Nat Genet. Author manuscript; available in PMC 2012 August 28.

Published in final edited form as:
Nat Genet. ; 43(8): 735–737. doi:10.1038/ng.885.

Exome sequencing identifies *NBEAL2* as the causative gene for Gray Platelet Syndrome

Cornelis A Albers[1,2,11], **Ana Cvejic**[1,2,11], **Rémi Favier**[3,4], **Evelien E Bouwmans**[2], **Marie-Christine Alessi**[5], **Paul Bertone**[6], **Gregory Jordan**[6], **Ross NW Kettleborough**[1], **Graham Kiddle**[2], **Myrto Kostadima**[6], **Randy J Read**[7], **Botond Sipos**[6], **Suthesh Sivapalaratnam**[8], **Peter A Smethurst**[2], **Jonathan Stephens**[2], **Katrin Voss**[2], **Alan Nurden**[9], **Augusto Rendon**[2,10], **Paquita Nurden**[9,11], and **Willem H Ouwehand**[1,2,11]

[1]Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1HH/1SA, UK. [2]Department of Haematology, University of Cambridge & NHS Blood and Transplant, Cambridge, CB2 0PT, UK. [3]Department d'Hématologie, Assistance-Publique Hôpitaux de Paris, Centre de Référence des Pathologies Plaquettaires, Hôpital Armand Trousseau, 75012, Paris, France. [4]INSERM U1009, 94805, Villejuif, France. [5]INSERM U626, Faculté de Marseille and 13385 Marseille, Cedex 05, Marseille, France. [6]EMBL - European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. [7]Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, CB2 0XY, UK. [8]Department of Vascular Medicine, Academic Medical Center, Amsterdam, 1105 AZ, The Netherlands. [9]Laboratoire d'Hématologie, Centre de Référence des Pathologies Plaquettaires, Hopital Xavier Arnozan, 33604, Pessac, France. [10]MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge, CB2 0SR, UK.

Abstract

Gray platelet syndrome (GPS) is a predominantly recessive platelet disorder characterized by a mild thrombocytopenia with large platelets and a paucity of α-granules; these abnormalities cause mostly moderate but in rare cases severe bleeding. We sequenced the exomes of four unrelated cases and identified as the causative gene *NBEAL2*, a gene with previously unknown function but a member of a gene family involved in granule development. Silencing of *nbeal2* in zebrafish abrogated thrombocyte formation.

Platelets are the second most abundant cell in the blood and maintain arterial vessel wall integrity. Formation of the platelet plug at the site of vessel wall damage is a multi-step process of tethering and attachment followed by activation that causes the release of both the α- and δ-granules. Platelets are anucleate and formed in large numbers in the bone marrow

Correspondence should be addressed to C.A.A. (caa@sanger.ac.uk) or W.H.O. (who1000@cam.ac.uk).
[11]These authors contributed equally to this work.

Accession codes The exome sequencing data can be found in the European Genotype Archive under accession numbers EGAN00001007686, EGAN00001001896, EGAN00001001894, and EGAN00001001895 for cases A-D respectively. The genomic reference sequence for *NBEAL2* can be found under the GenBank accession number NM_015175. The microarray expression data can be found under the ArrayExpress accession numbers E-TABM-633, E-MTAB-001, and E-MTAB-002.

Author contributions P.N. and W.H.O. conceived and designed the study. C.A.A. analyzed the sequencing data and performed statistical analysis. A.C. performed the zebrafish experiments. E.E.B., R.K. and J.S. performed the capillary sequencing. M.K. and P.B. analyzed the RNA sequencing and expression data. R.J.R. performed the protein structural modeling. P.S. generated the RNA sequencing and expression data. A.R. performed the microarray gene-expression analysis. G.J. and B.S. performed the evolutionary analyses. K.V. performed the megakaryocyte culture experiments. G.K. supervised the sample collection and exome sequencing. S.S. generated the whole genome expression data for platelets. R.F., M.A., A.N. and P.N. recruited and clinically characterized the cases. P.N. performed the electron-microscopy experiments. C.A.A., A.C., P.N. and W.H.O. wrote the manuscript.

Europe PMC Funders Author Manuscripts

by polyploid precursor cells, named megakaryocytes (MKs). The lack of adequate numbers or impaired function of platelets can result in bleeding. GPS was first defined in 1971 when it was observed that in some cases with an inherited bleeding disorder a May-Grünwald-Giemsa stained blood smear showed gray platelets with increased diameter[1]. Electron microscopic studies revealed a complete lack of α-granules in most cases[2] (Fig. 1a). Since the α-granule protein cargo is essential to the development of a robust platelet plug, GPS cases are symptomatic at platelet count levels that are typically not associated with bleeding. Proteins normally only released upon platelet activation are spontaneously released from MKs; most likely because of the lack of α-granules[3].

