



Longitudinal changes in the frequency of mosaic chromosome Y loss in peripheral blood cells of aging men varies profoundly between individuals

Marcus Danielsson¹ · Jonatan Halvardson¹ · Hanna Davies¹ · Behrooz Torabi Moghadam¹ · Jonas Mattisson¹ · Edyta Rychlicka-Buniowska^{1,2} · Janusz Jaszczynski³ · Julia Heintz¹ · Lars Lannfelt⁴ · Vilmantas Giedraitis⁴ · Martin Ingelsson⁴ · Jan P. Dumanski^{1,2} · Lars A. Forsberg^{1,5}

Received: 8 July 2019 / Revised: 10 October 2019 / Accepted: 13 October 2019 / Published online: 25 October 2019
© The Author(s) 2019. This article is published with open access

Abstract

Mosaic loss of chromosome Y (LOY) is the most common somatic genetic aberration and is associated with increased risk for all-cause mortality, various forms of cancer and Alzheimer's disease, as well as other common human diseases. By tracking LOY frequencies in subjects from which blood samples have been serially collected up to five times during up to 22 years, we observed a pronounced intra-individual variation of changes in the frequency of LOY within individual men over time. We observed that in some individuals the frequency of LOY in blood clearly progressed over time and that in other men, the frequency was constant or showed other types of longitudinal development. The predominant method used for estimating LOY is calculation of the median Log R Ratio of probes located in the male specific part of chromosome Y (mLRRY) from intensity data generated by SNP-arrays, which is difficult to interpret due to its logarithmic and inversed scale. We present here a formula to transform mLRRY-values to percentage of LOY that is a more comprehensible unit. The formula was derived using measurements of LOY from matched samples analysed using SNP-array, whole genome sequencing and a new *AMELX/AMELY*-based assay for droplet digital PCR. The methods described could be applied for analyses of the vast amount of SNP-array data already generated in the scientific community, allowing further discoveries of LOY associated diseases and outcomes.

These authors contributed equally: Jan P. Dumanski, Lars A. Forsberg

Supplementary information The online version of this article (<https://doi.org/10.1038/s41431-019-0533-z>) contains supplementary material, which is available to authorised users.

✉ Marcus Danielsson
marcus.danielsson@igp.uu.se

✉ Lars A. Forsberg
lars.forsberg@igp.uu.se

¹ Department of Immunology, Genetics and Pathology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden

² Faculty of Pharmacy and 3P Medicine Laboratory, International Research Agendas Programme, Medical University of Gdańsk, Gdańsk, Poland

³ Department of Urology, Maria Skłodowska-Curie Memorial Cancer Centre and Institute of Oncology, Kraków Branch, Kraków, Poland

⁴ Department of Public Health and Caring Sciences/Geriatrics, Uppsala University, 751 85 Uppsala, Sweden

⁵ Beijer Laboratory of Genome Research, Uppsala University, Uppsala, Sweden

Introduction

Mosaic loss of chromosome Y (LOY) refers to chromosome Y aneuploidy acquired during lifetime and it is the most common post-zygotic variant described in human blood cells, causing the absence of almost 2% of the haploid nuclear genome [1]. For over 50 years it has been known that LOY is frequent in cells of the hematopoietic system [2], but LOY in leukocytes was long viewed as a neutral event related to normal aging without phenotypical consequences [3]. However, recent studies suggest that the opposite as LOY has been found to be associated with increased risk for all-cause mortality [4, 5] as well as a growing list of disease such as various forms of cancer [4, 6–10], autoimmune conditions [11, 12], Alzheimer's disease [13], major cardiovascular events [5, 14], schizophrenia [15], diabetes [5] as well as age-related macular degeneration (AMD) [16]. As a male specific genetic risk factor for common disease, LOY in leukocytes might help explain why men live shorter lives [1, 4, 13, 17], likely as

an effect of compromised immune system functions in circulating immune cells without chromosome Y [18].

In single cells LOY is a binary event, and when measured in bulk DNA samples collected from peripheral blood, it is manifested as a continuous mosaicism ranging from zero to 100% of cells without a Y chromosome. Recent studies have established that the frequency of LOY in leukocytes increases with age, occurring in at least 10% of peripheral blood cells in about 5–10%, 15–20% and 20–40% of aging men around 60, 70 and 80 years of age, respectively [4, 8, 13, 19–21]. Furthermore, in a cohort of 93 year old men, 57% of the individuals had lost the Y chromosome in more than 10% of the leukocytes [22]. Although age is a strong risk factor, occurrence and phenotypic effects associated with LOY in blood cells have also been reported in younger men [9, 16, 20]. LOY has also been described in other non-cancerous tissues, such as ectodermally-derived buccal mucosa [22] and atherosclerotic plaque [14], although typically in lower frequency than in haematopoietic cells. In addition to age, known risk factors for LOY during lifetime include smoking [8, 16, 19, 23], exposure to air pollution [24] as well as genetic background [8, 19, 21]. Further studies are needed for additional insights regarding its variation between individuals, its prevalence in different tissues, its changes in frequency over time within tissues (and in different cell populations within tissues) as well as its functional and phenotypic effects during the entire lifespan of men.

