and ≥30%) using Fisher's exact tests. Age

403   was compared to LOY clone size using the Wilcoxon signed-rank test. The strength of the association

404   was reported as odds ratio (OR) with 95% CI. P-values were adjusted for 20 tests using the FDR

405   method. To test the relationship at the individual driver gene level, we initially focused on the 9

406   genes mutated in ≥3 males and cases with LOY in ≥30% of cells (Table 1), and then on genes mutated

407   in ≥10 cases+controls and cases with LOY at any level (Table S17). Self-reported cancer diagnosed by

408   doctor (UKB: Data-Field 2453) was used to identify participants with prevalent cancer (Table S15).

409

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

411   The co-existence of CH and LOY was tested by assessing the relationship between the BAF for LOY

412   and VAF of driver mutations. For cases with two or more mutations the highest VAF was used. To

413   avoid excess CH with VAF between 0.1-0.2, the analysis for myeloid and lymphoid CH was restricted

414   to VAFs detected for recurrent driver mutations as defined in our previous study (*24*). We assessed

415   the relationship with LOY in ≤10% and >10% of cells (Table S18), taking age, age squared, smoking

416   intensity and history, and 10 PC into account. The strength of the association was reported as β

417   coefficient with 95% CI. P-values were adjusted for 6 tests using the FDR method.

418

419   *The relationship between CH and sex hormones*

420   The association between CH and sex-hormone levels were tested using linear regression in R, with

421   normally transformed sex hormone as the independent variable. The dependent variables were

422   driver mutation state as a binary predictor, age, age squared, smoking status and intensity, 10 PC.

423   The association was reported as β coefficient with 95% CI. P-values were adjusted for 4 tests using

424   the FDR method.

425

426   *Expression quantitative trait analyses*

427   To examine the relationship between 19 SNPs associated with LOY (3) and SHBG levels we used the

428   eQTLGen database (55) to identify proxies associated with gene expression. Proxies were filtered for

429   cis-eQTLs within a distance <1 Mb and FDR < 0.05. The SNP with the smallest FDR value and no other

430   genes showing a stronger association were selected as proxies to minimize potential horizontal

431   pleiotropy. The analysis was restricted to directly genotyped SNPs with MAF >0.05 in UKB. The 19

432   SNPs were associated with 27 genes by position, biological function, expression, or nonsynonymous

433   variants in the gene (*3*). Our analysis was restricted to 13 of these genes for which an expression

434   proxy was identified.

435

436

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

438 The most significant eQTL SNP for each gene was encoded according to the risk genotype (0, 1, 2) in

439 an additive model. The following statistical tests were applied, and adjusted for age, age squared,

440 smoking history and intensity, 10 PC, and multiple tests by the FDR method. First, the relationship

441 between each eQTL and LOY was assessed using logistic regression in R where LOY (binary) was the

442 dependent variable and eQTL was the independent variable. Second, the relationship between

443 eQTLs and SHBG levels was assessed using linear regression in R where SHBG (continuous) was taken

444 as the dependent variable and transformed into a normal distribution using rank transformation,

445 and eQTLs were the independent variable. Finally, if an eQTL was significantly associated with both

446 LOY and SHBG, the interaction effect of the eQTL and LOY on SHBG and other sex hormones was

447 assessed by linear regression in R. Inverse normal rank transformed sex hormone was considered as

448 a continuous dependent variable, and each of eQTL, LOY, and eQTL x LOY (interaction effect) as the

449 independent variable.

450   **References**

451   1.   L. A. Forsberg, C. Rasi, N. Malmqvist, H. Davies, S. Pasupulati, G. Pakalapati, J. Sandgren, T. D.
452   de Ståhl, A. Zaghlool, V. Giedraitis, L. Lannfelt, J. Score, N. C. P. Cross, D. Absher, E. T. Janson,
453   C. M. Lindgren, A. P. Morris, E. Ingelsson, L. Lind, J. P. Dumanski, Mosaic loss of chromosome
454   Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nature*
455   *Genetics* **46**, 624-628 (2014).
456   2.   D. J. Thompson, G. Genovese, J. Halvardson, J. C. Ulirsch, D. J. Wright, C. Terao, O. B.
457   Davidsson, F. R. Day, P. Sulem, Y. Jiang, Genetic predisposition to mosaic Y chromosome loss
458   in blood. *Nature* **575**, 652-657 (2019).
459   3.   D. J. Wright, F. R. Day, N. D. Kerrison, F. Zink, A. Cardona, P. Sulem, D. J. Thompson, S.
460   Sigurjonsdottir, D. F. Gudbjartsson, A. Helgason, J. R. Chapman, S. P. Jackson, C. Langenberg,
461   N. J. Wareham, R. A. Scott, U. Thorsteindottir, K. K. Ong, K. Stefansson, J. R. B. Perry, Genetic
462   variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap
463   with cancer susceptibility. *Nature Genetics* **49**, 674-679 (2017).
464   4.   S. H. Lin, E. Loftfield, J. N. Sampson, W. Zhou, M. Yeager, N. D. Freedman, S. J. Chanock, M. J.
465   Machiela, Mosaic chromosome Y loss is associated with alterations in blood cell counts in UK
466   Biobank men. *Sci Rep* **10**, 3655 (2020).
467   5.   P. Noveski, S. Madjunkova, E. Sukarova Stefanovska, N. Matevska Geshkovska, M.
468   Kuzmanovska, A. Dimovski, D. Plaseska-Karanfilska, Loss of Y Chromosome in Peripheral
469   Blood of Colorectal and Prostate Cancer Patients. *PLoS One* **11**, e0146264 (2016).
470   6.   M. J. Machiela, C. L. Dagnall, A. Pathak, J. T. Loud, S. J. Chanock, M. H. Greene, K. A.
471   McGlynn, D. R. Stewart, Mosaic chromosome Y loss and testicular germ cell tumor risk.
472   *Journal of Human Genetics* **62**, 637-640 (2017).
473   7.   P. Baliakas, L. A. Forsberg, Chromosome Y loss and drivers of clonal hematopoiesis in
474   myelodysplastic syndrome. *Haematologica* **106**, 329-331 (2021).
475   8.   J. P. Dumanski, J. C. Lambert, C. Rasi, V. Giedraitis, H. Davies, B. Grenier-Boley, C. M.
476   Lindgren, D. Campion, C. Dufouil, F. Pasquier, P. Amouyel, L. Lannfelt, M. Ingelsson, L.
477   Kilander, L. Lind, L. A. Forsberg, Mosaic Loss of Chromosome Y in Blood Is Associated with
478   Alzheimer Disease. *Am J Hum Genet* **98**, 1208-1219 (2016).
479   9.   S. Haitjema, D. Kofink, J. van Setten, S. W. van der Laan, A. H. Schoneveld, J. Eales, M.
480   Tomaszewski, S. C. A. de Jager, G. Pasterkamp, F. W. Asselbergs, H. M. den Ruijter, Loss of Y
481   Chromosome in Blood Is Associated With Major Cardiovascular Events During Follow-Up in
482   Men After Carotid Endarterectomy. *Circ Cardiovasc Genet* **10**, e001544 (2017).
483   10.   C. Terao, A. Suzuki, Y. Momozawa, M. Akiyama, K. Ishigaki, K. Yamamoto, K. Matsuda, Y.
484   Murakami, S. A. McCarroll, M. Kubo, P. R. Loh, Y. Kamatani, Chromosomal alterations among
485   age-related haematopoietic clones in Japan. *Nature* **584**, 130-135 (2020).
486   11.   P. L. Greenberg, H. Tuechler, J. Schanz, G. Sanz, G. Garcia-Manero, F. Solé, J. M. Bennett, D.
487   Bowen, P. Fenaux, F. Dreyfus, H. Kantarjian, A. Kuendgen, A. Levis, L. Malcovati, M. Cazzola,
488   J. Cermak, C. Fonatsch, M. M. Le Beau, M. L. Slovak, O. Krieger, M. Luebbert, J. Maciejewski,
489   S. M. M. Magalhaes, Y. Miyazaki, M. Pfeilstöcker, M. Sekeres, W. R. Sperr, R. Stauder, S.
490   Tauro, P. Valent, T. Vallespi, A. A. van de Loosdrecht, U. Germing, D. Haase, Revised
491   International Prognostic Scoring System for Myelodysplastic Syndromes. *Blood* **120**, 2454-
492   2465 (2012).
493   12.   J. P. Dumanski, J. Halvardson, H. Davies, E. Rychlicka-Buniowska, J. Mattisson, B. T.
494   Moghadam, N. Nagy, K. Weglarczyk, K. Bukowska-Strakova, M. Danielsson, P. Olszewski, A.
495   Piotrowski, E. Oerton, A. Ambicka, M. Przewoznik, L. Belch, T. Grodzicki, P. L. Chlosta, S.
496   Imreh, V. Giedraitis, L. Kilander, J. Nordlund, A. Ameur, U. Gyllensten, A. Johansson, A.
497   Jozkowicz, M. Siedlar, A. Klich-Raczka, J. Jaszczynski, S. Enroth, J. Baran, M. Ingelsson, J. R. B.
498   Perry, J. Rys, L. A. Forsberg, Immune cells lacking Y chromosome show dysregulation of
499   autosomal gene expression. *Cell Mol Life Sci* **78**, 4019-4033 (2021).

