Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer

Lars A Forsberg1,2, Chiara Rasi1,2, Niklas Malmqvist1–3, Hanna Davies1,2, Saichand Pasupulati1,2, Geeta Pakalapati1,2, Johanna Sandgren4, Teresita Diaz de Stähl4, Ammar Zaghlool1,2, Vilmantas Giedraitis5, Lars Lannfelt6, Joanna Score6, Nicholas C P Cross6, Devin Absher7, Eva Tiensuu Janson3, Cecilia M Lindgren8,9, Andrew P Morris8, Erik Ingelsson2,3, Lars Lind3 & Jan P Dumanski1,2

Incidence and mortality for sex-unspecific cancers are higher among men, a fact that is largely unexplained1,2. Furthermore, age-related loss of chromosome Y (LOY) is frequent in normal hematopoietic cells3,4, but the phenotypic consequences of LOY have been elusive5–10. From analysis of 1,153 elderly men, we report that LOY in peripheral blood was associated with risks of all-cause mortality (hazards ratio (HR) = 1.91, 95% confidence interval (CI) = 1.17–3.13; 637 events) and non-hematological cancer mortality (HR = 3.62, 95% CI = 1.56–8.41; 132 events). LOY affected at least 8.2% of the subjects in this cohort, and median survival times among men with LOY were 5.5 years shorter. Association of LOY with risk of all-cause mortality was validated in an independent cohort (HR = 3.66) in which 20.3% of subjects showed LOY. These results illustrate the impact of post-zygotic mosaicism on disease risk, could explain why males are more frequently affected by cancer and suggest that chromosome Y is important in processes beyond sex determination. LOY in blood could become a predictive biomarker of male carcinogenesis.

Peripheral blood DNA from 1,153 participants of the Uppsala Longitudinal Study of Adult Men (ULSAM) was genotyped using the high-resolution 2.5M HumanOmni SNP BeadChip from Illumina. The population-based ULSAM cohort has extensive phenotypic information available on naturally aging men who were clinically followed for >40 years. We studied DNA sampled in an age window of 70.7–83.6 years. Scoring of structural genetic variants was focused on post-zygotically acquired changes such as deletions, copy number–neutral loss of heterozygosity (CNNLOH, also called acquired uniparental disomy, AUPD) and gains, as described previously11–13, with a minimum size of 2 Mb. Twelve subjects had a history of hematological malignancy before sampling, and these individuals were analyzed separately to avoid mixed analyses of normal blood and malignant clones (Supplementary Figs. 1 and 2). In the remaining 1,141 participants, we uncovered 40 autosomal somatic structural variants of ≥2 Mb in length occurring in 37 subjects (3.2%), including 13 deletions, 16 CNNLOH events and 11 gains (Fig. 1 and Supplementary Table 1).

Strikingly, the most frequent somatic variant was LOY (Figs. 1 and 2). The degree of LOY was calculated for each subject from the median log R ratio values (measuring copy number) for approximately 2,560 probes in the male-specific region of chromosome Y (mLRR-Y) and suggested considerable interindividual differences in the proportion of cells with nullisomy Y. A conservative estimate of the frequency of LOY in the ULSAM cohort of 8.2% (93/1,141) was based on the lowest value (~0.139) in a simulated distribution of experimental variation in mLRR-Y values (Fig. 2). At this threshold, ≥18% of cells in affected participants would be expected to have nullisomy Y. To calculate the fraction of cells affected by nullisomy Y, we implemented a new approach using B allele frequency (BAF) values for pseudoautosomal region 1 (PARI) on chromosomes X and Y from SNP array data (explained in Supplementary Fig. 3).

