nature genetics

Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit *RBM8A* causes TAR syndrome

Cornelis A Albers[1–3,17], Dirk S Paul[3,17], Harald Schulze[4,5,17], Kathleen Freson[6], Jonathan C Stephens[1,2], Peter A Smethurst[1,2], Jennifer D Jolley[1,2], Ana Cvejic[1–3], Myrto Kostadima[7], Paul Bertone[7], Martijn H Breuning[8], Najet Debili[9], Panos Deloukas[3], Rémi Favier[9], Janine Fiedler[5,10], Catherine M Hobbs[1,2], Ni Huang[3], Matthew E Hurles[3], Graham Kiddle[1,2], Ingrid Krapels[11], Paquita Nurden[12], Claudia A L Ruivenkamp[8], Jennifer G Sambrook[1,2], Kenneth Smith[13,14], Derek L Stemple[3], Gabriele Strauss[15], Chantal Thys[6], Chris van Geet[6,16], Ruth Newbury-Ecob[13,14,18], Willem H Ouwehand[1–3,18] & Cedric Ghevaert[1,2,18]

The exon-junction complex (EJC) performs essential RNA processing tasks[1–5]. Here, we describe the first human disorder, thrombocytopenia with absent radii (TAR)[6], caused by deficiency in one of the four EJC subunits. Compound inheritance of a rare null allele and one of two low-frequency SNPs in the regulatory regions of *RBM8A*, encoding the Y14 subunit of EJC, causes TAR. We found that this inheritance mechanism explained 53 of 55 cases ($P < 5 \times 10^{-228}$) of the rare congenital malformation syndrome. Of the 53 cases with this inheritance pattern, 51 carried a submicroscopic deletion of 1q21.1 that has previously been associated with TAR[7], and two carried a truncation or frameshift null mutation in *RBM8A*. We show that the two regulatory SNPs result in diminished *RBM8A* transcription *in vitro* and that Y14 expression is reduced in platelets from individuals with TAR. Our data implicate Y14 insufficiency and, presumably, an EJC defect as the cause of TAR syndrome.

The thrombocytopenia with absent radii (TAR) syndrome is characterized by a reduction in the number of platelets (the cells that make blood clot) (in TAR, platelet levels are generally below 50×10^9 platelets per liter, with the normal range being $150–350 \times 10^9$ platelets per liter) and the absence of the radius bone in the forearm, although there is preservation of the thumb, which distinguishes TAR from other syndromes that combine blood abnormalities with absence of the radius, such as Fanconi anemia[6,8]. Individuals with TAR have low numbers of megakaryocytes, the platelet precursor cells that reside in the bone marrow, and frequently present with bleeding episodes in the first year of life, which diminish in frequency and severity with age. The severity of skeletal abnormalities varies from absence of radii to virtual absence of upper limbs with or without lower-limb defects, such as malformations of the hip and knee[9]. An inherited or *de novo* deletion on chromosome 1q21.1 is found in the majority of affected individuals[7], but the apparent autosomal recessive nature of the syndrome requires the existence of an additional causative allele. This other allele has remained elusive, even with sequencing of the protein-coding exons of ten genes (including *RBM8A*) in the minimally deleted region (chr. 1: 145,399,075–145,594,214; hg19) (**Fig. 1a**, **Supplementary Fig. 1** and **Supplementary Note**)[7].

To identify the additional causative allele, we selected five individuals with TAR (cases) of European ancestry who had the 1q21.1 deletion and sequenced their exomes (see Online Methods). We were unable to find TAR-associated coding mutations in a gene. However, four of the cases carried the minor allele of a low-frequency SNP in the 5′ UTR of the *RBM8A* gene (rs139428292), while the remaining case carried a previously unknown SNP in the first intron of the same gene (**Fig. 1b**). Genotyping by Sanger sequencing of another 48 cases of European ancestry with the 1q21.1 deletion identified the two SNPs in 35 and 11 samples, respectively (**Fig. 1c**, **Supplementary Tables 1** and **2** and **Supplementary Note**). A mother of non-European ancestry with TAR and her fetus, aborted on the grounds of prenatal diagnosis with TAR, both did not carry the 5′ UTR or the intronic SNP

[1]Department of Haematology, University of Cambridge, Cambridge, UK. [2]National Health Service (NHS) Blood and Transplant, Cambridge, UK. [3]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. [4]Institute for Transfusion Medicine, Charité Universitätsmedizin, Berlin, Germany. [5]Laboratory for Pediatric Molecular Biology, Charité Universitätsmedizin, Berlin, Germany. [6]Center for Molecular and Vascular Biology, University of Leuven, Leuven, Belgium. [7]European Molecular Biology Laboratory (EMBL)–European Bioinformatics Institute (EBI), Hinxton, Cambridge, UK. [8]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands. [9]Institut National de la Santé et de la Recherche Médicale (INSERM) U790, Villejuif, France. [10]Department of Biology, Chemistry, and Pharmacy, Freie University Berlin, Berlin, Germany. [11]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands. [12]Laboratoire d'Hématologie, Centre de Référence des Pathologies Plaquettaires, Hopital Xavier Arnozan, Pessac, France. [13]Division of Child Health, University of Bristol, Bristol, UK. [14]Department of Clinical Genetics, St Michael's Hospital, Bristol, UK. [15]Department of Pediatric Oncology and Hematology, Charité Universitätsmedizin, Berlin, Germany. [16]Department of Pediatrics, Universitair Ziekenhuis Leuven, Leuven, Belgium. [17]These authors contributed equally to this work. [18]These authors jointly directed this work. Correspondence should be addressed to C.A.A. (caa@sanger.ac.uk) or C.G. (cg348@cam.ac.uk).

