CHAPTER 6

Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in *RBM8A* causes TAR syndrome.

This chapter is based on the following publication:

Albers, C.A., Paul, D.S., Schulze, H., et al. (2012). Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit *RBM8A* causes TAR syndrome. Nat. Genet. *44*, 435-439.

Collaboration note:

Section 6.2: Cornelis A. Albers[1–3] performed next-generation sequence, Sanger sequence, genetic and statistical analyses. Graham Kiddle[1,2] supervised exome sequencing. Jonathan C. Stephens[1,2] performed Sanger sequencing and analysed the data. Harald Schulze[4,5], Kathleen Freson[6], Janine Fiedler[5,7], Kenneth Smith[8,9], Chantal Thys[6] and Ruth Newbury-Ecob[8,9] ascertained deletion status for TAR cases. Harald Schulze, Martijn H. Breuning[10], Najet Debili[11], Rémi Favier[11], Ingrid Krapels[12], Paquita Nurden[13], Claudia A.L. Ruivenkamp[10], Gabriele Strauss[14], Chris van Geet[6,15], Ruth Newbury-Ecob and Cedric Ghevaert[1,2] clinically characterised TAR cases. I did not contribute to the analyses and experiments described in this section.

Section 6.3: Kathleen Freson and Chantal Thys performed luciferase assays. Harald Schulze, Kathleen Freson, Chantal Thys, Cedric Ghevaert and Catherine M. Hobbs[1,2] performed protein blot experiments. Myrto Kostadima[16] and Paul Bertone[16] analysed the megakaryocyte RNA sequencing data. I performed FAIRE-seq experiments and analysis, EMSA studies and *in silico* transcription factor binding analysis.

[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]Department of Biology, Chemistry, and Pharmacy, Freie University Berlin, Berlin, Germany; [8]Division of Child Health, University of Bristol, Bristol, UK; [9]Department of Clinical Genetics, St Michael's Hospital, Bristol, UK; [10]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; [11]Institut National de la Santé et de la Recherche Médicale (INSERM) U790, Villejuif, France; [12]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands; [13]Laboratoire d'Hématologie, Centre de Référence des Pathologies Plaquettaires, Hopital Xavier Arnozan, Pessac, France; [14]Department of Pediatric Oncology and Hematology, Charité Universitätsmedizin, Berlin, Germany; [15]Department of Pediatrics, Universitair Ziekenhuis Leuven, Leuven, Belgium; [16]European Molecular Biology Laboratory (EMBL)–European Bioinformatics Institute (EBI), Hinxton, Cambridge, UK.

6.1. Introduction

In **Chapters 3** and **4**, I applied FAIRE-generated maps of open chromatin to functionally assess sequence variants associated with complex traits. As a proof-of-concept, I subsequently investigated the molecular basis of the association between the non-coding GWA index SNP rs342293 and platelet volume and function at chromosome 7q22.3 (**Chapter 5**).

The final objective of this thesis was to explore the use of open chromatin maps to annotate low-frequency variants linked to a rare disease. For this purpose, I considered variants identified through exome sequencing of patients with thrombocytopenia with absent radii (TAR), a rare inherited blood and skeletal disorder. In this chapter, we functionally assess the candidate causal variants, found to be located in an NDR, and apply the experimental approach described in **Chapter 5** to establish the underlying biological mechanism.

6.1.1. Exome sequencing as a tool for gene discovery in rare diseases

In most cases, rare Mendelian diseases are caused by rare mutations, as selection acts strongly against these alleles. Strategies for identifying causal alleles depend on various factors including the structure of the pedigree or population, the mode of inheritance of a trait and the extent of locus heterogeneity.

Highly parallel sequencing has been successfully applied to identify causal mutations for monogenic disorders. This approach has been used to target genes within linkage intervals (Volpi et al., 2009; Nikopoulos et al., 2010), all protein-coding regions in the genome, referred to as 'exome' (Ng, Buckingham, et al., 2010; Ng, Bigham, et al., 2010), or whole genomes (Lupski et al., 2010; Roach et al., 2010). In recent years, many large-scale medical sequencing projects have focused on exome sequencing. One reason for this is cost, as whole-genome sequencing is still relatively expensive for large sample sizes. Another reason is biology, as most known examples of disease-causing variants alter the protein sequence, and functional assessment of non-coding genetic variation has been challenging.

There are four main strategies for identifying rare disease-causing variants through exome sequencing (Cirulli & Goldstein, 2010; Bamshad et al., 2011): (i) the sequencing of multiple affected but unrelated individuals; (ii) the sequencing of multiple affected individuals from the same pedigree; (iii) the sequencing of trios (parents and child) for identifying *de novo* mutations; and (iv) the sequencing of individuals at the extreme ends of a quantitative trait distribution.

