Supplementary Information

Inheritance of low-frequency regulatory SNPs and a rare null mutation in exon-junction complex subunit *RBM8A* causes TAR

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

[1]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. [2]Department of Haematology, University of Cambridge, UK. [3]NHS Blood and Transplant, 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 - European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, UK. [8]Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands. [9]Institut National de la Santé et de la Recherche Médicale, Villejuif, France. [10]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, the Netherlands. [11]Laboratoire d'Hématologie, Centre de Référence des Pathologies Plaquettaires, Hopital Xavier Arnozan, Pessac, France. [12]Division of Child Health, University of Bristol. [13]Department of Clinical Genetics, St Michael's Hospital, Bristol, UK. [14]Department Pediatric Oncology & Hematology, Charité Universitätsmedizin, Berlin, Germany. [15]Department of Pediatrics, Universitair Ziekenhuis Leuven, Leuven, Belgium.

*These authors contributed equally

Correspondence should be addressed to C.A.A (caa@sanger.ac.uk) or C.G. (cg348@cam.ac.uk)

Supplementary Note

A. Clinical phenotypes

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).

B. Minimal deleted region

Quantitative-PCR primers 4 and 7 from Table 2 in reference[1] were used to define the region that was heterozygously deleted in all 30 TAR cases tested by Klopocki *et al*. [1]. The primer sequences for these amplicons are respectively TGAGTGGTCTTCGGGTGATAGA/CCCATCCCACTGAAAACTGAA and GGAAAATGGTAAAGGGTGGTG/CATCTTCCCTAACCTGGGAGA. The minimal heterozygously deleted region as chr1:145399075-145594214 was defined by mapping of these primers to reference build Hg19.

C. Statistical significance of observed configurations

From the genotyping of 7504 individuals from the Cambridge BioResource, MAFs of 3.05% and 0.41% were estimated for 5'UTR and intronic SNP respectively. The probability of seeing the 5'UTR or the intronic SNP minor allele on one hapotype is 1-(1-0.0305)x(1-0.0041)= 0.0345. Furthermore, no chromosome 1q21.1 deletion was

observed in 5919 healthy individuals from the Wellcome Trust Case Control Consortium collection of shared controls[2]. Assuming these healthy individuals are drawn at random from a population of 200,000 individuals, if the population allele frequency of the deletion is 0.001, the probability of not seeing a deletion in 5919 samples is $6x10^{-6}$. (If the population allele frequency would be 0.0005, the probability of not seeing a deletion in this sample is 0.002). To obtain a conservative estimate of the statistical significance we therefore assumed a population frequency of 0.001 for 1q21.1 deletions.

We estimate that the probability of 53 of 55 TAR cases carrying a 1q21 deletion, and 51 of 53 with a deletion carrying a 5'UTR or the intronic SNP, by chance is $\binom{55}{53}\binom{53}{51}0.001^{53}0.0345^{51}(1-0.001)^2(1-.0345)^2 = 5\times10^{-228}$. This may be interpreted as the P-value for the association of the 1q21.1 deletion and 5'UTR and intronic SNPs jointly with TAR syndrome.

Assuming a null allele is present on one chromosome, we estimate that the probability of seeing one of the two non-coding SNPs in 53 of 55 TAR cases by chance is $\binom{55}{53}0.0345^{53}(1-0.0345)^2 = 4\times10^{-75}$. This may be interpreted as the p-value for the association of the non-coding alleles with TAR syndrome, assuming a null allele is present on the alternate chromosome (deletion or frameshift/nonsense mutation). The p-value for the independent association of the 5'UTR SNP and intronic SNP with TAR is respectively $\binom{55}{41}0.0305^{41}(1-0.0305)^{14} = 2\times10^{-50}$ and $\binom{55}{12}0.0041^{12}(1-0.0041)^{43} = 8\times10^{-18}$. These significance estimates may be compared to

a significance threshold corrected for multiple testing of 50 million sequence variants (approximately the number of SNPs identified by Phase 1 of the 1000 Genomes Project) is $0.05/50x10^6 = 10^{-9}$.

D. Trio with genetically unexplained TAR

Thirty-four trios of mother, father and child with one previously reported example of vertical transmission of TAR[1] were investigated. In all 25 trios of European ancestry where the deletion or frameshift insertion was not inherited *de novo*, the observed mutations were compatible with a compound autosomal recessive mode of inheritance. In the trio with vertical transmission both the affected mother, of non-European ancestry, and her fetus, which on ultrasound showed skeletal features of TAR, carried the typical 1q21.1 deletion. Sequencing of the entire coding fraction, the introns, the 5'UTR, and the promoter of the *RBM8A* gene, as well as a putative regulatory element 4 Kb upstream of the promoter of the *RBM8A* locus (see SOM Table S3 for the regions and primer sequences) showed an absence of the minor alleles of the 5'UTR or the intronic SNP in all three samples and we also 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 reason that another longer-distance cis-acting, or possibly a trans-acting modifier of the *RBM8A* locus may explain the disorder in this pedigree.

E. Haplotype analysis

We first determined the recombination rates in the 1q21.1 interval using the recombination rates estimated for the CEU population by the HapMap project, as

distributed as part of the Impute2 phasing software (http://mathgen.stats.ox.ac.uk/impute/impute_v2.html). From this we inferred that the LD block containing *RBM8A* is approximately given by the NCBI37 coordinates chr1:145200000-145700000, an interval of ~500 kb (**Supplementary Fig. 5A**).

To identify carriers of the 5'UTR and intronic SNP minor allele, we consulted a database of sequence variation for 809 European individuals from the TwinsUK cohort from the UK10K Project (http://www.uk10k.org), who were whole-genome sequenced to ~6X coverage. We looked if sequence variants (SNPs as well as small insertions and deletions) of similar minor allele frequencies segregated in the carriers together with the 5'UTR SNP or the intronic SNP minor allele in the LD block containing *RBM8A*, as these variants could potentially be in LD with the 5'UTR SNP or intronic SNP we identified in the TAR cases. Specifically, for the 5'UTR SNP we considered sequence variants with minor allele frequencies (MAFs) between 0.015 and 0.05; for the intronic SNP we considered all variants with a MAF below 0.0075. We then required that at least 70% of the individuals carrying the 5'UTR or intronic minor allele also carried the minor allele for the other variant.