Measurements of LOY mosaicism from DNA can be performed using technologies such as karyotyping, qPCR, DNA-arrays and next generation sequencing (NGS) (Supplementary Table 1). Recent studies have successfully re-analysed data generated with SNP-arrays in various genome wide association studies (GWAS) and described profound phenotypic effects associated with LOY in leukocytes [4, 5, 8, 10, 13, 14, 16]. During the last decades, several large-scale genome projects have characterised the human genome and generated data that are suitable for analysis of occurrence of somatic structural variants and aneuploidies. Hence, available data could be reanalysed to further investigate associations between LOY (and other disease-associated somatic mutations/variants) in leukocytes with various human diseases and outcomes.

LOY mosaicism is straightforward to quantify from SNP-array data by calculation of a continuous variable called mLRRY, as further explained in the Methods and elsewhere [4]. The mLRRY-value in subjects without LOY are close to zero and it decreases with increasing level of LOY mosaicism and this relationship is a shortcoming for intuitive interpretation of the mosaicism. To solve this problem, we present here a formula to transform mLRRY-values into a more intuitive unit, i.e. the percentage of cells with the aneuploidy. We applied this transformation in

comprehensive analyses of serially collected samples from aging men to characterise a previously unknown intra-individual variation of changes in the frequency of LOY within the blood of individuals studied over time. We further present a new effective method for estimation of LOY mosaicism based on quantification of the relative number of X and Y chromosomes using droplet digital PCR (ddPCR). The assay is targeting a 6 bp sequence difference present between the *AMELX* and *AMELY* genes using TaqMan-probes.

Materials and methods

Samples and DNA extraction

We analysed DNA from peripheral blood samples collected from participants of the cohort Uppsala Longitudinal Study of Adult Men (ULSAM, www.pubcare.uu.se/ulsam). For the longitudinal analyses, all available serially collected samples from ULSAM were included. This dataset comprised of 798 DNA samples collected from 276 men (median age = 81.9, range = 70–93) sampled 2–5 times over a period of up to 22.2 years (Supplementary Fig. 1). Furthermore, 121 DNA samples collected at 93 years of age from a subset of ULSAM participants were used for the pairwise analyses of LOY using three independent technologies (Supplementary Fig. 1). DNA was extracted from samples of peripheral blood nucleated cells using the QIAamp DNA Blood kit (51194, Qiagen) according to the manufacturer's instructions. The study had been approved by the Regional Ethical Committee in Uppsala, Sweden (reference numbers, i.e. dnrs: 02-018, 02-605, 2007/338 and 2013/350). All study participants provided written informed consent.

Measurements of LOY using three independent technologies

For comparing LOY measurements from different methods, we analysed the same set of DNA samples using three technologies, i.e. SNP-arrays ($n = 121$), whole genome sequencing (WGS, $n = 26$) and droplet digital PCR (ddPCR, $n = 121$). Description of how LOY estimations were performed using each readout is provided below. Briefly, for SNP-array data, the mLRRY variable was calculated as a median intensity of the probes located in the male-specific region of chromosome Y (MSY). For WGS, the frequency of cells with LOY was estimated from the ratio between the read depth on chromosome Y in relation to the full genome. For ddPCR, quantification of the relative number of X and Y chromosomes was performed by targeting a 6 bp sequence difference present between the *AMELX* and *AMELY* genes.

LOY from SNP-array data

The estimation of level of LOY mosaicism from SNP-array data was performed by calculation of the mLRRY from data generated by Illumina BeadChips (Illumina Inc., CA, USA) as previously described [4]. For each experiment passing strict quality control [13], the mLRRY-value was calculated as the median of the Log R Ratio (LRR) of the probes located in the MSY (chrY: 2.781.480–56.887.902, hg19/GRCh38.p12). The different versions of arrays used each contain sufficient number of probes for robust calculation of mLRRY (Supplementary Table 2). The mLRRY-values calculated for every sample was corrected for batch effects by adjustment using the local regression median of mLRRY-values from a kernel density estimation, using the density function in R on a smooth histogram of mLRRY-values, as previously described [4].

LOY from whole genome sequencing (WGS) data

Estimation of LOY mosaicism from WGS data was performed by comparing the read depth on chromosome Y in relation to the full genome. Sequencing libraries were prepared using the truseq Nano DNA sample preparation kit (TFC-121-4001/4002, Illumina Inc) extracting 100 ng DNA for each sample. Sequencing libraries were run on an Illumina HiSeq X instrument (version 2.5 sequencing chemistry) and sequenced to a depth of 30×. Each sequenced library had a read length of 150 bp with an insert size of 350 bp. Sequencing reads were aligned to the GRCh37 human reference genome with the BWA aligner (version 0.7.12). Copy number for chromosome Y was estimated by the Control-Freec software using read counts in non-overlapping windows across the genome. These were fitted by the GC content and mappability information and the median ploidy for the Y chromosome was calculated [25].

