# Rapid genetic diagnosis of heritable platelet function disorders with next-generation sequencing: proof-of-principle with Hermansky–Pudlak syndrome

M. L. JONES*, S. L. MURDEN*, D. BEM†, S. J. MUNDELL‡, P. GISSEN§, M. E. DALY¶, S. P. WATSON†, and A. D. MUMFORD,* ON BEHALF OF THE UK GAPP STUDY GROUP

**Bristol Heart Institute & School of Cellular and Molecular Medicine, University of Bristol, Bristol; †Centre for Cardiovascular Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham; ‡School of Physiology and Pharmacology, University of Bristol, Bristol; §Institute of Child Health and MRC Laboratory for Molecular Cell Biology, University College London, London; and ¶Department of Cardiovascular Science, University of Sheffield, Sheffield, UK*

**To cite this article:** Jones ML, Murden SL, Bem D, Mundell SJ, Gissen P, Daly ME, Watson SP, Mumford AD, on behalf of the UK GAPP study group. Rapid genetic diagnosis of heritable platelet function disorders with next-generation sequencing: proof-of-principle with HermanskyPudlak syndrome. *J Thromb Haemost* 2012; **10:** 306–9.

Platelet function disorders (PFDs) are common heritable causes of excessive mucocutaneous bleeding, but are genotypically diverse [1]. Genetic diagnosis of PFDs is desirable clinically, as identification of pathogenic mutations assists diagnosis in symptomatic index cases, and enables identification of PFDs in family members who have ambiguous clinical or laboratory phenotypes. However, this task is usually impractical, because clinical and laboratory phenotype in PFDs can seldom be used to select single candidate genes for conventional Sanger sequencing. In other heterogeneous heritable disorders, technologies such as in-solution enrichment of target DNA and next-generation sequencing (NGS) enable the simultaneous analysis of large groups of candidate genes, and may be useful for rapid genetic diagnosis. Here, we describe a strategy for genetic diagnosis of PFDs with Agilent SureSelect in-solution enrichment and Illumina sequencing of 216 candidate genes. We provide proof-of-principle that this approach is clinically useful by identifying a pathogenic single-nucleotide variation (SNV) in *HPS4* in a subject with Hermansky–Pudlak syndrome (HPS) but with a previously unknown genotype.

We first generated a candidate PFD gene list by selecting genes previously associated with PFDs in humans. This was extended to include human orthologs of genes linked to platelet dysfunction in selected animal models, and other genes encoding important mediators of platelet activation but with no previous association with PFDs (Data S1). In order to enrich these candidate genes from genomic DNA (gDNA), we used the eArray programme (all URLs are given in Data S2) to design a library of 120-mer overlapping baits (0.5–1.5-Mb Custom SureSelect product; Agilent Technologies, Wokingham, UK). The baits were designed to tile 1.36 Mb of gDNA sequence corresponding to the exons and splice sites in all known transcripts of the candidate genes identified in build NCB137 of the reference genome. The library design enabled most nucleotides within the target sequence to be tiled with at least four unique baits.

The bait library was tested by enriching the candidate PFD genes from gDNA obtained from a pilot group of 10 subjects recruited to the UK Genotyping and Phenotyping of Platelets (UK-GAPP) study. All subjects had lifelong excessive mucocutaneous bleeding and displayed abnormal platelet light transmission aggregation or ATP secretion responses, indicating a PFD. After informed written consent had been obtained (NHS REC ref. 06/MRE07/36), gDNA from venous blood was sheared into 300–500-bp fragments and tagged with unique multiplexing primers for each subject. Enrichment of target DNA was performed according to the manufacturer's instructions, and all 10 samples were sequenced in 76-bp reads in a single lane of an Illumina GAII. The sequence output was then demultiplexed to yield a single sequence dataset for each subject. The total Illumina output was $4.1 \times 10^7$ sequence reads (mean, $4.11 \times 10^6$ per study subject; standard deviation, $\pm\ 7.9 \times 10^5$), with a mean Phred score of $> 30$ at each position within the reads indicating high sequence quality.

