Pathogenic neurofibromatosis type 1 (NF1) RNA splicing resolved by targeted RNAseq

Koster, R.; Brandao, R.D.; Tserpelis, D.; Roozendaal, C.E.P. van; Oosterhoud, C.N. van; Claes, K.B.M.; ... ; Blok, M.J.

npj | Genomic Medicine

www.nature.com/npjgenmed

ARTICLE OPEN

Check for updates

Pathogenic neurofibromatosis type 1 (*NF1*) RNA splicing resolved by targeted RNAseq

R. Koster[1], R. D. Brandão[1,2], D. Tserpelis[1], C. E. P. van Roozendaal[1], C. N. van Oosterhoud[1], K. B. M. Claes[3], A. D. C. Paulussen[1,2], M. Sinnema[1], M. Vreeburg[1], V. van der Schoot[1], C. T. R. M. Stumpel[1], M. P. G. Broen[4], L. Spruijt[5], M. C. J. Jongmans[6,7], S. A. J. Lesnik Oberstein[8], A. S. Plomp[9], M. Misra-Isrie[9], F. A. Duijkers[10], M. J. Louwers[11], R. Szklarczyk[1], K. W. J. Derks[1], H. G. Brunner[1,2,5,12,13], A. van den Wijngaard[1], M. van Geel[1,2] and M. J. Blok[1,2]✉

Neurofibromatosis type 1 (NF1) is caused by loss-of-function variants in the *NF1* gene. Approximately 10% of these variants affect RNA splicing and are either missed by conventional DNA diagnostics or are misinterpreted by in silico splicing predictions. Therefore, a targeted RNAseq-based approach was designed to detect pathogenic RNA splicing and associated pathogenic DNA variants. For this method RNA was extracted from lymphocytes, followed by targeted RNAseq. Next, an in-house developed tool (QURNAs) was used to calculate the enrichment score (ERS) for each splicing event. This method was thoroughly tested using two different patient cohorts with known pathogenic splice-variants in *NF1*. In both cohorts all 56 normal reference transcript exon splice junctions, 24 previously described and 45 novel non-reference splicing events were detected. Additionally, all expected pathogenic splice-variants were detected. Eleven patients with NF1 symptoms were subsequently tested, three of which have a known *NF1* DNA variant with a putative effect on RNA splicing. This effect could be confirmed for all 3. The other eight patients were previously without any molecular confirmation of their NF1-diagnosis. A deep-intronic pathogenic splice variant could now be identified for two of them (25%). These results suggest that targeted RNAseq can be successfully used to detect pathogenic RNA splicing variants in *NF1*.

npj Genomic Medicine (2021)6:95 ; https://doi.org/10.1038/s41525-021-00258-w

INTRODUCTION

Neurofibromatosis type 1 (NF1; MIM 162200) is one of the most common autosomal dominant tumor-predisposition disorders affecting ~1 in 3000 individuals[1,2]. Patients typically present with the following clinical manifestations: café-au-lait spots, dermal neurofibromas, iris hamartomas (Lisch nodules), axillary and/or inguinal freckling, and subcutaneous or plexiform neurofibromas. Other manifestations include learning disabilities, optic glioma, skeletal abnormalities and an increased risk of specific malignancies[1–3]. Malignancies are an important cause of the 8–15-year reduction in average life expectancy[1–3]. Most patients are clinically diagnosed in childhood, according to NIH consensus criteria[4]. However, genetic testing to confirm a clinical diagnosis is still warranted, because of the clinical overlap with Legius syndrome (MIM 611431; *SPRED1)*[1,2,5]. In addition, an accurate genetic diagnosis facilitates appropriate screening and follow-up, reproductive options (prenatal testing and pre-implantation genetic diagnosis) and access to clinical studies and potential future trials and treatments.

