Heritability of skewed X-inactivation in female twins is tissue-specific and dependent on age

Antonino Zito¹, Matthew N. Davies², Pei-Chien Tsai¹, Susanna Roberts³, Stefano Nardone⁴, Jordana T. Bell¹, Chloe C. Y. Wong³ & Kerrin S. Small¹.

¹ Department of Twin Research & Genetic Epidemiology, King's College London, London SE1 7EH, UK

² The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

³ Institute of Psychiatry, Psychology & Neuroscience, King's College London, London SE5 8AF, UK

⁴ Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA

Abstract

To balance the X-linked transcriptional dosages between the sexes, one of the two X-chromosomes is randomly selected to be inactivated in somatic tissues of female placental mammals. Non-random, or skewed X-chromosome inactivation (XCI) toward one parental X has been observed in female somatic tissues, and this skewing effect has been associated with several complex human traits. However, the extent of the influence of genetic and environmental factors on XCI skewing is largely unknown. Here, we use RNA-seq and DNA-seq data taken from a large cohort of female twins to quantify the degree of skewing of XCI (DS) in multiple tissues and to study the relationship of XCI with age, genetic factors and complex traits. We show that the XCI patterns are highly tissue-specific with a higher prevalence of skewed XCI in blood-derived tissues than in fat or skin tissue. We also show that the DS in blood-derived tissues is associated with age and that the acquired DS occurs uniquely in blood-derived tissues with an inflection point at approximately 55 years of age. Heritability analysis indicates that the heritability of DS is both age and tissue specific; DS is heritable in blood tissue of females >55 years-old ($h^2 = 0.34$) but is not heritable in blood tissues of females <55 years-old ($h^2 = 0$), nor in skin and fat tissues at any age. We find a positive association between the DS and smoking status in blood tissues of older females ($P = 0.02$). The high tissue specificity of XCI patterns in human indicates the existence of tissue-specific mechanisms influencing XCI patterns, including genetic and environmental factors. We conclude that the heritability of XCI skewing in blood-derived tissues is dependent on age, representing a Gene x Age interaction that can shift the functional allelic dosage of an entire chromosome in a tissue-restricted manner.

27 Introduction

28 To balance the X-linked transcriptional dosages between the single X chromosome of males and the
 29 two X chromosomes of females, one X chromosome is silenced in female placental mammals¹. The
 30 X-chromosome inactivation (XCI) process starts during preimplantation phases of human embryonic
 31 development, presumably at around the 8-cell stage². XCI is initiated by the transcription of *XIST*, a
 32 17 kb, alternatively spliced long non-coding RNA mapped to Xq13.2 and exclusively expressed on
 33 the inactive X (Xi)³. Once transcribed, *XIST* molecules spread in *cis* along the X chromosome^{4,5}
 34 inducing a progressive epigenetic silencing through the recruitment of chromatin remodelling
 35 enzymatic complexes, which impose repressive histone and DNA changes on the Xi chromosome^{6,7}.
 36 Within each cell, the parental X chromosome selected for inactivation seems to occur at random,
 37 and the Xi is mitotically inherited to future somatic daughter cells. This random inactivation results
 38 in a mosaic of cells within an individual, where overall, a balanced expression (50:50) of both
 39 parental X-linked alleles is expected. Asymmetric selection of the X chromosome to inactivate
 40 causes the predominance of one parental Xi in a population of cells, unbalancing the X-linked
 41 transcriptional and allelic dosages toward one parental X chromosome. This phenomenon, known
 42 as skewed XCI (or non-random XCI), occurs when at least 80% of cells within a tissue inactivate the
 43 same parental X chromosome. The factors underlying primary skewed XCI are varied and several
 44 mechanisms are possible (reviewed in⁸). Secondary (or acquired) skewed XCI can result from
 45 positive selection of cells that after having inactivated a particular parental X, acquire a survival
 46 advantage over cells who inactivated the other parental X chromosome. Skewed XCI patterns can
 47 also be generated by the stochastic overrepresentation of cell clones in a given tissue, due for
 48 instance, to depletion of stem cell populations.

49

50 Comprised of 155 MB and containing >800 protein-coding genes, the X chromosome represents
 51 approximately 8% of the human genome. In heterozygous females with skewed XCI, the X-linked

transcriptional and allelic dosages of silenced genes are unbalanced and may be functionally homozygous. Skewed XCI is a major cause of discontinuity of dominance and recessiveness, as well as penetrance and expressivity of X-linked traits. How skewed XCI patterns modulate phenotypes in females, and whether they are a cause or a consequence of associated phenotypes is not fully understood. Skewed XCI patterns have been observed in females with X-linked diseases^{9,10}, autoimmune disorders^{11,12}, as well as in breast¹³ and ovarian cancer¹⁴. In autoimmune diseases with higher prevalence in females, including rheumatoid arthritis and systemic lupus erythematosus, XCI is hypothesized to play a role. Chromosome X is enriched for immune-related genes and skewed XCI patterns cause the breakdown of thymic tolerance induction processes¹⁵ conferring a high predisposition to develop autoimmunity (reviewed in¹⁶). XCI skewing levels in blood-tissues have been associated with ageing, with multiple studies indicating an increase after 50-60 years of age¹⁷⁻²². To date, the mechanisms underlying skewed XCI in humans remain to be determined and hypotheses are still controversial. Several twin studies have reported that genetic factors contribute to XCI skewing in blood derived cells^{21,23}, while other evidence indicated that most of the XCI skewing levels in human are acquired secondarily²⁴.

Nearly all studies of XCI skewing levels in humans have been carried out in peripheral blood samples or in very small sample sizes²⁵, while XCI patterns in other tissues have not been studied in great detail^{19,26,27}. In this study, we comprehensively assessed XCI patterns in a multi-tissue sample of nearly 800 female twins from the TwinsUK cohort²⁸. We quantified the degree of skewing of XCI using a metric based on *XIST* allele-specific expression from paired RNA-seq and DNA-seq data in four tissues. We examined the tissue-specific prevalence of skewed XCI patterns, compared the XCI skewing levels between tissues and evaluated the association between XCI skewing and lifestyle traits. In order to investigate the factors underlying the skewed XCI, we utilized classical twin models to characterize the extent of the influence of genetic and environmental factors on the tissue-

specific skewed XCI, showing that the XCI patterns have both a heritable and environmental (age) basis.