LOY from droplet digital PCR (ddPCR) data

Estimation of LOY with ddPCR was performed using Bio-Rad's QX200 Droplet Digital PCR System (Bio-Rad Laboratories, Inc., CA, USA). A TaqMan-based method was developed and used to quantify the relative number of X and Y chromosomes in a sample by targeting a 6 bp sequence difference present between the *AMELX* and *AMELY* genes (Supplementary Fig. 2). An advantage of this protocol compared with previous qPCR-based methods, is that the PCR amplification of the two target-genes (i.e. the test and reference loci on chromosomes Y and X, respectively) are performed using the same primer pair and would thus be relatively unbiased with regard to primer-properties. Primers and probes was purchased from Thermo Fisher Scientific (MA, USA)

article number C_990000001_10. DNA samples with concentrations ranging between 300 and 20 ng/μl were digested for 15 min in 37 °C with HindIII (Thermo Fischer, article number: #FD0504) and diluted with an equal volume of water. Subsequently 50 ng of the digested and diluted DNA sample was mixed in PCR supermix for probes without dUTP (BioRad, article number: 186–3023) together with TaqMan primers and probes. Following manufacturer's instructions, droplets were generated and PCR amplified using the following conditions: 95 °C for 10 min, 40 cycles of 94 °C for 30 s and 60 °C for 1 min. The PCR programme ended with 98 °C for 10 min and a 10 °C hold. The fluorescence of FAM (targeting *AMELY*) and VIC (for *AMELX*) was analysed for each droplet using a droplet reader and the ratio *AMELY/AMELX* was analysed using Bio-Rad's software QuantaSoft (version 1.7.4.0917). All samples were run in duplicates and the standard deviation of the measured ratios was calculated. Samples were re-analysed if the standard deviation was 1.2 or higher.

Transformation of mLRRY-values into percentage of cells with LOY

We developed a formula for conversion of mLRRY-values estimated from SNP-array data into the unit LOY (%) by taking advantage of LOY-estimates from pairwise studied DNA samples. First we determined that the percentage of LOY estimated in samples studied by WGS and ddPCR ($n = 26$) yielded essentially identical results. The readouts from these technologies could therefore be used as reference points for establishing a relationship between mLRRY and LOY (%). The latter was performed in parallel and independently using the data generated from pairwise studied samples using ddPCR and SNP-array ($n = 121$) as well as from samples analysed using WGS and SNP-array ($n = 26$) (Supplementary Fig. 3). Specifically, the mLRRY-values were first antilogged (2^{mLRRY}) and correlated with data generated from the same samples by WGS ($n = 26$) or ddPCR ($n = 121$). Calculation of power equations resulted in the relationships $y = 0.9242 \cdot (2^x)^{1.7703}$ and $y = 0.9695 \cdot (2^x)^{1.8779}$ (describing the relationships between level of LOY estimated by mLRRY and percentage of LOY estimated by WGS or ddPCR, respectively). The constants in these relationships were rounded to the nearest integer and used to adjust the antilog mLRRY, resulting in two equations: (1) *percent of cells with a Y chromosome* = $100 \cdot 2^{2 \cdot \text{mLRRY}}$ and (2) *LOY (%)* = $100 \cdot (1 - 2^{-2 \cdot \text{mLRRY}})$.

Statistical analyses

Calculations of Pearson's coefficient of determination (R^2) and linear regression models were performed using R. Standardised beta-values (β) was calculated using the R library lm.beta [26].

Results

Longitudinal analyses of LOY frequency in peripheral blood of 276 aging men

Analyses of the serially collected samples in the ULSAM study showed an overall higher level of LOY mosaicism in samples collected at higher ages (Fig. 1). The age-related accumulation of LOY was significant in linear regression models using the continuous mLRRY as response variable ($\beta = -0.21$, $p < 0.0001$) as well as using the new LOY (%) unit ($\beta = 0.19$, $p < 0.0001$). Furthermore, the serial analysis revealed a previously undescribed profound inter-individual variation in the changes of frequency of LOY in blood over time. For example, in about 1/3 of the individuals, the level of LOY mosaicism clearly increased with age, i.e. LOY progressors. In other subjects, the level did not change substantially during the follow-up time (Fig. 1b–d). Furthermore, in a few subjects the level of mosaicism decreased during the study and more complex and miscellaneous patterns were also observed, such as initial increase followed by a decrease but also the reverse with an initial decrease followed by increased level of mosaicism (Fig. 1d). The dotted line in panel Fig. 1b marks a threshold where 30% of the blood cells are without the Y chromosome and we identified 65 individuals that in at least one time point had a level of LOY on or above this threshold. In this subset, an increased level of LOY over time could be observed in a majority of subjects (Fig. 1c) while other subjects displayed different patterns (Fig. 1d). The longitudinal changes in the frequency of LOY mosaicism within each of the 276 studied individual subjects is provided in Supplementary Fig. 4.

New formula for transformation of mLRRY to percentage of LOY

We used data generated by SNP-array, whole genome sequencing (WGS) and droplet digital PCR (ddPCR targeting a sequence difference between the *AMELY* and *AMELX* genes) to estimate the level of LOY in DNA from whole blood samples. The LOY estimate from each experiment is provided in Supplementary Table 3. This dataset of pairwise analysed samples made it possible to compare measurements of LOY in individual samples estimated using the three independent technologies (Fig. 2). These comparisons exposed a non-linear estimation of LOY using the mLRRY calculated from SNP-array data (Fig. 2a, b) as well as a linear estimation of LOY by WGS and ddPCR. Specifically, a close to perfect fit to a linear regression line was observed between readouts from the same samples using WGS and ddPCR (Pearson's coefficient of determination, $R^2 = 0.998$, $n = 26$) (Fig. 2c).

Comparing the level of LOY estimated in the samples analysed using SNP-array and WGS ($n = 26$) as well as in the samples analysed using SNP-array and ddPCR ($n = 121$) also showed concordance with respect to level and direction of mosaicism (Fig. 2a, b). However, in contrast to the linear correlation between WGS and ddPCR readouts, a non-linear relationship (likely due to the logarithmic scale of the LRR-values) was observed and the fit to a linear regression line between the readouts from these technologies and mLRRY from SNP-arrays was lower ($R^2 = 0.896$ and 0.849 , respectively). Transformation of mLRRY-values into percentage of LOY using the equation $\text{LOY} (\%) = 100 * (1 - 2^{-2^{\text{mLRRY}}})$ resulted in improved fit to linearity ($R^2 = 0.965$ and 0.959 , respectively) (Fig. 3a, b).