5'UTR SNP

Among the five TAR cases that we exome sequenced, four carried the 5'UTR SNP minor allele. As these individuals all carried the 1q21.1 deletion as well, we could unambiguously determine the haplotypes for these four TAR cases (**Supplementary Fig. 5B**).

We then consulted the latest release of the 1000 Genomes Project (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20101123/interim_phase1_release/), which contains high-quality phased haplotypes for 381 individuals of European ancestry. We identified 24 individuals carrying the 5'UTR SNP, and we compared the haplotypes on which the 5'UTR SNP minor allele was present in these 1000 Genomes individuals to the haplotypes in the 4 exome sequenced TAR cases (**Supplementary Fig. 5B**).

In the 1000 Genomes Project individuals, a number of distinct haplotypes with the 5'UTR SNP segregate in the LD block containing *RBM8A*. Among our 4 TAR cases, at least two of these are present. Next, we identified one other sequence variant (SNP chr1:145417662 T/C) in the 809 individuals with minor allele frequency between 1.5% and 5% that occurred on the same haplotype as a 5'UTR SNP minor allele in the 1000 Genomes individuals. However, of the 24 1000 Genomes individuals carrying the 5'UTR SNP minor allele, 7 did not carry the chr1:145417662 minor allele. We therefore did not consider this as an alternative candidate causative mutation in the TAR cases carrying the 5'UTR SNP minor allele.

Given that the 5'UTR cases were unrelated and we observed multiple haplotypes, if the 5'UTR SNP locus was not causative we would expect to see independent and likely different mutation events that do not need to co-occur with the 5'UTR SNP in order to cause TAR syndrome. Taken together these data suggest that the 5'UTR SNP minor allele is the causative mutation in the TAR cases carrying this variant.

Intronic SNP

The intronic SNP was not present in the 1000 Genomes Release. We found that 5 individuals in the UKTwins cohort carried the intronic SNP minor allele, in agreement with the frequency we estimated from the Cambridge BioResource.

We then looked in the LD block for SNPs and small insertions and deletions with a similar low allele frequency, variants with an estimated population allele frequency below 0.0075 and present in at least 3 individuals of the 5 individuals (in the cohort of 809 individuals) who carry the intronic SNP minor allele.

We identified two variants satisfying these criteria: a chr1:145483747 C/T SNP (25 kb upstream of *RBM8A*) and a chr1:145273877 G/A SNP (200 kB upstream of *RBM8A*). Using Sanger sequencing we genotyped these SNPs in 11 TAR cases carrying the intronic SNP. We found that the chr1:145483747 SNP was present in all of these TAR cases together with the intronic SNP, but not the chr1:145273877 SNP. Thus, in contrast with the 5'UTR SNP, the intronic SNP seems to segregate on a single rare haplotype that contains at least one sequence variant of similar low frequency.

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 further corroborates the assumption that the intronic SNP is causative. We cannot exclude the possibility that the 5'UTR SNP or the intronic SNP are not the causative variants; however, in light of the biological and genetic evidence we believe this is unlikely.

Finally, because of the 10-fold lower frequency of the intronic SNP, we resequenced the complete *RBM8A* gene (exons, introns, 5'UTR and 3'UTR) in two TAR cases carrying the *RBM8A* intronic SNP to exclude variants that might have been undetected in the 809 individuals. However, we could not identify other variants in the *RBM8A* gene unit with similar frequency as the intronic SNP.

F. Association with platelet count

Platelet count was available for 6805 and 6938 of the 7504 healthy individuals genotyped for respectively the 5'UTR SNP and the intronic SNP. The log-transformed platelet count on genotype was regressed using an additive genetic model adjusted for age in years at date of venesection and gender. For both SNPs there was no statistically significant effect on platelet count (**Supplementary Table 3**). Although the SNPs were directly genotyped, power to detect subtle effects of the SNPs on platelet count is limited due to the low MAF of both SNPs. Also, since the low platelet count in TAR cases often recovers in adolescence, and >94% of the genotyped Cambridge BioResource individuals were older than 20 years, power to detect an effect of the SNP on platelet count is expected to be limited.

G. Cambridge BioResource

The Cambridge BioResource (ref. [3] and http://www.cambridgebioresource.org.uk) is funded by the National Institute for Health Research Cambridge Biomedical Research Centre. Informed consent was obtained from all volunteers upon recruitment to the Cambridge BioResource for the collection and use of DNA samples for genotyping and for association testing with phenotypic traits.