Results

Quantification of degree of skewing in TwinsUK

We assessed XCI patterns in multi-tissue samples from female twin volunteers from the TwinsUK cohort aged 38–85 years old (median age = 60; Figure S1)^{28,29}. We quantified the degree of skewing of XCI using a metric based on *XIST* allele-specific expression (ASE) from paired RNA-seq and DNA-seq data. *XIST* is uniquely expressed from the Xi³, so the relative expression of parental alleles within the *XIST* transcript are representative of XCI skewing levels in a sample²⁷. RNA-seq and genotype data were available for 814 Lymphoblastoid Cell Lines (LCL) samples, 395 whole-blood samples, 766 subcutaneous adipose tissue samples (herein referred as fat) and 716 skin samples. After stringent quality control, we obtained *XIST*_{ASE} calls for 422 LCL samples, 72 whole-blood samples, 378 fat samples and 336 skin samples. The smaller sample size for whole-blood was due to the relatively smaller size of the starting dataset and the relatively lower RNA-seq coverage of this tissue in our dataset. In order to have an absolute measure of the magnitude of the XCI skewing levels in each sample, we calculated the degree of skewing of XCI (DS) from the *XIST*_{ASE} calls. DS is defined as the absolute deviation of the *XIST*_{ASE} from 0.5 (see Materials and methods) and it has been similarly used in other investigations to assess XCI patterns^{20,30,31} and the XCI status of X-linked genes³². In line with previous investigations^{21,33} we classified samples with $DS < 0.3$ (corresponding to $0.2 < XIST_{ASE} < 0.8$) to have random XCI, and samples with $DS \geq 0.3$ (corresponding to $XIST_{ASE} \leq 0.2$ or $XIST_{ASE} \geq 0.8$) to have skewed XCI. Unless otherwise specified, DS was used in all the analyses performed.

We assessed the robustness of our estimates of the degree of skewing with an alternative DNA-based measure of XCI, the Human Androgen Receptor Assay (HUMARA)³⁴. HUMARA was and is still

102 the 'gold standard' technique to assess XCI patterns. A previous study has reported good replicability
 103 between HUMARA and expression-based quantification of XCI skewing³⁵. We used HUMARA to
 104 measure XCI skewing levels in 10 archived whole-blood DNA samples obtained at the same clinical
 105 visit as the LCLs samples. Spearman's correlation between the quantifications was 0.71, at $P = 0.031$,
 106 revealing a high degree of reproducibility between the both $XIST_{ASE}$ and HUMARA methods and the
 107 LCLs and whole-blood (Fig 1a).

108
 109 Previous investigations have reported that the XCI skewing levels increase with age in blood tissues,
 110 as discussed below. While it would be expected that increases in XCI skewing levels would be
 111 observed over relatively large time spans, we would expect minimal variations of XCI skewing levels
 112 between 2 close time points. We therefore reasoned that the sensitivity of our quantifications could
 113 also be assessed by comparing the XCI skewing levels in the same individuals at close time points.
 114 Briefly, using a publicly-available longitudinal whole-blood RNA-seq dataset from the TwinsUK
 115 cohort³⁶, we generated $XIST_{ASE}$ calls at 2 time points (1 to 2.7 years later) in 16 samples (see the
 116 Materials and methods). Spearman's correlation between the $XIST_{ASE}$ calls at the first time point and
 117 the $XIST_{ASE}$ calls at the second time point was 0.94 at $P < 2 \times 10^{-15}$ (Fig 1b). This indicates that $XIST_{ASE}$
 118 is a sensitive proxy when assessing the stability of XCI patterns over short time periods. Overall,
 119 these results indicate that $XIST_{ASE}$ is a reproducible, accurate and sensitive proxy of XCI skewing
 120 levels.

121 122 **Skewed XCI is tissue-specific with higher prevalence in blood-derived tissues**

123 We observed a wide range of DS values in the four tissues (Fig 2), with clear differences in the
 124 prevalence of skewed individuals between tissues. Blood-derived tissues had the highest incidence
 125 of skewed individuals, with skewed XCI observed in 34% of LCLs samples and 28% of whole-blood
 126 samples and a lower incidence in the primary tissues, where 12% of fat and 16% of skin samples

127 exhibited skewed XCI (Table 1). In order to examine the extent of similarities of XCI patterns
 128 between tissues, we compared the tissue-specific XCI skewing levels in a pairwise manner (Fig 3).
 129 For each tissue-tissue comparison, we included individuals with $XIST_{ASE}$ calls in both tissues (Table
 130 2). We found the strongest correlation on $XIST_{ASE}$ calls between LCLs and whole-blood ($N = 59$, $r =$
 131 0.78 , $P = 2 \times 10^{-13}$), indicating that blood-derived tissues share highly similar XCI skewing levels. We
 132 also found a good degree of similarity between the XCI skewing levels in fat and skin tissues ($N =$
 133 252 , $r = 0.47$, $P = 2 \times 10^{-15}$; Fig 3). However, low concordance was observed between skin and whole-
 134 blood ($N = 47$, $r = 0.3$, $P = 0.04$) and fat and whole-blood ($N = 57$, $r = 0.33$, $P = 0.02$). Our data
 135 demonstrate that tissue-specific XCI skewing within an individual is common in the population,
 136 indicating that XCI patterns are partially controlled by tissue-specific regulatory mechanisms.

137

The active or inactive state of each X chromosome in a cell is clonally passed on to daughter cells. In a pool of cells derived from a single clone (or patch), the XCI patterns are expected to be completely skewed. Patch size refers to the amount of cell clones in a pool of cells (e.g. in a tissue biopsy). We considered the possibility that patch size might bias our quantification of XCI patterns in fat and skin samples. This is likely to occur in biopsies that are smaller than the tissue patch size. However, several considerations led us to exclude the possibility that patch sizes in fat and skin biopsies might confound our $XIST_{ASE}$ calls. First, the biopsies included skin samples of $8mm^3$ in size, which were cut into 2 skin and 3 fat samples. As reported in another study, this size is large enough to measure the XCI ratio without being confounded by patch size³⁷. Second, most individuals exhibit random XCI patterns in fat and skin tissues, which is unlikely if patch size was larger than the biopsies. We therefore conclude that the biopsies used in this study are large enough to accurately assess the XCI patterns without being biased by patch size.

LCLs in this study are representative of XCI skewing *in-vivo* in blood tissues

Lymphoblastoid Cell Lines (LCLs) generated by Epstein-Barr virus mediated transformation of B lymphocyte cells have been and are widely used in gene expression studies. However, the possibility that the cell lines are monoclonal and/or polyclonal due to selection in the transformation process or clonal expansion in cell culture, and hence not be representative of the *in-vivo* XCI skewing levels, is a potential problem when using LCLs to assess XCI skewing³⁸. As the profiled RNA in this study was extracted from the LCLs very shortly after transformation with limited passaging or time in culture we expected this effect to be minimal, however, to address the possibility we performed the following analyses. First, as described above and shown in Figure 1a, the degree of skewing in LCLs were highly correlated with the HUMARA-based quantifications of XCI patterns in paired whole-blood samples ($r = 0.71$, $N = 10$). We would not expect such high similarity between the two quantifications if clonal propagation had occurred in LCLs samples after preparation. This was confirmed by the high correlation between LCLs and whole-blood $XIST_{ASE}$ values ($r = 0.78$, $N = 59$; Fig 3) and overall similarity in the prevalence of skewed XCI in LCLs and whole-blood (Table 1). Finally, we assessed the degree of skewing in monocytes, B, T-CD4⁺, T-CD8⁺ and natural-killer (NK) cells purified from two monozygotic twins exhibiting skewed XCI patterns in LCLs and from one individual exhibiting random XCI patterns in LCLs. We found that in both monozygotic twins showing skewed XCI in LCLs, the majority of immune cell types exhibited skewed XCI patterns. Conversely, none of the immune cell types purified from the non-skewed individual exhibited skewed XCI patterns (Table 3). We conclude that the XCI skewing levels of LCLs in this study are representative of XCI skewing *in-vivo* in blood tissues.

