



## King's Research Portal

DOI:

<https://doi.org/10.1038/s41467-019-10106-2>

*Document Version*

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Odhams, C. A., Roberts, A. L., Vester, S. K., S T Duarte, C., Beales, C. T., Clarke, A. J., Lindinger, S., Daffern, S. J., Zito, A., Chen, L., Jones, L. L., Boteva, L., Morris, D. L., Small, K. S., Fernando, M. M. A., Cunninghame, Graham, D. S., & Vyse, T. J. (2019). Interferon inducible X-linked gene CXorf21 may contribute to sexual dimorphism in Systemic Lupus Erythematosus. *Nature Communications*, *10*, 1-15. [2164].  
<https://doi.org/10.1038/s41467-019-10106-2>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

1 **Interferon inducible X-linked gene *CXorf21* may contribute to sexual**  
2 **dimorphism in Systemic Lupus Erythematosus**

3

4 Christopher A. Odhams<sup>1,5,10</sup>, Amy L. Roberts<sup>1,6,10</sup>, Susan K. Vester<sup>1</sup>, Carolina S.T. Duarte<sup>1</sup>,  
5 Charlie T. Beales<sup>1</sup>, Alexander J. Clarke<sup>2</sup>, Sonja Lindinger<sup>1,3,7</sup>, Samuel J. Daffern<sup>1,8</sup>, Antonino  
6 Zito<sup>4</sup>, Lingyan Chen<sup>1</sup>, Leonardo L. Jones<sup>1</sup>, Lora Boteva<sup>1,9</sup>, David L. Morris<sup>1</sup>, Kerrin S. Small<sup>4</sup>,  
7 Michelle M. A. Fernando<sup>1</sup>, Deborah S. Cunninghame Graham<sup>1</sup>, Timothy J. Vyse<sup>1,\*</sup>

8

9 <sup>1</sup>Department of Medical & Molecular Genetics, King's College London, SE1 9RT, UK

10 <sup>2</sup>Kennedy Institute of Rheumatology, University of Oxford, Oxford, OX3 7FY, UK

11 <sup>3</sup>University of Applied Sciences - FH Campus Wien Favoritenstrasse 226, 1100 Wien, Austria

12 <sup>4</sup>Department of Twin Research & Genetic Epidemiology, King's College London, London,  
13 SE1 7EH, UK

14 Current addresses: <sup>5</sup>Genomics England, Queen Mary University of London, Dawson Hall,

15 London, EC1M 6BQ, UK <sup>6</sup>Department of Twin Research and Genetic Epidemiology, King's

16 College London, SE1 7EH, UK <sup>7</sup>Institute of Biophysics, Johannes Kepler University Linz,

17 Gruberstrasse 40, 4020 Linz, Austria <sup>8</sup>Department of Genetics, University of Cambridge,

18 Downing Street, Cambridge CB2 3EH, UK <sup>9</sup>MRC Human Genetics Unit MRC IGMM,

19 University of Edinburgh Western General Hospital, Edinburgh, EH4 2XU, UK

20 <sup>10</sup>These authors contributed equally

21 \*Corresponding author email: [timothy.vyse@kcl.ac.uk](mailto:timothy.vyse@kcl.ac.uk)

22

23 **Abstract**

24 Systemic lupus erythematosus (SLE) is an autoimmune disease, characterised by increased  
25 expression of type I interferon (IFN) regulated genes and a striking sex imbalance towards  
26 females. Through combined genetic, in silico, in vitro, and ex vivo approaches, we define  
27 *CXorf21*, a gene of hitherto unknown function, which escapes X-chromosome inactivation, as  
28 a candidate underlying the Xp21.2 SLE association. We demonstrate that *CXorf21* is an IFN-  
29 response gene and that the sexual dimorphism in expression is magnified by immunological  
30 challenge. Fine-mapping reveals a single haplotype as a potential causal cis-eQTL for *CXorf21*.  
31 We propose that expression is amplified through modification of promoter and 3'-UTR  
32 chromatin interactions. Finally, we show that the CXORF21 protein colocalises with TLR7, a  
33 pathway implicated in SLE pathogenesis. Our study reveals modulation in gene expression  
34 affected by the combination of two hallmarks of SLE: *CXorf21* expression increases in a both  
35 an IFN-inducible and sex-specific manner.

36

## 37 **Introduction**

38 Females have a clear immunological advantage over males, with reduced susceptibility to  
39 infection at an early age and a superior ability to produce antibodies and serum IgM following  
40 immune challenge<sup>1,2</sup>. The immunological gain in females is thought to contribute to the striking  
41 sexual dimorphism observed in human autoimmune disease - where over 80% of sufferers are  
42 female<sup>3</sup> - and corroborates the hypothesis that genetic risk to autoimmunity is an evolutionary  
43 consequence of positive selection for favourable immune responses to infection<sup>4</sup>. Systemic  
44 lupus erythematosus (SLE), an autoimmune disease characterised by anti-nuclear  
45 autoantibodies and a type I interferon (IFN) signature, displays one of the most striking female-  
46 biased imbalances (9:1) in disease prevalence. Although the underlying mechanism has yet to  
47 be fully elucidated, a prominent role of X chromosome dosage is supported by the karyotypic  
48 risks for SLE. Males with Klinefelter's syndrome (47, XXY) have a 14-fold increased  
49 prevalence of SLE compared to 46, XY males, which approximates to the prevalence seen in  
50 46, XX females<sup>5</sup>. Furthermore, whereas 45, XO females have lower risk<sup>6</sup>, SLE prevalence in  
51 47, XXX females is ~2.5 times higher than in 46, XX females<sup>7</sup>. Indeed, mammalian X  
52 chromosomes, for which males are hemizygous, are enriched for immune-related genes<sup>8</sup>.

53

54 X chromosome inactivation (XCI) is a unique mammalian dosage-compensation mechanism  
55 which equalises expression of X-linked genes between sexes<sup>9</sup>. This random process results in  
56 either the paternally or maternally inherited X chromosome becoming inactivated (Xi) through  
57 enriched heterochromatic modifications, which promotes gene silencing to leave one  
58 transcriptionally active X chromosome (Xa) in females<sup>10</sup>. However, an estimated 15% of X-  
59 linked genes, preferentially found on the Xp arm, escape XCI and thus display expression from  
60 both chromosomes, although typically expression is still lower from the Xi compared with

61 Xa<sup>11</sup>. A further 10% of X-linked genes display variable expression from the Xi – an observation  
62 which itself is variable between both individuals and cell types, and throughout development  
63 and ageing<sup>12</sup>. It is these XCI-escaping genes, through their partial or complete lack of dosage  
64 compensation, that are thought to contribute to genetic sexual dimorphism and phenotypic  
65 differences in X-chromosome aneuploidies<sup>13</sup>. Furthermore, the relaxation of Xi silencing in  
66 female mammals includes increases in the expression of several immunity-related genes<sup>14</sup>.  
67 How genes that escape XCI contribute to sexually dimorphic diseases has not been thoroughly  
68 studied.

69 A SLE association at the Xp21.2 locus (rs887369;  $P_{\text{META}} = 5.26 \times 10^{-10}$ ; OR = 1.15) was  
70 recently identified in a European GWAS and replication study<sup>15</sup>. Intriguingly, this locus is  
71 encoded outside the pseudo autosomal region (PAR) and the lead SNP (a synonymous variant)  
72 resides in the final exon of *CXorf21*, a protein-coding gene of unknown function. *CXorf21* has  
73 been shown to escape XCI in lymphoblastoid cell lines (LCLs), and is one of only 14 X-linked  
74 genes that is differentially expressed between Klinefelter’s syndrome (47, XXY) and 46, XY  
75 males in LCLs<sup>16,17</sup>. A recent whole-blood gene expression study also identified *CXorf21* as one  
76 of seven genes upregulated in female SLE patients displaying disease flare relative to those  
77 with infection<sup>18</sup>.

78  
79 Despite the stark karyotypic risk, there remains a lack of understanding of the contribution of  
80 the X chromosome to SLE, which is a leading cause of death in females aged under 34 years of  
81 age<sup>19</sup>. Here we describe fine-mapping and characterisation of the association at Xp21.2 through  
82 complementary genetic, *in silico*, *in vitro* and *ex vivo* approaches using both existing and newly  
83 generated data (all methods are summarised as a flow chart in Supplementary Fig. 1). We  
84 demonstrate that the candidate gene, *CXorf21*, is an IFN-responder with both cell-type specific  
85 and sexually dimorphic expression amplified by cellular activation. Additionally, we provide

86 evidence at the protein-level of CXORF21 co-localisation with TLR7; a gene causatively  
87 linked to SLE and which also evades XCI. Our study demonstrates IFN-inducible  
88 magnification of sexual dimorphic gene expression contributing to SLE risk.

89

## 90 **Results**

### 91 **Genetic refinement of the Xp21.2 SLE susceptibility locus**

92 The source of all cohorts used within this manuscript along with the analyses performed are  
93 presented as a flow diagram in Supplementary Fig. 1. The UK10K-1000 Genomes Project  
94 Phase 3 reference panel<sup>20</sup> was firstly used to impute the Xp21.2 locus of the Bentham and  
95 Morris et al SLE GWAS (10,995 individuals of European ancestry)<sup>15</sup>. Logistic regression  
96 revealed a synonymous variant, rs887369 (MAF = 0.24), to be the most significantly associated  
97 SNP ( $P = 3.34 \times 10^{-7}$ ; OR = 1.43, 95% C.I = 1.23-1.66; Fig. 1a) and conditional analysis upon  
98 rs887369 showed no evidence of independent association signals (Fig. 1b). Haplotype analyses  
99 revealed that rs887369 tags a single, 1kb haplotype block comprising five near-perfectly  
100 correlated SNPs mapping to the 3' region of *CXorf21* (Fig. 1c)– encoding a small, 301-amino  
101 acid protein of unknown function. SNPs rs2529517 (distal) and rs887369 (proximal) define the  
102 boundaries of the associated haplotype, which map downstream of the 3'-UTR of *CXorf21*, and  
103 to the gene's third and final exon respectively (Fig. 1c). Three of the five associated SNPs are  
104 transcribed from *CXorf21*, with rs887369 effecting a synonymous change (V209) and both  
105 rs2532873 and rs2710402 residing in the 3'-UTR. The remaining two SNPs, rs2429517 and  
106 rs2429518, are located in the downstream intergenic region of *CXorf21*. The associated  
107 haplotype is distinctly separated from neighbouring haplotypes by high recombination ( $D' <$   
108  $0.6$ ,  $r^2 < 0.2$ ) and accordingly, the risk haplotype itself represents the only observed association  
109 with SLE ( $\chi^2 = 29.87$ ,  $P = 4.63 \times 10^{-8}$ ; Chi-Square test; Fig. 1d).

110

111 *CXorf21* is known to escape XCI<sup>16</sup>. We performed a statistical test on the association with  
112 rs887369 to see if a model that assumed the SNP was in an area that escaped inactivation fitted  
113 better than a model assuming full inactivation. A likelihood ratio test to fit both association

114 models failed to reject the model of full inactivation ( $P = 0.78$ ). Therefore, from our  
115 case/control data we have no evidence against the hypothesis that this association lies in an  
116 area of full inactivation. To extend these analyses, we determined the odds ratios of the risk  
117 alleles in females and males separately. We observed a higher odds ratio for females  
118 homozygous for the rs887369 [C] risk allele with respect to homozygous for [A] non-risk (OR  
119 = 1.58, 95% C.I. 1.29 – 1.93) compared to the males (OR = 1.46, 95% C.I. = 1.10 – 1.92), who  
120 are hemizygous for the risk or non-risk alleles. The higher odds ratio in females is likely to  
121 reflect a gene dosage effect secondary to some degree of loss of X inactivation.