A recent study established significant linkage of a locus on chromosome 3p21 to GPS, but the causative gene has remained elusive so far[4,5]. We therefore sequenced the exomes of four unrelated cases with GPS (Supplementary Table 1-2, Supplementary Fig. 1, Supplementary Note), using the Agilent SureSelect protocol to enrich for 39.3 Mb of coding sequence and the Illumina GAII platform[6]. GPS is an extremely rare disorder, illustrated by the fact that there are about 30 documented cases in France; cases typically occur in families consistent with a recessive disorder[4]. We hypothesized that the causative variants would be novel, and filtered out variants seen previously and those not likely to affect protein function (Online Methods). Assuming a recessive mode of inheritance we required at least two novel mutations per individual in the same gene.

We found that only for the Neurobeachin-like 2 gene (*NBEAL2*), encoding a member of the family of BEACH domain containing proteins, on chromosome 3 (3p21) all cases had at least two novel mutations with a predicted functional consequence for the protein, with two homozygous cases and two compound heterozygotes (Table 1, Fig. 1b). All mutations and genotypes were confirmed by capillary sequencing (not shown). Based on theoretical calculations we estimate that the probability of the observed mutation pattern occurring by chance is $\sim 5 \times 10^{-11}$ (Supplementary Note). Furthermore, we established correct segregation in a further five relatives, two unaffected and three affected ones, in pedigrees A, B and C and identified the affected father (B.I.2) of case B.II.3 as homozygous for the causative mutation, compatible with the known consanguinity in this pedigree (Supplementary Note). The mutations were at eight different locations in the 2,754 amino acid long Nbeal2 protein and were absent in 994 low-pass samples from the main phase of the 1000 Genomes Project[7]. Except for mutation G2553E in case D.II.3, all affected residues for the nsSNPs are highly conserved, as well as the splice site disrupted by the 7-base pair insertion in the same case (Table 1, Supplementary Fig. 2-9). The eight mutations affect multiple exons of *NBEAL2* and multiple protein domains, with the P2100L and S2269L variants being located in the BEACH domain itself. Modeling based on the fold of the homologous PH-BEACH structure from NBEA[8] (>50% identity at the amino acid level) shows that the environment of the former mutation would introduce clashes in a tightly-packed location of a hydrophobic pocket and the latter would potentially lead to local changes in conformation (Supplementary Fig. 10).

To verify the biological importance we silenced the *nbeal2* orthologous gene by injecting specific antisense morpholino oligonucleotides (MO) into one cell stage zebrafish embryos (Supplementary Fig. 11). This resulted in a lineage-specific effect with a complete abrogation of thrombocyte formation, but normal erythropoiesis (Fig. 1c, d). Spontaneous bleeding in the tail was observed in 41% of the embryos (Fig. 1e and Supplementary Fig. 11). The phenotype in MO-injected zebrafish is more severe than the observed phenotype of GPS cases, which may be expected due to the difference between a null-phenotype in zebrafish and a loss of function one in the GPS cases. Although the experiment did not address the function of α-granules in thrombocytes, these results support the essential role of the Nbeal2 protein in thrombopoiesis and the etiology of GPS.

Other members of the family of BEACH domain proteins include Nbea, Nbeal1, Lrba and Lyst (Supplementary Fig. 12) and are believed to be important in membrane protein trafficking[9]. Like other BEACH members except for *NBEAL1*, *NBEAL2* is expressed in MKs and indeed is upregulated during megakaryopoiesis (Supplementary Fig. 13-14). Mutations in other BEACH members affect vesicle function in both man and mice. Haplo-insufficiency of the *NBEA* gene caused by a chromosomal rearrangement in an individual with extreme autism was associated with atypical platelet δ-granules, suggestive of a possible role of Nbea in their formation[10], while the platelet α-granules were normal. In contrast with *Nbea*, which in mouse is required for the formation and functioning of central neuronal synapses[11], *NBEAL2* is not expressed in fetal and whole brain (Supplementary Fig. 15), consistent with the lack of neurological phenotypes in cases with GPS. Mutations in the *LYST* gene that result in the truncation of the BEACH/WD40 domains or the total absence of the Lyst protein lead to a dysregulation of lysosomal trafficking and are causative of Chediak-Higashi syndrome[12], for which symptoms include a platelet-type bleeding due to inadequate formation of δ-granules[13]. However, there was no previously known function for *NBEAL2* and it is has not been associated with common or rare disease, and is absent from any of the 68 quantitative trait loci that were identified in a recent GWAS meta-analysis for the number and volume of platelets in 68,000 healthy individuals (N. Soranzo, unpublished data). Taken together, these findings are consistent with a role for Nbeal2 in the development or secretion of platelet α-granules, whereby individuals with recessive mutations have impaired platelet activation and plug formation, resulting in a propensity for bleeding.