NF1 is caused by loss-of-function variants in the tumor-suppressor gene *NF1* (MIM 613113), located on chromosome 17q11.2 and consisting of ~350 kb of genomic DNA. The *NF1* gene produces a major 12 kb transcript NM_000267.3 that contains 57 exons and encodes for a Ras-guanosine triphosphatase (GTPase) activating protein: neurofibromin[1,6–8]. Pathogenic variants in *NF1* result in loss of function of neurofibromin causing an increase in Ras signaling, affecting cell proliferation and differentiation[9].

Currently, there are almost 3700 (likely) pathogenic *NF1* variants reported in HGMD (HGMD® Professional 2020.4) and 1890 in Clinvar (Jan 2021)[10,11], and at least half of these arose de novo[12–16]. The majority of (likely) pathogenic variants in *NF1* are predicted to produce a truncated form of neurofibromin, and of these variants ~30% cause splicing alterations affecting mRNA processing[12–16]. At least a third of these splicing alterations are either not found using standard DNA diagnostics (targeting only the coding region, several intronic nucleotides, but not including potential retrotransposon insertions) or are variants not predicted to affect splicing by current bioinformatics algorithms[12–17].

The diagnostics laboratory community has adopted the ACMG standards and guidelines for the interpretation of sequence variants[18] to determine their clinical relevance, including recommendations for splice variants. For variants adjacent to exon boundaries and not further than two nucleotides from the exon border, in silico tools often adequately predict changes in RNA splicing[19]. In genes were loss-of-function is a known mechanism of disease, a null variant such as a variant within the canonical splice site could get a pathogenic very strong 1 (PSV1) score. (See Richards et al[18] for the full explanation). However splice prediction tools for variants outside the consensus splice acceptor/donor

[1]Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands. [2]GROW Institute for Developmental Biology and Cancer, Maastricht University Medical Center, Maastricht, The Netherlands. [3]Centre for Medical Genetics, Ghent University Hospital, Ghent, Belgium. [4]Department of Neurology, Maastricht University Medical Center, Maastricht, The Netherlands. [5]Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands. [6]Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands. [7]Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands. [8]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands. [9]Department of Clinical Genetics, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [10]Department of Clinical Genetics, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands. [11]Pediatrics, Bernhoven hospital, Uden, The Netherlands. [12]Donders Center for Neuroscience, Radboud University Medical Centre, Nijmegen, The Netherlands. [13]MHENS School for Neuroscience, Maastricht University Medical Center, Maastricht, The Netherlands. ✉email: rien.blok@mumc.nl

Published in partnership with CEGMR, King Abdulaziz University

npj nature partner journals

sites perform poorly[20]. And although these tools are able to predict cryptic or novel splice sites, the in vivo outcome of competing splice sites for the splicing complex is difficult to predict[21]. Moreover, the specificity of predictions for the effect of genetic variants on splice enhancer or silencer sites, potential branch points[22] or even change in the branch point-exon distance[23] is limited. As a consequence, the effect of DNA variants on RNA splicing ultimately needs to be determined experimentally. A variant that shows an effect with a well-established functional study; especially a robust, reproducible and validated study in a clinical diagnostic laboratory will get a pathogenic strong 3 (PS3) score. As such a potential splice variant that shows a loss-of-function effect could get a PS3 score and in some cases a PVS1 score.

The functional study most often employed for splice variant validation is conventional Reverse Transcription (RT)-PCR, which has its drawbacks. First, it is laborious, time consuming and prone to preferential amplification of transcripts depending on primer choice. Another possible pitfall is the use of primers that amplify only a small part of the region surrounding the variant, while the genetic effect may involve skipping of multiple exons[21]. Furthermore, data are only semi-quantitative and, finally, almost every new splice isoform needs its own RT-PCR primer design. Targeted RNAseq on the other hand gives comprehensive information about RNA expression and splicing for genes/transcripts and can circumvent the aforementioned draw-backs[24].

A diagnostic procedure was designed that uses an in-house developed tool QURNAs (**Qu**antitative enrichment of aberrant splicing events in targeted **RNAs**eq) to facilitate the detection and quantification of normal and/or pathogenic RNA splicing events in targeted RNA sequencing data.