122

### 123 **The risk haplotype increases expression of *CXorf21* in LCLs**

124 As no protein-altering variants were identified through fine-mapping, we sought to establish  
125 whether the SLE risk alleles at *CXorf21* colocalised with *cis*-eQTLs for gene transcription.  
126 Non-random inactivation of the X chromosome (skewing) and variability in the degree of  
127 silencing of gene expression from the inactivated X in females complicates the identification  
128 of X chromosome eQTLs in females. Therefore, to study *cis*-eQTLs at the *CXorf21* locus, we  
129 employed two complementary methods of assessing the influence of the risk haplotype, tagged  
130 by rs887369, on the expression of genes within the Xp21.2 region: 1) using the hemizyosity  
131 of males to isolate the allelic effects; 2) removing females exhibiting strong evidence of  
132 extreme skewed XCI to reduce the variability in the degree of skewing of X-chromosome  
133 expression.

134

135 The associated haplotype, tagged by rs887369 [C], correlated with increased expression of  
136 *CXorf21* in LCLs from male samples in the Geuvadis RNA-Seq dataset ( $\beta = 1.56$ ,  $P = 1.94 \times$   
137  $10^{-03}$ ; linear-regression; Fig. 2a). The expression of neighbouring genes *GK* and *TAB3* showed



138 no significant association with rs887369 ( $P = 0.7$  and  $P = 0.09$ , respectively, linear-regression,  
139 Fig. 2a). Many variants may act as *cis*-eQTLs, however it is important to note that rs887369  
140 was the most significantly associated *cis*-eQTL for *CXorf21* (Fig. 2b) and the remaining *cis*-  
141 genes (the *MAGEB* family and *NROB1*; +/-1Mb from rs887369) were not expressed in LCLs  
142 (RPKM < 1).

143 The allelic effect on *CXorf21* expression was only nominally significant when performing the  
144 *cis*-eQTL analyses in female individuals from Geuvadis RNA-Seq dataset in LCLs ( $P = 0.02$ ;  
145 linear-regression; Supplementary Fig. 2a). In order to investigate *cis*-eQTL effects at rs887369  
146 in females, we interrogated an additional RNA-Seq gene expression dataset in LCLs  
147 constructed exclusively from female donors from the TwinsUK cohort<sup>21</sup>. This dataset was  
148 selected for analysis as it had been previously analysed for skewing of X chromosome  
149 inactivation using allele specific expression (ASE) of the *Xist* silencing lncRNA (manuscript  
150 in preparation). In order to study potential *cis*-eQTLs at the *CXorf21* locus, we removed  
151 individuals showing marked skewing, in whom the *Xist* ASE showed that one parental X  
152 chromosome contributed less than 20% of the *Xist* expression. In this subset of 412 non-skewed  
153 individuals, we observed a statistically significant increase of *CXorf21* expression with respect  
154 to the rs887369 [C] risk allele in females ( $P = 7.00 \times 10^{-03}$ ; linear-regression; Fig. 2c).

155  
156 We validated this effect *in vitro* by qPCR of independent LCL samples selected from the  
157 HapMap Project on the basis of their genotype at rs887369. In these cells a 1.9-fold increase  
158 of *CXorf21* mRNA was detected between rs887369 homozygous risk and non-risk females ( $P$   
159 =  $4.1 \times 10^{-5}$ ; *t*-test ;Supplementary Fig. 2b). Following validation of the anti-CXORF21  
160 antibody (Supplementary Fig. 3), the observed increase in expression by the risk allele was  
161 found to persist at protein-level ( $\beta = 0.49$ ,  $P = 2.88 \times 10^{-5}$ ; Fig. 2d; raw data are shown in  
162 Supplementary Fig. 2c).

163

164 **Risk variants increase *CXorf21* expression upon activation**

165 We expanded our analysis and interrogated a genotype-expression cohort from a range of  
166 human primary *ex vivo* immune cells. When assessing male samples only, we found that the  
167 associated haplotype, tagged by rs887369, was a significant *cis*-eQTL for *CXorf21* in both  
168 Lipopolysaccharide (LPS) stimulated ( $P = 1.08 \times 10^{-03}$ ) and IFN- $\gamma$ -stimulated ( $P = 1.10 \times 10^{-3}$ ;  
169 <sup>3</sup>; linear-regression) monocytes (Fig. 2e). The [C] risk allele once again correlated with  
170 increased *CXorf21* expression. Interestingly, no statistically significant *cis*-eQTL associations  
171 were observed in the unstimulated experiments: B cells, NK cells, neutrophils, and monocytes,  
172 which suggests an activation-state specificity of the *cis*-eQTL. When the same analysis was  
173 performed in the female samples of the same cohort, no significant *cis*-eQTLs were detected  
174 in any of the cell types (Supplementary Fig. 2d).

175

176 **Epigenetic fine-mapping of the Xp21.2 associated haplotype**

177 Using the Roadmap Epigenomes Project<sup>22</sup> (twelve different histone marks across 127 cell and  
178 tissue types), we used chromatin fine-mapping to functionally prioritize the five SNPs carried  
179 on the 1Kb associated haplotype. The associated SNPs localised only to a single histone  
180 modification, H3K36me3, across five cell types: blood mononuclear cells, peripheral blood B  
181 cells, monocytes, neutrophils, and the lymphoblastoid cell line GM12878. Analysis of the  
182 signal value distribution of H3K36me3, designating regions of active transcription, across  
183 these cell types revealed that rs887369 localised to the binding site summit of H3K36me3  
184 whilst the remaining four SNPs on the haplotype localised to the tails of the signal distribution  
185 (Fig. 3a). The greatest enrichment of H3K36me3 across the entire *CXorf21* gene locus was  
186 concentrated to +/-100bp of rs887369 in monocytes ( $P = 6.1 \times 10^{-14}$ ; MACS2) and neutrophils

187 ( $P = 2.0 \times 10^{-17}$ ; MACS2; Fig. 3b). The rs887369 SNP also localised to the binding site summit  
188 of H3K36me3 in primary B cells, LCLs, and blood mononuclear cells, with significant, albeit  
189 weaker enrichment.

190

191 As verification, we performed the same analysis using ChIP-Seq experiments (n=612) from  
192 the venous blood portion of the Blueprint Epigenetics consortium<sup>23</sup> (8 modifications across 24  
193 unique cell types from 83 donors). Only 22 ChIP-Seq experiments presented evidence of  
194 overlap with the SLE-associated haplotype, and strikingly, all of these intersections were again  
195 for the H3K36me3 chromatin modification. No other histone modifications intersected this  
196 region. All five SNPs on the 1kb SLE-associated haplotype were found to overlap with  
197 H3K36me3 in monocytes, B cells and neutrophils – corroborating the Roadmap Epigenomics  
198 data. We were unable to make robust conclusions on differential H3K36me3 signal between  
199 the sexes as the sample sizes per cell-type were too small (Supplementary Fig. 4,  
200 Supplementary Table 1).

201

202 Lastly, the associated SNPs in the 3'-UTR of *CXorf21* showed no evidence of disrupting a  
203 microRNA binding site after interrogation using miRDB<sup>24</sup>.

204

### 205 **The risk haplotype interacts with the promoter of *CXorf21***

206 We sought to investigate a conceivable molecular mechanism whereby the SLE-associated  
207 haplotype at the 3' end of *CXorf21* modulates expression through alteration of chromosome  
208 interactions. The promoter capture Hi-C dataset curated by the CHiCP resource<sup>25</sup> was  
209 interrogated. This resource comprises Hi-C data from 17 primary immune cell types taken from  
210 healthy donors. Three of the five SNPs (rs887369, rs2710402, and rs2532873) on the

211 associated haplotype, which are closest to *CXorf21*, reside within the chrX:30576528-  
212 30582605 target region. Across all primary immune cell types tested, the target region was  
213 found to interact with four baits (Fig. 3c): the promoter region of *CXorf21* (chrX:30595248-  
214 30603761); the promoter of *GK*; and two intronic antisense RNAs of *TAB3* (*TAB3-AS1* and  
215 *TAB3-AS2*). Significant bait-target region interactions (CHiCAGO score  $\geq 5$ ) were detected  
216 exclusively in neutrophils (Fig. 3d), where the *CXorf21* promoter bait interaction presented the  
217 greatest strength of interaction with the risk haplotype target region (6.09). Strong but sub-  
218 threshold interactions ( $3 \leq$  CHiCAGO score  $< 5$ ) were also detected for the risk haplotype  
219 target and the *CXorf21* promoter bait region in monocytes (3.72) and naïve B cells (3.15). The  
220 strength of the interaction score between the risk haplotype target region and the *CXorf21*  
221 promoter was found to correlate strongly with the signal strength of epigenetic marks (from  
222 ENCODE<sup>26</sup>) indicative of active gene-expression (H3K4me3 and H3K27ac) for matched cell  
223 types (Fig. 3e). These findings suggest that the 3'-promoter interaction of *CXorf21* is more  
224 pronounced in the cell types in which *CXorf21* is expressed, and the interaction itself is  
225 involved in regulation of expression. In fact, by assessing the transcription factor landscape at  
226 the *CXorf21* locus, we found significant binding events of RNA polymerase II (POLR2A) at  
227 the 3' SLE-associated region in immune cell types only; corroborating our hypothesis that the  
228 observed chromatin looping is necessary for transcriptional regulation (Supplementary Fig. 5).

229

### 230 **Sexual dimorphic expression is magnified upon activation**

231 GTEx RNA-Seq data<sup>27</sup> across 45 different cell/tissue types confirmed that there is significant  
232 sexual dimorphic expression of *CXorf21* in both LCLs and thyroid tissue (LCLs: 1.78-fold  
233 greater in females,  $P = 1.10 \times 10^{-5}$ , thyroid: 1.33-fold greater,  $P = 2.65 \times 10^{-3}$  following  
234 Bonferroni multiple testing correction; *t*-test; Supplementary Fig. 6a and Supplementary Table

235 2). Neighbouring genes *GK* and *TAB3* were equally expressed in both sexes, in both LCLs and  
236 in the cell types in which both genes are most expressed, suggesting escape from XCI at this  
237 locus is restricted to *CXorf21*. Using HapMap LCLs selected on the basis of their genotype at  
238 haplotype-tagging rs887369, we employed the validated anti-CXORF21 antibody  
239 (Supplementary Fig. 3) to quantify protein abundance by western blot. When we examined cell  
240 lines that all carried at least one risk haplotype, we confirmed that protein expression was  
241 higher in females (Supplementary Fig. 6b): females harboured 3.6 times more CXORF21 than  
242 males ( $P = 0.006$ ;  $t$ -test). These findings imply that the slight variation in *CXorf21* mRNA  
243 results in an amplified effect on overall protein abundance. To ensure these results were not a  
244 consequence of monoallelic expression of *CXorf21* in pauciclonal LCLs, we assayed *CXorf21*  
245 expression from microarray experiments across a range of primary *ex vivo* immune cells and  
246 found, as with other XCI escaping genes, the effect size of *CXorf21* expression between sexes  
247 was cell-type specific<sup>11</sup>. In resting B cells, NK cells, neutrophils, and monocytes, no significant  
248 difference in transcript abundance of *CXorf21* between sexes was observed (Supplementary  
249 Fig. 6c, Supplementary Table 3). However, though we see global increase of *CXorf21*  
250 expression in both sexes, a striking sexual dimorphic responses to LPS- or IFN- $\gamma$ -stimulation  
251 in monocytes was observed (Fig. 4;  $P_{\text{LPS}} = 1.41 \times 10^{-12}$ ;  $P_{\text{IFN-}\gamma} = 9.29 \times 10^{-8}$ ;  $t$ -test). Transcript  
252 abundance of *CXorf21* in monocytes is therefore greatest in females under immune-stimulated  
253 conditions.

