This is the accepted version of a paper published in *Clinical Genetics*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Delvallée, C., Nicaise, S., Antin, M., Leuvrey, A-S., Nourisson, E. et al. (2020) A BBS1 SVA F retrotransposon insertion is a frequent cause of Bardet-Biedl syndrome. *Clinical Genetics*  
https://doi.org/10.1111/cge.13878

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-428350
A BBS1 SVA F retrotransposon insertion is a frequent cause of Bardet-Biedl syndrome

Clarisse Delvallée1 | Samuel Nicaise1 | Manuela Antin2 | Anne-Sophie Leuvrey2 | Elsa Nourisson2 | Carmen C. Leitch3 | Georgios Kellaris3 | Corinne Stoetzel1 | Véronique Geoffroy1 | Sophie Scheidecker1,2 | Boris Keren4,5 | Christel Depienne4,6 | Joakim Klar7 | Niklas Dahi7 | Jean-François Deleuze8 | Emmanuelle Génin9 | Richard Redon10 | Florence Demurger11 | Koenraad Devriendt12 | Michèle Mathieu-Dramard13 | Christine Poitou-Bernert14 | Sylvie Odent15,16 | Nicholas Katansis3,17 | Jean-Louis Mandel2,18 | Erica E. Davis3,17 | Hélène Dollfus1,19,20 | Jean Muller1,2

1Laboratoire de Génétique Médicale, Institut de génétique médicale d’Alsace IGMA, INSERM U1112, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg UMRs_1112, Strasbourg, France
2Laboratoires de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France
3Advanced Center for Translational and Genetic Medicine (ACT-GeM), Stanley Manne Children’s Research Institute, Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois, USA
4Institut du Cerveau et de la Moelle épinière (ICM), Sorbonne Université, Paris, France
5AP-HP, Hôpital de la Pitié-Salpêtrière, Département de Génétique, Paris, France
6Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen, Essen, Germany
7Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden
8Centre National de Recherche en Génomique Humaine (CNRGH), Institut de biologie François Jacob, Evry, France
9Inserm UMR1078, CHRU Brest, Univ Brest, Brest, France
10Université de Nantes, CNRS, INSERM, l’Institut du thorax, Nantes, France
11Service de Génétique Médicale, Centre Hospitalier Bretagne Atlantique, Vannes, France
12Center for Human Genetics, University Hospital Leuven and KU Leuven, Leuven, Belgium
13Centre d’activité de génétique clinique, CLAD nord de France, CHU Amiens, Amiens, France
14Assistance Publique Hôpitaux de Paris, Nutrition Department Pitié-Salpêtrière Hospital; Sorbonne Université, INSERM, NutriOmics Research Unit, Paris, France
15Centre de Référence Maladies Rares CLAD-Ouest, Service de Génétique Clinique, CHU Rennes, Rennes, France
16CNRS, IGDR (Institut de Génétique et Développement de Rennes) UMR 6290, Université de Rennes, Rennes, France
17Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA
18Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS UMR 7104, INSERM U964, Université de Strasbourg, Dept Transl Med and Neurogenetics Illkirch, France
19Service de Génétique Médicale, Hôpitaux Universitaires de Strasbourg, Strasbourg, France
20Filière SENSGENE, Centre de Référence pour les affections rares en génétique ophtalmologique, CARGO, Hôpitaux Universitaires de Strasbourg, Strasbourg, France

Correspondence
Jean Muller, Laboratoire de Génétique Médicale, Institut de génétique médicale d’Alsace IGMA, INSERM U1112, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg

Abstract
Bardet-Biedl syndrome (BBS) is a ciliopathy characterized by retinitis pigmentosa, obesity, polydactyly, cognitive impairment and renal failure. Pathogenic variants in 24 genes account for the molecular basis of >80% of cases. Toward saturated...
discovery of the mutational basis of the disorder, we carefully explored our cohorts and identified a hominid-specific SINE-R/VNTR/Alu type F (SVA-F) insertion in exon 13 of BBS1 in eight families. In six families, the repeat insertion was found in trans with c.1169 T > G, p.Met390Arg and in two families the insertion was found in addition to other recessive BBS loci. Whole genome sequencing, de novo assembly and SNP array analysis were performed to characterize the genomic event. This insertion is extremely rare in the general population (found in 8 alleles of 8 BBS cases but not in >10 800 control individuals from gnomAD-SV) and due to a founder effect. Its 2435 bp sequence contains hallmarks of LINE1 mediated retrotransposition. Functional studies with patient-derived cell lines confirmed that the BBS1 SVA-F is deleterious as evidenced by a significant depletion of both mRNA and protein levels. Such findings highlight the importance of dedicated bioinformatics pipelines to identify all types of variation.