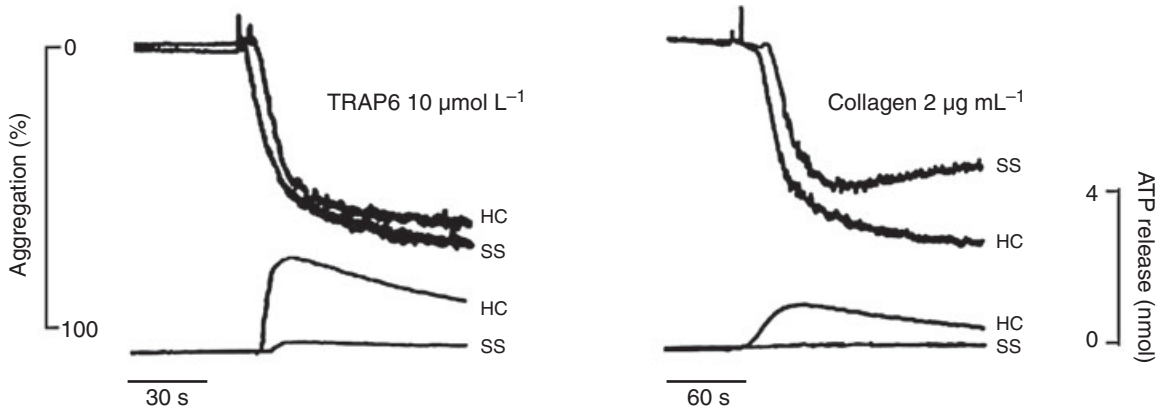
We then tested a strategy for mapping and filtering the Illumina output for SNVs within the candidate PFD genes in one subject from our pilot group. This subject was a 30-year-old male from a consanguineous family with oculocutaneous albinism in addition to a PFD. Platelets showed absent ATP secretion in response to all agonists, indicating defective dense granule number or release (Fig. 1A). These phenotype data indicated a background diagnosis of HPS. There are currently nine known human HPS genes and a larger group of other candidate HPS genes that encode components of the platelet secretion pathway. These genes cannot be distinguished with simple phenotype testing [2].

Correspondence: Matthew Jones, University of Bristol, Level 7, Bristol Royal Infirmary, Bristol BS2 8HW, UK.
Tel.: +44 117 3422655; fax: +44 117 3424036.
E-mail matthew.jones@bristol.ac.uk

