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, Habiba Bouhamed Chaabouni, Florence Petit, Marianne Till, Alain Verloes, Bernard Jost, 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), nephronophthisis (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 exon-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 2 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 hypogonadism. 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 BBS6.8

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 hypertriglyceridemia, 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

<table>
<thead>
<tr>
<th>Official gene symbol</th>
<th>BBS#</th>
<th>NPHP#</th>
<th>Other disease-related symbols</th>
<th># Exons in consensus</th>
<th># Total exons in all isoforms</th>
<th>Size of coding exons (bp)</th>
<th>Gene size (bp)</th>
<th>Targeted region size</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBS1*</td>
<td>BBS1</td>
<td>–</td>
<td>–</td>
<td>17</td>
<td>17</td>
<td>1782</td>
<td>22 966</td>
<td>23 285*</td>
</tr>
<tr>
<td>BBS2</td>
<td>BBS2</td>
<td>–</td>
<td>–</td>
<td>17</td>
<td>17</td>
<td>2166</td>
<td>35 750</td>
<td>3801</td>
</tr>
<tr>
<td>ARL6</td>
<td>BBS3</td>
<td>–</td>
<td>RP55</td>
<td>8</td>
<td>11</td>
<td>561</td>
<td>33 779</td>
<td>2123</td>
</tr>
<tr>
<td>BBS4*</td>
<td>BBS4</td>
<td>–</td>
<td>LCA</td>
<td>16</td>
<td>18</td>
<td>1560</td>
<td>52 292</td>
<td>52 611*</td>
</tr>
<tr>
<td>BBS5</td>
<td>BBS5</td>
<td>–</td>
<td>–</td>
<td>12</td>
<td>17</td>
<td>1026</td>
<td>27 160</td>
<td>3997</td>
</tr>
<tr>
<td>MKS</td>
<td>MKS</td>
<td>–</td>
<td>–</td>
<td>6</td>
<td>7</td>
<td>1713</td>
<td>29 034</td>
<td>3326</td>
</tr>
<tr>
<td>BBS7</td>
<td>BBS7</td>
<td>–</td>
<td>–</td>
<td>19</td>
<td>20</td>
<td>2148</td>
<td>46 008</td>
<td>5308</td>
</tr>
<tr>
<td>TTC8</td>
<td>TTC8</td>
<td>–</td>
<td>RP51</td>
<td>15</td>
<td>16</td>
<td>1518</td>
<td>53 358</td>
<td>3137</td>
</tr>
<tr>
<td>BBS9</td>
<td>BBS9</td>
<td>–</td>
<td>–</td>
<td>23</td>
<td>24</td>
<td>2559</td>
<td>476 529</td>
<td>5196</td>
</tr>
<tr>
<td>BBS10</td>
<td>BBS10</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>2</td>
<td>2172</td>
<td>3957</td>
<td>3941</td>
</tr>
<tr>
<td>TRIM32</td>
<td>TRIM32</td>
<td>BBS11</td>
<td>–</td>
<td>2</td>
<td>2</td>
<td>1962</td>
<td>13 999</td>
<td>4077</td>
</tr>
<tr>
<td>BBS12</td>
<td>BBS12</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>3</td>
<td>2133</td>
<td>12 242</td>
<td>3829</td>
</tr>
<tr>
<td>MKS1</td>
<td>MKS1</td>
<td>–</td>
<td>–</td>
<td>18</td>
<td>20</td>
<td>1680</td>
<td>14 170</td>
<td>3745</td>
