

Escape from X-inactivation in twins exhibits intra- and inter-individual variability across tissues and is heritable

Zito A.^{1,2}, Roberts A.L.^{1*}, Visconti A.^{1*}, Rossi N.¹, Andres-Ejarque R.³, Nardone S.⁴, Moustafa J.E.S.¹, Falchi M.¹, Small K.S.¹.

¹Department of Twin Research and Genetic Epidemiology, King's College London, London, UK;

²Present address: Department of Molecular Biology, Massachusetts General Hospital, Boston, USA; Department of Genetics, Harvard Medical School, Boston, USA;

³St John's Institute of Dermatology, Faculty of Life Science & Medicine, King's College London, London, UK;

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

(*) equal contribution

ABSTRACT

X-chromosome inactivation (XCI) silences one X-chromosome in female cells to balance sex-differences in X-dosage. A subset of X-linked genes escape XCI, but the extent to which this phenomenon occurs and how it varies across tissues and in a population is as yet unclear. In order to characterize the incidence and variability of escape across individuals and tissues, we conducted a large scale transcriptomic study of XCI escape in adipose, skin, lymphoblastoid cell lines (LCLs) and immune cells in 248 twins drawn from a healthy population cohort. We identify 159 X-linked genes with detectable escape, of which 54 genes, including 19 lncRNAs, were not previously known to escape XCI. Across tissues we find a range of tissue-specificity, with 11% of genes escaping XCI constitutively across tissues and 24% demonstrating tissue-restricted escape, including genes with cell-type specific escape between immune cell types (B, T-CD4⁺, T-CD8⁺ and NK cells) of the same individual. Escape genes interact with autosomal-encoded proteins and are involved in varied biological processes such as gene regulation. We find substantial variability in escape between individuals. 49% of genes show inter-individual variability in escape, indicating escape from XCI is an under-appreciated source of gene expression differences. We utilized twin models to investigate the role of genetics in variable escape. Overall, monozygotic (MZ) twin pairs share more similar escape than dizygotic twin pairs, indicating that genetic factors underlie differences in escape across individuals. However, we also identify instances of discordant XCI within MZ co-twin pairs, suggesting that environmental factors also influence escape. Thus, XCI escape may be shaped by an interplay of genetic factors with tissue- and cell type-specificity, and environment. These results illuminate an intricate phenotype whose characterization aids understanding the basis of variable trait expressivity in females.

1 Introduction

2 The X chromosome is a paradigmatic genetic model¹. It carries >1000 genes, representing
3 >5% of the haploid human genome. It is differentially inherited between the sexes. The unequal X-
4 linked transcriptional dosage between the sexes is partially compensated by random silencing of
5 one X in each female somatic cell². This process, known as X-chromosome inactivation (XCI), involves
6 synergistic DNA-RNA-protein interactions that mediate heterochromatinization of the X designated
7 for inactivation^{1,3,4} (known as "Barr Body"⁵). Non-coding RNAs play key roles in XCI. The master long
8 non-coding (lnc) RNA *XIST* spreads in *cis* along the inactive X chromosome (Xi) and promotes a
9 progressive epigenetic silencing^{4,6,7}. However, XCI is incomplete, with over 15% of X-genes reported
10 to escape silencing and exhibit expression from both parental alleles within a diploid cell^{8,9}. Mary
11 Lyon predicted that genes with Y-homologues (e.g. pseudo autosomal regions (PARs)), are naturally
12 dosage compensated and thus expected to escape¹⁰. Today, most known escapees lack functional
13 Y-homologues, thus being a potential source of sexual dimorphism^{9,11}. Chromosome X is enriched
14 for genes with immune- and neuro-modulatory functions^{12,13}; changes in escape can thus underpin
15 not only sexual dimorphism, but also phenotypic and disease risk variability across females^{12,14,15}.
16 Despite its biomedical relevance, the inter-individual variability of escape at population level, and
17 across cells and tissues within an individual, has not been systematically examined. Furthermore,
18 the extent to which genetics and environment influence the escape remains largely undefined.

19 Our current knowledge of XCI escape in humans largely rely on conventional studies of
20 male/female expression ratio, human/mouse hybrid cells, and epigenetic marks^{8,11,16,17}. In most
21 females, the random nature of XCI results in expression of both X-linked alleles at a tissue level. This
22 limits the ability to distinguish mono- from biallelic expression and thus identification of escape⁹. To
23 circumvent the problem, strategies like single-cell analyses (e.g. scRNAseq) or sex comparison have
24 been used to infer escape^{9,11,18,19}. However, scRNAseq is infeasible for large cohorts and is limited
25 to highly and consistently expressed genes due to allelic drop-out and transcriptional burst, which
26 both can inflate monoallelic expression ratio. On the other hand, sex differences may not directly
27 reflect the allelic expression ratio of X-genes in a tissue. These limitations can be circumvented by
28 using tissue samples exhibiting skewed XCI patterns – a common event in the female population²⁰
29 – which, as opposed to random XCI, enable detection and measurement of escape directly in skewed
30 females (Note S1). This strategy has been employed, but either sample sizes were limited (e.g. single
31 GTEx donor), or the study relied on arrays and other artificial biological models^{8,11,21}.

32 Here, we characterize XCI escape using paired bulk RNAseq and high-coverage DNaseq data
33 in a multi-tissue dataset sampled from 248 skewed female twins of the TwinsUK bioresource²². We
34 investigate escape prevalence and variability across adipose and skin tissue, lymphoblastoid cell
35 lines and purified immune cells (monocytes, B-cells, T-CD4⁺, T-CD8⁺ and NK cells), and individuals.
36 We identify novel genes exhibiting tissue- and immune cell type-specific escape, and genes escaping
37 XCI with high variability across tissues and individuals. We observe that escape varies across tissues
38 and immune cells within an individual and across individuals, a phenomenon with high biomedical
39 relevance. Using twins, we demonstrate that regulation of XCI escape has both heritable and
40 environmental components, implying a complex interplay between genetic and non-genetic factors.

41 RESULTS

42 Escape from XCI is more prevalent in solid tissues and wider than currently known

43 We quantified escape in multiple tissues concurrently sampled from female twins of the
44 TwinsUK cohort²². We determined XCI patterns using the gene-level *XIST* allele-specific expression
45 ($XIST_{ASE}$) from paired RNAseq and DNAseq data^{7,23,24}. From over 2200 tissue samples interrogated,
46 we obtained $XIST_{ASE}$ calls for 522 LCLs, 101 whole-blood, 421 adipose, and 373 skin samples. In
47 samples exhibiting skewed XCI ($0.2 \geq XIST_{ASE} \geq 0.8$, Methods) including 166 LCLs (32%), 26 whole-
48 blood (26%), 57 adipose (14%), and 64 skin (17%) samples, the levels of escape of each X-linked gene
49 were measured using a metric - herein referred to as Escape Score or 'EscScore' - derived from the
50 gene's allelic fold-change adjusted for the sample's XCI skew (Methods). EscScore values range from
51 0 (no escape, monoallelic expression) to 1 (full escape, equal expression from inactive (Xi) and active
52 X (Xa)). We interrogated a total of 551 genes, of which 85% are protein-coding and 15% non-coding
53 RNA genes. Based on a publicly available catalogue of XCI statuses²⁵ ('Balaton's list'), our interrogated
54 genes were categorized as XCI-silenced (n=326), fully or mostly escaping XCI (n=52), or variable
55 escapees (n=23). Variable escape refers to genes whose escape is variable across cells, tissues, or
56 individuals^{8,11,25}. We also included a subset of 41 genes whose XCI status was reported as discordant
57 across studies or undefined²⁵. The summary statistics indicated that EscScore differs between
58 different categories (Table S1). In all tissues, the EscScore of genes annotated as fully or mostly
59 escaping XCI significantly differed from genes annotated as either silenced or variable escapees
60 (Fig.1; Table S2), supporting the reliability of our escape metric to discriminate different XCI statuses.

61 By defining the allele with the lower RNAseq count as the Xi allele while the other as Xa
62 allele, we detected an average Xi/Xa expression ratio of 0.32, 0.36, and 0.37 across LCLs, adipose
63 and skin samples, respectively, with median value of 0.36. This is consistent with the GTEx survey of
64 XCI, which reported that the Xi expression is on average 33% of Xa expression¹¹. Based on these
65 observations, we classified genes with a median (across ≥ 3 skewed tissue samples) $EscScore \geq 0.36$
66 as escapees in that tissue, while genes with $EscScore < 0.36$ as silenced. In line with current
67 knowledge, most X-genes are subject to XCI in all tissues (84% in LCLs, 74% in adipose, 71% in skin).
68 We observed a higher incidence of escape in solid tissues than LCLs, with 16%, 26% and 29% of
69 genes escaping XCI in LCLs, adipose and skin tissues, respectively. Altogether, 159 X-genes exhibited
70 escape in at least one tissue in our dataset (Table S3). Expectedly, PAR-linked genes escaped XCI.
71 We used a hypergeometric test to assess overlap between the Balaton's list²⁵ and our list of
72 escapees, and found significant overlap (N=50; $P \leq 0.05$). Among our escape calls, about 60 genes
73 retain a Y-pseudogene or Y-homologue, and 13 are PAR-linked, supporting their escape. For a more
74 comprehensive comparison, we merged the Balaton's list of escapees with additional external lists
75 of escapees from other studies including (i) the GTEx XCI survey¹¹; (ii) Zhang *et al.*²⁶; (iii) Katsir
76 *et al.*¹⁹; (iv) Garieri *et al.*¹⁸. We found that 66% (N=105) of our escapees overlapped with this unified
77 list of escapees ($P \leq 0.05$), while the remaining (N=54) represent novel calls. We confirmed the
78 escape status of *CLIC2* in skin, as identified in GTEx¹¹, and also found it escapes in adipose in our
79 data. Notably, 35% of our novel escape calls are annotated as lncRNA, while the remaining are
80 protein-coding. We establish the first evidence that the lncRNA *AL683807.1* escapes XCI.
81 *AL683807.1* is PAR1-linked, explaining its escape ability. We examined the chromosomal distribution

82 of escape genes and confirmed a higher escape incidence on the short arm²⁷ (Note S2; Fig.S1).
83 Altogether, these data further support the suitability of our study design. We show that in our data,
84 escape is more prevalent in adipose and skin than LCLs, and is wider than currently known.

85 **Escape from X-inactivation exhibits both constitutive and tissue-specific patterns**

86 Presently, the extent to which tissue-specific escape occurs in humans is unclear. To study
87 this, we calculated the gene's median EscScore (across ≥ 3 tissue samples) as a measure for tissue-
88 specific levels of escape. We found significant differences across tissues (Kruskal-Wallis ('KW') *P*-
89 *value* $<10^{-10}$; Fig.S2), suggesting tissue-specific components of escape. Using a subset of 215 genes
90 with EscScore available in all tissues (Table S4), we identified 24 genes exhibiting escape in both
91 LCLs and solid tissues (Fig.1B), suggesting constitutive escape. We observed that tissue-specific
92 EscScore remained below 80% in most cases, in line with data showing that Xi/Xa expression ratio
93 would not exceed 80%¹¹. Notably, *PLCXD1*, *ASMTL*, *DHRX*, *SLC25A6* and *AKAP17A* are PAR-linked.
94 We show that *PUDP* and *PIN4*, whose escape status have remained so far unclear, show constitutive
95 escape in all tissues. Among the constitutive escapees, there are the highly biomedically-relevant
96 genes *DDX3X*, *KDM5C* and *KDM6A*, whose escape may contribute to lower cancer incidence in
97 females than males¹⁴. We also observed that *PRKX*, *PUPD*, *DDX3X* and *JPX* each had significantly
98 different escape between tissues (KW $P \leq 0.05$). Tissue-specificity of escape was further supported
99 by identification of 51 genes exhibiting escape restricted to a single tissue. Notably, 18 of these, of
100 which 3 non-coding RNAs and 15 protein-coding genes, are novel escape calls (Table S5).

101 We investigated whether the escapees may interact with other factors and be involved in
102 biological processes. To address this, we selected genes exhibiting escape in at least one tissue and
103 conducted protein-protein interaction network analysis using a recent human protein interactome
104 as reference²⁸. We found that protein-coding escape genes interact with other factors on a genome-
105 wide scale (Fig.1C). Gene ontology analysis revealed that members of this proteome network are
106 involved in distinct biological processes such as epigenetic regulation by chromatin assembly and
107 nucleosome organization, and regulation of steroid hormone signaling. These data were supported
108 by REACTOME pathway analysis which revealed multiple pathways for epigenetic control of genes,
109 such as histone methylation and acetylation, and DNA methylation (Fig.1D, Table S6). Altogether,
110 these data suggest that escape is shaped by an interplay of tissue-shared and tissue-specific factors,
111 and participates in genome-wide interactions involved in varied biological processes.

112 **Escape from X-inactivation exhibits intra- and inter-individual variability**

113 The extent to which escape varies across tissues within an individual is unclear. We
114 investigated this phenomenon in a subset of 6 donors exhibiting skewed XCI in LCLs, adipose and
115 skin. This strategy was employed in the GTEx survey using a single skewed female donor¹¹. Within
116 each donor, we examined all genes with available EscScore. We identified genes exhibiting escape
117 (EscScore ≥ 0.36 in a tissue within a donor) in 1, 2 or 3 tissues and found that their prevalence varied
118 between donors (Fig.2A; Table S7). Occurrence of genes escaping XCI in all tissues in multiple donors
119 suggests shared regulatory mechanisms across tissues and individuals. Genes exhibiting such a

120 behaviour included the zinc-finger protein *ZFX* and the histone demethylase *KDM5C* which is linked
121 to intellectual disability and autism^{29,30}. We also observed genes escaping XCI in all tissues but in
122 only 1 of the 6 donors. Examples are the leukaemia-protecting histone demethylase *KDM6A*^{14,31},
123 *FMR1*, a gene linked to Fragile-X and learning disability³², and the Duchenne muscular dystrophy
124 gene *DMD*³³. We observed that *ASMTL* escaped XCI in all tissues in a donor, and the degree to which
125 it escaped varied across tissues ranging from 0.6 to over 0.8. *ASMTL*'s behavior was also highlighted
126 in GTEX¹¹. Aside from these cases, most of the interrogated genes exhibited tissue-restricted escape,
127 supporting the occurrence of tissue-specific factors exerting dominant effects.

