Mapping eQTLs With RNA-Seq Reveals Novel SLE Susceptibility Genes, Non-Coding RNAs, and Alternative-Splicing Events That Are Concealed Using Microarrays

Christopher A. Odhams[1*], Andrea Cortini[1], Lingyan Chen[1], Amy L. Roberts[1], Alfonso Buil[2-3], Kerrin S. Small[4], Emmanouil T. Dermitzakis[2-3], David L. Morris[1], Timothy J. Vyse[1, 5*], Deborah S. Cunninghame Graham*[1, 5]

[1]Division of Genetics and Molecular Medicine, King's College London, London, SE1 9RT, UK. [2]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, CH – 1211 Geneva 4, Switzerland.

[3]Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, CH – 1211 Geneva 4 Switzerland.

[4]Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK.

[5]Division of Immunology, Infection and Inflammatory Disease, King's College London, London, SE1 1UL, UK.

*Correspondence: Dr Deborah S. Cunninghame Graham

Fax: +44 (0)207 188 2585

Telephone: +44 (0)207 848 8504

Email: deborah.cunninghame-graham@kcl.ac.uk

Abstract

Studies attempting to functionally interpret complex-disease susceptibility loci by GWAS and eQTL integration have predominantly employed microarrays to quantify gene-expression. RNA-Seq has the potential to discover a more comprehensive set of eQTLs and illuminate the underlying molecular consequence. We examine the functional outcome of 39 variants associated with Systemic Lupus Erythematosus (SLE) through integration of GWAS and eQTL data from the TwinsUK microarray and RNA-Seq cohort in lymphoblastoid cell lines. We use conditional analysis and a Bayesian colocalisation method to provide evidence of a shared causal-variant, then compare the ability of each quantification type to detect disease relevant eQTLs and eGenes. We discovered a greater frequency of candidate-causal eQTLs using RNA-Seq, and identified novel SLE susceptibility genes that were concealed using microarrays (e.g. *NADSYN1*, *SKP1*, and *TCF7*). Many of these eQTLs were found to influence the expression of several genes, suggesting risk haplotypes may harbour multiple functional effects. We pinpointed eQTLs modulating expression of four non-coding RNAs; three of which were replicated in whole-blood. Novel SLE associated splicing events were identified in the T-reg restricted transcription factor, *IKZF2*, the autophagy-related gene *WDFY4*, and the redox coenzyme *NADSYN1*, through asQTL mapping using the Geuvadis cohort. We have significantly increased our understanding of the genetic control of gene-expression in SLE by maximising the leverage of RNA-Seq and performing integrative GWAS-eQTL analysis against gene, exon, and splice-junction quantifications. In doing so, we have identified novel SLE candidate genes and specific molecular mechanisms that will serve as the basis for targeted follow-up studies.

Introduction

Genome-Wide Association Studies (GWAS) have successfully identified a large number of genetic loci that contribute to complex-disease susceptibility in humans (1). Evidence suggests these variants are enriched within regulatory elements of the genome and their effects play a central role in modulation of intermediate quantitative phenotypes such gene expression (1–6). Many expression quantitative trait loci (eQTL) mapping studies have since been conducted across a wide-range of ethnicities (7, 8), cell-types (9–16), disease states (17–22) and in response to various environmental stimuli (23, 24) - with each contributing to our understanding of the architecture of human regulatory variation in complex-disease.

In spite of diverse study designs, a significant constraint on the majority of such investigations is the use of 3´-targeted microarrays to profile gene expression. The effects of splicing are less likely to be detected through quantification of pre-defined probes that target common exons of a gene (25) and may explain why only a limited number of susceptibility loci localize to causal eQTL signals (26, 27). Technical limitations of microarrays and noise from the small probe design of exon-arrays, further hinder the accuracy of expression measurements (25, 28–30). RNA-Seq based eQTL mapping studies are beginning to emerge (31, 32) and, although large-scale analysis pipelines are still being streamlined, such types of investigations will greatly increase the likelihood of capturing disease associated eQTLs as quantification of overall gene and independent exon expression, and relative transcript abundance (including novel isoforms and non-coding RNAs) is possible (33–39).

Integrative studies using RNA-Seq to functionally annotate complex-disease susceptibility loci however have been limited (35, 40–44). Moreover, numerous investigations have aimed to explain the functional relevance of susceptibility loci by interrogation of GWAS SNPs themselves in eQTL datasets and simply testing for association with gene expression (45–47). Such inferential observations should be treated with caution as they may possibly be the

result of coincidental overlap between disease association and eQTL signal due to local LD and general ubiquity of regulatory variants (48). This has become particularly important as statistical power in eQTL cohorts grow and availability of summary-level data accession through eQTL data-browsers increases (49–51).

In this investigation, we integrate eQTL data derived from both microarray and RNA-Seq experiments with our GWAS results in Systemic Lupus Erythematosus (SLE [MIM: 152700]); a heritable autoimmune disease with undefined aetiology and over 50 genetically associated loci (52–54). We use summary-level *cis*-eQTL results in lymphoblastoid cell lines (LCLs) taken from the TwinsUK cohort to directly compare the microarray (9) and RNA-Seq (39) results in detecting SLE associated eQTLs along with their accompanying eGenes. We apply a rigorous two-step approach – a combination of conditional (55) and Bayesian colocalisation (56) analysis – to test for a shared causal variant at each locus. We demonstrate the benefits of using RNA-Seq over microarrays in eQTL analysis by identifying not only novel SLE candidate-causal eGenes but also putative molecular mechanisms by which SLE-associated SNPs may act; including differential exon usage, and expression modulation of non-coding RNA. Our investigation was extended to include RNA-Seq expression data in whole blood in order to validate the eQTL signals detected in LCLs and uncover the differences in genetic control of expression between cell-types. Finally, we interrogate the Geuvadis RNA-Seq cohort (35) to identify SLE associated alternative-splicing quantitative trait loci (asQTLs) and highlight the advantages of profiling at various resolutions to detect eQTLs that would otherwise remain concealed. Through functional annotation of SLE associated loci using microarray and RNA-Seq derived expression data, we have supplied comprehensive evidence of the need to use RNA-Seq to detect disease contributing eQTLs and, in doing so, have suggested novel functional mechanisms that serve as a basis for future targeted follow-up studies.

Results

Discovery and classification of SLE candidate-causal eQTLs and eGenes

The first part of this study integrated the 39 SLE associated SNPs taken from a recent GWAS in Europeans (**Table 1**) with eQTLs from the TwinsUK gene-expression cohort (n=856) profiled using microarray and RNA-Seq (at both gene-level and exon-level resolutions). To accomplish this, we implemented a two-step pipeline (**Fig. 1**), and subjected the genomic intervals within +/-1Mb of each of the 39 GWAS SNPs to eQTL association analysis against expression quantifications in LCLs (**Table 2**).

Full results of the conditional and colocalisation analysis for each significant association are presented in **S1 Table**, **S2 Table**, and **S3 Table** for microarray, RNA-Seq (gene-level), and RNA-Seq (exon-level), respectively. Statistically significant SLE-associated *cis*-eQTLs showing evidence of a shared causal variant or in very strong LD and close ranking between the disease and *cis*-eQTL signal following conditional and colocalisation analyses were classified as SLE candidate-causal eQTLs as stated in Methods. SLE candidate-causal eGenes were defined as genes whose expression is modulated by the eQTL. The final column of **S1-S3 Tables** indicates whether each GWAS SNP is deemed to be candidate-causal. These SLE candidate-causal eQTLs and eGenes are presented in separate tables contingent on the dataset from which they were generated: results from microarray assessment are listed in **Table 3**, from RNA-Seq (gene-level) in **Table 4**, and from RNA-Seq (exon-level) in **Table 5**. Effect sizes are with respect to the minor allele; risk alleles are highlighted in **Table 1**. Overall, exon-level analysis was the most effective quantification type for the discovery of eQTLs and eGenes compared with gene-level RNA-Seq or microarray analysis following an FDR cut-off of $q < 0.05$ and conditional and colocalisation thresholding as described.

Table 1

Independent allelic associations at SLE susceptibility loci following meta-analysis with replication study

GWAS SNP (Tag SNP)	GWAS SNP Position (hg19)	Variant Alleles	Odds Ratio (CI)	*P*-value	GENCODE v10 Annotation	Annotated eGenes from microarray LCLs
rs2476601	1:114377568	G/**A**	1.43 (1.34-1.53)	1.10E-28	*PTPN22* Missense	
rs1801274	1:161479745	**G**/A	1.16 (1.11-1.21)	1.04E-12	*FCGR2A* Missense	*FCGR2B*
rs10912578 (rs844663)	1:173251856	G/**A**	1.27 (1.22-1.33)	4.16E-19	*LOC100506023* Intronic	
rs10753074 (rs1935325)	1:173346343	**T**/C	1.21 (1.15-1.26)	5.82E-12	*LOC100506023* Intronic	
rs17849501	1:183542323	C/**T**	2.10 (1.95-2.26)	3.45E-88	*NCF2* Synonymous	*SMG7*
rs3024505	1:206939904	G/**A**	1.17 (1.11-1.24)	4.64E-09	1kb 3' of *IL10*	
rs9782955	1:236039877	**C**/T	1.16 (1.11-1.22)	1.25E-09	*LYST* Intronic	*LYST*
rs268134	2:65608363	**G**/A	1.21 (1.15-1.27)	1.14E-10	*SPRED* Intronic	*SPRED2*
rs2111485	2:163110536	**G**/A	1.15 (1.11-1.2)	1.27E-11	8.9kb 5' of *FAP*	
rs11889341	2:191943742	C/**T**	1.70 (1.64-1.75)	2.48E-75	*STAT4* Intronic	
rs6736175 (rs16833249)	2:191946322	T/**C**	1.24 (1.19-1.29)	9.17E-17	*STAT4* Intronic	
rs3768792	2:213871709	A/**G**	1.24 (1.17-1.31)	1.21E-13	*IKZF2* 3'-UTR	
rs9311676 (rs11714389)	3:58470351	**C**/T	1.17 (1.13-1.22)	3.06E-14	*LOC101929223* NC	*ABHD6*
rs564799	3:159728987	**C**/T	1.14 (1.09-1.18)	1.54E-09	*IL12A-AS1* Intronic	*IL12A*
rs10028805	4:102737250	**G**/A	1.20 (1.15-1.25)	4.31E-17	*BANK1* Intronic	*BANK1*
rs7726414 (rs17167273)	5:133431834	C/**T**	1.45 (1.32-1.58)	4.44E-16	19kb 5' of *TCF7*	
rs10036748	5:150458146	C/**T**	1.38 (1.32-1.45)	1.27E-45	*TNIP1* Intronic	
rs2431697	5:159879978	**T**/C	1.26 (1.21-1.31)	8.01E-28	15kb 5' of *hsa-mir-146a*	
rs6568431	6:106588806	C/**A**	1.21 (1.15-1.27)	5.04E-14	31kb 3' of *PRDM1*	
rs6932056	6:138242437	T/**C**	1.83 (1.65-2.02)	1.97E-31	22kb 3' of *RP11-10J5.1*	
rs849142	7:28185891	**T**/C	1.14 (1.1-1.19)	8.61E-11	*JAZF1* Intronic	
rs4917014	7:50305863	**T**/G	1.18 (1.13-1.24)	6.39E-14	8.5kb 3' of *AC020743.4*	
rs3757387 (rs4728142)	7:128576086	T/**C**	1.45 (1.4-1.5)	1.14E-48	1.6kb 5' of *IRF5*	
rs35000415 (rs10488631)	7:128585616	C/**T**	1.83 (1.76-1.9)	1.20E-60	*IRF5* Intronic	*IRF5, TNPO3*
rs2736340	8:11343973	C/**T**	1.29 (1.22-1.37)	6.28E-20	7.5kb 5' of *BLK*	*BLK*
rs2663052	10:50069395	**G**/A	1.16 (1.1-1.22)	5.25E-09	*WDFY4* Intronic	*WDFY4*
rs4948496	10:63805617	T/**C**	1.14 (1.1-1.19)	1.04E-10	*ARID5B* Intronic	
rs12802200 (rs2396545)	11:566936	**C**/A	1.23 (1.15-1.31)	8.81E-10	*MIR210HG* NC	
rs2732549	11:35088399	**A**/G	1.24 (1.19-1.29)	1.20E-23	46kb 3' of *PDHX*	
rs3794060	11:71187679	T/**C**	1.23 (1.18-1.29)	1.32E-20	*NADSYN1* Intronic	
rs7941765 (rs6590343)	11:128499000	**C**/T	1.14 (1.1-1.19)	1.35E-10	547bp 3' of *RP11-744N12.3*	
rs10774625	12:111910219	**A**/G	1.13 (1.08 - 1.18)	4.09E-09	*ATXN2* Intronic	
rs1059312	12:129278864	A/**G**	1.17 (1.12-1.21)	1.48E-13	*SLC15A4* Synonymous	
rs4902562	14:68731458	G/**A**	1.14 (1.09-1.19)	6.15E-10	*RAD51B* Intronic	
rs2289583	15:75311036	C/**A**	1.19 (1.14-1.24)	6.22E-15	*SCAMP5* Intronic	*CSK, ULK3, MPI*
rs9652601	16:11174365	**G**/A	1.21 (1.15-1.26)	7.42E-17	*CLEC16A* Intronic	
rs34572943 (rs9936831)	16:31272353	G/**A**	1.71 (1.61-1.81)	3.39E-76	*ITGAM* Intronic	
rs11644034	16:85972612	**G**/A	1.25 (1.19-1.32)	9.58E-18	16kb 3' of *IRF8*	
rs2286672	17:4712617	C/**T**	1.25 (1.16-1.35)	2.93E-09	*PLD2* Missense	*RNF167*
rs2941509	17:37921194	C/**T**	1.35 (1.22-1.49)	7.98E-09	*IKZF3* 3'-UTR	
rs2304256	19:10475652	**C**/A	1.24 (1.17-1.31)	3.50E-13	*TYK2* Missense	
rs7444	22:21976934	T/**C**	1.27 (1.21-1.33)	1.84E-22	*UBE2L3* 3'-UTR	*UBE2L3*

SLE GWAS SNPs taken forward for *cis*-eQTL association analysis. Results from post-replication meta-analysis as described (57). SNP with the lowest P-value post meta-analysis or the SNP with the greatest evidence of a missense effect as defined by a Bayes Factor reported. Autosomal, non-MHC SNPs with MAF > 0.05 included in analysis only (39 total). Risk alleles are highlighted in bold type – minor allele on right. Functional annotation from HaploReg v4.0 using GENCODE genes v10. Stated eGenes detected from microarray studies in LCLs listed as described (57).

Table 2
Details of genotype-expression (eQTL) cohorts used in study

Cohort Name	**TwinsUK (MuTHER)**				**Geuvadis**
Total subjects	856				373
Ethnicity	EUR (UK)				EUR (CEU, GBR, FIN, TSI)
Sex	F				M / F
Age	37–85				*NA*
Investigation	**Comparison of candidate-causal eQTL and eGene detection between microarray and RNA-Seq**			**Validation and comparison of LCL RNA-Seq discoveries in whole blood**	**Identification of asQTLs using RNA-Seq**
Citation	*Grundberg et. al* (9)	*Buil et. al* (39)	*Buil et. al* (39)	*Buil et. al* (39)	*Lappalainen et. al* (35)
Expression profile type	**Microarray**	**RNA-Seq**	**RNA-Seq**	**RNA-Seq**	**RNA-Seq**
Unit of expression	**Probe**	**Gene**	**Meta-exon**	**Meta-exon**	**Splice-junction**
Cell-type	LCL	LCL	LCL	Whole Blood	LCL
Subjects used in analysis	777	683	765	384	373
Data format	Genevar (summary results)	Read-count	Summary eQTL results	Summary eQTL results	Raw sequence alignments
RNA Platform	Illumina HT-12 V3	Illumina HiSeq2000		Illumina HiSeq2000	Illumina HiSeq2000
RNA-Seq mapper	*NA*	BWA v0.5.9 (GRCh37/hg19)		BWA v0.5.9 (GRCh37/hg19)	GEM v1.349 (GRCh37/hg19)
Reference transcriptome	*NA*	GENCODE V10		GENCODE V10	GENCODE V10
RNA-Seq read length	*NA*	49-bp PE		49-bp PE	75-bp PE

Breakdown of genotype-expression (eQTL) cohorts used in analysis. TwinsUK cohort in lymphoblastoid cell lines (LCLs) used for microarray and RNA-Seq comparison (profiled at gene and meta-exon resolution); meta-exons are described as non-redundant overlapping portions of exons generated flattening of the transcriptome annotation. All TwinsUK (MuTHER) samples used in analysis are derived from the original 856 individuals. Validation of LCL data in whole blood carried out at meta-exon level using 384 of the 856 individuals. Geuvadis cohort used for asQTL identification; splice-junction quantifications were generated by Altrans (57) from the raw sequence alignments. Summary eQTL results include only the eQTL association results per test (where full genotype and expression data were not obtainable).

123

Table 3
Candidate-causal eQTLs and associated eGenes detected using microarray (probe-level)

Risk Locus	GWAS SNP	eGene	Probe ID	β	*P*-Value	FDR (q)
1p13.2	rs2476601	*BCL2L15*	ILMN_1655722	-0.075	8.26E-04	2.73E-02
3q25.33	rs564799	*IL12A*	ILMN_1671353	0.113	4.91E-13	4.86E-11
3p14.3	rs11714389	*ABHD6*	ILMN_1706344	-0.130	1.41E-13	1.63E-11
		PDHB	ILMN_1739274	-0.055	2.55E-05	1.26E-03
4q24	rs10028805	*BANK1*	ILMN_1661646	0.195	7.98E-13	6.91E-11
8p23.1	rs2736340	*BLK*	ILMN_1668277	-0.407	7.69E-25	1.07E-22
		FAM167A	ILMN_1687213	0.412	1.48E-43	1.03E-40
15q24.2	rs2289583	*CSK*	ILMN_1754121	-0.059	9.56E-07	5.10E-05
		MPI	ILMN_1761262	0.058	1.45E-04	5.58E-03
		ULK3	ILMN_1679495	-0.071	1.78E-09	1.12E-07
17p13.2	rs2286672	*INCA1*	ILMN_1704380	0.035	7.69E-04	2.66E-02
22q11.21	rs7444	*UBE2L3*	ILMN_1677877	-0.183	4.97E-25	8.61E-23

GWAS SNPs deemed to be candidate-causal eQTLs using microarray expression data profiled from 777 individuals of the TwinsUK cohort in lymphoblastoid cell lines. 768 probes, corresponding to 559 genes, were tested against in *cis* to the 39 GWAS SNPs.