</tr>
<tr>
<td>CEP290</td>
<td>CEP290</td>
<td>BBS14</td>
<td>NPHP6</td>
<td>54</td>
<td>55</td>
<td>7440</td>
<td>93 204</td>
<td>10 510</td>
</tr>
<tr>
<td>WDPIC</td>
<td>WDPIC</td>
<td>BBS15</td>
<td>–</td>
<td>18</td>
<td>19</td>
<td>2469</td>
<td>467 317</td>
<td>4405</td>
</tr>
<tr>
<td>SDCCA8</td>
<td>SDCCA8</td>
<td>BBS16</td>
<td>NPHP10</td>
<td>18</td>
<td>21</td>
<td>2141</td>
<td>244 087</td>
<td>3599</td>
</tr>
<tr>
<td>ALMS1</td>
<td>ALMS1</td>
<td>–</td>
<td>–</td>
<td>23</td>
<td>24</td>
<td>12 510</td>
<td>224 161</td>
<td>13 682</td>
</tr>
<tr>
<td>NPHP1</td>
<td>NPHP1</td>
<td>–</td>
<td>JBTS4; SLSN1</td>
<td>20</td>
<td>22</td>
<td>2202</td>
<td>81 726</td>
<td>3264</td>
</tr>
<tr>
<td>INVS</td>
<td>INVS</td>
<td>–</td>
<td>–</td>
<td>17</td>
<td>19</td>
<td>3198</td>
<td>201 916</td>
<td>4103</td>
</tr>
<tr>
<td>NPHP3</td>
<td>NPHP3</td>
<td>–</td>
<td>MKS7; RHPD</td>
<td>27</td>
<td>27</td>
<td>3993</td>
<td>41 823</td>
<td>5328</td>
</tr>
<tr>
<td>NPHP4</td>
<td>NPHP4</td>
<td>–</td>
<td>SLSN4</td>
<td>30</td>
<td>30</td>
<td>4281</td>
<td>129 662</td>
<td>5693</td>
</tr>
<tr>
<td>IODCB1</td>
<td>IODCB1</td>
<td>NPHP5</td>
<td>SLSN5</td>
<td>15</td>
<td>15</td>
<td>1797</td>
<td>65 317</td>
<td>2585</td>
</tr>
<tr>
<td>GLIS2</td>
<td>GLIS2</td>
<td>–</td>
<td>–</td>
<td>8</td>
<td>8</td>
<td>1575</td>
<td>73 74</td>
<td>2175</td>
</tr>
<tr>
<td>RPGRIP1L</td>
<td>RPGRIP1L</td>
<td>NPHP8</td>
<td>MKS5; JBTS7; CS</td>
<td>27</td>
<td>27</td>
<td>3708</td>
<td>103 954</td>
<td>5243</td>
</tr>
<tr>
<td>NEX8</td>
<td>NEX8</td>
<td>–</td>
<td>–</td>
<td>15</td>
<td>19</td>
<td>2078</td>
<td>13 953</td>
<td>3906</td>
</tr>
<tr>
<td>TMEM67</td>
<td>TMEM67</td>
<td>NPHP11</td>
<td>MKS3; JBTS6; CS</td>
<td>28</td>
<td>30</td>
<td>2745</td>
<td>64 389</td>
<td>4979</td>
</tr>
<tr>
<td>TTC21B</td>
<td>TTC21B</td>
<td>NPHP12</td>
<td>JBTS11; AD4</td>
<td>29</td>
<td>29</td>
<td>3951</td>
<td>79 894</td>
<td>5414</td>
</tr>
<tr>
<td>TMEM216</td>
<td>TMEM216</td>
<td>–</td>
<td>MKS2; JBTS2</td>
<td>5</td>
<td>5</td>
<td>447</td>
<td>6504</td>
<td>1795</td>
</tr>
<tr>
<td>AH1</td>
<td>AH1</td>
<td>–</td>
<td>JBTS3</td>
<td>27</td>
<td>29</td>
<td>3591</td>
<td>213 794</td>
<td>5175</td>
</tr>
<tr>
<td>CDDC28B</td>
<td>CDDC28B</td>
<td>–</td>
<td>–</td>
<td>6</td>
<td>6</td>
<td>603</td>
<td>4790</td>
<td>1079</td>
</tr>
<tr>
<td>Total</td>
<td>Total</td>
<td>483</td>
<td>545</td>
<td>79 781</td>
<td>2 865 109</td>
<td>200 319</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Sequence of the entire gene (coding/non-coding exons, introns) was targeted.