Our result is supported by the findings of two previous linkage studies. Gunay *et al.* found that 19 affected individuals in 11 out of 14 pedigrees (one being case A.I.3 from this study) analyzed in total were homozygous in the region chr3:42688625-52061914 on 3p21, which contains *NBEAL2*, together with a significant linkage of log-odds (LOD) ratio of 7.2 in the largest pedigree[4]. They sequenced 69% of all exons in the homozygous interval, but no gene was identified. In a separate linkage study, Fabbro *et al.* reported a LOD score of 2.7 for the same locus and subsequently identified four affected individuals from two pedigrees that were completely homozygous and one individual that was mostly homozygous for the region chr3:48487338-50187790, which does not include *NBEAL2*[5]. Targeted next-generation resequencing of this region did not identify a causative gene. Interestingly, the segregation pattern inferred from microsatellite markers in the two pedigrees is consistent with compound heterozygosity for an interval that does contain *NBEAL2*, and based on mixed ethnicity in one pedigree the authors raised this as a possibility[5].

In summary, we identified multiple novel mutations in *NBEAL2*, a gene with no previously known function, in Gray Platelet Syndrome, a known recessive disorder. It is consistent with an important role for Nbeal2 in the formation of α-granules and this discovery may lead to the development of a novel generation of safer anti-platelet drugs for use in the treatment of individuals with heart attacks and stroke. A comparison with two previous linkage studies demonstrates the power of directly observing sequence variation for resolving complex segregation patterns, in particular combinations of compound heterozygosity and homozygosity.

Methods

Sequencing

We applied the Agilent SureSelect protocol (Agilent, South Queensferry, UK, catalogue no. G3362A) to enrich for 39.3 Mb of sequence covering 740K exons in 79K transcripts from a highly redundant set of 34,642 genes. The enriched DNA was sequenced on the Illumina GAII platform (Illumina, Little Chesterford, UK). We generated 11.7-13.4 Gb of sequence

per individual, resulting in a mean coverage of 103-112-fold and 88.3-92.3% of the sites within 25 bp of the targets covered at least ten-fold.

Sequence analysis

The sequence analysis was divided into three phases with the first one consisting of aligning and recalibrating the reads to produce a BAM file[15]. The second phase consisted of calling the sequence variants using various algorithms. The third phase consisted of filtering the set of variants.

First phase:

1. The FastQ files generated by the sequencing pipeline were aligned to reference NCBI build 37 using the Stampy read mapper[16].
2. For each lane independently, the reads were realigned around known insertions and deletions (indels) (1000 Genomes pilot project,[7]), followed by base quality recalibration using the Genome Analysis Toolkit (GATK)[17].
3. After merging the different lanes, duplicates were flagged using the Picard software package[18] and excluded from subsequent analyses.

Second phase:

1. Indels were called using Dindel[19] on each sample independently.
2. The reads were realigned in the BAM files produced in phase 1 around the indels called by Dindel using the GATK, in order to reduce alignment artefacts around all indels found in a sample.
3. The GATK software package and SAMtools were used to call Single Nucleotide Polymorphisms (SNPs)[15]. We called SNPs both for each sample independently and jointly using all samples. Furthermore, SNPs were called both from the BAM files produced in phase 1 and in step 2 of the second phase, so that SNPs were called using six different approaches. The joint calling was performed using exome sequencing results of 32 additional DNA samples from individuals with unrelated clinical phenotypes on the realigned BAM files produced in step 2 of the second phase.
4. We required a minimum variant quality of 20, which corresponds to a 99% confidence level. Furthermore, we removed variants with more than 10 reads with mapping quality 0 as a proxy for regions in the genome where mappings are likely unreliable.