XCI skewing levels are positively associated with age in blood-derived tissues

XCI skewing levels in peripheral blood have been shown to increase with age in multiple studies^{17-21,23,35,39,40}. The age-related increase of XCI skewing levels continues throughout life, since

centenarians exhibit higher XCI skewing levels than 95 years-old females²¹. However, there is very limited knowledge on the relationship between XCI patterns and ageing in tissues other than blood. In order to explore this, we investigated the association between age and degree of skewing in LCLs, fat and skin. Our whole-blood estimates were excluded from analysis due to low sample size (N = 72). Age was positively associated with XCI skew in LCLs (N = 422, $P < 0.01$), but we did not detect any association between XCI skew and age in skin (N = 336, $P = 0.4$) or in fat (N = 378, $P = 0.7$).

We next explored the dynamics of DS and age progression in each tissue, using the lowess procedure. Lowess curve detected an increase of DS beginning at around 55 years-old in LCLs (Fig 4), in agreement with what was found in other studies^{19,21}. Since the increase of DS starts at around 55 years, we divided LCLs samples into a younger group (N = 141, age < 55) and an older group (N = 281, age ≥ 55). We found that the mean DS in LCLs was significantly higher in older than in younger females (DS_{younger} = 0.2, DS_{older} = 0.24, $P = 0.03$; Figure 4). Accordingly, we found that the frequency of skewed XCI in LCLs was significantly higher in older (38%) than in younger (28%) females (χ^2 test, $P = 0.04$; Fig 4). In agreement with the lack of association between the DS and age, we did not detect significant differences between the mean DS in young and older females in fat (DS_{younger} = 0.15, DS_{older} = 0.15) or in skin tissues (DS_{younger} = 0.16, DS_{older} = 0.17). To acquire a more detailed view of the tissue-specific prevalence of skewed XCI in different groups of age, we categorized the samples into four age groups (40-50, 50-60, 60-70, >70) and calculated the frequency of skewed XCI in each category (Fig 4). We found that the frequency of skewed XCI increased with age in LCLs, with 41% of individuals >65 years-old demonstrating skewed XCI patterns. We did not observe any increase in the skewed XCI frequencies with age in fat and skin tissues. Overall, these data further confirm that XCI skewing levels increase with age in blood-derived tissues, supporting previous investigations. However, we find that there is no increase in XCI in fat and skin tissue from the same

individuals, suggesting that acquired XCI skewing with age is a distinctive feature of blood-derived tissues.

Heritability of skewed XCI is dependent on tissue and age

Twin studies are a powerful strategy to investigate the heritability of complex traits. Previous twin studies have reported that skewed XCI in blood-derived samples is heritable, with h^2 estimates of 0.68 in granulocytes of elderly twin pairs and 0.58 in peripheral blood cells²¹⁻²³, however these studies have not investigated heritability outside of blood. To estimate the influence of additive genetic effects (heritability) and environmental factors on the observed variance in XCI in the three tissues, we implemented the ACE twin model. The ACE statistical model quantifies the contribution of additive genetic effects (A), shared environment (C) and unique environment (E) to the phenotype variance. In order to investigate whether heritability varies with age, we stratified the twin pairs into a younger group (age < 55) and an older group (age $\geq$ 55; Table S1). Age 55 was chosen as it was identified as the inflection point at which XCI skew begins to increase in the lowess analysis above. We found that XCI skewing is heritable in LCLs of older females ($h^2 = 0.34$, $P = 9.6e-07$), but not younger females ($h^2 = 0$, $P = 1$). There was no evidence of heritability of XCI skew in fat or in skin tissues at any age (Table 4). The highest proportion of variance was explained by unique environmental factors in all tissues of both younger and older females ($E^2_{LCLs_younger} = 0.99$, $E^2_{LCLs_older} = 0.66$, $E^2_{Fat_younger} = 0.73$, $E^2_{Fat_older} = 0.92$, $E^2_{Skin_younger} = 1$, $E^2_{Skin_older} = 1$). As a complement to the heritability analysis, we calculated intraclass correlation (IC) of XCI skew within MZ and DZ twin pairs of all ages, and within younger and older MZ and DZ twin pairs (Table 5). IC analyses of twin pairs is often used to demonstrate the existence of genetic effect in smaller sample sizes. The IC of XCI skew within MZ twins pairs was positive and statistically significant ($IC_{MZ_allAges} = 0.31$, $P = 0.02$). We found significant IC of XCI skew within older MZ twin pairs ($IC_{MZ_older} = 0.42$, $P = 0.005$), but not within young MZ twin pairs ($IC_{MZ_younger} = 0.06$, $P = 0.8$). We did not detect significant IC within DZ twin pairs

at any age, in agreement with previous study in blood²¹. The higher IC of XCI skew within MZ twin pairs compared with DZ twin pairs indicates the involvement of genetic determinants in the regulation of XCI skew in blood-derived tissues. The increase of IC in older compared to younger MZ twin pairs and the fact that the heritability of XCI skew is observed only in females older than 55, confirm a role for genetic variants as age-dependent regulators of the acquired XCI skew in blood-derived tissues. Presumably, genetically-determined secondary cell selection processes act in haematopoietic cell lineages, with the high mitotic rates contributing to the manifestation of their effects in blood-derived tissues. Results also highlight an age-independent role for environmental factors as regulators of XCI skew in blood, fat and skin tissues.