ALMS, Alström syndrome; ATD, 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, 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 ophthalmogenetic 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 without known mutations, and recruited in Strasbourg, had been initially screened for BBS1 and BBS10 recurrent mutations, plus the entire coding sequence of BBS12.

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 SPRIworks 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 Illuma 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 computed, 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 nephropathies genes (NPHP1-t2), since retinal degeneration can often be observed in this kidney-specific disease. Coding sequences of AH11/JBTS3, TMEM216/MKS2/JBTS2, and of the proposed BBS-modifier CCDC28B/AGC1203, 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 baits-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 M390R with no second mutation detected, and multiple reports of BBS4 exon deletions in patients. This choice was dictated by two observations: an apparent excess of patients heterozygous for the BBS1-recurrent mutation M390R 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, introns of BBS1 and BBS4, besides a small number of 3'UTRs, and only 128 bp of protein coding regions (within first exons of ALMS1 and NPHP3, table S1).

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, nonsense, 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 BBS12 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.

Figure 1 Global flowchart of the bioinformatic pipeline implemented for mutation detection. Software acronyms: BWA, Burrows-Wheeler Aligner; SVA, Sequence Variant Analyser; SIFT, Polyphen2, HSF, Human Splicing Finder; MaxEntScan, Maximum Entropy Scanning; NNSplice, Mutation Taster; IGV, Integrative Genome Viewer.
similar drop in coverage was observed for a second patient, AHX91, with another complex mutation detected previously by Sanger sequencing (insertion/inversion in \textit{BBS5}).

In this first experiment, we almost systematically reached the maximal theoretical coverage of 144x illustrated by a mean coverage of 127.6 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 78.617 (283.153x before discarding duplicate reads) with 91.4\pm6.4\% 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 duplicate filtering), with very few of them in a systematic way in other patients (table S4). This only concerned 0.63\% of protein coding regions, and mostly included intronic GC-rich sequences (GC content: 68.3\% vs 40.2\% across all targeted regions), or some first exons (tables S3 and S4).

Table 2 Sequencing statistics of both coverage (in captured regions) and capture efficiency

<table>
<thead>
<tr>
<th></th>
<th>Mean coverage (x)</th>
<th>Targeted regions with 5x coverage</th>
<th>Targeted regions with 40x coverage</th>
<th>Targeted regions with 80x coverage</th>
<th>Reads in targeted regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before filters</td>
<td>After filters</td>
<td>Before filters</td>
<td>After filters</td>
<td></td>
</tr>
<tr>
<td>Pool of 4*</td>
<td>2208±416</td>
<td>130±3</td>
<td>100.0±0.0</td>
<td>99.9±0.5</td>
<td>98.3±0.6</td>
</tr>
<tr>
<td>Pool of 8*</td>
<td>1024±151</td>
<td>125±3</td>
<td>99.9±0.3</td>
<td>99.7±0.3</td>
<td>96.9±0.9</td>
</tr>
<tr>
<td>Pool of 12†</td>
<td>283±153</td>
<td>78±17</td>
<td>98.6±0.4</td>
<td>91.4±0.4</td>
<td>87.7±28%</td>
</tr>
</tbody>
</table>

*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.