**KEYWORDS**
Bardet-Biedl syndrome, BBS1, founder effect, Mobile element insertion, SVA F

1  |  INTRODUCTION

Bardet-Biedl syndrome (BBS, MIM# 209900) is a rare autosomal ciliopathy associating retinitis pigmentosa, polydactyly, obesity, cognitive impairment and renal dysfunction. Its prevalence ranged from 1/160000 live births\(^1\) to <1/50000 in consanguineous or isolated populations.\(^2\) Recent studies have delineated the antenatal and postnatal presentation of BBS patients\(^3,4\) with pathogenic variants in 24 genes accounting for the molecular basis of a majority of cases.\(^5\) Among the known BBS genes, BBS1 contributes the highest fraction of BBS patients (\(^\sim\)20%)\(^6\) thanks to a founder variant (c.1169T>G, p. Met390Arg).\(^7\) Among the unresolved cases in our cohort, several patients were heterozygous carriers of this variant suggesting two hypothesis: (1) a missed second primary driver allele in BBS1; or (2) participation of this variant as a second-site modifier in the presence of a different primary BBS driver locus.\(^8,9\) Notably, due to the high frequency of this variant in the general population (0.27% in non-Finnish Europeans; gnomAD) this variant could be detected coincidently.

In this study, we report the identification of a mobile element insertion of SVA F type in 8 families, a portion of which are in trans to the recurrent c.1169 T > G in BBS1 (n = 6), and a subset as secondary sites in addition to a known recessive BBS locus (n = 2). Already described as a single case report,\(^10\) here we estimate the frequency of this variant in BBS (131 families) identifying this variant as the second most common pathogenic variant in BBS1. We also determined its full-length sequence using WGS and performed functional analysis to understand its consequences. In our cohorts, the insertion always occurs at the same genomic location and its recurrence in BBS populations is the result of a founder effect.

2  |  SUBJECTS AND METHODS

2.1  |  Subjects

Written informed consent was obtained from each participant and/or parent or legal guardian. The study protocols were approved by local Institutional Review Boards: “Comité de Protection des Personnes” (EST IV, N°DC-20142222), or the Lurie Children’s Hospital IRB (IRB 2019–0357 or IRB 2019–2950). Our research complies with the Declaration of Helsinki. All patients were diagnosed with BBS according to Beales’ clinical diagnostic criteria\(^11\) (Supplementary Table 1). Patients (n = 564) have been assembled into several cohorts (Supplementary Table 2).

2.2  |  Methods

All experimental procedures are reported as supporting information (Appendix S1).

3  |  RESULTS

3.1  |  Identification of an SVA F insertion in BBS1

According to our tiered diagnostic paradigm, we screened the recurrent BBS variants (BBS1:p.Met390Arg and BBS10:p.Cys91Leufs*5) in cohort 1 (n = 217) and cohort 2 (n = 192) followed by targeted exome sequencing (TES) in cohort 1. We identified seven individuals who carry heterozygous p.Met390Arg changes without a second pathogenic BBS1 variant identified in trans. In one family (family 1), we...
FIGURE 1  Pedigree of Bardet-Biedl syndrome (BBS) families and variant analysis. (A) Eight families carrying the hominid-specific SINE-R/VNTR/Alu type F (SVA F) insertion in exon 13 of \textit{BBS1}. (B) Next generation sequencing data displayed from the Integrative Genomics Viewer\textsuperscript{23} surrounding the \textit{BBS1} locus from A.II-3 and one control individual. (C) Magnified view of the incorrectly called deletion insertion with the sorting and coloring of the reads according to the "insert size." The sequence reveals all reads ending at the exact same position with a "AG" deletion and a "T" insertion (c.1215_1216delinsT). (D) Region of interest with the "Show soft-clipped bases" turned off revealing multiple reads with aberrant alignments on their right side and corresponding to an SVA F insertion. The spacing between the reads on the left (including the 3' end of the SVA F with the poly(A) tail) and the reads on the right (including the 5' end of the SVA F) corresponds to the TSD. (E) Segregation analysis using Sanger sequencing for M1 (left) and M2 (right) [Colour figure can be viewed at wileyonlinelibrary.com]
identified a novel heterozygous deletion encompassing exons 4 to 11 (Supplementary Figure 1). Among the six p.Met390Arg variation-bearing cases, we leveraged a high depth of coverage to identify another putative pathogenic variant in exon 13 of BBS1 with the deletion/insertion of 2 bases, c.1215_1216delinsT (p.[Ala406Glnfs*4]) in family A.II-3 (Figure 1(B) and supplementary Figure 2) and 2 others (B. II-1, C.II-1). Careful analysis of the genomic region revealed a more complicated event leading to incorrect variant calling (Figure 1(C)). Indeed, soft clipped reads part (82 bp) revealed the insertion of an SVA F retrotransposon with a possible truncation of 311 bases 5' of the consensus sequence (Figure 1(D)).