For sequencing only a subset of the genome using next-generation DNA sequencing technology, 'genome-partitioning' methods are used. This requires the preparation of complex mixtures of sequencing templates that are highly enriched for the targeted genomic regions. In most exome sequencing protocols, the protein-coding fraction of the genome is selected by either solid-phase (Albert et al., 2007; Okou et al., 2007; Porreca et al., 2007) or liquid-phase hybridisation (Gnirke et al., 2009) to a complementary set of tiling oligonucleotide probes. After target enrichment, the regions are sequenced to great depth, i.e. a mean coverage of greater than 80-fold. In order to isolate pathogenic mutations from background polymorphisms, identified variants are filtered primarily based on function and frequency (reviewed in Stitziel et al., 2011 and Bamshad et al., 2011).

6.1.2. Genetics of thrombocytopenia with absent radii (TAR) syndrome

The thrombocytopenia with absent radii (TAR) syndrome is characterised by bilateral radial aplasia (the absence of the radius bones in the forearms) and severe thrombocytopenia (the reduction in the number of platelets). The incidence of TAR is estimated at 1:200,000–1:100,000 (http://www.ncbi.nlm.nih.gov/books/NBK23758/). However, many pregnancies are aborted if TAR is detected, therefore the real incidence may be higher. An excess of affected females has also been suggested (Greenhalgh et al., 2002). In contrast to other syndromes that combine absence of the radius with blood abnormalities, such as Fanconi anaemia, the thumb is preserved in TAR (Shaw & Oliver, 1959; Hall et al., 1969; Geddis, 2006). As illustrated in **Figure 6-1**, 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 (Greenhalgh et al., 2002). Individuals with TAR have low numbers of MKs and frequently present with bleeding episodes in the first year of life, which diminish in frequency and severity with age. In TAR, platelet levels are generally below $50x10^9$ platelets per litre, with the normal range being $150–350x10^9$ platelets per litre.

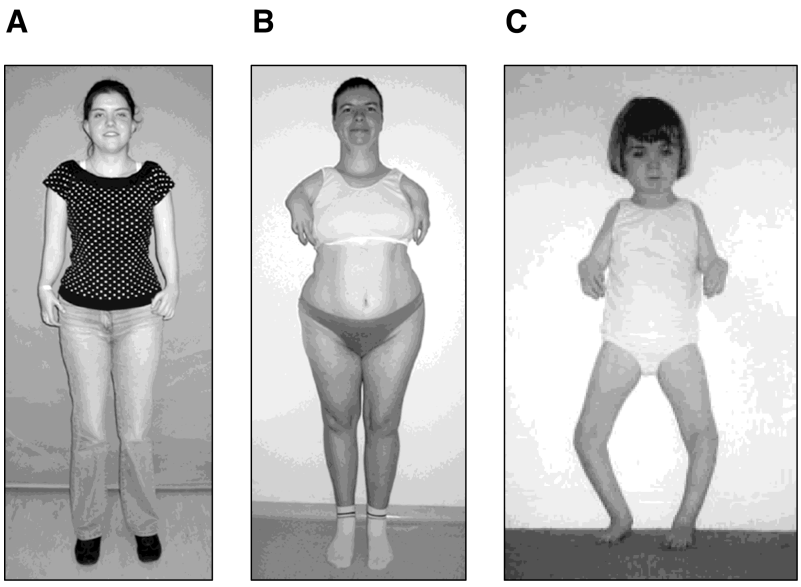

Figure 6-1. Skeletal abnormalities in TAR cases. (**A**) Patient shows mild upper-limb involvement with slightly reduced lengths of the arms. (**B**) Severe TAR phenotype with phocomelia (congenital absence of the proximal portion of a limb or limbs). (**C**) Lower limb involvement in a child with TAR, i.e. severe bowing of the legs. The pictures were adapted from Greenhalgh et al., 2002 and Klopocki et al., 2007.

An inherited or *de novo* deletion at chromosome 1q21.1 is found in the majority of affected individuals (Klopocki et al., 2007), 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 (chr1:145,399,075–145,594,214, build: hg19; 195 kb), as reported by Klopocki et al., 2007.