Figure 2 Detection of large deletions in three patients using a depth-of-coverage method. Black peaks: normalized depth of coverage from patients' DNA samples. Empty peaks: normalized mean depth of coverage across samples from the same sequencing lane. Grey squares: bait-covered regions. Black peaks: normalized depth of coverage from patients' DNA samples. Empty peaks: normalized mean depth of coverage across samples from the same sequencing lane. Grey squares: bait-covered regions. Highlighted squares: deleted regions. Gene representation: black squares: exons, dashed lines: introns. Genomic positions are given according to the human reference genome hg19/GRCh37. (A) Heterozygous deletion of BBS1 (exon 10, 11) in AMV5 patient. Corresponding Log2 ratios between both depths of coverage (normalised mean and AMV5 patient) further highlight the presence of the deletion. (B) Homozygous deletion of BBS3 (exon 1, 2a, 2b, 3) in ALG42 patient. (C) Homozygous deletion of BBS4 (exon 4, 5, 6) in P3 patient. (A and C): targeting intronic sequences allows restricting the deletion breakpoints. (B) and C: Log2 ratios between both depths of coverage (normalised mean and corresponding patients) could also allow detecting both deletions but are not shown (supplementary figure S2).
Variation 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: rs17936897, 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—\(^{20}\) 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.\(^{4,11,19}\)

Since we also targeted intronic sequences of BBS1 and BBS4, we were able to narrow the boundaries of subsequent deleted exons (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 1862–3841 bp (figure 2A, table S2). In patient F3, we could similarly reduce the assessed size of the BBS4 deletion from 4626–12975 bp based on exon positions down to 9876–10469 bp (figure 2C, table S3). Lastly, since the BBS3/ARL6 deletion in patient ALG42 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 exonic 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% (25/26) of mutated patients, coherently 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 ACL23, AKX44; table 3A).

Among the remaining 12 patients with no biallelic mutations identified (table 3B), one patient (AHR2) 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 framenshift in INVS/NPHP2 predicted to add 14 amino-acids to the C-terminus of the longer protein isoform, and a non-reported missense c2679L 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.\(^7,11\) 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%.\(^7\) 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 (\(\approx 20\%)\).\(^7\) 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 (BBS13–16) 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.\(^21\) Likewise, for BBS14/CEP290, a homozygous truncating mutation (p.E1903*) was found in a single BBS patient,\(^21\) 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/SDCCAG8.\(^22,23\) 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