To evaluate the ability of the formula to predict biologically relevant levels of LOY mosaicism we first tested it using a range of theoretically possible mLRRY-values representing varying degree of LOY (Fig. 4a). For comparison, we also applied another published formula [27] on the same dataset (i.e. Veitia's formula $F_{(\text{LOY})} = 1.8(1 - 2^{-\text{mLRRY}}) + 0.015$). The evaluation showed that the two formulas generates similar predictions of mosaicism at low levels of LOY but at higher levels of mosaicism, only the formula presented here asymptotically approaches the theoretical maximum of 100% mosaicism. In Fig. 4b, the ability of the formulas are assessed using authentic datasets generated from 121 samples studied with both SNP-array and ddPCR. Also in these comparisons, Veitia's formula tend to overestimate the level of LOY in samples with high levels of mosaicism, while the formula presented here predicts LOY (%) within relevant boundaries.

Discussion

During human lifespan, somatic cells acquire various forms of post-zygotic genetic variants that occasionally mediates a proliferative advantage to affected cell(s) [1, 20, 28–37]. Such processes of cell expansion are often referred to as clonal haematopoiesis (CH or CHIP) when affecting blood cell progenitors. A more general term to describe such cellular progressions, useful for all types of cells and not restricted to events in the haematopoietic system, is aberrant clonal expansions (ACE) [1]. LOY in leukocytes is the most common form of ACE showing a clear increase in frequency with age in the general population [4, 8, 13, 19–22]. The serial sampling applied in the present study revealed a previously unknown and profound variation in the dynamics of LOY-clone evolution within the peripheral blood of individual subjects over time. The participants of ULSAM study have been followed clinically for almost 50 years and blood samples have been collected up to five times from the

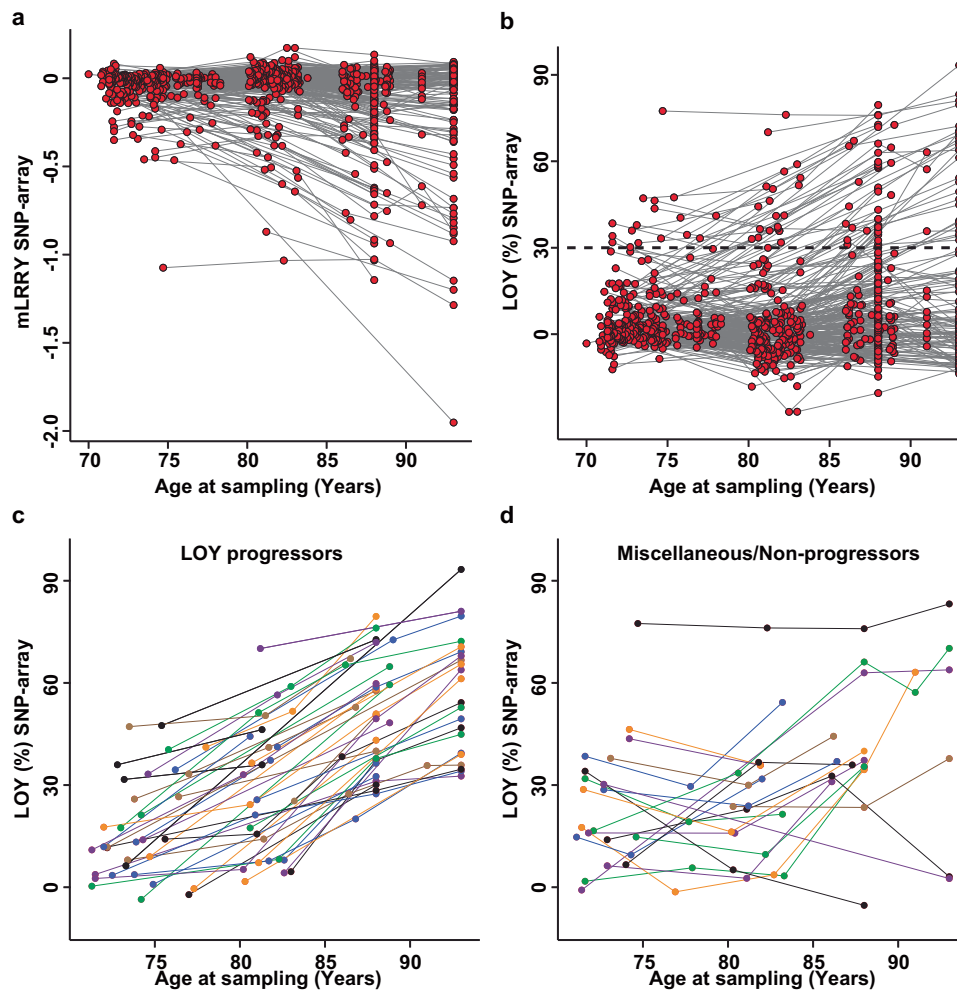


Fig. 1 Results from longitudinal analyses of LOY mosaicism in whole blood DNA from 276 aging individuals from the ULSAM cohort. Subjects were sampled serially 2–5 times over a period of up to 22.2 years. Every point represents a measurement of LOY in a subject at one time point with the level of LOY estimated from SNP-array data at the Y-axes and age of sampling on the X-axes. The dataset before (a) and after transformation (b) of the mLRRY-values using the equation: $LOY (\%) = 100 \times (1 - 2^{mLRRY})$. Grey lines connect the LOY measurements from the same individual at different time points. The dotted black line in b indicates a level of LOY mosaicism at which at least