Received 3 August 2011; accepted 21 December 2011; published online 26 February 2012; doi:10.1038/ng.1083

(**Supplementary Note**), and we suggest that, in this instance, there is a different causative allele that we have not been able to identify. In the 25 trios where the deletion in the child was not a *de novo* event, we confirmed that the deletion and the newly identified SNPs were inherited from different parents (**Supplementary Table 1**). The minor allele frequencies (MAFs) of the 5′ UTR and intronic SNPs were 3.05% and 0.42%, respectively, in 7,504 healthy individuals of the Cambridge BioResource[10] (**Supplementary Note**), and the deletion

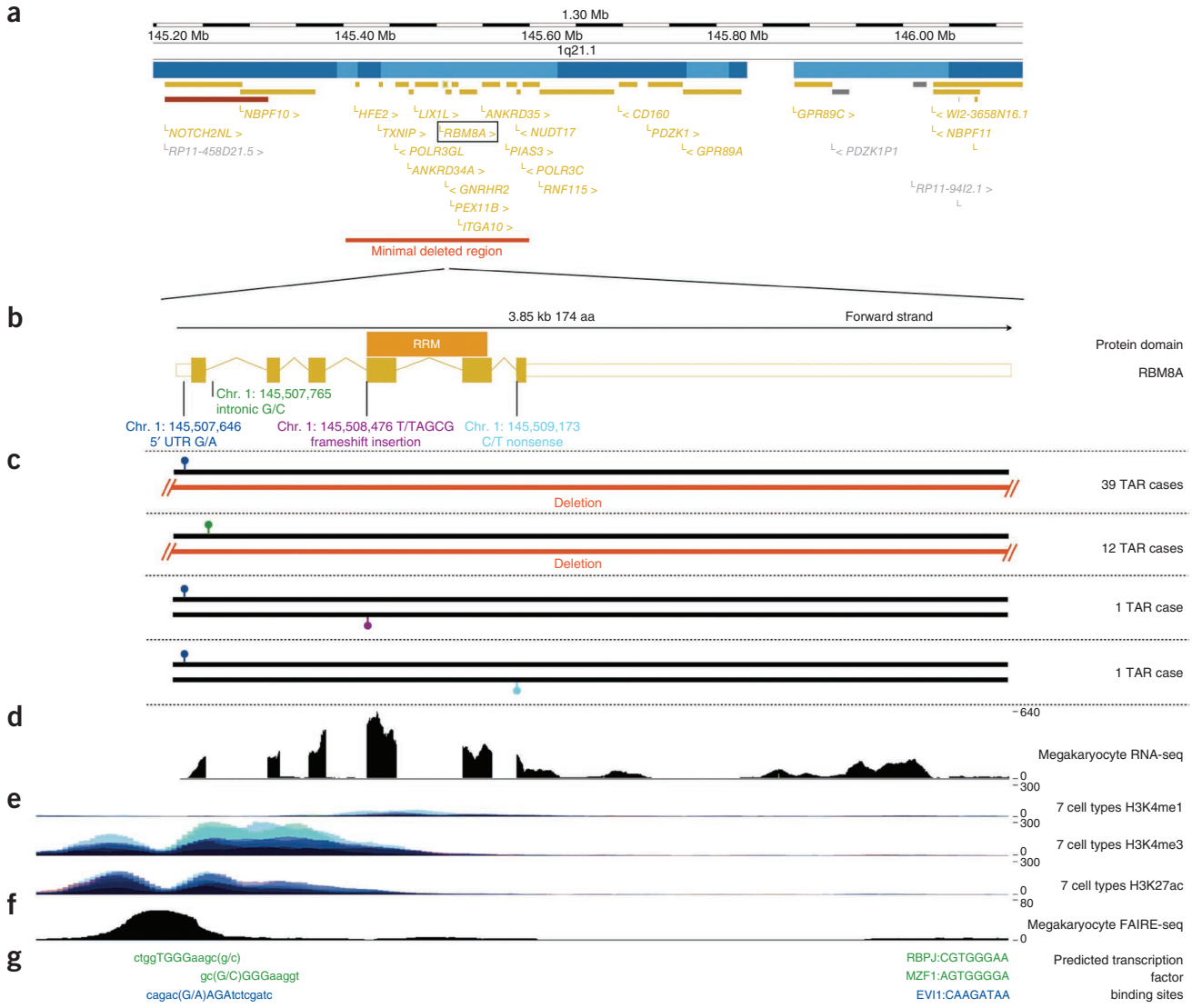

Figure 1 Most TAR syndrome cases have a low-frequency regulatory variant and a rare null allele at the *RBM8A* locus. (**a**) Fifty-three of 55 TAR cases were heterozygous carriers of a rare 1q21.1 deletion of varying size. The red bar indicates the region that was absent in all 53 cases having a deletion. Yellow bars, genes; grey bars, pseudogenes; blue bars, contigs. (**b**) The *RBM8A* transcript is shown in genomic coordinates with the sequence encoding the RNA-binding domain (RRM) indicated by the orange bar above the transcript. (**c**) We identified two low-frequency regulatory SNPs in 53 out of a total of 55 TAR cases studied. The first, at chr. 1: 145,507,646 (rs139428292), with a G or A allele, is located in the 5′ UTR of *RBM8A* and has a population MAF of 3.05% (dark blue). The second, at chr. 1: 145,507,765, with a G or C allele, is located in the first intron of *RBM8A* and has a population MAF of 0.41% (green). Thirty-nine TAR cases carried the minor allele of the 5′ UTR SNP on one chromosome and the 1q21.1 deletion on the other; 12 TAR cases carried the minor allele of the intronic SNP on one chromosome and the 1q21.1 deletion on the other. The compound inheritance of the 1q21.1 deletion and one of the two regulatory SNPs was strongly associated with TAR with $P < 5 \times 10^{-228}$. Two additional TAR cases were found to have the minor allele of the 5′ UTR SNP in combination with either a frameshift insertion (purple) or a nonsense mutation (light blue) instead of the 1q21.1 deletion, implicating *RBM8A* as the causative gene for TAR syndrome. (**d**) Sequencing of RNA from cord blood–derived megakaryocytes showing that *RBM8A* is transcribed in megakaryocytes. Shown is the sequencing read depth across the *RBM8A* locus. (**e**) Histone modifications from the ENCODE Project[13] in seven cell lines (GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK and NHLF) indicated the presence of regulatory elements in the promoter and first intron of *RBM8A*. Monomethylation of histone 3 at lysine 1 (H3K4me1) is often found near regulatory elements, trimethylation of this lysine (H3K4me3) is often found near promoters, and acetylation of histone H3 lysine 27 (H3K27ac) is often found near active regulatory elements (UCSC Genome Browser). Shown is the read depth resulting from sequencing the product of chromatin immunoprecipitation for the three histone modifications in the seven cell lines. Cell types are represented by different shades of blue and are superimposed. (**f**) FAIRE-seq, marking regions of open chromatin, showing that the 5′ UTR and intronic SNPs are accessible in megakaryocytes. (**g**) Computational modeling predicts that the minor allele of the 5′ UTR SNP creates a binding site for the EVI1 transcription factor, and the minor allele of the intronic SNP is predicted to disrupt binding of MZF1 and RBPJ. Capital letters indicate the consensus transcription factor binding sites, and the alleles for the SNPs are shown in parentheses.