Supplementary Tables

Unique Case Number	TAR diagnosed	Heterozygous 1q21 deletion	1q21 deletion origin (de novo, parent, Unkn.)	Genotype 5'UTR SNP 145507646 G/A	Genotype intronic SNP 145507765 G/C	Sex	Age (years)	Gestation / Delivery (weeks)	BW (g)	Neonatal Problems	Lowest platelet count (x 109/L)	Highest platelet count (x 109/L)	Upper limb abnormality	Lower limb abnormality	Cardio-vascular abnormality	Cows milk intolerance
1	Yes	Yes	De novo	A/Del	G/Del	M	31	Unkn.	Unkn.	Diarrhoea	110	111	Radii absent or hypoplastic	Yes	No	Unkn.
105 (mother of 1)	No	No		G/G	G/G											
106 (mother of 1)	No	No		G/A	G/G											
2	Yes	Yes	De novo	A/Del	G/Del	F	6 months	41	3549	Bruising	7	20	Radii absent or hypoplastic	Yes	No	Unkn.
109 (mother of 2)	No	No		G/G	G/G											
110 (father of 2)	No	No		G/A	G/G											
3	Yes	Yes	Maternal	G/Del	C/Del	F	14	Unkn.	3232	Bruising	90	140	Radii absent or hypoplastic	Yes	Unkn.	Yes
103 (mother of 3)	No	Yes		G/Del	G/Del											
104 (father of 3)	No	No		G/G	G/C											
4	Yes	Yes	Paternal	A/Del	G/Del	F	14	Unkn.	Unkn.	Unkn.	112		Radii absent or hypoplastic	Unkn.	Unkn.	Unkn.
107 (mother of 4)	No	No		G/A	G/G											
108 (father of 4)	No	Yes		G/Del	G/Del											
5	Yes	Yes	Paternal	A/Del	G/Del	F	29	40	3175	Tube fed	11	78	Radii absent or hypoplastic	no	No	No
111 (mother of 5)	No	No		G/A	G/G											
112 (father of 5)	No	Yes		G/Del	G/Del											
6	Yes	Yes	Paternal	A/Del	G/Del	F	28.5	40	2900	Unkn.	101	142	Radii absent or hypoplastic	Yes	No	Yes
81 (mother of 6)	No	No		A/A	G/G											
82 (father of 6)	No	Yes		G/Del	G/Del											
7	Yes	Yes	Maternal	A/Del	G/Del	F	22	Unkn.	Unkn.	Unkn.	12	59	Radii absent or hypoplastic	Yes	Yes	No
90 (mother of 7)	No	No		G/A	G/G											
91 (father of 7)	No	Yes		G/Del	G/Del											
8	Yes	Yes	De novo	A/Del	G/Del	F	2 days	Unkn.	Unkn.	Unkn.	Unkn.	20	Radii absent or hypoplastic	Yes	No	No
99 (mother of 8)	No	No		A/A	G/G											
100 (father of 8)	No	No		G/A	G/G											
10	Yes	Yes	Paternal	A/Del	G/Del	M	28	40	2900	Bleeding	8	120	Absence of radius with hypoplasia of humerus and ulna	No	No	Yes
11 (father of 10)	No	Yes		G/G	G/G											
12 (mother of 10)	No	No		G/A	G/G											
13	Yes	Yes	Maternal	A/Del	G/Del	F	5	40	2670	Phototherapy	37	43	Radii absent or hypoplastic	Yes	No	No
14 (father of 13)	No	No		A/A	G/G											
15 (mother of 13)	No	Yes		G/G	G/G											
16	Yes	Yes	Paternal	A/Del	G/Del	F	15	40	3150	Bleeding	29	64	Absence of radius, ulna and humerus. Hypoplasia of scapula	Yes	No	Yes
17 (father of 16)	No	Yes		G/G	G/G											
18 (mother of 16)	No	No		G/A	G/C											
19	Yes	Yes	Unkn.	A/Del	G/Del	Unkn.	23	40	Unkn.	None	26	133	Radii absent or hypoplastic	Yes	No	No
20	Yes	Yes	Unkn.	A/Del	G/Del	Unkn.	23	40	2750	Petechiae	10	110	Radii absent or hypoplastic	No	No	No
21	Yes	Yes	Unkn.	A/Del	G/Del	Unkn.	26	40	Unkn.	None	20	53	Radii absent or hypoplastic	Yes	No	No
22	Yes	Yes	Unkn.	A/Del	G/Del	Unkn.	27	40	Unkn.	None	10	40	Radii absent or hypoplastic	No	No	Yes
23	Yes	Yes	De novo	A/Del	G/Del	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.
24 (parent of 23)	No	No		G/G	G/G											
25 (parent of 23)	No	No		G/A	G/G											
33	Yes	Heterozygous frameshift insertion 145508476 T/TAGCG	N/A	G/A	G/G	M	29	41	3110	Petechiae	13	Unkn.	Unkn.	Unkn.	Unkn.	Yes
31(parent of 33)	No	Heterozygous frameshift insertion 145508476 T/TAGCG		G/G	G/G											
32 (parent of 33)	No	No		G/A	G/G											
40	Yes	Yes	Parent	A/Del	G/Del	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.
41	Yes	Yes	Parent	A/Del	G/Del	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.
38 (parent 1 of 40 and 41)	No	Yes		G/Del	G/Del											
39 (parent 2 of 40 and 41)	No	No		G/A	G/G											
42	Yes	Yes	Maternal	G/Del	C/Del	F	17	40	2750	Unkn.	9	58	Radii absent or hypoplastic	No	No	Yes
43 (mother of 42)	No	Yes		G/G	G/G											
44 (father of 42)	No	No		G/G	G/C											
47	Yes	Yes	Maternal	A/Del	G/Del	M	4 months	40	3450	Unkn.	30	233	Radii absent or hypoplastic	Yes	No	No
48 (mother of 47)	No	Yes		G/Del	G/Del											
49 (father of 47)	No	No		G/A	G/Del											
50	Yes	Yes	Maternal	A/Del	G/Del	M	26	Unkn.	Unkn.	Unkn.	Unkn.	163	Radii absent or hypoplastic	Yes	No	No
51 (mother of 50)	No	Yes		G/Del	G/Del											
52 (father of 50)	No	No		G/A	G/G											
53	Yes	Yes	Paternal	A/Del	G/Del	F	1	Unkn.	2610	None	10	28	Radii absent or hypoplastic	Yes	No	No
54 (mother of 53)	No	No		G/A	G/G											
55 (father of 53)	No	Yes		G/Del	G/Del											
56	Yes	Yes	De novo	A/Del	G/Del	F	34	Unkn.	Unkn.	Unkn.	Unkn.	Unkn.	Radii absent or hypoplastic	Yes	No	No
57 (mother of 56)	No	No		G/G	G/G											
58 (father of 56)	No	No		G/A	G/G											
59	Yes	Yes	De novo	A/Del	G/Del	M	26	Unkn.	2600	Unkn.	10	200	Radii absent or hypoplastic	Yes	Unkn.	Unkn.
60 (mother of 59)	No	No		G/G	G/G											
61 (father of 59)	No	No		G/A	G/G											