<table>
<thead>
<tr>
<th>Patient#</th>
<th>Principal mutations (p.)</th>
<th>Principal mutations (c.)</th>
<th>Sanger validation/ segregation</th>
<th>Mutation, as predicted by (among SIFT, PPheM2 &amp; Mutation Taster)</th>
<th>Additional potentially pathogenic variants</th>
<th>Frequency of additional variants in EVSdb</th>
<th>Geographic origin</th>
<th>Inbred</th>
<th>BBS inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>A) BBS1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL047</td>
<td>p.[R160Q]; [R160Q]</td>
<td>c.[479G→A]; [479G→A]</td>
<td>SV</td>
<td>PP, MT</td>
<td></td>
<td></td>
<td>Turkey</td>
<td>Yes</td>
<td>(BBS1, BBS4) No</td>
</tr>
<tr>
<td>P1</td>
<td>p. ?: ?</td>
<td>c.[1473+4A→G]; [1473+4A→G]</td>
<td>SV + S</td>
<td>MT [Ex14 splice site altered]</td>
<td></td>
<td></td>
<td>Tunisia</td>
<td>Yes *</td>
<td>Yes</td>
</tr>
<tr>
<td>P9</td>
<td>p. ?: ?</td>
<td>c.[1110G→A]; [1110G→A]</td>
<td>SV + S</td>
<td>MT [Ex11 splice site altered]</td>
<td></td>
<td></td>
<td>Tunisia</td>
<td>Yes *</td>
<td>No</td>
</tr>
<tr>
<td>AMK19</td>
<td>p. ?: [M390R]</td>
<td>c.[951+1G→A]; [1169T→G]</td>
<td>SV + S</td>
<td>MT [Ex10 splice site altered]; S</td>
<td></td>
<td></td>
<td>Tunisia</td>
<td>Yes *</td>
<td>Yes</td>
</tr>
<tr>
<td>P12</td>
<td>p.[A14Lfs<em>28]; [A14Lfs</em>28]</td>
<td>c.[39del]; [39del]</td>
<td>SV + S</td>
<td>MT</td>
<td></td>
<td></td>
<td>Tunisia</td>
<td>Yes *</td>
<td>Yes</td>
</tr>
<tr>
<td>P11</td>
<td>p.[R146*]; [R146*]</td>
<td>c.[565C→T]; [565C→T]</td>
<td>SV + S</td>
<td>MT</td>
<td></td>
<td></td>
<td>Tunisia</td>
<td>Yes *</td>
<td>Yes</td>
</tr>
<tr>
<td>AIO57</td>
<td>p.[E224K]; [E224K]</td>
<td>c.[670G→A]; [670G→A]</td>
<td>SV</td>
<td>MT</td>
<td></td>
<td></td>
<td>Morocco</td>
<td>Yes</td>
<td>(BBS1, BBS3, NPHP5, JBTS2) No</td>
</tr>
<tr>
<td>BBS2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>p. ?: ?</td>
<td>c.[345+5G→A]; [345+5G→A]</td>
<td>SV + S</td>
<td>MT [Ex2 splice site altered]</td>
<td></td>
<td></td>
<td>Tunisia</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AGL23</td>
<td>p.[H655Ts<em>11]; [H655Ts</em>11]</td>
<td>c.[1992del]; [1992del]</td>
<td>SV + S</td>
<td>MT</td>
<td></td>
<td></td>
<td>India</td>
<td>Yes</td>
<td>(ALMS1, NPHP1, NPHP3, NPHP6, NPHP8) No</td>
</tr>
<tr>
<td>P7</td>
<td>p.[R189*]; [R189*]</td>
<td>c.[565C→T]; [565C→T]</td>
<td>SV + S</td>
<td>MT</td>
<td></td>
<td></td>
<td>Tunisia</td>
<td>Yes *</td>
<td>Yes</td>
</tr>
<tr>
<td>AKX44</td>
<td>p.[R272*]; [R272*]</td>
<td>c.[814C→T]; [814C→T]</td>
<td>SV</td>
<td>MT</td>
<td></td>
<td></td>
<td>Algeria</td>
<td>Yes</td>
<td>(BBS7, BBS12, NPHP8)</td>
</tr>
<tr>
<td>ALG76</td>
<td>p.[L.209P]; [L.209P]</td>
<td>c.[626T→C]; [626T→C]</td>
<td>SV + S</td>
<td>S, PP, MT</td>
<td></td>
<td></td>
<td>Turkey</td>
<td>Yes</td>
<td>(BBS2, NPHP8) Yes</td>
</tr>
<tr>
<td>ALG42</td>
<td>p.[del Ex1-3]; [del Ex1-3]</td>
<td>c.[?→-30;[123+7]del];[?→-30;[123+7]del]</td>
<td>SV + S</td>
<td>—</td>
<td></td>
<td></td>
<td>Turkey</td>
<td>Yes</td>
<td>(ALMS1, BBS1, BBS3, BBS4, BBS5, BBS9, BBS15, NPHP2, NPHP3) Yes</td>
</tr>
<tr>
<td>BBS3/ARL6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BBS4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BBS5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>p.[L50R]; [L50R]</td>
<td>c.[1497→G]; [1497→G]</td>
<td>SV + S</td>
<td>S, PP, MT</td>
<td></td>
<td></td>
<td>Tunisia</td>
<td>Yes *</td>
<td>Yes</td>
</tr>
<tr>
<td>ALG5</td>
<td>p.[R139H]; [R139H]</td>
<td>c.[413G→A]; [413G→A]</td>
<td>SV + S</td>
<td>S, PP, MT</td>
<td></td>
<td></td>
<td>France</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>BBS5/NKKS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10</td>
<td>p. ?: ?</td>
<td>c.[1272+1G→A]; [1272+1G→A]</td>
<td>SV + S</td>
<td>MT [Ex5 splice site altered]</td>
<td></td>
<td></td>
<td>Tunisia</td>
<td>Yes *</td>
<td>Yes</td>
</tr>
<tr>
<td>BBS5/TTC8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14</td>
<td>p. ?: ?</td>
<td>c.[329+1G→A]; [329+1G→A]</td>
<td>SV + S</td>
<td>MT [Ex4 splice site altered]</td>
<td></td>
<td></td>
<td>Tunisia</td>
<td>Yes *</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Continued
### Methods