RESULTS

QURNAs detects normal and known pathogenic *NF1* splicing events

To validate our diagnostic procedure, targeted RNAseq was performed using blood samples of nine patients with a known pathogenic *NF1* splice-variant, and two control samples. (Table 1).

First, the presence of all normal exon-exon splice junctions belonging to *NF1* reference transcript NM_000267 was checked and found detectable with a median enrichment score (ERS) ~1 and >5000 reads per splice junction across all samples (Fig. 1a, Supplementary Data 1 and 2). Furthermore, with QURNAs an extensive list of (naturally occurring) splicing events for *NF1* was obtained (Supplementary Data 2) and compared to previously published naturally occurring events (Supplementary Data 3). From the 25 previously reported normal (rare) events, 24 were detected. Approximately half of these previously reported events are major splicing events, detected by QURNAs in almost all of the patient derived cells (median ERS~1 & reads >1000). The other half are minor events (median ERS < 1 & reads <700) and detected in a fraction of the patient derived cells (Supplementary Data 2). The only missed previously reported normal rare event is from a transcript 'ins 9a/9br' (legacy exon nomenclature, corresponding to insertion of 10 amino acids between exons 11–12 of *NF1;* c.1260 + 1617_1260 + 1646) that is supposed to be specific for neuronal tissue, and not detectable in blood[25–27]. Additionally 45 events not published before were detected in multiple samples (Supplementary Data 2).

For the nine samples with a known pathogenic splice-variant all expected (major) pathogenic splicing events were detected using QURNAs. The corresponding causal DNA variant could also be observed in the RNAseq reads (except for one - data not shown). The highest ERS per sample and concomitant major drop in ERS of the normal exon-exon splice-junctions due to alternative splicing are listed (Table 1, Supplementary Data 1, 2 & Fig. 1a). Samples 2,3

Table 1. Overview of validation patient samples, including nine samples with a known pathogenic *NF1* splice-variant and two anonymous patients without a clinical NF1 diagnosis.

Sample	NF1 variant (NM_000267.3)	Expected/published splicing events^	Top splicing event (effect)^	Top event ERS (reads)	RNA	Protein
Val. 01	c.7127−1 G>A	Δ1nt 5′exon 48 (FS)	Δ1nt 5′exon 48 (FS)	16.5 (999)	r.7128del	p.(Tyr2377Thrfs*20)
Val. 02	c.654 + 2dup	Δexon 6 (FS)	Δexon 6 (FS) Δexon 6,7 (IF)	23.5 (2676) 8.1 (1063)	r.587_654del r.587_730del	p.(Glu196Glyfs*12) p.(Thr197_Glu244del)
Val. 03	c.4773−2 A>G	use of cryptic acceptor site, Δ293nt 5′exon 36 (FS)[*1,2]	Δ293nt 5′exon 36 (FS) +multiple (incl Δ36)	15.1 (392) range 6.3–10.2 (295–400)	r.4773_5065del multiple	p.(Phe1592Leufs*7) Δ36=p.(Phe1592Leufs*8)
Val. 04	WT	—	Δ3′ part exon 36–5′ part exon 42 (IF)	5 (150)[#1]	r.4986_6506del	p.(Trp1664_Ser2170del)
Val. 05	WT	—	splicing 3′UTR	6.7 (90)[#2]	r.?	p.?
Val. 06	c.586 + 5 G>A	Δexon 5 (FS)[*3]	Δexon 5 (FS)	6.1 (6509)[#3]	r.480_586del	p.(Leu161Asnfs*4)
Val. 07	c.1466 A>G (p.Tyr489Cys)	Δ62nt 3′exon 13 (FS)[*4]	Δ45nt in exon 18 (IF) Δ62nt 3′exon 13 (FS)	23.4 (228)[#4] 22.4 (1792)	r.2106_2150del r.1466_1527del	p.(Val703_Ala717del) p.(Tyr489*)
Val. 08	c.4515−15_4515−13delGTT	ins 14nt 5′ to exon 34 (FS)	ins 14nt 5′ to exon 34 (FS)	19.2 (3123)	r.4514_4515ins4515−17_4515−1del4515−15_4515−13	p.(Arg1505Serfs*53)
Val. 09	c.1721 + 2 T>A	Δexon 15 (FS)	Δexon 15 (FS)	9.6 (2793)[#5]	r.1642_1721del	p.(Ala548Leufs*13)
Val. 10	c.5205 + 1 G>C	Δ54nt 3′exon 36 (IF)	Δ54nt 3′exon 36 (IF)	44.9 (9158)	r.5152_5205del	p.(Phe1719_Val1736del)
Val. 11	c.1885G>A (p.Gly629Arg)	Δ41nt 5′exon 17(FS)[*2]	Δ41nt 5′exon 17 (FS) Δ46nt 5′exon 17 (FS)	18.1 (1420) 5.4 (703)	r.1846_1886del r.1846_1891del	p.(Gln616Glyfs*4) p.(Gln616Aspfs*57)