254

### 255 ***CXorf21* is a likely interferon response gene**

256 Given the marked increase of *CXorf21* expression in stimulated immune cells (including LCLs  
257 which exhibit a partially activated phenotype<sup>28</sup>) and the observed up-regulation of IFN-  
258 regulated genes in SLE<sup>29</sup>, we investigated whether *CXorf21* is an interferon response gene by

259 profiling gene expression using in-house microarray data in primary *ex vivo* B cells taken from  
260 healthy females (n=49 in total, of which n=32 were treated with IFN- $\alpha$ ). We observed *CXorf21*  
261 is one of eighteen X chromosome genes (including *TLR7*, *IL13RA1*, and *ELF4*) which were  
262 up-regulated in response to IFN- $\alpha$  stimulation (fold-change: 2.41;  $P = 6.0 \times 10^{-9}$ ; ANOVA;  
263 Fig. 5a). No other Xp21.1 gene was modulated by IFN- $\alpha$ . We profiled the epigenetic landscape  
264 surrounding the *CXorf21* locus in ENCODE data and detected significant and localised binding  
265 events of NF- $\kappa$ B, STAT1, STAT2, STAT3, IRF4, and IRF3 at the immediate promoter region  
266 of *CXorf21* in LCLs (Fig. 5b). We also identified a single interferon-stimulated response  
267 element (ISRE) +25bp upstream of the *CXorf21* transcription start site (TSS). This sequence  
268 motif and the array of interferon regulatory factors was not detected in any of the promoters of  
269 other genes within the Xp21.2 locus (Fig. 5b).

270

### 271 **Functional characterisation of the Xp21.2 SLE risk locus**

272 Eight genes are encoded at the Xp21.2 SLE risk locus (rs887369;  $P = 3.34 \times 10^{-7}$ ; OR = 1.43):  
273 four Melanoma Antigen B (MAGEB) family genes (*MAGEB1-4*), *NROB1* encoding the DAX1  
274 nuclear receptor, *GK* (glycerol kinase), *TAB3* (TGF-beta activated kinase 1 and MAP3K7  
275 binding protein 3), and *CXorf21* (Fig. 1a; Supplementary Table 4). None of these eight genes  
276 had reported associations with immune-related phenotypes in human or mouse.

277

278 *CXorf21* is the only gene in the locus with a discrete immune-specific mRNA expression  
279 profile; being most highly expressed in the spleen, appendix, bone marrow, and lymph nodes  
280 (Protein Atlas; Supplementary Fig. 7, GTEx and FANTOM5 validation in Supplementary Fig.  
281 8). This suggests the mechanism by which the SLE-risk haplotype is affecting disease risk is  
282 through candidate gene *CXorf21*. To refine this analysis in terms of cellular expression, we

283 used data from Blueprint Epigenome (RNA-sequencing) and BioGPS (microarray) to show  
284 that within immune cell types, the expression of *CXorf21* is largely restricted to monocytes,  
285 neutrophils, and B cells (Supplementary Fig. 9). We corroborated these findings in RoadMap  
286 Epigenomics data and found a striking chromatin landscape around the transcription start site  
287 of *CXorf21*, indicative of epigenetic silencing in non-immune cell types (Supplementary Fig.  
288 10). The expression profile of *CXorf21* at protein-level was largely consistent with the mRNA  
289 data; though CXORF21 protein was found to be in equal abundance in certain secondary  
290 immune tissue such as the bowel and skin (Supplementary Fig. 11).

291  
292 RNA-Seq co-expression analysis across a range of human cell and tissue-types was undertaken  
293 using the COEXPRES algorithm<sup>30</sup>. The results indicate that *GPR65* (G-couple protein receptor  
294 65) tops the ranking, whose protein product is important in lysosomal function<sup>31</sup>. Examination  
295 of the top 100 ranked genes revealed the expression signature of *CXorf21* correlated with the  
296 Toll-like receptor (TLR) signalling pathway including *TLR7*, *TLR6*, *PIK3CG*, and *PIK3CD*  
297 (Supplementary Table 5<sup>30</sup>), of which *TLR7* was highest ranked. The correlation between the  
298 expression of the two X-linked genes, *CXorf21* and *TLR7*, was replicated in TwinsUK RNA-  
299 Seq data<sup>21</sup> from LCLs from non-skewed females (n= 271;  $\rho = 0.38$ ;  $P = 6 \times 10^{-11}$ ).

300  
301 In order to gain further insight into the potential function of CXORF21, we utilized high-  
302 throughput affinity-purification mass spectrometry data from BioPlex<sup>32</sup> and revealed a high  
303 confidence (quantitative score: 0.999) protein-protein interaction between CXORF21 and  
304 SLC15A4, encoded by the SLE susceptibility gene *SLC15A4* (rs1059312;  $P_{\text{META}}=1.48 \times 10^{-13}$ ;  
305 OR=1.17)<sup>15</sup>. *SLC15A4* is an immune-restricted lysosomal amino-acid transporter required for  
306 TLR7- and TLR9-mediated type I IFN production in dendritic cells and B cells in lupus<sup>33</sup>.

307 Interestingly, in the BioPlex data, CXORF21 was also found to interact with itself, suggesting  
308 probable oligomerization of this protein.

309

### 310 **Protein level correlates with disease activity in females**

311 In a modest cohort ( $n_{\text{cases}} = 19$ ;  $n_{\text{controls}} = 13$ ) we did not observe a statistically significant  
312 difference in CXORF21 protein abundance between female case and controls in CD14+  
313 monocytes or CD19+ B cells (Supplementary Fig. 12). However, we observed an age-  
314 dependent correlation between CXORF21 and SLE Disease Activity Index  
315 (SLEDAI). CXORF21 protein abundance is positively correlated with SLEDAI in SLE  
316 females < 35 years of age (CXORF21 ~ SLEDAI \* Stratified Age) in both CD14+ monocytes  
317 and CD19+ B cells (Supplementary Fig. 13). A likelihood ratio test (LRT) rejected the model  
318 of SLEDAI as a single variable (upper panels Supplementary Fig. 13) in favour of an  
319 interaction model in monocytes (LRT  $P = 0.0002$ ) and B cells (LRT  $P = 0.0006$ ). The rejection  
320 of the single variable models are also supported by BIC ( $\Delta\text{BIC}_{\text{monocytes}} = 8.1$ ;  $\Delta\text{BIC}_{\text{Bcells}} =$   
321  $5.9$ ). We observed a significant interaction term (SLEDAI \* Stratified Age) in monocytes ( $P =$   
322  $0.002$ ), though the interaction term in B cells did not pass multiple testing correction ( $P =$   
323  $0.011$ ; lower panels Supplementary Fig. 13).

324

### 325 **CXORF21 protein may act within endosomal pathway**

326 CXORF21 is a small protein of ~34kDa as identified by Western Blot. Very little of the  
327 secondary/tertiary protein structure of CXORF21 could be accurately determined by the Phyre  
328 bioinformatics prediction tool<sup>34</sup>. Thus, to gain insight into the protein's function we sought to  
329 determine its cellular location in *ex vivo* cells from healthy females and the GM12878  
330 lymphoblastoid cell line. We undertook multispectral imaging flow cytometry (MIFC) with a



331 range of labels for different organelles. The results demonstrated minimal co-localisation of  
332 CXORF21 with nuclear, Golgi or lysosomal markers in *ex vivo* PBMCs, and this was not  
333 affected by IFN stimulation (Supplementary Fig. 14 and 15). In view of these negative findings  
334 and the data showing co-expression of *CXorf21* with components of the Toll-like receptor  
335 signalling pathway (Supplementary Table 6), we utilised the increased resolution of structured  
336 illumination microscopy (SIM) to determine whether there was any evidence for colocalisation  
337 of CXORF21 with TLR7. Representative images for the staining in resting and stimulated *ex*  
338 *vivo* B cells are shown (Fig. 6a through 6d). We quantified the correlation between signals  
339 obtained from CXORF21 with TLR7 staining (Fig. 6e) and determined the colocalisation of  
340 the two staining signals in B cells using Mander's co-efficient (see Methods). These analyses  
341 were undertaken in both resting B cells and stimulated B cells (B cell receptor cross-linking  
342 and CD40) and in each case with and without exposure to IFN- $\alpha$ . We conclude that there is  
343 weak colocalisation between TLR7 and CXORF21 in *ex vivo* B cells (Pearson correlation 0.3  
344  $< \rho < 0.4$ ). No significant differences in colocalisation between CXORF21 and TLR7 were  
345 observed after IFN- $\alpha$  treatment of resting or IgM/CD40 stimulated B cells.

346

347 As the endosomal intracellular pathway interacts with the autophagy pathway (which has also  
348 been implicated in SLE pathogenesis)<sup>35</sup> we sought to determine whether CXORF21 colocalised  
349 with autophagosomes, once more utilising SIM. Using LC3 as a marker of the autophagosome,  
350 representative results of the joint staining (LC3 and CXORF21) are shown for Ig/CD40  
351 stimulated B cells (Fig. 7a) with exposure to the inhibitor of autophagic flux, bafilomycin (Fig.  
352 7b), and Ig/TLR7/8 stimulated B cells (Fig. 7c) with bafilomycin (Fig. 7d). The results from  
353 multiple cells are summarised in Fig. 7e and 7f, which show no colocalisation between LC3  
354 and CXORF21 in bafilomycin-treated *ex vivo* B cells when stimulated with Ig/CD40 or

355 Ig/TLR7/8. Assaying CXORF21 protein abundance by western blot in starved LCL (see  
356 methods) indicates that the amount of CXORF21 is not altered by the addition of bafilomycin  
357 and hence it is unlikely that CXORF21 is an autophagy substrate (Fig. 7g; left panel). The blot  
358 shows some elevation of sequestome 1 (p62), an autophagosome cargo protein, following  
359 exposure to bafilomycin, which would be expected (Fig. 7g; right panel).

360

## 361 **Discussion**

362 The underrepresentation of genetic associations on the X chromosome in autoimmune disease  
363 is highly paradoxical given the prominent sex bias towards females and the increased density  
364 of immune related genes compared to the autosomes. This is partly due to the paucity of sex  
365 chromosome data in genome-wide studies; only 33% of GWAS report sex chromosome data<sup>36</sup>.  
366 We sought to functionally investigate the undefined SLE susceptibility locus Xp21.2 from our  
367 own GWAS dataset (rs887369;  $P = 3.34 \times 10^{-7}$ ; OR = 1.43). Our investigation defines *CXorf21*  
368 – encoding a protein of hitherto unknown function – as the candidate gene and demonstrates  
369 its expression is upregulated through by a number of distinct factors: chromosome X dosage  
370 and loss of XCI, the risk haplotype (tagged by rs887369), and cellular activation by interferon  
371 (summarised in Fig. 8). Our study supports the hypothesis that altered expression of X-linked  
372 genes contributes to the sexual dimorphism in autoimmunity<sup>14</sup> and provides some preliminary  
373 evidence for the role of *CXorf21* in SLE, although this topic clearly warrants further  
374 investigation.

375

376 To date, six X-encoded SLE susceptibility loci have been identified, and four have been shown  
377 to harbour genes that escape XCI (*TLR7*, *TMEM187*, *IRAK1*, and *CXorf21*). Of these, *CXorf21*  
378 is the most robustly escaping; possessing evidence of escape in ~80% of individuals in contrast  
379 to the remaining genes that exhibit escape in <30%<sup>16</sup>. We show that escape from XCI is highly  
380 localised to *CXorf21* across a +/-1Mb window.

381

382 The Xp21.2 locus is not as strongly associated with SLE in individuals of non-European  
383 ancestry, although an association has been reported in Koreans<sup>37</sup>. This is partially explained by  
384 the marked disparity in allele frequency of risk allele rs887369 [major allele: C] between

385 populations (1000Genomes: 0.76 in Europeans, 0.92 in Africans, 0.95 in Asians). The lower  
386 minor allele frequency in non-Europeans may clearly impact on power, especially as non-  
387 European GWAS have been of smaller sample size. The fraction of individuals who exhibit  
388 XCI of *CXorf21* is reported to be diminished in individuals of African descent (relative to those  
389 of European descent<sup>16</sup>); however, lower allele frequencies of transcribed polymorphisms and  
390 limited samples sizes impede power. Whether allele frequency of rs887369 and reduced XCI  
391 escape are correlated or whether variation at rs887369 itself is causal to a degree of escape  
392 poses an interesting line of enquiry. Furthermore, the reduced level of escape in non-Europeans  
393 may mean the effect size will limit the power of this locus to be detected.