500 13. K. B. Jacobs, M. Yeager, W. Zhou, S. Wacholder, Z. Wang, B. Rodriguez-Santiago, A.
501   Hutchinson, X. Deng, C. Liu, M.-J. Horner, Detectable clonal mosaicism and its relationship to
502   aging and cancer. *Nature Genetics* **44**, 651–658 (2012).
503 14. C. C. Laurie, C. A. Laurie, K. Rice, K. F. Doheny, L. R. Zelnick, C. P. McHugh, H. Ling, K. N.
504   Hetrick, E. W. Pugh, C. Amos, Detectable clonal mosaicism from birth to old age and its
505   relationship to cancer. *Nature Genetics* **44**, 642 (2012).
506 15. S. Jaiswal, P. Natarajan, A. J. Silver, C. J. Gibson, A. G. Bick, E. Shvartz, M. McConkey, N.
507   Gupta, S. Gabriel, D. Ardissino, Clonal hematopoiesis and risk of atherosclerotic
508   cardiovascular disease. *New England Journal of Medicine* **377**, 111-121 (2017).
509 16. S. Jaiswal, P. Fontanillas, J. Flannick, A. Manning, P. V. Grauman, B. G. Mar, R. C. Lindsley, C.
510   H. Mermel, N. Burtt, A. Chavez, Age-related clonal hematopoiesis associated with adverse
511   outcomes. *New England Journal of Medicine* **371**, 2488-2498 (2014).
512 17. G. Genovese, A. K. Kähler, R. E. Handsaker, J. Lindberg, S. A. Rose, S. F. Bakhoum, K.
513   Chambert, E. Mick, B. M. Neale, M. Fromer, Clonal hematopoiesis and blood-cancer risk
514   inferred from blood DNA sequence. *New England Journal of Medicine* **371**, 2477-2487
515   (2014).
516 18. F. Zink, S. N. Stacey, G. L. Norddahl, M. L. Frigge, O. T. Magnusson, I. Jonsdottir, T. E.
517   Thorgeirsson, A. Sigurdsson, S. A. Gudjonsson, J. Gudmundsson, Clonal hematopoiesis, with
518   and without candidate driver mutations, is common in the elderly. *Blood* **130**, 742-752
519   (2017).
520 19. M. Xie, C. Lu, J. Wang, M. D. McLellan, K. J. Johnson, M. C. Wendl, J. F. McMichael, H. K.
521   Schmidt, V. Yellapantula, C. A. Miller, Age-related mutations associated with clonal
522   hematopoietic expansion and malignancies. *Nature Medicine* **20**, 1472 (2014).
523 20. A. G. Bick, J. S. Weinstock, S. K. Nandakumar, C. P. Fulco, E. L. Bao, S. M. Zekavat, M. D.
524   Szeto, X. Liao, M. J. Leventhal, J. Nasser, K. Chang, C. Laurie, B. B. Burugula, C. J. Gibson, A. E.
525   Lin, M. A. Taub, F. Aguet, K. Ardlie, B. D. Mitchell, K. C. Barnes, A. Moscati, M. Fornage, S.
526   Redline, B. M. Psaty, E. K. Silverman, S. T. Weiss, N. D. Palmer, R. S. Vasan, E. G. Burchard, S.
527   L. R. Kardia, J. He, R. C. Kaplan, N. L. Smith, D. K. Arnett, D. A. Schwartz, A. Correa, M. de
528   Andrade, X. Guo, B. A. Konkle, B. Custer, J. M. Peralta, H. Gui, D. A. Meyers, S. T. McGarvey, I.
529   Y.-D. Chen, M. B. Shoemaker, P. A. Peyser, J. G. Broome, S. M. Gogarten, F. F. Wang, Q.
530   Wong, M. E. Montasser, M. Daya, E. E. Kenny, K. E. North, L. J. Launer, B. E. Cade, J. C. Bis, M.
531   H. Cho, J. Lasky-Su, D. W. Bowden, L. A. Cupples, A. C. Y. Mak, L. C. Becker, J. A. Smith, T. N.
532   Kelly, S. Aslibekyan, S. R. Heckbert, H. K. Tiwari, I. V. Yang, J. A. Heit, S. A. Lubitz, J. M.
533   Johnsen, J. E. Curran, S. E. Wenzel, D. E. Weeks, D. C. Rao, D. Darbar, J.-Y. Moon, R. P. Tracy,
534   E. J. Buth, N. Rafaels, R. J. F. Loos, P. Durda, Y. Liu, L. Hou, J. Lee, P. Kachroo, B. I. Freedman,
535   D. Levy, L. F. Bielak, J. E. Hixson, J. S. Floyd, E. A. Whitsel, P. T. Ellinor, M. R. Irvin, T. E.
536   Fingerlin, L. M. Raffield, S. M. Armasu, M. M. Wheeler, E. C. Sabino, J. Blangero, L. K.
537   Williams, B. D. Levy, W. H.-H. Sheu, D. M. Roden, E. Boerwinkle, J. E. Manson, R. A. Mathias,
538   P. Desai, K. D. Taylor, A. D. Johnson, N. Abe, C. Albert, L. Almasy, A. Alonso, S. Ament, P.
539   Anderson, P. Anugu, D. Applebaum-Bowden, D. Arking, A. Ashley-Koch, S. Aslibekyan, T.
540   Assimes, D. Avramopoulos, J. Barnard, R. G. Barr, E. Barron-Casella, L. Barwick, T. Beaty, G.
541   Beck, D. Becker, R. Beer, A. Beitelshees, E. Benjamin, P. Benos, M. Bezerra, L. Bielak, R.
542   Bowler, J. Brody, U. Broeckel, K. Bunting, C. Bustamante, J. Cardwell, V. Carey, C. Carty, R.
543   Casaburi, J. Casella, P. Castaldi, M. Chaffin, C. Chang, Y.-C. Chang, D. Chasman, S. Chavan, B.-
544   J. Chen, W.-M. Chen, S. H. Choi, L.-M. Chuang, M. Chung, R.-H. Chung, C. Clish, S. Comhair, E.
545   Cornell, C. Crandall, J. Crapo, J. Curtis, C. Damcott, S. Das, S. David, C. Davis, M. DeBaun, R.
546   Deka, D. DeMeo, S. Devine, Q. Duan, R. Duggirala, S. Dutcher, C. Eaton, L. Ekunwe, A. E.
547   Boueiz, L. Emery, S. Erzurum, C. Farber, M. Flickinger, N. Franceschini, C. Frazar, M. Fu, S. M.
548   Fullerton, L. Fulton, S. Gabriel, W. Gan, S. Gao, Y. Gao, M. Gass, B. Gelb, X. Geng, M. Geraci,
549   S. Germer, R. Gerszten, A. Ghosh, R. Gibbs, C. Gignoux, M. Gladwin, D. Glahn, D.-W. Gong, H.
550   Goring, S. Graw, D. Grine, C. C. Gu, Y. Guan, N. Gupta, J. Haessler, M. Hall, D. Harris, N. L.