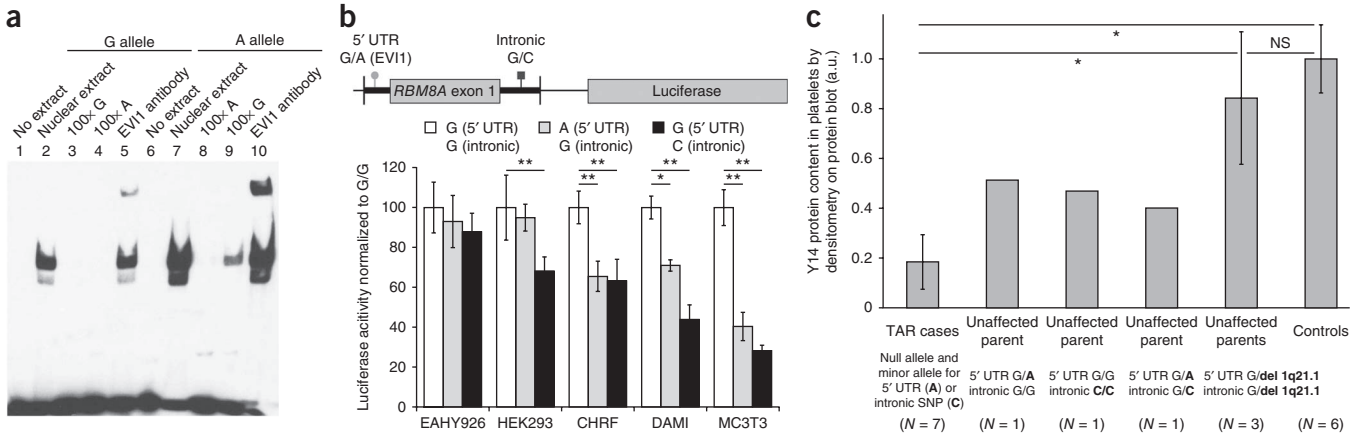

Figure 2 Effect of the regulatory SNPs on transcription factor binding, *RBM8A* promoter activity and protein expression in platelets. (**a**) EMSAs with nuclear protein extracts from the megakaryocytic cell line CHRF-288-11. Nuclear protein showed higher affinity for the probe with the A allele (lane 7) than the probe with the G allele (lane 2) of the 5′ UTR noncoding SNP. Binding of the A-allele probe was competed by a specific, unlabeled probe (100× A; lane 8) but not by a nonspecific, G-allele probe (100× G; lane 9). We observed a supershift with an antibody to EVI1 in DNA-protein complexes with the A-allele probe (lane 10), indicating that the minor allele of the 5′ UTR SNP increases binding affinity for the transcription factor EVI1 *in vitro*. (**b**) Luciferase reporter assays in cell lines representative of megakaryocytes (CHRF and DAMI) and osteoblasts (MC3T3). Top, schematic of the luciferase reporter construct with the 5′ UTR and intronic SNPs represented by circle and square symbols, respectively. Bottom, there was significantly decreased *RBM8A* promoter activity for the minor alleles of both the 5′ UTR and intronic noncoding SNPs relative to the major alleles. No effect of the 5′ UTR SNP was observed in EAHY926 and HEK293 human endothelial cells. Error bars, s.d.; $*P < 0.01$, $**P < 0.001$. Statistical analysis was performed using the Tukey-Kramer multiple comparisons test. Luciferase activity was normalized with respect to the construct consisting of the major G allele for both SNPs (indicated by G/G). (**c**) Densitometry analysis of immunoblot staining for Y14, the protein encoded by *RBM8A*, in platelet lysates from seven TAR cases, six parents (three with the 1q21.1 deletion, one heterozygous for the 5′ UTR SNP, one homozygous for the 5′ UTR SNP and one compound heterozygous for the 5′ UTR and intronic SNPs) and six controls. Results show significantly reduced Y14 protein levels in TAR cases compared to parental and control samples. Immunoblots are presented in **Supplementary Figure 3**. Error bars, s.d.; $*P < 0.01$, NS, not significant. Statistical analysis was performed using the heteroscedastic *t* test. Only genotype configurations indicated by lines were compared. The minor alleles of the 5′ UTR and intronic SNPs are shown in bold type. a.u., arbitrary units.

