

## SHORT REPORT

# Fine mapping genetic determinants of the highly variably expressed MHC gene *ZFP57*

Katharine Plant<sup>1</sup>, Benjamin P Fairfax<sup>1</sup>, Seiko Makino<sup>1</sup>, Claire Vandiedonck<sup>1,2,3</sup>, Jayachandran Radhakrishnan<sup>1</sup> and Julian C Knight<sup>\*1</sup>

**ZFP57 is an important transcriptional regulator involved in DNA methylation and genomic imprinting during development. Here we demonstrate that gene expression also occurs at a low level in adult peripheral blood cells and other tissues including the kidney and thymus, but is critically dependent on underlying local genetic variation within the MHC. We resolve a highly significant expression quantitative trait locus for *ZFP57* involving single-nucleotide polymorphisms (SNPs) in the first intron of the gene co-localizing with a DNase I hypersensitive site and evidence of CTCF recruitment. These data identify *ZFP57* as a candidate gene underlying reported MHC disease associations, notably for putative regulatory variants associated with cancer and HIV-1. The work highlights the role that *ZFP57* may play in DNA methylation and epigenetic regulation beyond early development into adult life dependent on genetic background, with important potential implications for disease.**

*European Journal of Human Genetics* (2014) 22, 568–571; doi:10.1038/ejhg.2013.244; published online 6 November 2013

**Keywords:** ZFP57; expression quantitative trait; genetic variation; MHC; disease susceptibility

## INTRODUCTION

ZFP57 is a Kruppel-associated box (KRAB) containing zinc-finger protein, preferentially expressed early in development.<sup>1</sup> ZFP57 has been shown in the mouse to act as a transcriptional regulator and is important in maintenance of imprinting,<sup>2,3</sup> regulating chromatin modifications and DNA methylation at murine imprinted loci in ES cells with the co-factor KRAB-associated protein 1 (KAP1/TRIM28).<sup>4</sup> ZFP57 loss-of-function mutations are known to cause a hypomethylation disorder presenting as transient neonatal diabetes (TND) associated with a unique epigenetic profile at the TND differentially methylated region and other imprinted loci such as *GRB10* and *PEG3*,<sup>5</sup> but the biology of ZFP57 in humans is not well characterized. Recently we showed that the expression of ZFP57 was dependent on underlying genetic variation.<sup>6</sup> Given the location of ZFP57 in the MHC class I region, we sought to resolve the association and investigate the relationship with disease.

## MATERIALS AND METHODS

Volunteer recruitment, cell purification, cell culture, RACE, genotyping, imputation, eQTL mapping and relationship with reported GWAS were performed as detailed in Supplementary Information.

## RESULTS

We aimed to define the genetic modulators of ZFP57 transcription by expression quantitative trait (eQTL) mapping. Alternatively spliced isoforms are well characterized for murine *Zfp57*<sup>1</sup> and a number of isoforms are annotated in humans (Supplementary Figure S1). In order to take account of this when quantifying ZFP57 expression, we first characterized transcription in lymphoblastoid cell lines (LCLs) and peripheral blood mononuclear cells (PBMCs) from volunteers identified as expressing ZFP57. Rapid amplification of cDNA ends

(RACE) using 3' and 5' adapted cDNA from the COX LCL, known to be a high expresser of ZFP57,<sup>6</sup> and PCR with exon spanning primers revealed a previously unrecognized isoform, in which exon 2 is skipped and predicted to have a significantly truncated KRAB domain (Supplementary Figure S1). Quantification of ZFP57 using isoform-specific primers or primers spanning exons 3/4 to capture both isoforms revealed low but detectable expression in PBMCs, ES cells and several adult tissues, notably the thymus and kidney (Supplementary Figure S1). Relative abundance of the different isoforms remained consistent between different tissues and across individuals (Supplementary Figure S1).