To screen additional patients not yet available by TES (cohort 2; 213 cases from 192 families), or an independent group with no pre-selection bias for BBS variations other than exclusion of BBS1 recessive cases (cohort 3; 134 cases from 126 families), we designed a duplex PCR including the SVA insertion point (Supplementary Figure 3). This led to the identification of 5 additional cases (family D to H) including 3 in cis to the p.Met390Arg and 2 as third or fourth allele (Figure 1(A)). Biallelic status of the variants and the exact same insertion point was confirmed for 7 individuals (Figure 1(E)).

To estimate the overall incidence of the SVA F as a contributor of the BBS1 mutational burden, we explored all our BBS1 positive cases (e.g. biallelic pathogenic variants). Data from 131 unrelated cases of European ancestry corresponding to 262 pathogenic alleles in total (cohort 4) could be retrieved and account for 44 different pathogenic variations. Among those, the c.1169T>G represents 72.1% (189 alleles) while the second most frequent variant is the SVA F with a frequency of 2.3% (6 alleles).

### 3.2 Determination of the full length SVA F sequence

To understand the consequences of this insertion, we determined its full-length sequence. Given the repeated sequence of the SVA F, alignment of short reads onto the human genome generate artifacts (multimapped reads, split reads, discordant read pairs) that can be rescued using a dedicated bioinformatics pipeline. The initial TES aligned reads allowed us to identify 99 bp from the SVA F (Figure 2(B)). Assembled reads from WGS data (A.II-3) output a 387 bp contig composed of BBS1 (38 bp before the breakpoint in exon 13), a TSD (17 bp) and 330 bp that matched the SVA F consensus sequence. In conclusion, WGS data enabled accurate characterization of an additional 250 bp SVA F sequence (Figure 2). Based on this and the consensus SVA F sequence,12 we walked on the genome and determined an inserted sequence of 2435 bp (Figure 2(B) and Supplementary Figures 4 and 5). With the exception of a missing 311 bases (Figure 2(A)) in the 5' part, hallmarks of L1 mediated retrotransposition could be identified (Figure 2). The variant can now be described as c.1214_1215ins[MT113356]. p.(Ala406Glnfs*47).

### 3.3 Functional impact of the SVA F insertion on BBS1

Predicted to create a premature stop codon, a reduced or absent mRNA expression due to nonsense-mediated decay potentially leading to a shorter protein (452 vs 593 AA) is likely to happen. Using

![FIGURE 2](https://wileyonlinelibrary.com)

**FIGURE 2** Schematic of the inserted hominid-specific SINE-R/VNTR/Alu type F (SVA F) element in exon 13 of BBS1. (A) Canonical SVA F structure composed of six parts: (1) hexameric CCCCTCT repeats (2) Alu-like domain, (3) a Variable Number Tandem Repeats (VNTR) domain, (4) a Short Interspersed Nucleotide Elements (SINE-R) domain, (5) a poly(A) tail and (6) the target site duplication (TSD) at both extremities. Numbers above the scheme indicate the size of each element. (B) Schematic of both BBS1 alleles in A.II-3 with primer positioning and amplification products highlighting each contribution to the resolution of the full-length sequence [Colour figure can be viewed at wileyonlinelibrary.com]
primary cells from individual A.II-3, we demonstrated a significantly reduced BBS1 expression corresponding a priori to the loss of the SVA F allele (Figure 3A). However, a transcript carrying the SVA F is expressed (Supplementary Figure 6), albeit likely minor compared to the p.Met390Arg allele. Western blot analysis on protein extracted from A.II-3 cells revealed a single band corresponding to the normal BBS1 protein (Figure 3(B)) demonstrating the absence of a putative BBS1 polypeptide carrying the SVA F.