DOI: 10.1111/j.1538-7836.2011.04569.x

Received 13 September 2011, accepted 16 November 2011

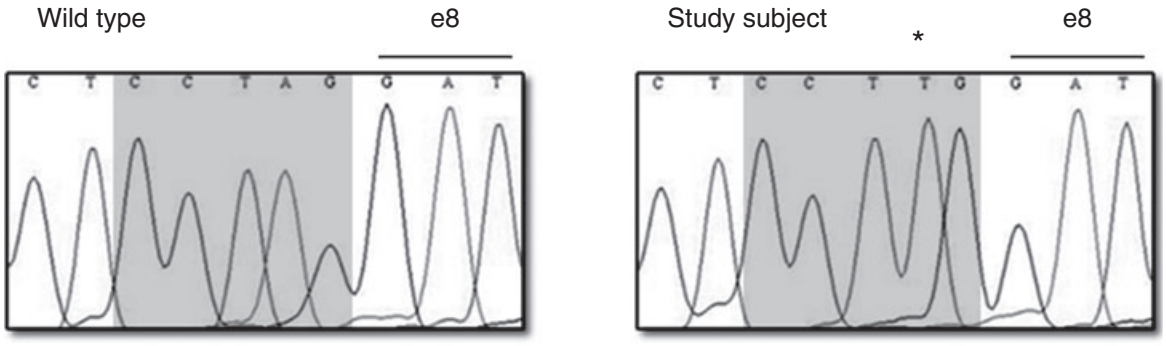
**A**




**B**

| Coordinate | Gene | Reference base | Observed base | Number of reads A | Number of reads C | Number of reads G | Number of reads T | Total reads | Confirmed by direct sequencing |
|---|---|---|---|---|---|---|---|---|---|
| Chr10:100182208 | *HPS1* | T | Y | 0 | 2 | 0 | 10 | 12 | No |
| Chr10:100185398 | *HPS1* | G | S | 0 | 2 | 2 | 0 | 4 | No |
| Chr15:91549893 | *VPS33B* | G | K | 0 | 0 | 10 | 2 | 12 | No |
| Chr19:47995334 | *NAPA* | G | R | 2 | 0 | 4 | 0 | 6 | No |
| Chr1235914653 | *LYST* | T | K | 0 | 0 | 2 | 14 | 16 | No |
| Chr22:26859902 | *HPS4* | G | K | 0 | 0 | 4 | 2 | 6 | No |
| Chr15:41186976 | *VPS18* | A | M | 4 | 2 | 0 | 0 | 6 | No |
| Chr20:2846067 | *VPS16* | G | R | 2 | 0 | 8 | 0 | 10 | No |
| Chr6:137245125 | *SCL35D3* | A | R | 4 | 0 | 2 | 0 | 6 | No |
| Chr22:26864591 | *HPS4* | A | T | 0 | 0 | 0 | 14 | 14 | Yes |

**C**




**D**

*Normal splicing*

e7 e8

...TCTATAAAGGACTGATTGTCAGCACCCAACTCCCGCCCTCCCTCACCGCCAAGGTCC

e9

TGCTTCACCGAACAGCACCTCAGGAGCAGAGACTCCCTACGGGAGAGGATGCCCCGCAAG

e10

GAACATGGAGCGGCATTGCCCCCGAATGTCCAGATTATCCCTGTTTTTGTGA...

*Aberrant splicing*

e7

...TCTATAAAGGACTA*GACTCCCTACGGGAGAGGATGCCCCGCAAGGAACATG*GAGCGGC

*ATTGCCCCCGAATGTCCAGATTATCCCTGTTTTTGTGA*...

**Fig. 1.** Platelet phenotyping and next-generation sequencing identifies a novel pathogenic single-nucleotide variation (SNV) at a splice acceptor site in *HPS4*. Platelet function testing by lumiaggregometry was performed according to the manufacturer's instructions (Chronolog). After being prewarmed for 180 s, platelet-rich plasma (PRP) was incubated for 60 s with Chronolume reagent. Thrombin receptor-activating peptide-6 (TRAP6) (10 μmol $L^{-1}$) or collagen (2 μg $mL^{-1}$) was then added to healthy control (HC) or study subject (SS) PRP. Simultaneous recording of aggregation (upper traces labeled HC or SS) and ATP release (lower traces labeled HC or SS) was performed for 180–300 s, until maximal aggregation was achieved for each agonist (A). Analysis of the Illumina sequencing reads from a study subject with a Hermansky–Pudlak syndrome phenotype identified 10 potential SNVs in genes involved in dense granule assembly or release (B). The chromosome coordinate of each potential SNV, gene name, reference sequence base and observed ambiguity (IUPAC nucleotide code convention relative to + strand) are shown in columns 1–4, respectively. Columns 5–9 indicate the number of Illumina sequence reads for each base and the total number of high-quality reads generated at the variant allele. Only the SNV at Chr22:26864591 (*HPS4* c.597-2 A > T in NM_022081.4) was called as variant in > 2 sequence reads and was confirmed by Sanger sequencing. This homozygous SNV (*) lay within the consensus splice acceptor site (shaded region) of intron 7 of *HPS4* transcript 1 (C), and is predicted to cause aberrant splicing of *HPS4*, in which exon 8 of the wild-type gene is skipped. This results in a frameshift downstream of exon 7 (italics) and a premature stop codon after 80 bp within the exon 10 sequence (D).

The 354 Mb of Illumina sequence generated from this subject was first mapped to the entire NCBI37 genome build, with both Bowtie and Burrows–Wheeler Aligner (bwa) tools within the Galaxy bioinformatics resource (Data S2). We then filtered the data for potential SNVs with the default quality thresholds (Phred score of > 20, and coverage of > 3). This yielded 22 087 (Bowtie) and 37 108 (bwa) potential SNVs. However, in-solution enrichment has previously captured significant quantities of off-target sequence [3,4]. Consistent with this, only 89 Mb of the total Illumina sequence from this subject was within the 216 candidate genes for PFDs (capture efficiency of 25%). When we restricted mapping to these candidate genes, the numbers of potential SNVs were reduced to 4164 (Bowtie; 18.9% of total mapped SNVs) and 4576 (bwa; 12.3% of total mapped SNVs).