Third phase:

1. Variants that did not overlap the targeted regions ±25 bp were filtered and not considered in further analyses.
2. We predicted the functional consequence of SNPs and small indels using the Ensembl Variation API[20].
3. We considered only non-synonymous (ns) SNPs, SNPs introducing a premature stop-codon, indels in protein coding sequence, and variants affecting 3′UTRs and essential splice sites.
4. Each variation was annotated for presence in databases of genetic variation. Any variant found to be present was not considered as a candidate for GPS. The following public databases were used: dbSNP131, SNPs from 8 HapMap exomes[21], the 1000 Genomes pilot project SNPs and indels[7]. In addition to these

Europe PMC Funders Author Manuscripts

public resources we made also use of the variation data observed in 354 exomes from the CoLaus cohort, which have been sequenced at the Wellcome Trust Sanger Institute as part of a partnership between the Sanger Institute, the CoLaus principal investigators (P Vollenweider, Department of Internal Medicine, University Hospital Center, University of Lausanne, Lausanne, Switzerland and GlaxoSmithKline (GSK) (V Mooser, GSK, Philadelphia, USA)[14]. We also used as a filter the variants called on 32 exomes sequenced on the same platform for unrelated studies, which allowed us to correct for systematic errors in the variant calling.

Statistical significance of observed mutations

We estimated the significance of the novel mutations identified in the four exome-sequenced GPS cases using a dynamic programming approach based on the coalescent. The coalescent is a mathematical description of the notion that ultimately all individuals are related through a single common ancestor[22].

Since the observed mutation pattern is extremely rare and generally would only expected to be seen by conditioning on a rare event, for instance the presence of a rare phenotype, permutation analyses would not provide an accurate estimate of the significance.

We therefore assume a population of N_a chromosomes of affected individuals, and N_u chromosomes of unaffected individuals. The infinite site model provides an estimate of the number of variable sites in genomic segment of length L through the population-scaled mutation rate θ. The relationship $\theta = 4N_e\mu$, relates the population scaled mutation rate to N_e, the effective population size, and μ, the per-generation mutation rate. Here, we assumed an effective population size of 23000, which is based on the mutation rate $\mu=1.1 \times 10^{-8}$ estimated from trios[7] and a heterozygosity of 0.001 for SNPs. We use the standard expression for the probability density of the number of variable sites in the infinite sites model and the allele frequency prior given that a site is variable (see e.g. Wakeley *et al.* [22]).

The purpose of the dynamic programming is to calculate the probability that given k previous sites, the next variable site $k+1$ is a site with frequency count at most f_{max}, and that the derived allele is present only on chromosomes of affected individuals; conditional on this event, we calculate the probability that such a mutation at site $k+1$ is present on the chromosomes of affected individuals that had no previous mutations (up to site k) satisfying the same criteria. Weighted by the probability of observing k variable sites, this then provides an estimate of the probability that all chromosomes of affected individuals carry at least one mutation of frequency at most f_{max} and not present in controls.

Zebrafish experiments

Zebrafish husbandry—General maintenance, collection, and staging of the wild type and transgenic *Tg(cd41:EGFP)* zebrafish were carried out according to the Zebrafish Book[23]. Embryos were maintained in egg water (60 mg/L Red Sea salts) at 28°C until the appropriate stage.

O-Dianisidine staining—O-Dianisidine staining for haemoglobin was performed as previously described[24]. In brief, unfixed embryos were dechorionated and stained for 10 minutes in the dark, with a solution consisting of o-Dianisidine (0.7 mg/ml), 0.01 M sodium acetate (pH 4.5), 0.65% hydrogen peroxide and 45% (vol/vol) ethanol. Photomicrographs were taken with a Zeiss camera AxioCam HRC attached to a LeicaMZ16 FA dissecting microscope.

Nbeal2 antisense oligonucleotides—Morpholinos (MO) targeting zebrafish *nbeal2* atg (5′-TCGAAGCCATTCTCCCTCGCCCTTC-3′) and standard control MO (5′-CCTCTTACCTCAGTTACAATTTATA-3′) were obtained from GeneTools LLC (Philomath, Oregon). The MO were resuspended in sterile water and approximately 0.8 nl was injected in zebrafish embryos, at the one cell stage. For both *nbeal2* atg MO and the standard control MO a concentration of 1 μg/μl was used.

Expression of BEACH domain containing genes in tissues

Transcript profiling of blood cells and blood cell precursors—Data about transcript levels of in the above cells were extracted from a compendium of expression data sets generated by the Bloodomics Consortium and which in part have been released via Array Express at the European Bioinformatics Institute (www.ebi.ac.uk). Two main datasets have been used in this study: i) the HaemAtlas[25] which encompasses the results of whole genome expression (WGE) studies performed with RNA samples from the eight main blood cell types, including megakaryocytes (MKs) and erythroblasts (EBs) and ii) the WGE results obtained with the RNA samples from CD34+ haematopoietic stem cells (HSCs) and from a study of paired HSC cultures that were differentiated towards MKs and EBs. Cells were harvested during the 10-day culture at five time points in time (days 3, 5, 7, 9 and 10).