Table 3 continued

<table>
<thead>
<tr>
<th>Patient#</th>
<th>Gene</th>
<th>Potential mutations (p.)</th>
<th>Potential mutations (c.)</th>
<th>Frequency in EVSdb</th>
<th>Mutation, as predicted by (among SIFT, PolyPhen2 &amp; Mutation T@ster)</th>
<th>Additional potentially pathogenic variants</th>
<th>Frequency of additional variants in EVSdb</th>
<th>Geographic origin</th>
<th>Inbred criteria</th>
<th>BBS inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>P8</td>
<td>BBS10</td>
<td>p. [396Lfs<em>6]; [S396Lfs</em>6];</td>
<td>c. [1171_1181dupGCATTATACCG]; [1171_1181dupGCATTATACCG];</td>
<td>SV + S</td>
<td>MT</td>
<td>Tunisia</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JSL</td>
<td>BBS10</td>
<td>[V707*]; [R95S];</td>
<td>c. [2119_2120delGT]; [285A → T]</td>
<td>SV + S</td>
<td>MT, S, PP, MT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BBH64</td>
<td>BBS10</td>
<td>[T483Nfs*8]; [7]</td>
<td>c. [1448→1452delCTGAA]; [?];</td>
<td>S</td>
<td>MT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMR64</td>
<td>BBS10</td>
<td>[L414S]; [L414S];</td>
<td>c. [1241T → C]; [1241T → C]</td>
<td>SV</td>
<td>S, PP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKR68</td>
<td>BBS10</td>
<td>[L414S]; [L414S];</td>
<td>c. [1241T → C]; [1241T → C]</td>
<td>SV</td>
<td>S, PP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIA84</td>
<td>ALMS1</td>
<td>[E1114Rfs<em>9]; [E1114Rfs</em>9];</td>
<td>c. [3343del]; [3343del]</td>
<td>SV</td>
<td>MT</td>
<td>Turkey</td>
<td>Yes</td>
<td>ALMS1, BBS14</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>AKO26</td>
<td>ALMS1</td>
<td>[R3629*]; [R3629*];</td>
<td>c. [10885C / T]; [10885C / T]</td>
<td>SV + S</td>
<td>MT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALB64</td>
<td>ALMS1</td>
<td>[S577*]; [S577*];</td>
<td>c. [1730C / G]; [1730C / G]</td>
<td>SV + S</td>
<td>MT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**BBS10**

- **P8**: p. [S396Lfs*6]; [S396Lfs*6]; c. [1171_1181dupGCATTATACCG]; [1171_1181dupGCATTATACCG]; SV + S; MT; Tunisia; Yes
- **JSL**: [V707*]; [R95S]; c. [2119_2120delGT]; [285A → T]; SV + S; MT, S, PP, MT
- **BBH64**: [T483Nfs*8]; [7]; c. [1448→1452delCTGAA]; [?]; S; MT
- **AMR64**: [L414S]; [L414S]; c. [1241T → C]; [1241T → C]; SV
- **AKR68**: [L414S]; [L414S]; c. [1241T → C]; [1241T → C]; SV

**ALMS1**

- **AIA84**: [E1114Rfs*9]; [E1114Rfs*9]; c. [3343del]; [3343del]; SV; MT; Turkey; Yes; ALMS1, BBS14
- **AKO26**: [R3629*]; [R3629*]; c. [10885C / T]; [10885C / T]; SV + S; MT