Table 4
Candidate-causal eQTLs and associated eGenes detected using RNA-Seq (Gene-level)

Risk Locus	GWAS SNP	eGene	β	*P*-Value	FDR (q)
1p13.2	rs2476601	*DCLRE1B*	-0.330	8.41E-04	1.91E-02
3q25.33	rs564799	*IL12A*	0.372	1.07E-11	9.70E-10
3p14.3	rs11714389	*ABHD6*	0.320	7.00E-09	2.28E-07
		RPP14	-0.285	4.15E-07	1.64E-05
4q24	rs10028805	*BANK1*	-0.316	2.37E-08	1.29E-06
5q31.1	rs17167273	*SKP1*	-0.564	1.57E-05	5.06E-04
		TCF7	-0.534	2.89E-05	8.27E-04
5q33.3	rs2431697	*MIR146A*	0.256	1.52E-06	6.89E-05
8p23.1	rs2736340	*FAM167A*	0.668	1.40E-33	7.62E-31
		BLK	-0.596	3.13E-28	8.51E-26
		RP11-148O21.2	-0.591	1.11E-26	2.01E-24
11p15.5	rs2396545	*HRAS*	-0.248	1.56E-04	4.04E-03
11q13.4	rs3794060	*NADSYN1*	-0.620	6.43E-24	8.74E-22
		DHCR7	-0.200	1.70E-03	3.43E-02
15q24.2	rs2289583	*ULK3*	0.303	3.62E-07	1.79E-05
		UBE2Q2	0.249	5.66E-05	1.54E-03
		FAM219B	-0.198	1.18E-03	2.57E-02
		CSK	0.194	1.64E-03	3.43E-02
22q11.21	rs7444	*UBE2L3*	0.345	2.28E-06	9.54E-05

GWAS SNPs deemed to be candidate causal eQTLs using RNA-Seq expression data profiled from 683 individuals of the TwinsUK cohort in lymphoblastoid cell lines at gene-level resolution. 520 genes were tested against in *cis* to the 39 GWAS SNPs.

Table 5
Candidate causal eQTLs and associated eGenes detected using RNA-Seq (Exon-level)

Risk Locus	GWAS SNP	eGene	Meta-exon ID (chr. start. end)	β	*P*-Value	FDR (q)
1p13.2	rs2476601	*BCL2L15*	1.114420790.114424619	0.226	2.71E-10	1.80E-08
			1.114429871.114430169	0.199	2.72E-08	1.28E-06
		DCLRE1B	1.114449618.114449783	0.110	2.23E-03	3.85E-02
		MAGI3	1.114225519.114228545	0.115	1.41E-03	2.62E-02
1q32.1	rs3024505	*FCAMR*	1.207133024.207134568	0.114	1.62E-03	2.94E-02
		IL10	1.206945616.206945839	0.174	1.26E-06	4.88E-05
		IL24	1.207076321.207077484	0.108	2.81E-03	4.63E-02
2q24.2	rs2111485	*IFIH1*	2.163123589.163123889	0.119	1.00E-03	1.98E-02
			2.163128736.163128897	0.108	2.69E-03	4.49E-02
			2.163130305.163130454	0.116	1.25E-03	2.37E-02
			2.163136506.163136622	0.138	1.31E-04	3.29E-03
			2.163137838.163138055	0.110	2.34E-03	4.01E-02
			2.163163219.163163680	0.127	4.45E-04	9.72E-03
3q25.33	rs564799	*IFT80*	3.159996981.159997152	0.109	2.65E-03	4.44E-02
		IL12A	3.159706537.159706961	0.202	1.66E-08	8.24E-07
			3.159710799.159711631	0.286	7.56E-16	8.33E-14
			3.159713191.159713806	0.281	2.23E-15	2.31E-13
		RP11-432B6.3	3.159996981.159997152	0.109	2.65E-03	4.44E-02
		SMC4	3.160129546.160129872	0.120	8.69E-04	1.74E-02
			3.160137146.160137331	0.112	1.97E-03	3.48E-02
			3.160141213.160141438	0.112	1.98E-03	3.48E-02
			3.160141549.160141628	0.110	2.36E-03	4.03E-02
			3.160149431.160149613	0.126	4.69E-04	1.02E-02
			3.160150071.160150303	0.122	7.51E-04	1.53E-02
			3.160150814.160150997	0.137	1.41E-04	3.47E-03
3p14.3	rs11714389	*ABHD6*	3.58223233.58223643	0.132	2.62E-04	6.10E-03
			3.58242289.58242432	0.173	1.57E-06	5.89E-05
			3.58252916.58253072	0.174	1.34E-06	5.11E-05
			3.58255048.58255161	0.207	7.18E-09	3.76E-07
			3.58256659.58256791	0.190	1.21E-07	5.27E-06
			3.58260385.58260542	0.200	2.47E-08	1.18E-06
			3.58270924.58271180	0.221	6.02E-10	3.75E-08
			3.58279316.58280461	0.345	7.89E-23	1.13E-20
4q24	rs10028805	*BANK1*	4.102946358.102946666	0.264	1.07E-13	9.91E-12
			4.102981368.102981546	0.253	1.19E-12	9.92E-11
5q31.1	rs17167273	*SKP1*	5.133492082.133494319	0.276	7.20E-15	7.17E-13
5q33.3	rs2431697	*MIR146A*	5.159912306.159914433	0.248	3.36E-12	2.52E-10
			5.159895275.159895447	0.151	2.74E-05	7.97E-04
8p23.1	rs2736340	*BLK*	8.11351510.11352100	0.462	9.76E-42	3.16E-39
			8.11366659.11367400	0.350	1.72E-23	2.55E-21
			8.11412252.11412398	0.469	5.06E-43	2.01E-40
			8.11414166.11414346	0.448	6.05E-39	1.65E-36
			8.11415471.11415547	0.445	2.13E-38	5.26E-36
			8.11417842.11418961	0.485	2.45E-46	2.11E-43
			8.11420488.11420619	0.440	1.49E-37	2.96E-35
			8.11421412.11422113	0.446	1.32E-38	3.43E-36
		FAM167A	8.11278972.11282145	0.469	3.48E-43	1.64E-40
			8.11301540.11302395	0.463	7.22E-42	2.49E-39
		RP11-148O21.2	8.11415975.11416256	0.454	3.63E-40	1.04E-37
			8.11416421.11416495	0.325	2.69E-20	3.67E-18
			8.11417293.11417529	0.395	5.49E-30	9.17E-28
		RP11-148O21.4	8.11413760.11414170	0.353	6.24E-24	9.50E-22
			8.11415399.11415531	0.391	2.15E-29	3.48E-27
11p15.5	rs2396545	*ANO9*	11.417933.418589	0.180	5.66E-07	2.32E-05
			11.418720.419415	0.207	7.70E-09	3.95E-07
			11.419582.420860	0.196	4.54E-08	2.06E-06
			11.420929.421042	0.163	6.11E-06	2.03E-04
			11.428088.428199	0.119	9.93E-04	1.97E-02
			11.428475.428639	0.181	4.99E-07	2.07E-05
			11.428722.428826	0.164	5.44E-06	1.83E-04
			11.429474.430179	0.203	1.52E-08	7.63E-07
			11.430269.430403	0.198	3.28E-08	1.52E-06
			11.431694.433459	0.144	6.20E-05	1.69E-03
			11.433815.433937	0.156	1.45E-05	4.48E-04
			11.434024.434098	0.127	4.24E-04	9.30E-03
		HRAS	11.533277.533358	0.136	1.64E-04	3.97E-03
			11.533766.533976	0.151	2.72E-05	7.97E-04
			11.534212.534375	0.138	1.29E-04	3.24E-03
		PHRF1	11.587259.588497	0.126	4.86E-04	1.05E-02
			11.597395.597570	0.118	1.03E-03	2.01E-02
			11.607066.609720	0.170	2.37E-06	8.56E-05
			11.611634.612222	0.142	8.00E-05	2.09E-03
		RASSF7	11.561547.561892	0.160	9.26E-06	2.98E-04

		RNH1	11.498457.498627	0.152	2.58E-05	7.72E-04
			11.504824.505881	0.309	2.28E-18	2.82E-16
			11.507113.507300	0.126	4.55E-04	9.90E-03
		TMEM80	11.700615.701127	0.131	2.75E-04	6.37E-03
11q13.4	rs3794060	*DHCR7*	11.71145460.71147019	0.246	5.41E-12	4.00E-10
			11.71148858.71148989	0.172	1.66E-06	6.15E-05
			11.71149795.71150129	0.179	6.25E-07	2.51E-05
			11.71152273.71152486	0.163	5.77E-06	1.93E-04
			11.71153309.71153399	0.207	7.53E-09	3.90E-07
			11.71155003.71155299	0.175	1.14E-06	4.45E-05
			11.71155901.71156004	0.150	3.12E-05	9.03E-04
		NADSYN1	11.71175099.71175554	0.227	2.04E-10	1.37E-08
			11.71185441.71186668	0.545	1.79E-60	9.26E-57
			11.71187079.71188484	0.538	1.06E-58	2.75E-55
			11.71189441.71190128	0.339	5.32E-22	7.45E-20
			11.71190340.71191320	0.419	7.54E-34	1.30E-31
			11.71191801.71193071	0.245	6.27E-12	4.57E-10
			11.71195358.71196694	0.280	2.94E-15	2.98E-13
			11.71202880.71202949	0.166	3.68E-06	1.28E-04
			11.71207481.71208657	0.248	3.30E-12	2.51E-10
			11.71209398.71211081	0.196	4.82E-08	2.15E-06
			11.71214910.71216920	0.297	5.05E-17	5.81E-15
		RP11-660L16.2	11.71159720.71159931	0.482	7.76E-46	5.74E-43
			11.71162736.71163203	0.530	1.44E-56	2.49E-53
12q24.12	rs10774625	*HECTD4*	12.112601913.112602079	0.113	1.81E-03	3.25E-02
15q24.2	rs2289583	*CSK*	15.75090574.75090653	0.153	2.27E-05	6.82E-04
			15.75091613.75091832	0.140	9.68E-05	2.48E-03
			15.75092753.75092846	0.143	7.38E-05	1.97E-03
			15.75093164.75093263	0.117	1.15E-03	2.20E-02
			15.75093863.75093936	0.140	1.08E-04	2.76E-03
			15.75094036.75094231	0.167	3.20E-06	1.14E-04
			15.75094338.75094424	0.129	3.65E-04	8.10E-03
		FAM219B	15.75192329.75195127	0.148	4.08E-05	1.15E-03
		MPI	15.75182867.75182995	0.150	3.29E-05	9.41E-04
			15.75185002.75185143	0.250	2.32E-12	1.87E-10
			15.75185479.75185661	0.191	1.08E-07	4.75E-06
			15.75189352.75189560	0.212	3.48E-09	1.92E-07
		ULK3	15.75128459.75129585	0.286	7.98E-16	8.61E-14
			15.75130607.75130685	0.130	3.14E-04	7.11E-03
			15.75130984.75131086	0.142	7.90E-05	2.09E-03
			15.75131351.75131391	0.145	5.99E-05	1.64E-03
			15.75131609.75131714	0.182	4.02E-07	1.69E-05
			15.75131899.75132054	0.150	3.16E-05	9.08E-04
			15.75132575.75132657	0.120	8.45E-04	1.70E-02
			15.75132839.75132982	0.133	2.32E-04	5.48E-03
			15.75133746.75133850	0.116	1.35E-03	2.54E-02
			15.75134416.75134536	0.129	3.46E-04	7.79E-03
			15.75134621.75134761	0.130	3.00E-04	6.88E-03
22q11.21	rs7444	*UBE2L3*	22.21965146.21965332	0.216	1.65E-09	9.46E-08
			22.21975804.21978323	0.422	2.19E-34	4.20E-32

GWAS SNPs deemed to be candidate causal eQTLs using RNA-Seq expression data profiled from 765 individuals of the TwinsUK cohort in lymphoblastoid cell lines at exon-level resolution. 4,786 exons, corresponding to 716 genes were tested against in *cis* to the 39 GWAS SNPs.

RNA-Seq better annotates SLE risk loci by revealing known and novel candidate genes

Fig. 2 illustrates the clear improvement of RNA-Seq relative to microarray in the discovery of candidate-causal eQTLs and their corresponding eGenes when annotating complex-disease susceptibility loci. With RNA-Seq, we replicated known SLE associated eQTLs and eGenes that were previously discovered using microarray. These included rs564799 for *IL12A*, rs2736340 for *BLK*, rs9311676 *ABHD6*, and rs2289583 for *ULK*, *CSK*, and *MPI* (**Tables 4**, and **5**; **Table 1** for previously reported associations in LCLs). Several eQTLs for eGenes that have been extensively studied in terms of their role in SLE pathogenesis were consistent across all three platforms: for example, rs10028805 for *BANK1* (57) and rs7444 for *UBE2L3* (58). **Fig. 2** also shows how exon-level RNA-Seq analysis led to the greatest frequency of candidate-causal eQTLs and eGenes than with either gene-level RNA-Seq or microarray (**S1 Fig.** and **S2 Fig.**). A total of 14 eQTLs modulating expression of 34 eGenes were detected using exon-level RNA-Seq contrasted to 11 eQTLs and 19 eGenes at gene-level RNA-Seq and only 8 eQTLs with 12 eGenes identified using microarray (**S1 Fig.** and **S2 Fig.**). Only one eQTL (rs2286672) and two eGenes (*PDHB*, *INCA1*) were found by microarray exclusively. These associations were either not significant post multiple testing using either RNA-Seq method, or were not deemed candidate-causal (**S1 Table S1** and **S3 Table**). In total 8 eQTLs regulating expression of 27 eGenes were detected using RNA-Seq but missed using microarray (**S3 Fig. S3** and **S4 Fig.**). Interestingly, exon-level analysis led to the greatest frequency of non-candidate-causal associations. Only 14 of the 34 significant eQTLs, $q < 0.05$, showed evidence of a shared causal variant post conditional and colocalisation testing (**S1 Fig.** and **S2 Fig.**).

Several eGenes known to be involved in the pathogenesis of SLE were identified using RNA-Seq exclusively and not reported in previous microarray-based eQTL studies in LCLs (**Table 1**). These include *IL10*, *IFIH1*, and the microRNA *MIR146A*. We believe a handful of

other eGenes unique to RNA-Seq to be novel SLE candidate genes. *NADSYN1* (NAD Synthetase), *HECTD4* (HECT Domain E3 Ubiquitin Protein Ligase 4), *SKP1* (S-Phase Kinase-Associated Protein 1) and *TCF7* (T-cell specific Transcription Factor 7) are examples of novel eGenes.

RNA-Seq eQTL analysis reveals eQTLs regulating multiple eGenes

Fig. 2 also illustrates that many of the eQTLs discovered using RNA-Seq regulate multiple eGenes. Exon-level analysis generated the greatest ratio (2.42) of candidate-causal eGenes to eQTLs (**S5 Fig.**); suggesting that disease-associated haplotypes may be more functionally potent and harbour multiple gene regulatory effects than previously thought.

One example of this effect is at the *TCF7-SKP1* locus where the disease-association signal encompasses both genes (**S6 Fig.**). Both *TCF7* and *SKP1* were classified as candidate-causal against the GWAS SNP rs7726414 using gene-level RNA-Seq (**Table 5**), and *SKP1* but not *TCF7* at exon-level (**Table 4**). Interestingly, a missense variant of *TCF7* has been implicated in Type 1 Diabetes risk (59), but there is only weak LD (r^2<0.4) between this missense variant and rs7726414 (19 kb upstream of *TCF7*) or any protein-coding variants of *TCF7*, suggesting that in SLE the causal mechanism may be dysregulated gene expression of *TCF7* rather than a missense change *per se*. *SKP1*, part of the ubiquitin ligase complex, is thought to stabilize the conformation of E3 ligases and its expression has been shown to be upregulated in lung-cancer (60). Neither of these effects were present in the microarray data as there is another more significant *cis*-eQTL, rs17167273, for the *TCF7* probe (ILMN_1676470), which is not correlated with the rs7726414[T] risk variant (r^2:0.02, **S1 Table**). In addition, rs7726414[T] was not a significant eQTL for the *SKP1* microarray probe (ILMN_1790710) (*P*=0.16). Our RNA-Seq eQTL analyses indicated that rs7726414[T] represents a novel eQTL for dysregulated gene expression of *SKP1* in SLE (**S6 Fig.**). We

believe both *TCF7* and *SKP1* to be highly plausible candidate genes as the literature suggests that knockdown of *TCF7* results in impaired stem cell potency and gene expression regulation of CD34+ cells (61), whilst the mouse knockout of *SKP1* develop highly penetrant T-cell lymphomas (62).

Other examples where either gene- or exon-level RNA-Seq analysis identified multiple eGenes are at rs3024505 (*IL10*, *IL24*, and *FCAMR*), rs12802200 and rs3794060 (**Fig. 2**). In the immune-gene concentrated 11p15.5 region (**S7 Fig.**), the GWAS SNP rs12802200, is an eQTL for six eGenes, (*HRAS**,** TMEM80*, *RNH1*, *ANO9*, *PHRF1* and *RASSF7*; **Tables 4 and 5**) following exon- and/or gene-level RNA-Seq analysis, supporting our recent observations of rs12802200 being a *cis*-eQTL for six genes across multiple immune cell-types at this locus (63) and earlier reports demonstrating that rs12802200 was an eQTL for *TMEM80* (whole blood(64)) and for both *RNH1* and *TMEM80* (LCLs (35)). 11p15.5 is also an important regulatory region in non-immune types, because a total of eight eGenes in this region have been identified in multiple non-immune cell types (65), four of which (*HRAS*, *TMEM80*, *RNH1*, *ANO9*) were captured by our RNA-Seq analysis in LCLs. Since SNPs within this susceptibility locus have also been previously shown to be correlated with increased autoantibody production and interferon-α activity in sufferers of SLE (66), further investigation will be required to elucidate a potential mechanism by which dysregulation of gene expression may contribute to these effects. In this gene-dense region (**S7 Fig.**), none of the six candidate-causal eGenes had an annotated microarray probe that passed quality control.

In the 11q13.4 region, the intronic GWAS SNP rs3794060[C], was classified using gene-level quantification as being a candidate-causal eQTL for both *NADSYN1* (NAD Synthetase 1) and *DHCR7* (7-Dehydrocholesterol Reductase) (**Table 4, Fig. 3A** and **S8 Fig.**). At exon-level resolution, *NADSYN1* was also deemed candidate-causal, as well as the non-coding RNA *RP11-660L16.2* (**Table 5**; described in following section). The risk variant rs3794060[C]

has been correlated with reduced circulating 25-hydroxy vitamin D concentrations (67) with reduced vitamin D levels being recently associated with increased disease activity of SLE (68)**.** Interestingly, although a mouse knockout of *DHCR7* showed reduced serum and tissue cholesterol levels (69), no autoimmune phenotype has been described at this locus. Given the role oxidative stress plays in promoting inflammation and triggering autoimmunity through tissue damage (70), it will be interesting to elucidate the role that dysregulation of *NADSYN1* expression plays in this process. Visualization of exon-level association data of *NADSYN1* against rs3794060 suggests a potential splicing mechanism affecting meta-exons 11 and 12 (P=1.79^{E-60}, 1.06^{E-58} respectively; **Fig. 3B**). The probe for *DHCR7* (ILMN_1815626) showed a weak, but not statistically significant, association with rs3794060 (P=2.9x10^{-03}), whilst *NADSYN1* had no annotated probe in the microarray dataset.