XCI skew is associated with smoking status in older females

Tobacco smoking has been reported to induce epigenomic changes including DNA methylation variation (reviewed in⁴¹). Smoking is a well characterized risk factor in cancer⁴² and, as more recently discovered, in the aetiology of autoimmunity⁴³. Although smoking-related X-linked DNA methylation sites have been discovered⁴⁴, no previous studies, to our knowledge, have investigated the relationship between smoking and XCI patterns. We reasoned that changes of XCI patterns may result from smoking, and affect in turn short-term and long-term health. In order to test our hypothesis, we used the 270 individuals in our dataset for which we had smoking status at the time of sample collection, including 233 never smokers and 37 current smokers⁴⁵. We found no difference in the frequency of skewed XCI patterns between never and current smokers (36% and 35% respectively) in LCLs. To take into account the effects of age on the degree of skewing in blood-derived tissues and to examine the relationship between smoking status and degree of skewing at different ages, we split the dataset into a younger (age < 55) and older group (age ≥ 55; Table S2). While the frequencies of skewed XCI were very similar between young smokers and young never smokers (27% and 28% respectively), we detected a higher prevalence of skewed XCI in older

smokers compared with older never smokers (47% and 40% respectively). Accordingly, we found an overall positive association between XCI skew and smoking status in older ($P = 0.02$), but not in younger individuals ($P = 0.5$). The data suggest a role for smoking as a modulator of XCI skew in blood-derived tissues of females older than 55. Presumably, the association between smoking and XCI skew changes is complex, and further investigations are needed to characterize the genetic and molecular mechanisms underlying this phenomenon.

Discussion

In this study, we used multi-tissue transcriptomic data from twins to comprehensively characterize XCI patterns in LCLs, whole-blood, fat and skin tissues from a healthy twin cohort. We show XCI patterns to be tissue-specific and that blood-derived tissues exhibited the highest prevalence of skewed XCI and share the highest similarity of XCI patterns. These findings indicate that XCI patterns are partially driven by tissue-specific mechanisms, and that the XCI skew measured in blood is not a reliable proxy for the skew in other tissues. Skewed XCI patterns limited to disease-relevant tissues and cells have been observed in multiple conditions^{9,10,13,14,46,47} but except for several cases of X-linked diseases, their roles in disease aetiology and predisposition remain largely unknown. Our results demonstrate that tissue-specific XCI patterns within an individual is common in this healthy population.

We show that XCI skewing levels in blood tissues increase with age, with an inflection point at around 55, confirming previous reports^{17-21,23,35,39,40}. In this study, more than 41% of females >65 years-old demonstrate skewed XCI patterns in blood-derived tissues, indicating that acquired skewed XCI is a highly prevalent phenotype in ageing populations. We show age-related increase in XCI skew is a distinctive feature of blood-derived tissues, with no evidence for an age-related increase in fat or skin. Age-related increase in XCI skew partially explains the higher incidence of

262 skewed XCI in blood than fat and skin tissues. The effects of age-related skewing of XCI on healthy
 263 ageing remain largely unknown, but may have a broad impact on the immune system.
 264 Hematopoietic stem cells and the immune system continue to develop throughout life. Presumably,
 265 imbalanced X-linked immune-related gene expression toward one parental haplotype leads to a
 266 reduced molecular diversity, which may translate in a decline of immune repertoire as well as poor
 267 sustenance of the immunological memory.

268

269 Previous twin studies have reported that XCI patterns in blood have a genetic component^{21,23}. To
 270 our knowledge, this is the first study to investigate heritability of XCI skewing levels in other tissues.
 271 We found that the heritability of XCI skewing level is limited to blood-derived tissues of females >55
 272 years-old ($h^2 = 0.34$), with no evidence of heritability in fat or skin or younger individuals in any
 273 tissue. The restriction of heritability to blood of older individuals is of interest given the link between
 274 skewed X-inactivation and clonal haematopoiesis. Positive selection of cells carrying an
 275 advantageous somatic mutation will lead to clonal haematopoiesis and skewed XCI patterns as the
 276 selected cells will carry the same inactivated parental X. Somatic mutation-driven clonal
 277 haematopoiesis is now known to be common in blood of healthy older individuals and is often
 278 referred to as clonal haematopoiesis of indeterminate potential (CHIP)^{31,48-50}. CHIP is associated
 279 with increased risk of both cancer and all-cause mortality^{51,52}. The increase in XCI skew in older
 280 smokers in our study is consistent with the increase in clonal haematopoiesis observed in
 281 smokers^{50,53,54}. It is unknown to what extent CHIP accounts for age-acquired XCI skew, however if it
 282 is a major driver this would suggest that like age-related XCI skew, CHIP has a significant germline
 283 genetic component. Stochastic selection of cells could also contribute to the variance of XCI skewing
 284 levels, but, in agreement with previous works^{21,23}, we reason that their contribution is minimal. If
 285 stochastic selection of cells was a dominant mechanism, the correlation of XCI patterns between
 286 twin pairs would decrease with age.

287

288 Overall, the data presented in this study indicate a Gene x Age interaction that shifts the functional
289 allelic dosages of chromosome X in a tissue-restricted manner. The high prevalence of skewed XCI
290 and tissue-restricted XCI in a healthy population could complicate discovery of Chromosome X
291 variants associated with a trait and subsequent genetic risk prediction, as an individual's genotype
292 may not match their functional genotypic dosage in the relevant tissue. Further investigations of
293 the heterogeneity of XCI patterns across tissues and how this is regulated are essential to clarify the
294 biomedical implications of skewed XCI and its role in healthy ageing in women.

295

296 **Materials and methods**

297 **Sample collection**

298 The study included 856 female twins from the TwinsUK registry^{28,29} who participated in the MuTHER
299 study⁵⁵. Study participants included both monozygotic (MZ) and dizygotic (DZ) twins, aged 38-85
300 years old (median age = 60; Suppl. Fig 1) and were of European ancestry. Volunteers received
301 detailed information regarding all aspects of the research project and gave a prior signed consent
302 to participate in the study. Peripheral blood samples were collected and lymphoblastoid cell lines
303 (LCLs) were generated via Epstein-Barr virus (EBV) mediated transformation of the B-lymphocyte
304 fraction. Punch biopsies of subcutaneous adipose tissue were taken from a photo-protected area
305 adjacent and inferior to the umbilicus. Skin samples were obtained by dissection from the punch
306 biopsies. Adipose and skin samples were weighed and frozen in liquid nitrogen. This project was
307 approved by the research ethics committee at St Thomas' Hospital London, where all the TwinsUK
308 biopsies were carried out. Volunteers gave informed consent and signed an approved consent form
309 prior to the biopsy procedure. Volunteers were supplied with an appropriate detailed information
310 sheet regarding the research project and biopsy procedure by post prior to attending for the biopsy.

311 **Genotyping and phasing**

312 Whole-Genome Sequence data (WGS) were generated within the UK10K project as previously
 313 described⁵⁶. 557 individuals had both X-chromosome sequence data and RNA-seq data for at least
 314 one tissue. For individuals with unavailable X chromosome sequence data, X-linked genotypes data
 315 were retrieved from the TwinsUK genotypes previously imputed into the 1000 Genomes Project
 316 phase 1 reference panel^{57,58}, as described⁵⁹. Haplotypes of X-linked SNPs with a MAF >5% were then
 317 phased using shapeit v2.r837^{60,61}, with the --chrX flag to set up all functionalities for the phasing of
 318 non pseudo-autosomal (non-PAR) regions of the X-chromosome, along with a phasing window of
 319 2Mb and 1000 conditional states. Phasing was performed using the genetic map b37 and the 1000
 320 Genomes Project phase 3 reference panel of non-PAR X-linked haplotypes⁶¹. Phased X-linked SNPs
 321 were used for further analysis.