128 To robustly investigate the inter-female diversity in escape, we used a subset of 127 genes
129 escaping XCI in at least 1 tissue, and with EscScore available in at least 10 individuals per tissue. For
130 a given gene, we defined its EscScore to be consistent within a tissue if in at least 80% of individuals
131 the gene's EscScore lay between ± 1 standard deviation from the gene's average EscScore in the
132 tissue. This strategy revealed both genes with consistent and genes with variable EscScore. We
133 identified 40 genes (~31% of the interrogated genes) showing consistent EscScore across individuals
134 in at least 1 tissue (Fig.S3). Representative examples are *BTK*, a gene involved in the control of
135 lymphocyte maturation, and *CD99L2*, involved in leucocyte homeostasis. Both genes exhibited
136 consistent EscScore across most donors in LCLs (Fig.2B). A subset of 3 genes (*ARHGAP6*, *SAT1*,
137 *RAP2C-AS1*) also exhibited consistency across most donors in both LCLs and a solid tissue (Fig.S3).
138 Genes exhibiting inter-female variability in EscScore in multiple tissues accounted for about 49%
139 (N=62) of the interrogated genes (Fig.S4). Examples are *DDX3X*, *KDM6A* and *UBA1*, which exhibited
140 variability in LCLs and at least one solid tissue (Fig.2B). Interestingly, inter-female variability occurred
141 more frequently in solid tissues (63% of cases) than LCLs (37% of cases). Altogether, these data are
142 indicative of complex escape patterns. Variable escape across females complements with and may
143 be driven by variable escape across tissues and cells within a female. Inter-female variation has high
144 biomedical relevance as it may underlie predisposition to and manifestation of X-linked traits.

145 **Escape from X-inactivation exhibits immune cell type-specificity**

146 Females have a higher risk of autoimmune disease than males, and such risk may correlate
147 with increased X-dosage^{34,35}. This has raised the hypothesis that XCI escape may contribute to
148 autoimmunity^{12,36}. The extent to which escape varies across different immune cells within an
149 individual is not well known. We addressed this question by interrogating 257 X-genes in multiple
150 immune cell types purified from two identical co-twins (Fig.S5). Monocytes, B-cells and T-CD8⁺ cells
151 were available from both co-twins, while T-CD4⁺ and NK-cells from one co-twin. Per each immune
152 cell type, when data were available from both co-twins, we calculated the average gene's EscScore
153 across the 2 co-twins as a proxy for immune cell type-specific escape. We observed differences
154 between cell types in the average EscScore (Av.EscScore_{Monocytes}=0.24; Av.EscScore_{B-cells}=0.27;
155 Av.EscScore_{T-CD4+}=0.24; Av.EscScore_{T-CD8+}=0.28; Av.EscScore_{NK-cells}=0.24; KW $P \leq 0.01$; Fig.3A). These
156 results were consistent when comparison was limited to a subset of 53 genes with EscScore data
157 available in all cell types (Table S8; Av.EscScore_{Monocytes}=0.24; Av.EscScore_{B-cells}=0.27; Av.EscScore_{T-}
158 _{CD4+}=0.26; Av.EscScore_{T-CD8+}=0.28; Av.EscScore_{NK-cells}=0.24; $P \leq 0.01$). The incidence of escape varied
159 between cell types, being 15% in monocytes, 20% in B-cells, 22% in T-CD4⁺, 25% in T-CD8⁺, and 28%

160 in NK-cells. Thus, in line with current knowledge, most X-genes are subject to XCI in immune cells.
161 In parallel, our data indicate that escape is heterogeneous across immune cell types, with overall
162 higher incidence in lymphocytes than monocytes. To investigate intra-lineage variation, we
163 compared the EscScore(s) between lymphoid cell types and also found substantial differences
164 ($P \leq 0.01$), indicating intra-lineage variation. Among the 53 genes with EscScore data available in all
165 cell types (Table S8), we identified 12 genes (*ARSD*, *PRKX*, *PUDP*, *CA5B*, *AP1S2*, *ZFX*, *USP9X*, *DDX3X*,
166 *CASK*, *KDM6A*, *JPX*, *DIAPH2*) escaping XCI in at least three immune cell types. *CASK* is a novel
167 candidate escapee. The gene subset *PRKX*, *ZFX*, *JPX* and *DIAPH2* escaped XCI in all 5 immune cell
168 types, in line with their behavior as constitutive escapees observed above. For most of these genes,
169 the escape status in immune cells is a novel finding. Interestingly, *KDM6A* exhibited highest EscScore
170 in T-CD8⁺ cells, possibly because of its roles in T-cells control³¹. We identified 9 genes exhibiting
171 escape restricted to one immune cell type, supporting immune cell type-specific factors (Fig.3B).
172 Intriguingly, immune cell type-specific events were restricted to lymphocytes but not monocytes.
173 This might suggest differences between lymphoid and myeloid lineages, and aligns with evidence of
174 increased X-linked biallelic expression in lymphocytes³⁷. Altogether, these data indicate that escape
175 varies between immune cell types within an individual. Presumably, this heterogeneity is driven by
176 mechanisms with immune cell type-specific effects.

177 **Escape from X-inactivation is influenced by heritable and environmental factors**

178 Twin studies are a unique strategy to assess the contribution of genetic factors to complex
179 traits. Using 27 complete twin pairs (17 monozygotic (MZ or identical); 10 dizygotic (DZ or
180 fraternal)), we quantified the concordance in the escape in LCLs between co-twins and compared
181 such concordance between MZ and DZ twins. We correlated the EscScore (using ≥ 5 genes) between
182 co-twins of each pair (Fig.4A-B), and found that the average correlation across MZ and DZ twins was
183 0.6 and 0.46, respectively (ρ 's t-test, $P \leq 0.05$; Fig.4C). These data indicate that MZ share significantly
184 more similar escape than DZ twins. To support this finding, we examined each interrogated gene in
185 a twin pair, and observed higher rates of discordant XCI (gene escaping XCI only in one of the two
186 co-twins) between DZ than MZ twin pairs (Av.Disc.Rate_{DZ}=27%; Av.Disc.Rate_{MZ}=19%). These data
187 suggest a significant genetic component of escape, in line with previous data on concordance in
188 methylation-based XCI status between MZ twins¹⁶. In parallel, discordance between MZ twins
189 suggests environmental influences. To gain insights at the cell type level, we next examined the
190 concordance of EscScore in immune cells between two MZ co-twins, and observed significant
191 correlation ($\rho_{\text{monocytes}}=0.8$; $\rho_{\text{B-cells}}=0.68$; $\rho_{\text{T-CD8}^+}=0.6$; $P < 1e-10$). Genes with discordant XCI status were
192 observed in all three immune cell types, and their prevalence differed between cells ranging from
193 6.4% in monocytes to 11% in B-cells, and 18% in T-CD8⁺ cells. Interestingly, the genes *CA5B* and
194 *ZNF81* exhibited discordant XCI between the two co-twins in both T-CD8⁺ and B-cells. In all other
195 cases, discordant XCI events concerned distinct gene subsets in distinct immune cell types (Fig 3C).
196 Taken together, our data indicate that genetic and environmental factors may interplay to regulate
197 XCI escape. Variability between immune cell types may also suggest an immune cell type-specific
198 response to environmental influences.

199 Discussion

200 Most knowledge of the extent of escape from X-inactivation in humans is currently based on
201 indirect proxies such as sex comparison, hybrid cell lines, or epigenome analyses^{8,11,16,17}. Here, we
202 studied escape in tissues and immune cells using paired transcriptomic and genotype data from
203 nearly 250 female twins from the TwinsUK bioresource²². The large sample size and strategy of using
204 tissue samples with skewed XCI^{8,9,11,21,38}, enabled us to systematically distinguish silenced from
205 escape genes and identify novel candidate escapees, as predicted⁹. We demonstrate that escape
206 from XCI is a complex phenotype. The incidence of escape varies across tissues, and is higher in solid
207 than blood tissues. Most of our escape calls align with previously annotated XCI statuses, however
208 we identified 54 novel candidate escapees, of which 35 (65%) are protein-coding and 19 (35%)
209 lncRNA genes. We demonstrated via protein network analysis that X-linked protein-coding genes
210 escaping XCI interact with other proteins on a genome-wide scale, and regulate varied biological
211 processes and pathways such as epigenetic changes and hormone signaling. These data indicate
212 that XCI escape has genome-wide effects, in line with recent findings on the effects of XCI changes
213 on global proteome³⁹. Thus, de-regulated escape caused by mutations or other disrupting events
214 may have complex phenotypic consequences. Future studies are needed to investigate the
215 functional effects of X-autosomal interactome. Our discovery set includes the PAR1 lncRNA
216 *AL683807.1*, which escaped XCI in LCLs, in line with its expression profile in GTEx. The X chromosome
217 is enriched for non-coding RNAs, yet their transcriptional modes and roles are unclear. Due to their
218 unique ability to recruit factors and target a genomic address¹, lncRNAs play critical roles in genome
219 regulation and health. The study of lncRNAs escaping XCI may reveal novel mechanisms of inter-
220 female phenotypic variation and sexual dimorphism.

221 We identified genes that constitutively escaped XCI in all tissues, and genes with tissue-
222 specific escape. Co-occurrence of both patterns suggests involvement of tissue-shared and tissue-
223 specific determinants. Presumably, genetic variants (e.g. eQTLs) with constitutive or tissue-specific
224 effects, modulate the escape dosages. We found genes with consistent and with heterogeneous
225 EscScore across individuals. We reasoned that genes with key physiological roles may be subject to
226 shared regulation across females. Examples are *BTK* and *CD99L2*, which, in line with their roles in
227 lymphoid cells, exhibited consistent EscScore in LCLs, but inter-female variability in adipose and skin.
228 Thus, the escape behaviour could depend on the gene's functional context. The tumorigenesis-
229 related genes *DDX3X* and *KDM6A* both escaped XCI, in line with previous studies^{11,25,26}, and showed
230 inter-female variability in all tissues. These genes may underlie sexual dimorphism in cancer¹⁴, and
231 plausibly their variable escape contributes to inter-female diversity in cancer risk. We found that
232 the histone demethylase *KDM5C* and the transcription factor *ZFX*, escaped in LCLs and solid tissues
233 in the same individual. Other genes manifested more composite behaviour, with escape restricted
234 to either LCLs or solid tissues in an individual, supporting tissue-specificity of escape. Altogether,
235 these patterns highlight the complexity of escape, and anticipate its roles as phenotype modulator.

236 The X chromosome plays key roles in innate and adaptive immunity^{12,27,36,40}. We found that
237 escape differed between immune cell types, with higher incidence in lymphocytes than monocytes.
238 Among 53 genes with data available in all immune cell types, 7% (*PRKX*, *DIAPH2*, *JPX*, *ZFX*) escaped
239 XCI in all cells, in line with their constitutive escape behavior across tissues. For *ZFX*, this aligns with
240 data showing its involvement in networks for X-linked dosage regulation⁴¹. 17% of genes (*TCEANC*,

241 *TAB3, MIR222HG, ARMCX4, AC234775.3, NTX, SLC25A43, INTS6L, GAB3*) exhibited escape restricted
242 to a single immune cell type which was always of lymphoid lineage. We also found significant
243 variation between lymphoid cells. Altogether, these data indicate an immune cell type-specific
244 propensity to escape. Myeloid and lymphoid lineages are subject to distinct regulation during
245 development. Possibly, cell type-specific factors, such as genetic variants and/or epigenetic marks
246 inherited during differentiation, establish distinct escape dosages across cells within an individual¹¹.
247 These data would have multiple biological significance. Firstly, different cell types and possibly single
248 cells, would provide a different contribution to the overall escape dosages in a tissue. This
249 phenomenon could establish the bases for an X-linked transcriptional mosaicism throughout the
250 female's immune system. Secondly, changes in cell type composition or proportion, which may
251 characterize pathological states^{42,43}, may alter the escape which in turn modifies disease risk. The
252 extent to which this phenomenon modulates inter-female variability in risk and expressivity of
253 immunological traits will require future larger studies. Immune cell type-specific escapees could
254 potentially serve as markers of disease-relevant cells, with applications for diagnostic purposes and
255 design of immunotherapy approaches⁴⁴⁻⁴⁶.

256 The extent to which genetics and environment influence escape in humans is unclear.
257 Concordance in methylation-based XCI status between MZ twins supported a dominant model of
258 cis-acting influences¹⁶. MZ twins share >99% of DNA, age, and multiple environmental traits such as
259 in-utero growth and early life. Variable escape may affect MZ twins differently, leading to different
260 trait expressivity. We found significantly more similar EscScore between MZ than DZ co-twins.
261 Congruently, we found overall higher rates of discordant XCI between DZ than MZ co-twins. These
262 data demonstrate a solid contribution owing to genotype, but also that DNA does not fully explain
263 such concordance patterns. Thus, escape has both heritable and environmental components, in line
264 with current knowledge on complex traits. An interplay between QTLs^{47,48}, differential epigenetic
265 control of parental alleles¹, and gene-environment effects may ultimately modulate the allelic
266 expression of X-genes and propensity to escape. These effects may have cell type- and tissue-
267 specific components, and underlie intra- and inter-female variation across cells and tissues. Our
268 observations on inter-female variability align with population differences in dose compensation^{21,49},
269 supporting the involvement of genetic factors. Cis-acting variation may also model the Xi vs Xa
270 haplotype expression across tissues, explaining intra-individual variation¹¹. The identification and
271 functional characterization of such genetic and environmental factors can aid understanding what
272 drives the inter-female variation in trait expressivity and disease risk, and sexual dimorphism. While
273 this is the first multi-tissue transcriptomic study of escape in twins, we acknowledge the need for
274 further strategies to characterize the determinants of XCI escape in humans.

275 The present study contributes toward a detailed characterization of escape in humans. Given
276 the paradigmatic roles of the X chromosome in epigenetics and clinical genetics, a full understanding
277 of XCI escape has implications on epigenetic research and therapeutics. Therapeutics may include
278 genetic counselling and design of treatments for X-linked conditions. Despite nearly 60 years after
279 Mary Lyon's landmark intuition on escape, a lot is yet to be learned. Future large-scale studies that
280 combine biomedical records and functional assays will be critical to disentangle the breadth of
281 variability of escape from XCI in humans and characterize its phenotypic impact.