RNA-Seq uncovers the role of non-coding RNA modulation at SLE susceptibility loci

Quantification of non-coding polyadenylated RNAs in the TwinsUK LCL cohort through RNA-Seq revealed three candidate-causal eQTLs influencing the expression of four non-coding eGenes (**Table 6**); none of which were captured using microarray.

We validated the effect at rs2431697 (5q33.3) where it is documented that the protective minor allele [C] is associated with expression upregulation of the miRNA *MIR146A*, a negative regulator of the type I Interferon pathway (71) (**Fig. 4A**). The best eQTL for *MIR146A* at gene-level was rs2431697 (P=1.5x10^{-06}) and also at exon level for both of its exons (P=3.4x10^{-12} and 1.2x10^{-04}). The decrease in rs2431697[T]-dependent expression of *MIR146A* reported in peripheral blood leukocytes of SLE patients disrupts binding of transcription factor Ets-1, uncouples of the type-1 IFN response (71), thereby increasing the inflammatory response.

Fig. 4B shows GWAS variant rs2736340, and other SNPs in tight LD (r^2>0.8), within the *FAM167A-BLK* (8p23.1) bi-directional promoter region. In this study we detected eQTLs at

the known eGenes *BLK* (B lymphocyte kinase) and *FAM167A* (Family with Sequence Similarity 167, Member A) by all quantification methods (**S9 Fig.**). The rs2736340[T] risk allele causes decreased expression of *BLK* (P=3.2x10^{-28}) and increased expression of *FAM167A* (P=1.4x10^{-33}) at RNA-Seq gene-level (**Table 4**). This effect has been previously described in microarray studies (72) - with reduced promoter activity of *BLK* leading to altered B-cell development (73). The eQTL lies within a region of the genome subject to multiple regulatory effects, with a 24bp region immediately around the variant containing strong chromatin marks (peak height>60) in LCLs for H3K9me3 (gene repression) and H3K4me3 (low expression when present in the promoter of CpG genes) assayed as part of the ENCODE project (**S9 Fig.**). Interestingly, exon-level RNA-Seq analysis revealed that rs2736340 also appeared to modulate the expression of two non-coding RNAs antisense to the 3′ region of *BLK*. These are: *RP11-148O21.2* and *RP11-148O21.4* (**Fig. 4B**). The SNP, rs2736340, significantly reduced the expression of both *BLK* and three exons of *RP11-148O21.2* and the two exons of *RP11-148O21.4* (**Table 6**). Our analyses indicated that expression disruption of these antisense RNAs caused by SLE risk variants represent a potential novel mechanism at the locus.

There is an rs3794060 allele-dependent expression modulation of both exons of another non-coding eGene, *RP11-660L16.2I* (**Fig. 4C**), which is located in the bi-directional promoter between *DHCR7* and *NADSYN1*. The best eQTL for those exons is highly correlated with the GWAS SNP (rs2282621, r^2:0.99). Both *DHCR7* and *NADSYN1* are candidate-causal eGenes regulated in the same downward direction at RNA-Seq gene-level with respect to the risk allele rs3794060[T] (P=1.7x10^{-03} and P=6.4x10^{-24}, respectively) (**S8 Fig.** and **Table 4**).

Table 6

Candidate-causal non-coding eGenes discovered using RNA-Seq (exon-level)

Risk Locus	GWAS SNP	eGene	Exon ID (chr. start. end)	β	*P*-Value	FDR (q)
5q33.3	rs2431697	*MIR146A*	5.159912306.159914433	0.248	3.36E-12	2.52E-10
			5.159895275.159895447	0.151	2.74E-05	7.97E-04
8p23.1	rs2736340	*RP11-148O21.2*	8.11415975.11416256	0.454	3.63E-40	1.04E-37
			8.11416421.11416495	0.325	2.69E-20	3.67E-18
			8.11417293.11417529	0.395	5.49E-30	9.17E-28
		RP11-148O21.4	8.11413760.11414170	0.353	6.24E-24	9.50E-22
			8.11415399.11415531	0.391	2.15E-29	3.48E-27
11q13.4	rs3794060	*RP11-660L16.2*	11.71159720.71159931	0.482	7.76E-46	5.74E-43
			11.71162736.71163203	0.530	1.44E-56	2.49E-53

GWAS SNPs deemed to be candidate causal eQTLs for non-coding eGenes using RNA-Seq expression data profiled from 765 individuals of the TwinsUK cohort in lymphoblastoid cell lines at exon-level resolution.

Confirmation of LCL candidate-causal eQTLs and eGenes using whole-blood RNA-Seq

To validate our LCL findings in a primary tissue-type, we extended our analytical pipeline to include an exon-level RNA-Seq dataset in 384 whole-blood samples from the TwinsUK cohort (**Table 2**). The full results of these analyses are provided in **S4 Table**.

We observed good correlation between LCLs and whole-blood effect-sizes (β) of GWAS SNPs against all matched *cis* exon-level associations (R^2=0.74; **S10 Fig.**). Seven of the 39 GWAS SNPs were classified as candidate-causal eQTLs in whole-blood, modifying 19 candidate-causal eGenes (**Table 7**). All seven of the whole-blood eQTLs and 15 of the 19 eGenes were deemed candidate-causal in LCLs, suggesting strong conservation across whole-blood cell types (**S11 Fig.**). The remaining four eGenes specific to whole-blood were: *PXK* (rs9311676); *IRF7* and *TALDO1* (rs12802200); and *SCAMP2* (rs2289583) (**Table 7**). Interestingly, the eQTLs regulating these four eGenes in whole-blood also regulated multiple eGenes in LCLs (**S11 Fig.**), implying that they tag highly regulatory haplotypes that may cause cell-type specific gene-expression disruption across the entire locus (same eQTL regulating different eGenes across different cell-types). Three of the four candidate-causal non-coding eGenes from LCLs were found in whole-blood (*RP11-148O21.2*, *RP11-148O21.4*, and *RP11-660L16.2*.). The LCL-specific *MIR146A* eGene association with GWAS SNP rs2431697 was not deemed to be significant (*P*=0.32), which is likely to be a result of its lymphocyte-specific gene expression profile (71) that is diluted in the heterogeneous population of whole-blood cell-types.

Inspection of specific exons modulated by the GWAS SNPs in each cell-type revealed instances of variability in the genetic control of exon usage between cell-types. A known splicing event in B-cells caused by branch-point SNP rs17266594 results in the loss of exon 2 in susceptibility gene *BANK1* which subsequently leads to B-cell hyper-responsiveness (57) (**S12 Fig.**). In whole-blood the GWAS variant, rs10028805, is associated with altered

expression of exon 2 (P=8.4x10^{-05}), with the best *cis*-eQTL for this effect being in near-perfect LD (rs4411998; r^2:0.98). Both rs10028805 and rs4411998 are in strong LD with the branch-point SNP (r^2:0.9). In LCLs however, the best *cis*-eQTL for exon 2, rs4572885 (P=9.74 x10^{-23}), has a large effect but is less correlated with the GWAS SNP (r^2:0.65) and conditional analysis judges the effect of the GWAS SNP to be independent to the best *cis*-eQTL for exon 2 (**S3 Table**). Interestingly, there is low correlation between the branch-point SNP rs17266594 and the best *cis*-eQTL for exon 2 in LCLs (r^2:0.42); suggesting the regulatory mechanism of exon 2 splicing in *BANK1* may be under two separate genetic influences between the two cell-groups (**S12 Fig.**).

We saw a near identical pattern of differential exon usage within eGene *NADSYN1* between LCLs and whole-blood driven by the GWAS SNP rs37940460 or tightly correlated variants (**S13 Fig.**, **Table 7**). Variation at rs37940460 appeared to drive extensive expression disruption of two meta-exons (11 and 12) of *NADSYN1* located near the centre of the gene (meta-exon 11: LCL P=1.79x10^{-60}; whole-blood P=1.28x10^{-27}; meta-exon 12: LCL P=1.06x10^{-58}; whole-blood P=6.30x10^{-26}). These two meta-exons were deemed to be candidate-causal for SLE across both cell-types. We believe the meta-exons in the 3′ end of *NADSYN1* that are candidate-causal in LCLs (**Fig. 3**), are not detected in whole-blood may be because of the smaller sample size or the mixed cell-type composition of the whole blood cohort. This novel instance of specific exon expression disruption found in a primary cell-type at *NADSYN1* may help to resolve the functional consequence of this locus.

Table 7

Candidate-causal eQTLs and eGenes detected using RNA-Seq (exon-level) in whole blood

Risk Locus	GWAS SNP	eGene	Exon ID (chr. start. end)	β	*P*-Value	FDR (q)
3p14.3	rs9311676	*PXK*	3.58383332.58383449	0.272	1.15E-07	2.03E-05
		RPP14	3.58303167.58305816	0.207	6.18E-05	5.45E-03
			3.58303167.58305920	0.211	4.52E-05	4.07E-03
4q24	rs10028805	*BANK1*	4.102750965.102751363	0.204	8.41E-05	7.11E-03
8p23.1	rs2736340	*BLK*	8.11400733.11400856	0.297	6.20E-09	1.40E-06
			8.11403240.11403612	0.268	1.74E-07	2.94E-05
			8.11405541.11405634	0.267	2.01E-07	3.26E-05
			8.11412252.11412398	0.282	3.72E-08	6.86E-06
			8.11412841.11412993	0.301	3.80E-09	9.63E-07
			8.11414166.11414346	0.290	1.50E-08	3.04E-06
			8.11415471.11415547	0.264	2.75E-07	4.29E-05
			8.11417842.11418961	0.373	1.26E-13	4.65E-11
			8.11420488.11420619	0.334	4.58E-11	1.33E-08
			8.11421412.11422113	0.344	1.16E-11	3.62E-09
		FAM167A	8.11278972.11282145	0.523	2.96E-27	6.00E-24
		RP11-148O21.2	8.11417293.11417529	0.222	1.79E-05	1.96E-03
			8.11415975.11416256	0.203	9.13E-05	7.56E-03
		RP11-148O21.4	8.11413760.11414170	0.220	2.09E-05	2.23E-03
11p15.5	rs12802200	*ANO9*	11.419582.420860	0.180	5.20E-04	3.35E-02
		HRAS	11.532242.532755	0.179	5.75E-04	3.64E-02
		IRF7	11.613785.614534	0.263	3.02E-07	4.54E-05
			11.614783.615728	0.216	2.89E-05	2.86E-03
		RNH1	11.504824.505881	0.238	3.84E-06	5.02E-04
		TALDO1	11.758950.759057	0.186	3.44E-04	2.36E-02
11q13.4	rs3794060	*NADSYN1*	11.71185441.71186668	0.526	1.28E-27	5.19E-24
			11.71187079.71188484	0.512	6.30E-26	8.52E-23
		RP11-660L16.2	11.71159720.71159931	0.437	1.48E-18	8.58E-16
			11.71162736.71163203	0.508	1.54E-25	1.56E-22
15q24.2	rs2289583	*CSK*	15.75094672.75095539	0.204	8.17E-05	7.05E-03
		MPI	15.75189853.75191798	0.252	1.02E-06	1.43E-04
		SCAMP2	15.75142855.75143014	0.188	2.83E-04	2.05E-02
		ULK	15.75128459.75129585	0.259	4.62E-07	6.69E-05
			15.75132839.75132982	0.189	2.58E-04	1.94E-02
			15.75134621.75134761	0.194	1.85E-04	1.42E-02
22q11.21	rs7444	*UBE2L3*	22.21975804.21978323	0.505	3.68E-25	2.99E-22

GWAS SNPs deemed to be candidate causal eQTLs using RNA-Seq expression data profiled from 384 individuals of the TwinsUK cohort in whole blood at exon-level resolution. In total, 3,793 exons were tested against, corresponding to 654 genes.

Splice-junction quantification reveals asQTLs and additional candidate eGenes

We extended our investigation to determine whether the GWAS SNPs (**Table 1**) had a direct influence on the alternative-splicing of transcripts (alternative-splicing quantitative trait loci; asQTL), and whether expression quantification at this resolution would reveal any additional candidate-genes or potential functional mechanisms. We undertook *cis*-asQTL analysis within a +/-1Mb window around each GWAS SNP against 33,039 splice-junction quantifications, corresponding to 817 genes, in the Geuvadis cohort (**Table 2**). We identified nine asQTLs significantly associated with 62 splice-junctions, corresponding to 10 eGenes (**S5 Table**). After testing for a shared causal variant between the GWAS and asQTL signal, six SLE candidate-causal asQTLs (26 splice-junctions) for seven eGenes remained (**Table 8**). Four eGenes (*TCF7*, *SKP1*, *BLK*, and *NADSYN1*) had been previously associated through either gene-level or exon-level eQTL mapping using the TwinsUK cohort. The remaining three candidate-causal eGenes detected using asQTL mapping (*IKZF2*, *WDFY4*, and *IRF5*), as well as the novel causal mechanism involving *NADSYN1*, are described below.

IKZF2

IKZF2 (Ikaros Family Zinc Finger 2) is novel SLE candidate-causal eGene detected only by asQTL analysis. The GWAS association signal around the 3′ end of *IKZF2* tagged by risk variant rs3768792[G] drove an increase in the fraction of splicing between exon 6A and exon 6B (P=3.8x10^{-05}); a bridge that is unique to the truncated isoform (ENST00000413091, 239 amino-acids) of *IKZF2* (**Fig. 5A**). Interestingly, this isoform possesses a premature termination codon found on exon 6B that is not found on the canonical isoform (ENST00000457361, 526 amino-acids) as in this isoform, exon 6A is spliced to exon 7 (**Fig. 5B**). This effect results in the premature truncation of the full-length protein and the subsequent loss of the two zinc-finger dimerization domains found on exon 8 (**Fig. 5B**). *IKZF2* is a transcription factor thought to play a key role in T-reg stabilisation in the presence of inflammatory responses (74). Since the Ikaros transcription factor family primarily regulate

gene expression through homo-/hetero-dimerization and DNA binding/protein-protein interactions, the rs3768792[G] dependent asQTL effect on exon 6A to 6B resulting in less functional *IKZF2* could be highly deleterious. *IKZF2* is known to regulate T-reg associated genes, including *IL-2* and *FoxP3* (75, 76), therefore a decrease in the amount of DNA binding *IKZF2* may result in loss of T-reg stability and a decrease of suppressive capacity with consequential autoimmune sequelae. Interestingly, we identified an additional asQTL variant (rs2291241) in near-perfect LD with the rs3768792 GWAS variant (r^2:0.99), located 9 bp upstream of exon 6B in truncated isoform ENST00000413091 (**S5 Table, Fig. 5C**). This second asQTL, located within the polypyrimidine tract in the exon 6A/exon 6B intron, is a highly plausible driving variant and may act through promotion of the described splicing event (**Fig. 5C**).

WDFY4

We also discovered a novel putative SLE-associated splicing mechanism involving *WDFY4* (WDFY Family Member 4), a gene belonging to a family thought to function as master conductors of aggregate clearance by autophagy (77). Risk variant rs2263052[G] or correlated SNPs (**Fig. 6A**) greatly increased the fraction of link-counts between exon 34A and exon 34B (P=3.3x10^{-19}) which are unique to the truncated isoform ENST00000374161 (**Fig. 6B**). This truncated isoform (552 amino-acids) lacks the two WD40 domains found in the full length isoform (ENST00000325239, 3184 amino-acids) that are essential to enzymatic activity (77). There is a consequential decrease in the fraction of link-counts between exon 34A and exon 35 (P=3.0x10^{-06}) that are unique to the canonical isoform of *WDFY4* (**Fig. 6B**). Interestingly, a known missense variant found in exon 31 of *WDFY4* (**Fig. 6B**), rs7097397 (Arg1816Gln), in strong LD (r^2:0.7) with rs2263052, has also been implicated in SLE through GWAS (78). SIFT and PolyPhen predict the amino-acid substitution to be tolerated (0.38) or benign (0.11) respectively (79); thereby suggesting the risk haplotype may harbour two functional mechanisms influencing *WDFY4* (amino-acid change and upregulation of a shorter isoform) that are both involved in SLE pathogenesis.

IRF5

The GWAS SNP and known asQTL, rs3757387, causes differential promoter usage of *IRF5* (Interferon regulatory factor 5); a molecular mechanism that has previously been reported in predisposition of SLE (80). Alteration of a consensus splice-site causes upregulation of a shorter isoform of *IRF5* which subsequently leads to erroneous activation of the type-1 IFN-pathway and pro-inflammatory cytokines (81). We replicated this known effect by observing an increased fraction of splicing of the shorter isoform of *IRF5*, ENST00000489702, with respect to the risk allele rs3757387[C] (**Table 8**). The risk allele increases splicing from the first exon to the penultimate exon of ENST00000489702 (P=2.2x10^{-08}); and from the first exon to the final exon of ENST00000489702 (P=5.9 x10^{-07}).

NADSYN1

Finally, using splice-junction quantification, we were able to pinpoint the specific transcript of *NADSYN1* that drives the exon-level association previously described (**Table 5**, **Fig. 3**). The GWAS SNP rs3794060 leads to substantial upregulation of the meta-exon 10 to meta-exon 12 splice-site (P=8.0x10^{-12}) which is unique the ENST00000528509 transcript of *NADSYN1* (**Table 8**, **S14 Fig.**). As a consequence of this splicing event, it appears the meta-exon 11 to meta-exon 12 splice-site is highly reduced (P=2.1x10^{-14}) with reference to the risk allele [C]. Meta-exons 11 and 12 were implicated in exon-level analysis but a specific transcript could not be isolated as the gene annotation was not collapsed to the granularity used in the asQTL analysis. Interestingly, transcript ENST00000528509 is translated to a 294 amino acid residue protein where the canonical transcript of *NADSYN1*, ENST00000319023, is 706 amino acids. The shorter protein lacks the NAD(+) Synthetase domain (located in positions 339-602aa) found in the canonical protein (Pfam: PF02540); thus implicating loss of this domain as a potential causal mechanism.