Expected and/or published splicing events are shown together with the top events found using QURNAs. RNA was enriched for *NF1* using PCR based (NuGEN SPET) approach. #1, 2, 4 are artefacts, they do not represent true splicing events; #3 ΔExon5 natural occurring event (ERS ~0.5; 731 reads across all samples); #5 Δ15 natural occurring event (ERS ~0.3 ERS; 144 reads across all samples). ^ FS = frame shift; IF = in frame. [*1]Wimmer et al. 2007[13]; [*2]Pros et al. 2008[29]; [*3]De Conti et al. 2012[30]; [*4]Messiaen et al. 1999[32]

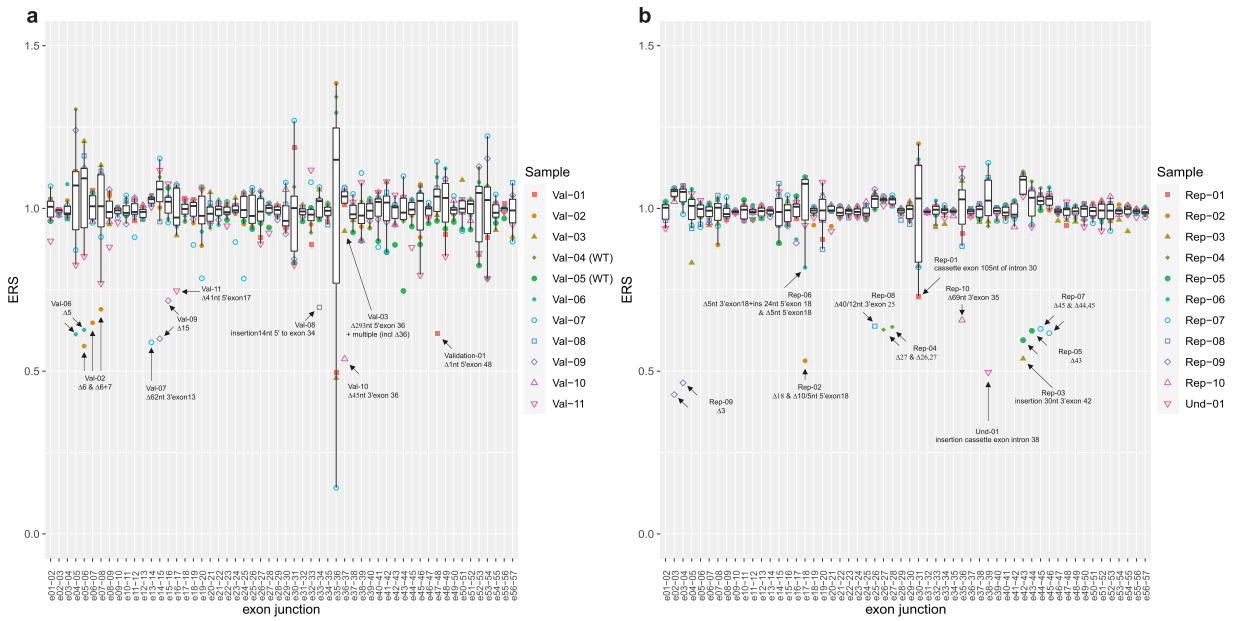

Fig. 1 Drop in enrichment score of normal exon splice junctions for all NF1 patients is clearly visible in close proximity to their specific pathogenic splicing event. Enrichment score of normal exon-exon splice junctions of the validation cohort (**a**) and the replication cohort (**b**). Shown are the ERS score for the normal exon splice junctions for the validation (**a**) and replication (**b**) cohort. Annotated in the figures are the drop in ERS due to the pathogenic splicing event. Normal reference exon splice junctions are expected to have a median ERS ~1. center line of the boxplot represents the median value (Q2), bounds of box are Q1 and Q3 respectively and whiskers Q1–1.5xIQR and Q3 + 1.5xIQR respectively. Outliers are depicted with color and shape matching the specific sample (see inner figure legend).

and 11 had additional (minor) splicing events that are related to their major pathogenic splicing event albeit with lower ERS. Events with a maximum ERS of ~7 were found for the two control samples. However, these events had relatively low reads per event (< 150) and do not appear to be true splicing events based on in silico predictions (Supplementary Fig. 2). Also, the concomitant drop in ERS of the normal exon-exon splice junctions in close proximity was absent, indicating that wild-type splicing was not significantly altered (Fig. 1a). This indicates that different data sources have to be taken into account for correct interpretation of an increased ERS, to be able to discriminate artefacts/non-splicing related events from true alternative splicing events. This is further illustrated by samples 6 and 7.