282 **Materials and Methods**

283 **Sample collection**

284 The study included 856 female twins from the TwinsUK registry^{22,50} who participated in the MuTHER
285 study⁴⁷. Study participants included both monozygotic (MZ) and dizygotic (DZ) twins, aged 38-85
286 years old (median age = 60). All subjects are of European ancestry. Peripheral blood samples were
287 collected and lymphoblastoid cell lines (LCLs) generated via Epstein-Barr virus (EBV) mediated
288 transformation of B-lymphocytes. Punch biopsies of subcutaneous adipose tissue were taken from
289 a photo-protected area adjacent and inferior to the umbilicus. Skin samples were obtained by
290 dissection from punch biopsies. Adipose and skin samples were weighed and frozen in liquid
291 nitrogen. This project was approved by the research ethics committee at St Thomas' Hospital (UK).
292 Volunteers received detailed information sheet regarding all aspects of the research, gave informed
293 consent and signed an approved consent form prior to biopsy and to participate in the study.

294 **DNA sequencing data and variant calling**

295 30X whole genome sequencing (WGS) was carried out at Human Longevity, Inc. (HLI). Details on
296 sample and library preparation, clustering and sequencing have been reported elsewhere⁵¹. For the
297 purpose of this project, as all individuals were female, the DNA sequencing reads (pre-mapped to
298 the X chromosome by Illumina's ISIS Analysis Software v.2.5.26.13⁵²) were realigned to the GRCh38
299 X chromosome reference sequence using BWA-MEM in SpeedSeq v0.1.2⁵³. Base quality score
300 recalibration (BQSR) was performed in GATK v4.1.6⁵⁴. Following this, DNA variant calling was
301 performed using the gold-standard workflow in GATK v4.1.6⁵⁴. This included implementation of
302 HaplotypeCaller to call germline variants, GenomicsDBImport to create a unified gVCF repository,
303 and GenotypeGVCFs for joint genotyping to produce a multi-sample variant call set. Variants with a
304 VQSLOD (variant quality score odd-ratio) corresponding to a truth sensitivity of <99.9% and with a
305 HWE (Hardy-Weinberg equilibrium) *P-value* <1e-6 were removed. Data quality checks were further
306 performed with VCFtools⁵⁵ to check levels of transition/transversion ratio. Dataset comprised 621
307 female samples. For individuals with available RNA-seq but unavailable genotypes, chrX DNAseq
308 data were retrieved from the UK10K project⁵⁶.

309 **RNA sequencing data**

310 The Illumina TruSeq sample preparation protocol was used to generate cDNA libraries for
311 sequencing. Libraries were sequenced on a Illumina HiSeq2000 machine and 49 bp paired-end reads
312 were generated⁵⁰. Samples that failed library preparation (according to manufacturer's guidelines)
313 or had less than 10 million reads were discarded. As all individuals were female, for this manuscript
314 RNA-seq reads were aligned to a Y-masked⁵⁷ GRCh38 reference genome using STAR v.2.7.3⁵⁸.
315 Properly paired and uniquely mapped reads with a MAPQ of 255 were retained for further analyses.

316 **Purified immune cell RNA-sequencing data**

317 Monocytes, B, T-CD4⁺, T-CD8⁺ and NK cells were purified using fluorescence activated cell sorting
318 (FACS) from two monozygotic twins exhibiting skewed XCI patterns in LCLs. Gating strategy for cell
319 sorting is described in Fig.S5. Total RNA was isolated and cDNA libraries for sequencing were
320 generated using the Sureselect sample preparation protocol. Samples were then sequenced with
321 the Illumina HiSeq machine and 126 bp paired-end reads were generated. Adapter and polyA/T
322 nucleotide sequences were trimmed using trim_galore v.0.6.3 and PrinSeq tools v.0.20⁵⁹. RNA-seq
323 reads were aligned to Y-masked⁵⁷ GRCh38 reference genome using STAR v.2.7.3⁵⁸. Properly paired
324 and uniquely mapped reads were retained for further analyses.

325 **Correction of RNA-sequencing mapping biases**

326 To eliminate mapping biases in RNA-seq, the WASP pipeline for mappability filtering⁶⁰ in
327 STARv2.7.3⁵⁸ was used. In each read overlapping a heterozygous SNP, the allele is flipped to the
328 SNP's other allele and the read is remapped. Reads that did not remap to the same genomic location
329 are flagged as owing to mapping bias and were discarded.

330 **Haplotype phasing and gene-level haplotype expression**

331 WGS genomes were read-back phased using recent SHAPEIT2 implementation⁶¹ that takes
332 advantage of the phase information present in DNA-seq reads. Subsequently, phASER v.0.9.9.4^{62,63}
333 was used for RNAseq-based read-backed phasing and to generate gene-level haplotype expression
334 data. Only reads uniquely mapped and with a base quality ≥ 10 were used for phasing. Using
335 haplotype expression data, the gene's ASE in each sample can be calculated as follow:

$$336 \quad \text{Gene's Allele-Specific Expression (ASE)} \quad [1]$$
$$337 \quad ASE_{g,s} = \frac{AC_{g,s}}{TC_{g,s}}$$

338 Where, for a biallelic site:

339 A = haplotype A;

340 B = haplotype B;

341 $AC_{g,s}$ = RNAseq allelic count at haplotype A of gene g in sample s ;

342 $BC_{g,s}$ = RNAseq allelic count at haplotype B of gene g in sample s ;

343 $TC_{g,s}$ = Total RNAseq read depth at gene g in sample s ;

344 $TC_{g,s} = AC_{g,s} + BC_{g,s}$.

345
346 The gene ASE values range from 0 to 1, with 0 and 1 indicating monoallelic expression and 0.5
347 indicating completely balanced haplotypic expression. From [1], it follows:

$$348 \quad \frac{AC_{g,s}}{TC_{g,s}} + \frac{BC_{g,s}}{TC_{g,s}} = 1 \quad [2]$$

349
350 For each gene, the effect size of the allelic imbalance in expression was calculated as allelic fold
351 change (aFC), that is the ratio between the allele with the lower RNAseq count and the allele with
352 the higher RNAseq count, as similarly used in a previous study⁶⁴:

$$353 \quad \text{aFC of gene } g \text{ in sample } s \quad [3]$$
$$354 \quad aFC_{g,s} = \frac{\text{Min}(AC_{g,s}, BC_{g,s})}{\text{Max}(AC_{g,s}, BC_{g,s})}$$

355
356 aFC values range from 0 to 1, with 0 indicating monoallelic expression and 1 completely balanced
357 haplotypic expression. For the purpose of our study, a gene's aFC can be interpreted as X_i/X_a
358 expression ratio. The X_i is assumed to be the allele with the lower RNA-seq count, while the X_a the
359 allele with higher RNA-seq count.
360

361 **Quantification of XCI skewing levels**

362 In each sample, the *XIST* allele-specific expression ($XIST_{ASE}$) was used as proxy for XCI skewing levels.
363 *XIST* is uniquely expressed from the X_i , and thus the relative expression of parental alleles within
364 *XIST* transcript is representative of XCI skewing in a bulk sample^{7,23,24,65,66}. Within each sample, the
365 $XIST_{ASE}$ values range from 0 to 1, with 0 or 1 indicating complete inactivation of one parental
366 chromosome (completely skewed XCI, 100:0 XCI ratio), and 0.5 indicating balanced inactivation ratio
367 of the two parental X chromosomes (completely random XCI). To be consistent with previous

368 literature^{23,67,68}, we classified samples with $XIST_{ASE} \leq 0.2$ or $XIST_{ASE} \geq 0.8$ to have skewed XCI, and
369 samples with $0.2 < XIST_{ASE} < 0.8$ to have random XCI. To have an absolute measure of the magnitude
370 of XCI skewing levels in each sample, the degree of XCI skewing (DS) was calculated from the $XIST_{ASE}$
371 calls. DS is defined as the absolute deviation of $XIST_{ASE}$ from 0.5, and it has been similarly been used
372 to assess XCI patterns and XCI status of X-genes^{21,69,70}. In each sample, DS was calculated as follow:

$$373 \quad \quad \quad DS_s = |0.5 - XIST_{ASE}| \quad \quad \text{Degree of XCI skewing in sample } s \quad \quad [4]$$

374
375
376 DS is a proxy for the magnitude of the sample's XCI-skew. DS values range from 0 to 0.5; 0 indicates
377 random XCI and 0.5 completely skewed XCI. Samples with $DS \geq 0.3$ were classified to have skewed
378 XCI; samples with $DS < 0.3$ to have random XCI patterns.

379 **Quantification of XCI escape**

380 Bulk samples with random XCI patterns confound mono- and bi-allelic X-linked expression as both
381 X-alleles would be, on overall, expressed. Conversely, in skewed samples silenced genes will exhibit
382 monoallelic expression while escape genes biallelic expression^{8,11,21} (Note S1). Only genes with a
383 RNAseq read depth ≥ 8 reads were used. Furthermore, to increase the confidence that genotypes
384 were truly heterozygous, we considered only genes whose both haplotypes were detected in the
385 RNA-seq data at least once. We reasoned that variations in DS might influence allelic variation in X-
386 linked expression leading to biases in escape measurements across samples. To account for this, we
387 implemented a linear model of the sample's DS as explanatory variable and the gene's aFC (as
388 defined above) as response variable using our entire skewed cohort of 166 LCLs, 26 whole-blood,
389 57 adipose and 64 skin samples. We then derived the aFC's residuals (referred to as 'EscScore'),
390 which represent the gene's aFC adjusted for DS. The EscScore values were subsequently rescaled to
391 the [0,1] range via min-max normalization. Due to low number of skewed whole-blood samples with
392 available data in other tissues, we excluded our whole-blood estimates from further analyses.
393 EscScore 0 and 1 indicate complete monoallelic (silencing) and complete biallelic expression (full
394 escape), respectively. We detected average EscScore of 0.32, 0.36, and 0.37 across LCLs, adipose
395 and skin samples, respectively, whose median is 0.36. These data are consistent with the GTEx XCI
396 survey¹¹. Based on these observations, we classified genes exhibiting a median (across ≥ 3 tissue
397 samples) $EscScore \geq 0.36$ as escapees in that tissue, while all others as silenced.

398 **Protein-protein interaction (PPI) network and gene ontology analyses**

399 Genes with EscScore data available in all three tissues and escaping XCI in at least one tissue were
400 analysed for protein-protein interaction (PPI) with a new genome-wide protein interactome
401 database as reference (www.interactome-atlas.org)²⁸. PPI network was imported to STRING v.11⁷¹,
402 and proteins with at least one direct interaction with our genes and a PPI score (edge confidence)
403 ≥ 0.4 were selected. PPI network was imported to Cytoscape v.3.8.2⁷² for visualization and gene
404 ontology analyses of Biological Processes and REACTOME Pathways using ClueGO v.2.5.8⁷³. A term
405 was considered as significantly enriched when its Bonferroni-corrected *P-value* was ≤ 0.01 .

406 **REFERENCES**

- 407
- 408 1 Lee, J. T. & Bartolomei, M. S. X-inactivation, imprinting, and long noncoding RNAs in health and
409 disease. *Cell* **152**, 1308-1323, doi:10.1016/j.cell.2013.02.016 (2013).
- 410 2 Lyon, M. F. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372-
411 373, doi:10.1038/190372a0 (1961).
- 412 3 Galupa, R. & Heard, E. X-Chromosome Inactivation: A Crossroads Between Chromosome
413 Architecture and Gene Regulation. *Annu Rev Genet* **52**, 535-566, doi:10.1146/annurev-genet-
414 120116-024611 (2018).
- 415 4 Simon, M. D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-
416 chromosome inactivation. *Nature* **504**, 465-469, doi:10.1038/nature12719 (2013).
- 417 5 Barr, M. L. & Bertram, E. G. A morphological distinction between neurones of the male and female,
418 and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature* **163**,
419 676 (1949).
- 420 6 Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread
421 across the X chromosome. *Science* **341**, 1237973, doi:10.1126/science.1237973 (2013).
- 422 7 Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed
423 exclusively from the inactive X chromosome. *Nature* **349**, 38-44, doi:10.1038/349038a0 (1991).
- 424 8 Carrel, L. & Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene
425 expression in females. *Nature* **434**, 400-404, doi:10.1038/nature03479 (2005).
- 426 9 Carrel, L. & Brown, C. J. When the Lyon(ized chromosome) roars: ongoing expression from an
427 inactive X chromosome. *Philos Trans R Soc Lond B Biol Sci* **372**, doi:10.1098/rstb.2016.0355
428 (2017).
- 429 10 Lyon, M. F. Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet* **14**,
430 135-148 (1962).
- 431 11 Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**,
432 244-248, doi:10.1038/nature24265 (2017).
- 433 12 Libert, C., Dejager, L. & Pinheiro, I. The X chromosome in immune functions: when a chromosome
434 makes the difference. *Nat Rev Immunol* **10**, 594-604, doi:10.1038/nri2815 (2010).
- 435 13 Neri, G., Schwartz, C. E., Lubs, H. A. & Stevenson, R. E. X-linked intellectual disability update
436 2017. *Am J Med Genet A* **176**, 1375-1388, doi:10.1002/ajmg.a.38710 (2018).
- 437 14 Dunford, A. *et al.* Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex
438 bias. *Nat Genet* **49**, 10-16, doi:10.1038/ng.3726 (2017).
- 439 15 Clement-Jones, M. *et al.* The short stature homeobox gene SHOX is involved in skeletal
440 abnormalities in Turner syndrome. *Hum Mol Genet* **9**, 695-702 (2000).
- 441 16 Cotton, A. M. *et al.* Landscape of DNA methylation on the X chromosome reflects CpG density,
442 functional chromatin state and X-chromosome inactivation. *Hum Mol Genet* **24**, 1528-1539,
443 doi:10.1093/hmg/ddu564 (2015).
- 444 17 Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* **369**,
445 doi:10.1126/science.aba3066 (2020).