As the phenotype of the subject included absent platelet ATP secretion, we further refined the candidate gene list to a shortlist of 57 genes implicated in dense granule assembly and release, including the known and putative HPS genes. Within the coding exons and associated splice sites of genes in this shortlist, there were 321 (Bowtie) and 361 (bwa) potential SNVs. We then eliminated potential SNVs that were not identified by both Bowtie and bwa, and used PolyPhen-2 (Data S2) to eliminate SNVs that had been identified previously as population variants in dbSNP132 (Data S2) and to predict the pathogenicity of the remaining potential SNVs. This yielded a group of 35 potential SNVs in 18 candidate genes, 25 of which were synonymous in all reported transcripts, and 10 of which were non-synonymous or occurred at splice sites. In nine potential SNVs in this shortlist, the variant allele was identified in only two sequence reads in the Illumina output (Fig. 1B) and none appeared as likely homozygous variants. When we resequenced the exons containing these SNVs with the reference Sanger sequencing method, we identified wild-type sequence in all cases, indicating that these were false-positive SNV calls in the Illumina output.

The single remaining potential SNV at Chr22:26864591 was identified in 14/14 Illumina sequence reads for this locus (Fig. 1B) and was confirmed as a homozygous SNV by Sanger sequencing (Fig. 1C). This SNV lies within the *HPS4* gene, for which three transcripts encoding HPS4 protein isoforms have been identified in the Consensus CDS protein set (Data S2). In all three coding transcripts, this SNV occurs in a splice acceptor site predicted to disrupt mRNA assembly. In *HPS4* transcript variant 1 (NM_022081.4), this SNV corresponds to a c.597-2 A > T transversion in the intron 7 splice acceptor site. Analysis of the variant *HPS4* sequence using neural network, NetGene2 and Human Splicing Finder 2.4.1. splice prediction programmes (Data S2) showed that there were no other plausible splice acceptor sites within this region. Therefore, in this transcript, *HPS4* c.597-2 A > T is predicted to cause abnormal splicing between exons 7 and 9, leading to a frameshift and a premature stop codon in exon 10 (Fig. 1D). This is expected to cause expression of a truncated HPS4 protein or to prevent expression entirely through nonsense-mediated decay of the variant mRNA. Although this SNV has not previously been reported in association with HPS, other SNVs that prevent HPS4 expression have been recognized previously, and define the type 4 variant (OMIM #614073; Data S2) of the HPS group of disorders [5].

In this pilot study of 10 subjects with different PFDs, we have demonstrated that high-quality sequence data from a large group of platelet genes can be generated by Agilent SureSelect in-solution enrichment and Illumina sequencing. In common with previous applications of NGS for genetic diagnosis, mapping of the Illumina sequence data initially yielded large numbers of potential SNVs in our study subject. However, we were able to refine an initial yield of approximately 4500 potential SNVs in the 216 candidate genes for PFDs to a shortlist of 10 potentially pathologic SNVs. This elimination of irrelevant SNVs required a systematic filtering strategy, in which we made the prior assumptions that the pathogenic SNV in the study subject was: (i) within a candidate gene coding region or splice site; (ii) not a population variant identified in dbSNP132; and (iii) identified with two different bioinformatic tools. Crucially, we also used the clinical and laboratory phenotype of HPS in the subject to refine filtering for potential SNVs to a shortlist of 57 candidate genes implicated in platelet granule assembly or release. Within the shortlist of 10 potentially pathogenic SNVs, only one SNV was confirmed by Sanger sequencing, indicating a high false-positive call rate in the mapped SNVs. The single remaining *HPS4* c.597-2 A > T transversion was predicted to prevent HPS4 protein expression by disrupting gene splicing, and is likely to be the pathogenic SNV responsible for the PFD phenotype.