Hawley, B. Heavner, R. Hernandez, D. Herrington, C. Hersh, B. Hidalgo, B. Hobbs, J. Hokanson, E. Hong, K. Hoth, C. Hsiung, Y.-J. Hung, H. Huston, C. M. Hwu, R. Jackson, D. Jain, C. Jaquish, M. A. Jhun, C. Johnson, R. Johnston, K. Jones, H. M. Kang, S. Kelly, M. Kessler, A. Khan, W. Kim, G. Kinney, H. Kramer, C. Lange, M. LeBoff, S. S. Lee, W.-J. Lee, J. LeFaive, D. Levine, J. Lewis, X. Li, Y. Li, H. Lin, H. Lin, K. H. Lin, X. Lin, S. Liu, Y. Liu, K. Lunetta, J. Luo, M. Mahaney, B. Make, A. Manichaikul, L. Margolin, L. Martin, S. Mathai, S. May, P. McArdle, M.-L. McDonald, S. McFarland, D. McGoldrick, C. McHugh, H. Mei, L. Mestroni, J. Mikulla, N. Min, M. Minear, R. L. Minster, M. Moll, C. Montgomery, S. Musani, S. Mwasongwe, J. C. Mychaleckyj, G. Nadkarni, R. Naik, T. Naseri, S. Nekhai, S. C. Nelson, B. Neltner, D. Nickerson, J. O'Connell, T. O'Connor, H. Ochs-Balcom, D. Paik, J. Pankow, G. Papanicolaou, A. Parsa, M. Perez, J. Perry, U. Peters, N. T.-O. f. P. M. Consortium, Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature*, (2020).

21. V. Ljungström, J. Mattisson, J. Halvardson, T. Pandzic, H. Davies, E. Rychlicka-Buniowska, M. Danielsson, P. Lacaze, L. Cavelier, J. P. Dumanski, P. Baliakas, L. A. Forsberg, Loss of Y and clonal hematopoiesis in blood—two sides of the same coin? *Leukemia* **36**, 889-891 (2022).

22. A. G. Bick, K. Popadin, C. W. Thorball, M. M. Uddin, M. V. Zanni, B. Yu, M. Cavassini, A. Rauch, P. Tarr, P. Schmid, E. Bernasconi, H. F. Günthard, P. Libby, E. Boerwinkle, P. J. McLaren, C. M. Ballantyne, S. Grinspoon, P. Natarajan, J. Fellay, I. Abela, K. Aebi-Popp, A. Anagnostopoulos, M. Battegay, E. Bernasconi, D. L. Braun, H. C. Bucher, A. Calmy, M. Cavassini, A. Ciuffi, G. Dollenmaier, M. Egger, L. Elzi, J. Fehr, J. Fellay, H. Furrer, C. A. Fux, H. F. Günthard, A. Hachfeld, D. Haerry, B. Hasse, H. H. Hirsch, M. Hoffmann, I. Hösli, M. Huber, C. R. Kahlert, L. Kaiser, O. Keiser, T. Klimkait, R. D. Kouyos, H. Kovari, K. Kusejko, G. Martinetti, B. M. de Tejada, C. Marzolini, K. J. Metzner, N. Müller, J. Nemeth, D. Nicca, P. Paioni, G. Pantaleo, M. Perreau, A. Rauch, P. Schmid, R. Speck, M. Stöckle, P. Tarr, A. Trkola, G. Wandeler, S. Yerly, H. I. V. C. S. the Swiss, Increased prevalence of clonal hematopoiesis of indeterminate potential amongst people living with HIV. *Scientific Reports* **12**, 577 (2022).