- 446 18 Garieri, M. *et al.* Extensive cellular heterogeneity of X inactivation revealed by single-cell allele-
447 specific expression in human fibroblasts. *Proc Natl Acad Sci U S A* **115**, 13015-13020,
448 doi:10.1073/pnas.1806811115 (2018).
- 449 19 Wainer Katsir, K. & Linial, M. Human genes escaping X-inactivation revealed by single cell
450 expression data. *BMC Genomics* **20**, 201, doi:10.1186/s12864-019-5507-6 (2019).
- 451 20 Shvetsova, E. *et al.* Skewed X-inactivation is common in the general female population. *Eur J Hum*
452 *Genet* **27**, 455-465, doi:10.1038/s41431-018-0291-3 (2019).
- 453 21 Cotton, A. M. *et al.* Analysis of expressed SNPs identifies variable extents of expression from the
454 human inactive X chromosome. *Genome Biol* **14**, R122, doi:10.1186/gb-2013-14-11-r122 (2013).
- 455 22 Moayyeri, A., Hammond, C. J., Hart, D. J. & Spector, T. D. The UK Adult Twin Registry (TwinsUK
456 Resource). *Twin Res Hum Genet* **16**, 144-149, doi:10.1017/thg.2012.89 (2013).
- 457 23 Zito, A. *et al.* Heritability of skewed X-inactivation in female twins is tissue-specific and associated
458 with age. *Nat Commun* **10**, 5339, doi:10.1038/s41467-019-13340-w (2019).
- 459 24 Rupert, J. L., Brown, C. J. & Willard, H. F. Direct detection of non-random X chromosome
460 inactivation by use of a transcribed polymorphism in the XIST gene. *Eur J Hum Genet* **3**, 333-343
461 (1995).
- 462 25 Balaton, B. P., Cotton, A. M. & Brown, C. J. Derivation of consensus inactivation status for X-
463 linked genes from genome-wide studies. *Biol Sex Differ* **6**, 35, doi:10.1186/s13293-015-0053-7
464 (2015).
- 465 26 Zhang, Y. *et al.* Genes that escape X-inactivation in humans have high intraspecific variability in
466 expression, are associated with mental impairment but are not slow evolving. *Mol Biol Evol* **30**,
467 2588-2601, doi:10.1093/molbev/mst148 (2013).
- 468 27 Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325-337,
469 doi:10.1038/nature03440 (2005).
- 470 28 Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402-408,
471 doi:10.1038/s41586-020-2188-x (2020).
- 472 29 Vallianatos, C. N. *et al.* Altered Gene-Regulatory Function of KDM5C by a Novel Mutation
473 Associated With Autism and Intellectual Disability. *Front Mol Neurosci* **11**, 104,
474 doi:10.3389/fnmol.2018.00104 (2018).
- 475 30 Adegbola, A., Gao, H., Sommer, S. & Browning, M. A novel mutation in JARID1C/SMCX in a
476 patient with autism spectrum disorder (ASD). *Am J Med Genet A* **146A**, 505-511,
477 doi:10.1002/ajmg.a.32142 (2008).
- 478 31 Van der Meulen, J. *et al.* The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in
479 T-cell acute lymphoblastic leukemia. *Blood* **125**, 13-21, doi:10.1182/blood-2014-05-577270 (2015).
- 480 32 Hagerman, R. J. *et al.* Fragile X syndrome. *Nat Rev Dis Primers* **3**, 17065, doi:10.1038/nrdp.2017.65
481 (2017).
- 482 33 Duan, D., Goemans, N., Takeda, S., Mercuri, E. & Aartsma-Rus, A. Duchenne muscular dystrophy.
483 *Nat Rev Dis Primers* **7**, 13, doi:10.1038/s41572-021-00248-3 (2021).
- 484 34 Scofield, R. H. *et al.* Klinefelter's syndrome (47,XXY) in male systemic lupus erythematosus
485 patients: support for the notion of a gene-dose effect from the X chromosome. *Arthritis Rheum* **58**,
486 2511-2517, doi:10.1002/art.23701 (2008).

- 487 35 Seminog, O. O., Seminog, A. B., Yeates, D. & Goldacre, M. J. Associations between Klinefelter's
488 syndrome and autoimmune diseases: English national record linkage studies. *Autoimmunity* **48**, 125-
489 128, doi:10.3109/08916934.2014.968918 (2015).
- 490 36 Odhams, C. A. *et al.* Interferon inducible X-linked gene CXorf21 may contribute to sexual
491 dimorphism in Systemic Lupus Erythematosus. *Nat Commun* **10**, 2164, doi:10.1038/s41467-019-
492 10106-2 (2019).
- 493 37 Wang, J. *et al.* Unusual maintenance of X chromosome inactivation predisposes female lymphocytes
494 for increased expression from the inactive X. *Proc Natl Acad Sci U S A* **113**, E2029-2038,
495 doi:10.1073/pnas.1520113113 (2016).
- 496 38 Fieremans, N. *et al.* Identification of Intellectual Disability Genes in Female Patients with a Skewed
497 X-Inactivation Pattern. *Hum Mutat* **37**, 804-811, doi:10.1002/humu.23012 (2016).
- 498 39 Brenes, A. J. *et al.* Erosion of human X chromosome inactivation causes major remodeling of the
499 iPSC proteome. *Cell Rep* **35**, 109032, doi:10.1016/j.celrep.2021.109032 (2021).
- 500 40 Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat Rev Immunol* **16**, 626-638,
501 doi:10.1038/nri.2016.90 (2016).
- 502 41 Zhang, X. *et al.* Integrated functional genomic analyses of Klinefelter and Turner syndromes reveal
503 global network effects of altered X chromosome dosage. *Proc Natl Acad Sci U S A* **117**, 4864-4873,
504 doi:10.1073/pnas.1910003117 (2020).
- 505 42 Lee, D. S. W., Rojas, O. L. & Gommerman, J. L. B cell depletion therapies in autoimmune disease:
506 advances and mechanistic insights. *Nat Rev Drug Discov* **20**, 179-199, doi:10.1038/s41573-020-
507 00092-2 (2021).
- 508 43 Pascual, V., Chaussabel, D. & Banchereau, J. A genomic approach to human autoimmune diseases.
509 *Annu Rev Immunol* **28**, 535-571, doi:10.1146/annurev-immunol-030409-101221 (2010).
- 510 44 Wang, S., Cowley, L. A. & Liu, X. S. Sex Differences in Cancer Immunotherapy Efficacy,
511 Biomarkers, and Therapeutic Strategy. *Molecules* **24**, doi:10.3390/molecules24183214 (2019).
- 512 45 Liu, C. *et al.* Involvement of X-chromosome Reactivation in Augmenting Cancer Testis Antigens
513 Expression: A Hypothesis. *Curr Med Sci* **38**, 19-25, doi:10.1007/s11596-018-1842-0 (2018).
- 514 46 Emran, A. A. *et al.* Study of the Female Sex Survival Advantage in Melanoma-A Focus on X-
515 Linked Epigenetic Regulators and Immune Responses in Two Cohorts. *Cancers (Basel)* **12**,
516 doi:10.3390/cancers12082082 (2020).
- 517 47 Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat*
518 *Genet* **44**, 1084-1089, doi:10.1038/ng.2394 (2012).
- 519 48 Consortium, G. T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-
520 213, doi:10.1038/nature24277 (2017).
- 521 49 Johnston, C. M. *et al.* Large-scale population study of human cell lines indicates that dosage
522 compensation is virtually complete. *PLoS Genet* **4**, e9, doi:10.1371/journal.pgen.0040009 (2008).
- 523 50 Buil, A. *et al.* Gene-gene and gene-environment interactions detected by transcriptome sequence
524 analysis in twins. *Nat Genet* **47**, 88-91, doi:10.1038/ng.3162 (2015).
- 525 51 Long, T. *et al.* Whole-genome sequencing identifies common-to-rare variants associated with human
526 blood metabolites. *Nat Genet* **49**, 568-578, doi:10.1038/ng.3809 (2017).

- 527 52 Raczy, C. *et al.* Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing
528 platforms. *Bioinformatics* **29**, 2041-2043, doi:10.1093/bioinformatics/btt314 (2013).
- 529 53 Chiang, C. *et al.* SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat Methods* **12**,
530 966-968, doi:10.1038/nmeth.3505 (2015).
- 531 54 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation
532 DNA sequencing data. *Nat Genet* **43**, 491-498, doi:10.1038/ng.806 (2011).
- 533 55 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158,
534 doi:10.1093/bioinformatics/btr330 (2011).
- 535 56 Consortium, U. K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature*
536 **526**, 82-90, doi:10.1038/nature14962 (2015).
- 537 57 Olney, K. C., Brotman, S. M., Andrews, J. P., Valverde-Vesling, V. A. & Wilson, M. A. Reference
538 genome and transcriptome informed by the sex chromosome complement of the sample increase
539 ability to detect sex differences in gene expression from RNA-Seq data. *Biol Sex Differ* **11**, 42,
540 doi:10.1186/s13293-020-00312-9 (2020).
- 541 58 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21,
542 doi:10.1093/bioinformatics/bts635 (2013).
- 543 59 Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets.
544 *Bioinformatics* **27**, 863-864, doi:10.1093/bioinformatics/btr026 (2011).
- 545 60 van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific software for
546 robust molecular quantitative trait locus discovery. *Nat Methods* **12**, 1061-1063,
547 doi:10.1038/nmeth.3582 (2015).
- 548 61 Delaneau, O., Howie, B., Cox, A. J., Zagury, J. F. & Marchini, J. Haplotype estimation using
549 sequencing reads. *Am J Hum Genet* **93**, 687-696, doi:10.1016/j.ajhg.2013.09.002 (2013).
- 550 62 Castel, S. E., Mohammadi, P., Chung, W. K., Shen, Y. & Lappalainen, T. Rare variant phasing and
551 haplotypic expression from RNA sequencing with phASER. *Nat Commun* **7**, 12817,
552 doi:10.1038/ncomms12817 (2016).
- 553 63 Castel, S. E. *et al.* A vast resource of allelic expression data spanning human tissues. *Genome Biol*
554 **21**, 234, doi:10.1186/s13059-020-02122-z (2020).
- 555 64 Mohammadi, P., Castel, S. E., Brown, A. A. & Lappalainen, T. Quantifying the regulatory effect
556 size of cis-acting genetic variation using allelic fold change. *Genome Res* **27**, 1872-1884,
557 doi:10.1101/gr.216747.116 (2017).
- 558 65 Amos-Landgraf, J. M. *et al.* X chromosome-inactivation patterns of 1,005 phenotypically unaffected
559 females. *Am J Hum Genet* **79**, 493-499, doi:10.1086/507565 (2006).
- 560 66 Brown, C. J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains
561 conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527-542 (1992).
- 562 67 Naumova, A. K. *et al.* Heritability of X chromosome--inactivation phenotype in a large family. *Am J*
563 *Hum Genet* **58**, 1111-1119 (1996).
- 564 68 Kristiansen, M. *et al.* Twin study of genetic and aging effects on X chromosome inactivation. *Eur J*
565 *Hum Genet* **13**, 599-606, doi:10.1038/sj.ejhg.5201398 (2005).
- 566 69 Wong, C. C. *et al.* A longitudinal twin study of skewed X chromosome-inactivation. *PLoS One* **6**,
567 e17873, doi:10.1371/journal.pone.0017873 (2011).

- 568 70 Gentilini, D. *et al.* Age-dependent skewing of X chromosome inactivation appears delayed in
569 centenarians' offspring. Is there a role for allelic imbalance in healthy aging and longevity? *Aging*
570 *Cell* **11**, 277-283, doi:10.1111/j.1474-9726.2012.00790.x (2012).
- 571 71 Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage,
572 supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-
573 D613, doi:10.1093/nar/gky1131 (2019).
- 574 72 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular
575 interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
- 576 73 Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and
577 pathway annotation networks. *Bioinformatics* **25**, 1091-1093, doi:10.1093/bioinformatics/btp101
578 (2009).

579 **Acknowledgements**

580 This study was supported by MRC Project Grant (MR/R023131/1) to K.S.S. The TwinsUK study was
581 funded by the Wellcome Trust and European Community's Seventh Framework Programme
582 (FP7/2007-2013). The TwinsUK study also receives support from the National Institute for Health
583 Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre
584 based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. This
585 project was enabled through access to the MRC eMedLab Medical Bioinformatics infrastructure,
586 supported by the Medical Research Council [grant number MR/L016311/1]. The authors
587 acknowledge use of the research computing facility at King's College London, *Rosalind*
588 (<https://rosalind.kcl.ac.uk>), which is delivered in partnership with the National Institute for Health
589 Research (NIHR) Biomedical Research Centres at South London & Maudsley and Guy's & St. Thomas'
590 NHS Foundation Trusts, and part-funded by capital equipment grants from the Maudsley Charity
591 (award 980) and Guy's & St. Thomas' Charity (TR130505).

592 **Authors contributions**

593 A.Z. and K.S.S conceived and designed the project. A.Z. designed and performed bioinformatic
594 analyses. A.L.R. and J.S.E.M contributed data interpretation and discussion. A.V., N.R. and M.F.
595 contributed DNaseq data processing. S.N. performed network and gene ontology analyses and
596 contributed data graphics. R.A.E. performed RNA isolation and FACS experiments. A.Z. and K.S.S.
597 wrote the manuscript. All authors read and approved the manuscript.

598 **Competing interests**

599 All authors have nothing to declare.

600 **Materials and Correspondence**

601 Correspondence should be addressed to A.Z (antonino.zito@kcl.ac.uk) or K.S.S (kerrin.small@kcl.ac.uk).

602 **Data Availability**

603 TwinsUK RNAseq data are available from EGA (Accession number: EGAS00001000805). TwinsUK
604 genotypes and phenotypes are available upon application to TwinsUK Data Access Committee
605 (<https://twinsuk.ac.uk/resources-for-researchers/access-our-data/>). All other data are contained in
606 the manuscript and its supplementary information.