was absent from 5,919 shared healthy controls of the Wellcome Trust Case Control Consortium[10]. Thus, the concurrent presence of one of the two noncoding SNPs on one allele and the 1q21.1 deletion at the other is strongly associated with TAR syndrome, with an estimated *P* value of $< 5 \times 10^{-228}$ (**Supplementary Note**). Next, we sequenced all exons of *RBM8A* in two additional TAR cases who did not carry the 1q21.1 deletion but were found to carry the 5′ UTR SNP. We identified a 4-bp frameshift insertion at the start of the fourth exon in the first case and established that the noncoding SNP and insertion were on different chromosomes; in the second case, we identified a nonsense mutation in the last exon of *RBM8A* (**Fig. 1b**,**c**). Both mutations were absent from 458 exome samples of the 1000 Genomes Project[11] and 416 samples from the Cohorte Lausannoise (CoLaus)[12]. We conclude that, in the vast majority of cases, compound inheritance of a rare null allele (containing a deletion, frameshift mutation or encoded premature stop codon) and one of two low-frequency noncoding SNPs in *RBM8A* causes TAR syndrome.

On the basis of the genetic results, we postulated a hypomorphic mechanism for TAR, in which one copy of the *RBM8A* gene is not functional, due to a null allele, and expression of the other copy is reduced, as a result of noncoding SNPs in the 5′ UTR or first intron. Analysis of histone modifications in seven human cell lines from the ENCODE project indicated that both SNPs are localized to potential active regulatory elements (**Fig. 1d**,**e**)[13]. Annotation of open-chromatin structure using the formaldehyde-assisted isolation of regulatory elements (FAIRE) technique combined with sequencing (FAIRE-seq) provided additional evidence of this in megakaryocytes (**Fig. 1f**)[14]. Computational predictions suggest that the 5′ UTR SNP introduces a binding site for the transcriptional repressor EVI1 and that the intronic SNP disrupts a binding site for the transcription factors MZF1 and RBPJ (**Fig. 1g**). The prediction of EVI1 binding was confirmed by electrophoretic mobility shift assays (EMSAs) in the megakaryocytic cell line CHRF-288-11 (also known as CHRF), in which the EVI1 protein bound the minor allele but only weakly associated with the major allele (**Fig. 2a**). EMSA studies for the intronic SNP showed decreases in the binding of nuclear proteins to the minor allele, although we could not confirm the presence of either MZF1 or RBPJ in supershift experiments (**Supplementary Fig. 2**). The results of luciferase reporter assays in cell lines representative of megakaryocytes and osteoblasts showed that the differential binding detected by EMSA was functionally relevant and that both the 5′ UTR and intronic SNPs significantly reduced *RMB8A* promoter activity. The minor alleles, relative to the corresponding major alleles, were associated with significantly lower luciferase activity in human megakaryocytic CHRF and DAMI cell lines and the mouse osteoblast cell line MC3T3 (**Fig. 2b**). No effect of the minor allele of the 5′ UTR SNP was observed in human endothelial EAHY926 and HEK293 cells; the minor allele of the intronic SNP did exert an effect in HEK293 cells but not in EAHY926 cells (**Fig. 2b**). We next performed immunoblot staining of platelet lysates from three TAR cases (Unique Case Number (UCN) 10, 13 and 16, all with the 1.q21.1 deletion and 5′ UTR SNP combination) and their parents and an additional four cases for whom parental samples were not available: three with the 1.q21.1 deletion and either the 5′ UTR SNP (UCN 83 and 113) or the intronic SNP (UCN 64) and one with the 4-bp insertion in *RBM8A* in combination with the 5′ UTR SNP (UCN 33) (**Supplementary Fig. 3**). Densitometry analysis of the protein blots showed a significant reduction in the levels of Y14, the protein encoded by *RBM8A*, in TAR cases compared to parents and healthy controls (**Fig. 2c**). Taken together, the genetic and biological data strongly support our

hypothesis that TAR results from insufficiency of the Y14 protein. The results from the luciferase assays suggest that the minor allele of the 5′ UTR SNP may code for decreased transcription relative to the major allele. Expression assays in platelet RNA samples from 12 healthy volunteers heterozygous for the 5′ UTR SNP, however, did not reveal a significant difference between transcript levels of the two alleles ($P = 0.91$, paired t test on allelic ratios; **Supplementary Fig. 4**). Therefore, what the exact mechanism is by which the noncoding SNPs lead to the decreased protein expression observed in TAR cases is still an open question.

We investigated whether there are any variants in strong linkage disequilibrium (LD) with either the 5′ UTR or the intronic SNP (**Supplementary Fig. 5**). We could identify no such candidates for the 5′ UTR SNP, and, in haplotype analysis using the four exome-sequenced TAR cases carrying the minor allele of the 5′ UTR SNP, this allele was present on at least two distinct haplotype backgrounds. This provides an additional line of evidence that the minor allele of the 5′ UTR SNP is causative in TAR. We did identify a rare noncoding SNP (chr. 1: 145,483,747; C/T) 25 kb upstream of *RBM8A* in high LD with the intronic SNP; Sanger sequencing confirmed that this variant was present in all 11 genotyped TAR cases carrying the minor allele of the intronic SNP. The data from the ENCODE Project and our own FAIRE-seq open-chromatin data in megakaryocytes indicate that this additional SNP is not located in a regulatory region, whereas the intronic SNP is. Increased protein binding to the minor allele of the intronic SNP further corroborates the assumption that this particular SNP is causative. We cannot exclude the possibility that the 5′ UTR and intronic SNPs are not causative variants in TAR; however, in light of the biological and genetic evidence, we believe this is unlikely.