30% of the nucleated blood cells are without a Y chromosome. This threshold was used to identify subjects displaying a high level of LOY at any time point during the study and the longitudinal changes in the frequency of LOY in this subset are displayed in c and d. To visualise changes in LOY frequency over time, points and lines were colour-coded to connect multiple measurements from the same individual. c Displays the subjects showing a clear progression in the frequency of LOY over time and d shows non-progressing individuals with miscellaneous types of longitudinal change

same participants during the last decades. We observed that in certain subjects, ACEs with LOY clearly progressed with time and that in others, the frequency of LOY cells did not change substantially with age or showed other types of LOY-clone trajectories (Fig. 1 and Supplementary Fig. 4). The possible mechanism(s) behind these profound differences between individuals is currently not known but could be related to variation in proliferative rates and longevity of progenitor cells giving rise to LOY-clones. Furthermore, variation in exposures to external risk factors (smoking, air pollution etc.) as well as other possible confounders (disease, therapies etc.) could explain part of the observed variation in progression patterns, a topic that require further

studies. Moreover, some individuals in the longitudinal dataset showed indications of low-frequency mosaic gain of chromosome Y (GOY) at several measured time points (Fig. 1). It is not clear if all of these observations are of biological origin or represent technical variation. However, in one individual, GOY was detected using both SNP-array and ddPCR (Figs. 2b and 3b, i.e. subject 207 in Supplementary Fig. 4). To our knowledge, no phenotypic consequences from mosaic gain of chromosome Y in leukocytes has been described.

It should also be noted that the samples studied here was bulk DNA extracted from whole blood samples. It is therefore unknown which blood cell type(s) that were

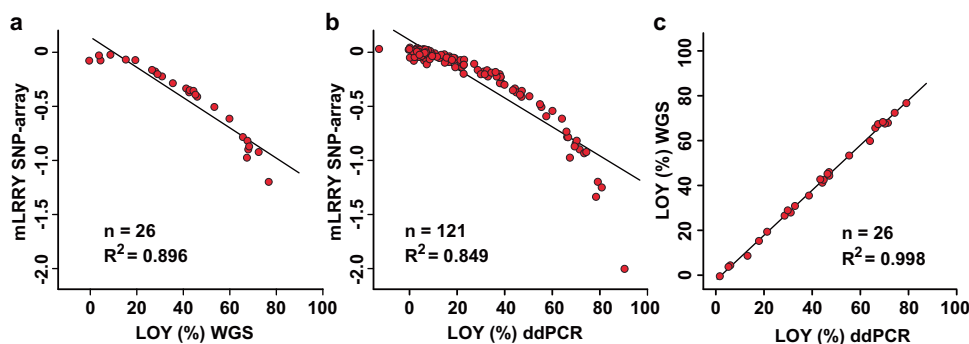


Fig. 2 Illustration of the non-linear estimation of LOY mosaicism by mLRRY calculated from SNP-array data by comparisons with LOY estimates generated from the same set of samples using independent technologies. **a**, **b** show the comparisons between mLRRY calculated from SNP-array data with the corresponding LOY estimates generated

from the pairwise studied samples using whole genome sequencing (WGS) and droplet digital PCR (ddPCR), respectively. **c** displays a high concordance between estimates of LOY in samples pairwise studied with WGS and ddPCR. A linear regression line with Pearson's coefficient of determination (R^2) is shown for each comparison

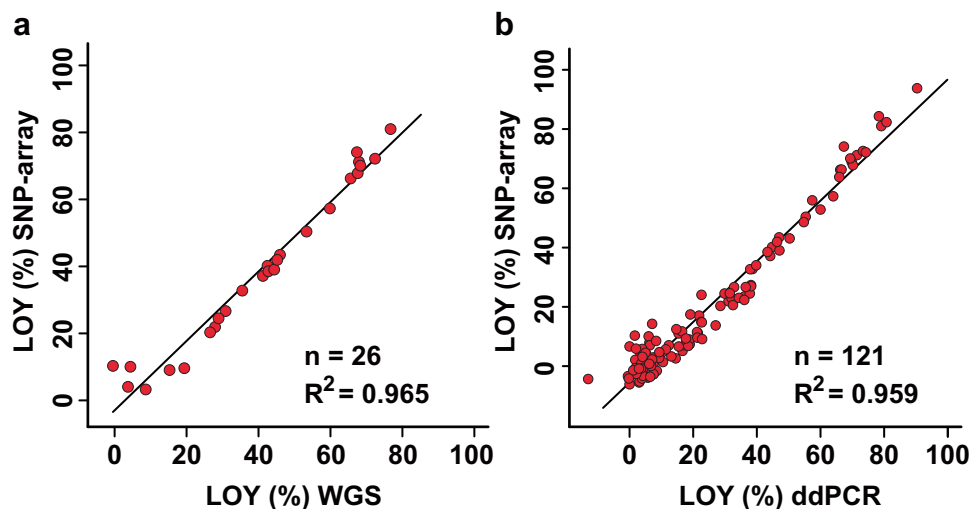


Fig. 3 Transformation of mLRRY-values linearises LOY estimates from SNP-array data using the equation: $LOY (\%) = 100 \times (1 - 2^{2mLRRY})$. **a** displays estimations of the level of LOY from 26 pairwise studied samples using SNP-array and whole genome sequencing (WGS). The Y-axis show the predicted LOY (%) from SNP-array data