HaemAtlas[25]: The purification of blood cells is described in[25]. In short, RNA was isolated from the six main blood cell types (CD4+ Th (CD4, n=7) and CD8+ Tc lymphocytes (CD8, n=7), CD14+ monocytes (CD14, n=7), CD19+ B lymphocytes (CD19, n=7), CD56+ natural killer (NK) cells (CD56, n=7) and CD66b+ granulocytes (CD66, n=7) obtained from seven healthy blood donors from the Cambridge BioResource and from MKs and EBs. The latter two cells were obtained by cultures of cord blood-derived $CD34^+$ HSCs. MKs were obtained by cultures for 7 days in a medium supplemented with human recombinant thrombopoietin (THPO) and interleukin-1ß (IL1B) and the latter by cultures for 10 days in the presence of erythropoietin (EPO), interleukin-3 (IL3) and stem cell factor (SCF). To ensure high level purity preparations of both MKs and EBs, cells were flow-sorted using monoclonal antibodies against CD markers: MKs were positive for CD41 and negative for CD34; EBs were negative for CD41 and positive for CD235a. RNA was prepared from two sets of four cell preparations each.

Platelet Whole-Genome Expression data: Platelet RNA was isolated using a Trizol standard protocol. Total RNA was quantified using the NanoDrop (Labtech International) and quality-checked using the 2100 Bioanalyzer (Agilent Technologies). Standard protocols were applied to generate biotinylated cRNA and hybridize to Illumina BeadChips HumanWG-6 v2. Then, arrays were washed and scanned. All experimental procedures were carried out according to the manufacturer's protocol.

Time-course study of MK and EB cultures: To capture changes in transcript levels over time during the proliferation and differentiation of HSCs towards MKs and EBs paired cultures were generated as described above using CD34+ HSCs from three cord blood donations. Umbilical cord blood of three healthy newborns was collected into cord blood collections bags (MacoPharma, Mouvaux, France) with parental informed consent. CD34+ HSCs were prepared as in[26] and 92-98% pure CD34+ HSCs were *in vitro* cultured (1×10^5/ml) for 10 days in serum-free media (CellGro-SCGM, Cellgenix, France) supplemented with 50 ng/ml THPO (CellGenix, France) and 10 ng/ml IL1B (Miltenyi Biotech, Surrey, UK) to differentiate into MKs. EBs were *in vitro* derived from HSCs (5×10^3/ml) in the presence of 6 U/ml EPO (R&D Systems, Abingdon, UK), 10 ng/ml IL3 (Miltenyi Biotech, Surrey, UK) and 100 ng/ml SCF (R&D Systems, Abingdon, UK). MKs and EBs were harvested at day 3, 5, 7, 9 and 10 as described in[26] and transferred for RNA isolation into a 5 ml tube,

centrifuged at 500 x g for 10 min at RT, resuspended in Trizol (Invitrogen, Paisley, UK) and stored at −80°C. Total RNA was isolated according to manufacturer's instructions. The following murine monoclonal antibodies were used for phenotyping of both lineages at day 10: Fluorescein isothiocyanate (FITC) IgG1 isotype control, Phycoerythrin (PE) IgG1 isotype control, Allophycocyanin (APC) IgG1 isotype control, PE-Cy5 IgG1 isotype control and Pacific Blue (PB) isotype control (BD Bioscience Pharmingen™, Becton Dickinson, Oxford, UK), anti-CD11c V450, anti-CD13 APC, anti-CD14 PB, anti-CD15 V450, anti-CD33 FITC (BD Bioscience Pharmingen™), anti-CD34 PE (Beckman Coulter, High Wycombe, UK), anti-CD36 PE, anti-CD41a APC, anti-CD42a FITC, anti-CD235a FITC and anti-CD66c PE (BD Bioscience Pharmingen™). For phenotyping of cells at all other days anti-CD34 PE and anti-CD41a APC, anti-CD42a FITC and anti-CD235a FITC with matching isotype controls were used. In addition, a ploidy stain of the MKs was performed at each day of harvest as described in[26]. In brief, cells were stained with anti-CD41a APC and with matched isotype control and incubated at 37°C for 30 min in 500 μl PBE buffer containing 0.1% (v/v) Tween 20, RNAse A (0.1 mg/ml) and propidium iodide (0.05 mg/ml, Sigma-Aldrich, Dorset, UK). Samples were analysed using a 9-colour Cyan-ADP flow cytometer running Summit software version 4.3.02 (Beckman Coulter).