6.2. Most TAR cases have a low-frequency regulatory variant and a rare null allele at the *RBM8A* locus

To identify the additional causative allele, we selected five individuals with TAR ('cases') of European ancestry, who had the 1q21.1 deletion (**Figure 6-2 A**), and sequenced their exomes using the SureSelect Human All Exon Kit [Agilent Technologies] (**Section 2.15**). All study subjects fulfilled the diagnostic criteria for TAR syndrome as described in **Section 6.1.2**. The clinical and genotype information of the TAR cases and their healthy parents are provided in **Appendix, Table 8-8**. Per individual, 13.1–13.5 Gb of sequence was generated, resulting in a mean coverage of 123–127-fold, with 89.9–90.5% of the targets covered by at least 10-fold.

We were unable to find recessive novel mutations in the protein-coding regions in the five TAR patients (**Section 2.15**). However, four of the cases carried the minor allele of a low-frequency SNP (chr1:145,507,646; rs139428292G>A) in the 5'-UTR of the *RBM8A* gene, while the remaining case carried a previously unknown SNP (chr1:145,507,765G>C) in the first intron of the same gene (**Figure 6-2 B**). Genotyping by Sanger sequencing of additional 48 cases of European ancestry with the 1q21.1 deletion identified rs139428292 ('5'-UTR SNP' hereafter) and chr1:145,507,765 ('intronic SNP' hereafter) in 35 and 11 samples, respectively (**Figure 6-2 C**; **Appendix, Table 8-8**).

In total, 34 trios of mother, father and child were investigated (**Appendix, Table 8-8**). In all 25 trios of European ancestry, where the deletion in the child was not inherited *de novo*, we confirmed that the deletion and the newly identified SNPs were inherited from different parents. Therefore, the observed mutations were compatible with a compound autosomal recessive mode of inheritance. Among the 34 trios, there was one previously reported example of vertical transmission of TAR (Klopocki et al., 2007). Both the affected mother and her (aborted) foetus, which showed skeletal features of TAR on ultrasound, carried the typical 1q21.1 deletion. It is important to note that, in contrast to all other cases studied, these patients were of non-European ancestry. Sequencing of the entire *RBM8A* gene, including exons, introns, 5'-UTR, promoter, as well as a putative regulatory element 4 kb upstream of the promoter, showed an absence of the minor alleles of both the 5'-UTR and intronic SNPs in all three samples. We did not identify an alternative sequence variant as a potential additional causative allele. Thus, we have failed to identify the second causative allele in this sporadic case of vertical transmission of TAR. We reasoned that another longer-distance *cis*-acting or possibly *trans*-acting modifier of the *RBM8A* locus may explain the disorder in this pedigree.

From the genotyping of 7,504 healthy individuals of the Cambridge BioResource, we estimated MAFs of 3.05% and 0.42% for the 5'-UTR SNP and the intronic SNP, respectively (**Table 6-1**). Analysis of copy number variants at the chromosome 1q21.1 locus in 5,919 healthy individuals from the Wellcome Trust Case Control Consortium did not reveal deletions of the *RBM8A* gene in these individuals, indicating a low frequency of the 1q21 deletions found in TAR cases and their healthy relatives. This is in agreement with the low incidence of TAR syndrome in the population. We observed five duplications, which suggested that overexpression of *RBM8A* is not deleterious (The Wellcome Trust Case Control Consortium, 2010; Huang et al., 2010). Thus, the concurrent presence of one of the two non-coding SNPs at one allele and the 1q21.1 deletion at the other is strongly associated with TAR syndrome (estimated $P<5x10^{-228}$; Albers et al., 2012).

Table 6-1. Genotyping of the 5'-UTR and intronic SNPs at the *RBM8A* locus in 7,504 healthy individuals of the Cambridge BioResource and association with platelet count. Platelet count data was available for 6,805 and 6,938 of the 7,504 individuals genotyped for the 5'-UTR SNP and the intronic SNP, respectively. The number of individuals for each genotype of the 5'-UTR and intronic SNPs with measured platelet count is indicated in parentheses. The log-transformed platelet count on genotype was regressed using an additive genetic model, adjusted for gender and age in years at date of venesection. Abbreviations: MAF: minor allele frequency; HWE: Hardy-Weinberg equilibrium.

	5'-UTR SNP (G/A)	Intronic SNP (G/C)
Genotypes passed QC (call rate)	7,317 (97.5%)	7,458 (99.4%)
Homozygous major	6,879 (6,402)	7,396 (6,879)
Heterozygous	431 (396)	62 (59)
Homozygous minor	7 (7)	0 (0)
Estimated MAF	3.05%	0.42%
Deviation from HWE (exact test)	*P*=0.84	*P*=1.00
Association with platelet count	*P*=0.87	*P*=0.99

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. In the first case, we identified a 4 bp frameshift insertion at the start of the fourth exon, and established that the non-coding SNP and insertion were at different chromosomes. By genotyping the parents of this case, we identified the 4 bp insertion in the mother and the 5'-UTR SNP in the father, both as heterozygous positions (**Appendix, Table 8-8**). In the second case, we identified a nonsense mutation in the last exon of *RBM8A* (**Figure 6-2 B,C**). Both mutations were absent from 458 exome samples of the 1000 Genomes Project (The 1000 Genomes Project Consortium, 2010) and 416 samples from the Cohorte Lausannoise ('CoLaus'; Firmann et al., 2008). We concluded 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 non-coding 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 non-coding SNPs in the 5'-UTR or first intron).