We proceeded to eQTL mapping in a cohort of 288 healthy volunteers<sup>7</sup> using primers spanning exons 3/4 to quantify transcript abundance in PBMCs. Following processing and quality control filtering, we analysed 651 210 SNP markers for 283 individuals. This revealed a major eQTL for ZFP57 with the most significant associated SNP rs375984 ( $P = 9.3 \times 10^{-50}$ ) in the second intron of ZFP57 (Figure 1a and b). Analysis of purified monocytes from the same volunteers confirmed a strong eQTL, the most significant association was to rs375984 ( $P = 3.2 \times 10^{-11}$ ; Supplementary Figure S2). To further resolve this, we imputed 19 129 additional SNPs within 250 kb, which revealed three more strongly associated variants in the first intron of ZFP57 in perfect LD (rs416568, rs365052 and rs2747431,  $P = 4.6 \times 10^{-52}$ ; Figure 1c and d). We determined the functional genomic landscape for these eSNPs using data from the ENCODE project.<sup>8</sup> Analysis of DNase-seq and ChIP-seq data sets resolved rs365052 is a candidate regulatory variant located in a DNase I hypersensitive site with evidence of CTCF binding (Figure 1d).

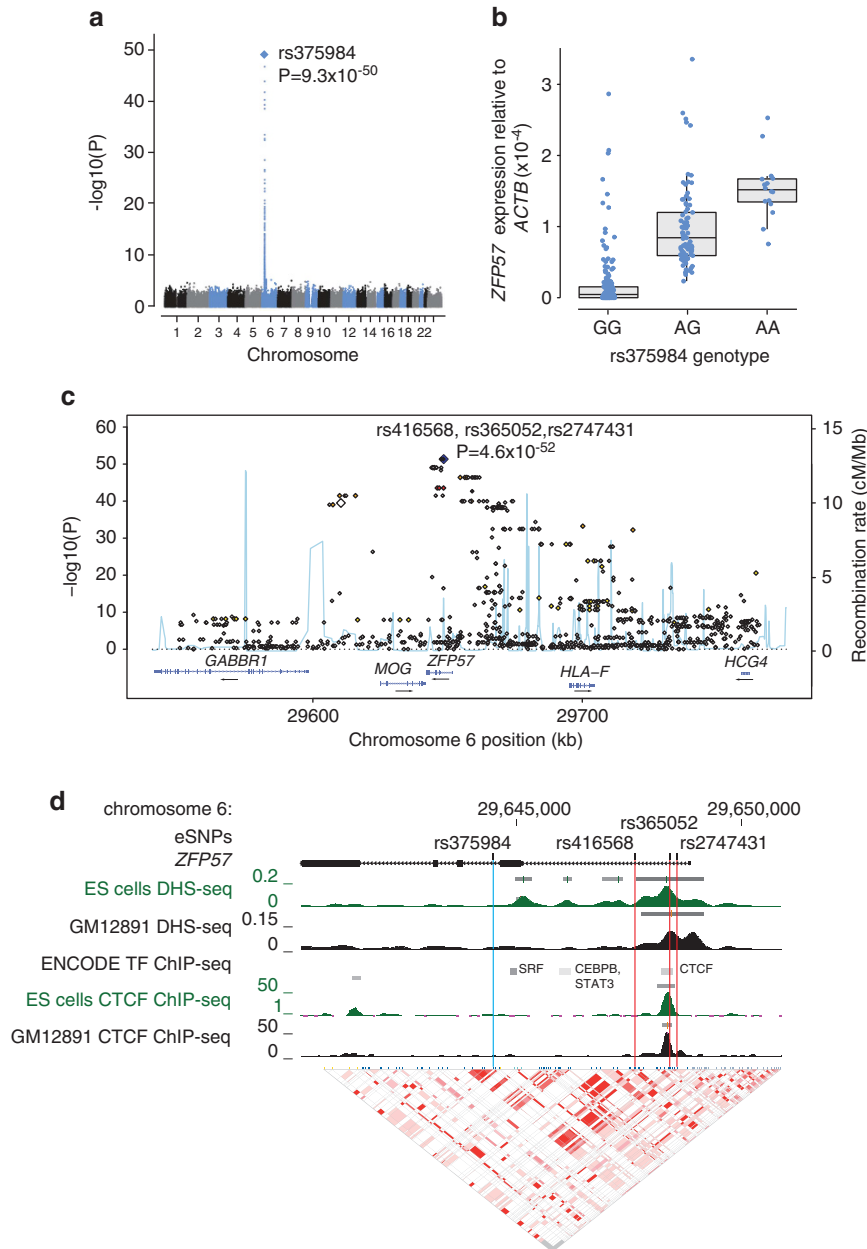
Analysis of the eQTL by HLA type showed association with HLA-A\*01 and \*23, but was not more informative than SNP

<sup>1</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; <sup>2</sup>INSERM UMR-S 958, F-75010 Paris, France; <sup>3</sup>Univ Paris Diderot, Sorbonne Paris Cité, F-75013 Paris, France

\*Correspondence: Dr JC Knight, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.