FIGURES

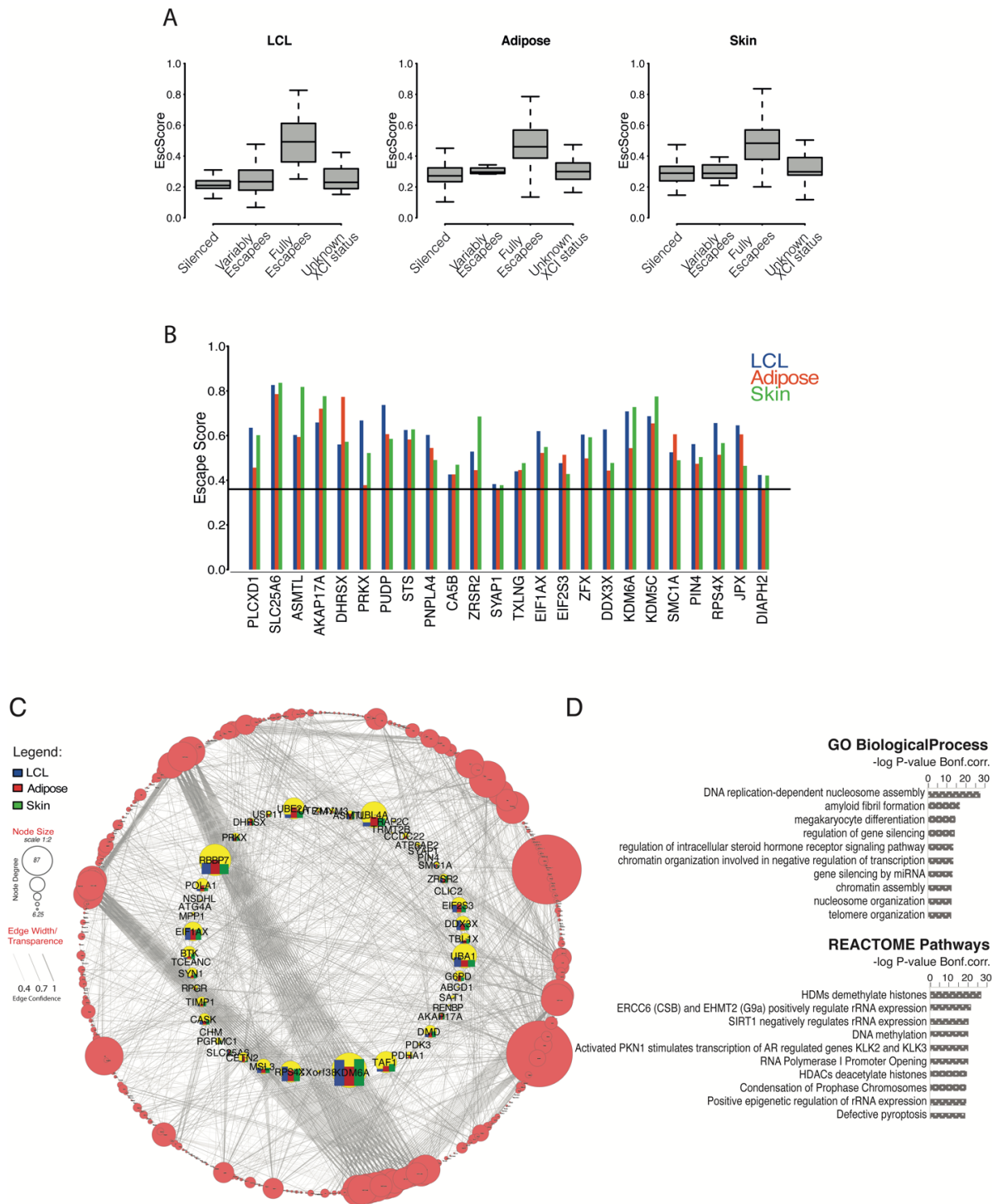


FIG.1 (A) EscScore distinguishes between silenced and escape genes. Boxplots show the distribution of median EscScore values (across skewed samples) of genes with different previously annotated XCI status²⁵. (B) Genes exhibiting constitutive XCI escape in all three tissues. Barplots show the tissue-specific gene's EscScore. Horizontal line denotes the EscScore cutoff to call XCI escape. (C-D) PPI, Gene Ontology (GO) and Pathway analysis. (C) PPI network between protein-coding X-genes escaping XCI in at least one tissue and a reference map of human proteome interactome²⁸. The network inner circle shows protein-coding escape genes characterized within the proteome interactome. The network outer circle shows proteins from the reference proteome²⁸ that have (i) at least 1 interaction with our escapee genes and (ii) edge confidence score ≥ 0.4 . The size of each node reflects the number of interactions (edges). (D) GO and REACTOME Pathways analysis of PPI network. The top 10 Biological Processes and REACTOME terms are shown.

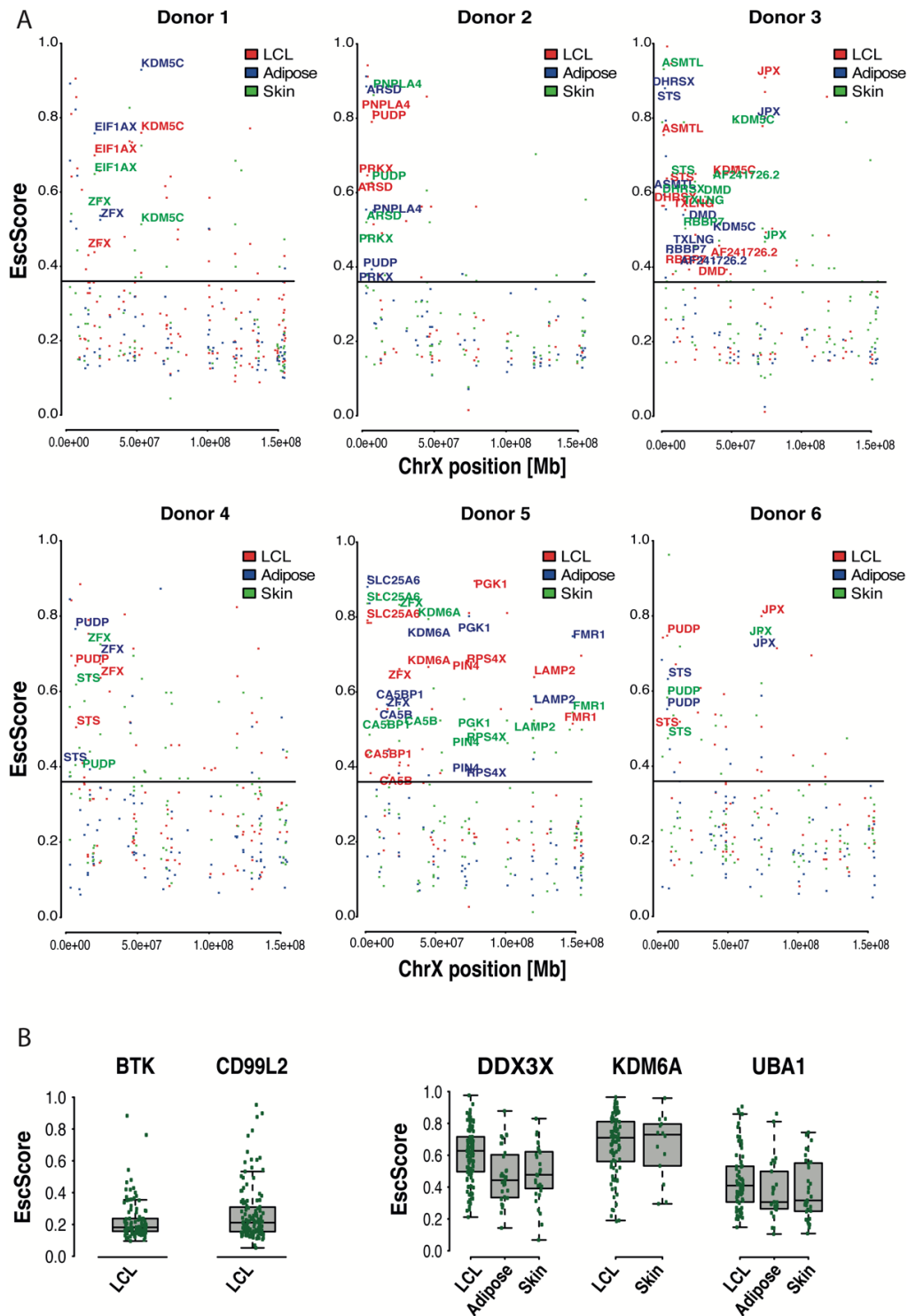
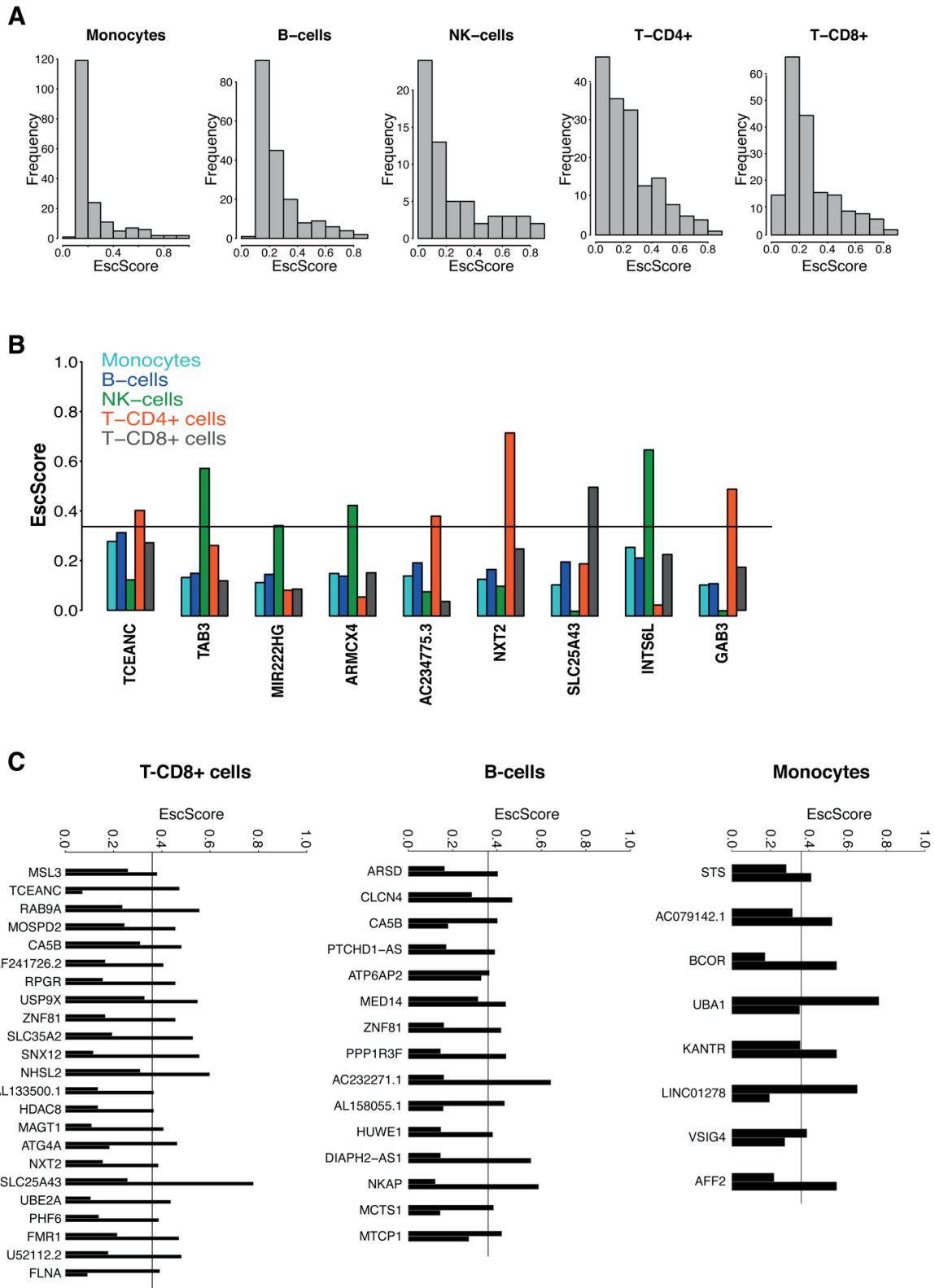


FIG.2 (A) Analysis of intra-individual EscScore variation within each of the 6 female donors exhibiting skewed XCI in all tissues. Each dot shows the gene's EscScore in each tissue. Genes exhibiting escape (EscScore ≥ 0.36 in a tissue in a donor) in all 3 tissues within a donor are highlighted in red (LCLs), blue (adipose), green (skin). (B) Representative subset of genes to assess inter-individual variation in EscScore. Each green dot shows the gene's EscScore in a tissue sample. *BTK* and *CD99L2* show consistent EscScore in LCLs across individuals. *DDX3X*, *KDM6A* and *UBA1* show inter-individual variability of EscScore in LCLs and a solid tissue. Similar plots for other genes are shown in Fig.S3 and Fig.S4.



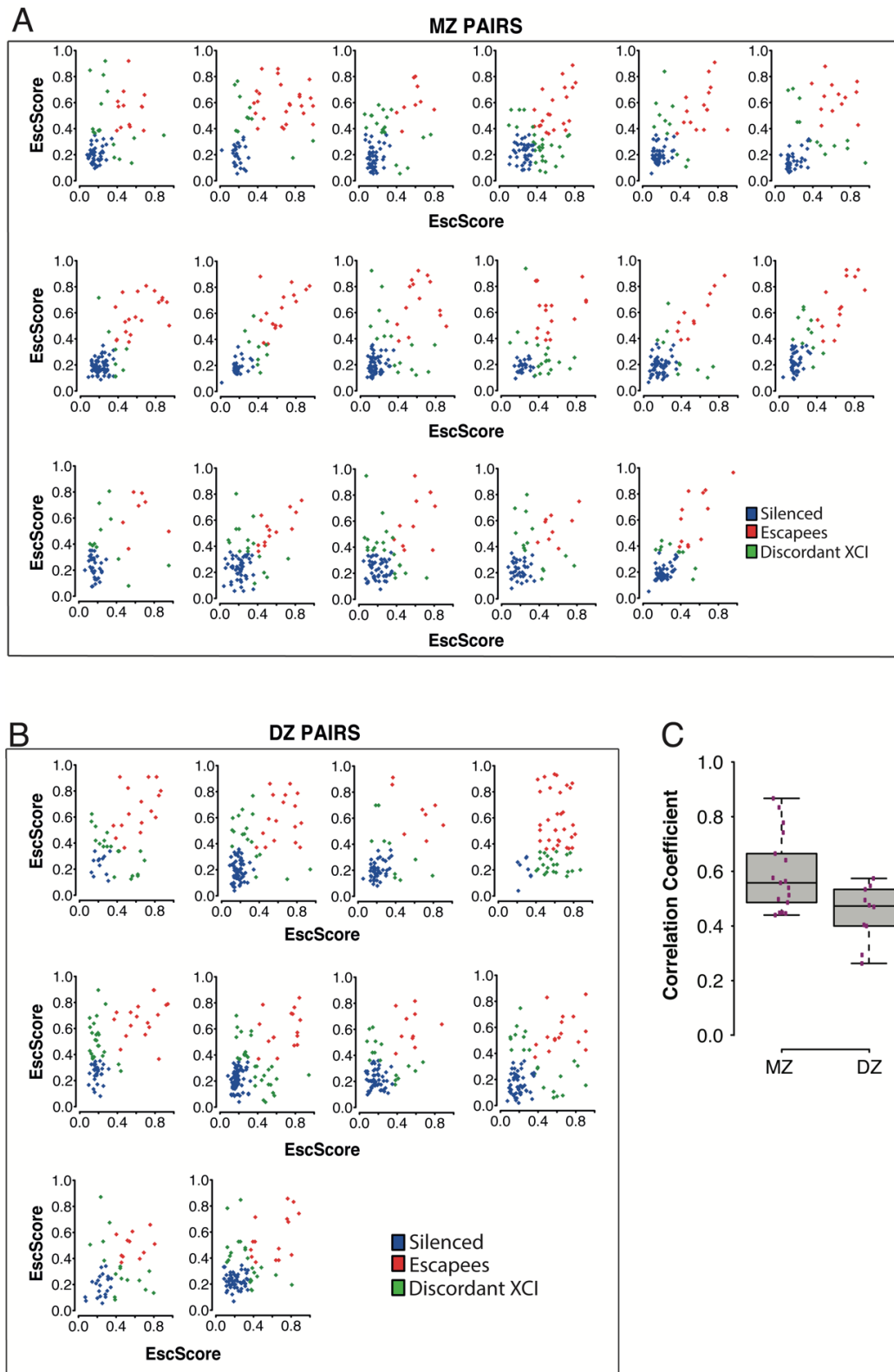


Fig.4 Scatterplots of EscScore(s) of genes (≥ 5) with data available in both co-twins of a pair. A total of 27 complete twin pairs (both co-twins exhibiting skewed XCI in LCLs) were used. Each dot represents a gene, colored in blue if silenced in both co-twins, red if escaping XCI in both co-twins, and green if exhibiting discordant XCI (escaping only in one co-twin). (A) Monozygotic (MZ) twin pairs (N=17); (B) Dizygotic (DZ) twin pairs (N=10). (C) Boxplots of coefficients of correlation between EscScore(s) in the two co-twins of each pair.

