Original Article

Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alström Syndromes

Claire Redin, Stéphanie Le Gras, Oussema Mhamdi, Véronique Geoffroy, Corinne Stoetzel, Marie-Claire Vincent, Pietro Chiurazzi, Didier Lacombe, Ines Ouertani, Florence Petit, Marianne Till, Alain Verloes, Bernard Jost, Habiba Bouhamed Chaabouni, Helene Dollfus, Jean-Louis Mandel, Jean Muller

ABSTRACT

Background Bardet-Biedl syndrome (BBS) is a pleiotropic recessive disorder that belongs to the rapidly growing family of ciliopathies. It shares phenotypic traits with other ciliopathies, such as Alström syndrome (ALMS), nephropathis (NPHP) or Joubert syndrome. BBS mutations have been detected in 16 different genes (BBS1-BBS16) without clear genotype-to-phenotype correlation. This extensive genetic heterogeneity is a major concern for molecular diagnosis and genetic counselling. While various strategies have been recently proposed to optimise mutation detection, they either fail to detect mutations in a majority of patients or are time consuming and costly.

Method We tested a targeted ex-on-capture strategy coupled with multiplexing and high-throughput sequencing on 52 patients: 14 with known mutations as proof-of-principle and 38 with no previously detected mutation. Thirty genes were targeted in total including the 16 BBS genes, the 12 known NPHP genes, the single ALMS gene ALMS1 and the proposed modifier CCDC28B.

Results This strategy allowed the reliable detection of causative mutations (including homozygous/ heterozygous exon deletions) in 68% of BBS patients without previous molecular diagnosis and in all proof-of-principle samples. Three probands carried homozygous truncating mutations in ALMS1 confirming the major phenotypic overlap between both disorders. The efficiency of detecting mutations in patients was positively correlated with their compliance with the classical BBS phenotype (mutations were identified in 81% of ‘classical’ BBS patients) suggesting that only a few true BBS genes remain to be identified. We illustrate some interpretation problems encountered due to the multiplicity of identified variants.

Conclusion This strategy is highly efficient and cost effective for diseases with high genetic heterogeneity, and guarantees a quality of coverage in coding sequences of target genes suited for diagnosis purposes.

INTRODUCTION

Bardet-Biedl syndrome (BBS; OMIM# 209900) is a pleiotropic recessive disorder with high non-allelic genetic heterogeneity. Its incidence varies from an estimated 1:160 000 in northern Europe to 1:13 500–17 000 in Bedouins and Newfoundlanders, respectively.1 BBS belongs to the large and growing family of ciliopathies and, therefore, shares phenotypic traits with Joubert (JBTs), Alström (ALMS) and Meckel (MKS) syndromes.1 3 Differential clinical diagnosis may thus be difficult, especially in young probands who do not yet show some later onset-specific manifestations.3 4 In particular, recent reports highlight a significant clinical overlap between BBS and ALMS.3 5

The main phenotypic features of BBS comprise retinal dystrophy, polydactyly, obesity, mild developmental delay, polycystic kidneys and hypogonitalism. Other minor features can also be observed in patients, such as cardiac abnormalities, other digit or eye anomalies, diabetes, hypertension, hearing defects, anosmia.6 7 Up to now, mutations have been detected in 16 different genes (BBS1-BBS16), but no clear genotype-to-phenotype correlation could be observed, besides the suggested exception of BBS16.8

Alström syndrome (OMIM #203800) was reported to be much less prevalent than BBS, with an estimated incidence of 1:1 000 000. Its phenotypic features overlap with those of BBS in early infancy and include: cone-rod dystrophy, obesity, type 2 diabetes mellitus, hearing loss but also hypertiglyceridaemia, dilated cardiomyopathy, and progressive pulmonary, hepatic, or renal dysfunction.9 To date, only one gene (ALMS1) has been identified, but recent reports showed some families with suggestive ALMS-carrying mutations in BBS genes.3 5 The large size of ALMS1 coding sequence appears to have impaired widespread diagnostic testing of ALMS.

Exhaustive conventional Sanger sequencing for BBS diagnosis is prohibitively expensive because of the large number of genes involved, and so also for ALMS due to the large size of ALMS1 coding sequence (12 kb, 24 exons; table 1). Alternative
Table 1 Genes included in the targeted enrichment strategy and their associated disorders

| Official gene symbol | BBS# | NPHP# | Other disease-related symbols | # Exons in consensus | # Total exons in all isoforms | Size of coding exons (bp) | Gene size (bp) | Targeted region size |
|----------------------|------|-------|------------------------------|----------------------|-------------------------------|--------------------------|---------------|---------------------|
| BBS1*                | BBS1 | –     | –                            | 17                   | 17                            | 1782                     | 22,966        | 23,285*            |
| BBS2                 | BBS2 | –     | –                            | 17                   | 17                            | 2166                     | 35,750        | 38,011             |
| ARL6                 | BBS3 | –     | RP55                         | 8                    | 11                            | 561                      | 33,779        | 2123               |
| BBS4*                | BBS4 | –     | LCA                          | 16                   | 18                            | 1560                     | 52,292        | 52,611*            |
| BBS5                 | BBS5 | –     | –                            | 12                   | 17                            | 1026                     | 27,160        | 3997               |
| MKS1                 | MKS1 | –     | –                            | 6                    | 17                            | 1713                     | 29,034        | 3326               |
| BBS7                 | BBS7 | –     | –                            | 19                   | 20                            | 2148                     | 46,008        | 5308               |
| TTC8                 | TTC8 | –     | –                            | 15                   | 16                            | 1518                     | 53,358        | 3137               |
| KK51                 | KK51 | –     | –                            | 18                   | 20                            | 1680                     | 14,170        | 3745               |
| CEP290               | CEP290 | –      | –                            | 54                   | 55                            | 7440                     | 93,204        | 10,510             |
| WDPCP                | WDPCP | –     | –                            | 2                    | 2                             | 2172                     | 3957          | 3941               |
| SDCAG8               | SDCAG8 | –     | –                            | 18                   | 19                            | 2469                     | 467,317       | 4405               |
| ALMS1                | ALMS1 | –     | ALMS; LCA                     | 23                   | 24                            | 2559                     | 476,529       | 5196               |
| NPHP1                | NPHP1 | –     | JBTS4; SLSN1                  | 2                    | 2                             | 2133                     | 12,242        | 3829               |
| NPHP3                | NPHP3 | –     | MKS7; RHPD                    | 2                    | 3                             | 3198                     | 201,916       | 4103               |
| NPHP4                | NPHP4 | –     | SLSN4                         | 3                    | 3                             | 3993                     | 41,823        | 5328               |
| NPHP5                | NPHP5 | –     | SLSN5                         | 15                   | 15                            | 1797                     | 65,317        | 2585               |
| GLIS2                | GLIS2 | –     | –                            | 8                    | 8                             | 1575                     | 73,74         | 2175               |
| RPGRIP1L             | RPGRIP1L | –    | MKS5; JBTS7; CS             | 27                   | 27                            | 3708                     | 103,954       | 5243               |
| NPHP8                | NPHP8 | –     | MKS5; JBTS7; CS             | 27                   | 27                            | 3708                     | 103,954       | 5243               |
| NPHP9                | NPHP9 | –     | –                            | 15                   | 15                            | 2078                     | 13,953        | 3096               |
| TMEM67               | TMEM67 | –    | MKS3; JBTS6; CS          | 28                   | 30                            | 2745                     | 64,389        | 4979               |
| TTC21B               | TTC21B | –   | JBTS11; AD4                  | 29                   | 29                            | 3951                     | 79,894        | 5414               |
| TMEV16               | TMEV16 | –   | MKS2; JBTS2                  | 5                    | 5                             | 447                      | 6504          | 1795               |
| AH11                 | AH11 | –     | JBTS3                        | 27                   | 29                            | 3591                     | 213,749       | 5175               |
| CDC28B               | CDC28B | –   | –                            | 6                    | 6                             | 603                      | 4790          | 1079               |
| Total                |       |       |                               | 483                  | 545                           | 79,781                   | 2,865,109     | 300,319            |

*Sequence of the entire gene (coding/non-coding exons, introns) was targeted.
ALMS, Alström syndrome; ATO, asphyxiating thoracic dystrophy; BBS, Bardet-Biedl syndrome; CS, COACH syndrome; JBTS, Joubert syndrome; LCA, Leber congenital amaurosis; LGMD, limb-girdle muscular dystrophy; MKS, McKusick-Kaufman syndrome; MKS, Meckel-Gruber syndrome; NPHP, nephronophthisis; RHPD, renal-hepatic-pancreatic dysplasia; RP, retinitis pigmentosa; SLSN, Senior-Loken syndrome; STM, sarcotubular myopathy.

Cost-conscious strategies have been proposed for BBS diagnosis, such as: initial screening of recurrent mutations and frequently mutated genes (BBS1, BBS10, and BBS12) combined with homozygosity mapping for consanguineous families; or primer extension arrays to test a series of known BBS mutations. Another approach recently proposed is the pooling of patients’ DNAs with subsequent PCR-amplification and massive parallel resequencing of BBS1-12 coding exons, followed by heteroduplex screening to identify the mutation carrier. Such a method presents some limitations as it will miss exon deletions and may not be suited for diagnostic purposes. Considering the clinical overlap with other ciliopathies, another approach would be to test, systematically and simultaneously, all corresponding genes for such overlapping syndromes, which would be particularly relevant for patients with atypical or incomplete clinical phenotypes. We describe here the results of such an approach, based on a targeted exon capture of 30 genes coupled to next-generation sequencing (NGS).

**SUBJECTS AND METHODS**

Detailed protocols are available in Supplementary Methods.

**Subjects**

DNA samples from 52 unrelated patients were collected. Most patients had been addressed to the diagnostic laboratory, or to the National Reference Center for rare ophthalmogenic diseases in Strasbourg. Eleven DNA samples stemmed from Tunisian patients included in an independent BBS epidemiology study.

The proof-of-principle cohort included 14 non-Tunisian patients with a confirmed BBS molecular diagnosis (identified prior to this study by Sanger sequencing). Twenty-six out of the 38 patients included in an independent BBS epidemiology study. Eleven DNA samples stemmed from Tunisian patients included in an independent BBS epidemiology study.

For all patients, a written consent for genetic testing was obtained, either from adult probands or from the legal representative in case of minors.

**Library preparation, targeted capture and sequencing**

DNA samples were prepared and controlled following standard procedures.

The capture design was performed with eArray following the manufacturer’s instructions (Agilent). DNAs (5 μg) were sheared mechanically using Covaris E220 (duty cycle: 10%; intensity: 5; cycles per burst: 200; time: 500 s).

For the proof-of-principle experiment, sequencing adaptors were added on 500 ng of sheared DNA using the SPRWorks Fragment Library System I (Beckman Coulter). After amplification and quality assessments, targeted capture was performed on individual samples using the in-solution SureSelect Target Enrichment System (Agilent) on 500 ng of DNA-prepped library. Additional
steps of washing, purification and elution were performed, and multiplexing adapters (truseq illumina DNA indexes) were added by PCR during the post-capture amplification step.

For all following experiments, multiplexing adapters were added simultaneously to sequencing adapters using the SPRI-works system. Equimolar amounts of two tagged libraries were then pooled prior to the capture reaction. All other following steps prior to sequencing remained identical. A 72-bp single-read sequencing was performed on a Genome Analyser IIx (GAIIx, Illumina).