23. S. M. Zekavat, S.-H. Lin, A. G. Bick, A. Liu, K. Paruchuri, C. Wang, M. M. Uddin, Y. Ye, Z. Yu, X. Liu, Y. Kamatani, R. Bhattacharya, J. P. Pirruccello, A. Pampana, P.-R. Loh, P. Kohli, S. A. McCarroll, K. Kiryluk, B. Neale, I. Ionita-Laza, E. A. Engels, D. W. Brown, J. W. Smoller, R. Green, E. W. Karlson, M. Lebo, P. T. Ellinor, S. T. Weiss, M. J. Daly, S. Koyama, K. Ito, Y. Momozawa, K. Matsuda, Y. Yamanashi, Y. Furukawa, T. Morisaki, Y. Murakami, K. Muto, A. Nagai, W. Obara, K. Yamaji, K. Takahashi, S. Asai, Y. Takahashi, T. Suzuki, N. Sinozaki, H. Yamaguchi, S. Minami, S. Murayama, K. Yoshimori, S. Nagayama, D. Obata, M. Higashiyama, A. Masumoto, Y. Koretsune, A. Palotie, A. Ziemann, A. Mitchell, A. Huertas-Vazquez, A. Salminen, A. Jussila, A. Havulinna, A. Mackay, A. Abbasi, A. Elliott, A. Cole, A. Shcherban, A. Mälarstig, A. Ganna, A. Loboda, A. Podgornaia, A. Lehtonen, A. Pitkäranta, A. Remes, A. Auranen, A. Hakanen, A. Palomäki, A. Jalanko, A. Loukola, A. Chhibber, A. Lertratanakul, A. Lehisto, A. Mannermaa, Å. Hedman, A. Chu, A. Madar, A. Ghazal, B. Challis, B. Sun, B. Cummings, B. Riley-Gillis, C. Fox, C.-Y. Chen, C. Wang, C. Chatelain, D. Gordin, D. Quarless, D. Oh, D. Choy, D. Close, D. Pulford, D. Rice, D. Waterworth, D. Rajpal, D. Baird, D. Jambulingam, D. Chang, D. Kulkarni, D. Paul, D. Liu, E. Teng, E. Punkka, E. Ekholm, E. Kangasniemi, E. Laakkonen, E. Wigmore, E. Järvensivu, E. Kilpeläinen, E. Widen, E. Tsai, E. Mohammed, E. Strauss, E. Kvikstad, E. Pitkänen, E. Kaiharju, E. Xu, F. Xu, F. Rahimov, F. Vaura, F. Auge, G. Brein, G. Lassi, G. Heap, H. Laivuori, H. Mattsson, H. Uusitalo-Järvinen, H. Kankaanranta, H. Uusitalo, H. Chen, H. Siirtola, H. Joensuu, H. Runz, H. Lehtonen, H. Heyne, H. Soininen, H. Jacob, H. Chen, H.-Y. Shen, H. Xu, I. Vähätalo, I. Kalliala, I. Tachmazidou, J. Kaprio, J. Parkkinen, J. Jacob, J. Kumar, J. van Adelsberg, J. Laukkanen, J. Ritari, J. Garcia-Tabuenca, J. Waring, J. Schutzman, J. Liu, J. Lee, J. Betts, J. Rämö, J. Huhtakangas, J. Mäkelä, J. Mattson, J. Schleutker, J. Kettunen, J. Eicher, J. Zierer, J. Chung, J. A. Turunen, J. E. Gordillo, J. Maranville, J. Karjalainen, J. Mehtonen, J. Rinne, J. Sinisalo, J. Junttila, J. Koskela, J. Partanen, J. Peltola, J. Hunkapiller, J. Pihlajamäki, J. Wade, J. Partanen, K. Mäkikallio, K.

602   Kaarniranta, K. Tasanen, K. Metsärinne, K. Pärn, K. S. King, K. Eklund, K. Linden, K. Nieminen,
603   K. Hannula-Jouppi, K. Call, K. Klinger, K. Donner, K. Hyvärinen, K. Kristiansson, K. Kivinen, K.
604   Kaukinen, K. Pylkäs, K. de Lange, K. Usiskin, K. Palin, K. Shkura, K. Auro, K. Kalpala, K. Sipilä, K.
605   Elenius, K. Tsuo, L. E. Lahtela, L. Addis, L. Huilaja, L. Kotaniemi-Talonen, L. Mustaniemi, L.
606   Pirilä, L. Morin-Papunen, L. Aaltonen, L. Koulu, L. Suominen, L. Kallio, L. McCarthy, L. Aoxing,
607   L. Männikkö, M. Obeidat, M. Rivas, M. Hautalahti, M. Pelkonen, M. Kaunisto, M. E. Niemi, M.
608   Siponen, M. Crohns, M. Kalaoja, M. Luodonpää, M. Vääräsmäki, M.-R. Taskinen, M.
609   Tuppurainen, M. J. Daly, M. McCarthy, M. Laakso, M. Laukkanen, M. Voutilainen, M. Juonala,
610   M. Perola, M. Hochfeld, M. Färkkilä, M. P. Reeve, M. Kanai, M. Brauer, M. Gossel, M. Peura,
611   M. Ehm, M. Miller, M. Liu, M. Aavikko, M. Koskinen, M. Helminen, M. Kähönen, M. Arvas, M.
612   Hiltunen, M. Kiviniemi, M. Caliskan, M. Karjalainen, M. Raivio, M. Koivusalo, M. Kurki, M.
613   Maasha, N. Bing, N. Bowers, N. Raghavan, N. Renaud, N. Välimäki, N. Hautala, N. Mars, N.
614   Pitkänen, N. Smaoui, O. Kaipiainen-Seppänen, O. Carpén, O. A. Dada, O. Soylemez, O.
615   Heikinheimo, O. Tuovila, O. Uimari, P. Gormley, P. Auvinen, P. Laiho, P. Mäntylä, P. Polo, P.
616   Bronson, P. Kauppi, P. Karihtala, P. Nieminen, P. The Biobank Japan, C. FinnGen,
617   Hematopoietic mosaic chromosomal alterations increase the risk for diverse types of
618   infection. *Nature Medicine* **27**, 1012-1024 (2021).
619 24.   A. A. Z. Dawoud, R. D. Gilbert, W. J. Tapper, N. C. P. Cross, Clonal myelopoiesis promotes
620   adverse outcomes in chronic kidney disease. *Leukemia* **36**, 507-515 (2022).
621 25.   M. Buscarlet, S. Provost, Y. F. Zada, A. Barhdadi, V. Bourgoin, G. Lépine, L. Mollica, N. Szuber,
622   M.-P. Dubé, L. Busque, DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate
623   benign phenotypes and different genetic predispositions. *Blood* **130**, 753-762 (2017).
624 26.   S. Sano, K. Horitani, H. Ogawa, J. Halvardson, N. W. Chavkin, Y. Wang, M. Sano, J. Mattisson,
625   A. Hata, M. Danielsson, E. Miura-Yura, A. Zaghlool, M. A. Evans, T. Fall, H. N. De Hoyos, J.
626   Sundström, Y. Yura, A. Kour, Y. Arai, M. C. Thel, Y. Arai, J. C. Mychaleckyj, K. K. Hirschi, L. A.
627   Forsberg, K. Walsh, Hematopoietic loss of Y chromosome leads to cardiac fibrosis and heart
628   failure mortality. *Science* **377**, 292-297 (2022).
629 27.   A. Niroula, A. Sekar, M. A. Murakami, M. Trinder, M. Agrawal, W. J. Wong, A. G. Bick, M. M.
630   Uddin, C. J. Gibson, G. K. Griffin, M. C. Honigberg, S. M. Zekavat, K. Paruchuri, P. Natarajan,
631   B. L. Ebert, Distinction of lymphoid and myeloid clonal hematopoiesis. *Nature Medicine* **27**,
632   1921-1927 (2021).
633 28.   R. Saiki, Y. Momozawa, Y. Nannya, M. M. Nakagawa, Y. Ochi, T. Yoshizato, C. Terao, Y.
634   Kuroda, Y. Shiraishi, K. Chiba, H. Tanaka, A. Niida, S. Imoto, K. Matsuda, T. Morisaki, Y.
635   Murakami, Y. Kamatani, S. Matsuda, M. Kubo, S. Miyano, H. Makishima, S. Ogawa, Combined
636   landscape of single-nucleotide variants and copy number alterations in clonal
637   hematopoiesis. *Nature Medicine* **27**, 1239-1249 (2021).
638 29.   M. M. Ouseph, R. P. Hasserjian, P. Dal Cin, S. B. Lovitch, D. P. Steensma, V. Nardi, O. K.
639   Weinberg, Genomic alterations in patients with somatic loss of the Y chromosome as the
640   sole cytogenetic finding in bone marrow cells. *Haematologica* **106**, 555-564 (2021).
641 30.   Y. Zhao, S. Stankovic, M. Koprulu, E. Wheeler, F. R. Day, H. Lango Allen, N. D. Kerrison, M.
642   Pietzner, P.-R. Loh, N. J. Wareham, C. Langenberg, K. K. Ong, J. R. B. Perry, GIGYF1 loss of
643   function is associated with clonal mosaicism and adverse metabolic health. *Nature
644   Communications* **12**, 4178 (2021).
645 31.   A. V. Jones, A. Chase, R. T. Silver, D. Oscier, K. Zoi, Y. L. Wang, H. Cario, H. L. Pahl, A. Collins,
646   A. Reiter, JAK2 haplotype is a major risk factor for the development of myeloproliferative
647   neoplasms. *Nature Genetics* **41**, 446-449 (2009).
648 32.   A. A. Z. Dawoud, W. J. Tapper, N. C. P. Cross, Clonal myelopoiesis in the UK Biobank cohort:
649   ASXL1 mutations are strongly associated with smoking. *Leukemia* **34**, 2660-2672 (2020).
650 33.   J. P. Dumanski, C. Rasi, M. Lönn, H. Davies, M. Ingelsson, V. Giedraitis, L. Lannfelt, P. K.
651   Magnusson, C. M. Lindgren, A. P. Morris, D. Cesarini, M. Johannesson, E. Tiensuu Janson, L.