394

395 *CXorf21* has a discrete expression pattern in immune cells, both adaptive and innate, with the  
396 greatest expression of *CXorf21* found in monocytes and neutrophils, primary B cells and LCLs.  
397 It appears to be epigenetically inert in non-immune cell types, suggesting the regulatory  
398 mechanisms driving expression of *CXorf21* are not present in non-immune cell types. These  
399 data align with the observation that other candidate genes of SLE and their accompanying  
400 causal variants exhibit a discrete expression signature and *cis*-regulatory landscape that is  
401 largely restricted to immune cell subsets; particularly B cells (including B-lymphoblastoid cell  
402 lines), T cells, and monocytes<sup>15,38-40</sup>.

403

404 We demonstrate that *CXorf21* expression is upregulated in LPS and IFN- $\gamma$ -stimulated  
405 monocytes, and in IFN- $\alpha$ -stimulated B cells, with the magnitude of increase greater in females  
406 leading to significant sexually dimorphic expression levels. We have also identified binding  
407 sites of respective transcription factors from these signalling cascades: IRF3, NF- $\kappa$ B and  
408 STAT1-3 at the immediate promoter of *CXorf21* suggesting *CXorf21* transcription could be a  
409 primary response gene of the TLR4 (IRF3) and IFN (STATs) signalling pathways. However,

410 *CXorf21* expression decreases following acute (2h) LPS-stimulation, suggesting *CXorf21* is in  
411 fact a late response gene induced by secondary activation of the TLR4-induced type I IFN feed  
412 forward loop<sup>41</sup>. Indeed, late response genes are characterised by STAT binding sites and  
413 ISRE<sup>41</sup>, which we also identify in the *CXorf21* promoter. Sex differences in the LPS-induced  
414 monocyte response have been previously reported, whereby females have heightened  
415 activation and cytokine release compared with males, although the underlying mechanism has  
416 yet to be delineated<sup>42,43</sup>.

417

418 rs887369 tags a short 1kb haplotype comprising five perfectly correlated SNPs. The haplotype  
419 is an eQTL for *CXorf21*, with the risk allele increasing the gene's expression; we hypothesise  
420 that a self-regulatory mechanism involving modification of H3K36me3 state and chromatin  
421 looping affects RNA polymerase II within the gene promoter (Fig. 8).

422

423 The expression of *CXorf21* transcript has previously been shown to be the most accurate  
424 delineator of disease flare from infection in SLE patients<sup>18</sup>. Interestingly, this previous study  
425 was conducted in largely non-European patients, suggesting the role of *CXorf21* is not limited  
426 to individuals of European ancestry. Further supporting our hypothesis that *CXorf21* is an IFN-  
427 inducible gene, the genes with dysregulated expression at exome-wide significant expression  
428 changes identified in this study are enriched for IFN-inducible genes<sup>18</sup>. We observed an age-  
429 dependent correlation between CXORF21 expression and disease activity using flow  
430 cytometry in a modest cohort, with CXORF21 protein abundance positively correlating with  
431 SLEDAI in patients <35 years of age. These data warrant further investigation and suggest age-  
432 stratified analysis in disease cohorts could be illuminating.

433

434 The CXORF21 protein has no known function and the primary amino acid sequence gives no  
435 clear clues in this regard. In an attempt to provide some insight into the function of CXORF21,  
436 we conducted a number of imaging studies to investigate its intracellular location. These  
437 studies showed that CXORF21 is present in both the nucleus and cytoplasm. Interestingly, we  
438 show that there was some colocalisation of CXORF21 with TLR7 in B cells. This intracellular  
439 toll-like receptor was selected for imaging as it is known to play a role in nucleic acid sensing  
440 in SLE and our analyses revealed some degree of co-expression of *TLR7* and *CXorf21* at the  
441 RNA level. Intracellular toll-like receptors operate in a complex system involving the  
442 endosomal and lysosomal compartments<sup>44</sup>. However, the precise points at which CXORF21  
443 and TLR7 may interact within these compartments is not clear on the basis of our data, but  
444 further exploration of this question should reveal more insights into the function of CXORF21  
445 and how it promotes systemic autoimmunity.

446

447 The female-biased sex imbalance of autoimmune diseases is not understood. Our study, which  
448 characterises the SLE association at *CXorf21*, an IFN-inducible gene which escapes XCI, adds  
449 support to the hypothesis that sex bias in immune function has a genetic basis and provides an  
450 underlying immunological mechanism that underpins the sexual dimorphism in SLE.

451

## 452 **Methods**

### 453 **European SLE GWAS data**

454 Genotype data from 10,995 individuals of matched European ancestry (4,036 SLE cases, 6,959  
455 controls) genotyped on the Illumina HumanOmni1 BeadChip from the Bentham and Morris et  
456 al<sup>15</sup> study were imputed as outlined below. These data had undergone quality control and PCA  
457 as described<sup>15</sup>.

458

### 459 **Imputation**

460 The European SLE GWAS<sup>15</sup>, Fairfax *et al*<sup>45,46</sup>, and Naranbhai *et al*<sup>47</sup> cohorts were imputed  
461 using UK10K-1000GP3 merged reference panel across the X:30077468-31077846 1Mb  
462 region, plus a 2Mb buffer region (GRCh37 assembly). A full imputation without pre-phasing  
463 was conducted using IMPUTE2 to increase the accuracy of imputed genotype calls<sup>48,49</sup>.  
464 Imputed genotypes were filtered using an info score (IMPUTE2) threshold of 0.5. The most  
465 likely genotype from IMPUTE2 was taken if its probability was > 0.5. If the probability fell  
466 below this threshold, it was set as missing.

467

### 468 **Allelic and haplotype association fine-mapping**

469 Imputed data from the European SLE GWAS were filtered to include variants with MAF >  
470 0.01 and HWE >  $1 \times 10^{-4}$ , and minimum genotype rate > 90%. SNPTEST 2.5.2<sup>50</sup> was used to  
471 test for additive models of allelic associations across the X:30077468-31077846 1Mb region,  
472 fitting a logistic regression model (including the first four covariates from the original  
473 GWAS<sup>15</sup>) with equal effect size between males and females<sup>50,51</sup>. Independent signals were  
474 assessed by including the genotype for the rs887369 SNP as a covariate using the SNPTEST

475 algorithm. Association plots were generated using LocusZoom<sup>52</sup>. Haplotype association  
476 analysis and LD calculations between SNPs were performed using Haploview 4.2<sup>53</sup>  
477 (implementing X-chromosome analysis) using the entire GWAS of 10,995 individuals.  
478 Specifically, haplotype blocks across a 100Kb region anchored on rs887369 were defined by  
479 the confidence intervals algorithm<sup>54</sup> and haplotype association testing performed by a Chi-  
480 Square test using marker thresholds of MAF > 0.01 and HWE >  $1 \times 10^{-4}$ , and minimum  
481 genotype rate > 90% (657 SNPs in total).

482

483 We fitted two models for association in SNPTEST. The inactivation model is the default in  
484 SNPTEST's newmlmethod with male genotypes coded as 0/1 and females coded as 0/0.5/1  
485 and one shared estimated effect (log odds ratio). In the escape model we used SNPTEST with  
486 the stratify\_on option which fits separate effects for males and females. In both models, we fit  
487 a different intercept for males and females (using sex as a covariate in the inactivation model)  
488 and so the two models only differ by one parameter (being the differing log odds ratio). A  
489 likelihood ratio test (LRT) on one degree of freedom was performed in R (using the likelihood  
490 values output by SNPTEST), where the escape model was tested against the simpler  
491 inactivation model. A statistically significant result (based on the p-values from the LRT)  
492 would therefore reject the inactivation model.

493

#### 494 **Genotype data for *ex vivo* cell eQTL cohorts**

495 X chromosome SNPs of the *Fairfax et al*<sup>45,46</sup> and *Naranbhai et al*<sup>47</sup> cohorts with an Illumina  
496 GenCall score of <0.7 and called on both X and Y were removed. PLINK v1.9<sup>55</sup> was used to  
497 remove samples that failed sex check assignments. Following separation of male and females,  
498 SNPs were removed if: HWE <  $1 \times 10^{-4}$ , MAF < 0.01, and SNP missingness > 10%. Individuals



499 were removed if >10% of SNPs were missing. Coordinates were converted from hg18 to hg19  
500 using the UCSC liftOver application<sup>56</sup>.

501

## 502 **Genotype-expression cohorts and *cis*-eQTL analysis**

503 Gene-level RNA-Seq data from LCLs were downloaded from ArrayExpress (Geuvadis;  
504 EGEUV-1)<sup>57</sup> and genotypes (X:30077468-31077846) of these individuals containing SNPs  
505 (MAF > 0.05) were taken from the 1000 Genomes Project Phase III<sup>58</sup>. Expression data of  
506 purified *ex vivo* primary immune cells were obtained from *Fairfax et al*<sup>45,46</sup> and *Naranbhai et*  
507 *al*<sup>47</sup>. Details are described in the respective articles. Data include resting B-cells, natural killer  
508 cells, and monocytes<sup>45</sup>; IFN- $\gamma$  stimulated monocytes after 24h, LPS stimulated monocytes after  
509 2h, LPS stimulated monocytes after 24h<sup>46</sup>; and resting neutrophils<sup>47</sup>. In all instances, *cis*-eQTL  
510 association analysis (1Mb of rs887369) was performed against expression residuals using the  
511 linear-model of the MatrxieQTL R package<sup>59</sup> including the number of PCs described in the  
512 respective articles.

513

514 The TwinsUK RNA-Seq eQTL cohort profiled in LCLs<sup>21</sup> was used for *cis*-eQTL association  
515 analysis in non-skewed females (n=412). Individuals were firstly assessed for skewed X-  
516 chromosome inactivation patterns using allele-specific expression of *Xist* to estimate the  
517 proportion of X inactivation from each parental X chromosome. Individuals were removed if  
518 the allele-specific expression of XIST-linked SNPs was <0.2 or >0.8, these parameters were  
519 chosen on the basis of precedence<sup>60-62</sup>. *cis*-eQTL analysis in the twins was performed as above  
520 against exon-count residuals corrected for probabilistic estimation of expression residuals  
521 (PEER) factors and family relatedness<sup>63</sup>.

522

523 Differential expression analysis of *CXorf21* between males and females using GTEx RNA-Seq  
524 (TMP) data across the 45 cell/tissue types where expression data were available for both sexes  
525 was performed using an unpaired *t*-test between males and females after grouping by cell/tissue  
526 type. Associations passing the Bonferroni adjusted *P*-value cut-off of  $P_{BF} < 0.05$  were deemed  
527 significant.

528

### 529 **Cell culture**

530 LCLs were obtained from Coriell Biorepository and cultured in suspension at 5% CO<sub>2</sub>, 37°C  
531 in RPMI 1640 medium supplemented with 2mM L-glutamine, 15% foetal bovine serum, 100  
532 units/ml penicillin and 100 µg/ml streptomycin. Cells were seeded every two days to a  
533 concentration of 300,000 viable cells/ml. Individuals used in functional assays were of  
534 European descent (GM12878, HG01702, HG01786, HG01746, HG0111, HG01628,  
535 HG00254, HG12878, HG01501, HG01507, HG01504, HG00269, HG00232).

536

### 537 **qPCR**

538 Total RNA was extracted with the RNeasy Mini Kit (QIAGEN) according to manufacturer's  
539 instructions. cDNA synthesised with the cDNA Synthesis Kit (Thermo Scientific) and  
540 quantified using the NanoDrop 2000 spectrophotometer. qPCR reactions were performed using  
541 the TaqMan® Universal PCR Master Mix and Universal ProbeLibrary System Technology  
542 (UPL) from Roche. Primers were purchased from Sigma and reactions performed using the  
543 Applied Biosystems 7500 and subsequent analysis with SDS 2.3. *CXorf21* F:  
544 GGATGTTTGACACAGACTTCAAA, R: CCGGATCAGATGAGCAGATT, UPL #65.  
545 *ACTB* F: AGAGCTACGAGCTGCCTGAC, R: CGTGGATGCCACAGGACT, UPL #9.  
546 Relative abundance and fold change was calculated using the  $\Delta\Delta C_t$  method.