using the above formula and the X-axis display the measured level of LOY using WGS. A linear regression line with Pearson's coefficient of determination (R^2) is shown. **b** shows a corresponding comparison between estimations of LOY from 121 pairwise studied samples using SNP-array and droplet digital PCR (ddPCR)

affected with LOY in the subjects of the dataset, and thus, if the observed variation in clonal trajectories are an effect from LOY in different cell types. However, results from a recent analysis [18] shows that the level of LOY varies substantially between different types of leukocytes in blood, with generally higher levels in myeloid compared with lymphoid lineages, and that LOY often occurs as an oligoclonal event in peripheral blood. To further elucidate if and how the developmental trajectories of LOY-clones varies between the different types of leukocytes in blood, studies of the frequency of LOY in different cells types in serially collected subjects will be informative. Interestingly, we observed in a subset of the studied men that the frequency

of LOY in whole blood samples increased in a non-linear fashion (Supplementary Fig. 4). Such expansions would likely be an effect of oligo or polyclonal processes, i.e. several hematopoietic progenitor cells giving rise to ACEs without the Y chromosome.

The predominant proxy used for estimation of LOY mosaicism is the mLRRY calculated from the median LRR-values of SNP-array probes positioned in the MSY [4]. Other methods to estimate LOY have also been proposed, such as calculation of the mean LRR of MSY-probes [8]. However, this approach could potentially be more sensitive to biases from probes in ampliconic regions compared with estimations using the median. Hence, the median would

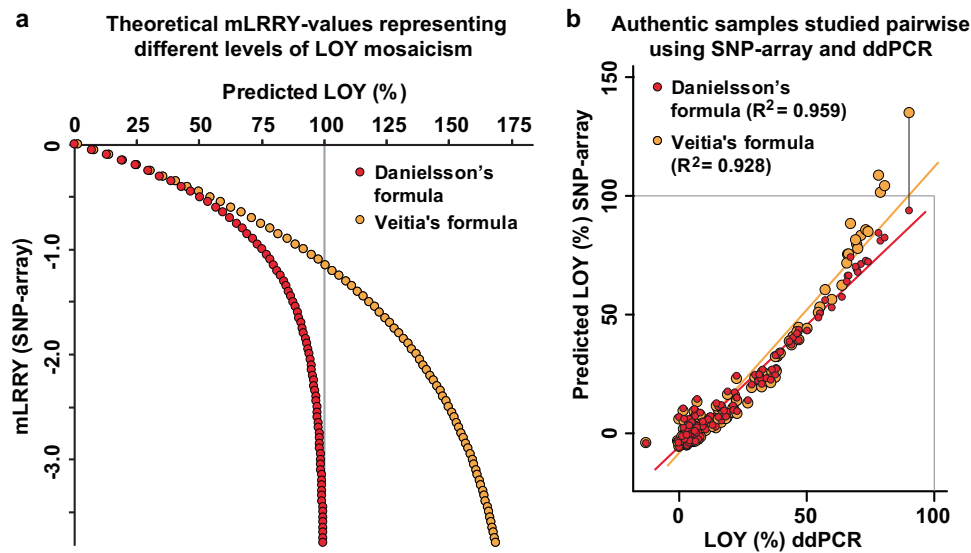


Fig. 4 Comparison of the performance of two formulas developed for prediction of percentage of LOY from mLRRY-values, i.e. Danielsson's formula presented here and the recently published Veitia's formula. In **a**, the estimates of mosaicism from each formula are compared in a range of theoretically possible mLRRY-values representing varying degree of LOY. The values of mLRRY plotted on the Y-axis were used to predict the percentages of LOY plotted on the X-axis. In male subjects without LOY, mLRRY-values are close to zero and lower values indicate increasing level of LOY mosaicism. The mLRRY calculated from female samples typically range -3 to -4 . The horizontal grey line at 100% represents an extreme level of

mosaicism (Y loss in all cells) and thus indicates a maximum theoretical limit of predicted LOY mosaicism in men. **b** shows a similar comparison using authentic data generated from 121 samples studied with both SNP-array and ddPCR. The Y-axis shows the predicted percentage of LOY from SNP-array data using each formula and the level of mosaicism measured by ddPCR in corresponding samples are plotted on the X-axis. Grey lines indicate theoretical upper limits of LOY estimations. To illustrate the overestimation of LOY mosaicism generated by the Veitia's formula, a black line connects the predictions from each formula in the sample with the highest level of LOY mosaicism