652   Lind, N. L. Pedersen, E. Ingelsson, L. A. Forsberg, Mutagenesis. Smoking is associated with
653   mosaic loss of chromosome Y. *Science* **347**, 81-83 (2015).
654 34.   A. Gray, H. A. Feldman, J. B. McKinlay, C. Longcope, Age, Disease, and Changing Sex
655   Hormone Levels in Middle-Aged Men: Results of the Massachusetts Male Aging Study*. *The*
656   *Journal of Clinical Endocrinology & Metabolism* **73**, 1016-1025 (1991).
657 35.   P. Emadi-Konjin, J. Bain, I. L. Bromberg, Evaluation of an algorithm for calculation of serum
658   "Bioavailable" testosterone (BAT). *Clinical Biochemistry* **36**, 591-596 (2003).
659 36.   J. F. Dunn, B. C. Nisula, D. Rodbard, Transport of Steroid Hormones: Binding of 21
660   Endogenous Steroids to Both Testosterone-Binding Globulin and Corticosteroid-Binding
661   Globulin in Human Plasma. *The Journal of Clinical Endocrinology & Metabolism* **53**, 58-68
662   (1981).
663 37.   M. J. Wheeler, The Determination of Bio-Available Testosterone. *Annals of Clinical*
664   *Biochemistry* **32**, 345-357 (1995).
665 38.   A. Vermeulen, T. StoÏCa, L. Verdonck, The Apparent Free Testosterone Concentration, An
666   Index of Androgenicity. *The Journal of Clinical Endocrinology & Metabolism* **33**, 759-767
667   (1971).
668 39.   A. Manni, W. M. Pardridge, W. Cefalu, B. C. Nisula, C. W. Bardin, S. J. Santner, R. J. Santen,
669   Bioavailability of albumin-bound testosterone. *J Clin Endocrinol Metab* **61**, 705-710 (1985).
670 40.   R. J. Marriott, K. Murray, G. J. Hankey, L. Manning, G. Dwivedi, F. C. W. Wu, B. B. Yeap,
671   Longitudinal changes in serum testosterone and sex hormone-binding globulin in men aged
672   40-69 years from the UK Biobank. *Clin Endocrinol (Oxf)* **96**, 589-598 (2022).
673 41.   E. Aribas, M. Kavousi, J. S. E. Laven, M. A. Ikram, J. E. Roeters van Lennep, Aging,
674   Cardiovascular Risk, and SHBG Levels in Men and Women From the General Population. *J*
675   *Clin Endocrinol Metab* **106**, 2890-2900 (2021).
676 42.   A. D. Coviello, R. Haring, M. Wellons, D. Vaidya, T. Lehtimäki, S. Keildson, K. L. Lunetta, C. He,
677   M. Fornage, V. Lagou, M. Mangino, N. C. Onland-Moret, B. Chen, J. Eriksson, M. Garcia, Y. M.
678   Liu, A. Koster, K. Lohman, L. P. Lyytikäinen, A. K. Petersen, J. Prescott, L. Stolk, L. Vandenput,
679   A. R. Wood, W. V. Zhuang, A. Ruokonen, A. L. Hartikainen, A. Pouta, S. Bandinelli, R. Biffar, G.
680   Brabant, D. G. Cox, Y. Chen, S. Cummings, L. Ferrucci, M. J. Gunter, S. E. Hankinson, H.
681   Martikainen, A. Hofman, G. Homuth, T. Illig, J. O. Jansson, A. D. Johnson, D. Karasik, M.
682   Karlsson, J. Kettunen, D. P. Kiel, P. Kraft, J. Liu, Ö. Ljunggren, M. Lorentzon, M. Maggio, M. R.
683   Markus, D. Mellström, I. Miljkovic, D. Mirel, S. Nelson, L. Morin Papunen, P. H. Peeters, I.
684   Prokopenko, L. Raffel, M. Reincke, A. P. Reiner, K. Rexrode, F. Rivadeneira, S. M. Schwartz, D.
685   Siscovick, N. Soranzo, D. Stöckl, S. Tworoger, A. G. Uitterlinden, C. H. van Gils, R. S. Vasan, H.
686   E. Wichmann, G. Zhai, S. Bhasin, M. Bidlingmaier, S. J. Chanock, I. De Vivo, T. B. Harris, D. J.
687   Hunter, M. Kähönen, S. Liu, P. Ouyang, T. D. Spector, Y. T. van der Schouw, J. Viikari, H.
688   Wallaschofski, M. I. McCarthy, T. M. Frayling, A. Murray, S. Franks, M. R. Järvelin, F. H. de
689   Jong, O. Raitakari, A. Teumer, C. Ohlsson, J. M. Murabito, J. R. Perry, A genome-wide
690   association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci
691   implicated in sex steroid hormone regulation. *PLoS Genet* **8**, e1002805 (2012).
692 43.   S. Burgess, S. G. Thompson, Use of allele scores as instrumental variables for Mendelian
693   randomization. *International Journal of Epidemiology* **42**, 1134-1144 (2013).
694 44.   C. Terao, Y. Momozawa, K. Ishigaki, E. Kawakami, M. Akiyama, P.-R. Loh, G. Genovese, H.
695   Sugishita, T. Ohta, M. Hirata, J. R. B. Perry, K. Matsuda, Y. Murakami, M. Kubo, Y. Kamatani,
696   GWAS of mosaic loss of chromosome Y highlights genetic effects on blood cell
697   differentiation. *Nature Communications* **10**, 4719 (2019).
698 45.   R. Neale. (Neale Lab, Boston, MA, 2018).
699 46.   U. Võsa, A. Claringbould, H.-J. Westra, M. J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha, R.
700   Kreuzhuber, S. Yazar, H. Brugge, R. Oelen, D. H. de Vries, M. G. P. van der Wijst, S. Kasela, N.
701   Pervjakova, I. Alves, M.-J. Favé, M. Agbessi, M. W. Christiansen, R. Jansen, I. Seppälä, L. Tong,
702   A. Teumer, K. Schramm, G. Hemani, J. Verlouw, H. Yaghootkar, R. Sönmez Flitman, A. Brown,