Unique Case Number	TAR diagnosed	Heterozygous 1q21 deletion	1q21 deletion origin (de novo, parent, unknown)	Genotype 5'UTR SNP 145507646 G/A	Genotype intronic SNP 145507765 G/C	Sex	Age(years)	Gestation/ Delivery (weeks)	BW (g)	Neonatal Problems	Lowest platelet count (x 109/L)	Highest platelet count (x 109/L)	Upper limb abnormality	Lower limb abnormality	Cardio-vascular abnormality	Cows milk intoleran ce
64	Yes	Yes	Unknown	G/Del	C/Del	M	23	Unknown	Unknown	Unknown	94	155	Radii absent or hypoplastic	Yes	Unknown	Unknown
65	Yes	Yes	Maternal	A/Del	G/Del	F	34	Unknown	Unknown	Unknown	79	142	Radii absent or hypoplastic	Unknown	Unknown	Unknown
66 (mother of 65)	No	Yes		G/Del	G/Del											
67 (father of 65)	No	No		G/A	G/G											
68	Yes	Yes	Maternal	G/Del	C/Del	M	1.5	Unknown	Unknown	Unknown	79	169	Radii absent or hypoplastic	Unknown	Unknown	Unknown
69 (mother of 68)	No	Yes		G/Del	G/Del											
70	Yes	Yes	Maternal	G/Del	C/Del	F	18	38	2600	None	34	154	Radii absent or hypoplastic	Yes	No	Unknown
71	Yes	Yes	Maternal	G/Del	C/Del	F	6 months	39	3510	Unknown	30	200	Radii absent or hypoplastic	Yes	Unknown	Unknown
72 (mother of 70 and 71)	No	Yes		G/Del	G/Del											
73 (father of 70 and 71)	No	No		G/G	G/C											
74	Yes	Yes	Unknown	G/Del	C/Del	M	39	Unknown	Unknown	Unknown	79	169	Radii absent or hypoplastic	Yes	No	No
75 (mother of 74)	No	Yes		G/Del	G/Del											
76	Yes	Yes	Maternal	G/Del	C/Del	M	2	40	3220	Bleeding	8	130	Radii absent or hypoplastic	Yes	No	Unknown
77 (mother of 76)	No	Yes		G/Del	G/Del											
78 (father of 76)	No	No		G/G	G/C											
83	Yes	Yes	Non Maternal	A/Del	G/Del	F	37	Unknown	Unknown	None	74	136	Radii absent or hypoplastic	Yes	Unknown	Unknown
84 (mother 83)	No	No		G/G	G/G											
85	Yes	Yes	De novo	A/Del	G/Del	F	17 days	37	2800	Unknown	Unknown	167	Radii absent or hypoplastic	no	Yes	No
87 (father of 85)	No	No		G/A	G/G											
90 (mother of 85)	No	No		G/A	G/G											
88	Yes	Yes	Unknown	A/Del	G/Del	F	4	Unknown	2720	Unknown	Unknown	34	Radii absent or hypoplastic	No	Unknown	Unknown
89	Yes	Yes	Unknown	G/Del	C/Del	M	8	Unknown	Unknown	Unknown	Unknown	88	Radii absent or hypoplastic	Yes	Unknown	Unknown
92	Yes	Yes	De novo	A/Del	G/Del	M	1.5	Unknown	Unknown	Unknown	7	65	Radii absent or hypoplastic	Unknown	Unknown	Unknown
93 (mother of 92)	No	No		G/G	G/G											
94 (father of 92)	No	No		G/A	G/G											
95	Yes	Yes	Maternal	A/Del	G/Del	M	6	39	2900	Unknown	18	180	Radii absent or hypoplastic	Yes	Yes	Unknown
96 (mother of 95)	No	Yes		G/Del	G/Del											
97 (father of 95)	No	No		G/A	G/G											
98	Yes	Yes	Unknown	A/Del	G/Del	F	4	40	2910	None	7	295	Radii absent or hypoplastic	No	No	Yes
101	Yes	Yes	Unknown	A/Del	G/Del	F	17	Unknown	Unknown	Unknown	31	91	Radii absent or hypoplastic	Yes	Unknown	Unknown
102 (mother of 101)	No	Yes		G/Del	G/Del											
113	Yes	Yes	Unknown	A/Del	G/Del	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
114	Yes	Yes	De novo	G/Del	C/Del	Unknown	8	42	2960	Bleeding	Unknown	43	Radii absent or hypoplastic	Yes	No	No
115 (parent of 114)	No	No		G/G	G/G											
131 (parent of 114)	No	No		G/G	G/C											
116	Yes	Yes	Maternal	A/Del	G/Del	F	8	40	3062	Bleeding	12	91	Radii absent or hypoplastic	Yes	No	No
117 (father of 116)	No	No		G/A	G/G											
118 (mother of 116)	No	Yes		G/Del	G/Del											
121	Yes	Yes	Maternal	G/Del	G/Del	F	36	Unknown	Unknown	Unknown	Unknown	Unknown	Radii absent or hypoplastic	no	No	Yes
145 (mother of 121)	Yes	Yes	Unknown	G/Del	G/Del	M	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
120 (father of 121)	No	No		G/G	G/G											
122	Yes	Yes	Unknown	A/Del	G/Del	M	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
123	Yes	Yes	Unknown	G/Del	C/Del	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
124	Yes	Yes	Unknown	A/Del	G/Del	F	22	Unknown	Unknown	Bruising	Unknown	Unknown	Radii absent or hypoplastic	Yes	No	Yes
125	Yes	Yes	Paternal	A/Del	G/Del	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
126 (father of 125)	No	Yes		G/Del	G/Del											
127 (mother of 125	No	No		A/G	G/G											
128	Yes	Yes	Unknown	G/Del	C/Del	F	8	41	3544	Bleeding	11	178	Radii absent or hypoplastic	No	No	Yes
129 (parent of 128)	No	No		G/G	G/C											
134	Yes	Yes	Maternal	A/Del	G/Del	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
132 (mother of 134	No	Yes		G/Del	G/Del											
133 (father of 132)	No	No		G/A	G/G											
136	Yes	Yes	Not Paternal	A/Del	G/Del	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
135 (father of 136)	No	No		G/G	G/G											
139	Yes	Yes	Unknown	A/Del	G/Del	F	17	42	3459	Bleeding	20	55	Radii absent or hypoplastic	No	No	Unknown
138 (parent of 138)	No	No		G/A	G/G											
140	Yes	Heterozygous for gain of stop codon 145509173 C/T	N/A	G/A	G/C	F	9	Unknown	Unknown	Bleeding	Unknown	Unknown	Unknown	Unknown	Yes	Yes
142	Yes	Yes	Paternal	A/Del	G/Del	M	16	40	Unknown	Bleeding	Unknown	78	Radii absent or hypoplastic	Yes	No	No
143 (mother of 142)	No	No		G/A	G/G											
144 (father of 142)	No	Yes		G/Del	G/Del											