Bioinformatic pipeline
Read mapping and variant calling were performed following standard procedures. Variant filtering was performed using VaRank, an in-house software which collects variant-specific information to rank them according to their predicted pathogenicity (figure 1, Supplementary Methods).

Copy-number variation (CNV) detection method
CNVs were identified using a depth-of-coverage method. For each patient, read counts in non-overlapping windows of 20 nucleotides were compared, normalised and then compared randomly with eight other samples from the same experiment (considered as replicates) using the Bioconductor package DEseq (initially designed for RNA-seq data). Candidate regions for CNVs were retrieved when log2 ratios (controls/sample) were either \( \geq 0.84 \) (fold change >1.8, potential deletion) or \( \leq -0.51 \) (fold change <0.7, potential duplication), and if \( p \) values adjusted for multiple testing (Benjamini and Hochberg procedure) were smaller than 0.1.

Statistical methods
Confidence intervals were computed for proportion estimates and indicated in brackets. Fisher’s exact test was computed to compare distributions of small populations. Subsequent \( p \) value is given at \( \alpha = 0.05 \).

RESULTS AND DISCUSSION
Targeted regions: design strategy
Our primary goal was to develop an efficient mutation-screening strategy for the diagnosis of patients with phenotypes evocative of BBS, or of clinically overlapping ciliopathies. We chose a target enrichment approach coupled with NGS in order to focus the sequencing on genomic regions of interest. We targeted all exons (including 5’ and 3’ UTRs) of the 16 known BBS genes (table 1).

Because of the known clinical overlap, we also included coding exons of ALMS1, and of all 12 known nephronophthisis genes (NPHP1–12), since retinal degeneration can often be observed in this kidney-specific disease. Coding sequences of AH1/JBTS3, TMEM216/MKS2/JBTS2, and of the proposed BBS-modifier CCDC28B/MGC1203, were also targeted. Because some of these genes are associated with multiple phenotypes, our design includes 6 MKS, 7 JBTS and 4 Senior-Loken syndrome (SLSN) genes (see table 1).

With this first design, we wanted to investigate whether including intronic sequences could favour both, the detection and sizing of exon deletions. We therefore included bait-targeting intronic sequences of BBS1 and BBS4. This choice was dictated by two observations: an apparent excess of patients heterozygous for the BBS1-recurrent mutation M590R, with no second mutation detected, and multiple reports of BBS4 exon deletions in patients. A maximal threshold of 200 kb for cumulated targeted regions was set because of the manufacturer’s pricing limits.

Presence of repeated sequences precluded bait tiling in 19.7% of initially targeted regions. This concerned, almost exclusively, of introns of BBS1 and BBS4, besides a small number of splice sites, large rearrangements (or known mutations).

Proof-of-principle and technical results
In our proof-of-principle experiment, we selected 16 DNA samples, of which 14 were with known BBS mutations. In this first trial, after barcoding the target-enriched libraries, we sequenced pools of four or eight libraries per lane of a GAIIx (see Supplementary Methods). This proof-of-principle analysis was carried blind, that is, without knowledge of implicated BBS genes and their associated mutations. A constellation of all mutation types (missenses, nonsenses, splice mutations, large deletions and complex rearrangements) at different allelic dosage was tested (figure S1). All 14 previously identified mutations, including two heterozygous BBS1-deletions (figure 2A), were detected in their correct heterozygous/homozygous state (table S2). In particular, in patient AKE12, we could detect an abnormal local drop of coverage in BRS12 due to a rare mutation type (insertion of an Alu sequence, figure S1A) although the exact nature of the mutation could only be determined by Sanger sequencing. A
similar drop in coverage was observed for a second patient, AHX91, with another complex mutation detected previously by Sanger sequencing (insertion/inversion in \(BBS5\)).

In this first experiment, we almost systematically reached the maximal theoretical coverage of 144x illustrated by a mean coverage of 127.6x after removal of duplicate reads (table 2). Due to this global saturating coverage when considering unique reads, we used all reads, including duplicates, when applying our depth-based method for the detection of CNVs.

These promising depth-of-coverage results (table 2, table S3) encouraged us to further increase the number of pooled samples. In the next experiments we used a single capture reaction for two barcoded libraries, allowing both cost and bench-time savings, and pooled 12 libraries per sequencing lane (maximum number of barcodes proposed at that time by Illumina).

This new protocol was performed on a second cohort of 36 patients with unknown mutations. Sequencing resulted in a mean coverage of 78.6x (283.153x before discarding duplicate reads) with 91.4±6.4% of targeted regions being covered more than 40x (table 2). This relative drop of coverage appears to be a consequence of a lower capture efficiency that might be due to: (1) an input amount of individual library reduced by half, due to the pre-capture pooling and (2) the addition of barcodes before capture, leading to less efficient blocking and unspecific hybridisation. The resulting coverage still guarantees a reliable detection of variants and of their homozygous/heterozygous state.

A small proportion of targeted regions was weakly covered in some patients (ie, depth <10x after duplicates filtering), with very few of them in a systematic way in other patients (table S4). This only concerned 0.63±0.68% of protein coding regions, and mostly included intronic GC-rich sequences (GC content: 68.3±5% vs 40.2±10% across all targeted regions), or some first exons (tables S3 and S4).

### Table 2

| Mean coverage \((\times)\) | Targeted regions \(\times 5\) coverage | Targeted regions \(\times 40\) coverage | Targeted regions \(\times 80\) coverage | Reads in targeted regions |
|--------------------------|---------------------------------------|----------------------------------------|----------------------------------------|--------------------------|
| Before filters           | After filters                          | After filters                          | After filters                          |                           |
| Pool of 4*               | 2208±416                               | 130±3                                  | 100±0                                  | 99.9±0.0                 | 98.3±0.6                 | 76±2%                    | 35±5%                    |
| Pool of 8*               | 1024±151                               | 125±3                                  | 99.9±0.3                               | 99.7±0.3                 | 96.8±0.9                 | 74±5%                    | 40±6%                    |
| Pool of 12†             | 283±153                                | 78±17                                  | 98.8±0.4                               | 91.4±6.4                 | 48.7±28%                 | 25±9%                    | 14±4%                    |

*First set of experiment. The capture-enrichment step was done individually, on untagged DNA libraries.
†Second set of experiment. The capture-enrichment step was realised on equimolar pools of 2 barcoded DNA libraries.
Variant filtering: importance of databases and frequency data

In targeted regions, we detected, on average, 170 variants (Single Nucleotide Variants (SNVs) and indels) per patient. All were systematically analysed for putative effect on protein structure and splice sites using VaRank (figure 1, Supplementary Methods). About 130 of these variants were recorded in dbSNP134 (table S5), but only 20 were validated with at least two independent methods and, therefore, filtered out. Indeed, in the context of a rare recessive disorder, some true mutations can be present at very low frequency in a heterozygous state in controls.

Potential pathogenicity of the remaining 150 variants was assessed using bioinformatic tools and considering their allele frequency in a European-American population, as reported in the Exome Variant Server database (EVSDb). This yielded from zero to six interesting variants per sample, among which were obvious truncating or known mutations in some patients.

The new ‘clinical significance’ field introduced in dbSNP134 has to be considered with caution since established mutations can now be reported in the database but are not systematically flagged as pathogenic (example: rs179568397, p.R138H mutation in BBS5). Conversely, we detected some false-positive annotations: rs4784677 (p.N70S) in BBS2—initially reported as a third allele according to the triallelic hypothesis—is flagged as pathogenic, but is too frequent to be a fully penetrant mutation (0.77% in EVSDb). Filters have to be carefully adapted to the disorder of interest, and to the constantly evolving updates of databases.

Detection of exon deletions

One advantage of NGS-based strategies, as opposed to Sanger sequencing, is the opportunity to detect—in addition to SNVs and small indels—CNVs affecting one or more exons (figure 2 and S2). In the proof-of-principle experiment, two heterozygous deletions could be detected in BBS1. Among the unknown samples, two homozygous deletions in BBS3/ARL6 and BBS4 were identified. To our knowledge, we provide here the first report of large deletions in BBS1 and BBS3/ARL6, while several deletions affecting BBS4 have been previously observed.

Since we also targeted intronic sequences of BBS1 and BBS4, we were able to narrow the boundaries of subsequent detected deletions (figure 2). For patient AMV5, by using coordinates of affected exons, the estimated size of BBS1 deletion would be between 466 and 4707 bp, while with our design, we could restrict it to 1802–3841 bp (figure 2A, table S2). In patient P3, we could similarly reduce the assessed size of the BBS4 deletion from 4626–12975 bp based on exon positions down to 9376–10469 bp (figure 2C, table S4A). Lastly, since the BBS3/ARL6 deletion in patient AGL42 encompasses the first three exons of the gene (figure 2B), we tested whether it may extend and affect EPHA6 located upstream, encoding an ephrin receptor. Direct PCR testing excluded such extended deletion (data not shown).

Thanks to this method, in the six patients in whom we detected a single heterozygous potentially pathogenic mutation, we can ascertain that no heterozygous deletion is present in trans, or at least none encompassing exon sequences.

Distribution of detected BBS mutations in the 38 unknown patients

Of the 38 samples with unknown BBS mutations (56+2 from the proof-of-principle experiment) we detected clearly pathogenic biallelic mutations in 26 cases (68.4%; table 3A). To our knowledge, seventeen of these mutations have not been reported previously, indicating that the BBS mutation spectrum is far from being saturated in spite of numerous BBS mutation reports. Homozygous mutations were found in 88% (23/26) of mutated patients, coherent with the large number of consanguineous probands included in the cohort (75%; 25/33). In two patients of consanguineous origin, the BBS mutation was located outside the homozygous regions detected by prior SNP array analysis and would, therefore, have been missed using a homozygosity mapping strategy (patients AGL23, AKX44; table 3A).

Among the remaining 12 patients with no biallelic mutations identified (table 3B), one patient (AH2) had a heterozygous clearly pathogenic splicing mutation in BBS3. Five patients had heterozygous missenses predicted to be damaging by SIFT, Polyphen2 and/or Mutation T@ster (AIY87, AIX45, AMQ77, AMA28, AHL86) with the latter two carrying such variants in two different genes. One consanguineous Melanesian patient, AKE98, presented with classical BBS-inclusion features, including polydactyly. He carried two homozygous variants which initially appeared as potentially pathogenic: a distal frameshift in INVS/NPHP2 predicted to add 14 amino-acids to the C-terminus of the longer protein isoform, and a non-reported missense F2679L affecting a conserved residue in ALMS1 (figure S3). Subsequent segregation analysis ruled out their implication in the disease since both variants were heterozygous in a similarly affected brother.

In five patients, no potentially pathogenic variant could be identified in any of the 30 targeted genes. These patients are thus candidate for exome sequencing which might either help in identifying novel genes or in reconsidering the clinical diagnosis.

Mutation load in BBS and other targeted genes: importance of ALMS1

The mutation load among BBS genes in our cohort appears consistent with previous reports. Observed occurrences for BBS1 (7/38, 18.4%) and BBS2 mutations (5/38, 13.2%), the most frequently mutated genes in our study, are similar to the respective reported figures of 16.9% and 12%.