E-mail: julian@well.ox.ac.uk

Received 10 August 2012; revised 5 September 2013; accepted 20 September 2013; published online 6 November 2013

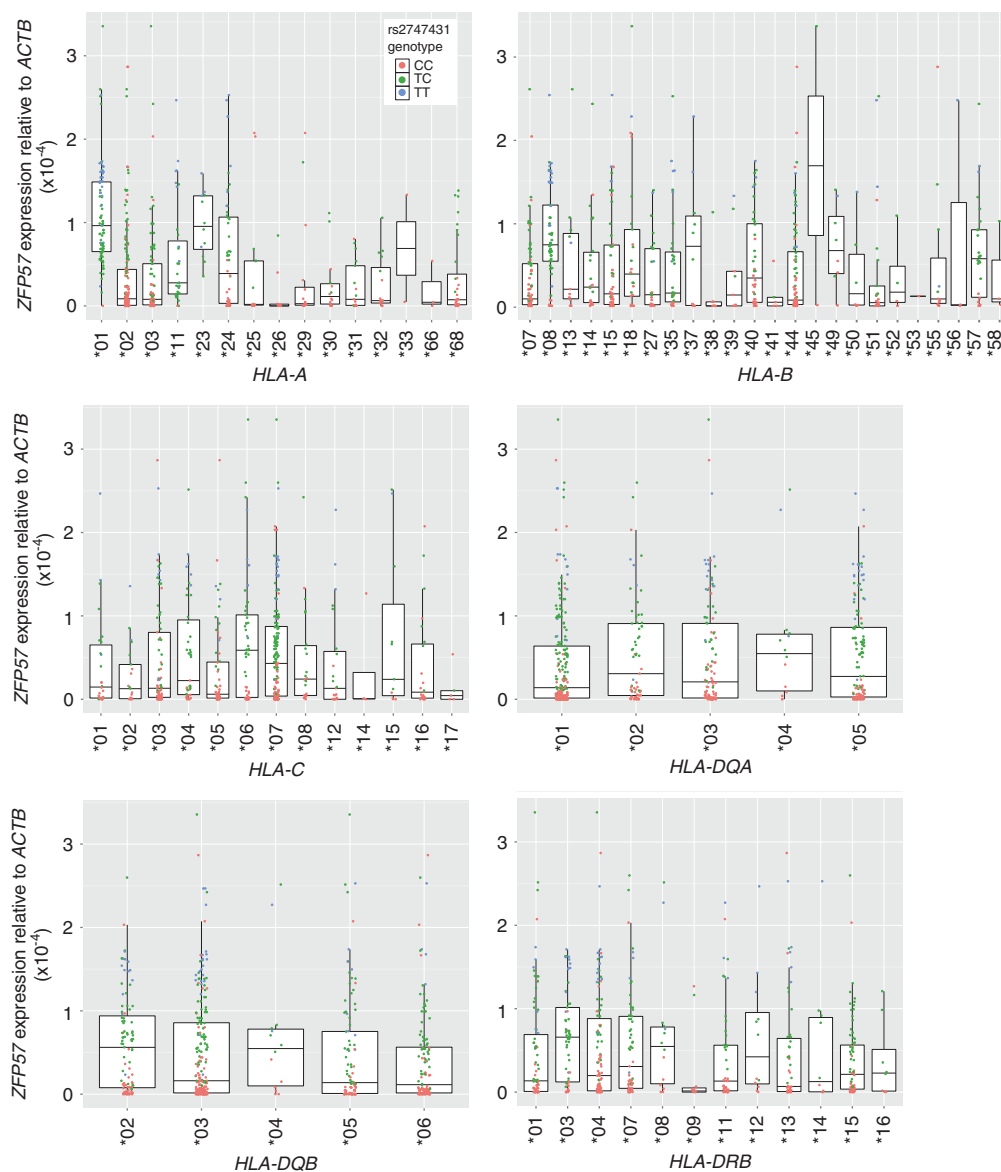


**Figure 1** Genetic modulators of *ZFP57* expression. (a) Manhattan plot showing strength of association plotted as  $-\log_{10}(P)$  values by chromosome for *ZFP57* expression. (b) Scatter/box and whiskers plot of *ZFP57* expression by rs375984 allele demonstrating significant differences between the different genotype groups ( $P < 0.0001$ , Kruskal–Wallis test). (c) Local association and recombination plot. Single marker allelic association results for a 215 kb region spanning *ZFP57* plotted as  $-\log_{10}(P)$  values (left y-axis) by genomic coordinate (x-axis). With reference to rs2747431 (which is in complete LD with rs416568 and rs365052), typed SNPs are shown in red ( $r^2 > 0.8$ ), orange (0.5–0.8), yellow (0.2–0.5) and white ( $< 0.2$ ). Imputed SNPs are shown in grey. Recombination rate is also plotted (right y-axis). (d) Functional genomic landscape for *ZFP57* (chr6:29640242–29650866) providing context for observed eSNPs, including rs375984, rs416568, rs365052 and rs2747431. Data are shown from the ENCODE project, accessed through the UCSC Genome Browser (<http://genome.ucsc.edu/>), resolving a DNase I hypersensitive site and evidence of CTCF binding in the region of rs365052 based on profiling of ES cells, LCLs (GM12878, GM12891) and CD20+ B cells. Linkage disequilibrium plot for the locus based on  $r^2$  is shown below including 115 SNPs (1000 Genomes CEU phase 1).

markers (Figure 2). For the two most common ancestral haplotypes among Europeans, we found that volunteers with a copy of HLA-A1-B8-DR3 ( $n = 19$ ) had higher expression of *ZFP57* compared to HLA-A3-B7-DR15 ( $n = 12$ ; Mann–Whitney,  $P < 0.0001$ ).

We investigated whether the genetic variants identified here as associated with *ZFP57* expression may be significant in common disease given the many disease associations reported involving the