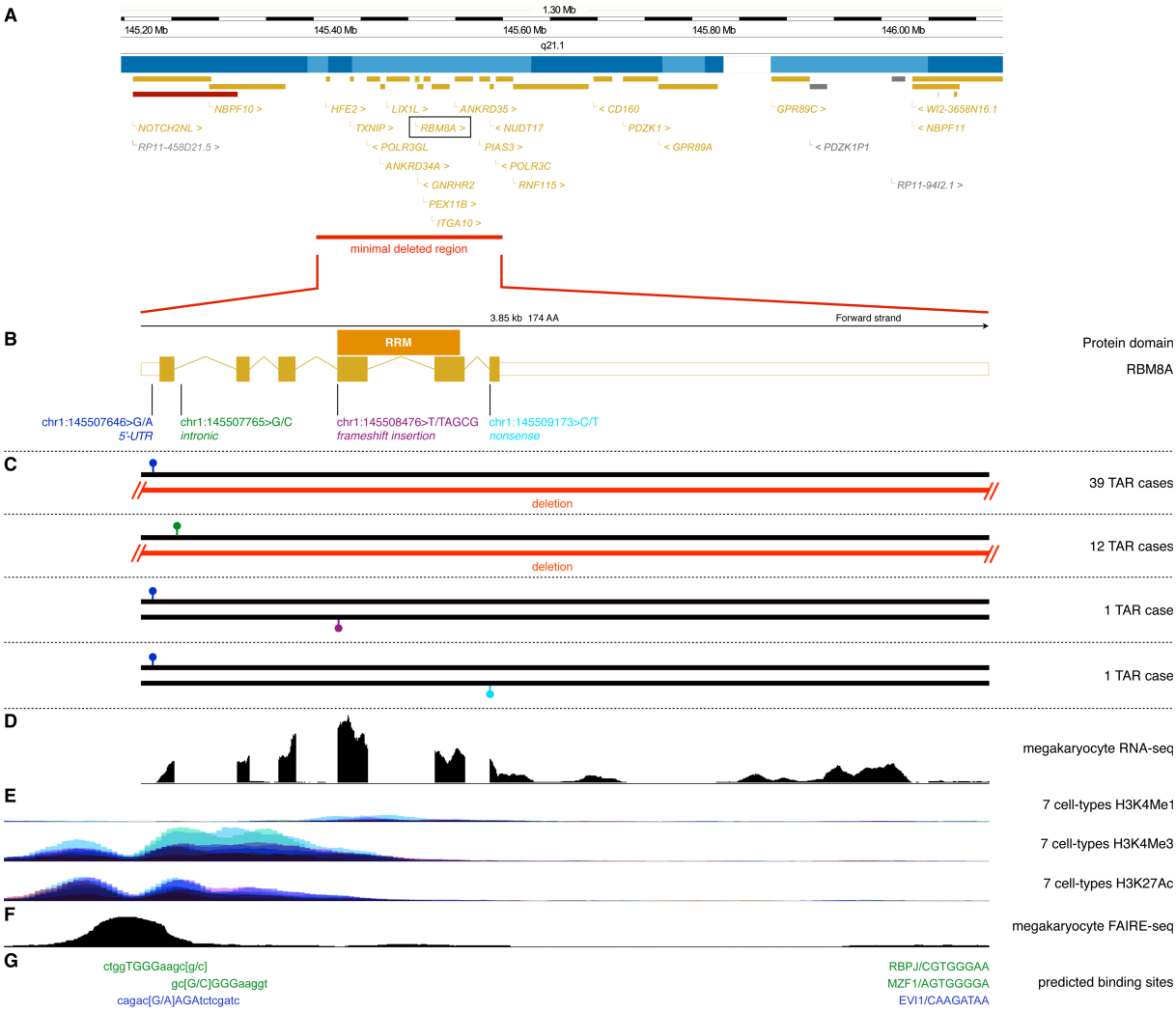

Figure 6-2. Low-frequency non-coding SNPs and a rare null mutation at the *RBM8A* locus. (**A**) Fifty-three of 55 TAR cases were heterozygous carriers of a rare 1q21.1 deletion of varying size. The red bar indicates the region that was absent in all 53 cases having a deletion. Legend: Yellow bars: genes; grey bars: pseudogenes; blue bars: contigs. (**B**) The *RBM8A* transcript is shown. The sequence encoding the RNA-binding domain (RRM) is indicated by the orange bar above the transcript. (**C**) We identified two low-frequency regulatory SNPs in 53 of a total of 55 TAR cases studied. The first, at chr1:145,507,646 (rs139428292G>A), is located at the 5'-UTR of *RBM8A* and has a population MAF of 3.05% (dark blue). The second, at chr1:145,507,765G>C, is located at 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 at one chromosome and the 1q21.1 deletion at the other. Twelve TAR cases carried the minor allele of the intronic SNP at one chromosome and the 1q21.1 deletion at the other. Compound inheritance of the 1q21.1 deletion and one of the two regulatory SNPs was strongly associated with TAR. 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 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 MKs provided evidence that *RBM8A* is transcribed in MKs. Shown is the sequencing read depth across the *RBM8A* locus. (**E**) Histone modifications in seven cell lines (GM12878, H1-hESC, HSMM,

HUVEC, K562, NHEK and NHLF) reported by the ENCODE Project (The ENCODE Project Consortium, 2007) indicated the presence of active regulatory elements at the promoter region, including the 5'-UTR and first intron of *RBM8A*. Shown are the read depths resulting from ChIP-seq experiments of the three histone marks in the seven cell lines. The coverage profiles of the different cell types are represented by different shades of blue and are superimposed. (**F**) Coverage profile of FAIRE-seq experiments showed that the 5'-UTR and intronic SNPs are accessible to regulatory factors in MKs. (**G**) *In silico* transcription factor binding site analysis predicted that the minor allele of the 5'-UTR SNP creates a binding site for the EVI1 transcription factor. The minor allele of the intronic SNP was predicted to disrupt binding of MZF1 and RBPJ. Capital letters indicate the consensus transcription factor binding sites, and the alleles of the SNPs are shown in parentheses.

6.3. The effect of the regulatory SNPs on transcription factor binding, *RBM8A* promoter activity and protein expression in platelets

Analysis of histone modifications in seven human cell lines from the ENCODE Project (The ENCODE Project Consortium, 2007) indicated that the 5'-UTR and the intronic SNP are located in potential active regulatory elements (**Figure 6-2 D,E**). Annotation of open chromatin structure using the FAIRE-seq technique provided additional evidence in MKs (**Figure 6-2 F**). Computational predictions suggested 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 (**Figure 6-2 G**).

I confirmed the prediction of EVI1 binding by EMSAs in the megakaryocytic cell line CHRF-288-11, in which the EVI1 protein bound to the oligonucleotide probe carrying the minor allele of the 5'-UTR SNP but only weakly associated with the major allele (**Figure 6-3 A**). EMSA studies for the intronic SNP showed a decrease in the binding of nuclear proteins to the minor allele, although I could not confirm the presence of either MZF1 or RBPJ in supershift experiments (**Figure 6-3 B**).