Y14 is one of the four components of the EJC, which is involved in basic cellular functions, such as nuclear export and subcellular localization of specific transcripts[2,4], translational enhancement[5] and nonsense-mediated RNA decay (NMD)[1,3,4]. The *RBM8A* transcript is widely expressed[15] and is present in all hematopoietic lineages (**Supplementary Fig. 6**), and its encoded protein sequence is highly conserved between species (**Supplementary Fig. 7**). Given the important functions of the EJC, it is likely that a complete lack of Y14 in humans is not viable. Indeed, in *Drosophila melanogaster*, knockdown of its ortholog *tsu* leads to major defects in abdomen formation[16], and we found that knockdown of the orthologous *rbm8a* transcript in *Danio rerio* using antisense morpholinos resulted in extreme malformations and death at 2 d post-fertilization (**Supplementary Fig. 8**). These findings are comparable with those from studies of a *Xenopus laevis* knockdown model of *Eif4a3*, which encodes an interacting EJC component, showing that EJC has a central role in vertebrate embryogenesis[17]. Considered in this context, our results are compatible with both a dose-effect phenomenon and a lineage-dependent deficiency in Y14. The possibility of a dose-effect phenomenon is supported by the observation that simple haploinsufficiency is not sufficient to create an aberrant phenotype, as evidenced by the seemingly healthy carriers of the 1q21.1 deletion. We also did not observe an effect on platelet count for either the 5′ UTR or the intronic SNP in the 403 and 59 individuals from the Cambridge BioResource who carried the minor allele for each SNP, respectively (**Supplementary Table 3**). This suggests that compound inheritance of a null allele together with the minor allele of one of the two regulatory SNPs brings Y14 levels below a critical threshold in certain tissues. The cell line–dependent effect shown in the luciferase assays is likely to be the result of differences in the regulation of *RBM8A* gene expression by combinatorial binding of transcription factors (including EVI1) in the context of the regulatory SNPs. An additional mechanism by which a deficiency in Y14 (and therefore in EJC function) may not be ubiquitous is suggested by studies showing that NMD not only targets nonsense mRNAs but also regulates physiological mRNA abundance in a gene-specific manner (reviewed in ref. 18). For example, hematopoietic-specific knockdown in the mouse of *Upf2*, which encodes a core NMD component, resulted in complete disappearance of the hematopoietic stem cell compartment, whereas more differentiated cells were only mildly affected[19]. Finally, in addition to a tissue-dependent effect, it is possible that the regulatory SNPs have developmental stage–dependent consequences: in mouse, the *Mecom* gene encoding Evi1 is expressed in a transient manner in emerging limb buds[20]. This may provide an explanation for the skeletal abnormalities observed in TAR.

In conclusion, we have used DNA sequencing to uncover the genetic basis of TAR syndrome, and we have identified a genetic mechanism of compound inheritance involving a null allele combined with a low-frequency regulatory variant. This compound inheritance mechanism reduces Y14 abundance, probably in a cell type– and developmental stage–dependent manner. Whether the same mechanism underlies other Mendelian disorders, in particular, other microdeletion syndromes showing variable penetrance and expression, remains to be established, but these results highlight the importance of analyzing regulatory regions for causative mutations. Although we have shown altered protein-binding affinity for the minor alleles of the regulatory SNPs, the mechanisms by which these SNPs lead to reduced levels of the Y14 protein in platelets are not clear and may be different for the 5′ UTR and intronic SNPs. Although genetic defects in the minor spliceosome[21,22] and NMD[23] have been linked to human disease, to the best of our knowledge, TAR syndrome is the first human disorder shown to be caused by a defect affecting one of the four EJC subunits.

URLs. Cambridge BioResource, http://www.cambridgebioresource.org.uk/; The European Genome-phenome Archive (EGA), http://www.ebi.ac.uk/ega/; UK10K, http://www.uk10k.org/; Simple Protocol for annealing oligonucleotides, http://www.piercenet.com/files/TR0045-Anneal-oligos.pdf.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. The *RBM8A* mRNA reference sequence is available at NCBI (NM_005105). Sequencing data have been submitted to the European Genotype-Phenotype Archive (EGAD00001000018).

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank C. Langford and P. Ellis for performing the enrichment for the exome sequencing, S. Balasubramaniam for assistance with data processing and the Wellcome Trust Sanger Institute sequencing core for sequencing. We thank V. Mooser from GlaxoSmithKline, G. Waeber and P. Vollenweider from CoLaus and J. Durham, C. Scott and colleagues at the Sanger Institute for providing access to their collection of whole-exome sequencing data for the CoLaus cohort[12]. We thank R. Durbin, B. Göttgens and I. Palacios for comments on the manuscript. This study makes use of data generated by the UK10K Consortium, which were derived from samples from the TwinsUK cohort. A full list of the investigators who contributed to the generation of the data is available from their website (see URLs). The study was supported by grants from the National Institutes for Health Research (NIHR; RP-PG-0310-1002 to C.G., G.K., P.A.S. and W.H.O.), the British Heart Foundation (FS/09/039 to C.G. and RG/09/12/28096 to C.A.A.), project grants from the Wellcome Trust (WT-082597/Z/07/Z to A.C. and WT-084183/2/07/2 to J.C.S.), grants by the Deutsche Forschungsgemeinschaft (SCHU1421/3-1 to H.S.) and the Sanitätsrat Dr. Emil Alexander Hübner-und-Gemahlin-Stiftung (T114/17644/2008/sm

to H.S.), by the Excellentie Financiering KULeuven (EF/05/013), by research grants from the Fonds Wetenschappelijk Onderzoek-Vlaanderen (G.0490.10N and G.0743.09) and by a grant from the Research Council of the University of Leuven (Onderzoeksraad–K.U. Leuven; GOA/2009/13 to C.T., K.F. and C.v.G.). The project made use of NHS Blood and Transplant donors from the Cambridge BioResource (see URLs). This local resource for genotype-phenotype association studies is supported by a grant from the NIHR to the Cambridge Biomedical Research Centre (RBAG096 to J.C.S. and J.D.J.). D.S.P. and M.K. were supported by a Marie-Curie NetSim Initial Training Network grant (EC-215820). The French cases were collected with support from the Gis-Maladies Rares, Diagnostic Différentiel des Purpuras Thrombopéniques Idiopathiques Chroniques et des Thrombopénies Congénitales (DIATROC) program and INSERM (ANR-08-GENO-028-03). Funding for UK10K was provided by the Wellcome Trust (WT091310).