322 **RNA-sequencing data**

323 RNA-sequencing data were generated as previously described²⁹. Briefly, the Illumina TruSeq sample
 324 preparation protocol was used to generate the cDNA libraries for sequencing. Samples were then
 325 sequenced on an Illumina HiSeq 2000 machine and 49 bp paired-end reads were generated.
 326 Sequencing reads were aligned to the UCSC GRCh37/hg19 reference genome with the Burrows-
 327 Wheeler Aligner v.0.5.9⁶². Genes were annotated using the GENCODE v10 reference panel⁶³.

328 **Longitudinal RNA-sequencing data**

329 Peripheral blood samples were collected 1 to 2.7 years apart from 114 female twins of the TwinsUK
 330 registry and were processed with the Illumina TruSeq protocol, sequenced on a HiSeq 2000 machine
 331 and 49 bp paired-end reads aligned as described in³⁶. Adapter and polyA/T nucleotide sequences
 332 were trimmed using trim_galore and PrinSeq tools⁶⁴, respectively. Reads were aligned to the UCSC
 333 GRCh37/hg19 reference genome with the STAR v.2.5.2a aligner⁶⁵. Alignments containing non-
 334 canonical and unannotated splice junctions were discarded. Properly paired and uniquely mapped
 335 reads with a MAPQ of 255 were retained for further analysis.

336 **Purified immune cells RNA-sequencing data**

337 Monocytes, B, T-CD4⁺, T-CD8⁺ and NK cells were purified using fluorescence activated cell sorting
338 (FACS) from two monozygotic twins exhibiting skewed XCI patterns in LCLs and from 1 individual
339 exhibiting random XCI patterns. Total RNA was isolated and cDNA libraries for sequencing were
340 generated using the Sureselect sample preparation protocol. Samples were then sequenced in
341 triplicates on an Illumina HiSeq machine and 126 bp paired-end reads were generated. Adapter and
342 polyA/T nucleotide sequences were trimmed using trim_galore and PrinSeq tools⁶⁴ respectively.
343 Human and prokaryotic rRNAs were identified using sortmerna v.2.1⁶⁶ and removed. Reads were
344 aligned to the UCSC GRCh37/hg19 reference genome using STAR v.2.5.2a⁶⁵. Alignments containing
345 non-canonical and unannotated splice junctions were discarded. Properly paired and uniquely
346 mapped reads with a MAPQ of 255 were retained for further analysis.

347 **Correction of RNA-seq mapping biases**

348 To eliminate mapping biases, all RNA-seq data were re-aligned within the WASP pipeline for
349 mappability filtering⁶⁷. The WASP tool has an algorithm specifically designed to identify and correct
350 mapping biases in RNA-seq data. In each read overlapping a heterozygous SNP, the allele is flipped
351 to the SNP's other allele (generating all possible allelic combinations) and the read is remapped.
352 Reads that did not remap to the same genomic location indicate mapping bias and were discarded.
353 Reads overlapping insertions and deletions were also discarded. Properly paired and uniquely
354 mapped reads were retained for analysis.

355 **Quantification of allelic read counts and Allele Specific Expression**

356 Allelic read counts at heterozygous SNPs within *XIST* were quantified from paired RNA-seq and X-
357 linked genotypes data with GATK ASEReadCounter⁶⁸. Reads flagged by ASEReadCounter as having
358 low base quality were discarded. To increase the confidence that genotypes were truly
359 heterozygous, only X-linked SNPs with both alleles detected in RNA-seq data and with a read depth
360 of least 10 reads were retained for analysis. X-chromosome allelic read count tables of samples with

at least 1 *XIST*-linked SNP passing all quality filters were retained as informative of XCI skewing levels. To quantify the allele specific expression (ASE) of each heterozygous X-linked SNP, the read count at the major allele was divided by the read depth at the site.

Quantification of $XIST_{ASE}$ and degree of skewing of XCI (DS)

In each sample, the XCI skewing levels were quantified by averaging the ASE values of heterozygous SNPs within *XIST*. All SNPs were phased prior to averaging as detailed above. The measure, called $XIST_{ASE}$ is defined as follow:

$$XIST_{ASE} = \frac{\sum XIST_SNP_{ASE}}{n} \quad (XIST_{ASE})$$

where $XIST_SNP_{ASE}$ are the ASE values of heterozygous SNPs within *XIST* and n is the number of heterozygous SNPs within *XIST* in the sample. $XIST_{ASE}$ is a proxy of XCI skewing levels in a sample, with values ranging from 0 to 1. An $XIST_{ASE}$ value of 0.5 indicates equal inactivation of the two parental chromosomes, whereas a value of 0 or 1 indicates complete inactivation of one parental chromosome.

To be consistent with previous literature^{21,33}, we classified samples with $XIST_{ASE} \leq 0.2$ or $XIST_{ASE} \geq 0.8$ to have skewed XCI patterns, and samples with $0.2 < XIST_{ASE} < 0.8$ to have random XCI patterns. To have an absolute measure of the magnitude of the XCI skewing levels in each sample, (or effect size of $XIST_{ASE}$), the degree of skewing of XCI (DS) was calculated. DS is the absolute deviation of $XIST_{ASE}$ from 0.5. In each sample, DS was calculated as follow:

$$DS = |0.5 - XIST_{ASE}| \quad (\text{Degree of skewing of XCI})$$

DS does not take into account the direction of XCI skewing, but the degree of deviation from a 50% XCI patterns ($XIST_{ASE} = 0.5$). DS is then a measure of the magnitude of XCI skewing levels in a sample. DS values range from 0 to 0.5, where 0 means random XCI and 0.5 completely skewed XCI patterns. Samples with $DS \geq 0.3$ were classified to have skewed XCI, while samples with $DS < 0.3$ were classified to have random XCI.

Heritability analysis of DS

The relative contributions of additive genetic factors (A), shared (C) and unique environmental factors (E) to the tissue-specific variance of DS, were calculated using the *twinlm()* function in the *met*s R package⁶⁹. For each tissue, samples were split into a young (< 55) and an older ($\geq$ 55) group according to their ages (Table S1). Due to the low number of MZ and DZ twin pairs in each group, whole-blood was excluded from heritability analysis. To further assess the contribution of genetic effects, the intraclass spearman's correlation (IC) of DS in blood-derived tissues of young and older MZ and DZ twin pairs was also calculated.