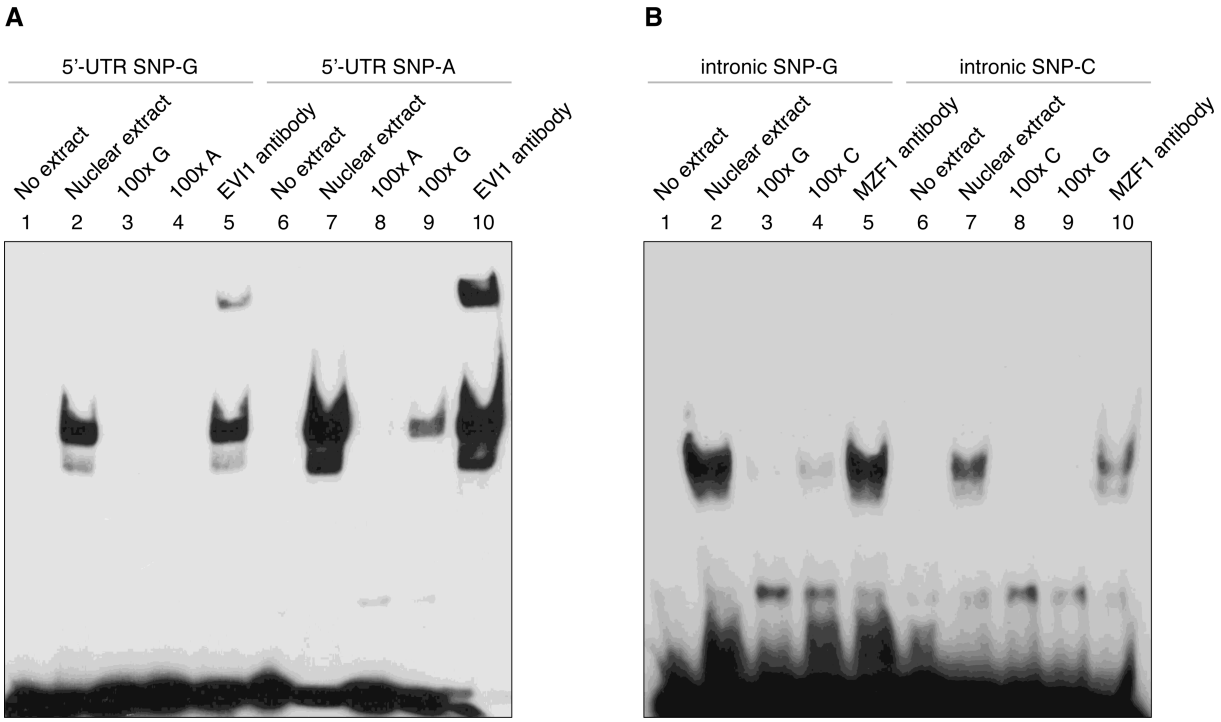

Figure 6-3. The effect of the regulatory 5'-UTR and intronic SNPs on transcription factor binding. (**A**) EMSAs in nuclear protein extracts from CHRF-288-11 cells showed higher protein affinity to probes with the A-allele (lane 7) than to probes with the G-allele (lane 2) of the 5'-UTR SNP. Binding of the A-allele probe was competed by a specific (lane 8) but not by an unspecific unlabelled probe (lane 9). I observed a supershift with an EVI1 antibody 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**) EMSAs for the intronic SNP showed higher protein affinity to probes containing the G-allele (lane 2) than the C-allele (lane 7). Protein binding of G-allele probes was competed by specific (lane 3) but not by unspecific unlabelled probes (lane 9). I performed supershift experiments with antibodies for the predicted transcription factors MZF1 and RBPJ. However, in my experiments none of the tested antibodies competed for binding and/or shifted the protein-DNA complex (lane 10; data not shown for RBPJ).

The results of luciferase reporter assays in cell lines representative of MKs and osteoblasts showed that the differential binding detected by EMSAs was functionally relevant and that both the 5'-UTR and intronic SNPs significantly reduced *RBM8A* promoter activity. The minor alleles, relative to the corresponding major alleles, were associated with significantly lower luciferase activity in human megakaryocytic CHRF-288-11 and DAMI cell lines and the mouse osteoblast cell line MC3T3 (**Figure 6-4**). 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 (**Figure 6-4**).