For sample 6 of the validation cohort with variant c.586 + 5 G > A, previously shown to cause skipping of exon 5[28–30], QURNAs indicated an ERS of 6.1 for exon 5 skipping (covered by 6509 reads). This is also reflected in Fig. 1a by a drop in ERS at the exon 4–5 junction (ERS~0.61) and 5–6 junction (ERS~0.63) compared to the other samples in the same run (median ERS ~1 for exon junction 4–5 and median ERS~1 for 5–6 across all samples—see Fig. 1a, Supplementary Fig. 3). The ERS of 6.1 for exon 5 skipping is smaller than expected, most likely because exon 5 skipping is a (previously published) naturally occurring event (Δ5 or 'NF1-ΔE4b')[31] that is detected with a median ERS of 0.52 and median 728 reads across all samples (Supplementary Fig. 3). Sample 7 of the validation cohort harbors variant c.1466 A > G (p.(Tyr489Cys)) in *NF1*. This specific variant has been reported many times as a disease-causing splice variant, since RNA studies have shown that it creates a new donor site leading to the deletion of the last 62nt of exon 13 (Supplementary Fig. 4)[15,28,32]. This is confirmed using RNAseq (high ERS of 22.4 and covered by 1792 reads). The event corresponds to a drop in ERS to 0.59 for sample 7 at the junction of exon 13–14 compared to median ERS~1 across all samples for this event in the same run. The observed drop in ERS is due to partial loss of normal exon 13 to exon 14 splicing. However, one event with a higher ERS score was also observed for this sample (ERS 23.4; 228 reads - Table 1, Supplementary Data 1). Again, this event corresponds to an artefact based on in silico predictions (Supplementary Fig. 4), the relatively low read number and the concomitant drop in ERS of the normal exon-exon splice junctions in close proximity is lacking (Fig. 1a).