Aberrations detected with Illumina 2.5M HumanOmni SNP arrays were validated using low-coverage (~5x) whole-genome next-generation sequencing, performed for 100 random participants. Of the 93 subjects with LOY in ≥18% of cells, sequencing was performed for 6, and all these cases of LOY were validated (Supplementary Figs. 4–7). Of the 37 subjects with autosomal events of ≥2 Mb in length, we performed whole-genome and/or whole-exome next-generation sequencing for 4 subjects with deletions, and all were validated (Supplementary Figs. 8–10). Moreover, the SNP array data suggested that some participants could have a gain of chromosome Y (GOY) (Fig. 2). However, next-generation sequencing did not confirm these observations in the three suspected cases of GOY that were sequenced (Supplementary Fig. 11). In summary, for autosomes and sex chromosomes in 1,141 subjects, we scored 133 somatic structural variants

1Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. 2Science for Life Laboratory, Uppsala University, Uppsala, Sweden. 3Department of Medical Sciences, Uppsala University, Uppsala, Sweden. 4Department of Oncology-Pathology, Cancer Center Karolinska, Karolinska Institutet, Stockholm, Sweden. 5Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden. 6Faculty of Medicine, University of Southampton, Southampton, UK. 7HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA. 8Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. 9Broad Institute of MIT and Harvard University, Cambridge, Massachusetts, USA. Correspondence should be addressed to L.A.F. (lars.forsberg@igm.uu.se) or J.P.D. (jan.dumanski@igp.uu.se).

Received 10 July 2013; accepted 3 April 2014; published online 28 April 2014; doi:10.1038/ng.2966
in 128 men (11.2%). Furthermore, participants with LOY did not show a difference in the frequency of autosomal structural variants compared to the rest of the cohort (data not shown), and we did not observe evidence of region-specific deletions on chromosome Y (Supplementary Figs. 5–7).

The effects of the identified structural variants on all-cause mortality, cancer-related mortality and non-cancer-related mortality were examined by Cox proportional hazards regression using the R package Survive. In survival analyses, we studied 982 participants free from cancer diagnosis before sampling (Supplementary Fig. 1), and survival was adjusted for nine potential confounders: age, hypertension, exercise, smoking, diabetes, body mass index (BMI), low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol and ancestry (Table 1 and Supplementary Table 2). Study entry was the date of DNA sampling, age was used as a timeline and the median follow-up time was 8.7 years (range of 0–20.2 years). Structural genetic variants were analyzed in three categories. LOY was analyzed separately as a continuous variable (mLRR-Y). Other deletions and CNNLLOH were grouped as autosomal loss of heterozygosity (LOH) variants, reflecting their similar biological effects.

The third category included autosomal gains. In primary survival analysis using the continuous estimator of LOY (mLRR-Y), men with a higher degree of LOY had increased risk of all-cause mortality (HR = 2.13, 95% CI = 2.04–2.22; Table 1). Further tests with the continuous estimator showed that LOY was a key risk factor for cancer-related mortality in the ULSAM cohort (HR = 3.76, 95% CI = 3.02–4.67; Table 1). However, LOY was not significantly associated with non-cancer-related mortality (95% CI = 0.62–1.34; Supplementary Fig. 1).

To plot the above results and to perform further exploratory tests, we scored participants on the basis of a defined threshold of mLRR-Y. Specifically, individuals with mLRR-Y of ≤–0.4 (Fig. 2) were scored as 1, and other subjects were scored as 0 (explained in Supplementary Fig. 12). These tests confirmed the effect of LOY on risk for all-cause mortality (HR = 1.91, 95% CI = 1.70–2.15; Fig. 3a) and showed that median survival times (with 50% probability of survival) in the group of men with LOY were 5.5 years shorter than for controls, representing half the survival time. The effect of LOY on risk of cancer-related mortality was also confirmed (HR = 3.29, 95% CI = 1.59–6.80; Fig. 3b). We further found that risk for non-hematological cancer mortality was clearly associated with LOY (HR = 3.62, 95% CI = 2.03–6.46; Supplementary Fig. 11). Moreover, risk of any cancer diagnosis, as well as risk of diagnosis with non-hematological cancer, was higher in participants with mLRR-Y of ≤–0.4 (HR = 2.47, 95% CI = 1.67–3.64; Fig. 3c). Therefore, risk of any cancer diagnosis, as well as risk of diagnosis with glioblastoma, was higher in participants with mLRR-Y of ≤–0.4 (HR = 2.47, 95% CI = 1.67–3.64; Fig. 3c). Moreover, risk of any cancer diagnosis, as well as risk of diagnosis with glioblastoma, was higher in participants with mLRR-Y of ≤–0.4 (HR = 2.47, 95% CI = 1.67–3.64; Fig. 3c).
We could not perform a valid test of the effect of LOY on the risk of mortality from hematological malignancies because only one participant scored with LOY died from such a cancer.