547

548 **Verification of anti-CXORF21 antibody by gene-knockdown**

549 Gene-knockdown of *CXorf21* in LCLs (GM12878) was performed by siRNA using the  
550 Nucleofector II Device (Lonza) and Amaxa Cell Line Nucleofector Kit V. Two days before  
551 transfection, cells were seeded to a concentration of  $0.5 \times 10^6$  cells/ml. In duplicate,  $2 \times 10^6$  cells  
552 were spun at 100g for 10mins and re-suspended in 100 $\mu$ l supplemented transfection solution  
553 and 20pmol Silencer Select Pre-Designed & Validated siRNA (Thermo Fisher Scientific)  
554 against *CXorf21* (#4392420). The Silencer Select Negative Control No. 1 siRNA (#4390843)  
555 was used as a non-targeting negative control at the same concentration. Cell/siRNA  
556 suspensions were transferred to a Nucleofector cuvette and electroporated using the X-001  
557 programme. Samples were cultured in 1.5ml medium in a 12-well plate and harvested 48 h  
558 post-transfection.

559

560 **Immunoblot**

561 Cell lysates were prepared in RIPA buffer (Sigma-Aldrich) and run on a SDS polyacrylamide  
562 gel for electrophoresis. Protein was transferred onto a nitrocellulose membrane and blocked in  
563 5% milk-PBS solution. The rabbit polyclonal anti-CXORF21 antibody (Atlas Antibodies;  
564 HPA001185) was used at a concentration of 1:1,000 and the secondary polyclonal swine anti-  
565 rabbit immunoglobulins/HRP (Dako; P0217) at 1:1,000. Membranes were stripped by  
566 Restore™ Western Blot Stripping Buffer (Thermo Fisher) and re-probed with mouse  
567 monoclonal  $\beta$ -Actin antibody (Santa Cruz Biotechnology; sc-47778) at 1:4,000 and anti-mouse  
568 IgG HRP conjugate (Promega; W4028) at 1:5,000 or secondary goat anti-mouse IgG HRP  
569 conjugate (Invitrogen; A16078) at 1:10,000. ImageJ was used to calculate the density of the  
570 bands relative to the loading control. Rabbit anti-SQSTM1/p62 (Cell Signalling, 5114) was

571 used at a concentration of 1:1,000 and detected with secondary goat anti-rabbit IgG HRP  
572 conjugate (Invitrogen; A16110) at 1:10,000. Raw blots are presented in accompanying Source  
573 Data file.

574

### 575 **Epigenetic fine-mapping**

576 SNPs in X:30077468-31077846 were downloaded from the 1000 Genomes Project Phase III<sup>58</sup>.  
577 Epigenetic data across all available cell types (n=127) in NarrowPeak format were obtained  
578 from the NIH Roadmap Epigenomics Project<sup>22</sup>. Peaks were filtered for genome-wide  
579 significance using an FDR threshold of 0.01, and peak widths harmonised to 2kb in length  
580 centred on the peak summit. SNPs were reported as being localised to an epigenetic mark if  
581 they overlapped the 2kb region. The signal value of the epigenetic mark was reported for the  
582 exact coordinate of the SNP using the signal track of the mark in bigWig format visualized  
583 using IGV v2.3.80<sup>64</sup>.

584

585 NarrowPeak files of ChIP-Seq experiments (H3K4me3, H3K27ac, H3K4me1, H3K36me3,  
586 H3K27me3, H3K9me3, H3K9/14ac, H2A.Zac) were downloaded from the Blueprint  
587 Epigenome Project ftp site  
588 ([http://ftp.ebi.ac.uk/pub/databases/blueprint/data/homo\\_sapiens/GRCh38/](http://ftp.ebi.ac.uk/pub/databases/blueprint/data/homo_sapiens/GRCh38/)). Only non-  
589 diseased cell-types from venous blood were selected for analysis (24 unique cell-types). Using  
590 the GRCh38 genomic positions of the 5 SNPs carried on the associated haplotype, intersection  
591 was performed against the genome-wide binding sites of the selected Blueprint ChIP-Seq  
592 experiments as per the Roadmap Epigenomics project (above). Fold-enrichment of the peaks  
593 that overlapped the associated haplotype were compared by unpaired *t*-test between males and  
594 female samples for H3K36me3 across different cell-types.

595

596 **Promoter capture Hi-C chromatin interaction data**

597 Chromatin interaction data across a 17 primary immune cell-types was assessed using Capture  
598 Hi-C Plotter (CHiCP; [www.chicp.org](http://www.chicp.org))<sup>25</sup>. The study focuses on autoimmune susceptibility loci  
599 from GWAS and ImmunoChip integrating promoter capture Hi-C datasets from three separate  
600 studies<sup>65-67</sup>. The bait to target coordinates and interaction scores were extracted from CHiCP  
601 manually. Scores were defined by the CHiCAGO algorithm<sup>68</sup>, where scores  $\geq 5$  were  
602 considered as significant interactions.

603

604 **B-cell isolation and cell stimulation**

605 CD19<sup>+</sup> B cells from healthy female subjects (n=49) were isolated by negative selection using  
606 the Dynabeads Untouched Human B Cells Kit (Invitrogen).  $1.5-3 \times 10^6$  cells/ml *ex vivo* B cells  
607 were cultured in RPMI 1640 medium, supplemented with 20% FCS, 2mM L-glutamine and  
608 100 U/mL penicillin/streptomycin. B cells from 32 of the 49 subjects were incubated with or  
609 without IFN- $\alpha$  2b (1000 U/ml; PBL Assay Science) at 37°C and 5% CO<sub>2</sub>. Cells were harvested  
610 after 6 h or 20 h as indicated.

611

612 For immunostaining, human B Cell Isolation Kit II (Miltenyi Biotec).  $1 \times 10^6$  cells/ml *ex vivo* B  
613 cells were cultured in RPMI 1640 medium, supplemented with 10% heat-inactivated FBS, 2  
614 mM L-glutamine and 100 U/mL penicillin/streptomycin. B cells were stimulated with 10  $\mu$ g/ml  
615 F(ab')<sub>2</sub> Fragment Anti-Human IgG+IgM (Jackson ImmunoResearch) and either 0.1  $\mu$ g/ml  
616 CD40L with 0.1  $\mu$ g/ml Enhancer (Enzo) or 5  $\mu$ g/ml resiquimod (Sigma). B cells were  
617 incubated with or without 10 nM bafilomycin A1 (Sigma) for 3 h before harvesting and with

618 or without 1000 U/ml IFN- $\alpha$  2b (PBL Assay Science) at 37°C and 5% CO<sub>2</sub>. Cells were  
619 harvested after 20 h or 27 h as indicated.

620

### 621 ***Ex vivo* B cell RNA extraction and array hybridization**

622 RNA was isolated using the RNeasy Mini kit (Qiagen) according to the manufacturer's  
623 instructions and integrity assessed using the Agilent 2100 Bioanalyzer (Agilent) with the RNA  
624 6000 Pico Kit (RIN < 8 excluded). cDNA was synthesised from 50ng of RNA using the High  
625 Capacity RNA-to-cDNA Kit (Applied Biosystems). Each sample was hybridised to Affymetrix  
626 Human Exon 1.0 ST arrays and expression data were obtained by fluorescence-based detection  
627 using the GeneChip Scanner 3000 7G (Affymetrix). Signal intensities were quantified and  
628 stored as CEL files.

629

### 630 **Quality control of exon array**

631 Quality control was carried out using the probe-set and transcript cluster annotation release  
632 33.1 (GRCh37 build). Probe and probe-set filters were applied to the data as recommended<sup>69</sup>.  
633 All probe sets targeting RefSeq-annotated RNA transcripts were included. Probes containing  
634 polymorphisms (MAF>0.01) from 1000Genomes were removed. Cross-hybridizing probes  
635 and probe sets containing less than three probes were also excluded. Detection above  
636 background noise (DABG) was calculated for all CEL files and probe-sets were filtered using  
637 Affymetrix Power Tools. Probe sets with DABG  $P>0.01$  in 50% of resting or IFN- $\alpha$  stimulated  
638 samples were removed. Probes and probe-sets that failed QC filters were removed from the  
639 data using Affymetrix Power Tools.

640

641 Intensity signals were normalized at exon-level and log<sub>2</sub>-transformed using the robust multi-  
642 array average algorithm in the Affymetrix Expression Console software (build 1.2.1.20). Array  
643 hybridization quality was verified using Affymetrix Expression Console according to the  
644 recommendations of the Affymetrix Quality Assessment of Exon and Gene Arrays White  
645 Paper. All arrays showed high hybridization quality and a normal distribution of probe intensity  
646 signals.

647 PCA was performed using Partek GS version 6.6 (Partek Incorporated) and sample outliers  
648 removed. Duplicate data for one monozygotic twin pair was processed in both batches to be  
649 used as technical replicates and sibling data from the same twin pair within each batch were  
650 used as biological replicates. Correlation between replicates was assessed using a Spearman  
651 correlation test in R. All replicates showed high correlation ( $r^2 > 0.89$ ). A total of 81 samples  
652 from 49 individual twins were included in the analyses.

653

#### 654 **Exon array data normalization and analysis**

655 Probe sets were summarized to generate gene-level data by calculating the winsorized mean  
656 (10% and 90%) using Partek GS. Batch effects were accounted for using the sva ComBat  
657 function<sup>70</sup>. Differential gene expression was calculated using Partek GS with a mixed-model  
658 analysis of variance (ANOVA) as follows:  $Y = \mu + \text{treatment} + \text{individual ID} + \text{twin ID} + \text{PC1}$   
659  $+ \text{PC3} + \text{error}$ . The fitted ANOVA model regressed expression levels at each gene (Y) on fixed-  
660 effect terms (treatment, explained by PC2) and on random-effect terms denoting individual ID,  
661 family structure and zygosity (twin ID) and PCs explaining most of the data variability (PC1  
662 and PC3).

663

## 664 **SLE Patients and Healthy Controls**

665 Female patients meeting the American College of Rheumatology (ACR) criteria for the  
666 definition of SLE active disease<sup>71</sup> were recruited from Louise Coote Lupus unit, Guy's Hospital  
667 (n=19), following informed consent and with ethical approval (Research Ethics Committee;  
668 REC 12/LO/1273 and REC 07/H0718/49) and SLE Disease Activity Index (SLEDAI) scores  
669 were calculated<sup>72</sup>. The investigator was blinded to SLEDAI scores during measurement of  
670 CXORF21 protein abundance. Healthy female controls were recruited from the TwinsUK  
671 Bioresource. The TwinsUK study is approved by the research ethics committee at St Thomas  
672 Hospital, London. Volunteers gave informed consent and signed an approved consent form  
673 prior to participation. Volunteers were supplied with an appropriate detailed information sheet  
674 regarding the research project and procedure by post prior to attendance.

675

## 676 **PBMC isolation**

677 20ml of whole blood in EDTA anti-coagulant was taken from female volunteers (SLE or  
678 healthy controls). Peripheral blood mononuclear cells (PBMCs) were separated from whole  
679 blood using Histopaque-1077 Hybri-Max (Sigma-Aldrich) density centrifugation and plated at  
680  $2 \times 10^6$  cells/ml in RPMI 1640-medium (Gibco) supplemented with 10% foetal calf serum  
681 (FCS), 2mM L-glutamine and 100 U/mL penicillin/streptomycin (all from Invitrogen).