Considering frequently mutated genes, our study was strongly biased against BBS1, BBS10 and BBS12, since two-thirds of the patients’ DNAs were previously tested negative for BBS1 and BBS10 recurrent mutations, plus all BBS12 protein coding sequence. This explains the total absence of BBS12 mutations in our cohort, and the relatively low contribution of BBS10 (5/38, 13.2%) as compared with the literature (≥20%). The contribution of other BBS genes was low, with frequently only one proband involved.

We did not find any mutation in the ‘new’ BBS genes (BBS3–16) suggesting that, cumulatively, they have a small contribution to the total mutation load. BBS13/MKS1 was indeed shown to be mostly implicated in MKS since only one BBS patient was reported with two heterozygous mutations p.[C492W];[F371del] others carried only heterozygous missenses, sometimes in addition to homozygous truncating BBS1 mutations. Likewise, for BBS14/CEP290, a homozygous truncating mutation (p.E1903*) was found in a single BBS patient, while other mutations are much more often implicated in Joubert, Senior Loken, Leber Congenital Amaurosis or Meckel syndromes. Similar observations can be made for BBS15/WDPCP and BBS16/SYCCAGA. Like in all other studies of BBS cohorts, no mutation was identified in BBS11/TRIM32 raising the question of its real implication in BBS: only one homozygous missense mutation was described in a single...
Table 3  Identified mutations and other potentially pathogenic variants in the 38 patients with previously unknown genotype. A) Patients with two clearly pathogenic variants in one gene; B) Patients with a single or no clear pathogenic variant in one gene. Mutations are described according to the latest nomenclature conventions described in HGVS.

| Patient# | Principal mutations (p.) | Principal mutations (c.) | Sanger validation/ segregation | Mutation, as predicted by (among SIFT, PPheM2 & Mutation Taster) | Additional potentially pathogenic variants | Frequency of additional variants in EVSdb | Geographic origin | Inbred | BBS inclusion criteria |
|----------|--------------------------|--------------------------|-------------------------------|-------------------------------------------------|------------------------------------------|------------------------------------------|-----------------|--------|------------------------|
| A)       |                          |                          |                               |                                                  |                                          |                                          |                  |        |                        |
| BBS1     |                          |                          |                               |                                                  |                                          |                                          |                  |        |                        |
| AL047    | p.[R160Q]; [R160Q]       | c.[479G→A]; [479G→A]    | SV                            | PP, MT                                          |                                          |                                          | Turkey          | Yes    | (BBS1, BBS4) No        |
| P1       | p.? ; ?                  | c.[1473+4A→G]; [1473+4A→G] | SV + S                        | MT [Ex14 splice site altered]                   |                                          |                                          | Tunisia         | Yes    | *                       |
| AMK19    | p.? ; ?                  | c.[1110G→A]; [1110G→A]  | SV + S                        | MT [Ex11 splice site altered]                   |                                          |                                          | Tunisia         | Yes    | *                       |
| P12      | p.[A14Lfs*28]; [A14Lfs*28] | c.[39del]; [39del]       | SV                            | MT                                              |                                          |                                          | Tunisia         | Yes    | *                       |
| P11      | p.[R146*]; [R146*]       | c.[436C→T]; [436C→T]    | SV + S                        | MT                                              |                                          |                                          | Tunisia         | Yes    | *                       |
| BBS2     |                          |                          |                               |                                                  |                                          |                                          |                  |        |                        |
| P2       | p.? ; ?                  | c.[345+5G→A]; [345+5G→A] | SV + S                        | (Ex2 splice site altered)                       |                                          |                                          | Tunisia         | Yes    | *                       |
| AGL23    | p.[H65Ts*11]; [H65Ts*11] | c.[1992del]; [1992del]   | SV + S                        | MT                                              |                                          |                                          | India           | Yes    | (ALMS1, NPHP1, NPHP3, NPHP6, NPHP8) No |
| P7       | p.[R189*]; [R189*]       | c.[565C→T]; [565C→T]    | SV + S                        | MT                                              |                                          |                                          | Tunisia         | Yes    | *                       |
| AKX44    | p.[R272*]; [R272*]       | c.[814C→T]; [814C→T]    | SV                            | MT                                              |                                          |                                          | Algeria         | Yes    | (BBS7, BBS12, NPHP8)   |
| ALG76    | p.[L209P]; [L209P]       | c.[626T→C]; [626T→C]    | SV + S                        | S, PP, MT                                       |                                          |                                          | Tunisia         | Yes    | (BBS2, NPHP8) Yes      |
| BBS3/ARL6|                          |                          |                               |                                                  |                                          |                                          |                  |        |                        |
| ALG42    | p.[del Ex1-3]; [del Ex1-3] | c.[?_30..123+7|del]; [?_30..123+7|del] | SV + S                        | —                                               |                                          |                                          | Tunisia         | Yes    | (ALMS1, BBS3, BBS4, BBS9, BBS14, NPHP2, NPHP3) Yes |
| BBS4     |                          |                          |                               |                                                  |                                          |                                          |                  |        |                        |
| P3       | p.[del Ex4-5-6];[del Ex4-5-6] | c.[(157—7..405+7|del]; [(157—7..405+7|del]; (g.[33651..34496|del); (g.[33651..34496|del); | SV + S                        | —                                               |                                          |                                          | Tunisia         | Yes    | *                       |
| BBS5     |                          |                          |                               |                                                  |                                          |                                          |                  |        |                        |
| P13      | p.[L50R]; [L50R]         | c.[1497→G]; [1497→G]    | SV + S                        | S, PP, MT                                       |                                          |                                          | Tunisia         | Yes    | *                       |
| ALG5     | p.[R138H]; [R138H]       | c.[413G→A]; [413G→A]    | SV + S                        | S, PP, MT                                       |                                          |                                          | France          | No     | No                     |
| BBS6/MKKS|                          |                          |                               |                                                  |                                          |                                          |                  |        |                        |
| P10      | p.? ; ?                  | c.[1272+1G→A]; [1272+1G→A] | SV + S                        | MT [Ex5 splice site altered]                     |                                          |                                          | Tunisia         | Yes    | *                       |
| BBS8/TTC8|                          |                          |                               |                                                  |                                          |                                          |                  |        |                        |
| P14      | p.? ; ?                  | c.[329+1G→A]; [329+1G→A] | SV + S                        | MT [Ex4 splice site altered]                     |                                          |                                          | Tunisia         | Yes    | *                       |

Continued
### Table 3

Continued

| Patient# | Gene       | Potential mutations (p.) | Potential mutations (c.) | Frequency in EVSdb | Mutation, as predicted by (among SIFT, PolyPhen2 & Mutation T@ster) | Additional potentially pathogenic variants | Frequency of additional variants in EVSdb | Geographic origin | BBS inclusion criteria |
|----------|------------|--------------------------|--------------------------|--------------------|----------------------------------------------------------------|---------------------------------------------|--------------------------------------------|------------------|------------------------|
|          |            |                          |                          |                    |                                                                |                                             |                                            |                  |                        |
| BBS10    |            |                          |                          |                    |                                                                |                                             |                                            |                  |                        |
| P8       | [S396Lfs*6]; [S396Lfs*6] | c.[1171_1181dupGCATTATACC]; [1171_1181dupGCATTATACC] | SV + S                 | MT                 |                                                                |                                             |                                            | Tunisia          | Yes*                   |
| JSL      | [P710*]; [R955]   | [2119_2120delGT]; [285A→T] | SV + S                 | MT, S, PP, MT      |                                                                |                                             |                                            | NA               | No                     |
| BBH64    | [T483Nfs*8]; [?]  | c.[1448→1452delCTCAA]; [?] | S                     | MT                 |                                                                | CCDC28B: Ex2 splice site altered (F110F) ; [=]             | 2.07%                               | NA               | No                     |
| AMR64    | [L1415S]; [L1415S] | c.[1241→C]; [1241→C]      | SV                     | S, PP              |                                                                | CCDC28B: Ex2 splice site altered (F110F) ; [=]             | 2.07%                               | NA               | Yes                    |
| AKR68    | [L1415S]; [L1415S] | c.[1241→C]; [1241→C]      | SV                     | S, PP              |                                                                | CEP290/BBS14- p.[K1870Nfs*4]                        | NA                                   | NA               | Yes                    |
| ALMS1    | [E1141Rfs*9]; [E1141Rfs*9] | [3340del]; [3340del]       | SV                     | MT                 |                                                                |                                             |                                            | Turkey           | Yes (ALMS1, BBS14)     |
| AKO26    | [R3629*]; [R3629*] | [10885C→T]; [10885C→T]    | SV + S                 | MT                 |                                                                | WDFCP/BBS15: p.[V329M]; [=]                            | 0.13%, 1.65%                     | France           | Yes (ALMS1, BBS12)     |
| ALB64    | [S577*]; [S577*]  | [1730C→G]; [1730C→G]      | SV + S                 | MT                 |                                                                | IQCB1/NPHP5: p.[E481K]                              | 0.17%                               | Portugal         | Yes (BBS7, BBS12, BBS14, ALMS1) |

*Patients from consanguineous families, but not genotyped on affymetrix SNP arrays. In bold: previously reported mutations in other studies. Mutation prediction software equivalents: possibly/probably damaging (polyphen2), damaging (SIFT), disease causing (mutation taster). Ex, exon; S, SIFT; PP, PolyPhen2; MT, mutation T@ster. (T): The exact nature of the second heterogeneous mutation could not be identified by high-throughput sequencing. An abnormal loss of coverage is observed at the very end of BBS10. The exact nature of this apparently complex mutation is still under investigation by direct sequencing, but appears to involve an alu insertion coupled with a duplication/inversion. Amino acid conservation of non-reported missense mutations is shown in supplementary figure S4. SV: Sanger validation; S: segregation validated. Consanguinity was documented by clinicians, and in most cases, BBS patients from consanguineous families were genotyped on Affymetrix 250k SNP arrays. When so, BBS and other targeted genes located within homozygous regions are thus indicated in ( ).