AUTHOR CONTRIBUTIONS

C.A.A. performed next-generation sequence, Sanger sequence, genetic and statistical analyses. D.S.P. performed EMSA experiments, FAIRE-seq experiments and analysis, and *in silico* transcription factor binding analysis under the supervision of P.D. H.S., K.F., J.F., K.S., C.T. and R.N.-E. ascertained deletion status for TAR cases. K.F. and C.T. performed luciferase assays. H.S., K.F., C.T., C.G. and C.M.H. performed protein blot experiments. J.C.S. performed the Sanger sequencing and analyzed the data. P.A.S. performed quantitative PCR (qPCR) and allele-specific expression experiments. J.D.J. performed allele-specific expression experiments. A.C. performed the zebrafish knockdown study with input from D.L.S. M.K. analyzed the megakaryocyte RNA sequencing (RNA-seq) data under the supervision of P.B. G.K. supervised exome sequencing. J.G.S. supervised the Cambridge BioResource study. N.H. and M.E.H. performed the CNV analyses. H.S., M.H.B., N.D., R.F., I.K., P.N., C.A.L.R., G.S., C.v.G., R.N.-E. and C.G. clinically characterized TAR cases. C.A.A., K.F., W.H.O. and C.G. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Kim, V.N., Kataoka, N. & Dreyfuss, G. Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. *Science* **293**, 1832–1836 (2001).
2. Le Hir, H., Gatfield, D., Izaurralde, E. & Moore, M.J. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20**, 4987–4997 (2001).
3. Lykke-Andersen, J., Shu, M.-D. & Steitz, J.A. Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* **293**, 1836–1839 (2001).
4. Palacios, I.M., Gatfield, D., St Johnston, D. & Izaurralde, E. An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. *Nature* **427**, 753–757 (2004).
5. Wiegand, H.L., Lu, S. & Cullen, B.R. Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. *Proc. Natl. Acad. Sci. USA* **100**, 11327–11332 (2003).
6. Hall, J.G. *et al.* Thrombocytopenia with absent radius (TAR). *Medicine (Baltimore)* **48**, 411–439 (1969).
7. Klopocki, E. *et al.* Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. *Am. J. Hum. Genet.* **80**, 232–240 (2007).
8. Geddis, A.E. Inherited thrombocytopenia: congenital amegakaryocytic thrombocytopenia and thrombocytopenia with absent radii. *Semin. Hematol.* **43**, 196–203 (2006).
9. Greenhalgh, K.L. *et al.* Thrombocytopenia-absent radius syndrome: a clinical genetic study. *J. Med. Genet.* **39**, 876–881 (2002).
10. Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat. Genet.* **41**, 1182–1190 (2009).
11. The 1000 Genomes Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
12. Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* **8**, 6 (2008).
13. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
14. Paul, D.S. *et al.* Maps of open chromatin guide the functional follow-up of genome-wide association signals: application to hematological traits. *PLoS Genet.* **7**, e1002139 (2011).
15. Salicioni, A.M. *et al.* Identification and structural analysis of human RBM8A and RBM8B: two highly conserved RNA-binding motif proteins that interact with OVCA1, a candidate tumor suppressor. *Genomics* **69**, 54–62 (2000).
16. Hachet, O. & Ephrussi, A. *Drosophila* Y14 shuttles to the posterior of the oocyte and is required for *oskar* mRNA transport. *Curr. Biol.* **11**, 1666–1674 (2001).
17. Haremaki, T., Sridharan, J., Dvora, S. & Weinstein, D.C. Regulation of vertebrate embryogenesis by the exon junction complex core component Eif4a3. *Dev. Dyn.* **239**, 1977–1987 (2010).
18. Nicholson, P. *et al.* Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cell Mol. Life Sci.* **67**, 677–700 (2010).
19. Weischenfeldt, J. *et al.* NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev.* **22**, 1381–1396 (2008).
20. Perkins, A.S., Mercer, J.A., Jenkins, N.A. & Copeland, N.G. Patterns of *Evi-1* expression in embryonic and adult tissues suggest that *Evi-1* plays an important regulatory role in mouse development. *Development* **111**, 479–487 (1991).
21. Edery, P. *et al.* Association of TALS developmental disorder with defect in minor splicing component *U4atac* snRNA. *Science* **332**, 240–243 (2011).
22. He, H. *et al.* Mutations in *U4atac* snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science* **332**, 238–240 (2011).
23. Tarpey, P.S. *et al.* Mutations in *UPF3B*, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. *Nat. Genet.* **39**, 1127–1133 (2007).

ONLINE METHODS

Samples. All study subjects fulfill the diagnostic criteria for TAR syndrome: bilateral radial aplasia in the presence of both thumbs and thrombocytopenia. Further clinical details are given in **Supplementary Table 1**. Informed consent was obtained from all study subjects with approval from the ethics committees of the following institutions: University Hospital Bristol (MREC/00/6/72), Universitair Ziekenhuis Leuven (ML-3580), University of Cambridge (REC 10/H0304/66, REC 10/H0304/65), INSERM (RBM 1-14) and Charité Universitätsmedizin Berlin (EA2/170/05).

Exome sequencing. We applied the Agilent SureSelect protocol (G3362A) to enrich for 39.3 Mb of exonic sequence[24]. The enriched DNA was sequenced on the Illumina Genome Analyzer II platform. We generated 13.1–13.5 Gb of sequence per individual, resulting in a mean coverage of 123–127-fold, and 89.9–90.5% of the targets were covered by at least tenfold.