Sequencing of megakaryocyte RNA

Total RNA was prepared from MKs that were obtained using the culture protocol described above and extraction of RNA was according to the Trizol method (Invitrogen, Paisley, UK). The RNA pellet was resuspended in nuclease-free water (Applied Biosystems, Warrington, UK) and analysis by Agilent Bioanalyser 2100 (Agilent, Waldbronn, Germany) gave a RNA integrity number (RIN)[27] number of 8.4. Following DNAse treatment (Turbo DNA-Free, Applied Biosystems), 5 μg of total RNA was applied to the mRNA Sequencing kit (Illumina) following the manufacturer's instructions, except that PCR was performed before gel extraction of a band range of 300-450 bp, to obtain the purified library. This was quantified by qPCR followed by paired-end sequencing by GAII analyser (Illumina).

Paired-end 76 bp RNA-seq reads were obtained for the *in-vitro* derived MK samples (above). We sequenced one lane on the Illumina GAIIx and obtained 40,539,150 reads. The data were initially assessed using the FASTX-Toolkit version 0.0.13[28] and the ShortRead[29] package from version 2.7 of the Bioconductor project. Reads were then aligned to the February 2009 Homo sapiens high coverage assembly (Hg19) using TopHat v1.2.0[30]. Default parameters allow for up to 40 alignments per read with a maximum of 2 mismatches. The alignments here were produced using more stringent parameters where only a single unique alignment was allowed per read. We used BEDTools[31] to calculate the read coverage from the alignments, and visualized the results on the IGV browser[32] using the Ensembl annotation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge Senduran Balasubramaniam, Alison Coffey, the principal investigators of the CoLaus cohort[14], Richard Durbin, Jillian Durham, Peter Ellis, Kathleen Freson, GlaxoSmithKline collaborators, Cordelia Langford, Carol Scott for supporting different aspects of this study; see Supplementary Note for their specific contributions and for funding support.

References

1. Raccuglia G. Gray platelet syndrome. A variety of qualitative platelet disorder. Am J Med. 1971; 51:818–28. [PubMed: 5129551]
2. Breton-Gorius J, Vainchenker W, Nurden A, Levy-Toledano S, Caen J. Defective alpha-granule production in megakaryocytes from gray platelet syndrome: ultrastructural studies of bone marrow cells and megakaryocytes growing in culture from blood precursors. Am J Pathol. 1981; 102:10–9. [PubMed: 7468753]
3. Maynard DM, Heijnen HF, Gahl WA, Gunay-Aygun M. The alpha-granule proteome: novel proteins in normal and ghost granules in gray platelet syndrome. J Thromb Haemost. 2010; 8:1786–96. [PubMed: 20524979]
4. Gunay-Aygun M, et al. Gray platelet syndrome: natural history of a large patient cohort and locus assignment to chromosome 3p. Blood. 2010; 116:4990–5001. [PubMed: 20709904]
5. Fabbro, S., et al. Blood. 2011. Homozygosity mapping with SNP arrays confirms 3p21 as a recessive locus for gray platelet syndrome and narrows the interval significantly. First edition paper, DOI 10.1182/blood-2010-12-322990
6. Coffey AJ, et al. The GENCODE exome: sequencing the complete human exome. Eur J Hum Genet. 2011
7. Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–73. [PubMed: 20981092]
8. Jogl G, et al. Crystal structure of the BEACH domain reveals an unusual fold and extensive association with a novel PH domain. EMBO J. 2002; 21:4785–95. [PubMed: 12234919]
9. Wang X, et al. Neurobeachin: A protein kinase A-anchoring, beige/Chediakhigashi protein homolog implicated in neuronal membrane traffic. J Neurosci. 2000; 20:8551–65. [PubMed: 11102458]
10. Castermans D, et al. SCAMP5, NBEA and AMISYN: three candidate genes for autism involved in secretion of large dense-core vesicles. Hum Mol Genet. 2011; 19:1368–78. [PubMed: 20071347]
11. Medrihan L, et al. Neurobeachin, a protein implicated in membrane protein traffic and autism, is required for the formation and functioning of central synapses. J Physiol. 2009; 587:5095–106. [PubMed: 19723784]
12. Nagle DL, et al. Identification and mutation analysis of the complete gene for Chediak-Higashi syndrome. Nat Genet. 1996; 14:307–11. [PubMed: 8896560]
13. Spritz RA. Genetic defects in Chediak-Higashi syndrome and the beige mouse. J Clin Immunol. 1998; 18:97–105. [PubMed: 9533653]
14. 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. 2008; 8:6. [PubMed: 18366642]
15. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–9. [PubMed: 19505943]
16. Lunter G, Goodson M, Stampy. A statistical algorithm for sensitive and fast mapping of illumina sequence reads. Genome Research. 2011 doi10.1101/gr.111120.110.
17. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–303. [PubMed: 20644199]
18. [Accessed 12/2010] Web-based application Picard. http://picard.sourceforge.net
19. Albers CA, et al. Dindel: Accurate indel calls from short-read data. Genome Res. 2010
20. Hubbard T, et al. The Ensembl genome database project. Nucleic Acids Res. 2002; 30:38–41. [PubMed: 11752248]
21. Ng SB, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009; 461:272–6. [PubMed: 19684571]
22. Wakeley, J. Coalescent Theory: An Introduction. Roberts and Company Publishers; Greenwood Village, Colorado: 2009.
23. Westerfield, M. The Zebrafish Book. University of Oregon Press; Eugene, OR: 1994.
24. Detrich HW 3rd, et al. Intraembryonic hematopoietic cell migration during vertebrate development. Proc Natl Acad Sci U S A. 1995; 92:10713–7. [PubMed: 7479870]