RNAseq replication of RT-PCR based molecular diagnosis

Ten RNA samples from an external diagnostic laboratory that uses RT-PCR to detect pathogenic *NF1* splicing, were subsequently retested using the targeted RNAseq assay and processed with QURNAs. Initially, the RT-PCR results were blinded to us to test the performance of the targeted RNAseq assay in a simulated diagnostic setting. The RT-PCR results could be fully replicated with some minor differences and could even be complemented with splicing events that are more difficult to detect or extract from conventional RT-PCR and Sanger sequencing data (see Table 2; further discussed below). Using the targeted RNAseq diagnostic procedure all pathogenic splicing events were detected with high ERS for the aberrant event (Table 2) and the concomitant drop in ERS of the normal exon-exon splice junctions in close proximity (Fig. 1b). For these samples, as for the validation set, we could detect all normal reference exon splice junctions (median ERS~1, median reads > 800) and 24 published naturally occurring splicing events that are detectable in blood derived RNA (Supplementary Data 2, 4). Besides detection of the causal splicing events using QURNAs, the corresponding causal DNA variant could be observed in the RNAseq reads using the BAM-file (data not shown).

A minor discrepancy between the RT-PCR and RNAseq data was observed for replication sample 6. RT-PCR determined skipping of exon 18, while RNAseq in combination with QURNAs showed deletion of the first 5nt in combination with the intronic insertion of 24nt (Δ5nt 5′exon 18 + ins24nt intron 17; ERS 4.3 with 124 reads) and deletion of the first 5nt only (Δ5nt 5′exon 18; ERS 2.3 with 53 reads). Skipping of exon 18 is only observed with ERS 0.5 and 16 reads. After thoroughly reinvestigating the RT-PCR Sanger traces, the forementioned events could be observed in the background. An explanation for this difference may be preferential amplification of the smaller Δexon 18 fragment with RT-PCR.

Table 2. Overview of replication patient samples, including ten samples with a known pathogenic *NF1* splice-variant (initially we were blinded for the variants).

Sample	NF1 variant (NM_000267.3)	splicing events determined by RT-PCR (effect)^	Top splicing event determined using RNAseq/QURNAs (effect)^	Top event ERS Nugen SPET (reads)	RNA	Protein	Top event ERS Agilent Sureselect (reads)
Repl. 01	c.4110 + 945 A>G	cassette exon 105nt of intron 30 (FS)	cassette exon 105nt of intron 30 (FS)	34.6 (754) & 21.2 (611)	r.4110_4111ins4110 + 836_4110 + 940	p.(Gln1370_Val1371ins9*)	n/a
Repl. 02	c.2002−2 A>G	Δ10nt 5′exon 18 (FS)	Δ10nt 5′exon 18 (FS) Δexon 18 (FS) Δ5nt 5′exon 18 (FS)	13.8 (160) 7.3 (75) 3.0 (38)	r.2002_2011del r.2002_2251del r.2002_2006del	p.(Asp668Glnfs*17) p.(Asp668Glufs*9) p.(Asp668Cysfs*30)	19.5 (7227) 2.0 (734) 3.1 (1561)
Repl. 03	c.6579 + 31 T>G	ins 30nt 3′exon 42 (FS)	ins 30nt 3′exon 42 (FS)	37.1 (181)	r.6579_6580ins6579 + 1_6579 + 30	p.?	n/a
Repl. 04	c.3708 + 3 A>T	Δexon 27 (FS)	Δexon 27 (FS) Δexon 26,27 (FS)	32.0 (471) 1.8 (34)	r.3497_3708del r.3315_3708del	p.(Leu1167*) p.(Tyr1106Metfs*3)	n/a
Repl. 05	c.6641 + 1 G>C	Δexon 43 (FS)	Δexon 43 (FS)	33.2 (417)	r.6580_6641del	p.(Ala2194Ilefs*6)	40.2 (5941)
Repl. 06	c.2002−3 C>G	Δexon 18 (FS)	Δ5nt 5′exon 18 + ins24nt intron 17 (FS) Δ5nt 5′exon 18 (FS)	4.3 (124) 2.3 (53)	r.2002_2006delins2001 + 1_2001 + 24 r.2002_2006del	p.(Asp668Valfs*38) p.(Asp668Cysfs*30)	6.6 (3252) 4.4 (1894)
Repl. 07	c.6792 C>A; p.(Tyr2264*)	Δexon 45 (IF)	Δexon 45 (IF) Δexon 44,45 (FS)	27.8 (615) 4.5 (56)	r.6757_6858del r.6642_6858del	p.(Ala2253_Lys2286del) p.(Phe2215Thrfs*11)	26.0 (9097) 2.2 (1529)
Repl. 08	c.3314 + 1 G>C	Δ40nt 3′exon 25 (FS)	Δ40nt 3′exon 25 (FS) Δ12nt 3′exon 25 (IF)	9.8 (201) 7.4 (85)	r.3275_3314del r.3303_3314del	p.(Gly1092Aspfs*7) p.(Leu1102_Lys1105del)	7.5 (3795) 4.2 (1473)
Repl. 09	c.288 + 1 G>A	Δexon 3 (IF)	Δexon 3 (IF)	55.5 (1836)	r.205_288del	p.(Arg69_Gly96del)	28.3 (8086)
Repl. 10	c.4772 + 2 T>C	Δ69nt 3′exon 35 (IF)	Δ69nt 3′exon 35 (IF)	11.8 (116)	r.4704_4772del	p.(Leu1569_Arg1591del	9.2 (3348)