represent the average probe-intensity more accurately since the intensity values of outliers will have less weight on the calculated mLRRY. Furthermore, a data type called B allele frequency (BAF) is generated by SNP-arrays (in addition to SNP-calls and the LRR-data) and several algorithms have been developed for detection of autosomal chromosomal alterations using imbalances in BAF [28, 35, 37]. Recently, the BAF has also been used to estimate LOY mosaicism by mapping of BAF deviations in probes located in the pseudo-autosomal regions, a region that is shared between the X and Y chromosomes [21]. Although haplotype information is incorporated in this approach, it is currently not known if uncontrolled variation (such as potential X-mosaicism) could potentially influence the estimates of low level LOY mosaicism using this algorithm. In contrast, the mLRRY is calculated from Y-specific variation in the MSY. However, the mLRRY-estimate has other shortcomings, such as its logarithmic and inversed scale (Fig. 2). Here we present a formula to transform mLRRY-values from SNP-arrays to the unit percentage of cells with LOY, which is a more intuitive unit on a linearised scale. The formula for mLRRY-transformation was established by analysing the same DNA samples using three independent technologies, i.e. SNP-array, WGS and ddPCR. A detailed description of how LOY was estimated

using each technology is provided in the Materials and Methods. We first established that LOY estimation using WGS and ddPCR yielded close to identical results from the same set of DNA samples (Fig. 2c). The LOY estimates from these technologies could therefore be used as a baseline to derive a formula for transformation of mLRRY to percentage of LOY. Of note, the same formula was independently derived from analyses of the WGS and ddPCR datasets, i.e. $LOY (\%) = 100 * (1 - 2^{2 * mLRRY})$. After transformation of mLRRY-values with this formula, the concordance of the SNP-array data improved substantially in comparisons with readouts from the other technologies (Fig. 3).

It should be noted that the results presented here are based on data generated by different versions of Illumina genotyping arrays (Supplementary Table 2). Further studies are needed to evaluate the performance of the formula using data generated from different arrays and from other manufacturers, for example by the here presented ddPCR approach. The unit percentage of LOY represents the studied biological event in a more comprehensible way compared with mLRRY, since a higher percentage of LOY is indicative of a higher level of mosaicism. We also established that the formula predicts the level of LOY mosaicism within a theoretically and biologically relevant continuum

(between zero and 100% mosaicism) in contrast to a recently published formula [27] (Fig. 4).

In conclusion, we describe a dynamic nature of changes in the frequency of LOY within in the hematopoietic system of serially studied men. A group of individuals were identified as LOY progressors with a clear expansion over time, while in others, the longitudinal frequency remained unchanged or showed other types of trajectories. We also present a formula for transformation of mLRRY-values calculated from SNP-array data into percentage of LOY and describe a new TaqMan/ddPCR-based method for efficient LOY analysis. The pipelines and methods described will be useful to further investigate associations between LOY in leukocytes and various outcomes.

Acknowledgements We acknowledge all men of the ULSAM study for their participation and K. Ström for collection of samples. Thanks goes to U.Landegren for discussions and feedback on the manuscript.

Funding The study was sponsored by grants, for the purpose of investigating LOY, from the Swedish Cancer Society, the Swedish Research Council, Konung Gustav V:s och Drottning Viktorias Frimurarestiftelse, the Science for Life Laboratory Uppsala, Alzheimerfonden and the Foundation for Polish Science under the International Research Agendas Programme (Award nr. MAB/2018/6) to JPD, and by grants from the European Research Council (ERC) Starting Grant, the Swedish Research Council, the Olle Enqvist Byggnästars Foundation, the Kjell and Märta Beijers Foundation to LAF. Genotyping and sequencing were performed by the SNP&SEQ Technology Platform in Uppsala, which is part of the Science for Life Laboratory at Uppsala University and is supported as a national infrastructure by the Swedish Research Council.