We replicated the result of LOY association with increased risk of all-cause mortality in an independent cohort of slightly younger males (age range of 69.8–70.7 years) from the PIVUS (Prospective Investigation of the Vasculature in Uppsala Seniors) study. The longest and median follow-up times were >10 and 7.0 years, respectively. Blood DNA from 488 was genotyped using the Illumina OmniExpress chip, and we scored aberrations on autosomes and sex chromosomes as described for the ULSAM cohort. The experimental noise from SNP genotyping was considerably lower in the PIVUS cohort, and, in scoring of LOY, 100 males (20.5%) had LOY ≥13% of nucleated blood cells (Supplementary Fig. 13). Furthermore, we scored 12 autosomal somatic aberrations of ≥2 Mb in length in the PIVUS cohort, including 6 gains, 3 deletions and 3 CNNLOH events (Supplementary Table 3). Cox proportional hazards regression with the same nine confounders was performed in the PIVUS cohort as described for the ULSAM cohort (Supplementary Table 4). Analysis using the continuous estimator of LOY (mLRR-Y) confirmed that men with a higher degree of LOY had increased risk of all-cause mortality (HR = 5.24, P = 0.022). Furthermore, after subjects were scored as 1 or 0 on the basis of LOY status (Supplementary Fig. 14), survival analysis showed that men with LOY had higher risk for all-cause mortality (HR = 3.66, 95% CI = 1.27–10.54; P = 0.016; number of events = 59). In summary, the results from the ULSAM cohort showing that LOY increased the risk for all-cause mortality were validated in the independent PIVUS cohort. It has recently been reported that it is not a gene desert. Furthermore, variation on chromosome Y has been associated with the expression of hundreds of autosomal and X-linked genes in Drosophila melanogaster18,19. Our results support the notion that human chromosome Y is important in biological processes beyond sex determination and reproduction. It has been known for more than half a century that elderly males frequently lose chromosome Y in normal hematopoietic cells3–4. The clinical consequences of this aneuploidy have been unclear, and the prevailing consensus suggests that this aberration should be considered to be phenotypically neutral and related to normal aging5–10. Furthermore, nullisomy Y has also been shown to occur in human cancer cells, and the literature provides a list of up to 20 different malignancies with LOY in combination with numerous other structural aberrations of the genome. Employing various methods, these studies showed that LOY often occurs at high frequency in tumors20–23. Recent comprehensive analysis showed that chromosome Y is one of the most commonly deleted chromosomes in the human cancer genomes24. Moreover, azoosperma is associated with higher cancer risk25, suggesting a connection between chromosome Y and cancer. Also, reintroduction of chromosome Y into a prostate cancer cell line suppresses its tumorigenicity in nude mice26. In summary, the above findings indicate that chromosome Y has a role in tumor suppression, and our results on expanding cell clones with LOY are in agreement with this hypothesis. Previously published data have shown that clonal expansions are common among the elderly, as a result of autosomal structural aberrations31,12,14,15. We examined the dynamics of changes in the proportion of cells with LOY by longitudinal analyses of five ULSAM subjects with data collected at two time points 6–14 years

somatic structural variants of autosomes in blood cells are connected with risk of hematological cancers, resulting in 5.5- to 35.4-fold higher risk12,13,15. Thus, the question of whether LOY is also connected with increased risk of hematological malignancies needs to be further investigated. Furthermore, Jacobs et al.12 found that autosomal somatic aberrations were more frequent in subjects with solid tumors (OR = 1.25), which is in agreement with the results for chromosome Y reported here.