Escape from X-inactivation in twins exhibits intra- and inter-individual variability across tissues and is heritable

Zito A.^{1,2}, Roberts A.L.^{1*}, Visconti A.^{1*}, Rossi N.¹, Andres-Ejarque R.³, Nardone S.⁴, Moustafa J.E.S.¹, Falchi M.¹, Small K.S.¹.

¹Department of Twin Research and Genetic Epidemiology, King's College London, London, UK;

²Present address: Department of Molecular Biology, Massachusetts General Hospital, Boston, USA; Department of Genetics, Harvard Medical School, Boston, USA;

³St John's Institute of Dermatology, Faculty of Life Science & Medicine, King's College London, London, UK;

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

(*) equal contribution

SUPPLEMENTAL DATA

Note S1

XIST plays essential roles in XCI¹⁻³. *XIST* spreads in *cis* from the Xq, triggering an epigenetic silencing of the X designated for inactivation (Xi)^{2,4,5}. *XIST* RNA is exclusively expressed from the Xi^{3,6,7}, thus the relative expression of parental *XIST* haplotypes can be used to determine the sample's XCI-skew^{8,9}. As XCI occurs at random within each cell, a somatic tissue with random XCI patterns is a mosaic of cells with either parental X silenced. In such a scenario, all X-genes would exhibit biallelic expression, confounding silenced and escape genes¹⁰. Conversely, in bulk samples with skewed XCI, most expression will be restricted to one haplotype, enabling distinguishment of monoallelic (XCI-silenced genes) from biallelic (escape genes) expression^{7,10,11}.

Note S2

We assessed differences in the incidence of escape between short (Xp) vs long X-arm (Xq). In line with literature^{7,12}, we found a higher prevalence of escape on Xp (Fig. S1). This phenomenon has biological explanations as (i) Xp-genes have had Y-paralogs; (ii) the centromere might be a 'physical barrier' against the spreading in *cis* of *XIST* from Xq to Xp.

SUPPLEMENTARY TABLES

Table S1: Summary statistics (mean, median, standard deviation) of EscScore values of genes with different annotated XCI status according to the Balaton's list¹³. A) LCLs; B) adipose; C) skin. Each statistics is computed across ≥ 3 tissue samples.

A				B				C			
XCI status	Mean	Median	SD	XCI status	Mean	Median	SD	XCI status	Mean	Median	SD
Silenced	0.26	0.21	0.16	Silenced	0.33	0.27	0.2	Silenced	0.34	0.29	0.21
Variable Escapees	0.33	0.29	0.18	Variable Escapees	0.33	0.3	0.17	Variable Escapees	0.32	0.28	0.17
Escapees	0.49	0.48	0.21	Escapees	0.5	0.49	0.22	Escapees	0.48	0.47	0.21
Unknown status	0.33	0.27	0.2	Unknown status	0.34	0.28	0.21	Unknown status	0.37	0.33	0.21

Table S2: Statistical comparison between the EscScore(s) of different gene categories previously annotated¹³. Escapees refer to genes annotated as fully or mostly escaping XCI. For each interrogated gene, the median EscScore value across ≥ 3 tissue samples was used for comparison. Tables report the p-value of the Wilcoxon test between two gene categories A) LCLs; B) adipose; C) Skin.

A				B				C			
XCI status	Silenced	Variably Escapees	Escapees	XCI status	Silenced	Variably Escapees	Escapees	XCI status	Silenced	Variably Escapees	Escapees
Silenced	*	0.1	9e-24	Silenced	*	0.3	4e-14	Silenced	*	0.7	1e-14
Variably Escapees	0.1	*	9e-8	Variably Escapees	0.3	*	1e-4	Variably Escapees	0.7	*	2e-4
Escapees	9e-24	9e-8	*	Escapees	4e-14	1e-4	*	Escapees	1e-14	2e-4	*

Tables S3, S4, S5, S6, S8 are described below and provided as separate files:

Table_S3: X-linked genes (N=159) exhibiting escape in at least one of the three studied tissues (LCLs, adipose, skin) in our dataset. The table lists the gene's EscScore in each tissue, computed as the median EscScore across ≥ 3 samples.

Table_S4: X-linked genes (N=215) with EscScore available in all three studied tissues (LCLs, adipose, skin). The table lists the gene's EscScore in each tissue, computed as the median EscScore across ≥ 3 samples.

Table_S5: X-linked genes (N=51) escaping XCI in only one of the three studied tissues (LCLs, adipose, skin). This is a subset of Table_S4. Column 2 indicates whether the escape status of the gene was either known (previously reported) or is a novel call.

Table_S6: Results from ClueGO¹⁴ enrichment analysis of genes in the PPI network.

Table_S7 Fraction of X-linked genes exhibiting escape (EscScore ≥ 0.36 in a tissue in a donor) in 1, 2 or all 3 studied tissues in each of the 6 female donors exhibiting skewed XCI in all three studied tissues (LCLs, adipose, skin).

DONOR	% escape in 1 tissue	% escape in 2 tissues	% escape in 3 tissues
Donor 1	17.8%	5.6%	1.6%
Donor 2	17.4%	3.5%	3.5%
Donor 3	26.6%	7.8%	7%
Donor 4	35%	6%	2%
Donor 5	27.5%	3.8%	7.6%
Donor 6	30%	2.6%	2.6%

Table_S8: X-linked genes (N=53) with EscScore available in all five studied immune cell types (Monocytes, B-cells, T-CD4⁺ cells, T-CD8⁺ cells, NK-cells).

SUPPLEMENTARY FIGURES

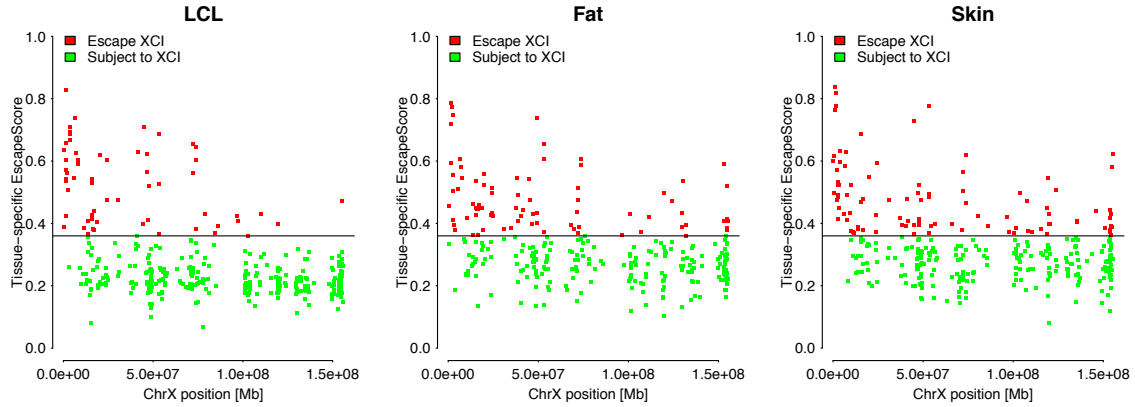


Fig.S1: Relationship between the gene's tissue-specific EscScore and gene position on chrX (GRCh38). Each dot represents a gene. Red and green dots are escapee and silenced genes, respectively.

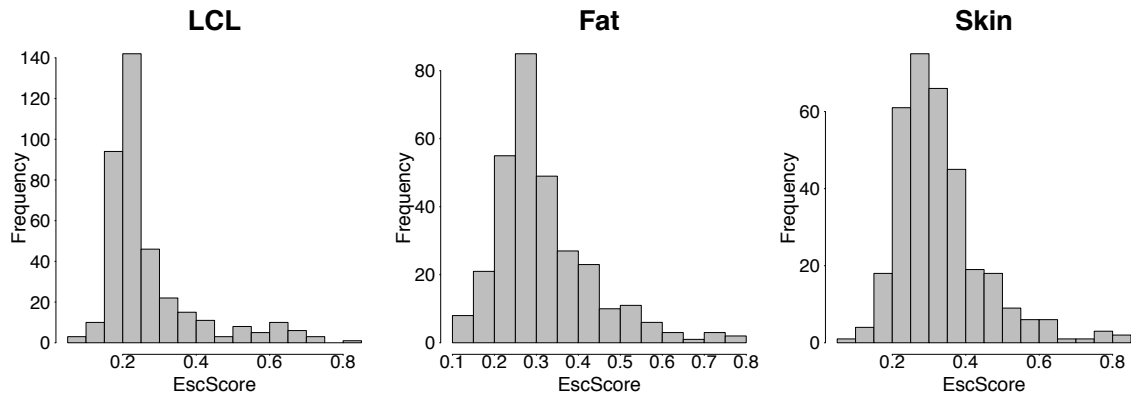


Fig.S2: Distribution of median EscScore values in each of the three studied tissues (LCLs, adipose, skin). Median values were calculated per gene across ≥ 3 skewed tissue samples.