**Patient#**

- **Bi**: AHR2, AY67, AKE98
- **ALMS1**: AIA84, AKO26, ALB64

**Gene**

- **BBS3/ARL6**: p. [D179N]; [E1114Rfs*9];
- **BBS9**: p. [P641S]; [P641S];
- **ALMS1**: p. [P2679L]; [P2679L];
- **NPHP4**: p. [V1122G]; [V1122G];
- **NPHP2**: p. [A1959G]; [A1959G];
- **NPHP1**: p. [E569G]; [E569G];
- **TMEM67/NPHP11**: p. [Y486C]; [Y486C];
- **MKKS/BBS6**: p. [A2423S]; [A2423S];
- **NPHP1**: p. [E569G]; [E569G];
- **TTCC18/NPHP12**: p. [R863W]; [R863W];
- **AH11/JBTS3**: p. [R830W]; [R830W];
- **INVS/NPHP2**: p. [T122P]; [T122P];
- **NPHP1**: p. [E569G]; [E569G];
- **NPHP2**: p. [V1122G]; [V1122G];
- **NPHP3**: p. [A519G]; [A519G];
- **NPHP4**: p. [V1122G]; [V1122G];
- **NPHP5**: p. [A684]; [A684];
- **ALMS1**: p. [E1114Rfs*9]; [E1114Rfs*9];
- **NPHP6**: p. [A1959G]; [A1959G];

**Potential mutations**

- **p. [D179N]**; [E1114Rfs*9];
- **p. [P641S]**; [P641S];
- **p. [P2679L]**; [P2679L];
- **p. [V1122G]**; [V1122G];
- **p. [A1959G]**; [A1959G];
- **p. [E569G]**; [E569G];
- **p. [Y486C]**; [Y486C];
- **p. [A2423S]**; [A2423S];
- **p. [E569G]**; [E569G];
- **p. [R863W]**; [R863W];
- **p. [R830W]**; [R830W];
- **p. [T122P]**; [T122P];

**Mutation, as predicted by**

- **Sanger**
- **PolyPhen2**
- **Mutation T@ster**

**Additional potentially pathogenic variants**

- **SV + S**: MT
- **SV**: S, PP
- **SV**: S, PP
- **SV**: S
- **SV**: S, PP

**Frequency of additional variants in EVSdb**

- **0.13%**, **0.65%**, **2.07%**, **1.65%**, **0.17%**

**Geographic origin**

- **Turkey**
- **France**
- **Portugal**

**Inbred criteria**

- **Yes**
- **Yes**
- **Yes**

**BBS inclusion criteria**

- **Yes**
- **Yes**
- **Yes**

**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, exons; SIFT, PolyPhen2; MT, mutation T@ster.**

**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.

One noteworthy result is the finding of homozygous truncating ALMS1 mutations in 5/38 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. 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, and the reverse situation, such as in our study, of ALMS1 mutations in patients with suspected BBS.

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). 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, 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 (16 20 29 30–32). 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 AKR68 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. 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 BBSome or the BBS-chaperonin complex) might contribute to the disease state in a digenic mode of inheritance proposed in few BBS families.

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...
been a subject of discussion.\textsuperscript{37–39} Analysis in zebrafish indicated 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{44} 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.550C→T in CCDC28B/MGC1203,\textsuperscript{16} 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{41–43} or targeted sequencing, either by using multiplex PCR,\textsuperscript{12} multiple singleplex PCR\textsuperscript{44–46} or capture enrichment approaches.