Sequence analysis. Sequence analysis was performed as described previously[25], with the main difference being that, here, we considered sequence variants with allele frequency of up to 5%, as inferred from variation data from dbSNP131, the 1000 Genomes Project[11] and 354 exomes from the CoLaus cohort[12].

Genotyping of the 5′ UTR and intronic SNPs in Cambridge BioResource samples. The *RBM8A* 5′ UTR and intronic SNPs were genotyped in 7,504 individuals from the Cambridge BioResource with custom TaqMan SNP Genotyping Assays (Applied Biosystems) according to the manufacturer's protocols. All genotyping data were scored twice by different operators. The genotype counts and the corresponding estimated MAFs for both variants are shown in **Supplementary Table 2**. There was no evidence for deviation from Hardy-Weinberg equilibrium (**Supplementary Table 2**).

Preparation of primary human megakaryocytes. Megakaryocytes were obtained from cord blood–derived $CD34^+$ hematopoietic stem cells by culture for 7 d in CellGro SCGM medium (CellGenix) supplemented with human recombinant thrombopoietin (THPO) and interleukin-1β (IL-1β)[26].

Sequencing of megakaryocyte RNA. Megakaryocyte RNA was sequenced as described previously[25]. Then, reads were aligned to the February 2009 *Homo sapiens* high-coverage assembly (hg19) using GSNAP[27] version 2011-03-28. Read trimming was disabled, and we allowed for up to five mismatches and newly identified splicing sites being at most 100,000 bp apart. Visualization with the Integrative Genomics Viewer[28] showed that the *RBM8A* gene is transcribed in megakaryocytes, as confirmed by quantitative PCR (qPCR; see **Supplementary Note**).

Formaldehyde-assisted isolation of regulatory elements (FAIRE)-seq. Primary megakaryocytes from three unrelated individuals were obtained as described above. For each sample, we cross-linked approximately 15 million primary megakaryocytes with 1% formaldehyde for 12 min at room temperature and then performed FAIRE experiments as previously described[14,29]. DNA from FAIRE was processed following the Illumina paired-end library generation protocol. Libraries were sequenced with 54-bp paired-end reads on the Illumina Genome Analyzer II and aligned as described previously[25]. In order to reduce experimental noise from individual preparations, we pooled the read fragments from the three individuals. The coverage profile on the combined data was created using the R packages ShortRead[30] and rtracklayer[31].

Predicted transcription factor binding site analysis. Transcription factor binding sites were annotated using MatInspector[32] software with the following parameters: library version, 8.3 (October 2010); matrix group, general core promoter elements and vertebrates; core, 1.00; and matrix, optimized+0.02.

Electrophoretic mobility shift assays (EMSAs). CHRF-288-11 cells were cultured as previously described[14]. Nuclear protein extracts were prepared using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher Scientific) following the manufacturer's instructions. Oligonucleotides for gel-shift assays were as follows: for the 5′ UTR SNP on chr. 1 at 145,507,646, 5′-biotin-AGTGTCTGAGCGGCACAGAC(G/A)AGATCTCGATCGAAGG-′3, and, for the intronic SNP on chr. 1 at 145,507,765, 5′-biotin-AGACGGCTGGTGGGAAGC(G/C)GGGAAGGTGCGAGAGAAGG-′3. Competitor probes were prepared without biotin labels. The labeled strands were annealed with the unlabeled complementary strands using the 'Simple Protocol' (see URLs). All oligonucleotides were obtained from Sigma-Aldrich. We performed gel-shift assays as previously decribed[14]. For competition assays, we used 100-fold molar excess of the unlabeled probes. For the 5′ UTR SNP, supershift experiments were performed with antibody to EVI1 (2 µl; sc-8707 X, Santa Cruz Biotechnology). Reactions were incubated for 45 min at room temperature. Reaction products were separated by electrophoresis for 75 min on 6% DNA Retardation Gels (Invitrogen) in 0.5× Novex TBE Running Buffer (Invitrogen) and were then transferred onto nylon membranes (Biodyne B, Thermo Fisher Scientific) and detected using the Chemiluminescent Nucleic Acid Detection Module (Thermo Fisher Scientific). For the intronic SNP, supershift experiments were performed using antibodies to MZF1 (2 µl; sc-46179 X and sc-66991 X) and RBPJ (2 µl; sc-28713 X) antibodies (all from Santa Cruz Biotechnology) with 0.1 mM EDTA in the binding reaction and 120 min incubation at room temperature.

***RBM8A* promoter activity by luciferase reporter assay.** Cotransfection experiments in different cell lines (EAHY926, HEK296, MC3T3, CHRF-288-11 and DAMI) were performed with pEGFP vector (Clontech) and *RBM8A* reporter plasmid (wild type or with the 5′ UTR or intronic SNP) constructed from the pGL3-Basic luciferase vector (Promega). The *RBM8A* promoter region, starting at −303 nt upstream of the transcription start site and including exon 1 and the first 142 nt of intron 1, was cloned 5′ to the luciferase gene. For each cotransfection assay, cells were transfected using Lipofectamine (Life Technologies) with 2 µg of pEGFP and 4 µg of *RBM8A*-pGL3 plasmid for HEK293, EAHY926 and MC3T3 cells. DAMI and CHRF cells were transfected using the Amaxa electroporation system (method X-01; Lonza). Luciferase activity was determined as described[33]. Each plasmid was assayed in six separate transfection experiments, and firefly luciferase activity was normalized to EGFP expression. Statistical analysis was performed using InStat 3.01 software (GraphPad).