Chromosome Y is recognized for its role in sex determination and normal sperm production, but it has long been considered to be a genetic wasteland, and, for several reasons, its characterization has lagged behind that of the rest of the genome. However, recent studies in humans and in a few other primates have shown that human chromosome Y contains a large number of genes16,17, indicating that it is not a gene desert.
Longitudinal LOY analyses in five elderly men showing progressive accumulation of cells containing LOY with increasing age. (a–e) ULSAM subjects 33 (a), 41 (b), 311 (c), 1655 (d) and 102 (e). Each subject was analyzed at two different ages, and the lower part of each panel shows a time axis with ages at genotyping, the time point of cancer diagnosis and cancer type as well as age and cause of death, when applicable. The red line in each panel shows mLRR-Y as estimated from SNP array experiments performed on blood collected at different ages in each subject. The text in red indicates the estimated percentage of nucleated blood cells affected with LOY. This number was calculated using MAD software from the analysis of SNP array data for PAR1 of chromosomes X and Y (Online Methods and Supplementary Fig. 3).

Figure 4

One of our most noteworthy findings is that men with LOY in peripheral blood are at risk of cancer outside the hematopoietic system (Fig. 3). These results could be explained by two hypotheses that are not mutually exclusive. First, a normal feature of the immune system is immune surveillance, which suppresses tumor development in other tissues. In participants with a high proportion of leukocytes with LOY, this normal function could be disturbed, leading to higher cancer risk. This explanation is supported by a higher frequency of cancer in individuals with a history of treatment with immunosuppressive drugs. The second hypothesis is that LOY also occurs in additional tissues in parallel with the chromosome Y aneuploidy in blood, stimulating the neoplastic proliferation of cancer progenitors outside the hematopoietic system. This hypothesis is in agreement with the briefly described frequent LOY observed in a wide range of cancers. We investigated the spectrum of cancer diagnoses in men scored with LOY in comparison to the rest of the cohort, finding no apparent differences (Supplementary Fig. 15). Regardless of the mechanism(s) underlying increased cancer risk in men with LOY, our results demonstrate the importance of post-zygotic aberrations, acquired during lifetime variation in normal cells, for the risk of cancer development. As mentioned, men have higher incidence and mortality rates for most sex-unspecific cancers, a fact that has largely been unexplained by known risk factors. Thus, our findings could help in explaining why males are more frequently affected by cancer. We further anticipate that extension of our study with respect to post-zygotic LOY in blood cells and possibly also in other tissues could in the future become a useful predictive biomarker of male carcinogenesis.

URLs. ULSAM (Uppsala Longitudinal Study of Adult Men), http://www2.pubcare.uu.se/ULSAM/invest/indexinv.htm; PIVUS (Prospective Investigation of the Vasculature in Uppsala Seniors), http://www.medsci.uu.se/ pivus; Survival R package, http://cran.r-project.org/web/packages/survival/index.html.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. Data on structural variants detected in this study are available at dbVar under accession nstd92.
ACKNOWLEDGMENTS

We thank G. Arnqvist and K. Lindblad-Toh for advice and F. Steele for review of the manuscript. This study was sponsored by funding from the Olle Enqvist Byggmästare Foundation to L.A.F. and by the Swedish Cancer Society, the Swedish Research Council, the Swedish Heart-Lung Foundation and funding from Sci-Life-Lab-Uppsala to J.P.D. Genotyping and next-generation sequencing were performed by the SNP&SEQ Technology Platform in Uppsala, Sweden, and were supported by Wellcome Trust grants WT0998017, WT064890 and WT090532 and by Uppsala University, Uppsala University Hospital, the Swedish Research Council and the Swedish Heart-Lung Foundation. The SNP&SEQ Technology Platform is part of the Science for Life Laboratory at Uppsala University and is supported as national infrastructure by the Swedish Research Council. C.M.L. is a Wellcome Trust Senior Research Fellow (grant 086596/Z/08/Z). A.P.M. is a Wellcome Research Career Development Fellow (grant 086596/Z/08/Z). A.P.M. is a Wellcome Trust Senior Research Fellow (grants WT098017, WT090532 and WT064890).