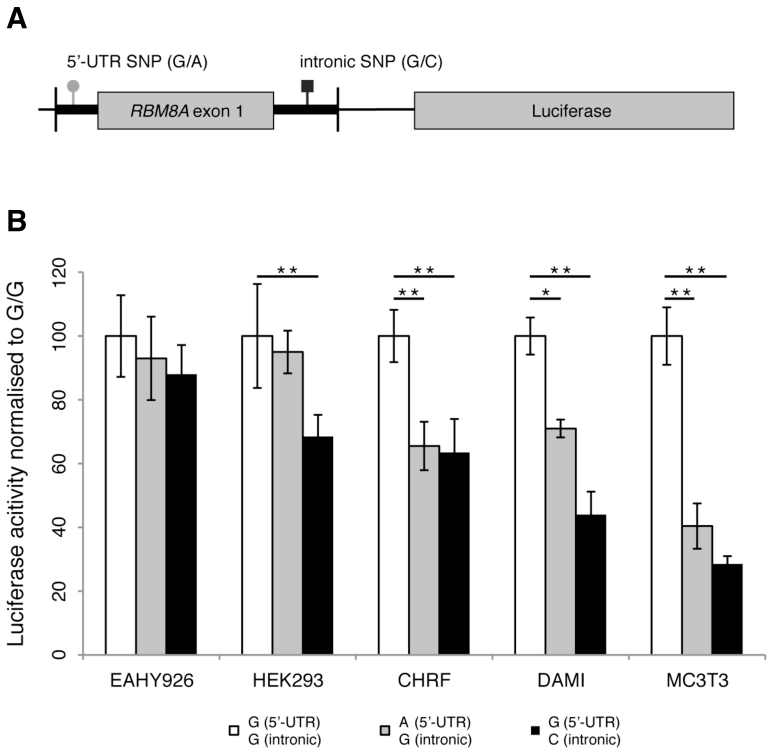

Figure 6-4. Luciferase reporter assays in cell lines representative of MKs (CHRF-288-11 and DAMI) and osteoblasts (MC3T3). (**A**) Schematic of the luciferase reporter construct with the 5'-UTR and intronic SNPs represented by circle and square symbols, respectively. (**B**) We observed significantly decreased *RBM8A* promoter activity for the minor alleles of both the 5'-UTR and intronic non-coding SNPs relative to the major alleles. No effect of the 5'-UTR SNP was observed in EAHY926 and HEK293 human endothelial cells. Error bars indicate standard deviations (s.d.). Statistical analysis was performed using the Tukey-Kramer multiple comparisons test, indicating $^{*}P<0.01$ and $^{**}P<0.001$. Luciferase activity was normalised with respect to the construct consisting of the major (G-) allele of both SNPs (indicated by G/G).

We next performed immunoblot staining of platelet lysates from seven TAR cases (all carrying the 1q21.1 deletion and either the 5'-UTR or intronic SNP), six unaffected 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), as well as six controls (**Figure 6-5 A**). 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 parental and healthy control samples (**Figure 6-5 B**).