**Methods**

| BBS10 |
|-------|
| Patient# | Principal mutations (p.) | Principal mutations (c.) | Sanger validation/ segregation | Mutation, as predicted by (among SIFT, PolyPhen2 & Mutation T@ster) | Additional potentially pathogenic variants | Frequency of additional variants in EVSdb | Geographic origin | Inbred | BBS inclusion criteria |
|--------|--------------------------|--------------------------|-------------------------------|----------------------------------------------------------------|---------------------------------------------|--------------------------------------------|------------------|-------|------------------------|
| P8     | p.[S396Lfs*6]; [S396Lfs*6] | c.[1171_1181dupGCATTATACC]; [1171_1181dupGCATTATACC] | SV + S                        | MT                                                            |                                             |                                            | Tunisia          | Yes*  | Yes                    |
| JSL    | [P710*]; [R955] | [2119_2120delGT]; [285A→T] | SV + S                        | MT, S, PP, MT                                                |                                             |                                            | NA               | No    | No                     |
| BBH64  | [T483Nfs*8]; [?] | c.[1448→1452delCTCAA]; [?] | S                             | MT                                                            | CCDC28B: Ex2 splice site altered (F110F) ; [=] | 2.07%                               | NA               | No    | No                     |
| AMR64  | [L1415S]; [L1415S] | c.[1241→C]; [1241→C] | SV                            | S, PP                                                        |                                             |                                            | 2.07%                      | NA    | Yes                    |
| AKR68  | [L1415S]; [L1415S] | c.[1241→C]; [1241→C] | SV                            | S, PP                                                        |                                             |                                             | NA               | No    | Yes                    |

**BBS10**

| Patient# | Gene       | Potential mutations (p.) | Potential mutations (c.) | Frequency in EVSdb | Mutation, as predicted by (among SIFT, PolyPhen2 & mutation T@ster) | Geographic origin | Consanguinity | BBS inclusion criteria |
|----------|------------|--------------------------|--------------------------|--------------------|----------------------------------------------------------------|------------------|---------------|------------------------|
|          |            |                          |                          |                    |                                                                |                  |               |                        |
| B)       |            |                          |                          |                    |                                                                |                  |               |                        |
| AHR2     | BBS3/ARL6  | p.[D179N]; [=]           | c.[535G→A]; [=]          | NF                 | MT (Ex8 splice site altered)                                    | Reunion          | Yes. BBS3, JBT3S | No                     |
| AY97     | BBS9       | p.[P641S]; [=]           | c.[1921C→T]; [=]         | NF                 | S, PP, MT                                                      | NA               | Yes. BBS5, ALMS1, NPHP12, NPHP5, BBS14 | No                     |
| AKE98    | ALMS1      | p.[P2679L]; [P2679L]     | c.[8036C→T]; [8036C→T]   | NF                 | S, P                                                           | Melanesia        | Yes. BBS12, BBS9, CEP290, NPHP1, MKS1 | Yes                     |
| NPHP4    |            |                          |                          |                    |                                                                |                  |               |                        |
| NPHP2    |            |                          |                          |                    |                                                                |                  |               |                        |
| NPHP3    |            |                          |                          |                    |                                                                |                  |               |                        |
| INVS/NPHP2 |            |                          |                          |                    |                                                                |                  |               |                        |
| AIX45    | TMEM67/ NPHP11 | p.[V486C]; [=]       | c.[1457A→G]; [=]          | 0.01%              | P, MT                                                          | NA               | Yes. JBT3S | Yes                     |
| AMA28    | MKKS/BBS6  | p.[A242S]; [=]           | c.[724G→T]; [=]           | 0.58%              | S, P                                                           | France           | No                   | Yes                     |
| NPHP1    |            |                          |                          |                    |                                                                |                  |               |                        |
| AMD77    | TTC21B/ NPHP12 | p.[R863W]; [=]        | c.[2587C→T]; [=]          | 0.23%              | S, P, MT                                                       | NA               | Yes                   | No                      |
| AHL86    | AHI1/JBTS3 | p.[R830S]; [=]           | c.[2481C→T]; [=]         | 2.83%              | S, P, MT                                                       | NA               | No                    | No                      |
| ALA60    |            |                          |                          |                    |                                                                |                  |               | No                      |
| AKT96    |            |                          |                          |                    |                                                                |                  |               | No                      |
| AMT10    |            |                          |                          |                    |                                                                |                  |               | No                      |
| AIF95    |            |                          |                          |                    |                                                                |                  |               | No                      |
| AKK32    |            |                          |                          |                    |                                                                |                  |               | No                      |

*Patients from consanguineous families, but not genotyped on affymetrix SNP arrays. In bold: previously reported mutations in other studies. Mutation prediction software equivalents: possibly/probably damaging (polyphen2), damaging (SIFT), disease causing (mutation taster). Ex, exon; S, SIFT; PP, PolyPhen2; MT, mutation T@ster. (T): The exact nature of the second heterogeneous mutation could not be identified by high-throughput sequencing. An abnormal loss of coverage is observed at the very end of BBS10. The exact nature of this apparently complex mutation is still under investigation by direct sequencing, but appears to involve an alu insertion coupled with a duplication/inversion. Amino acid conservation of non-reported missense mutations is shown in supplementary figure S4. SV: Sanger validation; S: segregation validated. Consanguinity was documented by clinicians, and in most cases, BBS patients from consanguineous families were genotyped on Affymetrix 250k SNP arrays. When so, BBS and other targeted genes located within homozygous regions are thus indicated in ( ).
consanguineous family, while several other mutations were identified in recessive forms of limb girdle muscular dystrophy.24

One noteworthy result is the finding of homozygous truncating ALMS1 mutations in 3/38 patients (AIA84, AKO26, ALB64; 7.9%). In particular, the nonsense found in AKO26 patient p.R3629* seems to be a recurrent ALMS1 mutation, since already reported in five other ALMS patients.25–27 The phenotypic overlap between BBS and ALMS seems to be larger than previously thought, as recently suggested with examples of Alström patients with mutations in BBS genes,5 and the reverse situation, such as in our study, of ALMS1 mutations in patient with suspected BBS.3

Lastly, no clearly pathogenic mutation was found in any NPHP or JBTS genes in the cohort.

Correlation between mutation detection efficiency and clinical phenotype
Comparison of clinical phenotype between patients with two clearly pathogenic mutated alleles (n=26) and those with either a single possible pathogenic variant or no suspicious variant detected (n=12) showed a clear correlation between the number of major BBS clinical features and the probability of detecting two BBS mutated alleles in patients (figure 3). Biallelic mutations were detected in 81% (CI (60% to 92%)) of patients meeting BBS inclusion criteria. In particular, in Tunisian patients recruited upon strict clinical criteria, mutations were found in 11/11 cases and in seven different BBS genes, ruling out a potential founder effect. On the contrary, for some of the 12 patients without clear mutations, BBS was only one suspected diagnosis among others. Furthermore, our initial selection of patients without recurrent mutations in BBS1 or BBS10, and without any mutation in BBS12 may have enriched our cohort in patients with non-typical BBS phenotypes. The current widely quoted estimation that known BBS genes account for only 70–75% of the total mutation load in BBS patients may thus be underestimated if considering only patients with strictly defined BBS phenotype.

The distribution of BBS inclusion features appears different between patients with two BBS mutations, two ALMS1 mutations or no biallelic mutation identified (table 4). Patients with no detected mutation presented with significantly less polydactyly, a major BBS clinical sign: only 25% versus 70% in patients with detected BBS mutations (p=0.029*). The other clinical features seem to follow the trend of classical BBS patients.

Regarding ALMS1-mutated probands, 2/3 had been sent for suspected BBS (Prader-Willi or ALMS) were also considered for AIA84 and AKO26, respectively) and satisfied BBS diagnostic criteria; the last one (ALB64) was addressed for syndromic retinal dystrophy (table 4).5 AIA84 presents a classical BBS with retinal dystrophy, obesity, cognitive defects, hypogonadism and brachydactyly. AKO26 presents an atypical BBS with the same features along with abnormal severe deafness, specific for ALMS. Lastly, ALB64 presents a typical ALMS with severe deafness and retinal dystrophy. None of them presented with polydactyly. As previously suggested,3 both, the absence of polydactyly and the prevalence of deafness in ALMS1-mutated patients, are keys for genotype-phenotype discrimination between ALMS and BBS mutated patients.

Assessment of oligogenism in BBS
The presence and potential effect of triallelism or oligogenism in BBS has been widely discussed and appears controversial (18 20 29 vs30–32). In our approach, the simultaneous sequencing of all 16 BBS genes, and of 14 other genes involved in overlapping ciliopathies, allows the systematic detection of most additional potentially pathogenic variants in those genes and, consequently, an unbiased assessment of oligogenism.

Out of the 52 patients analysed, we found only one heterozygous truncating mutation (p.K1870Nfs*4) as a third allele in BBS14/CEP290 in patient AKR65 who carries a pathogenic missense mutation in BBS10. Such a frequency is in fact in the range of what to expect by chance. We previously calculated the probability of carrying a true BBS-pathogenic mutation to be about 1:50.31 Since we also included in our design other ciliopathy genes, the probability to carry a pathogenic mutation in one of the 30 genes is rather between 1:20 and 1:30 (calculation based on disease incidence and reported contributions of targeted genes in the mutation load). Potentially, pathogenic heterozygous missenses (not previously reported in patients or in the EVSdb) were also found in eight patients (three of the proof-of-principle, table S2; five of the ‘unknown’ cohort, table 3). Such variants might act as modifiers, but it is unlikely that they are required for full expression of a classical BBS phenotype. Conversely, in some patients where a single clearly pathogenic mutation was found, variants in other genes of the same pathway (especially those encoding proteins of the same complex, such as the BSosome or the BBS-chaperonin complex)33 34 might contribute to the disease state in a digenic mode of inheritance proposed in few BBS families.20 35 36

Potential case of triallelism is illustrated by patient AIZ62, who is compound heterozygous or a nonsense p.E191* and a missense p.A242S in BBS6/MKKS. The pathogenicity of A242S variant has
that it affects BBS6/MKKS function and suggested a dominant negative effect.\textsuperscript{40} EVSdb allows to infer its frequency at 0.59\% (CI (0.45 to 0.80\%)), higher than the most frequent BBS mutation, M390R in BBS1 (0.24\%, CI (0.15 to 0.39\%)). A242S cannot thus be a highly penetrant mutation since it should then be found more frequently in patients than the M390R mutation, which is not the case. In patient AIZ62, a third heterozygous variant was identified in BBS12 (p.Q620R, residue conserved in mammals, but not in more distant vertebrates) thus affecting another subunit of the BBS-chaperonin complex.\textsuperscript{34} We suggest that A242S is a hypomorphic allele that may lead to a phenotype when in trans, with a complete null mutation, and could be further potentiated by a hypomorphic allele affecting another subunit of the same complex. Segregation analysis in AIZ62 family could not be performed to test this hypothesis.

Lastly, we looked in our cohort for the allelic frequency of the previously proposed BBS-modifier variant c.330C\textsubscript{T} in CCDC28B/MGC1203,\textsuperscript{15} and found a frequency of 5.85\% (CI (1.56 to 9.47\%)), which is not significantly different (p=0.17) from the 2.07\% (CI (1.76 to 2.45\%)) observed in EVSdb.

CONCLUSIONS

The extensive non-allelic genetic heterogeneity of Bardet-Biedl syndrome has been a major problem for molecular diagnostic and genetic counselling applications. Various strategies have been proposed in recent years to optimise mutation detection,\textsuperscript{5,10–12} but have either low sensitivity or are too time consuming and expensive in diagnostic settings. This problem is shared by other disease entities, such as cardiomyopathies, hearing loss, Usher syndrome or Charcot-Marie tooth neuropathies. Those NGS-based alternative strategies can be divided in whole genome/exome sequencing,\textsuperscript{11} which would be selected RNA sequencing (enriched in cognate gene transcripts).\textsuperscript{44} Conclusions are more exhaustive, in terms of gene coverage in current implementations, 20\%–50\% of targeted exons are not sufficiently covered for diagnostic accuracy, that is, to ensure low rates of false-positive/false-negative findings and reliable detection of heterozygous mutations or exon deletions. Moreover, the informatics resources needed for exome/genome sequencing data analysis and storage are considerably more important than for targeted sequencing, and can often be a limitation.

This strategy, however, presents some limited pitfalls. Few protein coding regions were not well covered, either because of failure in bait design (presence of repeat motifs) or poor capture efficiency (mostly GC-rich sequences and first exons).