We implemented an in-solution targeted capture strategy for the 16 known BBS genes and 14 other genes implicated in ciliopathies that share overlapping clinical features with BBS. We show here that this is very efficient since in a single sequencing lane of an Illumina HiSeq2000 with even higher coverage than the one obtained in previous experiments with the GAIIx (data not shown). While exome sequencing is clearly 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). Then, 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{51} 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 systematic 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

<table>
<thead>
<tr>
<th>Retinitis pigmentosa</th>
<th>Obesity</th>
<th>Polydactyly</th>
<th>Brachydactyly</th>
<th>Hypogonadism</th>
<th>Cystic kidney</th>
<th>Mild intellectual disabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 BBS mutations</td>
<td>95.65% (22/23)</td>
<td>86.96% (20/23)</td>
<td>69.57% (16/23)</td>
<td>43.48% (10/23)</td>
<td>39.13% (9/23)</td>
<td>21.74% (5/23)</td>
</tr>
<tr>
<td>2 ALMS1 mutations</td>
<td>3/3</td>
<td>1/3 (AI484)</td>
<td>0/3</td>
<td>2/3 (AI484, AKO26)</td>
<td>2/3 (AI484, AKO26)</td>
<td>1/3 (ALB64)</td>
</tr>
<tr>
<td>0 or 1 mutation</td>
<td>81.82% (9/11)</td>
<td>81.82% (9/11)</td>
<td>25.00% (3/12)</td>
<td>33.33% (4/12)</td>
<td>41.67% (5/12)</td>
<td>50% (5/10)</td>
</tr>
</tbody>
</table>

ALMS1 patients: AI484 was addressed to Strasbourg Diagnostic Laboratory for Bardet–Biedl or Prader-Willi syndromes, AKO26 for suggestive BBS or ALMS with abnormal cognitive defects 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 to triallelic in our cohort of BBS patients.

Author affiliations
1 Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR7104, Université de Strasbourg, Illkirch, France
2 Microarrays and Sequencing Platform, IGBMC, Illkirch, France
3 Laboratory of Human genetics, University of Medicine of Tunis, Tunis, Tunisia
4 Bioinformatics Platform, IGBMC, Illkirch, France
5 Laboratoire de Génétique Médicale EA3949 Inserm, Université de Strasbourg, Strasbourg, France
6 Laboratoire de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France
7 Istituto di Genetica Medica, Universita' Cattolica, Roma, Italy
8 CHU Bordeaux, University of Bordeaux, Department of Medical Genetics, Bordeaux, France
9 Service de Génétique Clinique, Hôpital Jeanne de Flandre, CRU de Lille, Lille, France
10 Service de Cytogénétique Constituionnelle, Hospices Civils de Lyon, CBPE, Bron Cedex, France
11 Department of Genetics, INSERM U676, Assistance Publique Hôpitaux de Paris (AP-HPI), Robert Debré University Hospital, Paris, France
12 Service de Génétique Médicale, Centre de Référence pour les Aeffections Rares en Génétique Ophthalmologique (CARGO), Hôpitaux Universitaires de Strasbourg, Strasbourg, France
13 Chaire de Génétique Humaine, Collège de France, Illkirch, France

Acknowledgements
We thank Ngoc-Hoan Nguyen for his help in the bioinformatics setup, and Cécile Pizot for the development of VaRank. We warmly thank Gérardine Gref, Anne-Sophie Jaeger, Manuela Antin, Elisabeth Scherrer, Vincent Vacic and Muriel Philips 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-Jaoude, C Baumann, H Flodrops, T Frebourg, D Genevieve, A Goldenberg, B Lephoon, P Parent, P Pettijean, C Poitou-Bernert and S Taque.

Contributors
CR, JLM and JM designed the study; CR, OM, CS, BJ performed experiments; CR, SLG, VG, JM performed the bioinformatics studies; CR, SLG, OM, VG, CS, JLM 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, JLM and JM wrote the manuscript; MCV, BJ, HBC, HD, JLM and JM provided technical support, conceptual advice and project coordination.

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

Competing interests
None.

Patient consent
For children, consent was signed by the parents.

Provenance and peer review
Not commissioned; externally peer reviewed.

REFERENCES