Table 8

Candidate-causal asQTLs and associated eGenes detected using RNA-Seq (Splice-junction level)

Risk Locus	GWAS SNP	eGene	Spice Junction (chr: donor-acceptor exon)	β	*P*-Value	FDR (q)
2q34	rs3768792	*IKZF2*	2:213872084.213872808-213886368.213886444	-0.04	3.64E-09	5.51E-06
			2:213878515.213880002-213886368.213886444	-0.03	1.30E-06	1.07E-03
			2:213881647.213881768-213886368.213886444	-0.04	1.48E-07	1.41E-04
			2:213886717.213886854-213886368.213886444	0.07	3.80E-05	1.93E-02
			2:213914437.213914604-213886368.213886444	-0.04	4.55E-06	3.15E-03
5q31.1	rs7726414	*SKP1*	5:133541305.133541822-133561451.133561762	0.02	1.42E-06	1.16E-03
			5:133541645.133541822-133561451.133561762	-0.03	1.28E-06	1.07E-03
		TCF7	5:133478412.133478791-133474642.133474729	0.01	1.90E-06	1.49E-03
7q32.1	rs3757387	*IRF5*	7:128585899.128586088-128577666.128577888	0.02	2.18E-08	2.49E-05
			7:128586555.128586616-128577666.128577888	0.05	9.51E-07	8.13E-04
8p23.1	rs2736340	*BLK*	8:11351510.11352100-11403561.11403612	0.10	4.16E-05	2.08E-02
			8:11400733.11400856-11403561.11403612	-0.04	6.24E-05	2.96E-02
			8:11407668.11407771-11403240.11403612	0.02	1.74E-09	3.14E-06
			8:11412252.11412398-11403240.11403612	0.02	2.77E-09	4.61E-06
			8:11414167.11414297-11407503.11407771	0.01	1.36E-08	1.64E-05
			8:11415471.11415547-11407503.11407771	0.01	2.52E-06	1.90E-03
			8:11415471.11415547-11414166.11414346	-0.04	3.24E-06	2.32E-03
			8:11420488.11420619-11414166.11414346	0.02	1.09E-13	7.06E-10
10q11.23	rs2663052	*WDFY4*	10:50031227.50032214-50030425.50030582	-0.02	3.31E-19	2.15E-14
			10:50034716.50034954-50030425.50030582	0.02	3.01E-06	2.17E-03
11q13.4	rs3794060	*NADSYN1*	11:71171087.71171259-71171768.71171993	-0.06	2.95E-05	1.55E-02
			11:71184324.71184412-71187247.71188271	0.17	8.00E-12	2.08E-08
			11:71184615.71184732-71187079.71188484	0.06	1.02E-04	4.51E-02
			11:71185441.71186668-71187247.71188271	-0.19	2.13E-14	1.73E-10
			11:71185441.71185572-71189441.71189515	0.02	1.06E-05	6.71E-03
			11:71185441.71186668-71187079.71188484	-0.06	1.02E-04	4.51E-02

GWAS SNPs deemed to be candidate causal asQTLs using RNA-Seq expression data profiled from 373 individuals of the Geuvadis cohort in lymphoblastoid cell lines at splice-junction resolution. 33,039 splice-junctions, corresponding to 817 genes were tested against in cis to the 39 GWAS SNPs.

Discussion

Detailed characterization of the functional effects of human regulatory genetic variation associated with complex-disease is paramount to our understanding of molecular aetiology and poised to make significant contributions to translational medicine (82). Use of eQTL mapping studies to interpret GWAS findings have proved fundamental in our progression towards this goal - through prioritization of candidate genes, refinement of causal variants, and illumination of mechanistic relationships between disease-associated genetic variants and gene expression (82, 83). However, there is often a disparity between disease-associated genetic variation and phenotypic alteration, which historically may be due to the use of microarray-based technologies to profile genome-wide gene expression. With the advent of RNA-Seq, we can achieve more accurate quantification of the mRNA output of genes, individual exons, and isoform abundance, as well as unannotated and non-coding transcripts. Detection of splicing variants at susceptibility loci using RNA-Seq has the potential to uncover the role of specific isoforms implicated in disease risk, which are likely to have remained concealed by microarray, as a largely independent subset of variants control alternative splicing of isoforms compared to overall gene abundance (35).

Our two major motivations for this study were firstly to directly compare the ability of RNA-Seq with microarrays to detect candidate-causal eQTLs and their associated eGenes from GWAS data, and then to assess each platform's effectiveness in explaining the potential causal genes and mechanisms implicated by SLE risk alleles. A previous investigation of the same SLE risk alleles used in this study with eQTL data from LCL microarray datasets revealed that only 13 of the 39 risk alleles were eQTLs for a total of 15 eGenes (**Table S6**). Therefore, in an attempt to increase our understanding of the extent to which eQTLs can explain the functional consequences of our risk alleles and to achieve both of our major aims for this investigation, we set up an analytical pipeline to compare the results of eQTL annotation for 39 SLE susceptibility loci with *cis*-eQTL data from both microarray and RNA-

Seq experiments from the TwinsUK cohort in LCLs. We incorporated steps to minimize false-positive associations through conditional and colocalisation analysis, as approximately 34% of all genes have a second independently associated *cis*-eQTL for any of their exons when conditioning on the best *cis*-eQTL (35). The analytical pipeline we present will be applicable to the functional annotation of susceptibility loci from a wide range of human diseases.

Fig. 7 summarises the data generated in this manuscript illustrating which of the GWAS SNPs show evidence of a candidate-causal eQTL association across the quantification types. Our data analysis revealed that RNA-Seq is a more powerful in the identification of candidate-causal eQTLs and their accompanying eGenes than corresponding microarray datasets (**S6 Table**). Many of the published SLE candidate eGenes and associated mechanisms were well-replicated by performing *cis*-eQTL analysis of RNA-Seq datasets at various resolutions for each of the GWAS variants. These eGenes included, among others, the effect of risk variants at *BLK* and *FAM167A* (**S9 Fig.**), *MIR146A* (**Fig. 4A**), *BANK1* (**S12 Fig.**) and *IRF5*. Microarray studies were unable to detect the novel SLE candidate-causal eGenes identified from RNA-Seq data including *NADSYN1, TCF7*, *SKP1*, *WDFY4*, *IKZF2*, and the non-coding RNA genes: *RP11-148O21.2, RP11-148O21.4*, and *RP11-660L16.2* (**Fig. 2**).

Our results also demonstrate that RNA-Seq analysis is much better than microarrays in identifying multiple eGenes for a single SNP (that may tag multiple functional variants). An increased ratio of eQTLs to eGenes (average number of eGenes per candidate-causal eQTL) was observed using RNA-Seq at exon-level (2.42) compared with gene-level (1.72) and which were both greater than microarray (1.5) (**S5 Fig.**). The ability of RNA-Seq exon-level analysis to identify multiple target eGenes for a specific eQTL is supported by recent observations from capture Hi-C (cHi-C) experiments to functionally annotate chromatin interactions, such as enhancer–promoter interfaces (84, 85). It has been shown that

chromatin interactions can control transcription in both *cis* and *trans* in a largely sequence-specific manner, thus it is likely that some GWAS variants may functionally act through the disruption of chromatin dynamics resulting in perturbation of expression of multiple genes (84, 86, 87). Specific instances of this type of effect are seen in colorectal cancer risk loci where the risk locus 11q23 mapped to interactions with genes *C11orf53*, *C11orf92* and *C11orf93,* and separately, the risk SNP rs6983267 within 8q24 disrupts a chromatin regulatory network involving interactions between three genes *CCAT2*, *CCAT1* and *MYC* (84). Our results support this notion of multiple perturbed genes at a single susceptibility locus. At 1q32.1, for example, rs3024505 was found to be associated with three plausible candidate-causal eGenes: *IL10*, *IL24*, and *FCAMR* (located 1 kb, 130 kb, and 191 kb away from rs3024505 respectively). These chromatin capture data also support the argument of using RNA-Seq and extending the *cis*-eQTL distance (typically +/-0.25–1Mb) to a larger region (+/-5Mb) around the associated SNP to identify effects caused by chromatin interactions over a larger distance than commonly designated as *trans* from eQTL-type analyses (88). Integration of eQTL data with epigenetic regulation (promoter methylation, histone modification and expression of non-coding RNA) will allow the identification of the potential mechanism of action (disruption of epigenetic landscape) and disease biology of associated variants.

We have also demonstrated the power of RNA-Seq compared with microarray in the discovery of alternative-splicing events. This is of significant importance as approximately 80% of all human genes undergo alternative splicing and it is estimated that 20-30% of disease-associated mutations modify the configuration of expressed isoforms (89, 90). A recent study has concluded that regulatory variants controlling gene splicing are major contributors to complex traits (91). Two examples of this taken from our analyses are the risk alleles at the *IKZF2* and *WDFY4* loci which drive up-regulation of short isoforms. Neither of these effects were captured by the microarray probes that targeted the 3′-UTR of the canonical longer isoforms. At *WDFY4*, the splicing variant rs2263052[G] is in strong LD with

an (Arg1816Gln) missense variant rs7097397[G], which we showed to be associated with SLE (**Fig. 6**). These two potentially causal signals may reinforce each other. The novel association identified in our group's recent GWAS study (63) at the *IKZF2* locus implicated a risk haplotype tagged by the risk allele of rs3768792[G]. We identified, by RNA-Seq splice-junction quantification exclusively, that variation at the rs3768792[G] risk allele led to increased production of a shorter isoform of *IKZF2* (**Fig. 5**). Interestingly, other members of this gene family, *IKZF1* and *IKZF3*, are also associated with SLE (63). An associated variant in the 3′-UTR of *IKZF1* has been associated with increased expression of genes of the type 1 IFN pathway and decreased expression of complement genes, both mediated in *trans* (64). The functional effect of the *IKZF3* common variant is less well documented, although *IKZF3* knockout mice develop spontaneous autoantibodies and B-cell lymphoma (92). The *IKZF2* knockout mouse has not been characterised for immune-deficient phenotypes. We hypothesize that upregulation of the shorter isoform of *IKZF2* caused by rs3768792[G], which lacks the dimerization domain, reduces translocation of the protein into the nucleus and regulation of transcription of target genes. We believe this aberrant mechanism may result in loss of T-reg stability. The asQTL discoveries described in this manuscript are examples of how RNA-Seq can suggest a potential causal mechanism that can be easily validated experimentally.

Our study also demonstrates that RNA-Seq is able to identify disease-relevant non-coding RNAs. These type of transcripts have long been known to be of relevance to human disease, however their detection and functional importance may have been under-estimated. We identified three novel non-coding anti-sense RNAs using RNA-Seq: two regulated by rs2736340 (*RP11-148O21.2* and *RP11-148O21.4*) (**Fig. 4B**), and one regulated by rs3794060 (*RP11-660L16.2*) (**Fig. 4C**). These findings were replicated in whole blood (**Table 7**).

The data we present in this manuscript demonstrate that a comprehensive integrated approach for eQTL analysis should be undertaken at gene-, exon- and exon-junction level quantification. Excluding one or more levels of analysis will mean that eQTLs and/or eGenes may be missed. This is illustrated by a number of examples in **Fig. 7**. At some loci, there was an exon-level effect, which was not observed at the gene-level. For example, the risk allele rs3024505[A] was an eQTL for *IL10*, *IL24*, and *FCAMR* at exon-level resolution only (**Table 5**), but none of these eGenes were deemed to be significant at gene-level. This suggests that the exon-level effect is more targeted than gene-level quantification, where multiple different effects across the gene may dilute out the signal. However, some eGenes exhibit a probable whole gene-level effect (*UBE2L3*, *BLK*, and *FAM167A)* as every expressed exon showed a candidate-causal association for the respective risk alleles. Hierarchical clustering tests could be designed to distinguish between these genuine gene-level and exon-level effects, such as at *NADSYN1* or *BANK1*, where the gene-level effect is likely to be driven by only a subset of exons (**Fig. 3**, **S12 Fig.**). There were occurrences where significant candidate-causal eGenes were detected but there was no effect at the exon-level. At *TCF7*, variants in low LD with rs7726414 exhibited significant exon-level *cis*-eQTLs (**S3 Table**) and were deemed to be independent of the disease-association. However, gene-level analysis revealed that the risk rs7726414 variant was candidate-causal for total *TCF7* expression (**S2 Table**). These results emphasise that for any given eGene there may be multiple genetic effects at different resolutions of quantification.

We understand the limitation of LCLs for transcript profiling studies. There is an inherent limitation analysing LCL expression datasets, such as those available for our study, because although LCLs are a good surrogate model for primary B-cells, the effect of EBV transformation is likely to disrupt their underlying epigenetic and transcriptomic background. The percentage of asQTLs in LCLs will exhibit significantly less replication in primary cell types because of cell-type variability in the genetic control of isoform usage, approximately 70% of *cis*-eQTL detected in LCLs can be replicated in a primary tissue type (33). The

significance of alternative-splicing in genomic medicine will become better understood once large RNA-Seq based eQTL cohorts emerge across a multitude of disease-relevant cell-types. A gold standard of candidate-causal eQTL mapping strategies using RNA-Seq across datasets using an explicit set of quantification types (gene-, exon-, splice-junction, isoform) and analytical pipelines, will accelerate this process.

In summary, we have demonstrated the effectiveness of eQTL analysis using RNA-Seq by increasing the numbers of candidate eGenes regulated by SLE associated alleles (**Fig. 2**, **Fig. 7**, and **S6 Table**). We have shown that the power of RNA-Seq in eQTL annotation of GWAS loci lies not only the assessment of the variants regulating the expression of candidate genes, but also in uncovering putative molecular mechanisms, which allow for more refined targeted follow-up studies to assess the phenotypic consequence of the disease-associated variant. These studies could include knocking down target gene or introducing recombinant vectors *in vitro* to overexpress target genes to evaluate phenotypic consequence of expression changes of specific loci. Site directed mutagenesis could be used to introduce candidate causal splice-sites and over-express target isoforms. The CRISPR/Cas9 system for targeted genome editing presents an exciting opportunity for eQTL/RNA-Seq targeted follow-up studies and the investigation of the effect that specific variants have on the expression profile across different cell types.

Materials and Methods

Selection of SLE associated SNPs

SLE associated SNPs were taken from our recent 2015 publication (63). The study comprised a primary GWAS, with validation through meta-analysis and replication study in an external cohort (7,219 cases, 15,991 controls in total). The independently-associated susceptibility loci taken forward for this investigation were those that passed either genome-wide significance ($P < 5x10^{-08}$) in the primary GWAS or meta-analysis and/or those that reached significance in the replication study (False Discovery Rate, q<0.01). We defined the 'GWAS SNP' at each locus as either being the SNP with the lowest *P*-value post meta-analysis or the SNP with the greatest evidence of a missense effect as defined by a Bayes Factor. We omitted non-autosomal associations and those within the Major Histocompatibility Complex (MHC), and SNPs with a MAF <0.05. In total, 39 GWAS SNPs were taken forward (**Table 1**).

TwinsUK cohort eQTL datasets

Expression profiling by microarray (9) and RNA-Seq (39) of individuals from the UK Adult Twin Registry (TwinsUK) was carried out in two separate studies on the MuTHER (Multiple Tissue Human Expression Resource) cohort (**Table 2**). The MuTHER cohort is composed of 856 healthy female individuals of European descent aged between 37-85 years. We considered expression quantification data from both resting LCLs and whole blood. Profiling by microarray was performed using the Illumina Human HT-12 V3 BeadChips. For RNA-Seq, samples were sequenced using the Illumina HiSeq2000 and the 49-bp paired-end reads mapped with BWA v0.5.9 to the GRCh37 reference genome. Exons ('meta-exons' created by merging all overlapping exonic portions of a gene into non-redundant units) were quantified using read-counts against the GENCODE v10 annotation; with gene quantification defined as the sum of all exon quantifications belonging to the same gene. Full quality control and normalization procedures are described in the respective articles. Data from

each of the TwinsUK eQTL studies (**Table 2**) were provided in different formats. In each instance it was necessary to generate summary *cis*-eQTL statistics per GWAS SNP (SNP, expression-unit, β, standard error of β, and *P*-value of association) for integration analysis. Per quantification type (microarray, RNA-Seq gene-level, and exon-level), each GWAS SNP was subject to *cis*-eQTL analysis against all expression-units within +/-1Mb using no *P*-value threshold. If the GWAS SNP was not found in an eQTL dataset, the most highly correlated, closest tag SNP with $r^2 \geq 0.7$, common to all datasets, was used as its proxy (**Table 1**). Adjustment for multiple testing of *cis*-eQTL results per quantification type were undertaken using FDR with $q < 0.05$ deemed significant.

Microarray cis-eQTL mapping

We used the Genevar (GENe Expression VARiation) portal to generate summary-level *cis*-eQTL results (50). We ran the association between normalized expression data of the 777 individuals and each GWAS SNP implementing the external algorithm option (two-step mixed model–based score test). In total 768 probes (559) genes, were tested.

RNA-Seq (gene-level) cis-eQTL mapping

RNA-Seq gene-level quantification was provided as residualized read-counts (effect of family structure and other covariates regressed out). We had full genetic data for 683 individuals and performed the analysis of each GWAS SNP against the transformed residuals using the linear-model function within the MatrixeQTL R package (93). 520 genes were tested against in *cis*.

RNA-Seq (exon-level) cis-eQTL mapping

P-values from the association of all SNPs against exon-level quantifications for 765 individuals using linear-regression were provided. We generated the t-statistic using the lower-tail quantile function t-distribution function in R with 763 degrees of freedom. The

standard error and β were derived from the t-statistic. We then extracted the summary *cis*-eQTL results for each GWAS SNP. 4,786 exons, corresponding to 716 genes for testing.

SLE candidate-causal *cis*-eQTL classification

Conditional analysis

We used the COJO (conditional and joint genome-wide association analysis) function of the GCTA (Genome-wide Complex Trait Analysis) application to determine whether the GWAS SNP had an independent effect on expression from that of the best *cis*-eQTL (55). For each significant association (q<0.05), we re-performed the analysis using all SNPs within +/-1Mb of the expression-unit in hand. We used the available genotype information of the 683 TwinsUK individuals to extract allele coding along with the MAF, and integrated this with the *cis*-eQTL summary data. We discarded SNPs with: MAF < 0.05, imputation call-rates < 0.8, and HWE $P<1x10^{-04}$. We used these individuals as the reference panel to calculate local pairwise linkage disequilibrium (LD) between variants. Per significant association, all *cis*-eQTLs were conditioned on by the best *cis*-eQTL. We then extracted the conditional *P*-value of the GWAS SNP and considered associations to be independent to the best *cis*-eQTL if $P_{cond}<0.05$.