Our protocol with initial barcoding of libraries followed by capture on pooled samples may be cost effective, but at present, limits the capture efficiency and needs further optimisation, especially if applied to larger gene sets. Finally, like for all targeted exon strategies, deep intronic mutations will be missed. The alternative to targeting entire genes would still miss a high proportion of intronic regions containing repetitive sequences, and would also disproportionately increase the number of rare variants to analyse with splice-prediction bioinformatic tools that are currently not highly reliable. For genes expressed in leucocytes or fibroblasts, another alternative would be selected RNA sequencing (enriched in cognate gene transcripts).

While similar capture strategies have been recently developed for other diseases, most of them included a much smaller cohort, and reported only proof-of-principle analysis.\textsuperscript{47–49} With the exception of Walsh et al.\textsuperscript{50} Multiplex PCR approaches may have the potential of covering exons more exhaustively,\textsuperscript{49} given that primer design is more flexible than hybridisation bait tiling, but is limited to smaller gene panels than for exon capture. PCR pooling without barcoding has been used for BBS and NPHP,\textsuperscript{12,51} a strategy which may be cost effective for analysing large numbers of samples in epidemiological studies, but appears unsuited for diagnosis where a key preoccupation is to limit false-positive/negative rates.

Regarding the specific case of BBS, our study suggests that when strict clinical criteria are complied with, the frequency of detected mutations is higher than the generally quoted 70\% figure.\textsuperscript{7,11} There may thus be only few strict BBS genes remaining to be identified, especially considering that with most strategies, with the exception of whole-gene sequencing, one will miss deep intronic pathogenic mutations. Additional genes to be discovered may correspond rather to variant BBS-like phenotypes than to strictly defined BBS. Indeed, for patients with BBS16/SDCCAG8 mutations, genotype-phenotype analysis showed for the first time a clear departure from the typical BBS phenotype with absence of polydactyly and systemic and severe renal manifestations (usually present in only 50\% of BBS patients).\textsuperscript{7,8} Our finding of ALMS1 mutations in three patients

### Table 4 Report of major BBS clinical features in the 38 patients without previously known molecular diagnosis, with or without detected mutations

| Retinitis pigmentosa | Obesity | Polydactyly | Brachydactyly | Hypogonadism | Cystic kidney | Mild intellectual disabilities |
|----------------------|---------|------------|--------------|--------------|--------------|-----------------------------|
| 2 BBS mutations      | 95.65\% (22/23) | 86.96\% (20/23) | 69.57\% (16/23) | 43.48\% (10/23) | 39.13\% (9/23) | 21.74\% (5/23) | 73.91\% (17/23) |
| 2 ALMS1 mutations    | 3/3     | 1/3 (AI48A) | 0/2          | 2/3 (AI48A, AK026) | 2/3 (AI48A, AK026) | 1/3 (ALB64) | 2/3 (AI48A, AK026) |
| 0 or 1 mutation      | 81.82\% (9/11) | 81.82\% (8/11) | 25.00\% (3/12) | 33.33\% (4/12) | 41.67\% (5/12) | 50\% (5/10) | 81.82\% (8/11) |

ALMS1 patients: AI48A was addressed to Strasbourg Diagnostic Laboratory for Bardet-Biedl or Prader-Willi syndromes, AK026 for suggestive BBS or ALMS with abnormal cognitive deficits and ALB64 for a syndromic retinal dystrophy or suggested ALMS. Patients with a proposed Alström syndrome presented with early deafness: at 5 (ALB64) or 6 (AKO26) years of age.
confirms the major clinical overlap with BBS. Finally, we found no evidence for triallelism in our cohort of BBS patients.

**Author affiliations**

1. Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR7104, INSERM U964, Université de Strasbourg, Illkirch, France
2. Laboratory of Human Genetics, University of Medicine of Tunis, Tunisia
3. Bioinformatics Platform, IGBMC, Illkirch, France
4. Laboratoire de Génétique Médicale EA3949 Inserm, Université de Strasbourg, Strasbourg, France
5. Laboratoire de Diagnostic Génétiques, Hôpitaux Universitaires de Strasbourg, Strasbourg, France
6. Instituto de Genética Médica, Universita’ Cattolica, Roma, Italy
7. CHU Bordeaux, University of Bordeaux, Department of Medical Genetics, Bordeaux, France
8. Service de Génétique Clinique, Hôpital Jeanne de Flandre, CHRU de Lille, Lille, France
9. Service de Cytogénétique Constitutionnelle, Hospices Civils de Lyon, CBPE, Bron Cedex, France
10. Department of Genetics, INSERM U676, Assistance Publique Hôpitaux de Paris (AP-HP), Robert Debré University Hospital, Paris, France
11. Service de Génétique Médicale, Centre de Référence pour les Affections Rares en Génétique Ophthalmologique (CARGO), Hôpitaux Universitaires de Strasbourg, Strasbourg, France
12. Chaire de Génétique Humaine, Collège de France, Illkirch, France

**Acknowledgements**

We thank Ngoc-Hoan Nguyen for his help in the bioinformatics setup, and Céline Pizot for the development of VaRank. We warmly thank Gérardine Greff, Anne-Sophie Jaeger, Manuela Antin, Elisabeth Scherrer, Serge Vacic and Muriel Philipp for their technical assistance. Lastly, we wish to thank all patients and families included in this study, and our clinician colleagues who addressed patients for diagnostic analysis and provided clinical information, in particular: Drs P Abou-Jaoudé, C Baumann, H Flodrops, T Frèbourg, D Genevieve, A Goldenberg, B Leheup, P Parent, P Pettigean, C Poitou-Bernert and S Taque.

**Contributors**

CR, JLJ and JM designed the study; CR, OM, CS, BJ performed experiments; CR, SLG, VG, JM performed the bioinformatics studies; CR, SLG, OM, VG, CS, JLJ and JM collected and analysed data; OM, CS, MCV, PC, DL, IO, FP, MT, AV, HBC and HD provided DNA samples and clinical information; CR, JLJ and JM wrote the manuscript; MCV, BJ, HBC, HD, JLJ and JM provided technical support, conceptual advice and project coordination.

**Funding**

This work was partially supported by a grant from Agence de Biomédecine to JLJ and JM, by funds from APLM and by the Association Française contre les Myopathies (AFM) thanks to its support to the IGBMC sequencing platform.

**Competition interests**

None.

**Patient consent**

For children, consent was signed by the parents.

**Provenance and peer review**

Not commissioned; externally peer reviewed.

**REFERENCES**

1. Zhanghoul NA, Katsanis N. Mechanistic insights into Bardet-Biedl syndrome, a model ciliopathy. J Clin Invest 2009;119:428–37.

2. Hildebrandt F, Benzing T, Katsanis N. Cilopathies. N Engl J Med 2011;364:1533–43.

3. Alliferis K, Helle S, Gyapay G, Duchateau S, Stoetzel C, Mandel JL, Dollfus H. Differentiating diastrom from Bardet-Biedl syndrome (BBS) using systematic ciliopathy genes sequencing. Ophthalmic Genet 2011;32:18–37.

4. Karmous-Benayli H, Martinovic J, Gubier MC, Sirot Y, Checl C, Ouzi C, Auge J, Brahim N, Etchever Ay, Drett E, Escolупat C, Audollent S, Godefroye G, Gonzales M, Tantart J, Logart P, Jobert M, Gallard C, Jeanne-Pasque C, Delezode AL, Peters GD, Dressel G, Simon-Bouy B, Dollfus H, Le Memer M, Munchin A, Encha-Razavi F, Vecsemans M, Attic-Bitch T. Antenatal presentation of Bardet-Biedl syndrome may mimic Meckel syndrome. Am J Hum Genet 2005;76:493–504.

5. Pereira I, Hoskins BE, Marshall JD, Collin GB, Naggert JK, Finelino-Salles O, Otmam EA, Katsanis N, Vakevrde C, Beales PL. Arrayed primer extension technology: simplified mutation detection in Bardet-Biedl and Alstrom syndrome. Eur J Hum Genet 2011;19:485–8.

6. Beales PL, Eicjobu N, Woolf A, Parker D, Finter FA. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. J Med Genet 1999;36:47–46.

7. Deveault C, Billsingley D, Duncan JS, Ben J, Theal R, Vincent A, Fieggen KJ, Gerth C, Nordeti T, Traboubl E, Fishman GA, Keon E. BBS genotype-phenotype assessment of a multicentric patient cohort calls for a revision of the disease definition. Hum Mol 2011;20:310–19.

8. Schaefer E, Zaloszy C, Lauer J, Durand M, Stutzmann F, Perdomo-Trujillo Y, Redin C, Bennouna Greene CE, Toutain A, Perrin L, Gerard Y, Caillard S, Bei X, Lewis RA, Chiastmann D, Lzech JS, Krbs M, Mutter C, Muller J, Stoetzel C, Fischbach M, Marion V, Katsanis N, Dollfus H. Mutations in SDC2/CNH/P10 cause Bardet-Biedl syndrome and are associated with penetrant renal disease and absent polydactyly. Mol Sydromol 2011;1:273–81.

9. Marshall JD, Malpie F, Beck S, Barrett TG, Paisby CE. Clinical utility gene card for: Alstrom syndrome. Eur J Hum Genet 2011;19.

10. Billsingley D, Deveault C, Heon E. BBS mutational analysis: a strategic approach. J Med Genet 2011;48:3181–7.

11. Muller J, Stoetzel C, Vincent MC, Leitch CC, Laurier V, Danske JM, Helle S, Marion V, Bennouna Greene V, Vicarie S, Megabame A, Kaplan J, Drouin-Graud J, Hamidiri M, Sigaudy S, Francarret C, Roane J, Bitoun P, Goldenberg A, Philip N, Odet S, Green J, Cosse M, Davis EE, Katsanis N, Bonneau D, Verloes A, Poch C, Mandel JL, Dollfus H. Identification of 28 novel mutations in the Bardet-Biedl syndrome genes: the burden of private mutations in an extensively heterogeneous disease. Hum Mol 2010;19:583–93.

12. Janssen S, Ramaswamy G, Davis EE, Hur J, Anik R, Kasunski JM, Van Der Kraak L, Allen SJ, Beales PL, Katsanis N, Otto EA, Hildebrandt F. Mutation analysis in Bardet-Biedl syndrome by DNA pooling and massively parallel sequencing in 105 individuals. Hum Mol 2011;19:79–90.

13. Medvedev P, Stanciute M, Brudno M. Computational methods for discovering structural variation with next-generation sequencing. Nat Methods 2009;6(11 Suppl):S13–20.

14. Nord AS, Lee M, King MC, Walsh T. Accurate and exact CNV identification from targeted high-throughput sequence data. BMC Genomics 2011;12:184.

15. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106.

16. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Methodological) 1995;57:289–300.

17. Simms RA, Hynes AM, Ely L, Sayer JA. Nephronophthisis: an genetically diverse ciliopathy. J Int Nept Genet 2011;57:717–26.

18. Badano JL, Leitch CC, Ansley SJ, May-Simera H, Lawson S, Lewis RA, Beales PL, Dietz HC, Fisher S, Katsanis N. Dissection of epistasis in oligogenic Bardet-Biedl syndrome. Nature 2006;439:326–30.