Association between degree of skewing and smoking status

Association between the degree of skewing in LCLs and self-reported smoking status was tested in the 270 individuals with reliable smoking status recorded⁴⁵. Dataset included 270 females classified either as current smokers (N = 37) or never smokers (N = 233; Table S2). To examine the association between DS and smoking status, the smoking status was converted into a binary trait (0 = no smoker, 1 = smoker). A linear model of the DS as a function of the smoking status was then implemented for younger (age < 55) and older (age $\geq$ 55) individuals separately. Age was used as covariate. A *P* value $\leq$ 0.05 was considered to be statistically significant.

References

- 1 Lyon, M. F. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372–373 (1961).
- 2 van den Berg, I. M. *et al.* X chromosome inactivation is initiated in human preimplantation embryos. *Am J Hum Genet* **84**, 771–779, doi:10.1016/j.ajhg.2009.05.003 (2009).
- 3 Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44, doi:10.1038/349038a0 (1991).
- 4 Simon, M. D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465–469, doi:10.1038/nature12719 (2013).

- 413 5 Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to
414 spread across the X chromosome. *Science* **341**, 1237973, doi:10.1126/science.1237973
415 (2013).
- 416 6 Pinter, S. F. *et al.* Spreading of X chromosome inactivation via a hierarchy of defined
417 Polycomb stations. *Genome Res* **22**, 1864–1876, doi:10.1101/gr.133751.111 (2012).
- 418 7 Wutz, A. Gene silencing in X-chromosome inactivation: advances in understanding
419 facultative heterochromatin formation. *Nat Rev Genet* **12**, 542–553, doi:10.1038/nrg3035
420 (2011).
- 421 8 Gendrel, A. V. & Heard, E. Fifty years of X-inactivation research. *Development* **138**, 5049–
422 5055, doi:10.1242/dev.068320 (2011).
- 423 9 Puck, J. M., Nussbaum, R. L. & Conley, M. E. Carrier detection in X-linked severe combined
424 immunodeficiency based on patterns of X chromosome inactivation. *J Clin Invest* **79**, 1395–
425 1400, doi:10.1172/JCI112967 (1987).
- 426 10 Migeon, B. R. *et al.* Adrenoleukodystrophy: evidence for X linkage, inactivation, and
427 selection favoring the mutant allele in heterozygous cells. *Proc Natl Acad Sci U S A* **78**, 5066–
428 5070 (1981).
- 429 11 Ozbalkan, Z. *et al.* Skewed X chromosome inactivation in blood cells of women with
430 scleroderma. *Arthritis Rheum* **52**, 1564–1570, doi:10.1002/art.21026 (2005).
- 431 12 Chabchoub, G. *et al.* Analysis of skewed X-chromosome inactivation in females with
432 rheumatoid arthritis and autoimmune thyroid diseases. *Arthritis Res Ther* **11**, R106,
433 doi:10.1186/ar2759 (2009).
- 434 13 Kristiansen, M. *et al.* High frequency of skewed X inactivation in young breast cancer
435 patients. *J Med Genet* **39**, 30–33 (2002).
- 436 14 Buller, R. E., Sood, A. K., Lallas, T., Buekers, T. & Skilling, J. S. Association between
437 nonrandom X-chromosome inactivation and BRCA1 mutation in germline DNA of patients
438 with ovarian cancer. *J Natl Cancer Inst* **91**, 339–346 (1999).
- 439 15 Chitnis, S. *et al.* The role of X-chromosome inactivation in female predisposition to
440 autoimmunity. *Arthritis Res* **2**, 399–406, doi:10.1186/ar118 (2000).
- 441 16 Libert, C., Dejager, L. & Pinheiro, I. The X chromosome in immune functions: when a
442 chromosome makes the difference. *Nat Rev Immunol* **10**, 594–604, doi:10.1038/nri2815
443 (2010).
- 444 17 Busque, L. *et al.* Nonrandom X-inactivation patterns in normal females: lyonization ratios
445 vary with age. *Blood* **88**, 59–65 (1996).
- 446 18 Hatakeyama, C. *et al.* The dynamics of X-inactivation skewing as women age. *Clin Genet* **66**,
447 327–332, doi:10.1111/j.1399-0004.2004.00310.x (2004).
- 448 19 Knudsen, G. P., Pedersen, J., Klingenberg, O., Lygren, I. & Orstavik, K. H. Increased skewing
449 of X chromosome inactivation with age in both blood and buccal cells. *Cytogenet Genome*
450 *Res* **116**, 24–28, doi:10.1159/000097414 (2007).
- 451 20 Wong, C. C. *et al.* A longitudinal twin study of skewed X chromosome-inactivation. *PLoS One*
452 **6**, e17873, doi:10.1371/journal.pone.0017873 (2011).

- 453 21 Kristiansen, M. *et al.* Twin study of genetic and aging effects on X chromosome inactivation.
454 *Eur J Hum Genet* **13**, 599–606, doi:10.1038/sj.ejhg.5201398 (2005).
- 455 22 Christensen, K. *et al.* X-linked genetic factors regulate hematopoietic stem-cell kinetics in
456 females. *Blood* **95**, 2449–2451 (2000).
- 457 23 Vickers, M. A., McLeod, E., Spector, T. D. & Wilson, I. J. Assessment of mechanism of
458 acquired skewed X inactivation by analysis of twins. *Blood* **97**, 1274–1281 (2001).
- 459 24 Bolduc, V. *et al.* No evidence that skewing of X chromosome inactivation patterns is
460 transmitted to offspring in humans. *J Clin Invest* **118**, 333–341, doi:10.1172/JCI33166
461 (2008).
- 462 25 Gale, R. E., Wheadon, H., Boulos, P. & Linch, D. C. Tissue specificity of X-chromosome
463 inactivation patterns. *Blood* **83**, 2899–2905 (1994).
- 464 26 Bittel, D. C. *et al.* Comparison of X-chromosome inactivation patterns in multiple tissues
465 from human females. *J Med Genet* **45**, 309–313, doi:10.1136/jmg.2007.055244 (2008).
- 466 27 Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature*
467 **550**, 244–248, doi:10.1038/nature24265 (2017).
- 468 28 Moayyeri, A., Hammond, C. J., Hart, D. J. & Spector, T. D. The UK Adult Twin Registry
469 (TwinsUK Resource). *Twin Res Hum Genet* **16**, 144–149, doi:10.1017/thg.2012.89 (2013).
- 470 29 Buil, A. *et al.* Gene–gene and gene–environment interactions detected by transcriptome
471 sequence analysis in twins. *Nat Genet* **47**, 88–91, doi:10.1038/ng.3162 (2015).
- 472 30 Gentilini, D. *et al.* Age-dependent skewing of X chromosome inactivation appears delayed in
473 centenarians' offspring. Is there a role for allelic imbalance in healthy aging and longevity?
474 *Aging Cell* **11**, 277–283, doi:10.1111/j.1474-9726.2012.00790.x (2012).
- 475 31 Busque, L. *et al.* Recurrent somatic TET2 mutations in normal elderly individuals with clonal
476 hematopoiesis. *Nat Genet* **44**, 1179–1181, doi:10.1038/ng.2413 (2012).
- 477 32 Cotton, A. M. *et al.* Analysis of expressed SNPs identifies variable extents of expression from
478 the human inactive X chromosome. *Genome Biol* **14**, R122, doi:10.1186/gb-2013-14-11-
479 r122 (2013).
- 480 33 Naumova, A. K. *et al.* Heritability of X chromosome—inactivation phenotype in a large
481 family. *Am J Hum Genet* **58**, 1111–1119 (1996).
- 482 34 Allen, R. C., Zoghbi, H. Y., Moseley, A. B., Rosenblatt, H. M. & Belmont, J. W. Methylation of
483 Hpall and HhaI sites near the polymorphic CAG repeat in the human androgen–receptor gene
484 correlates with X chromosome inactivation. *Am J Hum Genet* **51**, 1229–1239 (1992).
- 485 35 Mossner, M. *et al.* Skewed X-inactivation patterns in ageing healthy and myelodysplastic
486 haematopoiesis determined by a pyrosequencing based transcriptional clonality assay. *J Med*
487 *Genet* **50**, 108–117, doi:10.1136/jmedgenet-2012-101093 (2013).
- 488 36 Bryois, J. *et al.* Time-dependent genetic effects on gene expression implicate aging
489 processes. *Genome Res* **27**, 545–552, doi:10.1101/gr.207688.116 (2017).
- 490 37 de Hoon, B., Monkhorst, K., Riegman, P., Laven, J. S. & Gribnau, J. Buccal swab as a reliable
491 predictor for X inactivation ratio in inaccessible tissues. *J Med Genet* **52**, 784–790,
492 doi:10.1136/jmedgenet-2015-103194 (2015).