Colocalisation Analysis

We employed the 'coloc' Bayesian statistical method using summary data implemented in R to test for colocalisation between *cis*-eQTL and disease causal variants derived from the GWAS (56). The method makes the assumption of there being a single causal variant for each trait (disease association and gene-expression from two separate studies) per locus and calculates the posterior probabilities under five different causal variant hypotheses: association with neither trait (H0), association with one trait but not the other (H1, H2), association with both traits but from independent signals, and association with both traits with a shared causal signal (H4). We extracted the necessary SNP statistics for the disease-associated regions from our own GWAS and applied the same SNP filters used in the

conditional analysis. We tested for colocalisation between the GWAS summary data and *cis*-eQTL data for each significant association within a +/-1Mb window of the GWAS SNP. We assigned the prior probabilities, p1 and p2 (SNP is associated with GWAS and gene expression respectively), as 1 x 10^{-04} i.e. 1 in 10,000 SNPs are causal to either trait, with p12 (SNP is associated with both traits) as 1x10^{-06} or 1 in 100 SNPs associated with one trait are also associated with the other. For each *cis*-eQTL association colocalisation test, if the posterior probability PP3 (two distinct causal variants, one for each trait) is greater than PP4 (single causal variant common to both traits), then greater posterior support is given to the hypothesis that independent causal variants exist in both traits and thus the eQTL is unlikely to be attributed to SLE genetic association.

Definition of SLE candidate-causal cis-eQTL and eGene

We defined a GWAS SNP as an SLE candidate-causal *cis*-eQTL if it met the following criteria: significant post-multiple testing adjustment ($q < 0.05$), not independent to the best *cis*-eQTL from conditional analysis ($P_{cond} > 0.05$), and supporting evidence of a shared causal variant between gene expression and the primary GWAS signal based on colocalisation (PP3 < PP4). The gene whose expression is modulated by the candidate-causal eQTL is defined as an SLE candidate-causal eGene (**Fig. 1**).

Validation of LCL SLE candidate-causal cis-eQTLs in whole blood

Cis-eQTL summary data from whole blood at RNA-Seq exon-level were made available for 384 individuals of the 856 TwinsUK cohort individuals (**Table 2**). Expression profiling and genotyping were identical to that as described for LCLs. We applied the same methodology to this dataset to generate full *cis*-eQTL summary statistics, perform conditional and colocalisation analysis, and classify SLE candidate-causal eQTLs and associated eGenes (**Fig. 1**). In total, 3,793 exons were tested against, corresponding to 654 genes.

Geuvadis SLE candidate-causal *cis*-asQTL analysis

We investigated SLE disease-associated alternative splicing QTLs (asQTLs) using European samples from the raw alignment files of the Geuvadis (35) 1000 Genomes RNA-Seq project profiled in LCLs (**Table 2**). Genotype data and read-alignments were downloaded from ArrayExpress for the 373 Europeans (comprising 91 CEU, 95 FIN, 94 GBR, and 93 TSI). We performed PCA on chromosome 20 using the R/Bioconductor package SNPRelate (94) and decided to include the first three principle components as covariates in the eQTL model as well as the binary imputation status (mixture of Phase 1 and Phase 2 imputed individuals). We removed SNPs with MAF<0.05, imputation call-rates<0.8, and HWE $P<1x10^{-04}$. We removed non-uniquely mapped, non-properly paired reads, and reads with more than eight mismatches for read and mate using Samtools(95). We used the Altrans (96) method against GENCODE v10 to generate relative quantifications (link-counts) of splicing events; which in brief, utilizes split and paired-end reads to count links between exon-boundaries, which themselves are created by flattening the annotation into unique non-redundant exon-groups. Following PCA of the link-counts, we decided to normalize all link-counts with the first 10 principle components then removed exon-boundaries with zero links in more than 10% of individuals. Link-counts were converted to link-fractions (coverage of the link over the sum of the coverage of all the links that the first exon makes) and merged in both 5′-3′ and 3′-5′ directions. Per GWAS SNP we performed *cis*-eQTL analysis against the normalized link-fractions in MatrixeQTL with a linear-model (93). 33,039 link-fractions were tested against corresponding to 817 genes in total. After FDR multiple-testing adjustment we considered associations with q<0.05 as significant. As full genetic and expression data were available, we decided to use the Regulatory Trait Concordance (RTC) method to assess the likelihood of a shared functional variant between the GWAS SNP and the asQTL signal (48). For each significant asQTL association we extracted the residuals of the linear-regression of the best *cis*-eQTL against normalized link-fractions and re-performed the analysis using all SNPs within the defined hotspot interval against this pseudo-phenotype. The RTC score was defined as $(N_{SNPs}\text{-}Rank_{GWAS\ SNP})/N_{SNPs}$ where N_{SNPs} is the number of SNPs in the interval, and $Rank_{GWAS\ SNP}$ is the rank of the

GWAS SNP association *P*-value against all other SNPs in the interval. We classified an SLE candidate-causal *cis*-asQTL as a GWAS SNP with a significant association ($q < 0.05$) with link-fraction quantification and an RTC score > 0.9.

Statistical analysis and data visualisation

We performed statistical analysis, graphics and data handling in R version 3.2.0 and ggplot2. Genetic plots were generated using LocusZoom v1.1 (97). Karyotype diagrams were modified from Ensembl (98). GWAS association plots and gene annotation graphic visualisations were created using the UCSC Genome Browser(99).

Acknowledgements

DSCG and TJV were awarded an Arthritis Research UK PhD Studentship (for CO) (20332) "SLE pathogenesis: the molecular basis for multiple association signals in *IKZF2*" to undertake the analyses described in this manuscript. DSCG, DLM and TJV were awarded an Arthritis Research UK Project Grant (20265) "Post GWAS: A Functional Genomic Analysis of the SLE Susceptibility Gene *IKZF1*", which supported AC. TJV received funds from China Scholarship Council, number 201406380127 for LC. ALR was supported by MRC project grant L002604/1 "Functional Genomics of SLE: A Trans-ancestral approach" and an ARUK project grant (20580) "Targeted DNA sequencing in Indian and European samples contributes to causal allele identification for lupus". KSS was awarded funding by the MRC (MR/L01999X/1 and MR/M004422/1).

The Geuvadis 1000 Genomes RNA-Seq data was downloaded from the EBI ArrayExpress Portal (accessions E-GEUV-1). The RNA-Seq data on the TwinsUK samples in both LCLs and whole blood was made available through the EUROBATS project (EGAS00001000805 from the European Genome-Phenome Archive). The expression microarray data on the TwinsUK samples was downloaded from the EBI ArrayExpress Portal (E-TABM_1140).

The TwinsUK study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR.

Conflict of Interest Statement

The authors declare that there are no conflicts of interest.

References

1. Hindorff,L.A., Sethupathy,P., Junkins,H.A., Ramos,E.M., Mehta,J.P., Collins,F.S. and Manolio,T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 9362–7.

2. Maurano,M.T., Humbert,R., Rynes,E., Thurman,R.E., Haugen,E., Wang,H., Reynolds,A.P., Sandstrom,R., Qu,H., Brody,J., *et al.* (2012) Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science (80-.).*, **337**, 1190–1195.

3. The ENCODE Project Consortium, Dunham,I., Kundaje,A., Aldred,S.F., Collins,P.J., Davis,C. a, Doyle,F., Epstein,C.B., Frietze,S., Harrow,J., *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.

4. Kundaje,A., Meuleman,W., Ernst,J., Bilenky,M., Yen,A., Heravi-Moussavi,A., Kheradpour,P., Zhang,Z., Wang,J., Ziller,M.J., *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.

5. Forrest,A.R.R., Kawaji,H., Rehli,M., Kenneth Baillie,J., de Hoon,M.J.L., Haberle,V., Lassmann,T., Kulakovskiy,I. V., Lizio,M., Itoh,M., *et al.* (2014) A promoter-level mammalian expression atlas. *Nature*, **507**, 462–470.

6. Nicolae,D.L., Gamazon,E., Zhang,W., Duan,S., Dolan,M.E. and Cox,N.J. (2010) Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.*, **6**, e1000888.

7. Stranger,B.E., Montgomery,S.B., Dimas,A.S., Parts,L., Stegle,O., Ingle,C.E., Sekowska,M., Smith,G.D., Evans,D., Gutierrez-Arcelus,M., *et al.* (2012) Patterns of cis regulatory variation in diverse human populations. *PLoS Genet.*, **8**, e1002639.

8. Spielman,R.S., Bastone,L.A., Burdick,J.T., Morley,M., Ewens,W.J. and Cheung,V.G.

(2007) Common genetic variants account for differences in gene expression among ethnic groups. *Nat. Genet.*, **39**, 226–231.

9. Grundberg,E., Small,K.S., Hedman,Å.K., Nica,A.C., Buil,A., Keildson,S., Bell,J.T., Yang,T.-P., Meduri,E., Barrett,A., *et al.* (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.*, **44**, 1084–1089.

10. Fairfax,B.P., Makino,S., Radhakrishnan,J., Plant,K., Leslie,S., Dilthey,A., Ellis,P., Langford,C., Vannberg,F.O. and Knight,J.C. (2012) Genetics of gene expression in primary immune cells identifies cell type–specific master regulators and roles of HLA alleles. *Nat. Genet.*, **44**, 502–510.

11. Naranbhai,V., Fairfax,B.P., Makino,S., Humburg,P., Wong,D., Ng,E., Hill,A.V.S. and Knight,J.C. (2015) Genomic modulators of gene expression in human neutrophils. *Nat. Commun.*, **6**, 7545.

12. Nica,A.C., Parts,L., Glass,D., Nisbet,J., Barrett,A., Sekowska,M., Travers,M., Potter,S., Grundberg,E., Small,K., *et al.* (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.*, **7**, e1002003.

13. Myers,A.J., Gibbs,J.R., Webster,J. a, Rohrer,K., Zhao,A., Marlowe,L., Kaleem,M., Leung,D., Bryden,L., Nath,P., *et al.* (2007) A survey of genetic human cortical gene expression. *Nat. Genet.*, **39**, 1494–9.

14. Field,J.M., Hazinski,M.F., Sayre,M.R., Chameides,L., Schexnayder,S.M., Hemphill,R., Samson,R. a., Kattwinkel,J., Berg,R. a., Bhanji,F., *et al.* (2010) Part 1: Executive summary: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation*, **122**, 640–657.

15. Emilsson,V., Thorleifsson,G., Zhang,B., Leonardson,A.S., Zink,F., Zhu,J., Carlson,S., Helgason,A., Walters,G.B., Gunnarsdottir,S., *et al.* (2008) Genetics of gene expression and its effect on disease. **452**.

16. Schadt,E.E., Molony,C., Chudin,E., Hao,K., Yang,X., Lum,P.Y., Kasarskis,A., Zhang,B., Wang,S., Suver,C., *et al.* (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.*, **6**, e107.

17. Hao,K., Bossé,Y., Nickle,D.C., Paré,P.D., Postma,D.S., Laviolette,M., Sandford,A., Hackett,T.L., Daley,D., Hogg,J.C., *et al.* (2012) Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.*, **8**, e1003029.

18. Li,Q., Seo,J.H., Stranger,B., McKenna,A., Pe'Er,I., Laframboise,T., Brown,M., Tyekucheva,S. and Freedman,M.L. (2013) Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell*, **152**, 633–641.

19. Ongen,H., Andersen,C.L., Bramsen,J.B., Oster,B., Rasmussen,M.H., Ferreira,P.G., Sandoval,J., Vidal,E., Whiffin,N., Planchon,A., *et al.* (2014) Putative cis-regulatory drivers in colorectal cancer. *Nature*, **512**, 87–90.

20. Zou,F., Chai,H.S., Younkin,C.S., Allen,M., Crook,J., Pankratz,V.S., Carrasquillo,M.M., Rowley,C.N., Nair,A.A., Middha,S., *et al.* (2012) Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet.*, **8**, e1002707.

21. Albert,F.W. and Kruglyak,L. (2015) The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.*, **16**, 197–212.

22. Gibson,G., Powell,J.E. and Marigorta,U.M. (2015) Expression quantitative trait locus analysis for translational medicine. *Genome Med.*, **7**, 60.

23. Barreiro,L., Tailleux,L., Pai,A., Gicquel,B., Marioni,J.C. and Gilad,Y. (2012) Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. *Proc. Natl. Acad. Sci.*, **109**, 1204–1209.

24. Fairfax,B.P., Humburg,P., Makino,S., Naranbhai,V., Wong,D., Lau,E., Jostins,L.,

Plant,K., Andrews,R., McGee,C., *et al.* (2014) Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science (80-.).*, **343**, 1246949–1246949.

25. Marioni,J.C., Mason,C.E., Mane,S.M., Stephens,M. and Gilad,Y. (2008) RNA-seq□: An assessment of technical reproducibility and comparison with gene expression arrays. 10.1101/gr.079558.108.

26. Zhang,X., Johnson,A.D., Hendricks,A.E., Hwang,S.-J., Tanriverdi,K., Ganesh,S.K., Smith,N.L., Peyser,P.A., Freedman,J.E. and O'Donnell,C.J. (2014) Genetic associations with expression for genes implicated in GWAS studies for atherosclerotic cardiovascular disease and blood phenotypes. *Hum. Mol. Genet.*, **23**, 782–795.

27. Zhang,X., Joehanes,R., Chen,B.H., Huan,T., Ying,S., Munson,P.J., Johnson,A.D., Levy,D. and O'Donnell,C.J. (2015) Identification of common genetic variants controlling transcript isoform variation in human whole blood. *Nat. Genet.*, **47**, 345–352.

28. Zhao,S., Fung-Leung,W.-P., Bittner,A., Ngo,K. and Liu,X. (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One*, **9**, e78644.

29. Majewski,J. and Pastinen,T. (2011) The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet.*, **27**, 72–79.

30. Russo,G., Zegar,C. and Giordano,A. (2003) Advantages and limitations of microarray technology in human cancer. *Oncogene*, **22**, 6497–6507.

31. Sun,W. and Hu,Y. (2013) eQTL Mapping Using RNA-seq Data. *Stat. Biosci.*, **5**, 198–219.

32. 't Hoen,P. a C., Friedländer,M.R., Almlöf,J., Sammeth,M., Pulyakhina,I., Anvar,S.Y., Laros,J.F.J., Buermans,H.P.J., Karlberg,O., Brännvall,M., *et al.* (2013) Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nat.*

Biotechnol., **31**, 1015–22.

33. Pickrell,J.K., Marioni,J.C., Pai,A.A., Degner,J.F., Engelhardt,B.E., Nkadori,E., Veyrieras,J.-B., Stephens,M., Gilad,Y. and Pritchard,J.K. (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, **464**, 768–772.

34. Montgomery,S.B., Sammeth,M., Gutierrez-Arcelus,M., Lach,R.P., Ingle,C., Nisbett,J., Guigo,R. and Dermitzakis,E.T. (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature*, **464**, 773–777.

35. Lappalainen,T., Sammeth,M., Friedländer,M.R., ’t Hoen,P. a C., Monlong,J., Rivas,M. a, Gonzàlez-Porta,M., Kurbatova,N., Griebel,T., Ferreira,P.G., *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, **501**, 506–11.

36. Battle,A., Mostafavi,S., Zhu,X., Potash,J.B., Weissman,M.M., McCormick,C., Haudenschild,C.D., Beckman,K.B., Shi,J., Mei,R., *et al.* (2014) Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.*, **24**, 14–24.

37. Lalonde,E., Ha,K.C.H., Wang,Z., Bemmo,A., Kleinman,C.L., Kwan,T., Pastinen,T. and Majewski,J. (2011) RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression. *Genome Res*, 10.1101/gr.111211.110.21.

38. Kumar,V., Westra,H.-J., Karjalainen,J., Zhernakova,D. V., Esko,T., Hrdlickova,B., Almeida,R., Zhernakova,A., Reinmaa,E., Võsa,U., *et al.* (2013) Human Disease-Associated Genetic Variation Impacts Large Intergenic Non-Coding RNA Expression. *PLoS Genet.*, **9**, e1003201.

39. Buil,A., Brown,A.A., Lappalainen,T., Viñuela,A., Davies,M.N., Zheng,H.-F., Richards,J.B., Glass,D., Small,K.S., Durbin,R., *et al.* (2014) Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat.*

Genet., **47**, 88–91.

40. Conde,L., Bracci,P.M., Richardson,R., Montgomery,S.B. and Skibola,C.F. (2013) Integrating GWAS and Expression Data for Functional Characterization of Disease-Associated SNPs: An Application to Follicular Lymphoma. *Am. J. Hum. Genet.*, **92**, 126–130.

41. Simon,L.M., Chen,E.S., Edelstein,L.C., Kong,X., Bhatlekar,S., Rigoutsos,I., Bray,P.F. and Shaw,C.A. (2016) Integrative Multi-omic Analysis of Human Platelet eQTLs Reveals Alternative Start Site in Mitofusin 2. *Am. J. Hum. Genet.*, 10.1016/j.ajhg.2016.03.007.

42. Li,H., Pouladi,N., Achour,I., Gardeux,V., Li,J., Li,Q., Zhang,H.H., Martinez,F.D., Skip Garcia,J.G.N. and Lussier,Y.A. (2015) eQTL networks unveil enriched mRNA master integrators downstream of complex disease-associated SNPs. *J. Biomed. Inform.*, **58**, 226–234.

43. Huang,L., Xu,W., Yan,D., Dai,L. and Shi,X. (2016) Identification of expression quantitative trait loci of RPTOR for susceptibility to glioma. *Tumor Biol.*, **37**, 2305–2311.

44. Jeong,S., Patel,N., Edlund,C.K., Hartiala,J., Hazelett,D.J., Itakura,T., Wu,P.C., Avery,R.L., Davis,J.L., Flynn,H.W., *et al.* (2015) Identification of a novel mucin gene HCG22 associated with steroid-induced ocular hypertension. *Investig. Ophthalmol. Vis. Sci.*, **56**, 2737–2748.

45. Moffatt,M.F., Kabesch,M., Liang,L., Dixon,A.L., Strachan,D., Heath,S., Depner,M., von Berg,A., Bufe,A., Rietschel,E., *et al.* (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*, **448**, 470–473.

46. Barrett,J.C., Hansoul,S., Nicolae,D.L., Cho,J.H., Duerr,R.H., Rioux,J.D., Brant,S.R., Silverberg,M.S., Taylor,K.D., Barmada,M.M., *et al.* (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.*, **40**,

955–962.

47. Li,X., Hastie,A.T., Hawkins,G. a., Moore,W.C., Ampleford,E.J., Milosevic,J., Li,H., Busse,W.W., Erzurum,S.C., Kaminski,N., *et al.* (2015) eQTL of bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma genes. *Allergy*, **70**, n/a–n/a.

48. Nica,A.C., Montgomery,S.B., Dimas,A.S., Stranger,B.E., Beazley,C., Barroso,I. and Dermitzakis,E.T. (2010) Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.*, **6**, e1000895.

49. Veyrieras,J.-B., Kudaravalli,S., Kim,S.Y., Dermitzakis,E.T., Gilad,Y., Stephens,M. and Pritchard,J.K. (2008) High-Resolution Mapping of Expression-QTLs Yields Insight into Human Gene Regulation. *PLoS Genet.*, **4**, e1000214.