25. Watkins NA, et al. A HaemAtlas: characterizing gene expression in differentiated human blood cells. Blood. 2009; 113:e1–9. [PubMed: 19228925]
26. Macaulay IC, et al. Comparative gene expression profiling of in vitro differentiated megakaryocytes and erythroblasts identifies novel activatory and inhibitory platelet membrane proteins. Blood. 2007; 109:3260–9. [PubMed: 17192395]
27. RNA integrity Number (RIN) - Standardization of RNA Quality Control. Waldbronn, Germany: Agilent.
28. [Accessed 12/2010] Web-based application. http://hannonlab.cshl.edu/fastx_toolkit/index.html
29. Morgan. ShortRead: a Bioconductor package for input, quality assessment, and exploration of high throughput sequence data. Bioinformatics. 2011; 25:2607–9. [PubMed: 19654119]
30. Trapnell. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25:1105–11. [PubMed: 19289445]
31. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–2. [PubMed: 20110278]
32. Robinson JT, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29:24–6. [PubMed: 21221095]

Europe PMC Funders Author Manuscripts

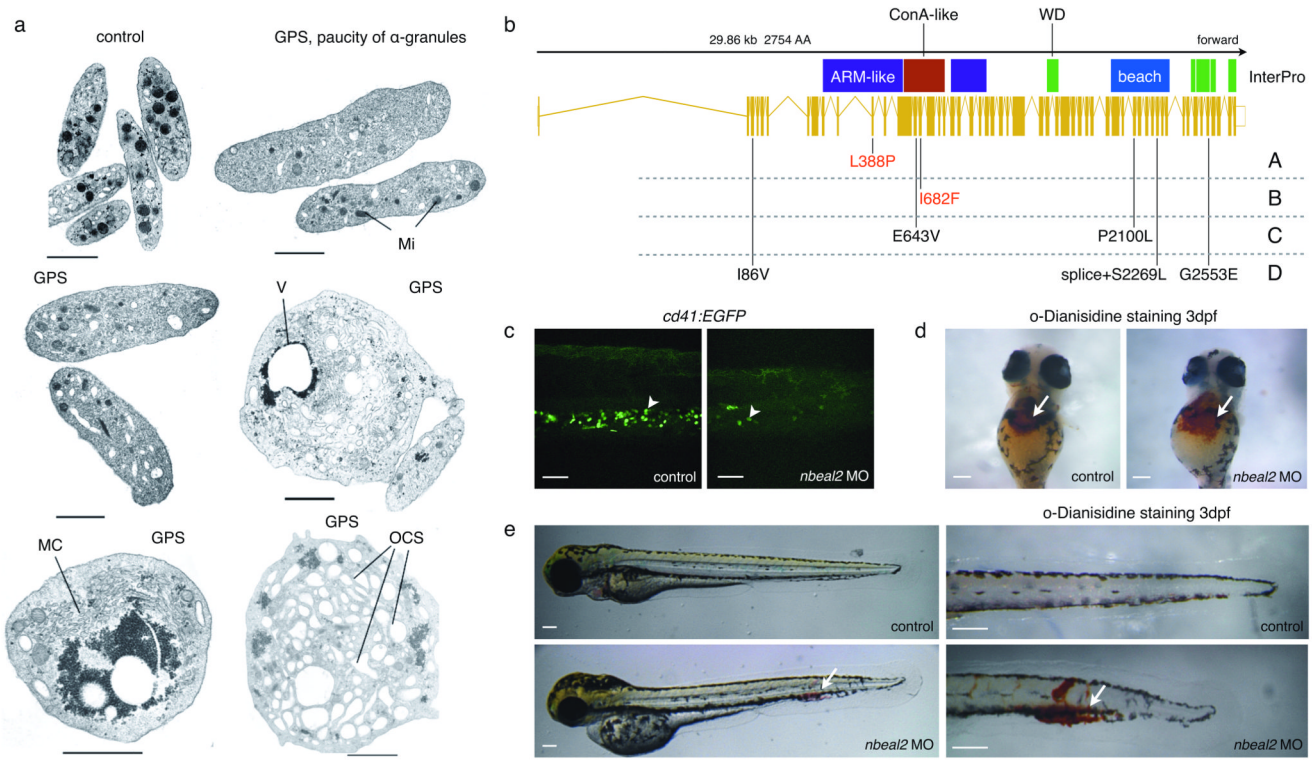

Figure 1.

Mutations in *NBEAL2* in Gray Platelet Syndrome cases. (**a**) Electron micrographs showing discoid normal platelets with abundant α-granules in comparison with a selection of platelets from Gray Platelet Syndrome (GPS) cases A.II.3 and B.II.3. The GPS platelets have heterogeneous shapes and characteristically lack α-granules. They normally contain mitochondria (Mi); morphological abnormalities include the presence of membrane complexes (MC), occasional large vacuoles (V) and an overdeveloped Open Canalicular System (OCS). Bars represent 1μm. (**b**) Novel variants in neurobeachin-like 2 (*NBEAL2)* in 4 whole-exome sequenced index cases (A-D, Supplementary Table 1) explain GPS. Heterozygous mutations are shown in black and homozygous ones in red. InterPro protein domains are shown as colored bars above the transcript, where blue indicates the characteristic BEACH domain. Sequencing reads showed that the splice and S2268L variants occurred on the same chromosome. (**c**) Silencing of *nbeal2* in zebrafish. To assess the function of *nbeal2* in thrombopoiesis we investigated cd41 expression in caudal haematopoietic tissue (CHT) of *Tg(cd41:EGFP)* transgenic embryos, at 3 days post-fertilisation (dpf). *nbeal2* morpholino (MO) knockdown resulted in a complete abrogation