493 38 Pastinen, T. *et al.* A survey of genetic and epigenetic variation affecting human gene
494 expression. *Physiol Genomics* **16**, 184–193, doi:10.1152/physiolgenomics.00163.2003
495 (2004).

496 39 Gale, R. E., Fielding, A. K., Harrison, C. N. & Linch, D. C. Acquired skewing of X-chromosome
497 inactivation patterns in myeloid cells of the elderly suggests stochastic clonal loss with age.
498 *Br J Haematol* **98**, 512–519 (1997).

499 40 Tonon, L. *et al.* Unbalanced X-chromosome inactivation in haemopoietic cells from normal
500 women. *Br J Haematol* **102**, 996–1003 (1998).

501 41 Gao, X., Jia, M., Zhang, Y., Breitling, L. P. & Brenner, H. DNA methylation changes of whole
502 blood cells in response to active smoking exposure in adults: a systematic review of DNA
503 methylation studies. *Clin Epigenetics* **7**, 113, doi:10.1186/s13148-015-0148-3 (2015).

504 42 Ma, Y. & Li, M. D. Establishment of a Strong Link Between Smoking and Cancer Pathogenesis
505 through DNA Methylation Analysis. *Sci Rep* **7**, 1811, doi:10.1038/s41598-017-01856-4
506 (2017).

507 43 Mahdi, H. *et al.* Specific interaction between genotype, smoking and autoimmunity to
508 citrullinated alpha-enolase in the etiology of rheumatoid arthritis. *Nat Genet* **41**, 1319–1324,
509 doi:10.1038/ng.480 (2009).

510 44 Klebaner, D. *et al.* X chromosome-wide analysis identifies DNA methylation sites influenced
511 by cigarette smoking. *Clin Epigenetics* **8**, 20, doi:10.1186/s13148-016-0189-2 (2016).

512 45 Tsai, P. C. *et al.* Smoking induces coordinated DNA methylation and gene expression changes
513 in adipose tissue with consequences for metabolic health. *Clin Epigenetics* **10**, 126,
514 doi:10.1186/s13148-018-0558-0 (2018).

515 46 Andreu, N. *et al.* Wiskott–Aldrich syndrome in a female with skewed X-chromosome
516 inactivation. *Blood Cells Mol Dis* **31**, 332–337 (2003).

517 47 Li, G. *et al.* Skewed X chromosome inactivation of blood cells is associated with early
518 development of lung cancer in females. *Oncol Rep* **16**, 859–864 (2006).

519 48 Abelson, S. *et al.* Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature*
520 **559**, 400–404, doi:10.1038/s41586-018-0317-6 (2018).

521 49 Loh, P. R. *et al.* Insights into clonal haematopoiesis from 8,342 mosaic chromosomal
522 alterations. *Nature* **559**, 350–355, doi:10.1038/s41586-018-0321-x (2018).

523 50 Zink, F. *et al.* Clonal hematopoiesis, with and without candidate driver mutations, is common
524 in the elderly. *Blood* **130**, 742–752, doi:10.1182/blood-2017-02-769869 (2017).

525 51 Steensma, D. P. *et al.* Clonal hematopoiesis of indeterminate potential and its distinction
526 from myelodysplastic syndromes. *Blood* **126**, 9–16, doi:10.1182/blood-2015-03-631747
527 (2015).

528 52 Steensma, D. P. Clinical consequences of clonal hematopoiesis of indeterminate potential.
529 *Blood Adv* **2**, 3404–3410, doi:10.1182/bloodadvances.2018020222 (2018).

530 53 Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA
531 sequence. *N Engl J Med* **371**, 2477–2487, doi:10.1056/NEJMoa1409405 (2014).

532 54 Coombs, C. C. *et al.* Therapy-Related Clonal Hematopoiesis in Patients with Non-
533 hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem*
534 *Cell* **21**, 374–382 e374, doi:10.1016/j.stem.2017.07.010 (2017).

535 55 Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in
536 twins. *Nat Genet* **44**, 1084–1089, doi:10.1038/ng.2394 (2012).

537 56 Consortium, U. K. *et al.* The UK10K project identifies rare variants in health and disease.
538 *Nature* **526**, 82–90, doi:10.1038/nature14962 (2015).

539 57 Zheng, H. F. *et al.* Performance of genotype imputation for low frequency and rare variants
540 from the 1000 genomes. *PLoS One* **10**, e0116487, doi:10.1371/journal.pone.0116487
541 (2015).

542 58 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human
543 genomes. *Nature* **491**, 56–65, doi:10.1038/nature11632 (2012).

544 59 Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research.
545 *Nature* **477**, 54–60, doi:10.1038/nature10354 (2011).

546 60 Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease
547 and population genetic studies. *Nat Methods* **10**, 5–6, doi:10.1038/nmeth.2307 (2013).

548 61 Delaneau, O., Marchini, J., Genomes Project, C. & Genomes Project, C. Integrating sequence
549 and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat*
550 *Commun* **5**, 3934, doi:10.1038/ncomms4934 (2014).