AUTHOR CONTRIBUTIONS

L.A.F. and J.P.D. conceived the study. L.A.F., C.R., E.T.J., E.I., L. Lind and J.P.D. were involved in study design. D.A., L. Lannfelt, V.G., C.M.L., A.P.M., E.I. and L. Lind provided materials, genotyping data and epidemiological data. H.D., S.P., G.P., A.Z. and J. Score performed wet-lab analyses. C.R., L.A.F., N.M., J. Sandgren, T.D.d.S. and J. Score implemented bioinformatics analyses. L.A.F. performed statistical analyses. L.A.F., C.R., E.I. and L. Lind were involved in survival analyses. L.A.F., N.C.P.C. and J.P.D. analyzed data. L.A.F. and J.P.D. coordinated the work and wrote the manuscript. All authors discussed the results and commented on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Cook, M.B., McGlynn, K.A., Devesa, S.S., Freedman, N.D. & Anderson, W.F. Sex disparities in cancer mortality and survival. Cancer Epidemiol. Biomarkers Prev. 20, 1629–1637 (2011).
2. Edgren, G., Liang, L., Adami, H.O. & Chang, E.T. Enigmatic sex disparities in cancer incidence. Eur. J. Epidemiol. 27, 187–196 (2012).
3. Jacobs, P.A., Brunton, M., Court Brown, W.M., Doll, R. & Goldstein, H. Change of human chromosome count distribution with age: evidence for a sex differences. Nature 197, 1080–1081 (1963).
4. Pierre, R.V. & Hoagland, H.C. Age-associated aneuploidy: loss of Y chromosome from human bone marrow cells with aging. Cancer 30, 889–894 (1972).
5. Nowinski, G.P. et al. The frequency of aneuploidy in cultured lymphocytes is correlated with age and gender but not with reproductive history. Am. J. Hum. Genet. 46, 1101–1111 (1990).
6. United Kingdom Cancer Cytogenetics Group. Loss of the Y chromosome from normal and neoplastic bone marrows. United Kingdom Cancer Cytogenetics Group (UKCCG). Genes Chromosom. Cancer 5, 83–88 (1992).
7. Wiktor, A. et al. Clinical significance of Y chromosome loss in hematologic disease. Genes Chromosom. Cancer 27, 11–16 (2000).
8. Wong, A.K. et al. Loss of the Y chromosome: an age-related or clonal phenomenon in acute myelogenous leukemia/myelodysplastic syndrome? Arch. Pathol. Lab. Med. 132, 1329–1332 (2008).
9. Wiktor, A.E., Van Dyke, D.L., Hodnefield, J.M., Eckel-Passow, J. & Hanson, C.A. The significance of isolated Y chromosome loss in bone marrow metaphase cells from males over age 50 years. Leuk. Res. 35, 1297–1300 (2011).
10. Jacobs, P.A. et al. Male breast cancer, age and sex chromosome aneuploidy. Br. J. Cancer 108, 959–963 (2013).
11. Forsberg, L.A. et al. Age-related somatic structural changes in the nuclear genome of human blood cells. Am. J. Hum. Genet. 90, 217–228 (2012).
12. Jacobs, K.B. et al. Detectable clonal mosaicism and its relationship to aging and cancer. Nat. Genet. 44, 651–658 (2012).
13. Schick, U.M. et al. Confirmation of the reported association of clonal chromosomal mosaicism with an increased risk of incident hematologic cancer. PLoS ONE 8, e59823 (2013).
14. Forsberg, L.A., Absher, D. & Dumanaki, J.P. Non-heritable genetics of human disease: spotlight on post-zygotic genetic variation acquired during lifetime. J. Med. Genet. 50, 1–10 (2013).
15. Laurie, C.C. et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat. Genet. 44, 642–650 (2012).
16. de Carvalho, C.M. & Santos, F.R. Human Y-chromosome variation and male dysfunction. J. Mol. Genet. Med. 1, 63–75 (2005).
17. Hughes, J.F. & Rozen, S. Genomics and genetics of human and primate Y chromosomes. Annu. Rev. Genomics Hum. Genet. 13, 83–108 (2012).
18. Lemos, B., Arariplo, L.O. & Hartl, D.L. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. Science 319, 91–93 (2008).
19. Lemos, B., Branco, A.T. & Hartl, D.L. Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. Proc. Natl. Acad. Sci. USA 107, 15826–15831 (2010).
20. Vijayakumar, S. et al. Detection of recurrent copy number loss at Yp11.2 involving TSPY gene cluster in prostate cancer using array-based comparative genomic hybridization. Cancer Res. 66, 4055–4064 (2006).
21. Zhang, L.J., Shin, E.S., Yu, Z.X. & Li, S.B. Molecular genetic evidence of Y chromosome loss in male patients with hematological disorders. Chin. Med. J. (Engl.) 120, 2002–2005 (2007).
22. Bianchi, N.D. Y chromosome structural and functional changes in human malignant diseases. Mutat. Res. 682, 21–27 (2009).
23. Veiga, L.C.S., Bergamo, N.A., Reis, P.P., Kowalski, L.P. & Rogatto, S.R. Loss of Y chromosomes does not correlate with age at onset of head and neck carcinoma: a case-control study. Braz. J. Med. Biol. Res. 45, 172–178 (2012).
24. Duijf, P.H., Schultz, N. & Benesra, R. Cancer cells preferentially lose small chromosomes. Int. J. Cancer 132, 2316–2326 (2013).
25. Eisenberg, M.L., Betts, P., Herder, D., Lamb, D.J. & Lipshultz, L.I. Increased risk of cancer among azoospermic men. Fertil. Steril. 90, 681–685 (2013).
26. Vijayakumar, S. et al. The human Y chromosome suppresses the tumorigenicity of PC-3, a human prostate cancer cell line, in athymic nude mice. Genes Chromosom. Cancer 44, 365–372 (2005).
27. Dunn, G.P., Bruce, A.T., Ikeda, H., Old, L.J. & Schreiber, R.D. Cancer immunoeediting: from immunosurveillance to tumor escape. Nat. Immunol. 3, 991–998 (2002).
28. Vial, T. & Descotes, J. Immunosuppressive drugs and cancer. Toxicology 185, 229–240 (2003).
29. Vajdic, C.M. & van Leeuwen, M.T. Cancer incidence and risk factors after solid organ transplantation. Int. J. Cancer 125, 1747–1754 (2009).
Online Methods