50. Yang,T.-P., Beazley,C., Montgomery,S.B., Dimas,A.S., Gutierrez-Arcelus,M., Stranger,B.E., Deloukas,P. and Dermitzakis,E.T. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics*, **26**, 2474–2476.

51. The GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–5.

52. Harley,I.T.W., Kaufman,K.M., Langefeld,C.D., Harley,J.B. and Kelly,J.A. (2009) Genetic susceptibility to SLE: new insights from fine mapping and genome-wide association studies. *Nat. Rev. Genet.*, **10**, 285–90.

53. Cui,Y., Sheng,Y. and Zhang,X. (2013) Genetic susceptibility to SLE: Recent progress from GWAS. *J. Autoimmun.*, **41**, 25–33.

54. Deng,Y. and Tsao,B.P. (2014) Advances in lupus genetics and epigenetics. *Curr. Opin. Rheumatol.*, **26**, 1–11.

55. Yang,J., Lee,S.H., Goddard,M.E. and Visscher,P.M. (2011) GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.*, **88**, 76–82.

56. Giambartolomei,C., Vukcevic,D., Schadt,E.E., Franke,L., Hingorani,A.D., Wallace,C. and Plagnol,V. (2014) Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.*, **10**, e1004383.

57. Kozyrev,S. V, Abelson,A.-K., Wojcik,J., Zaghlool,A., Linga Reddy,M.V.P., Sanchez,E., Gunnarsson,I., Svenungsson,E., Sturfelt,G., Jönsen,A., *et al.* (2008) Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat. Genet.*, **40**, 211–216.

58. Lewis,M.J., Vyse,S., Shields,A.M., Boeltz,S., Gordon,P.A., Spector,T.D., Lehner,P.J., Walczak,H. and Vyse,T.J. (2015) UBE2L3 polymorphism amplifies NF-κB activation and promotes plasma cell development, linking linear ubiquitination to multiple autoimmune diseases. *Am. J. Hum. Genet.*, **96**, 221–234.

59. Noble,J.A., White,A.M., Lazzeroni,L.C., Valdes,A.M., Mirel,D.B., Reynolds,R., Grupe,A., Aud,D., Peltz,G. and Erlich,H.A. Associated With Type 1 Diabetes. **06**, 1579–1582.

60. Liu,Y., Wang,X., Cheng,X., Lu,Y. and Wang,G. (2015) Skp1 in lung cancer□: Clinical significance and therapeutic efficacy of its small molecule inhibitors. **6**.

61. Wu,J.Q., Seay,M., Schulz,V.P., Hariharan,M., Tuck,D., Lian,J., Du,J., Shi,M., Ye,Z., Gerstein,M., *et al.* (2012) Tcf7 is an important regulator of the switch of self-renewal and differentiation in a multipotential hematopoietic cell line. *PLoS Genet.*, **8**.

62. Zhou,W., Wei,W. and Sun,Y. (2013) Genetically engineered mouse models for functional studies of SKP1-CUL1-F-box-protein (SCF) E3 ubiquitin ligases. *Cell Res*, **23**, 599–619.

63. Bentham,J., Morris,D.L., Cunninghame Graham,D.S., Pinder,C.L., Tombleson,P.,

Behrens,T.W., Martín,J., Fairfax,B.P., Knight,J.C., Chen,L., *et al.* (2015) Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.*, 10.1038/ng.3434.

64. Westra,H.-J., Peters,M.J., Esko,T., Yaghootkar,H., Schurmann,C., Kettunen,J., Christiansen,M.W., Fairfax,B.P., Schramm,K., Powell,J.E., *et al.* (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.*, **45**, 1238–43.

65. Ward,L.D. and Kellis,M. (2012) HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, 930–934.

66. Salloum,R., Franek,B.S., Kariuki,S.N., Rhee,L., Mikolaitis,R. a., Jolly,M., Utset,T.O. and Niewold,T.B. (2010) Genetic variation at the IRF7/PHRF1 locus is associated with autoantibody profile and serum interferon-alpha activity in lupus patients. *Arthritis Rheum.*, **62**, 553–561.

67. Wang,T.J., Zhang,F., Richards,J.B., Kestenbaum,B., van Meurs,J.B., Berry,D., Kiel,D.P., Streeten,E.A., Ohlsson,C., Koller,D.L., *et al.* (2010) Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet*, **376**, 180–188.

68. Yap,K.S., Northcott,M., Hoi, a. B.-Y., Morand,E. and Nikpour,M. (2015) Association of low vitamin D with high disease activity in an Australian systemic lupus erythematosus cohort. *Lupus Sci. Med.*, **2**, e000064–e000064.

69. Wassif,C. a, Zhu,P., Kratz,L., Krakowiak,P. a, Battaile,K.P., Weight,F.F., Grinberg,A., Steiner,R.D., Nwokoro,N.A., Kelley,R.I., *et al.* (2001) Biochemical, phenotypic and neurophysiological characterization of a genetic mouse model of RSH/Smith--Lemli--Opitz syndrome. *Hum. Mol. Genet.*, **10**, 555–564.

70. Shah,D., Mahajan,N., Sah,S., Nath,S.K. and Paudyal,B. (2014) Oxidative stress and its biomarkers in systemic lupus erythematosus. *J. Biomed. Sci.*, **21**, 23.

71. Luo,X., Yang,W., Ye,D.-Q., Cui,H., Zhang,Y., Hirankarn,N., Qian,X., Tang,Y., Lau,Y.L., de Vries,N., *et al.* (2011) A functional variant in microRNA-146a promoter modulates its expression and confers disease risk for systemic lupus erythematosus. *PLoS Genet.*, **7**, e1002128.

72. Hom,G., Ph,D., Graham,R.R., Modrek,B., Taylor,K.E., Ortmann,W., Garnier,S., Lee,A.T., Chung,S.A., Ferreira,R.C., *et al.* (2008) Association of Systemic Lupus Erythematosus with. *Genetics*.

73. Guthridge,J.M., Lu,R., Sun,H., Sun,C., Wiley,G.B., Dominguez,N., MacWana,S.R., Lessard,C.J., Kim-Howard,X., Cobb,B.L., *et al.* (2014) Two functional lupus-associated BLK promoter variants control cell-type- and developmental-stage-specific transcription. *Am. J. Hum. Genet.*, **94**, 586–598.

74. Kim,H., Barnitz,R.A., Kreslavsky,T., Brown,F.D., Moffett,H., Lemieux,M.E., Kaygusuz,Y., Meissner,T., Holderried,T.A.W., Chan,S., *et al.* (2009) egulatory T cells (CD4 and CD8 T.

75. Baine,I., Basu,S., Ames,R., Sellers,R.S. and Macian,F. (2013) Helios induces epigenetic silencing of IL2 gene expression in regulatory T cells. *J. Immunol.*, **190**, 1008–16.

76. Getnet,D., Grosso,J.F., Goldberg,M. V., Harris,T.J., Yen,H.R., Bruno,T.C., Durham,N.M., Hipkiss,E.L., Pyle,K.J., Wada,S., *et al.* (2010) A role for the transcription factor Helios in human CD4+CD25+ regulatory T cells. *Mol. Immunol.*, **47**, 1595–1600.

77. Isakson,P., Holland,P. and Simonsen,A. (2013) The role of ALFY in selective autophagy. *Cell Death Differ.*, **20**, 12–20.

78. Yang,W., Shen,N., Ye,D.Q., Liu,Q., Zhang,Y., Qian,X.X., Hirankarn,N., Ying,D.,

Pan,H.F., Mok,C.C., *et al.* (2010) Genome-wide association study in asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet.*, **6**.

79. McLaren,W., Pritchard,B., Rios,D., Chen,Y., Flicek,P. and Cunningham,F. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, **26**, 2069–2070.

80. Graham,R.R., Kozyrev,S. V, Baechler,E.C., Reddy,M.V.P.L., Plenge,R.M., Bauer,J.W., Ortmann,W. a, Koeuth,T., González Escribano,M.F., Pons-Estel,B., *et al.* (2006) A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat. Genet.*, **38**, 550–555.

81. Graham,R.R., Kyogoku,C., Sigurdsson,S., Vlasova,I.A., Davies,L.R.L., Baechler,E.C., Plenge,R.M., Koeuth,T., Ortmann,W.A., Hom,G., *et al.* (2007) Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 6758–63.

82. Lappalainen,T. (2015) Functional genomics bridges the gap between quantitative genetics and molecular biology. *Genome Res.*, **25**, 1427–1431.

83. Battle, a. and Montgomery,S.B. (2014) Determining causality and consequence of expression quantitative trait loci. *Hum. Genet.*, **133**, 727–735.

84. Jäger,R., Migliorini,G., Henrion,M., Kandaswamy,R., Speedy,H.E., Heindl,A., Whiffin,N., Carnicer,M.J., Broome,L., Dryden,N., *et al.* (2015) Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat. Commun.*, **6**, 6178.

85. Zhao,Z., Tavoosidana,G., Sjölinder,M., Göndör,A., Mariano,P., Wang,S., Kanduri,C., Lezcano,M., Sandhu,K.S., Singh,U., *et al.* (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and

interchromosomal interactions. *Nat. Genet.*, **38**, 1341–1347.

86. Kelsell,D.P., Norgett,E.E., Unsworth,H., Teh,M.-T., Cullup,T., Mein,C.A., Dopping-Hepenstal,P.J., Dale,B.A., Tadini,G., Fleckman,P., *et al.* (2005) Mutations in ABCA12 underlie the severe congenital skin disease harlequin ichthyosis. *Am. J. Hum. Genet.*, **76**, 794–803.

87. Steidl,U., Steidl,C., Ebralidze,A., Chapuy,B., Han,H., Will,B., Rosenbauer,F., Becker,A., Wagner,K., Koschmieder,S., *et al.* (2007) A distal single nucleotide polymorphism alters long-range regulation of the PU . 1 gene in acute myeloid leukemia. *J. Clin. Invest.*, **117**, 2611–2620.

88. Freedman,M.L., Monteiro,A.N.A., Gayther,S.A., Coetzee,G.A., Risch,A., Plass,C., Casey,G., De Biasi,M., Carlson,C., Duggan,D., *et al.* (2011) Principles for the post-GWAS functional characterization of cancer risk loci. *Nat. Genet.*, **43**, 513–518.

89. Graveley,B.R. (2008) The haplo-spliceo-transcriptome: common variations in alternative splicing in the human population. *Trends Genet.*, **24**, 5–7.

90. Wang,G.-S. and Cooper,T.A. (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.*, **8**, 749–761.

91. Li,Y.I., van de Geijn,B., Raj,A., Knowles,D.A., Petti,A.A., Golan,D., Gilad,Y. and Pritchard,J.K. (2016) RNA splicing is a primary link between genetic variation and disease. *Science*, **352**, 600–4.

92. Wang,J.H., Avitahl,N., Cariappa,A., Friedrich,C., Ikeda,T., Renold,A., Andrikopoulos,K., Liang,L., Pillai,S., Morgan,B.A., *et al.* (1998) Aiolos regulates B cell activation and maturation to effector state. *Immunity*, **9**, 543–553.

93. Shabalin,A.A. (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, **28**, 1353–1358.

94. Zheng,X., Levine,D., Shen,J., Gogarten,S.M., Laurie,C. and Weir,B.S. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.

95. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Subgroup,1000 Genome Project Data Processing (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

96. Ongen,H. and Dermitzakis,E.T. (2015) Alternative Splicing QTLs in European and African Populations. *Am. J. Hum. Genet.*, **97**, 567–575.

97. Pruim,R.J., Welch,R.P., Sanna,S., Teslovich,T.M., Chines,P.S., Gliedt,T.P., Boehnke,M., Abecasis,G.R. and Willer,C.J. (2010) LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.

98. Cunningham,F., Amode,M.R., Barrell,D., Beal,K., Billis,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fitzgerald,S., *et al.* (2015) Ensembl 2015. *Nucleic Acids Res.*, **43**, D662–D669.

99. Kent,W.J., Sugnet,C.W., Furey,T.S. and Roskin,K.M. (1976) The Human Genome Browser at UCSC W. *J. Med. Chem.*, **19**, 1228–31.

Legends to figures

Fig. 1: Two-stage *cis*-eQTL annotation pipeline for SLE susceptibility loci.

SLE susceptibility variants (**Table 1**) were annotated using residualized expression or summary-level eQTL statistics from three expression datasets: microarray probe-level expression data, and both gene-level and exon-level RNA-Seq quantifications. Each expression dataset was generated from LCLs from individuals of the TwinsUK cohort. **A**) We undertook *cis*-eQTL analysis of +/-1Mb intervals around each SNP and associations with q<0.05 after FDR adjustment were taken forward. **B**) Summary-level data from significant *cis*-eQTLs were tested for evidence of a shared causal variant using firstly conditional analysis using the TwinsUK genetic data as a reference panel, then colocalisation analysis to test for a single causal variant common to both traits. Associations passing these thresholds (described fully in methods) were classified as candidate-causal eQTLs and eGenes. Summary results per quantification type for significant and candidate-causal associations are shown in **Table 3** for microarray, **Table 4** for RNA-Seq (gene-level), and **Table 5** for RNA-Seq (exon-level). Full summary results are available in **Table S1**, **S2**, and **S3** respectively.

Fig. 2: Candidate-causal eQTLs and eGenes detected across quantification types. Candidate-causal eQTLs and eGenes detected using microarray (probe-level), RNA-Seq (gene-level), and RNA-Seq (exon-level) quantifications from the TwinsUK datasets are represented in their genomic context (**Tables 3**, **4**, and **5** respectively). Only chromosomes harbouring one or more candidate-causal eQTL or eGene per quantification type are shown.

Fig. 3: RNA-Seq gene-level and exon-level analysis implicate *NADSYN1* as a novel candidate-causal eGene.

A) *Cis*-eQTL analysis of GWAS SNP rs3794060 reveals the risk variant [C] leads to substantial down-regulation at the gene-level of novel susceptibility gene *NADSYN1* in the TwinsUK cohort that was not detected using microarray. **B**) The same analysis using exon-level quantification leads to the inference of the gene-level effect being driven by considerable expression disruption of two meta-exons of *NADSYN1* (meta-exon 11 and meta-exon 12). Association *P*-values of rs3794060 against exon quantifications are plotted with reference to the specific exon in the collapsed-gene model of *NADSYN1* (all annotated transcripts combined).

Fig. 4: Non-coding candidate-causal eGenes detected using exon-level RNA-Seq.

The three figures denote the eQTLs and corresponding non-coding eGenes identified from *cis*-eQTL analysis of GWAS SNPs against TwinsUK exon-level quantifications in LCLs. The top panels display the signal from the GWAS association plotted as $-\log_{10}(P)$, below which the exon-level eQTL *P*-values for the effects showing colocalisation with the GWAS signal are illustrated. The bar across the middle of the figure denotes the boundaries of the eQTL, below which there is a panel showing the association *P*-value of the GWAS SNP against the candidate-causal non-coding exons. The bottom panel shows LCL RNA-Seq alignments from ENCODE to show that the regions containing the candidate-causal eQTLs are expressed. **A**) GWAS SNP rs2431697 is a candidate-causal eQTL for non-coding eGene *MIR146A*. **B**) GWAS SNP rs2736340 is a candidate-causal eQTL for non-coding eGenes *RP11-148O21.4* and *RP11-148O21.2*. **C**) GWAS SNP rs3794060 is a candidate-causal eQTL for non-coding eGene *RP11-660L16.2*.

Fig. 5: Identification of novel eGene *IKZF2* and potential causal mechanism using RNA-Seq splice-junction quantification.

Cis-asQTL analysis of GWAS SNP rs3768792 against splice-junction quantifications classified *IKZF2* as a candidate-causal eGene with risk variant [G] causing upregulation of the exon 6A–exon 6B junction that is unique to truncated isoform ENST00000413091. **A**) GWAS association signal across the *IKZF2* locus (chr2q34), tagged by rs3768792 localised in the 3′-UTR of *IKZF2*. *Cis*-asQTL association signal of rs3768792 against splice-junction quantification of exon 6A–exon 6B shows significance and colocalisation with the GWAS signal. **B**) The exon 6A–exon 6B junction is unique to truncated isoform ENST00000413091. Exon 6B harbours a premature stop-codon and therefore is not translated into the full-length protein that contains the dimerization domains in exon 8. **C**) Close-up of the exon 6A–exon 6B junction and association (P=3.80E^{-05}) with GWAS SNP rs3768792. A potential causal asQTL in near-perfect LD was identified that is located within the polypyrimidine tract of the junction and may induce splicing (rs2291241, P=1.70^{E-05}).

Fig. 6: Identification of potential disease-associated mechanisms at the *WDFY4* susceptibility locus using asQTL mapping.

A) Our SLE GWAS indicates *WDFY4* as the candidate gene at the chr10q11.23 locus tagged by intronic variant rs2663052, as well as the missense coding variant rs7097397 in exon 31 that is in strong LD. *Cis*-eQTL showed rs2663052 is correlated with upregulation of the exon 34A–34B junction of *WDFY4* (signal is colocalised with GWAS) that is unique to the short isoform (ENST00000374161). This isoform lacks the two enzymatic WD40 domains of the full length isoform (ENST00000325239). **B**) Two potential functional mechanisms may occur when harbouring the risk haplotype that carries both risk alleles. Firstly an Arg to Gln amino-acid substitution by rs7097397 in exon 31 that is shared by both the canonical and short isoforms of *WDFY4*, and secondly an upregulation of the short isoform (P=3.31^{E-19})

that lacks functional domains, caused by rs2663052 or correlated variants, with corresponding down-regulation of the full-length isoform (P=3.01^{E-06}).

Fig. 7: Heatmap of candidate-causal eQTLs and eGenes detected across all expression-quantification types.

Heatmap of all candidate-causal *cis*-eQTL associations across the four quantification types (microarray, RNA-Seq gene-level, RNA-Seq exon-level, and RNA-Seq splice-junction level). The first column is the key showing the relative P value of the eQTLs within each platform. For the platform-specific columns, if an eQTL-eGene association is candidate-causal in at least one quantification type, the data is displayed across all platforms. Rows are or ordered by decreasing cumulative significance across quantification types. To normalize across quantification types, relative significance of each association per column was calculated as the $-\log_2 (P/P_{max})$; where P_{max} is the most significant association per quantification type. If an association is deemed to be candidate-causal within a particular profiling-type, it is highlighted with an asterisk.

Supporting Information

S1 Table. All significant eQTLs (q < 0.05) and associated eGenes detected at microarray (probe-level) with conditional and colocalisation results.

S2 Table. All significant eQTLs (q < 0.05) and associated eGenes detected at RNA-Seq (gene-level) with conditional and colocalisation results.

S3 Table. All significant eQTLs (q < 0.05) and associated eGenes detected at RNA-Seq (exon-level) with conditional and colocalisation results.