of thrombocytes (the zebrafish equivalent of human platelets) when compared to control (white arrowhead). White bars represent ~100μm. (**d**) Whole mount o-Dianisidine staining for mature erythrocytes at 3 dpf showed that the total number of mature erythrocytes was not affected in *nbeal2*-depleted embryos when compared to control (white arrow). (**e**) *nbeal2* MO-injected embryos showed normal morphological development and vigorous circulation at 3 dpf, similar to control embryos. However, 41% of *nbeal2* depleted embryos (N=78), and no control embryos (N=78) developed spontaneous bleedings visible within the tail of the embryo (left column). These spontaneous bleedings were confirmed with o-Dianisidine staining (right column, white arrows indicate the location of bleeding).

Europe PMC Funders Author Manuscripts

Table 1

Novel variants in *NBEAL2*. Chromosomal coordinates are given relative to build NCBI37 of the reference genome. Functional consequences were predicted from the CCDS transcript, which was confirmed by the sequencing of MK RNA (see Supplementary Note), and were either non-synonymous (ns) or disrupting a splice site. SIFT and Polyphen were used to predict deleterious impact of the mutations. Conservation is defined as the number of species where the residue was identical to that in human out of the number of species that had a protein alignment at the position of the mutation, using a total of 37 species including primates, placental mammals and vertebrates. The last four columns show the genotype of the four cases, where 'ref', 'het' and 'hom-alt' indicates a homozygote for the reference allele, a heterozygote and a homozygote for the alternate non-reference allele, respectively.

Position	Variant	SIFT	Polyphen	Exon	Type	AA change	Conservation	A	B	C	D
chr3:47030447	A/G	Tolerated	Benign	3	ns	I86V	23/27	ref	ref	ref	het
chr3:47035476	T/C	Damaging	Possibly damaging	11	ns	L388P	33/34	hom-alt	ref	ref	ref
chr3:47037233	A/T	Damaging	Possibly damaging	14	ns	E643V	31/31	ref	ref	het	ref
chr3:47037434	A/T	Damaging	Benign	15	ns	I682F	30/32	ref	hom-alt	ref	ref
chr3:47046466	C/T	Damaging	Probably damaging	39	ns	P2100L	29/29	ref	ref	het	ref
chr3:47047434	A/AGGAGTAG	NA	NA	42-43	splice site	Intronic	31/31	ref	ref	ref	het
chr3:47047440	C/T	Damaging	Possibly damaging	43	ns	S2269L	31/31	ref	ref	ref	het
chr3:47049615	G/A	Tolerated	Benign	50	ns	G2553E	6/33	ref	ref	ref	het