Studied cohorts. The ULSAM study (Uppsala Longitudinal Study of Adult Men) was initiated in 1970 (ref. 30), with 2,322 men born in Uppsala in 1920–1924 participating at the age of 50. The study is investigating a wide range of phenotypes, including cancer history from the National Cancer Registry and the Swedish Civil Registry. Major reexaminations have been undertaken at ages 60, 70, 77, 82 and 88 years. Here we have used DNA sampled in the age window of 70.7–83.6 years, and the longest follow-up time was >20 years. DNA was controlled for quality as described31.

The PIVUS study (Prospective Investigation of the Vasculature in Uppsala Seniors) was begun in 2001 with the primary aim of investigating the predictive power of various measurements of endothelial function and arterial compliance. Eligible participants were all aged 70 and were living in the community of Uppsala, Sweden. The subjects were randomly chosen from the community register, and 1,016 men and women participated. Two reinvestigations of the cohort were undertaken, starting in the spring of 2006 and in the spring of 2011 at the ages of 75 and 80 years, respectively31. Here we have used 488 male DNA samples collected at the age of 70 years, which were successfully genotyped. The longest and median follow-up times were >10 and 7.0 years, respectively. Survival analysis included the same nine confounders as described for the ULSAM cohort (age at sampling, hypertension, exercise habits, smoking, diabetes, BMI, LDL cholesterol, HDL cholesterol and education level). All of these factors, except smoking, were analyzed as for the ULSAM cohort. In PIVUS, two classes of smokers were defined, current smokers and former smokers or non-smokers. The ULSAM and PIVUS studies were approved by local research ethics committees, and participants have given their informed consent.