V. Kukushkina, A. Kalnapenkis, S. Rüeger, E. Porcu, J. Kronberg, J. Kettunen, B. Lee, F. Zhang, T. Qi, J. A. Hernandez, W. Arindrarto, F. Beutner, P. A. C. 't Hoen, J. van Meurs, J. van Dongen, M. van Iterson, M. A. Swertz, M. Jan Bonder, J. Dmitrieva, M. Elansary, B. P. Fairfax, M. Georges, B. T. Heijmans, A. W. Hewitt, M. Kähönen, Y. Kim, J. C. Knight, P. Kovacs, K. Krohn, S. Li, M. Loeffler, U. M. Marigorta, H. Mei, Y. Momozawa, M. Müller-Nurasyid, M. Nauck, M. G. Nivard, B. W. J. H. Penninx, J. K. Pritchard, O. T. Raitakari, O. Rotzschke, E. P. Slagboom, C. D. A. Stehouwer, M. Stumvoll, P. Sullivan, P. A. C. 't Hoen, J. Thiery, A. Tönjes, J. van Dongen, M. van Iterson, J. H. Veldink, U. Völker, R. Warmerdam, C. Wijmenga, M. Swertz, A. Andiappan, G. W. Montgomery, S. Ripatti, M. Perola, Z. Kutalik, E. Dermitzakis, S. Bergmann, T. Frayling, J. van Meurs, H. Prokisch, H. Ahsan, B. L. Pierce, T. Lehtimäki, D. I. Boomsma, B. M. Psaty, S. A. Gharib, P. Awadalla, L. Milani, W. H. Ouwehand, K. Downes, O. Stegle, A. Battle, P. M. Visscher, J. Yang, M. Scholz, J. Powell, G. Gibson, T. Esko, L. Franke, B. Consortium, Q. T. L. C. i, 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 (2021).

47. C. M. Mendel, The Free Hormone Hypothesis: A Physiologically Based Mathematical Model*. *Endocrine Reviews* **10**, 232-274 (1989).

48. S. T. d. Rocha, C. A. Edwards, M. Ito, T. Ogata, A. C. Ferguson-Smith, Genomic imprinting at the mammalian Dlk1-Dio3 domain. *Trends in Genetics* **24**, 306-316 (2008).

49. A. Chase, W. Leung, W. Tapper, A. V. Jones, L. Knoops, C. Rasi, L. A. Forsberg, P. Guglielmelli, K. Zoi, V. Hall, Profound parental bias associated with chromosome 14 acquired uniparental disomy indicates targeting of an imprinted locus. *Leukemia* **29**, 2069-2074 (2015).

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

51. M. D. Kessler, A. Damask, S. O'Keeffe, M. Van Meter, N. Banerjee, S. Semrau, D. Li, K. Watanabe, J. Horowitz, Y. Houvras, C. Gillies, J. Mbatchou, R. R. White, J. A. Kosmicki, M. G. LeBlanc, M. Jones, R. G. Center, G.-R. D. Collaboration, D. J. Glass, L. A. Lotta, M. N. Cantor, G. S. Atwal, A. E. Locke, M. A. R. Ferreira, R. Deering, C. Paulding, A. R. Shuldiner, G. Thurston, W. Salerno, J. G. Reid, J. D. Overton, J. Marchini, H. M. Kang, A. Baras, G. R. Abecasis, E. Jorgenson, Exome sequencing of 628,388 individuals identifies common and rare variant associations with clonal hematopoiesis phenotypes. *MedRxiv*, 2021.2012.2029.21268342 (2022).

52. M. Buscarlet, S. Provost, Y. F. Zada, A. Barhdadi, V. Bourgoin, G. Lépine, L. Mollica, N. Szuber, M.-P. Dubé, L. Busque, 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 (2017).

53. E. Mitchell, M. Spencer Chapman, N. Williams, K. J. Dawson, N. Mende, E. F. Calderbank, H. Jung, T. Mitchell, T. H. H. Coorens, D. H. Spencer, H. Machado, H. Lee-Six, M. Davies, D. Hayler, M. A. Fabre, K. Mahbubani, F. Abascal, A. Cagan, G. S. Vassiliou, J. Baxter, I. Martincorena, M. R. Stratton, D. G. Kent, K. Chatterjee, K. S. Parsy, A. R. Green, J. Nangalia, E. Laurenti, P. J. Campbell, Clonal dynamics of haematopoiesis across the human lifespan. *Nature* **606**, 343-350 (2022).

54. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).

55. A. Vermeulen, L. Verdonck, J. M. Kaufman, A Critical Evaluation of Simple Methods for the Estimation of Free Testosterone in Serum. *The Journal of Clinical Endocrinology & Metabolism* **84**, 3666-3672 (1999).

56. W. Kim, M. Moll, D. Qiao, B. D. Hobbs, N. Shrine, P. Sakornsakolpat, M. D. Tobin, F. Dudbridge, L. V. Wain, C. Ladd-Acosta, N. Chatterjee, E. K. Silverman, M. H. Cho, T. H. Beaty,

753 Interaction of Cigarette Smoking and Polygenic Risk Score on Reduced Lung Function. *JAMA*
754 *Network Open* **4**, e2139525-e2139525 (2021).
755 57. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar,
756 P. I. De Bakker, M. J. Daly, PLINK: a tool set for whole-genome association and population-
757 based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575 (2007).
758 58. G. D. Smith, N. M. Davies, N. Dimou, M. Egger, V. Gallo, R. Golub, J. P. Higgins, C.
759 Langenberg, E. W. Loder, J. B. Richards, "STROBE-MR: guidelines for strengthening the
760 reporting of Mendelian randomization studies" (PeerJ Preprints, 2019).
761 59. G. Hemani, J. Zheng, B. Elsworth, K. H. Wade, V. Haberland, D. Baird, C. Laurin, S. Burgess, J.
762 Bowden, R. Langdon, The MR-Base platform supports systematic causal inference across the
763 human phenome. *Elife* **7**, e34408 (2018).
764 60. Z. Zhang, Propensity score method: a non-parametric technique to reduce model
765 dependence. *Ann Transl Med* **5**, 7 (2017).
766 61. K. Cibulskis, M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M.
767 Meyerson, E. S. Lander, G. Getz, Sensitive detection of somatic point mutations in impure
768 and heterogeneous cancer samples. *Nature Biotechnology* **31**, 213-219 (2013).
769 62. Z. Sondka, S. Bamford, C. G. Cole, S. A. Ward, I. Dunham, S. A. Forbes, The COSMIC Cancer
770 Gene Census: describing genetic dysfunction across all human cancers. *Nature Reviews*
771 *Cancer* **18**, 696-705 (2018).