19. Mykytyn K, Braun T, Camii R, Haider NB, Seary CC, Shastri M, Beck G, Wright AF, lammarcane A, Edebeur K, Riese R, Bald A, Baas-Rothchild A, Gorman SW, Dahn OM, Jacobson CG, Sasavant T, Stone EM, Sheffield VC. Identification of the gene that, when mutated, causes the human obesity syndrome BBS4. Hum Mol 2005;14:3265–70.

20. Katsanis N, Ansley SJ, Badano JL, Eichers ER, Lewis RA, Hoskins BE, Scambler PJ, Davidson WS, Beales PL, Lupsik JR. Triallelic inheritance in Bardet-Biedl syndrome: a Mendelian recessive disorder. Science 2001;293:2256–9.

21. Leitch CC, Zhanghoul NA, Davis EE, Stoetzel C, Diaz-Fort A, Rix S, Addled Y, Lewis RB, Katsanis N, Yeadal WB, Banin E, Dollfus H, Beales PL, Badano JL, Katsanis N. Hypomorphic mutations in syndromic encephalocoele genes are associated with Bardet-Biedl syndrome. Nat Genet 2008;40:443–8.

22. Kim SK, Shindo A, Park T-J, Oh EC, Ghosh S, Gray R, Lewis RA, Johnson CA, Attie-Bitacht T, Katsanis N, Wallfing JB. Planar cell polarity acts through septins to control collective cell movement and ciliogenesis. Science 2001;293:337–40.

23. Otto EA, Hurd TW, Arin R, Chaki M, Zhou W, Stoetzel C, Patel SB, Levy S, Ghosh AK, Murga-Zamallowa CA, van Reeuwijk L, Betterbee SJ, Sang L, Giles RL, Liu Q, Cene KL, Estrada-Cuzzaco A, Collins RW, Mclaughin HM, Held S, Katsanis JM, Ramaswamy G, Conte J, Lopez I, Washburn J, Macdonald J, Hu J, Yamashita Y, Meher EA, Gun-Woodford L, Neumann HP, Oomenberg R, Koenekoop RK, Bergman C, Beiss X, Lewis RA, Katsanis N, Lopes V, Rivas LA, Williams DS, Lyons RH, Dang CV, Brito DA, Dias MB, Zhang X, Cawalcalo JD, Nurnberg G, Nummon N, Pierce EA, Jackson PK, Antignac C, Sauner S, Roepman H, Dollfus H, Khanna H, Hildebrandt F.
Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. Nat Genet 2010;42:340–50.

24. Schoser BG, Frost P, Engel AG, Kutzny U, Lochmüller H, Wrognemann K. Commonality of TBM and X-linked hydrocephalus in causing sarcotubular myopathy and LGMD2H. Ann Neurol 2005;57:591–5.

25. Bond J, Flintoff K, Higgins J, Scott S, Bennett C, Parsons J, Mannion J, Jafri H, Rashid Y, Barrow M, Trembath R, Woodford G, Rossa E, Lynch S, Shields J, Newbury-Ecob R, Falconer A, Holland P, Cockburn D, Karbani G, Malik S, Ahmed M, Roberts E, Taylor G, Woods GD. The importance of seeking ALMS1 mutations in infants with dilated cardiomyopathy. J Med Genet 2005;42:e10.

26. Marshall JD, Himman EG, Collin GB, Beck S, Cerquaeria R, Maffeii P, Milan G, Zhang W, Wilson DI, Hearn T, Tavare P, Vettor R, Veronese C, Martin M, So W, Nishina PM, Nagger JK. Spectrum of ALMS1 variants and evaluation of genotype-phenotype correlations in Alstrom syndrome. Hum Mutat 2007;28:1114–23.

27. Mintoa JA, Owen KR, Ricketts CJ, Crabtree N, Shaikh G, Ehtisham S, Porter JR, Carey C, Hodge D, Paisley W, March B, Barrett TG. Syndromic obesity and diabetes: changes in body composition with age and mutation analysis of ALMS1 in 12 United Kingdom kindreds with Alstrom syndrome. J Clin Endocrinol Metab 2006;91:3110–16.

28. Katsanis N. The oligogenic properties of Bardet-Biedl syndrome. Hum Mol Genet 2004;13:R65–71.

29. Katsanis N, Echers ER, Ansley SJ, Lewis RA, Kayserili H, Hoskins BE, Scambler PJ, Beales PL, Lupsir JR. BBS4 is a minor contributor to Bardet-Biedl syndrome and may also participate in triallelic inheritance. Am J Hum Genet 2002;71:22–9.

30. Hichri H, Stoetzel C, Laurier V, Caron S, Siaudy S, Sarra P, Hamel C, Martin-Coignard D, Gilles M, Leheup B, Holder M, Kaplan J, Bitoun P, Lacombe D, Verloes A, Bonhomme D, Perin-Schmitt F, Brandt C, Beasancon AF, Mandel JL, Dousse M, Doffins M. Testing for triallelism: analysis of six BBS genes in a Bardet-Biedl syndrome family cohort. Eur J Hum Genet 2005;13:807–18.

31. Laurier V, Stoetzel C, Muller J, Thibault C, Corbani S, Jalikh N, Saleyn L, Chouery E, Poch P, Licaire S, Dansie JM, Amati-Bonneau P, Bonneau D, Megabirane A, Mandel JL, Doffins H. Fifths of homoyzogosity mapping: an extended consanguineous Bardet-Biedl syndrome family with two mutant genes (BBS2, BBS10), three mutations, but no trialleliasm. Eur J Hum Genet 2006;14:1195–203.

32. Smaoula H, Chaabouni M, Serrouche Y, Kalief H, Li S, Mahfoudh N, Maazoul F, Kammoun H, Gondura N, Bouzai A, Nouari E, Ntrad R, Chaabouni H, Jaffrezik JF. Screening of the eight BBS genes in Tunisian families: no evidence of triallelic inheritance. Invest Ophthalmol Vis Sci 2006;47:3847–95.

33. Nachury MV, Loktev AV, Zhang Q, Westlake CJ, Peranen J, Merdes A, Slusarski DC. BBSome assembly. Cell 2007;129:1201–13.

34. Seo S, Baye LM, Schulz NP, Beck JS, Zhang D, Slusarski DC, Sheffield VC, Jackson PK. A core complex of BBS proteins cooperates with the GTPase Rab11 to promote ciliary membrane biogenesis. Cell 2007;127:1195–203.

35. de Ligt JB, Swisher EM, Lee MK, Casadei S, Thornton AM, Stray SM, Mandell RB, Biesecker LG. Mutation analysis of the MKKS gene in McKusick-Kaufman syndrome and may also participate in triallelic inheritance. Eur J Hum Genet 2004;11:561–7.

36. Zoghbi NA, Liu Y, Gerdes JM, Gassue C, Oh EC, Leitch CC, Bromberg Y, Binkley J, Leibl RL, Sidow A, Badano JL, Katsanis N. Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. Hum Mol Genet 2002;11:561–7.

37. Choi M, Scholl UI, Ji W, Liu J, Tikhonova IR, Zumbo P, Naray A, Bakkaloglu A, Ozen S, Sanjaj S, Sanjad S, Williams C, Farhi A, Mane S, Linton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 2009;106:19096–101.

38. Lupsir JR, Reid JG, Gonzaga Jauregui C, Rio Deiros D, Chen DC, Nazareth I, Banbridge D, Hing J, Ting C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Ehtisham S, Tom W, Fantin NJ, Munzy DM, Gibbs RA. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N Engl J Med 2010;362:1181–91.
APPENDIXES

SUPPLEMENTARY METHODS

DNA Samples & Quality Control
Genomic DNA was extracted automatically either from peripheral blood or saliva extracts following standard procedures (QIAcube, QIAGEN). Integrity of DNA samples was checked on agarose-gel. DNA concentration and quality were assessed at each quality-control step using either Nanodrop®, Qubit® (Life Technologies), or/and Bioanalyzer (either DNA1000 or High Sensitivity chips; Agilent).

Targeted Capture Design
Targeted genomic regions covered exons from all isoforms of 30 ciliopathy genes (Table 1). Exonic coordinates were extended to an additional 20 bp in flanking intronic sequences. Entire 5’ and 3’ untranslated regions were included for the BBS genes only. Lastly, the design aimed at targeting entire BBS1 and BBS4. Genomic coordinates were retrieved from the UCSC RefGene table (hg19/GRCh37)[1] and uploaded into eArray for bait design (Agilent). This resulted into 571 non-overlapping bait-covered regions, for a total of 189 kb. Some requested targeted regions were not bait-covered in eArray due to the presence of repetitive regions (supplementary Table S1).

Sequencing
72 bp single-read sequencing was performed on a GAIIx (Illumina) at the IGBMC sequencing platform. The first sequencing run was performed as follow: 2 lanes of 4 multiplexed DNA samples and 1 lane of 8 multiplexed DNA-samples. For the second run, 36 samples were equally distributed and sequenced in 3 lanes of 12 multiplexed DNA samples.

Read Mapping & Variant Calling
Image analysis and base calling were performed using the Illumina Pipeline Real-Time Analysis version 1.9. Reads were mapped to the human reference genome hg19/GRCh37 with BWA v.0.5.9 (Burrows-Wheeler
Aligner)[2] allowing the detection of small indels until 50-nucleotides long. Multi-mapped reads were discarded and only unique reads (with different start positions, sequences and strands) were kept. Variant calling was performed using Samtools pileup v0.1.7-4 (minimum mapping quality: 25, minimum depth: 3, minimum consensus quality: 20). Other additional filtering parameters were used either specific to single nucleotide variation (SNV; minimum SNP quality of 20, window size of 10 bp for filtering dense SNPs) or indel calling (minimum indel quality of 20, window size of 30 bp for filtering adjacent). Genomic annotation for SNV/indel was performed using Sequence Variant Analyzer v1.02.[3]

The mean coverage per nucleotide was computed both per sample and across samples subjected to the same protocol, before and after duplicates discarding (Table 2). Regions were considered as low-covered when the mean coverage per nucleotide was smaller than 10x (supplementary Table S2).

**Variant Filtering**

Initial filtering included the removal of variants present in dbSNP134 and validated by at least two methods. Subsequent variant ranking was performed using VaRank (manuscript in preparation), which uses Alamut (Interactive Biosoftware, Rouen, France) to collect for each variant genomic annotations and prediction effect at both nucleotide and protein levels. VaRank gathers variant-specific information such as genomic conservation (PhastCons)[4], potential functional effects of amino acid changes on the protein (SIFT and PolyPhen2)[5, 6] and splicing effect (Human Splicing Finder, MaxEntScan, NNsplice)[7-9]. Known mutations are highlighted by extracting either reported SNPs flagged as “probably-pathogenic”/“pathogenic” in the newly introduced “Clinical significance” field of dbSNP134. VaRank summarizes all such information to compute a score for each SNV/indel, generating as an output a list of ranked potentially pathogenic variants. Potential mutations are ranked starting from the one most likely to be pathogenic, as follow: known mutation, nonsense, frameshift, essential splice site (affecting conserved consensus motif), start loss, stop loss, missense, predicted splice site mutation (outside of consensus sites), in-frame indel, synonymous coding. VaRank produces such ranking according to the proposed mode of inheritance of the considered disease. Lastly, Mutation T@ster[10] was used independently to predict the consequence on the protein of any nucleotide substitution/insertion/deletion within the genomic sequence, in particular computing splicing prediction scores.
from ESE and NNSplice.[8, 11] When no obvious pathogenic variant was detected using the above strategy, we manually screened each targeted gene using IGV (Integrative Genome Viewer)[12] and/or lowered some filtering criteria. The global bioinformatic pipeline for variant filtering is summarized in Figure 1.