Table 1 ***Genotype and phenotype for the TAR cases and healthy parents.*** **Healthy individuals are on gray background. Unknown (Unkn.)**

Forward primer sequence	Reverse primer sequence	Chr1 region	Description
CACGCCAGCCTCTGAGTT	CCCTAATCTCAAACCACTTCCT	145501836-145502420	upstream regulatory element
GGACGAGCAGGAGACAGATG	CGCACCTGGCCTAAAAATTC	145502151-145502744	upstream regulatory element
CAAAGCACACCCTGCACA	CACCTCCTGGGTTCAGGTAA	145502469-145502949	upstream regulatory element
GGCCAGCCTGGGTAGTATAA	CCAGTCTGGGGGACAAGAG	145506316-145506883	promoter
TGCACCACTGCACTCTTAGC	TTTAGGCAGCGTGGTGTATG	145506650-145507248	promoter
GTCTCCCGGGTTCAACTG	TCTAAATCCCTCCCTCTGCAC	145506920-145507559	*RBM8A*
GCCCAGCTAATCAGCTTCC	TCCCTCTGCACGGTAAAAAC	145506991-145507549	promoter
TTTCCCAGTTTGGGATGAAG	GGGCGGAATCTCTAATCCAC	145507301-145507871	*RBM8A*
GCCGGGCCTCACTGTTAAT	TCAGTTTGTGAATGCTCTCTGG	145507354-145508033	*RBM8A*
GCCGCGGTTAAGAGGAAG	TTGTGAATGCTCTCTGGAACC	145507395-145508028	*RBM8A*
ATGGCCACAGAAACACTTCC	CACCGCCTCCAGTCTTAGTG	145507474-145507924	*RBM8A*
AGTTAGCCTTTGATTGGTCAGC	ACCCGTAGCTCCTGCCCTA	145507474-145508174	*RBM8A*
ATGGCCACAGAAACACTTCC	TCCTCCTTTCTCCCATTGTTC	145507474-145508174	*RBM8A*
ATGGCCACAGAAACACTTCC	CCACAGACACGGATACCTCA	145507474-145508324	*RBM8A*
CGGGTCTTGGGTGGATTA	CCACAGACACGGATACCTCA	145507842-145508324	*RBM8A*
GGGTCTTGGGTGGATTAGAGA	TTTAAGCAGGCTCACAGGAA	145507843-145508430	*RBM8A*
GGGTTCCAGAGAGCATTCAC	GATATCCTGTTCGCCTGTCG	145508007-145508606	*RBM8A*
CCTAGTAGGGCAGGAGCTACG	CCAACCACAGCAAACACAGA	145508105-145508692	*RBM8A*
CGCAGTAGGAATGGGTTCAG	CCTGGGCTTCCTTGTATGTT	145508355-145508958	*RBM8A*
GGCCAAGAGCAAAGTTGAAA	CCCAGTCCTATTTGTCCAAGG	145508691-145509284	*RBM8A*
TTGTCAGACACGCCAAAGAG	CAATGATCCATACAGCCTTGC	145508736-145509438	*RBM8A*
TGGGTGAAGGGAATACGAAC	ATGGTGGCATGTGCCTGTA	145509079-145509626	*RBM8A*
GTGTTACCCAGGGTGGATTG	CATGCCTTTAGACAGCTGGA	145509508-145509946	*RBM8A*
GGGAGGGACTTCAGTTAGCA	CCTGTTGCCTCTAGCATCATT	145510103-145510687	*RBM8A*
TGATAGAAATATGAAGCCACCAAG	AAGGATGAATTGGGAGGAGAC	145510458-145511028	*RBM8A*
AAGAGGCAGCAGAAGGTGAA	CAGCCCAATAGCATTTGGAA	145510814-145511453	*RBM8A*
GGCTTGAATATGATGCTGAACA	GCCTGATCGTAACTCCAAACA	145511127-145511721	*RBM8A*
CCCCTCTGCGACAGTTTC	GTCCCCATCCTCATCCATG		Allele-specific expression experiment
GTAAAACGACGGCCAG	CAGGAAACAGCTATGAC		Allele-specific expression experiment
CCCCTCTGCGACAGTTTCC	CGCCATCTCGCCTTCGA		Allele-specific expression experiment (genotyping)

Table 2 ***Primer pair sequences used for Sanger sequencing of RBM8A locus and upstream regulatory element, and primer sequences for the allele-specific expression experiment and genotyping.***

N=7504 individuals	Genotypes passed QC (call rate)	Homozygous major*	Heterozygous*	Homozygous minor*	Estimated minor allele frequency	Deviation from HWE (exact test)	Association with platelet count
5'UTR G/A	7317 (97.5%)	6879 (6402)	431 (396)	7 (7)	3.05%	P=0.84	P=0.87
Intronic G/C	7458 (99.4%)	7396 (6879)	62 (59)	0 (0)	0.42%	P=1.0	P=0.99

Table 3 ***Genotyping of the 5'UTR and intronic SNPs of the RBM8A gene in 7504 healthy individuals from the Cambridge BioResource and association with platelet count.*** ***Number of individuals with measured platelet count is indicated between parentheses.**