682

## 683 **Flow Cytometry**

684 PBMCs were first incubated with Human TrueStain FcX (5ul; BioLegend) to block Fc  
685 receptors, before cell-surface staining with 1µl anti-human CD14 PerCP-Cy5.5 (eBioscience;  
686 45-0149-42) and 1µl anti-human CD19 PE (eBioscience;12-0198-42) for 20min on ice. Cells  
687 were fixed with 200ul 1X stabilising fixative (BD biosciences) and then permeabilized in 0.1%



688 Triton X-100 (Sigma-Aldrich). Fc blocker was again added before intracellular staining of  
689 0.1µg rabbit polyclonal anti-human CXORF21 (Atlas antibodies; HPA001185) or 0.1µg rabbit  
690 monoclonal IgG isotype control (Abcam; ab172730), as appropriate, for 60min on ice.  
691 Following washing, cells were incubated with secondary goat anti-rabbit-Alexa Fluor 488  
692 (Abcam; ab150077) antibody at 1:2000. Cells were washed and resuspended in 250µl PBS for  
693 analysis on BD FACSCanto™ II cytometer (BD Biosciences) using BD FACSDiva software  
694 (version 8.0.1; BD Biosciences). Compensation was performed using compensation beads (BD  
695 Biosciences), and cytometer settings were standardised using Cytometer Setup and Tracking  
696 Beads (BD Biosciences). Following data acquisition, FlowJo v.10.1 software was used to  
697 calculate the Median fluorescent intensity (MFI). An unpaired Student's t-test was used for  
698 case-control analyses. Logistic regression models were fitted for CXORF21 abundance as a  
699 function of SLEDAI, and as a function of SLEDAI stratified by age (under/over 35 years of  
700 age) with an interaction term. The models were compared using a likelihood ratio test (LRT;  
701 d.f. = 5) and BIC using R. Multiple testing was corrected using Bonferroni correction.  
702 Preliminary results showed no expression of CXORF21 on cell surface.

703

#### 704 **ImageStream Analysis**

705 Multispectral imaging flow cytometry (MIFC) was performed on an ImageStreamX (Amnis)  
706 instrument. Golgi colocalisation:  $2 \times 10^6$  cells were fixed with 200ul 1X stabilising fixative  
707 (BD biosciences) and then permeabilized in 0.1% Triton X-100 (Sigma-Aldrich). Fc blocker  
708 was added before intracellular staining with 0.1µg rabbit polyclonal anti-CXORF21 antibody  
709 (Atlas antibodies; HPA001185) and secondary goat anti-rabbit Alexa Fluor 488 (Abcam;  
710 ab150077) at 1:2000. Cells were then incubated for 60min on ice with 0.1µg anti-GM130-  
711 Alexa Fluor 647 (Abcam; ab195303). Lysosomal and nuclear colocalisation:  $2 \times 10^6$  cells were

712 incubated at 37°C for 15mins in 1X Assay Buffer and 0.1µl Lyso-ID Red Detection Reagent  
713 and 0.1µl Hoechst 33342 Nuclear Stain (Lyso-ID Red Detection Kit; Enzo; anti-ENZ-51005-  
714 0100). Cells were then fixed with 200ul 1X stabilising fixative (BD biosciences) and  
715 permeabilized in 0.1% Triton X-100 (Sigma-Aldrich). Fc blocker was added before  
716 intracellular staining with 0.1µg rabbit polyclonal anti-CXORF21 antibody (Atlas antibodies;  
717 HPA001185) and secondary goat anti-rabbit Alexa Fluor 488 (Abcam; ab150077) at 1:2000.  
718 Cells were resuspended in 60µl PBS. Up to 100,000 images were acquired per sample. Cells  
719 were gated on aspect ratio to include only singlets, and the gradient root-mean-square feature  
720 to include focused cells. Using the co-localisation mask on the IDEAS software (Amnis), we  
721 calculated the overlap of CXORF21 and organelle markers for cellular localisation.

722

### 723 **Immunostaining of autophagic LC3-II and CXORF21**

724  $1 \times 10^6$  cells/ml LCL were starved in EBSS with or without 10 nM or 100 nM bafilomycin A1  
725 (Sigma) for 3 h at 37°C and 5% CO<sub>2</sub> before harvesting, when starvation was required. For LC3  
726 staining, the cells were selectively permeabilized with 0.05% saponin prior to fixation. Cells  
727 were fixed in 4% formaldehyde for 20 min at room temperature, then permeabilized with 0.1%  
728 Triton X-100 and 2% goat serum (both Sigma-Aldrich) in PBS for 30 min on ice. After  
729 overnight incubation in 5% goat serum, cells were Fc receptor blocked (Human TruStain FcX,  
730 Biolegend) and incubated with 2 µg/ml rabbit anti-human CXORF21 (Atlas antibodies;  
731 HPA001185) and either 2 µg/ml mouse anti-human TLR7 (Novus Biologicals, NBP2-27332)  
732 or 40 µg/ml mouse anti-human LC3 (MBL, M152-3) in 5% goat serum for 1 h on ice.  
733 Following washing, cells were stained with goat anti-rabbit Alexa Fluor 488 (Abcam;  
734 ab150077) and goat anti-mouse Alexa Fluor 594 (Abcam; ab150116), both at 1:2000, in 5%  
735 goat serum for 30 min on ice. Cells were washed and mounted in ProLong™ Gold Antifade  
736 Mountant containing DAPI (Invitrogen).

737

### 738 **Imaging and Analysis**

739 Imaging was performed at the Nikon Imaging Centre at King's College London. Z stacks were  
740 acquired at 0.12  $\mu\text{m}$  step size on an Eclipse Ti-2 Inverted microscope with Vt-iSIM scan head  
741 and Hamamatsu Flash4.0 sCMOS camera using a 100x oil immersion objective. Laser settings,  
742 image capture and Richardson-Lucy deconvolution were managed in NIS-Elements. Images  
743 were further processed and Pearson's correlation coefficient and Mander's colocalisation  
744 coefficient were calculated using the Colocalization Studio plugin<sup>73</sup> in Icy software. Maximum  
745 intensity projections are shown for better visualisation. A one-way ANOVA with Tukey  
746 multiple comparison correction was performed to test for statistical significance in GraphPad  
747 Prism v7.04.

748

### 749 **Data availability**

750 Summary statistics on 10,995 individuals of matched European ancestry (4,036 SLE cases,  
751 6,959 controls) genotyped on the Illumina HumanOmni1 BeadChip are available at  
752 <http://insidegen.com/insidegen-LUPUS-data.html>. TwinsUK RNASeq data are deposited in  
753 European Genome-Phenome Archive (EGAS00001000805). The UK10K (REL-2012-06-02)  
754 plus 1000 Genomes Project Phase3 data (release 20131101.v5) merged reference panel  
755 (UK10K-1000GP3) was accessed through the European Genome-phenome Archive  
756 (EGAD00001000776). All other data are contained within the article and its supplementary  
757 information or upon reasonable request from the corresponding author. The source data  
758 underlying Figures 2d, 6e-f, 7e-g, and Supplementary Figures 2b, 3, 6b, 12 and 13 are provided  
759 as a Source Data file.



761 **Acknowledgements**

762 We thank all volunteers for their contribution to this research. The work was funded by the  
763 Medical Research Council L002604/1, Arthritis Research UK (project grants 20265 and 20580,  
764 and PhD studentship 21252 to S.V. and studentship 19983 to C.S.T.D.). K.S.S. receives support  
765 from MRC grants MR/R023131/1 and MR/M004422/1. The TwinsUK study was funded by  
766 the Wellcome Trust and European Community's Seventh Framework Programme (FP7/2007-  
767 2013). The TwinsUK study also receives support from the National Institute for Health  
768 Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research  
769 Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's  
770 College London. The NIHR funded the Biomedical Research Centre (BRC) based at Guy's  
771 and St Thomas' NHS Foundation Trust and King's College London. The NIHR funded the  
772 BRC Flow Cytometry Core, Guy's Hospital at Guy's and St Thomas' NHS Foundation Trust  
773 in partnership with King's College London. We acknowledge Julian Knight and Ben Fairfax  
774 for providing X chromosome data on the expression datasets (references 42 and 43). We thank  
775 the Nikon Imaging Centre at King's College London for help with light microscopy.

776

777 **Author Contributions**

778 C.A.O. performed expression experiments, analysed gene expression and epigenetic data,  
779 performed genetic analysis, wrote the manuscript; A.L.R. performed gene expression and  
780 Image Stream experiments, performed genetic analysis, analysed data, wrote the manuscript;  
781 S.K.V. performed super resolution microscopy experiments, analysed data; C.S.T.D.  
782 performed gene expression experiments in B cells and analysed these data; S.K.V. and  
783 C.S.T.D. contributed equally; C.T.B. performed microscopy experiments, analysed data;  
784 A.J.C. designed the microscopy experiments and analysed these data; S.L. performed  
785 CXORF21 expression studies in SLE patients; S.D. analysed CXORF21 protein expression;  
786 L.C. analysed gene expression data; D.L.M. analysed genetic data; L.J. validated the anti-  
787 CXORF21 antibody; L.B. performed gene expression studies in resting and IFN-stimulated B  
788 cells; A.Z. and K.S.S. analysed X chromosome skewing and expression of *CXorf21* in  
789 TwinsUK; M.M.A.F. designed and supervised the B cell expression studies; D.S.C.G. designed  
790 and supervised the study; T.J.V. designed and supervised the study and contributed to writing  
791 the manuscript.

792

793 **Competing Interests**

794 The authors declare no competing interests .

795

796 **References**

- 797 1. Spolarics, Z. The X-Files of inflammation: Cellular Mosaicism of X-linked  
798 polymorphic genes and the female advantage in the host response to injury and  
799 infection.. *Shock* **27**, 597–604 (2007).
- 800 2. Gezon, H. M., Rogers, K. D., Yee, R. B. & Hatch, T. F. Excess risk of staphylococcal  
801 infection and disease in newborn males 1. **84**, 314–328 (2000).
- 802 3. Libert, C., Dejager, L. & Pinheiro, I. The X chromosome in immune functions: when a  
803 chromosome makes the difference. *Nat. Rev. Immunol.* **10**, 594–604 (2010).
- 804 4. Ramos, P. S. *et al.* A comprehensive analysis of shared loci between systemic lupus  
805 erythematosus (SLE) and sixteen autoimmune diseases reveals limited genetic overlap.  
806 *PLoS Genet.* **7**, e1002406 (2011).
- 807 5. Scofield, R. H. *et al.* Klinefelter’s syndrome (47,XXY) in male systemic lupus  
808 erythematosus patients: support for the notion of a gene-dose effect from the X  
809 chromosome. *Arthritis Rheum.* **58**, 2511–7 (2008).
- 810 6. Cooney, C. M. *et al.* 46,X,del(X)(q13) Turner’s syndrome women with systemic lupus  
811 erythematosus in a pedigree multiplex for SLE. *Genes Immun.* **10**, 478–81 (2009).
- 812 7. Liu, K. *et al.* X Chromosome Dose and Sex Bias in Autoimmune Diseases: Increased  
813 47,XXX in Systemic Lupus Erythematosus and Sjögren’s Syndrome. *Arthritis Rheum.*  
814 **68**, 1290–1300 (2016).
- 815 8. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–  
816 337 (2005).
- 817 9. Lyon, M. F. Gene Action in the X-chromosome of the Mouse. *Nature* **190**, 372–373  
818 (1961).
- 819 10. Payer, B. & Lee, J. T. X Chromosome Dosage Compensation: How Mammals Keep



- 820 the Balance. *Annu. Rev. Genet.* **42**, 733–772 (2008).
- 821 11. Carrel, L. & Willard, H. F. X-inactivation profile reveals extensive variability in X-  
822 linked gene expression in females. *Nature* **434**, 400–404 (2005).
- 823 12. Carrel, L., Cottle, A. A., Goglin, K. C. & Willard, H. F. A first-generation X-  
824 inactivation profile of the human X chromosome. *Proc. Natl. Acad. Sci. U. S. A.* **96**,  
825 14440–4 (1999).
- 826 13. Lyon, M. F. *X-chromosome Inactivation and Disease*. *Encyclopedia of Life Sciences*  
827 (2007). doi:10.1038/npg.els.0005480
- 828 14. Wang, J. *et al.* Unusual maintenance of X chromosome inactivation predisposes  
829 female lymphocytes for increased expression from the inactive X. *Proc. Natl. Acad.*  
830 *Sci.* **113**, E2029–E2038 (2016).
- 831 15. Bentham, J. *et al.* Genetic association analyses implicate aberrant regulation of innate  
832 and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus.  
833 *Nature Genetics* **47**, (2015).
- 834 16. Zhang, Y. *et al.* Genes That Escape X-Inactivation in Humans Have High Intraspecific  
835 Variability in Expression, Are Associated with Mental Impairment but Are Not Slow  
836 Evolving. *Mol. Biol. Evol.* **30**, 2588–2601 (2013).
- 837 17. Vawter, M. P., Harvey, P. D. & DeLisi, L. E. Dysregulation of X-linked gene  
838 expression in Klinefelter’s syndrome and association with verbal cognition. *Am. J.*  
839 *Med. Genet. Part B Neuropsychiatr. Genet.* **144**, 728–734 (2007).
- 840 18. Mackay, M. *et al.* Molecular signatures in systemic lupus erythematosus: distinction  
841 between disease flare and infection. *Lupus Sci. Med.* **3**, e000159 (2016).
- 842 19. Yen, E. Y. & Singh, R. R. Brief Report: Lupus—An Unrecognized Leading Cause of  
843 Death in Young Females: A Population-Based Study Using Nationwide Death  
844 Certificates, 2000–2015. *Arthritis Rheumatol.* **70**, 1251–1255 (2018).