Scoring of structural aberrations. SNP genotyping in the ULSAM and PIVUS cohorts was performed using 2.5M HumanOmni- and HumanOmniExpress BeadChips, respectively, according to the recommendations of the manufacturer. Illumina genotyping results passed strict quality control, as described31. Output files were analyzed using Nexus-Copy-Number-6.1 (BioDiscovery), which applies the rank segmentation algorithm, based on circular binary segmentation32, as described31. The applied version, SNPRank segmentation, an extended algorithm in which LRR as well as BAF is included in the segmentation32, as described11. The applied version, SNPRank segmentation, an extended algorithm in which LRR as well as BAF is included in the segmentation process, generated both copy number and allelic event calls. The size cut-off for the scoring of autosomal aberrations was set to ±2 Mb to minimize the false classification of constitutional structural variants as somatic events32. All copy number variant calls made by the software were manually inspected by two investigators before the data were entered into a MySQL database. Circos plots were used for visualization (Fig. 1 and Supplementary Fig. 2)33.

Estimations of LOY frequency in the cohorts and the fraction of cells with nullisomy Y. In the ULSAM cohort, LOY was scored from the median LRR values of approximately 2,560 SNP probes (2.5M chip) in the male-specific region of chromosome Y (mLRR-Y) in the 56-Mb region between pseudoautosomal regions 1 and 2 (PAR1 and PAR2) on chromosome Y (chr. Y: 2,694,521–59,034,049, hg19/GRCh37)16,17. In the PIVUS cohort, analogous scoring was performed using the 1,690 SNP probes in the region on the HumanOmniExpress chip. Total variation in mLRR-Y in the cohorts (Fig. 2 and Supplementary Fig. 13) consisted of a signal from LOY and a signal from experimental variation. To estimate LOY frequencies, the contribution from experimental factors was estimated. We assumed that experimental variation in mLRR-Y was distributed in a non-skewed fashion and that variation in the positive tail of the mLRR-Y distribution was all experimentally induced (gray bars in Fig. 2 and Supplementary Fig. 13). The latter assumption was reasonable, as validation experiments did not confirm all suspected cases of GOY (Supplementary Fig. 11). Experimental noise (white bars in Fig. 2 and Supplementary Fig. 13) was generated by imposing the observed variation in the positive tail into a reflected negative tail in two steps. First, the peak in the histogram of mLRR-Y (thick red line in Fig. 2 and Supplementary Fig. 13) was determined with a kernel density estimation method (kernel medial), using the Density function in R. Specifically, the bandwidth SJ was used, and the bin with the highest value (i.e., distribution peak) in the smoothed distribution was selected as the local median14. Next, the observations in the positive tail were mirrored over the kernel-derived median to create the negative tail. Finally, for the most conservative estimate of the frequency of LOY, the lowest value in the simulated noise distribution was used as the threshold (mLRR-Y values of −0.139 in ULSAM and −0.154 in PIVUS).

The percentage of cells containing a structural variant was calculated using MAD software33. MAD delivered the R package R-GADA36 and is used in SNP data analysis to identify regions containing deletions, gains and CNNOH events. R-GADA detects allelic imbalances caused by genetic abnormalities using a circular binary segmentation–based BAF segmentation algorithm37, producing calls from the BAF values typical for heterozygous probes (0.5). The extent of deviation (Bdev) is a measure of the percentage of mosaic cells, calculated using the published formula38. To calculate the fraction of cells with nullisomy Y, we implemented a new approach using BAF values from PAR1 on chromosomes X and Y (explained in Supplementary Fig. 3). We took advantage of LRR and BAF values derived from PAR1 of chromosomes X and Y, which was possible because chromosome X was never involved in a region-specific aberration in the 1,153 participants (Fig. 1). It was not feasible to use the MAD algorithm to calculate the percentage of cells in the PIVUS cohort, as in the ULSAM cohort, because of the few SNP probes in the PAR1 region on the platform used (HumanOmniExpress). This platform has 35 SNP probes in PAR1 in comparison to the 1,111 SNP probes in the same region on the 2.5M HumanOmni array. However, after adjusting for differences in median LRR between the two cohorts, the percentages of cells at the thresholds in the PIVUS cohort (23.6% at −0.5 and >13% at −0.154) could be estimated as in the ULSAM cohort (Supplementary Fig. 3).