551 62 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
552 *Bioinformatics* **25**, 1754–1760, doi:10.1093/bioinformatics/btp324 (2009).

553 63 Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE
554 Project. *Genome Res* **22**, 1760–1774, doi:10.1101/gr.135350.111 (2012).

555 64 Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets.
556 *Bioinformatics* **27**, 863–864, doi:10.1093/bioinformatics/btr026 (2011).

557 65 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21,
558 doi:10.1093/bioinformatics/bts635 (2013).

559 66 Kopylova, E., Noe, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs
560 in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217,
561 doi:10.1093/bioinformatics/bts611 (2012).

562 67 van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific software for
563 robust molecular quantitative trait locus discovery. *Nat Methods* **12**, 1061–1063,
564 doi:10.1038/nmeth.3582 (2015).

565 68 Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and best
566 practices for data processing in allelic expression analysis. *Genome Biol* **16**, 195,
567 doi:10.1186/s13059-015-0762-6 (2015).

568 69 Scheike, T. H., Holst, K. K. & Hjelmberg, J. B. Estimating heritability for cause specific
569 mortality based on twin studies. *Lifetime Data Anal* **20**, 210–233, doi:10.1007/s10985-013-
570 9244-x (2014).

571

Acknowledgements

This study was supported by MRC Project Grant (MR/R023131/1) to K.S.S. The TwinsUK study was funded by the Wellcome Trust and European Community's Seventh Framework Programme (FP7/2007-2013). The TwinsUK study also receives support from the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. This project was enabled through access to the MRC eMedLab Medical Bioinformatics infrastructure, supported by the Medical Research Council [grant number MR/L016311/1]

Authors contributions

A.Z. M.D. and K.S.S conceived and designed the project. A.Z performed analysis. SN contributed data discussion. P.C.T and J.T.B. contributed data. SR and C.Y.W performed HUMARA experiments. A.Z. and K.S.S. wrote the manuscript. All authors read and approved the manuscript. The authors also thank Julia El-Sayed Moustafa and Amy Roberts for providing feedback on the manuscript.

Competing interests

The authors declare that they have no competing interests.

Materials and Correspondence

Correspondence should be addressed to Kerrin S. Small (kerrin.small@kcl.ac.uk)

Data Availability

TwinsUK RNAseq data is available from EGA (Accession number: EGAS00001000805).
TwinsUK genotypes are available upon application to TwinsUK (www.twinsuk.ac.uk).

FIGURES

Fig. 1 Comparison to HUMARA and longitudinal quantifications of XCI skewing levels reveal that $XIST_{ASE}$ is a reproducible and accurate proxy of XCI skewing levels. **a** Scatter plot showing the comparison between the XCI skewing levels in 10 whole-blood and LCLs samples quantified with HUMARA and $XIST_{ASE}$ respectively. **b** Scatter plot of the $XIST_{ASE}$ at time point 1 and at time point 2 in 16 whole-blood samples.

Fig. 2 Skewed chromosome X inactivation varies across tissues with a higher prevalence of skewed samples in blood-derived tissues than fat and skin tissues. Distribution of the degree of skewing (top row) and $XIST_{ASE}$ (bottom row) in LCLs (422 samples), whole-blood (72 samples), fat (378 samples) and skin (336 samples) tissues.

Fig. 3 Pairwise tissue-tissue comparison of the $XIST_{ASE}$ calls reveal highest similarity of XCI skewing levels between blood-derived tissues. Each plot shows the comparison of $XIST_{ASE}$ between two tissues from an individual. Each dot represents an individual. In each comparison all individuals with available $XIST_{ASE}$ calls for both tissues were included.

Fig. 4 Skewed X-chromosome inactivation (XCI) increases with age in blood-derived tissues but not in fat or skin tissues. Left panel, *boxplots* show the degree of skewing in LCLs samples <55 and ≥ 55. Right panel, *line plots* show the frequencies of skewed XCI at different ages in LCLs (blue), fat (red) and skin (green) tissues.

TABLES

Tissue	Informative individuals	Skewed XCI individuals	Prevalence of skewed XCI
LCLs	422	145	34%
Whole-blood	72	20	28%
Fat	378	45	12%
Skin	336	54	16%

Table 1. Prevalence of skewed X-inactivation (XCI) differs across tissues in the TwinsUK cohort.

Tissue	LCLs	Whole-blood	Fat	Skin
LCLs	*	0.78 [2e-13]	0.42 [5e-14]	0.3 [1e-06]
Whole-blood	0.78 [2e-13]	*	0.33 [1e-02]	0.3 [4e-02]
Fat	0.42 [5e-14]	0.33 [1e-02]	*	0.47 [2e-15]
Skin	0.3 [1e-06]	0.3 [4e-02]	0.47 [2e-15]	*

Table 2. Correlation of XCI-skew varies between tissue pairs, and is highest between blood-derived tissues. Coefficient of correlation [and its p-value] of XCI-skew between tissue pairs.

Cell types	Degree of skewing in Twin A	Degree of skewing in Twin B	Degree of skewing in Individual C
LCLs	0.45	0.45	0.1
Monocytes	0.43	0.42	0.05
B-cells	0.4	0.42	NA
T-CD4 ⁺ cells	0.28	NA	0.05
T-CD8 ⁺ cells	0.44	0.3	0.06
NK-cells	0.4	NA	0.01

Table 3. XCI-skew in LCLs in this study is representative of XCI in purified primary immune cells.

Degree of skewing of XCI in immune cell types purified from 2 monozygotic twins (Twin A, Twin B) exhibiting skewed XCI patterns in LCLs, and from 1 individual (Individual C) exhibiting random XCI pattern in LCLs. Degree of skewing ≥ 0.3 indicates skewed XCI patterns.

Tissue [age group]	Additive genetics (h^2)	Common Environment	Unique Environment
LCLs [<55]	0	0.01	0.99
LCLs [≥ 55]	0.34	0	0.66
Fat [<55]	0	0.27	0.73
Fat [≥ 55]	0	0.08	0.92
Skin [<55]	0	0	1
Skin [≥ 55]	0	0	1

Table 4. XCI-skew is heritable in blood-derived tissues of older females. Estimates of the relative contribution of additive genetics, common environment and unique environmental factors to the tissue-specific XCI-skew in age-stratified twins.

	MZ_{allAges}	MZ_{young}	MZ_{older}	DZ_{allAges}	DZ_{young}	DZ_{older}
n. pairs	61	18	43	63	25	38
Intra-class correlation	0.31	0.06	0.42	0	0.1	-0.1
P-value	0.02	0.8	0.005	0.9	0.6	0.7

Table 5. Intra-class correlations of XCI skew in age stratified twin pairs. Twin pairs < 55 years-old are classified as young, twin pairs ≥ 55 years-old are classified as old.