**Mutation validation**

Presence, copy number and segregation of identified variants via targeted high-throughput sequencing were confirmed by direct Sanger sequencing on a 3130xl Genetic Analyzer® (Applied Biosystems) using PCR amplicons from patients’ genomic DNA samples and their parents’ or/and probands’.
**SUPPLEMENTARY TABLES**

Supplementary Table S1: List of initially targeted regions but left bait-orphans after eArray design.

| Region | Start | End   | Length | Exon/intron # |
|--------|-------|-------|--------|---------------|
| **BBS1** |       |       |        |               |
|        | 66278294 | 66278399 | 105   | I1            |
|        | 66279072 | 66279585 | 513   | I3            |
|        | 66279941 | 66279987 | 46    | I3            |
|        | 66280180 | 66281229 | 1049  | I3            |
|        | 66281590 | 66281753 | 163   | I3            |
|        | 66282306 | 66282574 | 268   | I4            |
|        | 66283607 | 66283872 | 265   | I7            |
|        | 66284065 | 66284695 | 630   | I7            |
|        | 66284936 | 66286986 | 2050  | I7            |
|        | 66287299 | 66288618 | 1319  | I8            |
|        | 66289099 | 66289696 | 597   | I9            |
|        | 66289817 | 66290406 | 589   | I9            |
|        | 66290767 | 66290825 | 58    | I9            |
|        | 66291450 | 66293040 | 1590  | I11           |
|        | 66293161 | 66293464 | 303   | I11           |
|        | 66293849 | 66293942 | 93    | I11           |
|        | 66294562 | 66296533 | 1971  | I13           |
|        | 66296774 | 66297092 | 318   | I13           |
|        | 66298893 | 66299055 | 162   | I15           |
| **BBS3/ARL6** |       |       |        |               |
|        | 97517199 | 97517248 | 49    | E9 (3'UTR)    |
| **BBS4** |       |       |        |               |
|        | 72979482 | 72980400 | 918   | I1            |
|        | 72980761 | 72981366 | 605   | I1            |
|        | 72981799 | 72982439 | 640   | I1            |
|        | 72982584 | 72982614 | 30    | I1            |
|        | 72982922 | 72987199 | 4277  | I1            |
|        | 72987632 | 72987918 | 286   | I2            |
|        | 72988159 | 72989573 | 1414  | I2            |
|        | 72989718 | 72989909 | 191   | I2            |
|        | 72990582 | 72990988 | 406   | I2            |
|        | 72991373 | 72991718 | 345   | I2            |
|        | 72991839 | 72993879 | 2040  | I2            |
|        | 72994191 | 72994486 | 295   | I2            |
|        | 72994607 | 72994760 | 153   | I2            |
|        | 72995073 | 72995748 | 675   | I2            |
|        | 72995941 | 72996021 | 80    | I2            |
|        | 72996958 | 72997627 | 669   | I2            |
|        | 72997940 | 72998212 | 272   | I2            |
|        | 72998477 | 72998755 | 278   | I2            |
|        | 72999188 | 72999319 | 131   | I2            |
|        | 72999512 | 72999813 | 301   | I2            |
|        | 73000317 | 73001664 | 1347  | I2            |
|        | 73002145 | 73002359 | 214   | I3            |
|        | 73002514 | 73002905 | 391   | I3            |
|        | 73003146 | 73004068 | 922   | I3            |
|        | 73004213 | 73004504 | 291   | I3            |
|        | 73004745 | 73006185 | 1440  | I4            |
|        | 73006570 | 73006850 | 280   | I4            |
|        | 73008128 | 73008521 | 393   | I5            |
|        | 73009266 | 73009621 | 355   | I6            |
|        | 73009814 | 73010095 | 281   | I6            |
|        | 73010216 | 73010666 | 450   | I6            |
|        | 73011027 | 73011078 | 51    | I6            |
|       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 73011199 | 73011351 | 152   | I6    |
| 73011615 | 73011862 | 247   | I6    |
| 73012103 | 73013128 | 1025  | I6    |
| 73013489 | 73013692 | 203   | I6    |
| 73013837 | 73013994 | 157   | I6    |
| 73014354 | 73014525 | 171   | I6    |
| 73017223 | 73017267 | 44    | I8    |
| 73017388 | 73018743 | 1355  | I8    |
| 73020568 | 73020841 | 273   | I9    |
| 73021514 | 73021872 | 358   | I9    |
| 73022353 | 73022928 | 575   | I10   |
| 73023409 | 73023427 | 18    | I10   |
| 73024988 | 73026553 | 1565  | I12   |
| 73028594 | 73028816 | 222   | I14   |

**BBS5**

| 170361518 | 170363042 | 1524 | E12 (3'UTR) |

**BBS7**

| 122746484 | 122746674 | 190  | E19 (3'UTR) |

| 122745925 | 122745955 | 30   | E19 (3'UTR) |

**BBS10**

| 76738714 | 76738749 | 35   | E2 (3'UTR) |

| 76738382 | 76738401 | 19   | E2 (3'UTR) |

**BBS12**

| 123665322 | 123666099 | 777  | E2 (3'UTR) |

**MKS1/BBS13**

| 56282796 | 56282853 | 57   | E18 (3'UTR) |

**WDPCP**

| 63348535 | 63348791 | 256  | E18 (3'UTR) |

**ALMS1**

| 73612937 | 73613078 | 141 (60) | E1 (5'UTR+ coding) |

**NPHP3**

| 132441249 | 132441304 | 55   | E1 (5'UTR) |

| 132441011 | 132441056 | 45   | E1 (coding) |

Only 2 small coding regions (in grey) are concerned: first exon of ALMS1 (presence of poly-Glutamate and poly-Alanine sequences), and first exon of NPHP3.

Genomic positions are given according to the human reference genome hg19/GRCh37.

**In bold:** bait orphans regions but well-covered (>10x) due to the capture of flanking regions.

E: exon, I: intron. ( ): size of the protein coding region.
**Supplementary Table S2:** Spectrum of previously identified mutations within the *proof of principle* cohort. Mutations are described according to the latest nomenclature conventions described in HGVS.

| Patient # | Principal mutations (p.) | Principal mutations (c.) | # Reads with mutation/ Total reads | Additional potentially pathogenic variants | Frequency of additional variants in EVSdb |
|-----------|---------------------------|---------------------------|-----------------------------------|-------------------------------------------|------------------------------------------|
| **BBS1** |                           |                           |                                   |                                           |                                          |
| AHZ63     | p.? (alteration within Ex14 | c.[1473+4A>G]; [=]       | 69/136                            |                                           |                                          |
|           | splice site) ; [=]         |                           |                                   |                                           |                                          |
| AKH61     | p.[M390R] ; [M390R]        | c.[1169T>G] ; [1169T>G]   | 137/137                           |                                           |                                          |
| ALD6      | p.[R146*]; [del Ex8-9]     | c.[436C>T] ; [(592-?) (830+?)del] |                        |                                           |                                          |
| AMV5      | p.[del Ex10-11]; [=]       | c.[(831-?) (1110+?)del]; [=] |                        |                                           |                                          |
| **BBS2** |                           |                           |                                   |                                           |                                          |
| AGA99     | p.? (Ex2 splice site altered); ? (Ex2 splice site altered) | c.[118-1G>C] ; [118-1G>C] | 120/120                           |                                           |                                          |
| **BBS5** |                           |                           |                                   |                                           |                                          |
| AHX91b    | p.[H190Yfs*2]; [H190Yfs*2] | c.[572_594inv[ins567_568]] ; [572_594inv[ins567_568] |                        | BBS2: p.[Q508P]; [=] | NF                                          |
| **MKKS/BBS** |                      |                           |                                   |                                           |                                          |
| 6         |                           |                           |                                   |                                           |                                          |
| AIZ46     | p.[M1?]; [Y37C]            | c.[3G>A] ; [110A>G]       | 70/143 ; 62/141                   |                                           |                                          |
| AIZ62     | p.[E191*]; [A242S]         | c.[571G>T] ; [724G>T]    | 78/142 ; 77/142                   | BBS12: p.[Q620R]; [=] | 0.09%                                           |
| **BBS9** |                           |                           |                                   |                                           |                                          |
| ALB60     | p.[W285*]; [W285*]         | c.[854del] ; [854del]     | 116/129                           | CEP290/BBS14: p.[L906W]; [=] | 0.06%                                           |
| **BBS10** |                           |                           |                                   |                                           |                                          |
| ALS67     | p.[C91Lfs*5]; p.[K243fs*15] | c.[271_272insT] ; c.[728_731delAAGA] | 61/142 ; 61/142 | BBS7: p.[A443G]; [=] | NF                                          |
| AMA70     | p.[C91Lfs*5]; [G401*]      | c.[271_272insT] ; [1201G>T] | 59/135 ; 61/138                   |                                           |                                          |
| **BBS12** |                           |                           |                                   |                                           |                                          |
| Gene   | Mutation Details | Alleles | Coverage | Identification |
|--------|------------------|---------|----------|---------------|
| AKE12  | c.[1590_1591insAluYa5] | -       | AHI1/JBTS3: p.[R830W] | 2.83% |
| ALP79  | p.[A289P]; [L69F] | c.[865G>C]; [205C>T] | 70/144; 78/144; 76/143 | - |
| ALMS1  | p.[N2636Qfs*59] | c.[7904insC] | 126/143 | CCDC28B: p.? (F110F, Ex2 splice site altered) | 2.07% |

Parents of BBS patients: heterozygous mutations expected. In bold: previously reported mutations in other BBS patients.

Observation of a fall in coverage, but the identification of the exact nature of this complex rearrangement could not be detected directly. The exact nature was identified previously by Sanger sequencing. Ex: exon.