MHC class I region. We interrogated GWAS data sets and found intersection of *ZFP57* eSNPs variants with reported disease associations involving malignancy, HIV/AIDS and autoimmunity (Table 1). These included nasopharyngeal carcinoma and prostate cancer, the latter involving disease risk based on gene–gene interaction with the tumour suppressor gene *NKX3-1*. Associations were also noted involving HIV-1 viral set point and disease progression to AIDS.



**Figure 2** Association between *ZFP57* expression, eSNPs and classical HLA types. *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB*, *HLA-DQA* and *HLA-DQB* shown at two-digit resolution with *ZFP57* expression in PBMCs quantified by qPCR. Expression values are plotted for each individual corresponding to each HLA allele and coloured based on rs2747431 genotype (individuals with CC genotype at rs2747431 shown in red, CT in green and TT in blue). Two *ZFP57* expression values are plotted for each individual corresponding to each allele. There was evidence of association for *HLA-A*\*01 and *HLA-A*\*23 alleles ( $P < 0.0001$  when analysed using a Mann–Whitney test).

## DISCUSSION

We found that a significant minority of people show low-level transcription of *ZFP57* in adult cells and tissues, where it may modulate epigenetic processes, and that this is dependent on a strong local eQTL for *ZFP57*. Further work is required to resolve the functional basis for this, but a potential mechanism involves modulation of a novel regulatory element involving CTCF binding in the first intron of *ZFP57*. Our data also highlighted a potential role for *ZFP57* eSNPs in traits including cancer and HIV/AIDS. KRAB-ZNF genes play a role in epigenetic processes critical to cancer, including silencing of tumour suppressor genes, while the co-factor KAP1 is involved in oncogenesis.<sup>9</sup> DNA methylation is involved in establishing latency by retroviruses with hypermethylation of the viral 5' long terminal repeat characteristic of HIV-1 aviraemic patients.<sup>10</sup>

To date in humans, *ZFP57* has only been associated with maintenance and not establishment of DNA methylation,<sup>5</sup> though a role in *de novo* methylation has been reported in mice.<sup>2</sup> The identification of a novel shorter isoform of *ZFP57* may be functionally significant given the resulting severely truncated KRAB domain, which is likely to limit interaction with KAP1.