Supplementary Figures

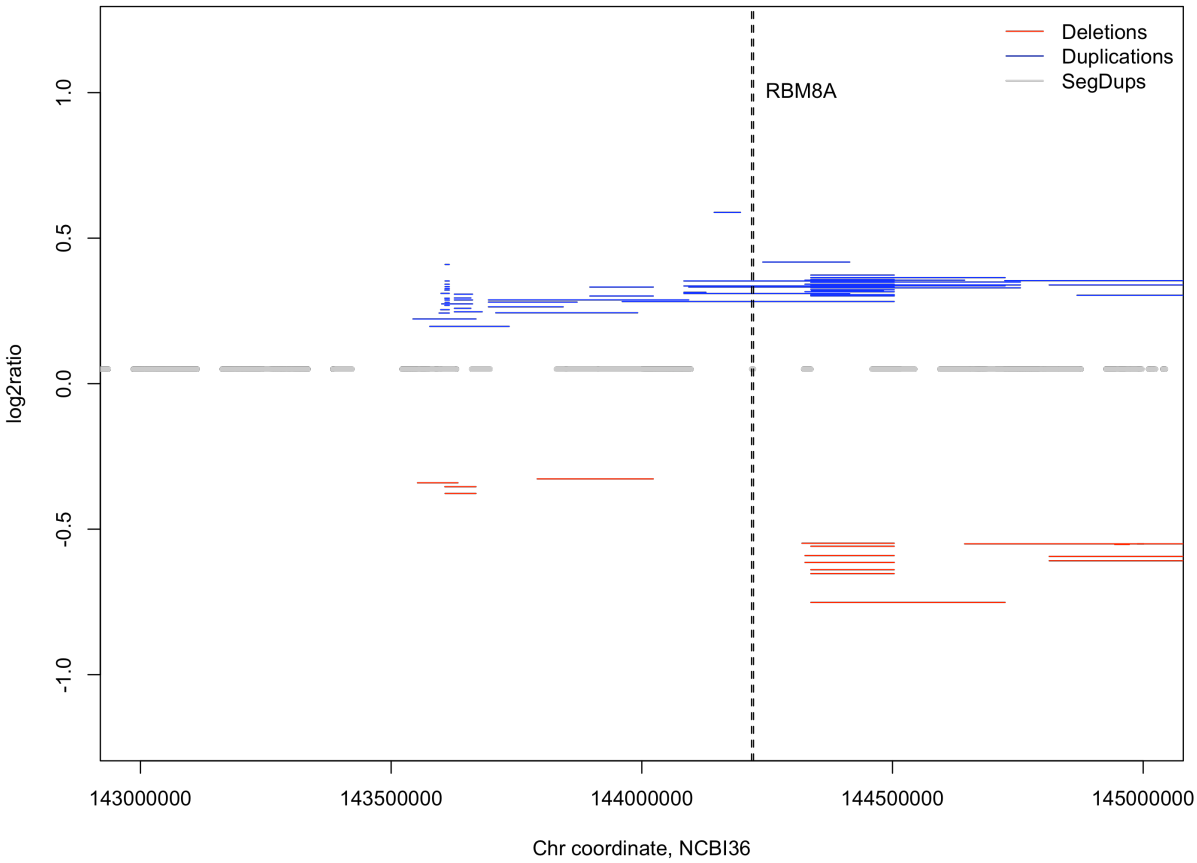

Figure 1 ***Copy number variants in the chromosome 1q21 locus.*** **The figure shows log2-ratios of Affy6 SNP array probe intensities for 5919 healthy Welcome Trust Case Control Consortium shared controls. See ref. [4] for details on the calling of the copy number variants (CNV). There are no deletions of the *RBM8A* gene in these individuals, indicating a low frequency of the 1q21 deletions found in TAR cases and their healthy relatives. Five duplications were observed, which suggests that over-expression of *RBM8A* is not deleterious.**

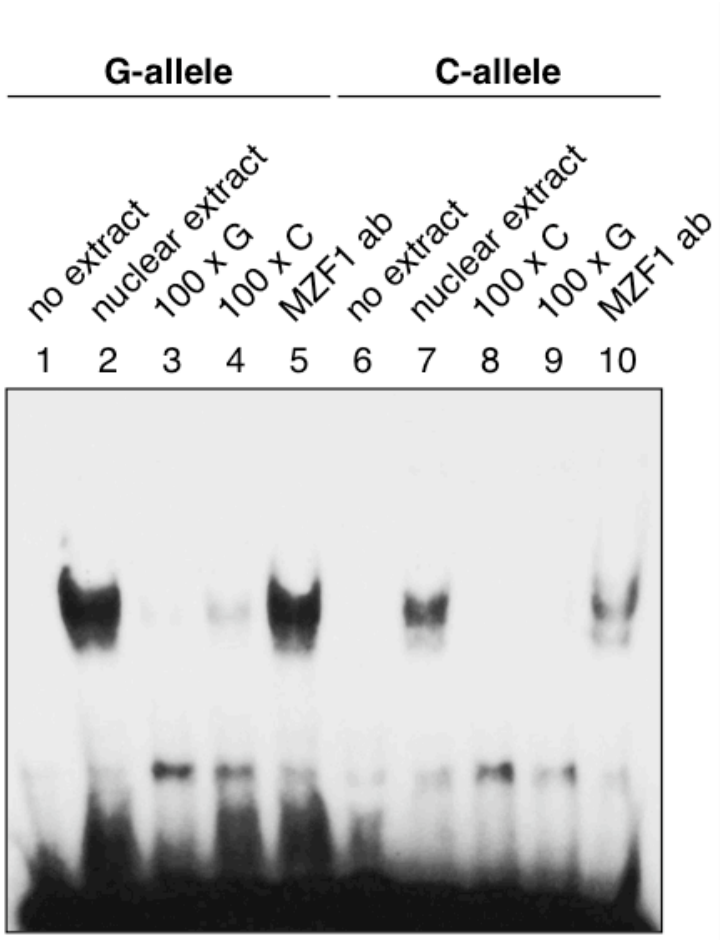

Figure 2 ***Differential binding of nuclear proteins at the intronic SNP.*** **Electrophoretic mobility shift assays (EMSA) in nuclear protein extracts from CHRF-288-11 cells showed higher protein affinity of the probe containing the G-allele (major allele of the RBM8A intronic SNP, lane 2) than the C-allele (minor allele, lane 7). Protein binding of G-allele-probes was competed by specific (lane 3), but not by unspecific unlabeled probes (lane 4). We performed supershift experiments with antibodies for the predicted transcription factors MZF1 and RBPJ. However, in our experiments none of the tested antibodies competed for binding and/or shifted the protein-DNA complex (lane 10, data not shown for RBPJ).**