772

773

Acknowledgments

777

Funding

780

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.

784

Declaration of interests

The authors declare no competing interests

787

Data and materials availability

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 (version 0.5.5) guide is available from

(https://mrcieu.github.io/TwoSampleMR/). The code used for calculating FT and BAT is available

793   from (https://labrtorian.com/tag/free-testosterone/). Further details regarding the biochemical

794   assay methods and external quality assurance are available at

795   https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/serum_biochemistry.pdf. All other data

796   are available in the main text or the supplementary materials.

Figure 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, 10 PC 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) Analysis using 8 independent SNPs associated with SHBG (genome wide significant or derived from conditional analysis) from multiple cohorts of men and women (*42*) used as instrumental variables in BBJ. The IVW test estimated a significant positive effect of SHBG on LOY ($P = 6.58 \times 10^{-4}$). (B) Conservative analysis using a subset of 4 SNPs associated with SHBG at genome wide significance in men ($P = 7.28 \times 10^{-4}$). (C) Analysis using 40 independent SNPs associated with LOY in BBJ assessed as instrumental variables in UKB. The IVW test estimated no effect of LOY on SHBG ($P = 0.21$). The line of regression is indicated in blue, and the axes show β coefficients for SNP effects on SHBG and LOY. rs7910927 at 10q21.3 within *JMJD1C* is highlighted with a red circle in panels (A) and (B), but not in panel (C) as it was not associated with LOY at a genome wide significant level in BBJ.

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 assessed as predictors for LOY (panel A) and the 8 significant SNPs were compared to SHBG levels (panel B) incorporating age, age squared, smoking history and intensity, and 10 PC as covariates.

831 **Figure 5. The relationship between LOY and CH**

832 LOY was stratified according to the clonal size and the proportion of participants with CH within each

833 group was compared with controls. (A) myeloid CH, (B) lymphoid CH, (C) unknown driver CH.

834

835 **Figure 6. The relationship between LOY clonal size and CH VAFs**

836 Boxplots summarizing the distribution of VAFs of somatic mutations in controls and cases with LOY

837 broken down by clone size. (A) myeloid CH, (B) lymphoid CH, (C) unknown driver CH. The red lines

838 connect median values.

839

840 **Figure 7. The distribution of B-allele frequencies and estimated clone sizes for the 44,558 detected**

841 **LOY events.**

842 **Table 1: The relationship between LOY with clone size ≥30% and driver mutations on the gene**

843 **level**

Driver mutations	Control		LOY		OR	95% confidence intervals		P_{FDR}
	No CH	CH	No CH	CH				
TET2	16199	253	727	30	2.64	1.73	3.90	9.58×10^{-5}
TP53	16199	16	727	5	6.96	1.99	19.95	7.62×10^{-3}
CBL	16199	9	727	3	7.43	1.29	29.84	0.04
NF1	16199	15	727	3	4.46	0.83	15.80	0.09
DNMT3A	16199	637	727	18	0.63	0.37	1.01	0.11
SF3B1	16199	36	727	4	2.48	0.64	6.93	0.14
STAG2	16199	113	727	8	1.58	0.66	3.23	0.33
SRSF2	16199	40	727	3	1.67	0.33	5.27	0.49
ASXL1	16199	167	727	7	0.93	0.37	1.98	1.00

844

845 Note: other variants associated with CH such as *JAK2* V617F do not appear on this table as they were not

846 mutated in ≥3 males with LOY in ≥30% of cells

847

848

849

850 **Supplementary Materials**

851 Tables S1 – S18

852

A

	β	95% CI		P_FDR
SHBG	0.12	0.10	0.13	7.44×10^{-36}
Testosterone	0.09	0.07	0.11	2.23×10^{-20}
HDL cholesterol	0.08	0.06	0.09	5.16×10^{-14}
IGF-1	0.06	0.04	0.08	3.37×10^{-11}
Apolipoprotein A	0.05	0.03	0.07	4.07×10^{-6}
Vitamin D	0.04	0.02	0.06	2.54×10^{-5}
Oestradiol	0.04	-0.02	0.10	0.32
Phosphate	0.03	0.02	0.05	7.15×10^{-4}
Lipoprotein A	0.02	0.00	0.04	0.11
Free Testosterone	0.01	-0.01	0.03	0.46
Cholesterol	0.01	-0.01	0.02	0.52
LDL direct	0.01	-0.01	0.02	0.52
Bioavailable Testosterone	0.00	-0.01	0.02	0.75
Direct bilirubin	0.00	-0.02	0.02	0.84
Calcium	-0.01	-0.03	0.01	0.51
Urea	-0.01	-0.03	0.01	0.38
Total protein	-0.01	-0.03	0.01	0.38
Glycated haemoglobin (HbA1c)	-0.01	-0.03	0.01	0.35
Apolipoprotein B	-0.01	-0.03	0.00	0.20
Total bilirubin	-0.02	-0.03	0.00	0.15
Albumin	-0.02	-0.04	0.00	0.06
Alkaline phosphatase	-0.02	-0.04	0.00	0.06
C-reactive protein	-0.02	-0.04	-0.01	0.02
Creatinine	-0.03	-0.05	-0.01	1.77×10^{-3}
Glucose	-0.05	-0.07	-0.03	8.27×10^{-7}
Aspartate aminotransferase	-0.05	-0.07	-0.03	1.72×10^{-7}
Cystatin C	-0.05	-0.07	-0.04	1.16×10^{-9}
Alanine aminotransferase	-0.07	-0.09	-0.05	3.16×10^{-14}
Urate	-0.08	-0.10	-0.06	1.29×10^{-16}
Gamma glutamyltransferase	-0.10	-0.11	-0.08	5.43×10^{-26}
Triglycerides	-0.11	-0.12	-0.09	4.49×10^{-30}

-0.12   -0.05   0   0.05   0.12

β coefficient (LOY)