S1 Fig. Number of eQTL discoveries per quantification type. Including significant associations (q<0.05), and candidate-causal associations (significant, and not-independent and colocalised with GWAS).

S2 Fig. Number of eGene discoveries per quantification type. Including significant associations (q<0.05), and candidate-causal associations (significant, and not-independent and colocalised with GWAS).

S3 Fig. Shared candidate-causal eQTLs per quantification type.

S4 Fig. Shared candidate-causal eGenes per quantification type.

S5 Fig. Ratio of eQTLs to candidate-causal eGenes per quantification type.

S6 Fig. Gene-level and exon-level candidate-causal associations with *TCF7* and *SKP1* against rs7726414. *Cis*-eQTL analysis at gene-level and exon-level using RNA-Seq implicate novel SLE-associated eGenes *TCF7* and *SKP1*.

S7 Fig. High gene-density over associated variants tagged by GWAS SNP rs12802200. The six candidate-causal eGenes against rs1280220 discovered using RNA-Seq at either quantification method are marked with an asterisk.

S8 Fig. Candidate-causal eGenes *DHCR7* and *NADSYN1* for rs3794060, and non-

coding eGene *RP11-660L16.2*. The GWAS SNP rs3794060 is a candidate-causal eGene for *DHCR7* and *NADSYN1*, and also the non-coding eGene *RP11-660L16.2* at exon-level; which is located between *DCHR7* and *NADSYN1*.

S9 Fig. Candidate-causal eGenes *FAM167A*, *BLK* and two non-coding RNAs (*RP11-138021.4* and *RP11-138021.2*) driven by rs2736340. Associated variant rs2736340 lies in a region of intense regulatory chromatin marks located in the bi-directional promoter of *BLK* and *FAM167A* which are both candidate-causal eGenes at RNA-Seq gene-level. At exon-level, the non-coding eGenes *RP11-138021.4* and *RP11-138021.2* are also candidate-causal.

S10 Fig. Effect-size correlation of GWAS SNP associations with matched *cis*-exons between LCL and whole-blood.

S4 Table. All significant eQTLs (q < 0.05) and associated eGenes detected at RNA-Seq (exon-level) l with conditional and colocalisation results in whole-blood.

S11 Fig. Candidate-causal eQTLs and eGenes in whole blood. Comparison with LCL associations.

S12 Fig. Whole-blood exon-level eQTL effect on *BANK1* exon 2. Correlation between GWAS SNP and the best whole-blood eQTL for *BANK1* exon 2. Both are highly correlated with known branch-point SNP. All are weakly correlated with best LCL eQTL for *BANK1* exon 2.

S13 Fig. Exon-level eQTL analysis of *NADSYN1* in whole-blood and LCLs reveal near-identical splicing effect. Meta-exons 11 and 12 are substantially disrupted with reference to GWAS SNP rs3794060 in both LCLs and whole-blood.

S5 Table. All significant asQTLs (q < 0.05) and associated eGenes detected at RNA-Seq (splice-junction level).

S14 Fig. Proposed splicing mechanism of *NADSYN1* caused by risk haplotype tagged by rs3794060. *NADSYN1* Ensembl transcript annotation displayed. *Cis*-asQTL identified the meta-exon 10 to meta-exon 12 junction is upregulated with risk allele [C] and consequently the meta-exon 11 to meta-exon 12 junction is downregulated.

S6 Table. Comparison of eQTLs and eGenes for SLE risk alleles between previously reported in microarray studies and from RNA-Seq in current study.

Abbreviations

GWAS: Genome-Wide Association Study

eQTL: expression Quantitative Trait Loci

RNA-Seq: RNA-Sequencing

SLE: Systemic Lupus Erythematosus

asQTL: alternative-splicing Quantitative Trait Loci

SNP: Single Nucleotide Polymorphism

LCL: Lymphoblastoid cell line

Fig 1.

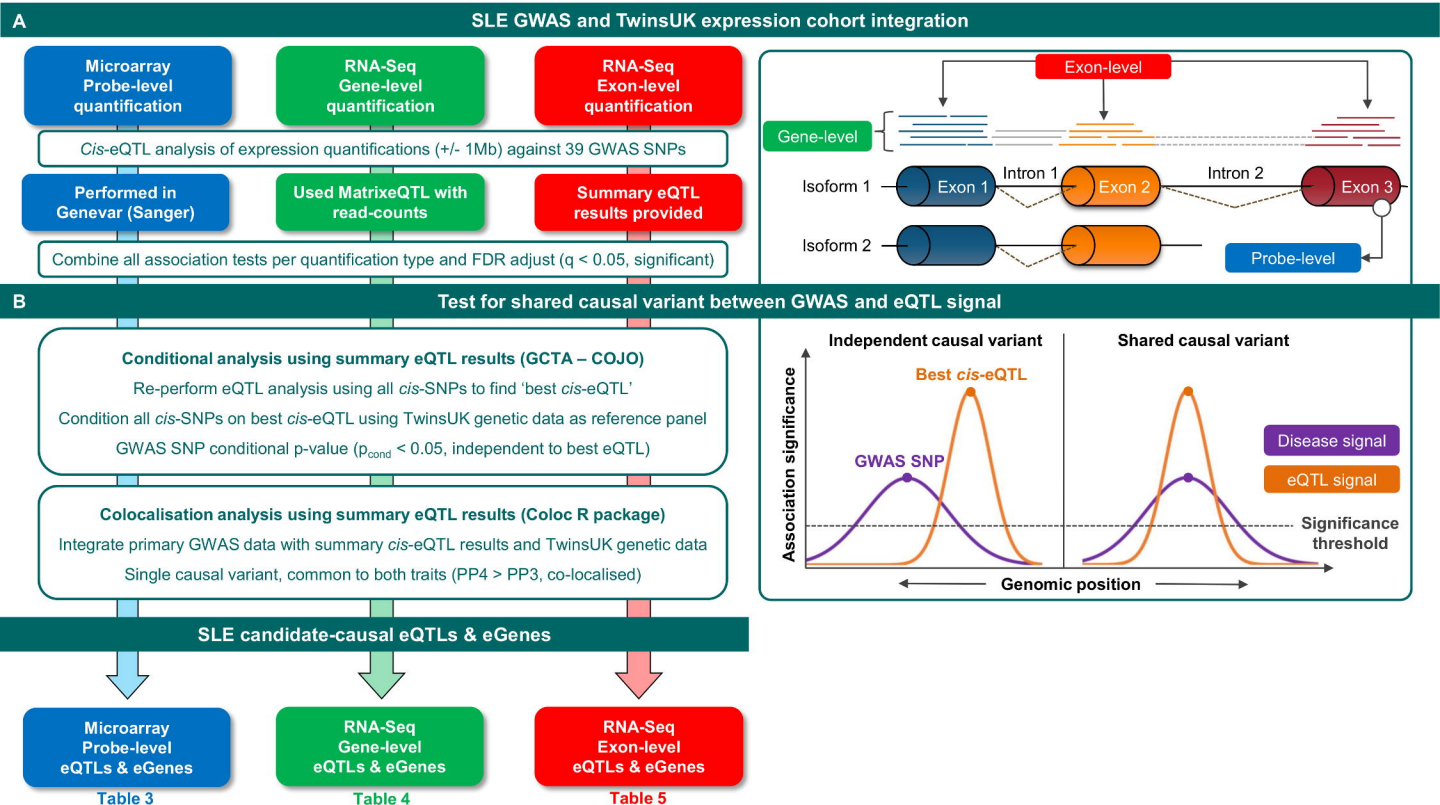

A
SLE GWAS and TwinsUK expression cohort integration
Microarray Probe-level quantification
RNA-Seq Gene-level quantification
RNA-Seq Exon-level quantification
Cis-eQTL analysis of expression quantifications (+/- 1Mb) against 39 GWAS SNPs
Performed in Genevar (Sanger)
Used MatrixeQTL with read-counts
Summary eQTL results provided
Combine all association tests per quantification type and FDR adjust (q < 0.05, significant)
Exon-level
Gene-level
Isoform 1
Exon 1
Intron 1
Exon 2
Intron 2
Exon 3
Isoform 2
Probe-level
B
Test for shared causal variant between GWAS and eQTL signal
Conditional analysis using summary eQTL results (GCTA – COJO)
Re-perform eQTL analysis using all cis-SNPs to find 'best cis-eQTL'
Condition all cis-SNPs on best cis-eQTL using TwinsUK genetic data as reference panel
GWAS SNP conditional p-value (p_{cond} < 0.05, independent to best eQTL)
Colocalisation analysis using summary eQTL results (Coloc R package)
Integrate primary GWAS data with summary cis-eQTL results and TwinsUK genetic data
Single causal variant, common to both traits (PP4 > PP3, co-localised)
Independent causal variant
Shared causal variant
Best cis-eQTL
GWAS SNP
Association significance
Disease signal
eQTL signal
Significance threshold
Genomic position
SLE candidate-causal eQTLs & eGenes
Microarray Probe-level eQTLs & eGenes
RNA-Seq Gene-level eQTLs & eGenes
RNA-Seq Exon-level eQTLs & eGenes
Table 3
Table 4
Table 5

Fig 2.

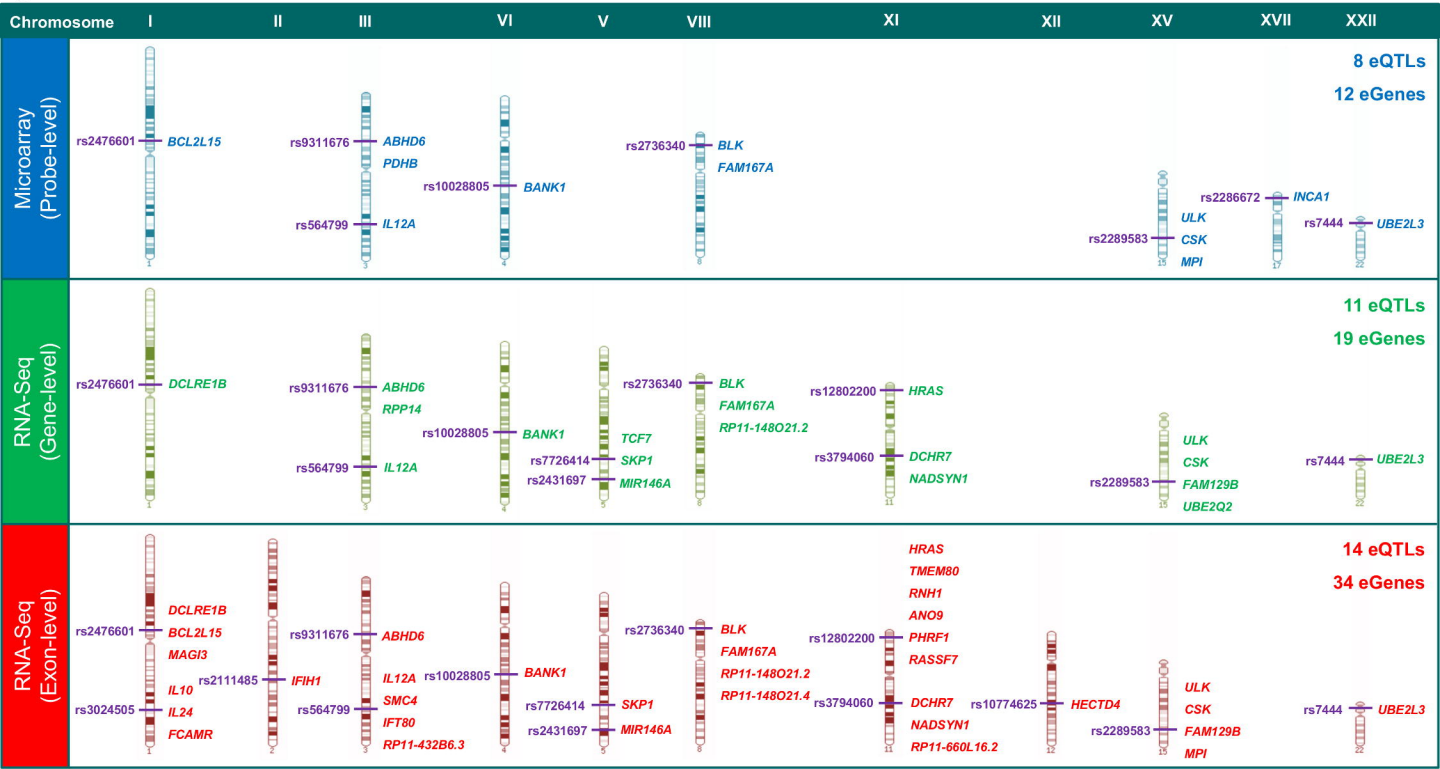

Chromosome
I
II
III
VI
V
VIII
XI
XII
XV
XVII
XXII
Microarray (Probe-level)
8 eQTLs
12 eGenes
rs2476601
BCL2L15
rs9311676
ABHD6
PDHB
rs564799
IL12A
rs10028805
BANK1
rs2736340
BLK
FAM167A
rs2286672
INCA1
ULK
rs2289583
CSK
MPI
rs7444
UBE2L3
RNA-Seq (Gene-level)
11 eQTLs
19 eGenes
rs2476601
DCLRE1B
rs9311676
ABHD6
RPP14
rs564799
IL12A
rs10028805
BANK1
TCF7
rs7726414
SKP1
rs2431697
MIR146A
rs2736340
BLK
FAM167A
RP11-148O21.2
rs12802200
HRAS
rs3794060
DCHR7
NADSYN1
ULK
CSK
rs2289583
FAM129B
UBE2Q2
rs7444
UBE2L3
RNA-Seq (Exon-level)
14 eQTLs
34 eGenes
rs2476601
DCLRE1B
BCL2L15
MAGI3
rs2111485
IL10
rs3024505
IL24
FCAMR
rs9311676
ABHD6
IFIH1
rs564799
IL12A
SMC4
IFT80
RP11-432B6.3
rs10028805
BANK1
rs7726414
SKP1
rs2431697
MIR146A
rs2736340
BLK
FAM167A
RP11-148O21.2
RP11-148O21.4
rs12802200
HRAS
TMEM80
RNH1
ANO9
PHRF1
RASSF7
rs3794060
DCHR7
NADSYN1
RP11-660L16.2
rs10774625
HECTD4
ULK
CSK
rs2289583
FAM129B
MPI
rs7444
UBE2L3

Fig 3.

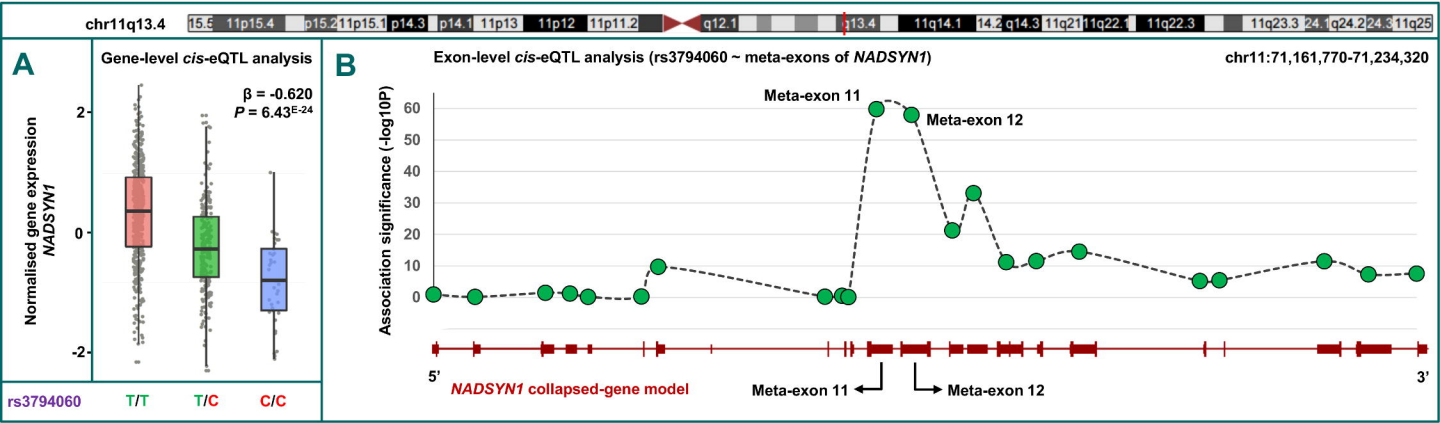

chr11q13.4
A
Gene-level cis-eQTL analysis
β = -0.620
P = 6.43E-24
Normalised gene expression NADSYN1
2
0
-2
rs3794060
T/T
T/C
C/C
B
Exon-level cis-eQTL analysis (rs3794060 ~ meta-exons of NADSYN1)
chr11:71,161,770-71,234,320
Association significance (-log10P)
60
50
40
30
20
10
0
Meta-exon 11
Meta-exon 12
5'
3'
NADSYN1 collapsed-gene model
Meta-exon 11
Meta-exon 12

Fig 4.