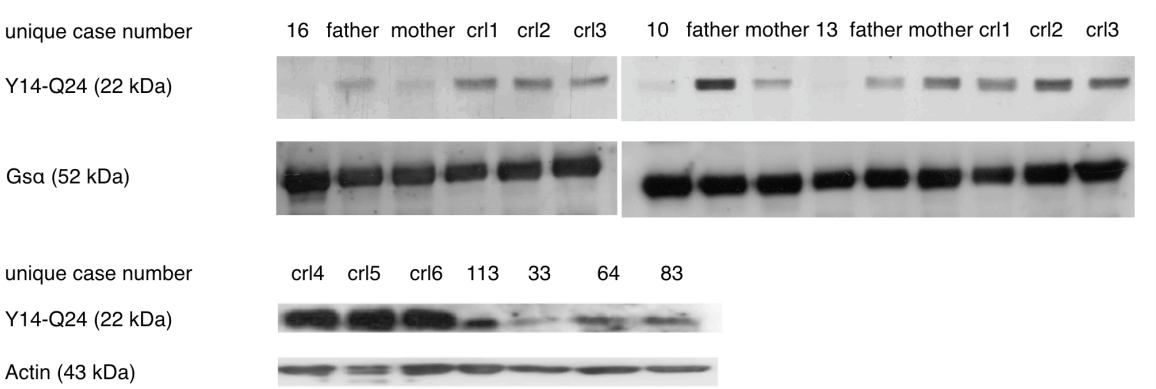

Figure 3 ***Y14 protein expression in platelets.*** **Western blots are shown for TAR cases (indicated by numbers), their parents (indicated by father and mother to the right of the unique case number) and controls (Crl). Protein expression of Gsα and Actin were used as a loading control (bottom rows).**

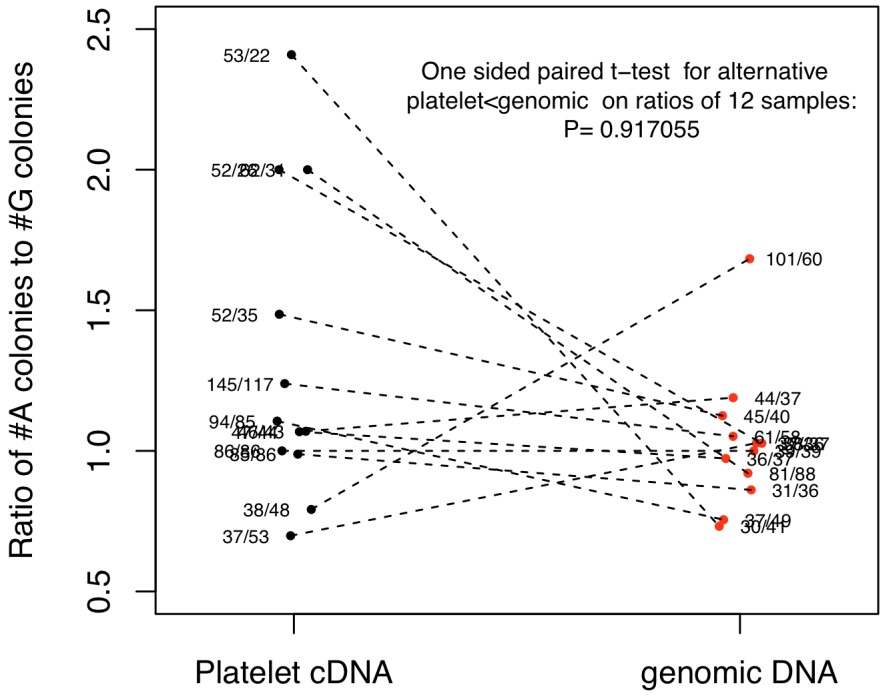

Figure 4 ***Allele-specific expression in platelet RNA for the 5'UTR SNP.*** **Expression of the major allele (G) and the minor allele (A) was determined for 12 healthy blood donors heterozygous for the 5'UTR SNP in platelet cDNA (black points) and in genomic DNA (red points) using a colony-PCR approach. Each data point represents one individual. A value of 1 on the vertical axis represents an expression ratio of one. The numbers next to the data points represent respectively the number of colonies with the minor allele and the number of colonies with the major allele. Jitter has been added to the positions on the horizontal axis. The paired t-test was performed on the ratios of the colony counts.**