B

	β	95% CI		P_FDR
SHBG	0.08	0.07	0.10	4.61×10^{-21}
Testosterone	0.05	0.04	0.07	4.13×10^{-9}
HDL cholesterol	0.04	0.02	0.05	2.05×10^{-4}
IGF-1	0.05	0.04	0.07	7.01×10^{-9}
Apolipoprotein A	0.02	0.00	0.03	0.15
Vitamin D	0.03	0.01	0.04	5.92×10^{-3}
Oestradiol	0.04	-0.02	0.11	0.24
Phosphate	0.03	0.01	0.05	3.49×10^{-3}
Lipoprotein A	0.02	0.00	0.04	0.15
Free Testosterone	0.00	-0.02	0.01	0.75
Cholesterol	0.00	-0.02	0.01	0.75
LDL direct	0.00	-0.02	0.02	0.91
Bioavailable Testosterone	-0.01	-0.03	0.01	0.36
Direct bilirubin	0.00	-0.02	0.01	0.75
Calcium	-0.01	-0.03	0.01	0.25
Urea	0.00	-0.02	0.02	0.97
Total protein	-0.01	-0.03	0.01	0.46
Glycated haemoglobin (HbA1c)	0.02	0.00	0.03	0.09
Apolipoprotein B	-0.01	-0.03	0.01	0.24
Total bilirubin	-0.02	-0.04	0.00	0.06
Albumin	-0.03	-0.05	-0.01	3.22×10^{-3}
Alkaline phosphatase	-0.02	-0.03	0.00	0.15
C-reactive protein	0.00	-0.01	0.02	0.70
Creatinine	-0.02	-0.04	0.00	0.04
Glucose	-0.03	-0.05	-0.01	2.23×10^{-3}
Aspartate aminotransferase	-0.04	-0.06	-0.02	8.67×10^{-5}
Cystatin C	-0.03	-0.05	-0.01	6.62×10^{-4}
Alanine aminotransferase	-0.04	-0.06	-0.02	8.01×10^{-6}
Urate	-0.05	-0.07	-0.04	7.01×10^{-9}
Gamma glutamyltransferase	-0.07	-0.09	-0.06	5.19×10^{-16}
Triglycerides	-0.07	-0.09	-0.06	1.81×10^{-16}

-0.1   -0.05   0   0.05   0.1

β coefficient (LOY)

Figure 1

Sex Hormone Binding Globulin (SHBG)

Total Testosterone (TT)

Free Testosterone (FT)

Bioavailable Testosterone (BAT)

Figure 2

A

Gene	SNP	Allele		OR	95% Confidence interval		P_{FDR}
HM13	rs6060260	T		0.83	[0.80	0.86]	2.60×10^{-22}
TCL1A	rs11849538	G		0.84	[0.80	0.87]	5.02×10^{-16}
SENP7	rs7628681	T		1.11	[1.08	1.14]	2.29×10^{-13}
SMPD2	rs7773095	C		0.89	[0.86	0.92]	2.01×10^{-12}
DLK1	rs7141210	T		0.90	[0.88	0.93]	2.22×10^{-12}
MAD1L1	rs10247428	A		0.92	[0.89	0.94]	6.00×10^{-10}
ACAT1	rs12361905	C		0.96	[0.93	0.98]	3.45×10^{-3}
SETBP1	rs11876015	C		0.93	[0.89	0.98]	7.11×10^{-3}
SEMA4A	rs6701295	C		1.03	[1.00	1.05]	0.10
RBPMS	rs2978263	T		1.03	[0.99	1.06]	0.14
BCL2	rs4940576	T		0.98	[0.95	1.01]	0.17
TSC22D2	rs1868673	A		0.98	[0.96	1.01]	0.30
QKI	rs1234977	C		1.02	[0.98	1.07]	0.35

B

Gene	SNP	Allele		β	95% Confidence interval		P_{FDR}
HM13	rs6060260	T		0.00	[-0.01	0.02]	0.67
TCL1A	rs11849538	G		-0.02	[-0.03	0.00]	0.15
SENP7	rs7628681	T		0.01	[0.00	0.02]	0.47
SMPD2	rs7773095	C		0.00	[-0.01	0.01]	0.79
DLK1	rs7141210	T		-0.01	[-0.03	0.00]	**0.05**
MAD1L1	rs10247428	A		0.02	[0.00	0.03]	**0.04**
ACAT1	rs12361905	C		0.00	[-0.01	0.01]	0.67
SETBP1	rs11876015	C		0.00	[-0.02	0.02]	0.98

Figure 4

Figure 3

A)

LOY	Control		LOY			OR	95% CI	P_FDR
	No CH	CH	NO CH	CH				
<10%	16199	1503	11836	942		0.86	[0.79, 0.93]	9.70×10^{-4}
10%-20%	16199	1503	2983	265		0.96	[0.83, 1.10]	0.62
20%-30%	16199	1503	826	84		1.10	[0.86, 1.38]	0.50
≥30%	16199	1503	727	96		1.42	[1.13, 1.78]	4.52×10^{-3}

Odds ratio
Mutations in myeloid driver genes
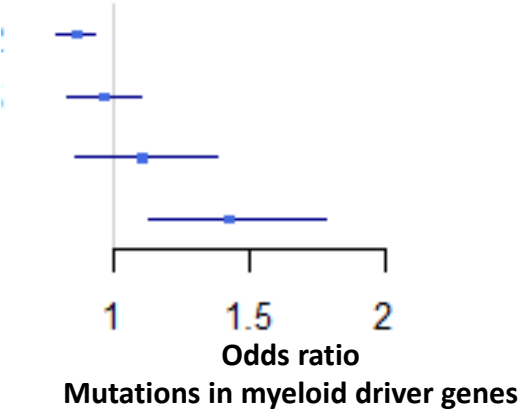
B)

LOY	Control		LOY			OR	95% CI	P_FDR
	No CH	CH	NO CH	CH				
<10%	17454	248	12598	180		1.01	[0.82, 1.22]	0.96
10%-20%	17454	248	3185	63		1.39	[1.04, 1.85]	0.04
20%-30%	17454	248	891	19		1.50	[0.88, 2.41]	0.17
≥30%	17454	248	801	22		1.93	[1.18, 3.01]	0.01

Odds ratio
Mutations in lymphoid genes
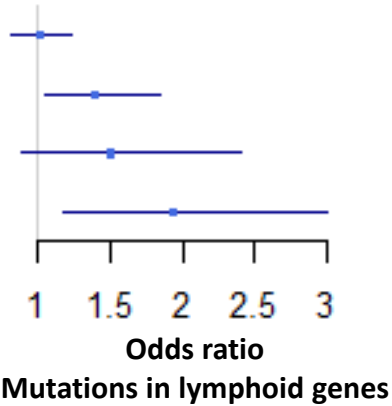
C)

LOY	Control		LOY			OR	95% CI	P_FDR
	No CH	CH	NO CH	CH				
<10%	10035	7667	7181	5597		1.02	[0.97, 1.07]	0.50
10%-20%	10035	7667	1722	1526		1.16	[1.08, 1.25]	3.16×10^{-4}
20%-30%	10035	7667	363	547		1.97	[1.72, 2.27]	2.03×10^{-22}
≥30%	10035	7667	286	537		2.46	[2.12, 2.86]	5.09×10^{-34}

Odds ratio
Mutations in unknown driver genes

Figure 5

Figure 6

Figure 7