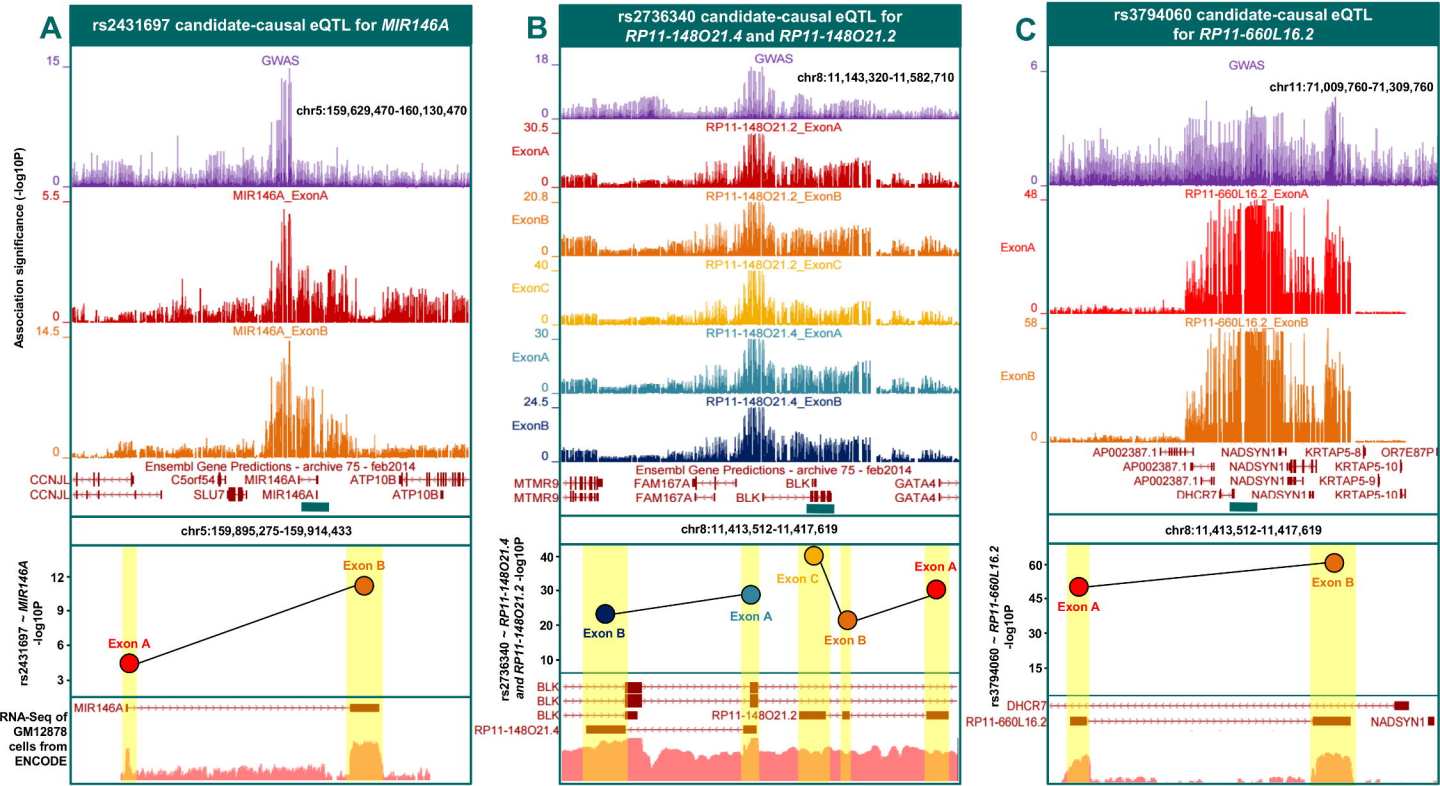

A
rs2431697 candidate-causal eQTL for MIR146A
GWAS
chr5:159,629,470-160,130,470
MIR146A_ExonA
MIR146A_ExonB
Association significance (-log10P)
Ensembl Gene Predictions - archive 75 - feb2014
CCNJL
C5orf54
MIR146A
ATP10B
SLU7
MIR146AI
chr5:159,895,275-159,914,433
rs2431697 ~ MIR146A -log10P
Exon A
Exon B
RNA-Seq of GM12878 cells from ENCODE
B
rs2736340 candidate-causal eQTL for RP11-148O21.4 and RP11-148O21.2
chr8:11,143,320-11,582,710
RP11-148O21.2_ExonA
RP11-148O21.2_ExonB
RP11-148O21.2_ExonC
RP11-148O21.4_ExonA
RP11-148O21.4_ExonB
MTMR9
FAM167A
BLK
GATA4
chr8:11,413,512-11,417,619
rs2736340 ~ RP11-148O21.4 and RP11-148O21.2 -log10P
Exon C
RP11-148O21.2
RP11-148O21.4
C
rs3794060 candidate-causal eQTL for RP11-660L16.2
chr11:71,009,760-71,309,760
RP11-660L16.2_ExonA
RP11-660L16.2_ExonB
AP002387.1
NADSYN1
KRTAP5-8
OR7E87P
KRTAP5-10
KRTAP5-9
DHCR7
chr8:11,413,512-11,417,619
rs3794060 ~ RP11-660L16.2 -log10P
RP11-660L16.2

Fig 5.

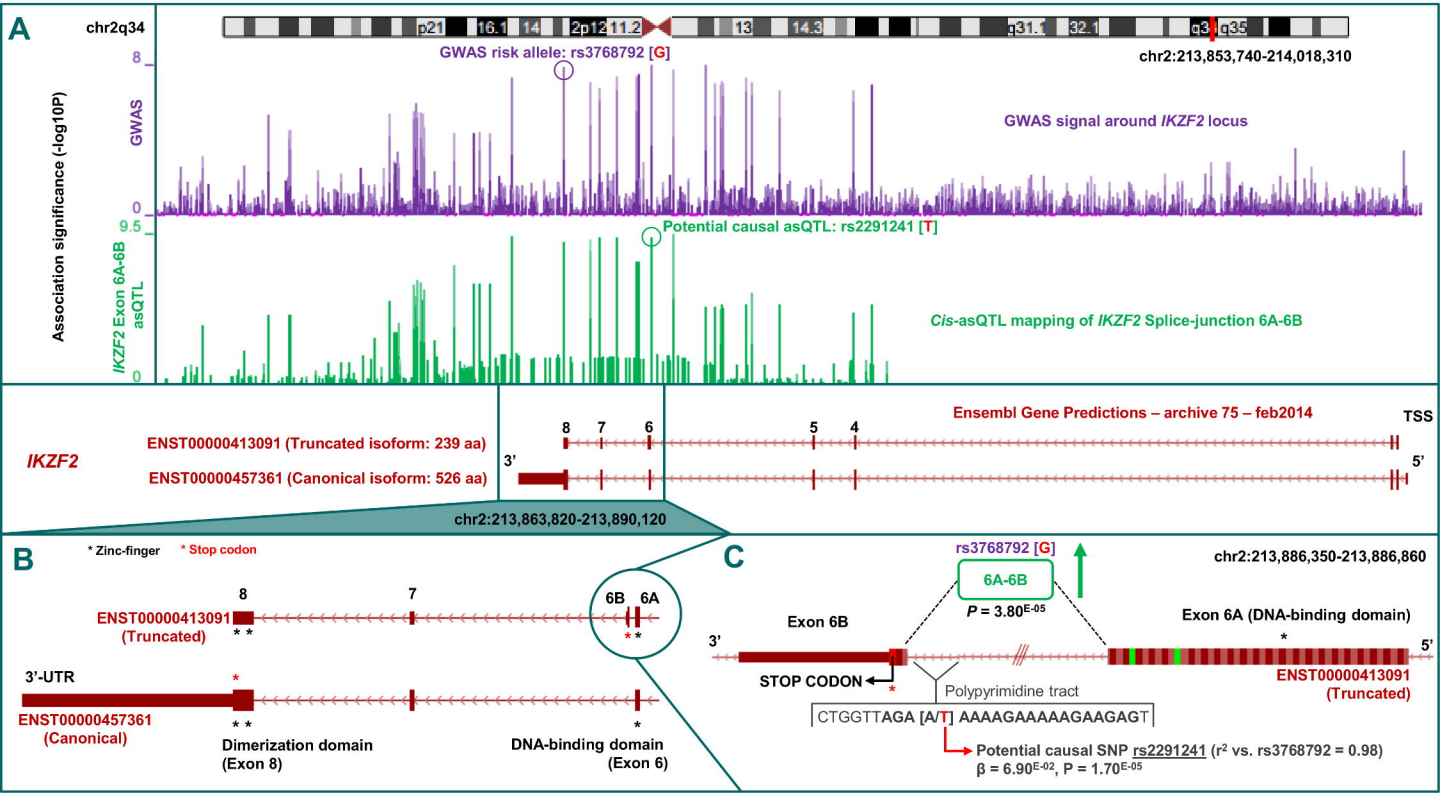

A
chr2q34
p21
16.1
14
2p12
11.2
13
14.3
q31.1
32.1
q3
q35
chr2:213,853,740-214,018,310
GWAS risk allele: rs3768792 [G]
GWAS signal around IKZF2 locus
Potential causal asQTL: rs2291241 [T]
Cis-asQTL mapping of IKZF2 Splice-junction 6A-6B
Association significance (-log10P)
GWAS
IKZF2 Exon 6A-6B asQTL
8
0
9.5
0
IKZF2
ENST00000413091 (Truncated isoform: 239 aa)
ENST00000457361 (Canonical isoform: 526 aa)
Ensembl Gene Predictions – archive 75 – feb2014
TSS
3'
5'
8
7
6
5
4
chr2:213,863,820-213,890,120
B
* Zinc-finger
* Stop codon
ENST00000413091 (Truncated)
3'-UTR
ENST00000457361 (Canonical)
Dimerization domain (Exon 8)
DNA-binding domain (Exon 6)
6B
6A
C
rs3768792 [G]
6A-6B
P = 3.80E-05
chr2:213,886,350-213,886,860
Exon 6B
Exon 6A (DNA-binding domain)
3'
5'
STOP CODON
Polypyrimidine tract
CTGGTTAGA [A/T] AAAAGAAAAAGAAGAGT
ENST00000413091 (Truncated)
Potential causal SNP rs2291241 (r2 vs. rs3768792 = 0.98)
β = 6.90E-02, P = 1.70E-05

Fig 6.

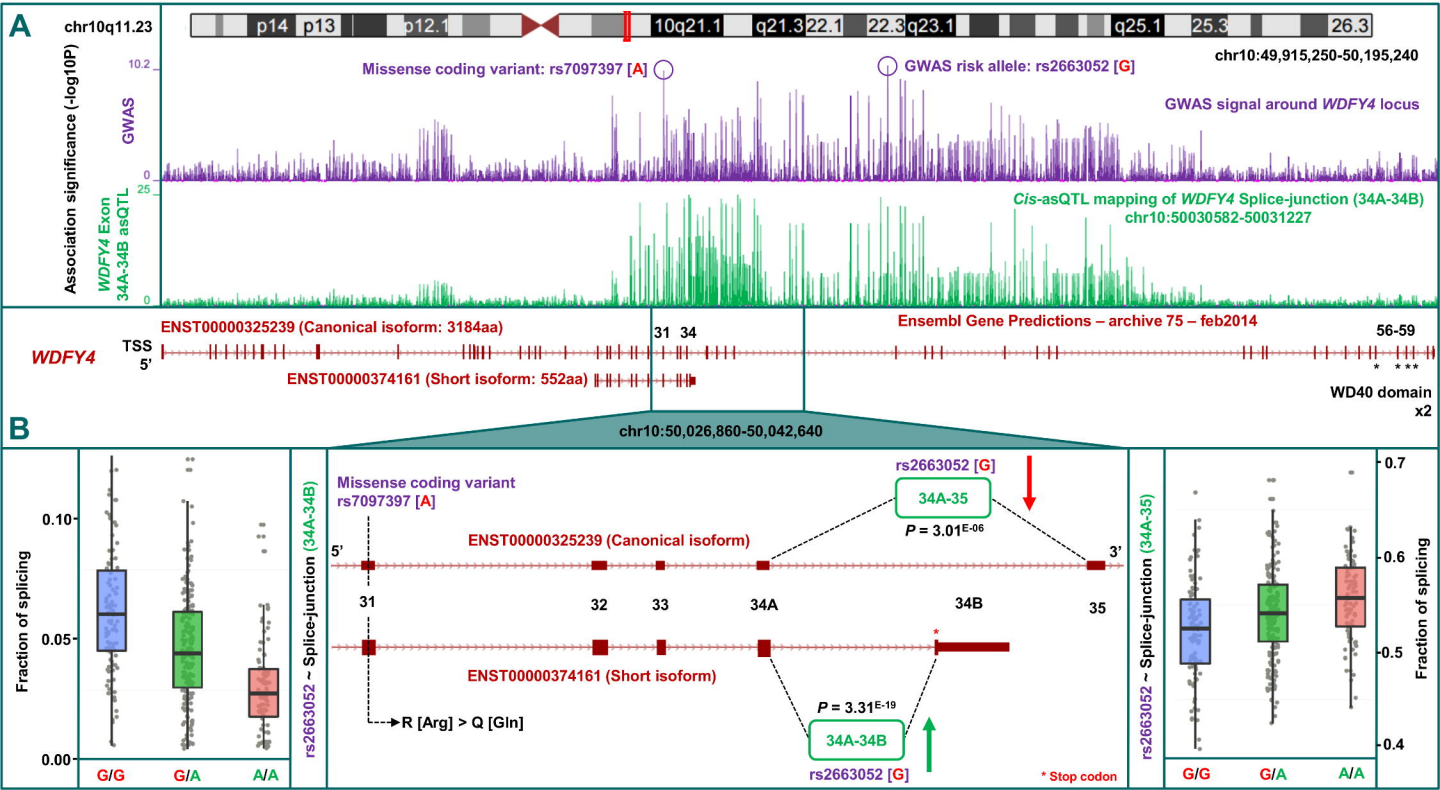
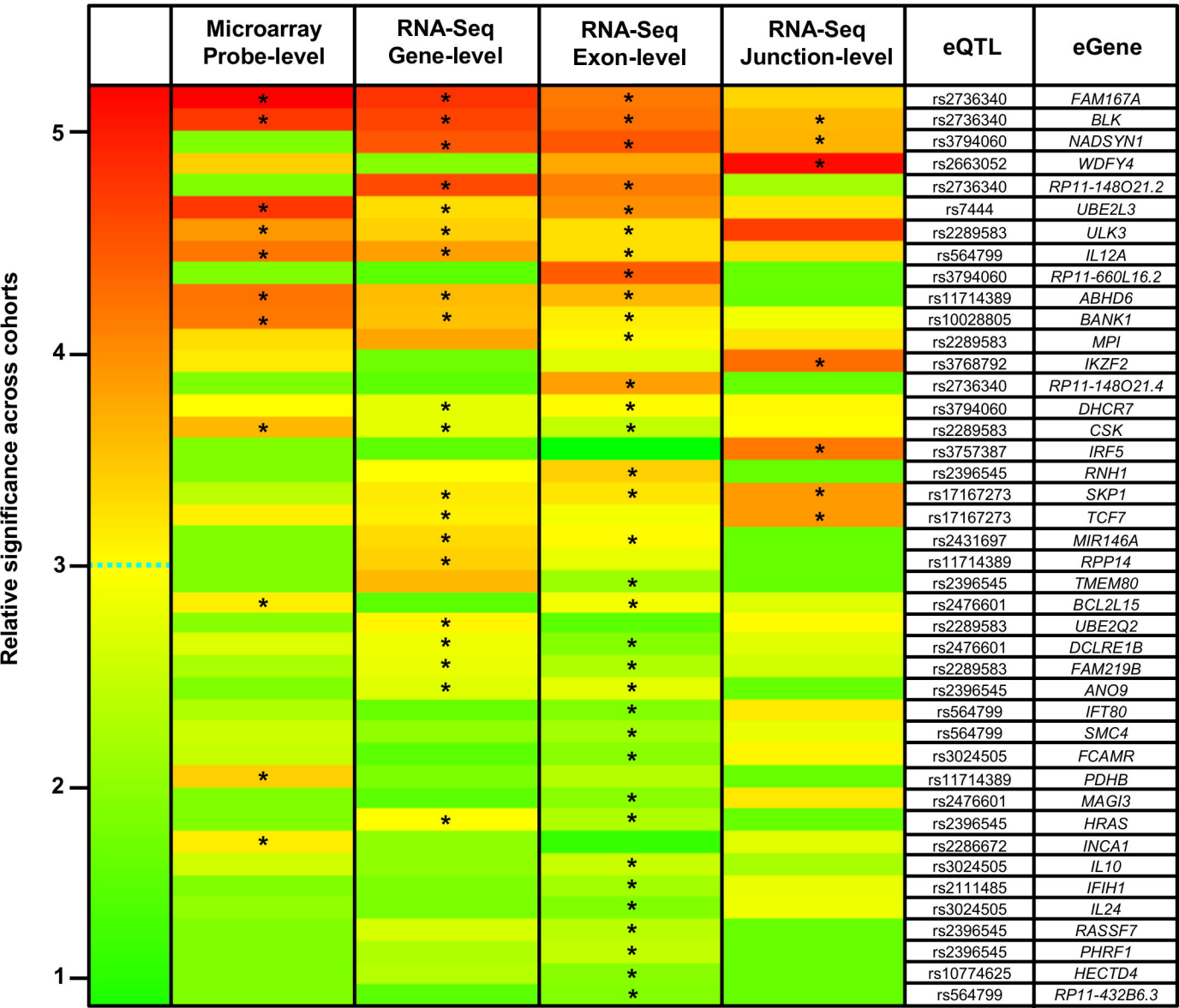

A
chr10q11.23
p14
p13
p12.1
10q21.1
q21.3
22.1
22.3
q23.1
q25.1
25.3
26.3
chr10:49,915,250-50,195,240
Missense coding variant: rs7097397 [A]
GWAS risk allele: rs2663052 [G]
GWAS signal around WDFY4 locus
Association significance (-log10P)
GWAS
10.2
0
WDFY4 Exon 34A-34B asQTL
25
0
Cis-asQTL mapping of WDFY4 Splice-junction (34A-34B)
chr10:50030582-50031227
WDFY4
TSS
5'
ENST00000325239 (Canonical isoform: 3184aa)
ENST00000374161 (Short isoform: 552aa)
31
34
Ensembl Gene Predictions – archive 75 – feb2014
56-59
WD40 domain x2
B
chr10:50,026,860-50,042,640
Fraction of splicing
0.10
0.05
0.00
G/G
G/A
A/A
rs2663052 ~ Splice-junction (34A-34B)
Missense coding variant
rs7097397 [A]
rs2663052 [G]
34A-35
P = 3.01E-06
ENST00000325239 (Canonical isoform)
5'
3'
31
32
33
34A
34B
35
ENST00000374161 (Short isoform)
R [Arg] > Q [Gln]
P = 3.31E-19
34A-34B
rs2663052 [G]
* Stop codon
rs2663052 ~ Splice-junction (34A-35)
0.7
0.6
0.5
0.4
Fraction of splicing
G/G
G/A
A/A

Fig 7.


Microarray Probe-level
RNA-Seq Gene-level
RNA-Seq Exon-level
RNA-Seq Junction-level
eQTL
eGene
Relative significance across cohorts
5
4
3
2
1
rs2736340 FAM167A
rs2736340 BLK
rs3794060 NADSYN1
rs2663052 WDFY4
rs2736340 RP11-148O21.2
rs7444 UBE2L3
rs2289583 ULK3
rs564799 IL12A
rs3794060 RP11-660L16.2
rs11714389 ABHD6
rs10028805 BANK1
rs2289583 MPI
rs3768792 IKZF2
rs2736340 RP11-148O21.4
rs3794060 DHCR7
rs2289583 CSK
rs3757387 IRF5
rs2396545 RNH1
rs17167273 SKP1
rs17167273 TCF7
rs2431697 MIR146A
rs11714389 RPP14
rs2396545 TMEM80
rs2476601 BCL2L15
rs2289583 UBE2Q2
rs2476601 DCLRE1B
rs2289583 FAM219B
rs2396545 ANO9
rs564799 IFT80
rs564799 SMC4
rs3024505 FCAMR
rs11714389 PDHB
rs2476601 MAGI3
rs2396545 HRAS
rs2286672 INCA1
rs3024505 IL10
rs2111485 IFIH1
rs3024505 IL24
rs2396545 RASSF7
rs2396545 PHRF1
rs10774625 HECTD4
rs564799 RP11-432B6.3