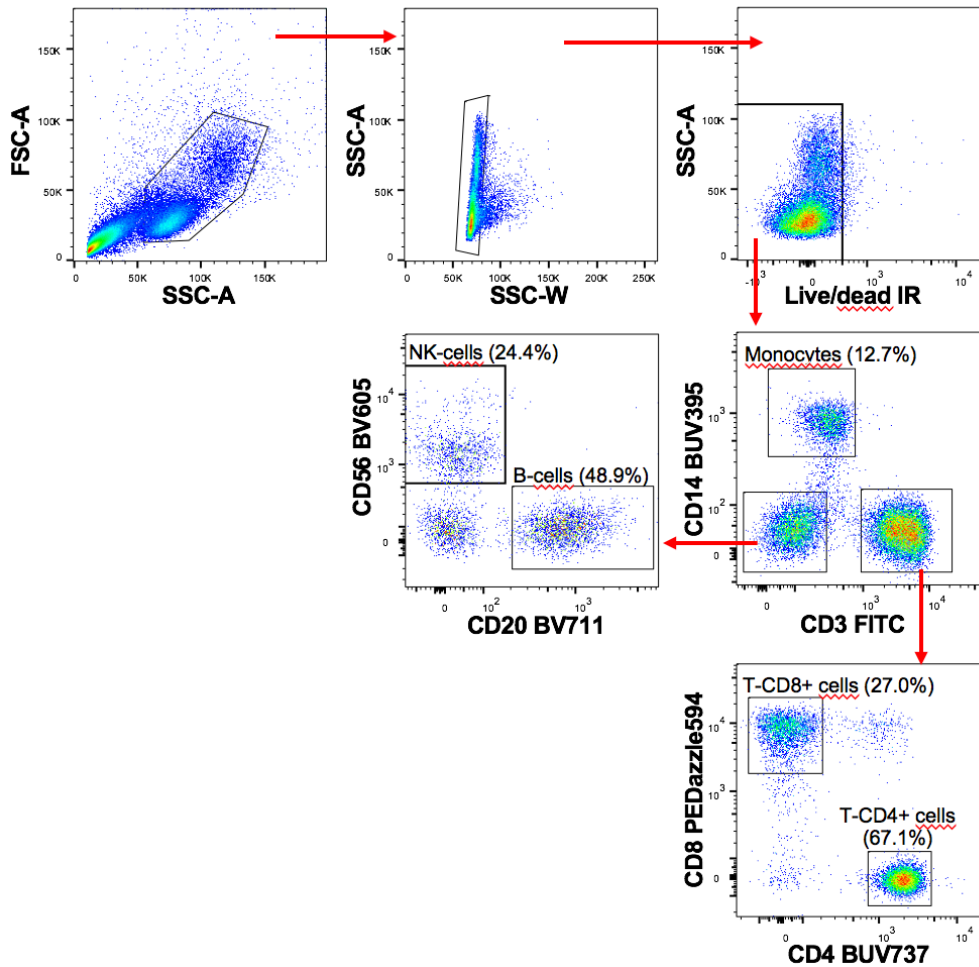
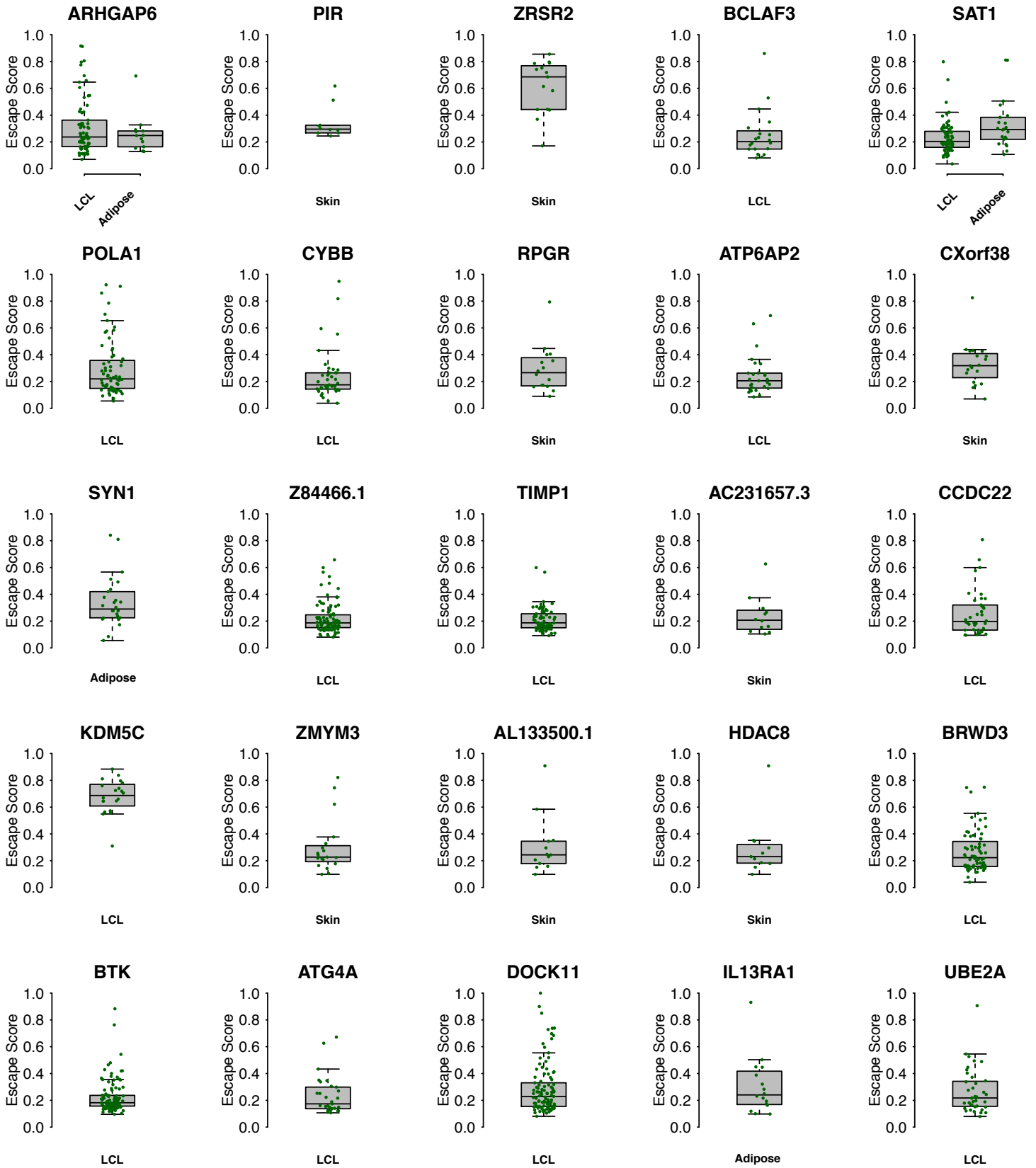
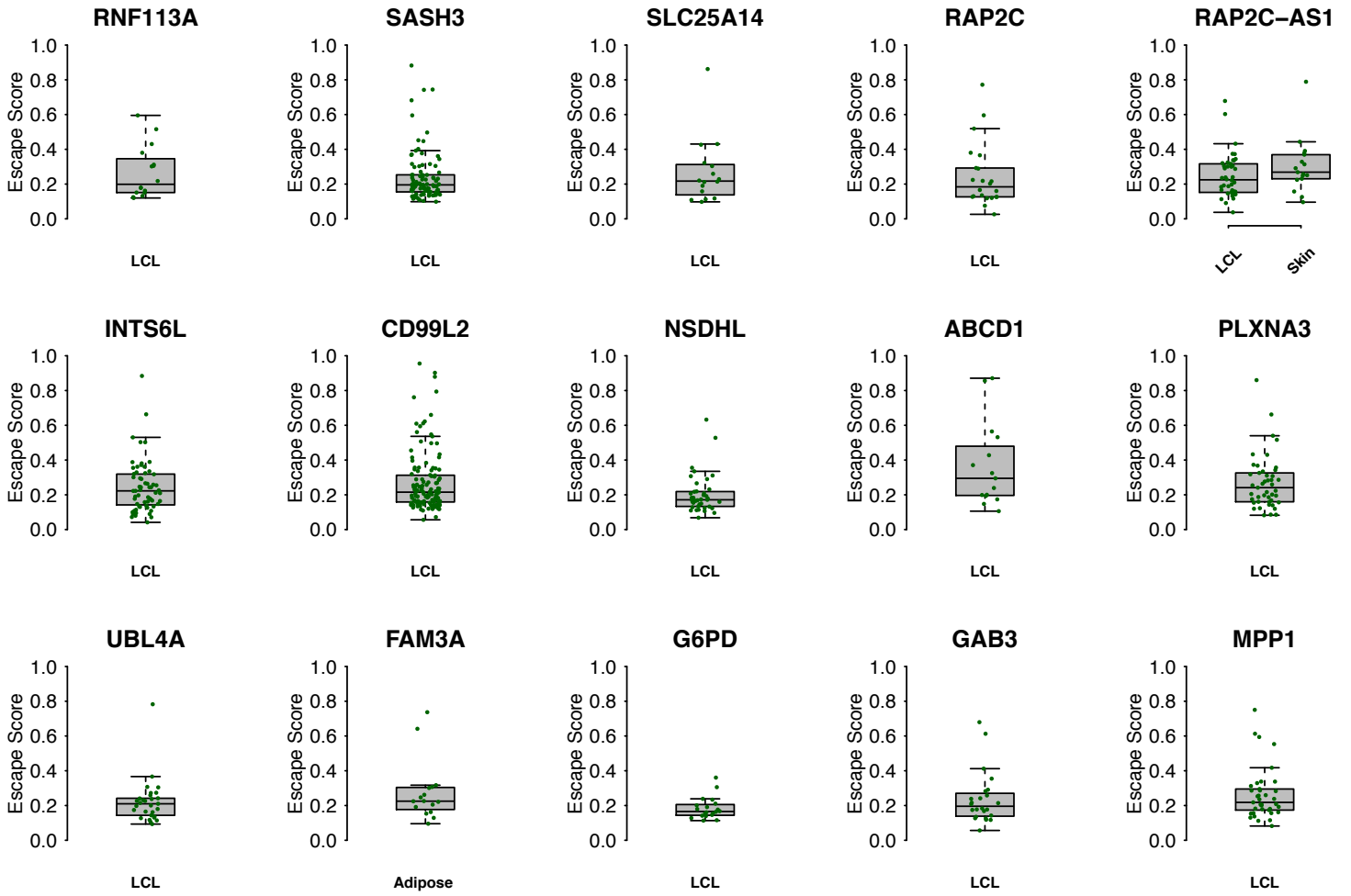


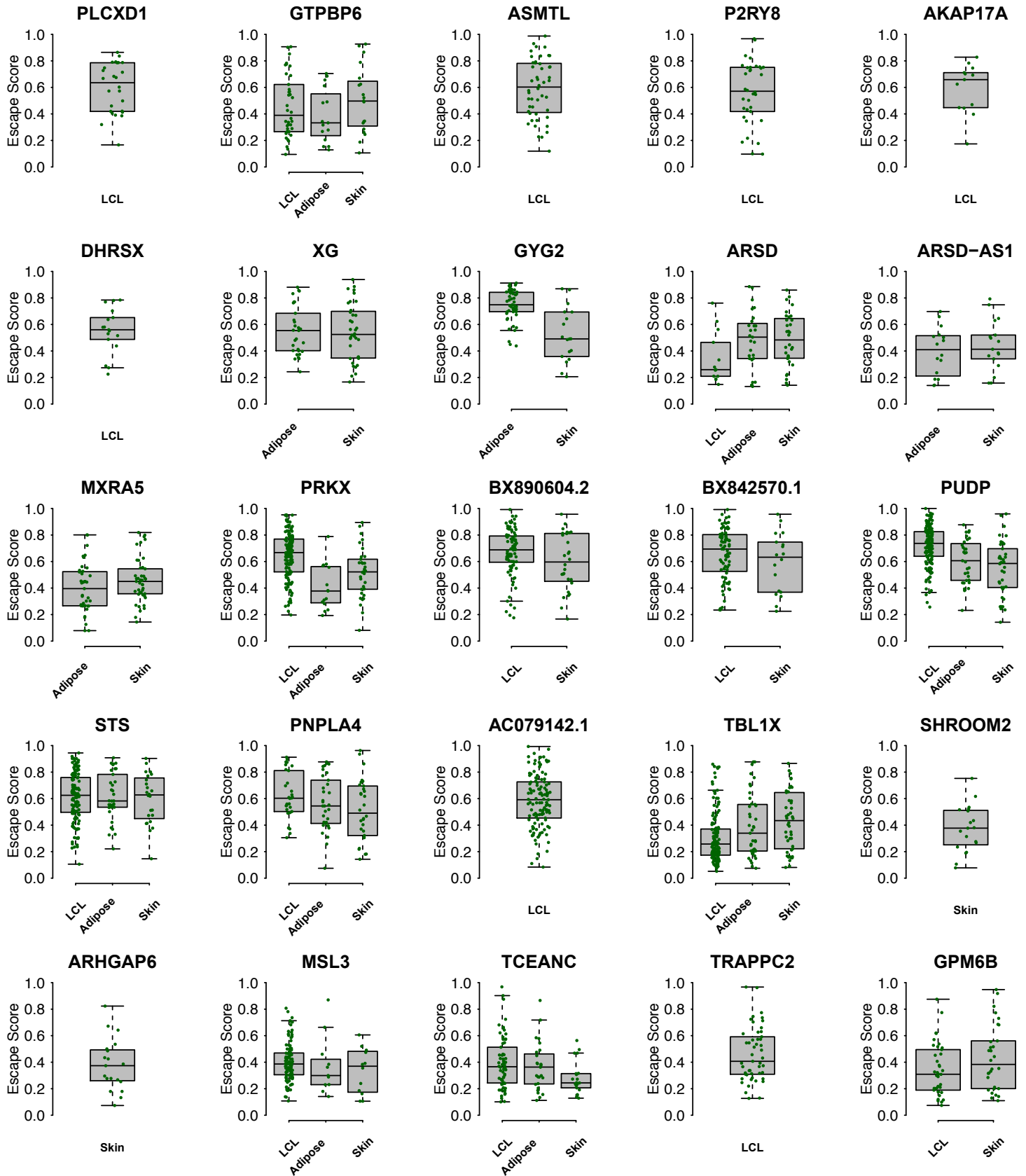
Fig.S5: Gating strategy for immune cell sorting. Gating strategy used to sort monocytes (CD14+), B (CD14-, CD3-, CD56-, CD20+), NK (CD14-, CD3-, CD20-, CD56+), T-CD4+ (CD14, CD3+, CD8-, CD4+) and T-CD8+ cells (CD14-, CD3+, CD4-, CD8+) from freshly isolated PBMCs from 2 monozygotic twins exhibiting skewed XCI in LCLs.

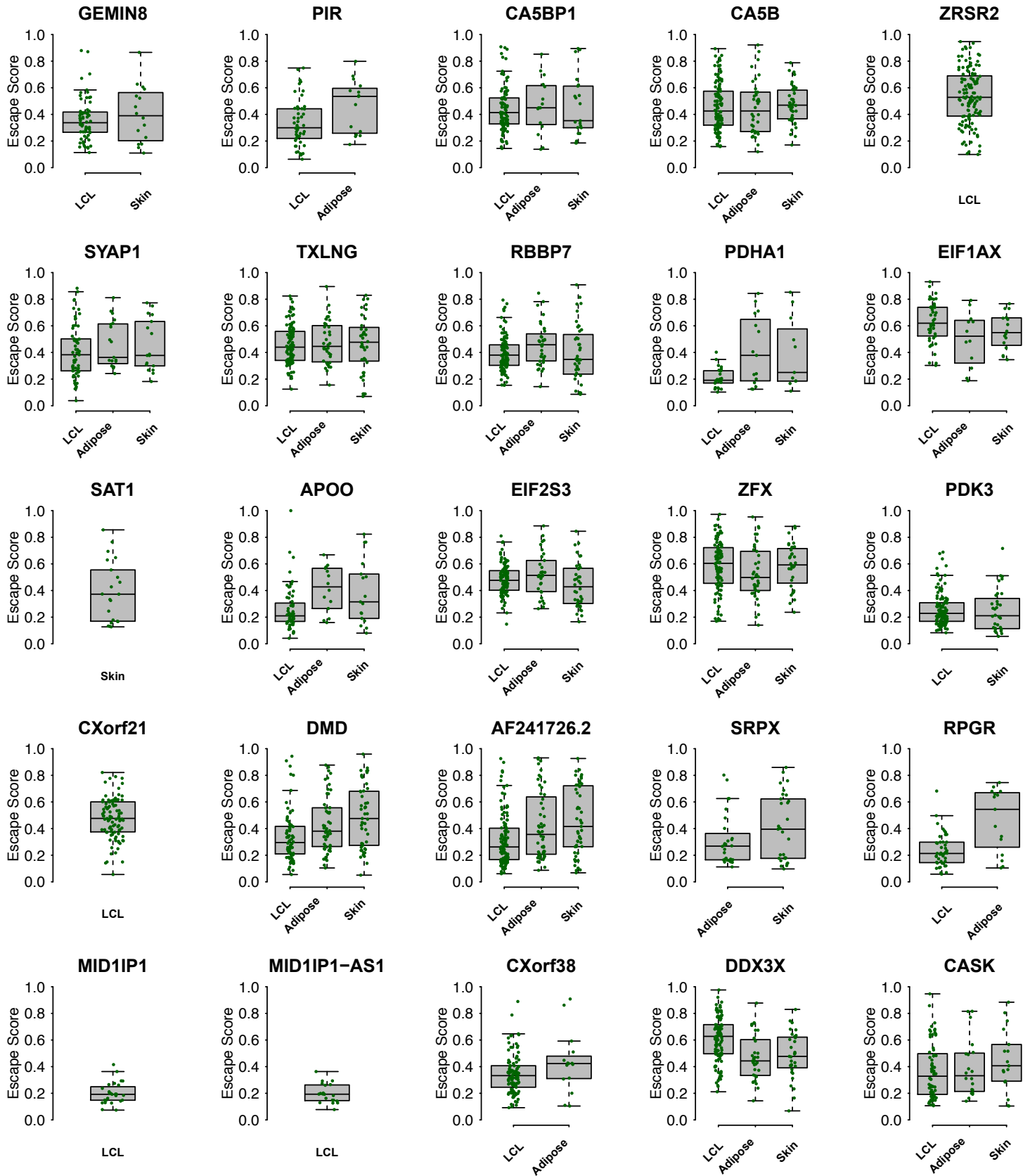
Fig_S3.pdf [Pages 26-27]: Boxplot of gene's EscScore in each skewed sample (≥ 10 tissue samples used for this analysis). Plotted are genes classified to have consistent EscScore across individuals. Each green dot is an individual.