Y14 protein expression analysis in platelet extracts. Blood (20 ml) anticoagulated with 3.8% trisodium citrate was centrifuged at 200*g*, and the platelet-rich plasma (PRP) was centrifuged at 700*g* with 0.1 volume of ACD buffer (2.5% trisodium citrate, 1.5% citric acid and 2% D-glucose, pH 4.5) to obtain a platelet pellet that was then lysed in ice-cold lysis buffer (1% IGEPAL (CA-630, Sigma Chemical), 1 mM EDTA, 2 mM DTE and one Complete protease inhibitor cocktail tablet (Roche) per 50 ml of PBS) and cleared of insoluble debris by centrifugation at 16,100*g* for 10 min at 4 °C. Protein fractions were mixed with 5% SDS reducing sample buffer, separated by SDS-PAGE and transferred to Hybond ECL-nitro-cellulose membrane (GE Healthcare). After being blocked with TBS with Tween-20 supplemented with 5% nonfat dry milk, the blots were incubated with primary antibody to Y14, Gsα or β-actin and were then incubated with horseradish peroxidase (HRP)-conjugated secondary antibody. Signal was detected with enhanced chemiluminescence (ECL) reagent (Thermo Scientific Pierce). The following primary antibodies were used: rabbit polyclonal antibody to Y14 (Q-24), mouse monoclonal antibody to Y14 (4C4) (both from Santa Cruz Biotechnology), mouse monoclonal antibody to Gsα (ref. 34) and mouse monoclonal antibody to β-actin (A5441, Sigma Chemical). Both Y14 antibodies were tested for their specificity using recombinant Y14-GST purified by sepharose beads as described[34]. Densitometry analysis was carried out using ImageJ64 software.

Allele-specific expression. Leukocyte-depleted platelet pellets were generated from EDTA-anticoagulated blood taken from Cambridge BioResource donors heterozygous for the 5′ UTR SNP through serial centrifugation and leukocyte depletion with antibody to CD45 conjugated to magnetic beads (Dynabeads CD45, 111.53D; Invitrogen) as described previously[35]. The pellets were resuspended in 2 ml of TRIzol (Invitrogen), and RNA was prepared essentially according to the manufacturer's instructions. After treatment of the samples with TURBO DNA-free reagent (Applied Biosystems), cDNA was generated using the Superscript III method with random hexamers (Invitrogen).

doi:10.1038/ng.1083

Genomic DNA was prepared from whole blood using the guanidine hydrochloride–chloroform method. PCR was performed to amplify exon 1 of *RBM8A* from genomic DNA and cDNA using AmpliTaq GOLD (Applied Biosystems), dNTPs (800 nM; GE Healthcare) and the primers described in **Supplementary Table 2**, with the following reaction conditions: 95 °C for 10 min; 95 °C for 15 s and 66 °C for 30 s for 5 cycles, with the temperature of the second step decreasing by 1 °C with each cycle; 95 °C for 15 s and 60 °C for 30 s for 30 cycles; and incubation at of 72 °C for 7 min and then 4 °C. PCR products were purified by spin column (D4014, Zymo Research) and ligated into the pCR2.1-TOPO vector (Invitrogen) at 20–25 °C for 2 h. Ligation mix (4 µl) was used to transform chemically competent TOP10 cells (Invitrogen), and cells were plated onto FastMedia Amp XGal Agar (InvivoGen). After overnight growth, white colonies were picked into separate wells of 96-well PCR plates, and colony PCR was performed with AmpliTaq GOLD and the primers described in **Supplementary Table 2**, using the following cycling conditions: 95 °C for 10 min; 95 °C for 15 s, 54 °C for 45 s and 72 °C for 15 s for 5 cycles, with the annealing temperature decreasing by 1 °C with each cycle; 95 °C for 15 s, 48 °C for 45 s and 72 °C for 15 s for 30 cycles; and incubation at 72 °C for 7 min and then 4 °C. PCR products were genotyped with custom TaqMan SNP Genotyping Assays (Applied Biosystems) according to the manufacturer's protocols.

qPCR in hematopoietic lineages. cDNA was prepared from leukocyte-depleted pellets as described above. cDNA preparation from the other hematopoietic lineage has been described[36]. TaqMan gene expression analysis was performed on cDNA using proprietary reagents, according to the manufacturer's instructions (Applied Biosystems). We used the GAPDH (Hs99999905_m1) and RBM8A (Hs4234933_g1) assays. Assays were conducted in 384-well format on a 7900HT Sequence Detection System (Applied Biosystems), and the threshold cycle number (C_T) for *GAPDH* was subtracted from that of the other genes assayed on that sample (ΔC_T) to normalize for reaction loading.

24. Coffey, A.J. *et al.* The GENCODE exome: sequencing the complete human exome. *Eur. J. Hum. Genet.* **19**, 827–831 (2011).
25. Albers, C.A. *et al.* Exome sequencing identifies *NBEAL2* as the causative gene for gray platelet syndrome. *Nat. Genet.* **43**, 735–737 (2011).
26. Tijssen, M.R. *et al.* Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev. Cell* **20**, 597–609 (2011).
27. Wu, T.D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
28. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
29. Giresi, P.G. & Lieb, J.D. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods* **48**, 233–239 (2009).
30. Morgan, M. *et al.* ShortRead: a Bioconductor package for input, quality assessment, and exploration of high throughput sequence data. *Bioinformatics* **25**, 2607–2609 (2011).
31. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).
32. Cartharius, K. *et al.* MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**, 2933–2942 (2005).
33. Freson, K. *et al.* 391 C to G substitution in the regulator of G-protein signalling-2 promoter increases susceptibility to the metabolic syndrome in white European men: consistency between molecular and epidemiological studies. *J. Hypertens.* **25**, 117–125 (2007).
34. Freson, K. *et al.* Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. *Blood* **98**, 85–92 (2001).
35. Gnatenko, D.V. *et al.* Transcript profiling of human platelets using microarray and serial analysis of gene expression. *Blood* **101**, 2285–2293 (2003).
36. Watkins, N.A. *et al.* A HaemAtlas: characterizing gene expression in differentiated human blood cells. *Blood* **113**, e1–e9 (2009).