Validation using next-generation sequencing. Sequencing libraries were prepared according to TruSeq DNA v2 sample preparation EUC 15026489 revA using reagents from the TruSeq DNA v2 Sample Prep Kit Set A and Set B (Illumina). A 14 pmol solution of five to six DNA libraries pooled in equimolar amounts was subjected to cluster generation on a cBot instrument (Illumina) using the TruSeq PE Cluster Kit v3. Paired-end sequencing was performed for 100 cycles using a HiSeq 2000 instrument (Illumina) with TruSeq SBS Chemistry v3, according to the manufacturer’s protocols. Base calling was performed on the instrument with RTA 1.13.48, and the resulting BCL files were filtered, demultiplexed, allowing for one mismatch base, and converted to FASTQ format with tools provided by CASAVA 1.8 (Illumina). Copy number profiles were computed from next-generation sequencing data using ControlFREEC software39. Sequence read depth was used to calculate an equivalent to LRR using a sliding window approach. We applied a fixed window size of 5 kb, and the other settings were default. The next-generation sequencing data set was composed of 100 whole genomes with Sx average coverage and additional exomes with an average coverage of 17×.

Survival and statistical analyses in the ULSAM cohort. Cox proportional hazards regression models were generated in R using the Survival package. The effects on survival of three types of structural genetic variants were tested, also adjusting for nine confounders (Supplementary Table 2). LOY was the first category. In the primary survival analyses, a continuous explanatory variable was used as a proxy for LOY (mLRR-Y). To plot the results and perform further exploratory tests, subjects were scored on the basis of a defined threshold of mLRR-Y. Specifically, individuals with mLRR-Y of ≤0.4 (Fig. 2) were scored as 1, and other subjects were scored as 0 (explained in Supplementary Fig. 12). The second type of structural variation investigated in the survival analyses included autosomal deletions and CNNOH events, grouped together as autosomal LOH mutations. The third category was autosomal gains. LOH events and gains were analyzed as binary variables. Study entry was the date of DNA sampling, age was used as a timeline and the median follow-up time was 8.7 years, with a range of 0–20.2 years. All participants with a history of cancer before sampling (n = 171) were excluded from survival analyses as were samples not passing quality control. The remaining 982 subjects were included in analyses using various endpoints (Supplementary Fig. 1). We used the statistical software R (versions 2.15.0–3.0.1)40 for data mining and statistical analyses. The described survival and statistical analyses in the ULSAM cohort were performed in an analogous manner in the independent PIVUS cohort (Supplementary Figs. 13 and 14, and Supplementary Tables 3 and 4).
30. Hedstrand, H. A study of middle-aged men with particular reference to risk factors for cardiovascular disease. Ups. J. Med. Sci. Suppl. 19, 1–61 (1975).
31. Lind, L., Fors, N., Hall, J., Marttala, K. & Stenborg, A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. Arterioscler. Thromb. Vasc. Biol. 25, 2368–2375 (2005).
32. Olishen, A.B., Venkatraman, E.S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5, 557–572 (2004).
33. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645 (2009).
34. Sheather, S. & Jones, M. A reliable data-based bandwidth selection method for kernel density estimation. J. R. Stat. Soc. Ser. A Stat. Soc. 53, 683–690 (1991).
35. González, J.R. et al. A fast and accurate method to detect allelic genomic imbalances underlying mosaic rearrangements using SNP array data. BMC Bioinformatics 12, 166 (2011).
36. Pique-Regi, R., Caceres, A. & Gonzalez, J.R. R-Gada: a fast and flexible pipeline for copy number analysis in association studies. BMC Bioinformatics 11, 380 (2010).
37. Pique-Regi, R. et al. Sparse representation and Bayesian detection of genome copy number alterations from microarray data. Bioinformatics 24, 309–318 (2008).
38. Rodriguez-Santiago, B. et al. Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. Am. J. Hum. Genet. 87, 129–138 (2010).
39. Boeva, V. et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28, 423–425 (2012).
40. R Development Core Team. R: A Language and Environment for Statistical Computing (Vienna, 2010).