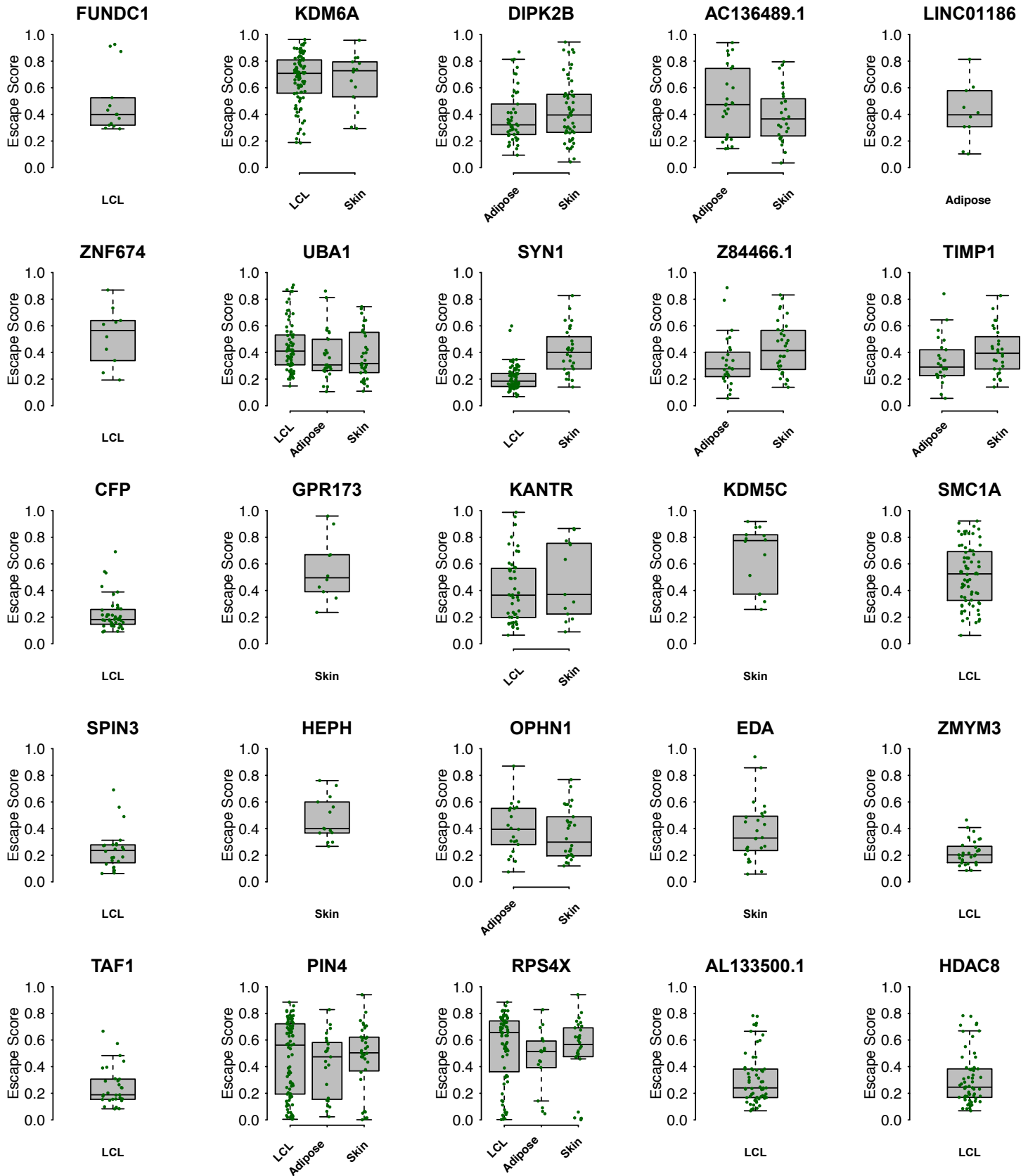
Fig_S4.pdf [Pages 28-32]: Boxplot of gene's EscScore in each skewed sample (≥ 10 tissue samples used for this analysis). Plotted are genes classified to have variable EscScore across individuals. Each green dot is an individual.

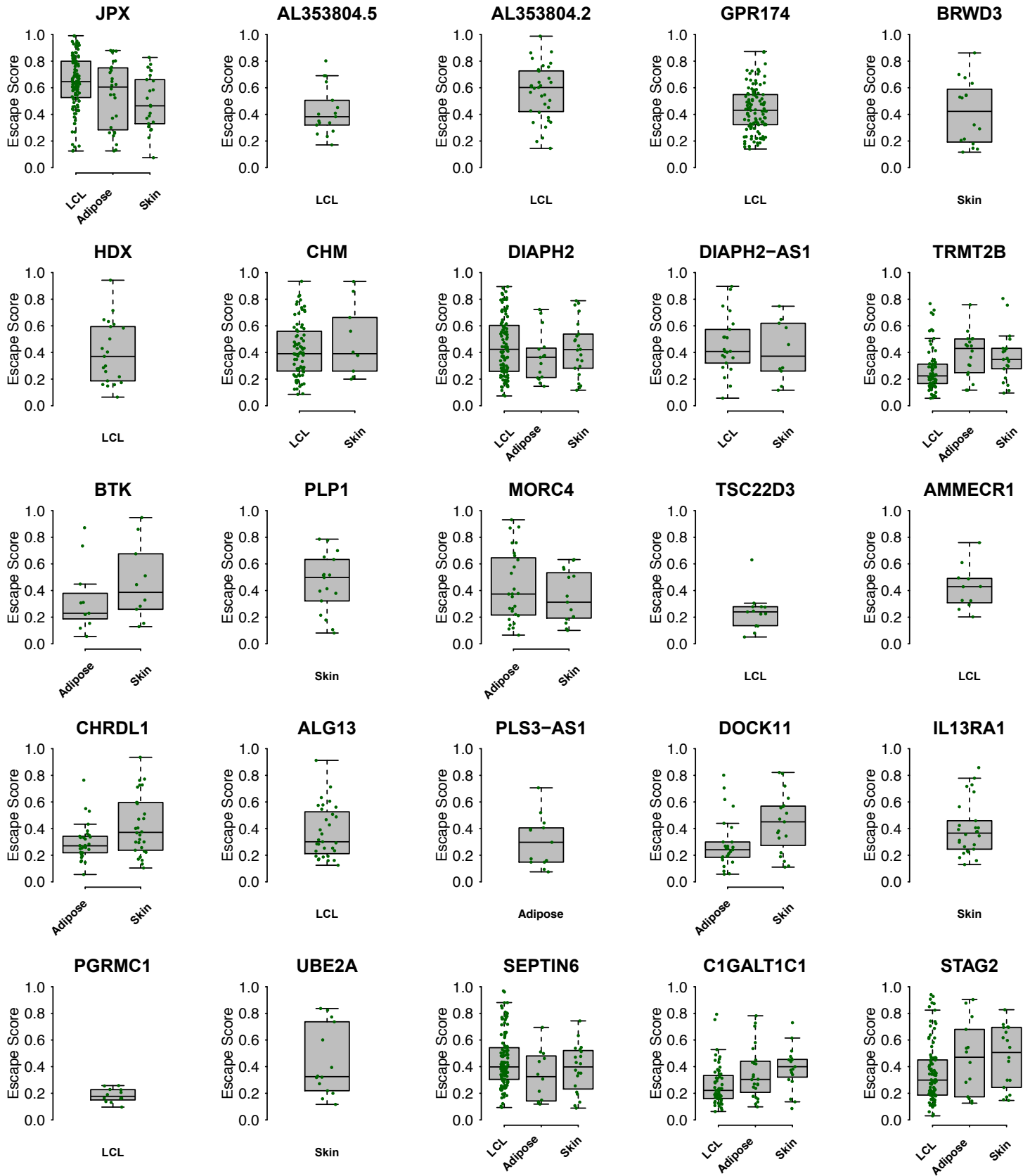


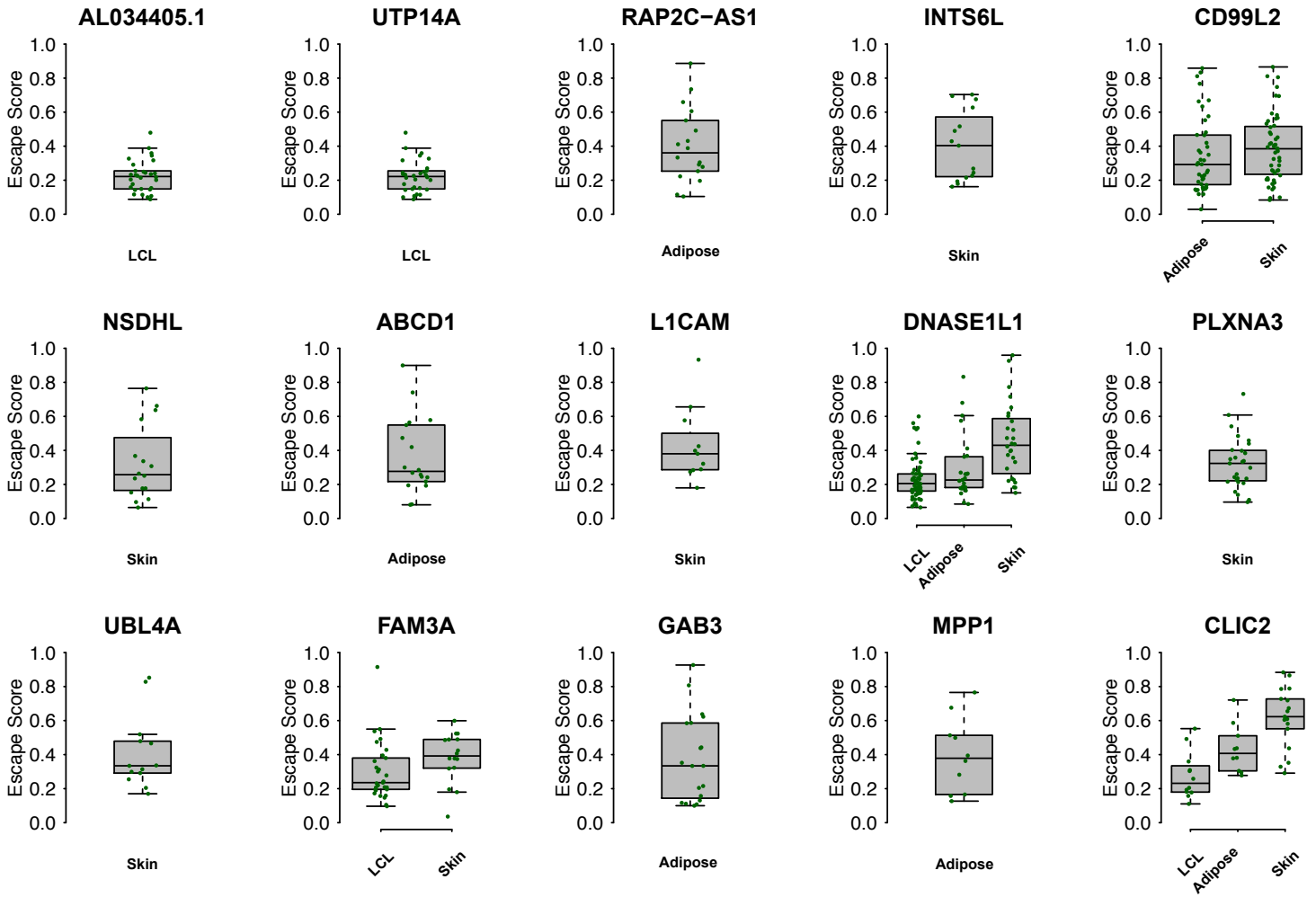












S. REFERENCES

- 1 Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-137, doi:10.1038/379131a0 (1996).
- 2 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).
- 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 Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973, doi:10.1126/science.1237973 (2013).
- 5 Pinter, S. F. *et al.* Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome Res* **22**, 1864-1876, doi:10.1101/gr.133751.111 (2012).
- 6 Brown, C. J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527-542 (1992).
- 7 Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244-248, doi:10.1038/nature24265 (2017).
- 8 Rupert, J. L., Brown, C. J. & Willard, H. F. Direct detection of non-random X chromosome inactivation by use of a transcribed polymorphism in the XIST gene. *Eur J Hum Genet* **3**, 333-343 (1995).
- 9 Zito, A. *et al.* Heritability of skewed X-inactivation in female twins is tissue-specific and associated with age. *Nat Commun* **10**, 5339, doi:10.1038/s41467-019-13340-w (2019).
- 10 Cotton, A. M. *et al.* Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol* **14**, R122, doi:10.1186/gb-2013-14-11-r122 (2013).
- 11 Carrel, L. & Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400-404, doi:10.1038/nature03479 (2005).
- 12 Carrel, L. & Brown, C. J. When the Lyon(ized chromosome) roars: ongoing expression from an inactive X chromosome. *Philos Trans R Soc Lond B Biol Sci* **372**, doi:10.1098/rstb.2016.0355 (2017).
- 13 Balaton, B. P., Cotton, A. M. & Brown, C. J. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biol Sex Differ* **6**, 35, doi:10.1186/s13293-015-0053-7 (2015).
- 14 Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-1093, doi:10.1093/bioinformatics/btp101 (2009).