- 845 20. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease.  
846 *Nature* **526**, 82–90 (2015).
- 847 21. Alfonso Buil, Andrew Anand Brown, Tuuli Lappalainen, Ana Viñuela, Matthew N  
848 Davies, Hou-Feng Zheng, J Brent Richards, Daniel Glass, Kerrin S Small, Richard  
849 Durbin, T. D. S. & E. T. D. Gene-gene and gene-environment interactions detected by  
850 transcriptome sequence analysis in twins. *Nat. Genet.* **47**, 88–91 (2015).
- 851 22. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human  
852 epigenomes. *Nature* **518**, 317–330 (2015).
- 853 23. Martens, J. H. A. & Stunnenberg, H. G. BLUEPRINT: Mapping human blood cell  
854 epigenomes. *Haematologica* **98**, 1487–1489 (2013).
- 855 24. Wong, N. & Wang, X. miRDB: An online resource for microRNA target prediction  
856 and functional annotations. *Nucleic Acids Res.* **43**, D146–D152 (2015).
- 857 25. Schofield, E. C. *et al.* CHiCP: A web-based tool for the integrative and interactive  
858 visualization of promoter capture Hi-C datasets. *Bioinformatics* **32**, 2511–2513 (2016).
- 859 26. The ENCODE Project Consortium *et al.* An integrated encyclopedia of DNA elements  
860 in the human genome. *Nature* **489**, 57–74 (2012).
- 861 27. The GTEx Consortium *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat.*  
862 *Genet.* **45**, 580–585 (2013).
- 863 28. Hussain, T. & Mulherkar, R. Lymphoblastoid Cell lines: a Continuous in Vitro Source  
864 of Cells to Study Carcinogen Sensitivity and DNA Repair. *Int. J. Mol. Cell. Med.* **1**,  
865 75–87 (2012).
- 866 29. Bezalel, S., Guri, K. M., Elbirt, D., Asher, I. & Sthoeger, Z. M. Type I interferon  
867 signature in systemic lupus erythematosus. *Isr. Med. Assoc. J.* **16**, 246–9 (2014).
- 868 30. Okamura, Y. *et al.* COXPRESdb in 2015 : coexpression database for animal species by  
869 DNA-microarray and RNAseq-based expression data with multiple quality assessment

- 870 systems. **43**, 82–86 (2015).
- 871 31. Lassen, K. G. *et al.* Genetic Coding Variant in GPR65 Alters Lysosomal pH and Links  
872 Lysosomal Dysfunction with Colitis Risk. *Immunity* **44**, 1392–1405 (2016).
- 873 32. Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human  
874 Interactome. *Cell* **162**, 425–440 (2015).
- 875 33. Kobayashi, T. *et al.* The histidine transporter SLC15A4 coordinates mTOR-dependent  
876 inflammatory responses and pathogenic antibody production. *Immunity* **41**, 375–388  
877 (2014).
- 878 34. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the Web: a case  
879 study using the Phyre server. *Nat. Protoc.* **4**, 363–371 (2009).
- 880 35. Clarke, A. J. *et al.* Autophagy is activated in systemic lupus erythematosus and  
881 required for plasmablast development. *Ann. Rheum. Dis.* 1–9 (2014).  
882 doi:10.1136/annrheumdis-2013-204343
- 883 36. Wise, A. L., Gyi, L. & Manolio, T. A. EXclusion: Toward integrating the X  
884 chromosome in genome-wide association analyses. *Am. J. Hum. Genet.* **92**, 643–647  
885 (2013).
- 886 37. Kwon, K.-S., Cho, H.-Y. & Chung, Y.-J. Recapitulation of Candidate Systemic Lupus  
887 Erythematosus-Associated Variants in Koreans. *Genomics Inform.* **14**, 85 (2016).
- 888 38. Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease  
889 variants. *Nature* **518**, 337–343 (2015).
- 890 39. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping  
891 complex trait variants. *Nat. Genet.* **45**, 124–30 (2013).
- 892 40. Odhams, C. A. *et al.* Mapping eQTLs with RNA-Seq Reveals Novel Susceptibility  
893 Genes, Non-Coding RNAs, and Alternative-Splicing Events in Systemic Lupus  
894 Erythematosus. *Hum. Mol. Genet.* **0**, ddw417 (2017).

- 895 41. Escoubet-Lozach, L. *et al.* Mechanisms establishing TLR4-responsive activation states  
896 of inflammatory response genes. *PLoS Genet.* **7**, (2011).
- 897 42. Jiang, W., Zhang, L., Lang, R., Li, Z. & Gilkeson, G. Sex differences in monocyte  
898 activation in Systemic Lupus Erythematosus (SLE). *PLoS One* **9**, 1–17 (2014).
- 899 43. Suurmond, J. *et al.* Repeated Fc $\gamma$ RI triggering reveals modified mast cell function  
900 related to chronic allergic responses in tissue. *J. Allergy Clin. Immunol.* **138**, 869–880  
901 (2016).
- 902 44. Blasius, A. L. & Beutler, B. Intracellular Toll-like Receptors. *Immunity* **32**, 305–315  
903 (2010).
- 904
- 905
- 906
- 907
- 908
- 909

## 910 **Figure Legends**

### 911 **Figure 1: Genetic refinement of the Xp21.2 (rs887369) SLE susceptibility locus.**

912 (A) Association plot of the 1Mb region (X:30,077,846-31,077,845) of SLE-associated region  
913 Xp21.2 following genotype imputation to the level of UK10K-1000G Phase III and association  
914 testing as described (n = 10,995 individuals of European ancestry). rs887369 is shown as the  
915 most significantly associated lead SNP. Genetic association plots were generated using  
916 LocusZoom. (B) Association plot of the 1Mb region following conditional analysis on lead  
917 SNP rs887369. (C) Haplotype construction and visualisation of the Xp21.2 SLE susceptibility  
918 locus conducted in Haploview 4.2. The top panel shows the structure of the three blocks and  
919 haplotypes surrounding the lead SNP rs887369 (highlighted in red, block B, SNP #15). Blocks  
920 are separated by regions of high recombination as specified by  $D'$  and  $r^2$ . The frequency of  
921 each haplotype is denoted. The middle panel presents the colour-coded haplotypes and  
922 individual SNPs by their genomic coordinates around *CXorf21*. The bottom panel shows the  
923 LD structure and pair-wise correlation ( $r^2$ ) of SNPs, and length of each block. (D) Right table:  
924 case-control association analysis of each haplotype using Haploview 4.2.

925

### 926 **Figure 2: eQTL association analysis of SLE associated risk SNP rs887369 in immune cell** 927 **types.**

928 (A) *Cis*-eQTL analysis of rs887369 using male samples of the Geuvadis RNA-Seq expression  
929 cohort profiled in LCLs. The MAGEB family of genes and *NR0B1* were not expressed in LCLs  
930 (RPKM < 1). Allele [C] of rs887369 tags the risk haplotype. The number underneath each box-  
931 plot represents the mean of the group and the number underneath the x-axis refers to the number  
932 of individuals in each group. (B) *Cis*-eQTL analysis performed for all SNPs in *cis* (+/-1Mb) to

933 rs887369 against *CXorf21* expression using the males of the Geuvaris cohort. The coordinate  
934 of each SNP is plotted on the x-axis and the  $-\log_{10}(P)$  value of association on the y-axis;  
935 rs887369 is highlighted as the best eQTL. (C) *cis*-eQTL analysis of rs887369 against *CXorf21*  
936 expression in LCLs from the TwinsUK cohort using only females who exhibit non-skewed  
937 patterns of X-chromosome inactivation (see methods). (D) Relative protein abundance of  
938 CXORF21 in LCLs from females stratified on genotype at the rs887369 SNP. Relative  
939 abundance normalized against beta-actin loading control. Source data are provided in the  
940 Source Data file (E) *Cis*-eQTL analysis of rs887369 using the microarray data from the *Fairfax*  
941 *et al*<sup>45,46</sup> and *Naranbhai et al*<sup>47</sup> cohorts in primary *ex vivo* immune cell types (see Methods).  
942 The remaining *cis*-genes did not pass quality control. Box-plots show minimum (Q1-1.5\*IQR),  
943 25<sup>th</sup> percentile (Q1), Median, 75<sup>th</sup> percentile (Q3), and maximum Q3+1.5\*IQR.

944

945 **Figure 3: Functional prioritization of causal variants at the Xp21.2 SLE susceptibility locus.**

946 The five SNPs carried on the risk haplotype attributed to SLE susceptibility and modulation of  
947 *CXorf21* gene expression were epigenetically fine-mapped using chromatin data from the  
948 Roadmap Epigenomes Project (twelve different marks across 127 cell/tissue types). (A) The  
949 five SNPs localised to significant H3K36me3 modification sites in five immune cell types. The  
950 heatmap shows the fold-enrichment of H3K36me3 between cell-types across SNP positions.  
951 (B) Signal tracks of H3K36me3 in primary monocytes (blue) and primary neutrophils (red)  
952 from peripheral blood across the *CXorf21* susceptibility locus. Only rs887369 localises to the  
953 binding site summit of H3K36me3 in these two cell types. (C) Promoter-capture Hi-C  
954 interaction of the rs887369 target locus (chrX :30576528-30582605) with four bait loci across  
955 17 primary immune cell types from healthy human donors (Note that the majority of the  
956 samples are pooled from multiple donors making it impossible to deconvolute the sex and

957 genotypes of the individuals). Interaction #1 is the interaction between the association target  
958 region (at the 3' end of *CXorf21*) and the *CXorf21* promoter region. **(D)** Heatmap of strength  
959 of interaction (CHiCAGO score) of the four interactions across immune cell types. **(E)**  
960 Correlation of interaction score for interaction #1 (3' of *CXorf21* and *CXorf21* promoter) with  
961 Roadmap Epigenomes Project chromatin marks found at the *CXorf21* promoter across different  
962 immune cell types. Higher interactions are correlated with greater enrichment of active  
963 chromatin marks suggesting the interaction to regulate gene expression is cell-type specific.  
964

965 **Figure 4: Expression of *CXorf21* in primary *ex vivo* cells stratified on sex and cellular**  
966 **activation.**

967 *CXorf21* expression data in resting and stimulated monocytes from healthy individuals of  
968 European ancestry derived from the Fairfax *et al* studies (see Methods; n = 322 biologically  
969 independent samples). Samples were separated based on sex and activation condition:  
970 following stimulation with interferon gamma (IFN- $\gamma$ ) or with lipopolysaccharide (LPS) and  
971 harvested after 24 hours. For each group, the mean is reported in the corners and the effect size  
972 (Cohen's d) is reported along the corresponding three-dimensional regression plane. Plots were  
973 constructed using plot3D for R. Source data are provided as a Source Data file.  
974

975 **Figure 5: *CXorf21* as an interferon response gene**

976 **(A)** Differential gene expression of X chromosome genes in response to IFN- $\alpha$  stimulation  
977 (harvested after 6 h) in primary *ex vivo* B cells from healthy females of European ancestry (in-  
978 house data). Genes highlighted in red are significantly differentially expressed ( $q < 0.01$ ;  
979 absolute fold-change  $> 2$ ). **(B)** Epigenetic landscape of *CXorf21* using ENCODE transcription

980 factor binding data in LCLs (GM12878 cell line). All five transcription factors have genome-  
981 wide significant binding sites at the *CXorf21* promoter. Heat colour is a function of signal  
982 strength (fold-change over input).