Compliance with ethical standards

Conflict of interest JPD and LAF are cofounders and shareholders in Gray Innovation AB. All other authors declare no competing interest.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Forsberg LA, Gisselsson D, Dumanski JP. Mosaicism in health and disease—clones picking up speed. *Nat Rev Genet.* 2017;18:128–42.
- Jacobs PA, Brunton M, Court Brown WM, Doll R, Goldstein H. Change of human chromosome count distribution with age: evidence for a sex differences. *Nature.* 1963;197:1080–1.
- UKCCG. Loss of the Y chromosome from normal and neoplastic bone marrows. United Kingdom Cancer Cytogenetics Group (UKCCG). *Genes Chromosome Cancer.* 1992;5:83–8.
- Forsberg LA, Rasi C, Malmqvist N, Davies H, Pasupulati S, Pakalapati G, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat Genet.* 2014;46:624–8.
- Lofffield E, Zhou W, Graubard BI, Yeager M, Chanock SJ, Freedman ND, et al. Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank. *Sci Rep.* 2018;8:12316.
- Ganster C, Kampfe D, Jung K, Braulke F, Shirmeshan K, Machherndl-Spandl S, et al. New data shed light on Y-loss-related pathogenesis in myelodysplastic syndromes. *Genes Chromosomes Cancer.* 2015;54:717–24.
- Noveski P, Madjunkova S, Sukarova Stefanovska E, Matevska Geshkovska N, Kuzmanovska M, Dimovski A, et al. Loss of Y chromosome in peripheral blood of colorectal and prostate cancer patients. *PLoS ONE.* 2016;11:e0146264.
- Zhou W, Machiela MJ, Freedman ND, Rothman N, Malats N, Dagnall C, et al. Mosaic loss of chromosome Y is associated with common variation near TCL1A. *Nat Genet.* 2016;48:563–8.
- Machiela MJ, Dagnall CL, Pathak A, Loud JT, Chanock SJ, Greene MH, et al. Mosaic chromosome Y loss and testicular germ cell tumor risk. *J Hum Genet.* 2017;62:637–40.
- Lofffield E, Zhou W, Yeager M, Chanock SJ, Freedman ND, Machiela MJ. Mosaic Y loss is moderately associated with solid tumor risk. *Cancer Res.* 2019;79:461–6.
- Persani L, Bonomi M, Lleo A, Pasini S, Civardi F, Bianchi I, et al. Increased loss of the Y chromosome in peripheral blood cells in male patients with autoimmune thyroiditis. *J Autoimmun.* 2012;38:J193–6.
- Lleo A, Oertelt-Prigione S, Bianchi I, Caliarì L, Finelli P, Miozzo M, et al. Y chromosome loss in male patients with primary biliary cirrhosis. *J Autoimmun.* 2013;41:87–91.
- Dumanski JP, Lambert JC, Rasi C, Giedraitis V, Davies H, Grenier-Boley B, et al. Mosaic loss of chromosome Y in blood is associated with Alzheimer disease. *Am J Hum Genet.* 2016;98:1208–19.
- Haitjema S, Kofink D, van Setten J, van der Laan SW, Schoneveld AH, Eales J, et al. Loss of Y chromosome in blood is associated with major cardiovascular events during follow-up in men after carotid endarterectomy. *Circ Cardiovasc Genet.* 2017;10:e001544.
- Hirata T, Hishimoto A, Otsuka I, Okazaki S, Boku S, Kimura A, et al. Investigation of chromosome Y loss in men with schizophrenia. *Neuropsychiatr Dis Treat.* 2018;14:2115–22.
- Grassmann F, Kiel C, den Hollander AI, Weeks DE, Lotery A, Cipriani V, et al. Y chromosome mosaicism is associated with age-related macular degeneration. *Eur J Hum Genet.* 2019;27:36–41.
- Forsberg LA. Loss of chromosome Y (LOY) in blood cells is associated with increased risk for disease and mortality in aging men. *Hum Genet.* 2017;136:657–63.
- Dumanski JP, Halvardson J, Davies H, Rychlicka-Buniowska E, Mattisson J, Moghadam BT, et al. Loss of Y in leukocytes, dysregulation of autosomal immune genes and disease risks. *BioRxiv.* 2019; p 673459. <https://doi.org/10.1101/673459>.
- Wright DJ, Day FR, Kerrison ND, Zink F, Cardona A, Sulem P, et al. Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility. *Nat Genet.* 2017;49:674–9.
- Zink F, Stacey SN, Norddahl GL, Frigge ML, Magnusson OT, Jonsdottir I, et al. Clonal hematopoiesis, with and without

- candidate driver mutations, is common in the elderly. *Blood*. 2017;130:742–52.
21. Thompson DJ, Genovese G, Halvardson J, Ulirsch JC, Wright DJ, Terao C, et al. Genetic predisposition to mosaic Y chromosome loss in blood is associated with genomic instability in other tissues and susceptibility to non-haematological cancers. *BioRxiv*. 2019; p 514026. <https://doi.org/10.1101/514026>.
 22. Forsberg LA, Halvardson J, Rychlicka-Buniowska E, Danielsson M, Moghadam BT, Mattisson J, et al. Mosaic loss of chromosome Y in leukocytes matters. *Nat Genet*. 2019;51:4–7.
 23. Dumanski JP, Rasi C, Lonn M, Davies H, Ingelsson M, Giedraitis V, et al. Smoking is associated with mosaic loss of chromosome Y. *Science*. 2015;347:81–3.
 24. Wong JYY, Margolis HG, Machiela M, Zhou W, Odden MC, Psaty BM, et al. Outdoor air pollution and mosaic loss of chromosome Y in older men from the Cardiovascular Health Study. *Environ Int*. 2018;116:239–47.
 25. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*. 2012;28:423–5.
 26. Behrendt S. lm.beta: add standardized regression coefficients to lm-Objects. 2014. <https://CRAN.R-project.org/package=lm.beta>.
 27. Grassmann F, International AMDGC, Weber BHF, Veitia RA. Insights into the loss of the Y chromosome with age in control individuals and in patients with age-related macular degeneration using genotyping microarray data. *Hum Genet*. 2019. <https://doi.org/10.1007/s00439-019-02029-1>.
 28. Forsberg LA, Rasi C, Razzaghian H, Pakalapati G, Waite L, Stanton Thilbeault K, et al. Age-related somatic structural changes in the nuclear genome of human blood cells. *Am J Hum Genet*. 2012;90:217–28.
 29. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet*. 2012;44:651–8.
 30. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet*. 2012;44:642–50.
 31. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*. 2014;371:2477–87.
 32. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014;371:2488–98.
 33. Machiela MJ, Zhou W, Sampson JN, Dean MC, Jacobs KB, Black A, et al. Characterization of large structural genetic mosaicism in human autosomes. *Am J Hum Genet*. 2015;96:487–97.
 34. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*. 2015;348:880–6.
 35. Vattathil S, Scheet P. Extensive hidden genomic mosaicism revealed in normal tissue. *Am J Hum Genet*. 2016;98:571–8.
 36. Abelson S, Collord G, Ng SWK, Weissbrod O, Mendelson Cohen N, Niemeyer E, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature*. 2018;559:400–4.
 37. Loh PR, Genovese G, Handsaker RE, Finucane HK, Reshef YA, Palamara PF, et al. Insights into clonal hematopoiesis from 8,342 mosaic chromosomal alterations. *Nature*. 2018;559:350–5.