This proof-of-principle study illustrates that NGS enables rapid genetic diagnosis of a PFD in a single test. In this example, we were able to restrict SNV mapping to a subgroup of 57 genes implicated in secretion, so that it was feasible to use Sanger sequencing to determine whether each potential SNV was a true-positive or a false-positive call. Although successful in HPS, restricting analysis to a subgroup of genes is likely to reduce the diagnostic yield of NGS for other PFDs where it is less easy to select a candidate gene list. Strategies such as whole-exome sequencing may circumvent this difficulty by increasing the overall sensitivity of NGS for pathogenic SNVs. However, increasing the number of mapped genes is also expected to yield significantly larger numbers of irrelevant SNVs and false-positive calls. Alternative NGS strategies require further evaluation in other PFDs to determine the optimum diagnostic and cost-effective approach.

### Acknowledgements

Preparation and enrichment of gDNA and Illumina sequencing was performed within the University of Bristol Transcriptomics facility (Data S2). This work was funded by the British Heart Foundation (RG/09/007). S. P. Watson holds a British Heart Foundation Chair (CH/03/003).

### Disclosure of Conflict of Interest

The authors state that they have no conflict of interest.

### Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Data S1.** The 216 candidate PFD genes.
**Data S2.** The accessible internet-based resources used in this work.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

### References

1 Nurden A, Nurden P. Advances in our understanding of the molecular basis of disorders of platelet function. *J Thromb Haemost* 2011; **9**(Suppl. 1): 76–91.
2 Huizing M, Helip-Wooley A, Westbroek W, Gunay-Aygun M, Gahl WA. Disorders of lysosome-related organelle biogenesis: clinical and molecular genetics. *Annu Rev Genomics Hum Genet* 2008; **9**: 359–86.
3 Tewhey R, Nakano M, Wang X, Pabon-Pena C, Novak B, Giuffre A, Lin E, Happe S, Roberts DN, LeProust EM, Topol EJ, Harismendy O, Frazer KA. Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biol* 2009; **10**: R116.
4 Shearer AE, DeLuca AP, Hildebrand MS, Taylor KR, Gurrola J II, Scherer S, Scheetz TE, Smith RJ. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. *Proc Natl Acad Sci USA* 2010; **107**: 21104–9.
5 Anderson PD, Huizing M, Claassen DA, White J, Gahl WA. Hermansky–Pudlak syndrome type 4 (HPS-4): clinical and molecular characteristics. *Hum Genet* 2003; **113**: 10–17.

# RGS2 deficiency in mice does not affect platelet thrombus formation at sites of vascular injury

F. BANNO,* T. NOJIRI,* S. MATSUMOTO,* K. KAMIDE*† and T. MIYATA*
**Department of Molecular Pathogenesis, National Cerebral and Cardiovascular Center, Suita, Osaka; and †Department of Geriatric Medicine, Osaka University Graduate School of Medicine, Suita, Osaka, Japan*

**To cite this article:** Banno F, Nojiri T, Matsumoto S, Kamide K, Miyata T. RGS2 deficiency in mice does not affect platelet thrombus formation at sites of vascular injury. *J Thromb Haemost* 2012; **10:** 309–11.

Correspondence: Fumiaki Banno, Department of Molecular Pathogenesis, National Cerebral and Cardiovascular Center, 5-7-1 Fujishirodai, Suita, Osaka 565-8565, Japan.
Tel.: +81 6 6833 5012; fax: +81 6 6835 1176.
E-mail: banno@ri.ncvc.go.jp

DOI: 10.1111/j.1538-7836.2011.04575.x

Received 12 October 2011, accepted 12 November 2011



RGS2 is a member of the regulator of G-protein signaling (RGS) protein family that plays a key role in the regulation of G-protein-coupled receptors. RGS2 inhibits Gq-mediated and Gi-mediated signaling by activating the intrinsic GTPase of the Gα subunit [1,2]. RGS2 also interacts with Gs and adenylyl cyclase, and suppresses Gs signaling independently of GTPase-activating protein activity [2,3]. In platelets, the activation of the Gq-coupled or Gi-coupled receptors by thrombin, thromboxane $A_2$ and ADP stimulates platelet aggregation, whereas the activation of the Gs-coupled prostacyclin receptor inhibits aggregation [4,5]. Recently, active roles of RGS proteins in regulating platelet G-protein-coupled receptors have been