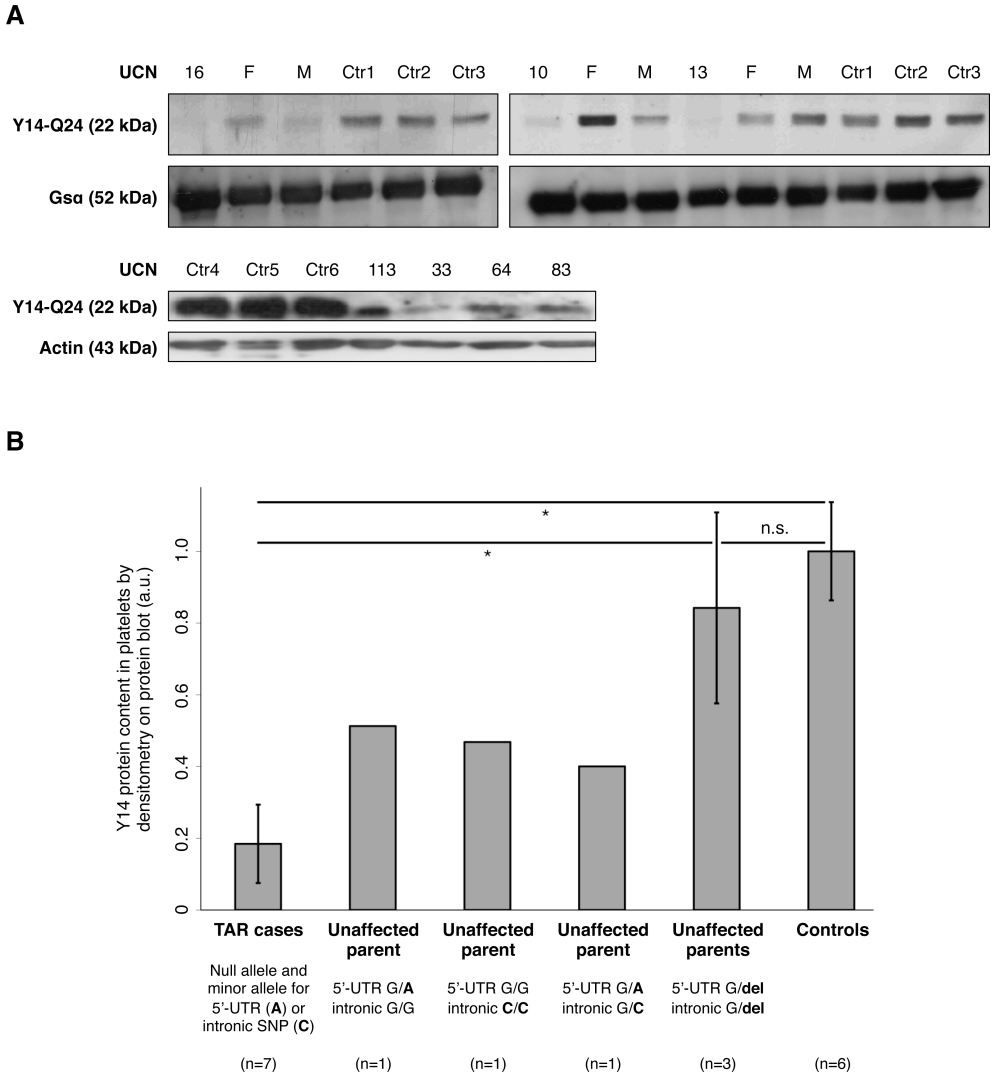

Figure 6-5. Immunoblot staining for Y14, the protein encoded by *RBM8A*, and densitometry analysis. (**A**) Western blot analyses of Y14 protein expression were performed in platelet lysates. We selected three TAR cases, all with the 1q21.1 deletion and the 5'-UTR SNP (UCNs 10, 13 and 16; **Appendix, Table 8-8**), and their six parents (labelled as 'F' and 'M' on the lane to the right of the TAR cases on the gel). In addition, we selected four TAR cases for which parental samples were not available: three with the 1q21.1 deletion and either the 5'-UTR SNP (UCNs 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). Protein expression of Gsα or β-actin was used as a loading control. (**B**) Densitometry analysis showed significantly reduced Y14 protein levels in TAR cases compared to parental and control samples. Error bars indicate s.d. Statistical analysis was performed using the heteroscedastic t-test, marking *P<0.01 and n.s. (not significant). Only genotype configurations indicated by lines were compared. The minor alleles of the 5'-UTR and intronic SNPs are shown in bold type. Abbreviations: UCN: unique case number; F: father; M: mother; Ctr: control; a.u.: arbitrary units.

Taken together, the genetic and biological data strongly supported our hypothesis that TAR results from insufficiency of the Y14 protein. The results from the luciferase assays suggested that the minor

allele of the 5'-UTR SNP may cause decreased transcription relative to the major allele. Expression assays in platelet RNA samples from twelve 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; Albers et al., 2012). Therefore, what the exact mechanism is by which the non-coding 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 LD with either the 5'-UTR SNP or the intronic SNP (Albers et al., 2012). We could identify no such candidates for the 5'-UTR SNP. 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 provided an additional line of evidence that the minor allele of the 5'-UTR SNP is causative in TAR. We identified a rare non-coding SNP (chr1:145,483,747C>T) 25 kb upstream of *RBM8A* in high LD with the intronic SNP. Sanger sequencing confirmed that this variant was present in all eleven 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 MKs indicated that this additional SNP was not located in a regulatory region, in contrast to the intronic SNP. Increased protein binding to the minor allele of the intronic SNP further corroborated 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 genetic and biological evidence, we believe this is unlikely.

6.4. Discussion