983

#### 984 **Figure 6: Super resolution microscopy of CXORF21 and TLR7**

985 Structured Illumination Microscopy data showing colocalisation of TLR7 and CXORF21 in *ex*  
986 *vivo* B cells. Representative results on individual cells are shown in panels A through D with  
987 TLR7 staining in the first column, CXORF21 in the second column, DAPI nuclear staining in  
988 column three, and in the fourth column all three stains are merged: TLR7 (magenta), CXORF21  
989 (green) and DAPI (blue). The B cells are under different conditions in the panels: (A) resting,  
990 (B) resting and IFN- $\alpha$  treated (1,000U/ml), (C) Ig/CD40 stimulated, and (D) Ig/CD40  
991 stimulated and IFN- $\alpha$  treated *ex vivo* B cells at 20 hours. Maximum intensity projections are  
992 shown. Scale bar in white on bottom left hand corner is 2  $\mu$ m. (E) Plot showing the correlation  
993 co-efficients ( $\rho$ ) between TLR7 and CXORF21 staining of multiple B cells quantified using  
994 the results from Z-stack images from individual cells (represented as open circles). From left  
995 to right: unstimulated cells (n=84), cells stimulated with IFN- $\alpha$  (n=60), B cells stimulated with  
996 Ig/CD40 (n=32), B cells stimulated with Ig/CD40 and IFN- $\alpha$  (n=22). The horizontal bar  
997 represents the mean correlation co-efficient ( $\mu^{\rho}$ ) and the bars above and below this denote the  
998 standard deviation of the distribution. (F) Mander's colocalisation coefficient ( $M^2$ ) between  
999 TLR7 and CXORF21 are shown from Z-stack images from single B cells (represented as open  
1000 circles). From left to right: unstimulated cells (n=84), cells stimulated with IFN- $\alpha$  (n=60), B  
1001 cells stimulated with Ig/CD40 (n=32), B cells stimulated with Ig/CD40 and IFN- $\alpha$  (n=22). The  
1002 horizontal bar represents the mean colocalisation co-efficient ( $\mu^{M^2}$ ) and the bars above and



1003 below this denote the standard deviation of the distribution. Source data are provided as a  
1004 Source Data file.

1005

1006 **Figure 7: CXORF21 and the autophagosome**

1007 Structured Illumination Microscopy data showing colocalisation of LC3 and CXORF21 in *ex*  
1008 *vivo* B cells. Representative results on individual cells are shown in panels A through D with  
1009 LC3 staining in the first column, CXORF21 in the second column, DAPI nuclear staining in  
1010 the third column, and in the fourth column all three stains are merged: LC3 (magenta),  
1011 CXORF21 (green) and DAPI (blue). In panel (A) B cells were Ig/CD40 stimulated, (B)  
1012 Ig/CD40 stimulated and bafilomycin-treated, (C) Ig and TLR7/8 stimulated, and (D) Ig and  
1013 TLR7/8 stimulated and bafilomycin-treated after 27 hours. Maximum intensity projections are  
1014 shown. Scale bar in white on bottom left hand corner is 2  $\mu\text{m}$ . (E) Plot showing the correlation  
1015 co-efficients ( $\rho$ ) between LC3 and CXORF21 staining quantified using the results from Z-stack  
1016 images, individual B cells are represented as open circles. From left to right: Ig/CD40  
1017 stimulated cells (n=17), Ig/CD40 stimulated cells in the presence of 10nM bafilomycin (n=22),  
1018 B cells stimulated with Ig and resiquimod (n=21), B cells stimulated with Ig and resiquimod  
1019 in the presence of 10nM bafilomycin (n=32). The horizontal bar represents the mean  
1020 correlation co-efficient ( $\mu^{\rho}$ ) and the bars above and below this horizontal bar denote the  
1021 standard deviation of the distribution. (F) Mander's colocalisation coefficient ( $M2$ ) between  
1022 LC3 and CXORF21 are shown from Z-stack images, individual B cells are represented as open  
1023 circles. From left to right: Ig/CD40 stimulated cells (n=17), Ig/CD40 stimulated cells in the  
1024 presence of 10nM bafilomycin (n=23), B cells stimulated with Ig and resiquimod (n=21), B  
1025 cells stimulated with Ig and resiquimod in the presence of 10nM bafilomycin (n=32). The  
1026 horizontal bar represents the mean colocalisation co-efficient ( $\mu^{M2}$ ) and the bars above and

1027 below this denote the standard deviation of the distribution. (G) Western blot analysis of  
1028 protein extract from starved LCL, in the left-hand blot CXORF21 is quantified in the absence  
1029 of bafilomycin and after 10nM and 100nM treatment. The amount of CXORF21 was quantified  
1030 by densitometry and the relative abundance shown against a beta actin control, using the  
1031 unstimulated conditions as a reference point. In the right-hand blot sequestosome 1 (p62) is  
1032 quantified in the absence of bafilomycin and after 10nM and 100nM treatment. The amount of  
1033 Sequestosome-1 was quantified by densitometry and the relative abundance shown against a  
1034 beta actin control, using the unstimulated conditions as a reference point. Source data are  
1035 provided as a Source Data file.

1036

1037 **Figure 8: Summary of factors influencing expression of *CXorf21* at RNA and protein**  
1038 **level.**

1039 We summarise five factors increasing the cellular abundance of *CXorf21* either at RNA level  
1040 or protein level across a range of immune cell types. These are: (1) genetic variation at SLE  
1041 susceptibility haplotype - tagged by SNP rs887369 - where the risk haplotype [C] may drive  
1042 up-regulation by modulation of chromatin interaction and/or modification of H3K36me3 state;  
1043 (2) female sex, in which escape from X-inactivation results in an increased amount of transcript  
1044 and protein in females; (3) X chromosome aneuploidy; (4) type I and type II interferons, and  
1045 LPS, increase the expression of *CXorf21* in *ex vivo* B cells and monocytes; (5) ancestry –  
1046 potentially linked to the minor allele frequency of rs887369 – in which higher levels of *CXorf21*  
1047 transcript is observed in LCLs derived from donors with European ancestry. We hypothesise  
1048 that elevation of CXORF21 is a risk factor for developing SLE and that this is may be mediated  
1049 through it's role in the endosomal pathway. Figure generated by C.A.O.

1050

1051 **Methods References**

- 1052 45. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell  
1053 type-specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510  
1054 (2012).
- 1055 46. Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants  
1056 upon monocyte gene expression. *Science* **343**, 1246949 (2014).
- 1057 47. Naranbhai, V. *et al.* Genomic modulators of gene expression in human neutrophils.  
1058 *Nat. Commun.* **6**, 7545 (2015).
- 1059 48. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and  
1060 accurate genotype imputation in genome-wide association studies through pre-phasing.  
1061 *Nat. Genet.* **44**, 955–959 (2012).
- 1062 49. Roshyara, N. R. *et al.* Comparing performance of modern genotype imputation  
1063 methods in different ethnicities. *Sci. Rep.* **6**, 34386 (2016).
- 1064 50. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies.  
1065 *Nat. Rev. Genet.* **11**, 499–511 (2010).
- 1066 51. Marchini, J., Howie, B. N., Myers, S., McVean, G. & Donnelly, P. A new multipoint  
1067 method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*  
1068 **39**, 906–13 (2007).
- 1069 52. Pruim, R. J. *et al.* LocusZoom: Regional visualization of genome-wide association  
1070 scan results. *Bioinformatics* **26**, 2336–2337 (2010).
- 1071 53. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: Analysis and visualization  
50

- 1072 of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
- 1073 54. Gabriel SB *et al.* The structure of haplotype blocks in the human genome. *Science* (80-  
1074 .). **296**, 2225–2229 (2002).
- 1075 55. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and  
1076 richer datasets. 1–16 (2014). doi:10.1186/s13742-015-0047-8
- 1077 56. Kent, W. J., Sugnet, C. W., Furey, T. S. & Roskin, K. M. The Human Genome  
1078 Browser at UCSC *W. J. Med. Chem.* **19**, 1228–31 (2002).
- 1079 57. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional  
1080 variation in humans. *Nature* **501**, 506–11 (2013).
- 1081 58. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74  
1082 (2015).
- 1083 59. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations.  
1084 *Bioinformatics* **28**, 1353–1358 (2012).
- 1085 60. Naumova, A. K. *et al.* Heritability of X chromosome-inactivation phenotype in a large  
1086 family. *Am. J. Hum. Genet.* **58**, 1111–1119 (1996).
- 1087 61. Amos-Landgraf, J. M. *et al.* X Chromosome–Inactivation Patterns of 1,005  
1088 Phenotypically Unaffected Females. *Am. J. Hum. Genet.* **79**, 493–499 (2006).
- 1089 62. Kristiansen, M. *et al.* Twin study of genetic and aging effects on X chromosome  
1090 inactivation. *Eur. J. Hum. Genet.* **13**, 599–606 (2005).
- 1091 63. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation

- 1092 of expression residuals (PEER) to obtain increased power and interpretability of gene  
1093 expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).
- 1094 64. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer  
1095 (IGV): high-performance genomics data visualization and exploration. *Brief.*  
1096 *Bioinform.* **14**, 178–192 (2013).
- 1097 65. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and  
1098 Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369–1384.e19  
1099 (2016).
- 1100 66. Mifsud, B. *et al.* Sup Mapping long-range promoter contacts in human cells with high-  
1101 resolution capture Hi-C. *Nat. Genet.* **47**, 598–606 (2015).
- 1102 67. Martin, P. *et al.* Capture Hi-C reveals novel candidate genes and complex long-range  
1103 interactions with related autoimmune risk loci. *Nat. Commun.* **6**, 1–17 (2015).
- 1104 68. Cairns, J. *et al.* CHiCAGO: Robust Detection of DNA Looping Interactions in Capture  
1105 Hi-C data. *Genome Biol.* **17**, 28068 (2016).
- 1106 69. Lockstone, H. E. Exon array data analysis using Affymetrix power tools and R  
1107 statistical software. *Brief. Bioinform.* **12**, 634–644 (2011).
- 1108 70. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The SVA  
1109 package for removing batch effects and other unwanted variation in high-throughput  
1110 experiments. *Bioinformatics* **28**, 882–883 (2012).
- 1111 71. Tan, E. M. *et al.* The 1982 revised criteria for the classification of systemic lupus  
1112 erythematosus. *Arthritis Rheum.* **25**, 1271–1277 (1982).

- 1113 72. Bombardier, C., Gladman, D. D., Urowitz, M. B., Caron, D. & Chang, C. H.  
1114 Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee  
1115 on Prognosis Studies in SLE. *Arthritis Rheum.* **35**, 630–640 (1992).
- 1116 73. Lagache, T., Sauvonnet, N., Danglot, L. & Olivo-Marin, J. C. Statistical analysis of  
1117 molecule colocalization in bioimaging. *Cytom. Part A* **87**, 568–579 (2015).
- 1118