3.4 Founder effect or recurrent mechanism?

Given the multiple carriers of this insertion in our cohorts (8 families), we wondered whether this MEI occurred as an independent (recurrent) mechanism or as the same mutational event derived from common ancestry (founder effect). First, no occurrence could be observed either from 1299 healthy individuals of European ancestry or from the 14 891 gnomAD SV controls.13 Second, genotyping of four different families with SVA F alleles, along with three patients carrying the c.1169T>G change in the homozygous state (Supplementary Table 5) revealed the common c.1169 T > G haplotype (\[C24 \text{1.614 Mb}\]) as well an SVA F allele of a much smaller size (\[C24 \text{56.6 kb}\]) (Supplementary Table 5). PCA analysis confirmed the geographical distribution of the patients (Supplementary Figure 7). Haplotypes shared by the four SVA F carriers were compared to find the position where recombination events are likely to have happened. The estimation of the most recent common ancestor age was 74 generations 95% CI \([33;190]\). If we assume that one generation lasts 25 years, this will indicate an age of 1850 years 95% CI \([825;4750]\). Overall, this MEI is an ultra-rare event that arose from a common ancestor on one of the most common haplotypes at this locus and more recently than the Met390Arg haplotype.

4 DISCUSSION

Although the diagnostic yield in BBS is relatively high, (60–100%), a substantial proportion of cases remain unresolved.3,5 This could be related to the attributes of the cohort (sample size and geographic origin), the clinical inclusion criteria (strict or more inclusive) or the genetic test used (TES, WES, WGS). In addition to the first BBS case with MEI reported by our group14 and the recent SVA F description,10 this is the largest description of such cases in BBS patients. Although rarely described, a number of MEI cases (>100) have been reported in rare diseases15 or in cancer.16 The majority are Alu insertions17 and only a few SVA (n = 16). Recently, large studies have evaluated the impact of MEI in patients between 0.04 and 0.15% of the cases.18,19 BBS1 is often the target of structural variations.20 Interestingly, the BBS1 locus is GC rich and has a high density of L1 and Alu insertions that might explain the SVA F insertion (Supplementary methods). Given the nature of the SVA F sequence, which contains highly repeated elements, GC rich content and a poly(A) tail, complete amplification was not easily performed. We thus used the breakpoints identified by the TES and WGS to walk across the patient’s genome and determine a 2435 bp insertion in exon 13 of BBS1. This inserted sequence is identical to the one already described.10 Unlike most of the SVA elements in the human genome (63%),21 this inserted SVA F is incomplete lacking 311 bases in 5' and possibly as almost all SVA it is inactive.17 Notably, beside short read standard WGS we also performed linked-read sequencing (10x genomics) to see whether this technology could overcome any of the technical challenges encountered through WGS and analysis (Supplementary Figure 8). Consistent with other reports, we did not find any significant advantage and it is now discontinued.22 Interestingly sequencing of the insertion point revealed the exact same sequence for all patients and the genotyping of some individuals revealed a common shared haplotype. All families
carrying the SVA F are of European ancestry: 4 originated from the western part of France (family A, C, E and F), one from the northern part of France (family B) and one from Denmark and Sweden (family D). These observations were highly suggestive of a single founder genomic event that we confirmed and estimated its age close to 1850 years.

In conclusion, we identified the recurrence of a specific SVA F element in exon 13 of BBS1 in multiple BBS families highlighting the importance of detecting such genetic events even by TES. We determined the full length sequence of the inserted SVA F, showed its pathogenicity and demonstrated the founder effect. This rare allele is the second most frequent cause of BBS1 after the c.1169T>G variant. We have also developed a duplex PCR, offering a simple and inexpensive way to detect this variant prior to any other large genetic screen. Accordingly we have modified our BBS diagnostic strategy (Supplementary Figure 10).

ACKNOWLEDGEMENTS
We thank the patients and their family members for their participation, and the French Bardet-Biedl association for their support. Part of TES was performed by the Plateforme GenomEast (IGBMC, Illkirch-Graffenstaden, France) and for this, we thank Serge Vicaire, Murielle Philipp and Bernard Jost. We would like to thank Amélie Piton for scientific discussions and acknowledge Pierre Lindenbaum and every member of the FranceGenRef consortium.

CONFLICT OF INTEREST
Nicholas Katsanis is a significant shareholder in Rescindo Therapeutics. The other authors declare no conflict of interest.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1111/cge.13878.

DATA AVAILABILITY STATEMENT
Data generated or analyzed during this study are included in the published article and the corresponding supplementary data. The raw sequencing data generated in the course of this study are not publicly available due to the protocol and the corresponding consents used that did not include such information. The variants are available in ClinVar (SCV001427246, SCV001427247) and the SVA F sequence in GenBank (MT113356).

ORCID
Joakim Klar https://orcid.org/0000-0003-4185-7409
Jean Muller https://orcid.org/0000-0002-7682-559X

REFERENCES
SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.