**Supplementary Table S3:** Weakly or non-covered regions (i.e. coverage ≤10x) in both initially targeted and effectively captured regions for the 38 patients with previously unknown molecular diagnosis.

|                    | Initially targeted regions | Captured (bait-covered) regions |
|--------------------|---------------------------|---------------------------------|
|                    | Size (bp) | ≤10x (%) | >10x (%) | Size (bp) | ≤10x (%) | >10x (%) |
| **Coding regions** | 81,710     | **0.96±0.68%** | **99.04±0.68%** | 81,582 | **0.63±0.68%** | **99.37±0.68%** |
| **Non coding regions (UTRs, introns)** | 118,609 | **33.74±0.68%** | **66.26±0.68%** | 106,948 | **0.74±0.75%** | **99.26±0.75%** |
| **Total**          | 200,319   | **20.37±0.67%** | **79.63±0.67%** | 188,530 | **0.69±0.71%** | **99.31±0.71** |
Supplementary Table S4: List of the most weakly covered regions in which coverage per nucleotide was ≤10x in some patients, with their relative size and GC content.

| Gene   | Start  | End    | Region Length (pb) | Exon/Intron #       | GC content (%) | Nucleotide Coverage (Mean) | Nucleotide Coverage (SD) |
|--------|--------|--------|--------------------|---------------------|----------------|---------------------------|--------------------------|
| ALMS1  | 73612867 | 73612936 | 70                 | E1 (5’UTR)          | 0.701          | 12.80                     | 5.17                     |
| BBS1   | 66296471 | 66297041 | 571                | E1 + I13            | 0.536          | 15.76                     | 5.29                     |
| BBS5   | 170361286 | 170361449 | 164               | E12 (3’UTR)         | 0.665          | 30.93                     | 11.04                    |
| CCDC28B| 32670764 | 32670883 | 120 (54)           | Exon/Intron #       | 0.664          | 16.90                     | 10.91                    |
| GLIS2  | 4382364 | 4382528 | 165 (90)           | E1 + I1             | 0.667          | 20.28                     | 8.58                     |
|         | 4383478 | 4383554 | 77 (43)            | E2 + I2             | 0.688          | 26.41                     | 12.17                    |
|         | 4386945 | 4387606 | 662 (580)          | E6 (coding + 3’UTR) | 0.702          | 17.52                     | 7.41                     |
| MKS1   | 10414712 | 10414885 | 174               | E1 (5’UTR)          | 0.695          | 24.89                     | 6.67                     |
| NEK8   | 27064992 | 27065096 | 105 (27)           | end E7 + I7         | 0.686          | 28.05                     | 11.94                    |
| NPHP1  | 110962579 | 110962662 | 84               | E1 (5’UTR)          | 0.774          | 21.87                     | 8.67                     |
| NPHP3  | 132440928 | 132441249 | 322 (272)         | Exon/Intron #       | 0.742          | 9.01                      | 4.29                     |
|         | 132440772 | 132440815 | 44 (9)           | end E1 + I1         | 0.75           | 23.15                     | 9.06                     |
| NPHP4  | 6027330 | 6027450 | 121 (65)           | E5                  | 0.69           | 19.57                     | 9.26                     |
|         | 6012708 | 6012772 | 65 (13)            | end E7 + I7         | 0.692          | 22.13                     | 8.97                     |
|         | 5969184 | 5969203 | 120 (62)           | E12                 | 0.675          | 13.44                     | 6.30                     |
|         | 5967110 | 5967185 | 76 (11)            | end E13 + I13       | 0.632          | 29.20                     | 11.42                    |
|         | 5934671 | 5934743 | 73 (46)            | end I21 + E22       | 0.651          | 20.74                     | 9.04                     |
|         | 5927921 | 5927998 | 78 (36)            | end I23 + E24       | 0.641          | 24.83                     | 10.71                    |
|         | 5925129 | 5925368 | 240 (172)          | E27                 | 0.629          | 21.91                     | 9.28                     |
|         | 5924369 | 5924608 | 240 (180)          | E28                 | 0.638          | 19.75                     | 10.24                    |
| TMEM216 | 61159934 | 61160212 | 279 (34)          | Exon/Intron #       | 0.692          | 19.18                     | 7.48                     |
| TRIM32  | 119449442 | 119449680 | 239               | E1 (5’UTR)          | 0.728          | 6.34                      | 3.13                     |
| TTC21B | 166810195 | 166810535 | 341 (20)         | Exon/Intron #       | 0.703          | 13.28                     | 5.71                     |
| TTC8   | 89290956 | 89291029 | 74                | E1 (5’UTR)          | 0.73           | 21.33                     | 11.42                    |
| Total  |        |        | 4810              | Mean                | 0.683          | 6.83                      | 0.05                     |

The most weakly covered regions are significantly enriched in GC content: 68.3±5% versus 40.2±10% across all targeted regions. Genomic positions are given using human reference genome hg19/GRCh37. E: exon, I: intron. Highlighted in grey: coding regions. ( ): size (bp) of the protein coding region.
**Supplementary Table S5:** Nature of detected variants in the total cohort of 52 patients.

|                  | Mean (/52 patients) | Included in dbSNPv134 | % in dbSNPv134 |
|------------------|---------------------|-----------------------|----------------|
| Frameshifts      | 0.3±0.5             | 0±0                   | 0.0%           |
| Splice           | 2±0.7               | 0.8±0.5               | 40.6%          |
| Non-sense        | 0.2±0.4             | 0±0                   | 0.0%           |
| Missenses        | 17.8±4.6            | 16.9±4.5              | 94.8%          |
| Inframe          | 0.8±0.4             | 0±0                   | 0.0%           |
| Synonymous       | 17.2±3.9            | 16.8±3.9              | 97.5%          |
| Intronic         | 134.9±11.8          | 95.6±10.5             | 70.8%          |
| **Total**        | **173.2±16.1**      | **130.1±14.9**        | **75.1%**      |
SUPPLEMENTARY FIGURES

Supplementary Figure S1: Various mutations types can be detected with our targeted high-throughput sequencing method. A) AKE12: abnormal loss of coverage and multiple mismatches in aligned reads suggest the real mutation, the homozygous insertion of an Alu element in BBS12; B) ADC44: insertion of a single C residue leading to an homozygous frameshift in ALMS1 (indicated by the vertical purple bars); C) AMA70: insertion of a single A residue causing an heterozygous frameshift in trans with a nonsense in BBS10.

Screening of mutations was done in the Integrative Genome Viewer. For each sample, the above curve in grey represents the depth of coverage (unique reads only), and the underneath grey arrows represent the unique 72 bp aligned reads. Colored (blue, red, green, brown) bars indicate nucleotide mismatches, purple vertical bars represent insertions/deletions. For some complex mutations (AKE12), the exact nature of the mutation was detected by Sanger sequencing.

A

| p1 6.1 | p15.2 | p14 | p11 | q13.1 | q21.1 | q24 | q26 | q28.2 | q31.21 | q32.1 | q33 | q35.1 |
|-------|-------|-----|-----|-------|-------|-----|-----|-------|--------|-------|-----|-------|
| 123 864 500 bp | 123 864 600 bp | 123 864 700 bp | 123 864 bp |
| 347 bp |

| p. 144 |
|-------|
| AKE12 |
| c.[1590_1591insAluYa5] ; [1590_1591insAluYa5] |

| p. 144 |
|-------|
| Negative Control |
| c. [=] ; [=] |

10/01/12
ADC44

c.[7904insC] ; [7904insC]
p.[N2636Qfs*59] ; [N2636Qfs*59]
AMA70

\[c.271_272\text{insT} ; [1201G>T] \]
\[p.[C91Lfs*5] ; [G401*] \]
Supplementary Figure S2: Log2 ratios between depths of coverage of normalized mean and corresponding patient highlight the presence of homozygous deletions. A) Homozygous deletion of BBS3 (exon #1, 2a, 2b, 3) in ALG42 patient. B) Homozygous deletion of BBS4 (exon #4, 5, 6) in P3 patient. Indicated positions are genomic positions given according to the human reference genome hg19/GRCh37.
Chromosome 15 (BBS4)
**Supplementary Figure S3**: Interpretation of suspicious homozygous variants in ALMS1 and INVS/NPHP2 in a single patient (AKE98), fulfilling BBS diagnostic inclusion criteria. 

A) Pedigree and segregation of variants in the Melanesian consanguineous family. B) Sequence conservation of Proline 2679 in ALMS1. P2679 residue is conserved in placental mammals but not in more distant vertebrate species. C) Genomic conservation in vertebrates of the longest INVS/NPHP2 isoform a with 5 additional exons (in red frame), of which the last one is potentially affected by the frameshift N1061Kfs*20. These alternative exons seem somewhat less conserved than the others in shorter isoforms, highly conserved until zebrafish.

The star indicates the position of aminoacid P2679. Alignments are displayed in Jalview and coloured using “clustalx” scheme.
The star indicates the last exon in which the frameshift occurs.
**Supplementary Figure S4:** Multiple sequence alignments covering the 5 pathogenic missense mutations identified in patients of our cohort. A) E224K in BBS1; B) L209P in BBS2; C) L50R in BBS5; D) R138H in BBS5; E) R95S in BBS10; F) L69F in BBS12; G) Q620R in BBS10. The only variant reported in EVS is Q620R (0.09%).

| A) BBS1: E224K | B) BBS2: L209P |
|----------------|----------------|
| Human          | Human          |
| Mouse          | Mouse          |
| Rat            | Rat            |
| Dog            | Dog            |
| Danio rerio    | Danio rerio    |
| Tetradodon     | Tetradodon     |
| Fly            | Fly            |
| Mosquito       | Mosquito       |
| C. briggsae    | C. elegans     |
| C, elegans     | C, elegans     |

| C) BBS5: L50R | D) BBS5: R138H |
|---------------|----------------|
| Human         | Human          |
| Mouse         | Mouse          |
| Rat           | Rat            |
| Dog           | Dog            |
| Chicken       | Chicken        |
| Frog          | Frog           |
| Danio rerio   | Danio rerio    |
| Tetradodon    | Tetradodon     |
| C. intestinalis | C. intestinalis |
| Fly           | Fly            |
| Mosquito      | Mosquito       |
| C. elegans    | C. elegans     |
| C. briggsae   | C. briggsae    |

| E) BBS10: R95S | F) BBS12: L69F |
|---------------|---------------|
| Human         | Human         |
| Mouse         | Mouse         |
| Rat           | Rat           |
| Dog           | Dog           |
| Chicken       | Chicken       |
| Frog          | Frog          |
| Danio rerio   | Danio rerio   |
| Tetradodon    | Tetradodon    |
| C. intestinalis | C. intestinalis |
| Fly           | Fly           |
| Mosquito      | Mosquito      |
| C. elegans    | C. elegans    |
| C. briggsae   | C. briggsae   |

| G) BBS10: Q620R | |
|----------------|---|
| Human          |   |
| Mouse          |   |
| Rat            |   |
| Chicken        |   |
| Frog           |   |
| Danio rerio    |   |
| Tetradodon     |   |
| Fugu           |   |

When possible, aligned sequences for each BBS protein cover metazoan organisms from *Caenorhabditis elegans* to *Homo sapiens*. Alignments are displayed in Jalview and coloured using “clustalx” colouring scheme. New missense mutations are highlighted with a star. In A) the diamond indicates the position of variant E234K (rs35520756), reported as probably pathogenic as a third allele. The observed mutation E224K is clearly more conserved than variant E234K, supporting its probable pathogenicity.

10/01/12
REFERENCES

1. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Giardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH, Meyer LR, Pohl A, Raney BJ, Rosenbloom KR, Smith KE, Haussler D, Kent WJ. The UCSC Genome Browser database: update 2011. *Nucleic Acids Res* 2011;39(Database issue):D876-82.

2. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26(5):589-95.

3. Ge D, Ruzzo EK, Shianna KV, He M, Pelak K, Heinzen EL, Need AC, Cirulli ET, Maia JM, Dickson SP, Zhu M, Singh A, Allen AS, Goldstein DB. SVA: software for annotating and visualizing sequenced human genomes. *Bioinformatics* 2011;27(14):1998-2000.

4. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;15(8):1034-50.

5. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7(4):248-9.

6. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4(7):1073-81.

7. Desmet FO, Hamroun D, Lalande M, Collo-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009;37(9):e67.

8. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol* 1997;4(3):311-23.

9. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11(2-3):377-94.

10. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010;7(8):575-6.

11. Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet* 2006;15(16):2490-508.

12. Robinson JT, Thorvalsdottir H, Winckler W, Guttmann M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol* 2011;29(1):24-6.