Y14 is one of the four components of the exon-junction complex (EJC), which is involved in basic cellular functions, such as nuclear export and subcellular localisation of specific transcripts (Le Hir et al., 2001; Palacios et al., 2004), translational enhancement (Wiegand et al., 2003) and nonsense-mediated RNA decay (NMD) (Kim et al., 2001; Lykke-Andersen et al., 2001; Palacios et al., 2004). The *RBM8A* transcript is widely expressed (Salicioni et al., 2000) and is present in all haematopoietic lineages (Albers et al., 2012). Its encoded protein sequence is highly conserved between species (Albers et al., 2012). 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 (Hachet & Ephrussi, 2001), 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-fertilisation (Albers et al., 2012). 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 (Haremaki et al., 2010). 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 shown by the seemingly healthy carriers of the 1q21.1 deletion.

We did not observe an effect on platelet count for both the 5'-UTR and intronic SNPs in the respective 403 and 59 individuals of the Cambridge BioResource who carried the minor allele of each SNP (**Table 6-1**). 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. Although the SNPs were directly genotyped, power to detect subtle effects of the SNPs on platelet count was limited due to the low MAFs of both SNPs. In addition, since the low platelet count in TAR cases often recovers in adolescence, and >94% of the genotyped individuals of the Cambridge BioResource were older than 20 years, the power to detect an effect of the SNPs on platelet count was expected to be limited.

The cell line-dependent effect shown in the luciferase assays was 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 has been suggested by studies showing that NMD not only targets nonsense mRNAs but also regulates physiological mRNA abundance in a gene-specific manner (Nicholson et al., 2010). For example, haematopoietic-specific knockdown of *Upf2* in mouse, which encodes a core NMD component, resulted in complete disappearance of the haematopoietic stem cell compartment, whereas more differentiated cells were only mildly affected (Weischenfeldt et al., 2008). 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 (Perkins et al., 1991). This may provide an explanation for the skeletal abnormalities observed in TAR.

In conclusion, we applied next-generation sequencing to uncover the genetic basis of TAR syndrome, and 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. However, these results highlight the importance of analysing regulatory regions even when searching for causative mutations in rare diseases. 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 (Edery et al., 2010; He et al., 2011), and NMD (Tarpey et al., 2007) 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.