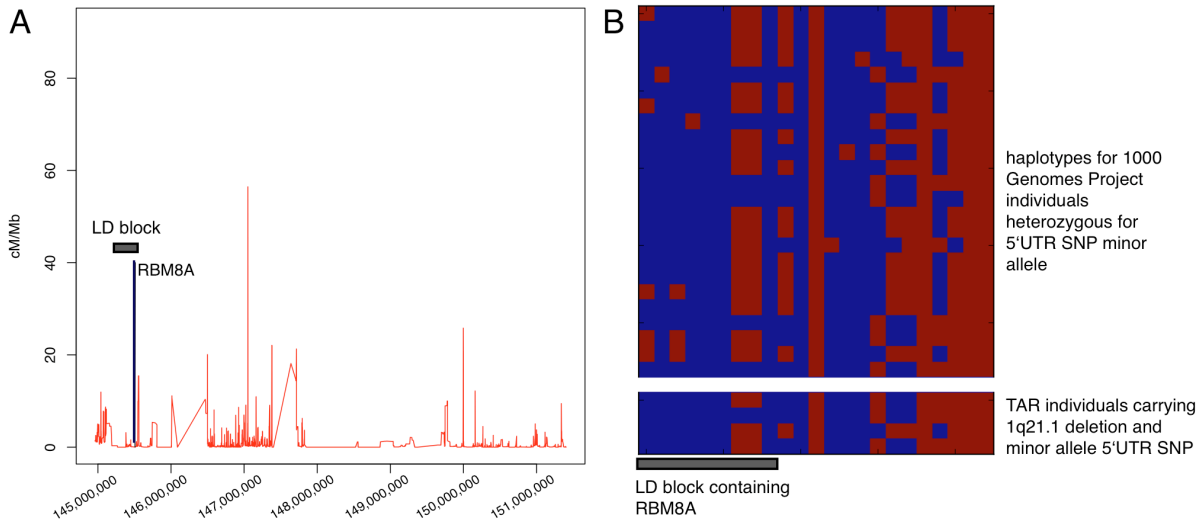

Figure 5 ***Haplotype analysis of RBM8A locus.*** **A) Recombination rates in the 1q21.1 locus in CEU individuals. The gray bar indicates the LD block containing *RBM8A*. B) 24 individuals of European descent from Phase 1 of the 1000 Genomes Project are heterozygous for the 5'UTR SNP (top) showing distinct haplotypes segregating in the LD block containing *RBM8A*. The haplotype on which the 5'UTR SNP minor allele present is shown. At least two of these haplotypes were present in the 4 exome sequenced TAR cases carrying the 5'UTR SNP minor allele (bottom).**

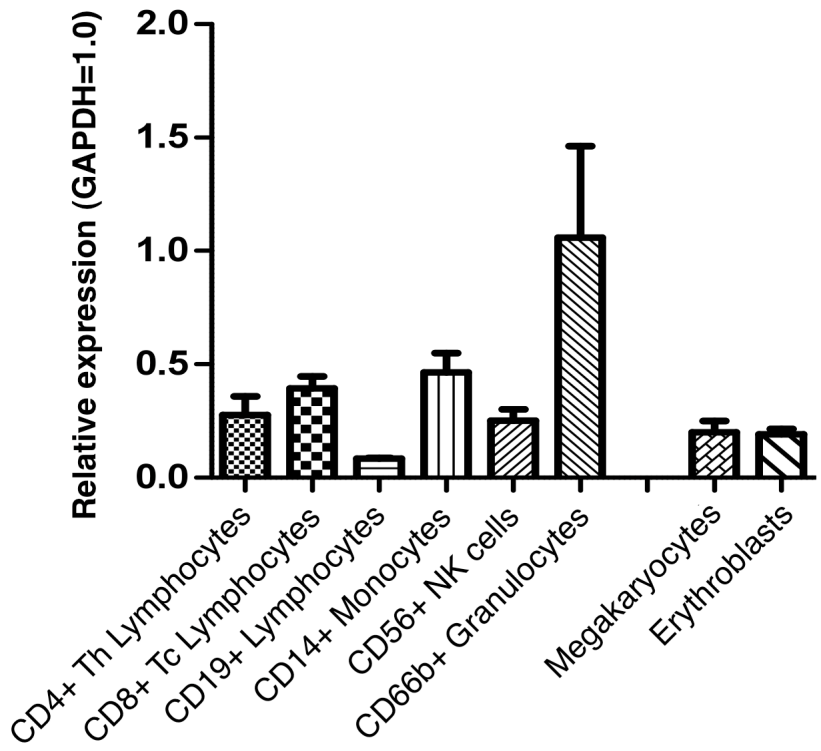

Figure 6 RBM8A *gene expression in 8 hemopoietic lineages.* Gene expression was determined using a TaqMan qPCR probe for *RBM8A* (catalog number *Hs04234933_g1, Invitrogen, US)* and normalized with respect to the *GAPDH* gene. *RBM8A* is expressed in all lineages.

Figure 7 ***Protein alignments of*** **Rbm8a** ***in 8 species.*** **Differences with the human protein sequence are highlighted in grey.**

Figure 8 ***The protein encoded by*** **rbm8a** ***is required for early zebrafish development.*** **To evaluate the biological importance of *rbm8a* in hemopoiesis in zebrafish embryos, we adopted the morpholino antisense (MO) knockdown approach in order to block the expression of the endogenous gene. First we performed dose response experiment and injected 0.5-, 1-, 3- and 6-ng of *rbm8a* MO into wild-type zebrafish embryos at the 1-2 cell stage. Embryos injected with 6-ng of MO died within 24 hours post-fertilization (hpf); 0.5-, 1- and 3-ng MO injected embryos had gross developmental defects at 48 hpf. Even at lower doses of the MO, 0.5-, 1- and 3-ng, morphological abnormalities and loss of viability were still evident, although some embryos did survive beyond 24 hpf. These effects from a single knock-down MO are atypical and dramatic, and certainly more extensive than those seen from similar experiments for other genes implicated in hemopoiesis[5-8]. These data are compatible with the notion that the protein encoded by *rbm8a* is required for early zebrafish development. Scale bars represent ~0.5 mm.**

References

1. Klopocki, E. et al. Complex Inheritance Pattern Resembling Autosomal Recessive Inheritance Involving a Microdeletion in Thrombocytopenia-Absent Radius Syndrome. *The American Journal of Human Genetics* **80**, 232-240 (2007).
2. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).
3. 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).
4. Huang, N., Lee, I., Marcotte, E.M. & Hurles, M.E. Characterising and Predicting Haploinsufficiency in the Human Genome. *PLoS Genet* **6**, e1001154 (2010).
5. Cvejic, A., Serbanovic-Canic, J., Stemple, D.L. & Ouwehand, W.H. The role of meis1 in primitive and definitive hematopoiesis during zebrafish development. *Haematologica* **96**, 190-198 (2011).
6. O'Connor, M.N. et al. Functional genomics in zebrafish permits rapid characterization of novel platelet membrane proteins. *Blood* **113**, 4754-4762 (2009).
7. Serbanovic-Canic, J. et al. Silencing of RhoA nucleotide exchange factor, ARHGEF3 reveals its unexpected role in iron uptake. *Blood* **118**, 4967-4976 (2011).
8. Tijssen, M.R. et al. Genome-wide Analysis of Simultaneous GATA1/2, RUNX1, FLI1, and SCL Binding in Megakaryocytes Identifies Hematopoietic Regulators. *Developmental Cell* **20**, 597-609 (2011).