Complex LD structure in the MHC together with differences in SNP coverage between genotyping platforms necessitates further work to establish whether the most significant *ZFP57* eSNPs are also most informative for disease association and to explore the biological significance and relative importance of this observation in the context of extensive haplotype-specific expression in the MHC for other genes<sup>6</sup> with multiple *cis*- and *trans*-eQTL identified for this genomic region.<sup>7</sup>

**Table 1 Diseases and traits from the Catalogue of Published Genome-Wide Association Studies (www.genome.gov/gwastudies, Accessed May 2012) in which reported GWAS SNPs are also eSNPs for ZFP57 expression**

Disease/trait	PUBMED ID	First author	SNP	GWAS P-value	eQTL P-value	r <sup>2</sup> with peak eSNP
<i>Cancer</i>						
Nasopharyngeal carcinoma	19664746	Tse KP	rs3129055	7.4 × 10 <sup>-11</sup>	5.4 × 10 <sup>-39</sup>	0.61
Nasopharyngeal carcinoma	20512145	Bei JX	rs2860580	4.9 × 10 <sup>-67</sup>	7.0 × 10 <sup>-07</sup>	0.22
Prostate cancer (gene × 10 gene interaction)	22219177	Tao S	rs2523395	1.5 × 10 <sup>-06</sup>	1.3 × 10 <sup>-13</sup>	0.35
<i>Autoimmune disease</i>						
Graves' disease	21900946	Nakabayashi K	rs3893464	1.9 × 10 <sup>-20</sup>	1.1 × 10 <sup>-18</sup>	0.47
Multiple sclerosis	19525953	De Jager PL	rs2523393	1.0 × 10 <sup>-17</sup>	1.3 × 10 <sup>-13</sup>	0.35
<i>HIV/AIDS</i>						
AIDS progression	19115949	Limou S	rs8321	4.7 × 10 <sup>-07</sup>	5.5 × 10 <sup>-14</sup>	0.30
HIV-1 control	20041166	Fellay J	rs259919	3.0 × 10 <sup>-07</sup>	9.8 × 10 <sup>-14</sup>	0.22
<i>Other</i>						
Drug-induced liver injury (amoxicillin-clavulanate)	21570397	Lucena MI	rs2523822	1.8 × 10 <sup>-10</sup>	4.6 × 10 <sup>-07</sup>	0.21
IgE levels	22075330	Granada M	rs2571391	1.2 × 10 <sup>-15</sup>	1.8 × 10 <sup>-06</sup>	0.23

r<sup>2</sup> between these SNPs and the peak ZFP57 eQTL, tagged by rs2747431, is shown.

We have presented evidence that ZFP57 is expressed more widely than previously appreciated, notably beyond development, and that this is dependent on underlying genetic variation. Further work is needed to investigate the role of ZFP57 in epigenetic regulation, notably in terms of cancer and HIV infection where expression associated SNPs may play a role and epigenetic mechanisms are known to be important.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGEMENTS

We are grateful to the volunteers, to Stephen Leslie and Alexander Dilthey for HLA-type imputation and to Tomas Malinauskas for assistance with amino-acid analysis. We thank Peter Humburg and other members of the Knight Lab for useful advice and discussion of this work. This study was supported by MRC Studentship (KP); Wellcome Trust (074318 JCK, 088891 BPE, 075491/Z/04 core); European Research Council (281824 JCK); MRC (98082 JCK); Alliance Franco-British Partnership (JCK, CV); NIHR Oxford Biomedical Research Centre.

- Li X, Ito M, Zhou F *et al*: A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Dev Cell* 2008; **15**: 547–557.
- Zuo X, Sheng J, Lau HT *et al*: Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. *J Biol Chem* 2012; **287**: 2107–2118.
- Quenneville S, Verde G, Corsinotti A *et al*: In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol Cell* 2011; **44**: 361–372.
- Mackay DJ, Callaway JL, Marks SM *et al*: Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nat Genet* 2008; **40**: 949–951.
- Vandiedonck C, Taylor MS, Lockstone HE *et al*: Pervasive haplotypic variation in the spliceo-transcriptome of the human major histocompatibility complex. *Genome Res* 2011; **21**: 1042–1054.
- Fairfax BP, Makino S, Radhakrishnan J *et al*: Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* 2012; **44**: 502–510.
- Dunham I, Kundaje A, Aldred SF *et al*: An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57–74.
- Iyengar S, Farnham PJ: KAP1 protein: an enigmatic master regulator of the genome. *J Biol Chem* 2011; **286**: 26267–26276.
- Blazkova J, Trejbalova K, Gondois-Rey F *et al*: CpG methylation controls reactivation of HIV from latency. *PLoS Pathog* 2009; **5**: e1000554.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>

1 Okazaki S, Tanase S, Choudhury BK *et al*: A novel nuclear protein with zinc fingers down-regulated during early mammalian cell differentiation. *J Biol Chem* 1994; **269**: 6900–6907.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)