The previously determined splicing events using RT-PCR and Sanger sequencing by an external clinical lab are shown, together with the top events found using QURNAs. RNA was enriched for *NF1* using PCR based (NuGEN SPET) approach or for *NF1* and *SPRED1* using hybridization based enrichment (Agilent SureSelect) as indicated. ^ FS=frame shift; IF=in frame.

For sample 2, RNAseq identified besides the expected with RT-PCR detected deletion of the first 10nt of exon 18 (ERS 13.8, reads 160), additional exon 18 skipping (ERS 7.3, reads 75) and deletion of the first 5nt of exon 18 (ERS 3.0, reads 38, see Table 2 & Supplementary Data 3). After reevaluating the RT-PCR Sanger traces, these events could also be observed as minor events in the background.

Targeted enrichment platform comparison

Part of the replication cohort samples was used to test the performance of QURNAs in combination with a different library preparation method to enrich for *NF1* transcripts. PCR based targeting (NuGEN SPET) to enrich for *NF1* was switched to a hybridization based enrichment (Agilent SureSelect). Comparable results were obtained with QURNAs using the different enrichment protocols. All splicing events were present using both methods with some variations in ERS and reads (Table 2 & Supplementary Fig. 5).

Testing molecular undiagnosed cases with a (suspected) clinical diagnosis of NF1

Eleven patients with a (suspected) clinical diagnosis of NF1, who lacked a definitive molecular diagnosis based on DNA diagnostics, were tested using targeted RNAseq (Sureselect, Agilent) and QURNAs data-analysis. All patients were checked for the highest ERS event(s) (Table 3 and Supplementary Data 4 and 5) and a possible drop in the relative expression of the normal exon junctions (Fig. 1b (patient Und-01), 2a (other patients)). When possible, heterozygous single nucleotide variants present in the coding DNA sequence were compared with the RNA sequence reads, to detect potential allelic imbalance (Table 3). For six patients no enriched pathogenic splicing event in *NF1* nor *SPRED1* (Supplementary Fig. 6—only ERS for normal exon-junctions of *SPRED1*) could be detected based on high ERS, nor a prominent drop in ERS of the normal exon-exon junction(s). Also, comparing the DNA and RNA sequencing reads, no evident allelic imbalance was observed for these patients. Interestingly, for two patients we were able to identify a deep intronic pathogenic splice variant or and for three patients resolve the exact effect on RNA splicing for variants (VUS, likely pathogenic) previously detected using DNA diagnostics (Table 3).

For undiagnosed patient 1 (in the same run as the replication samples), two major enriched splicing events between exons 38 and 39 were detected. These events introduce a cassette exon of 177 bp in intron 38 (acceptor site: ERS 51.3 and 1170 reads and donor site ERS 20.4 and 273 reads). In addition, a minor alternative acceptor site was also seen in the QURNAs output, resulting in a larger cassette exon of 208 bp at the 5′end but having the same 3′end (ERS 1.4 and 31 reads). The inclusion of these novel cassette exons, results in a drop in ERS of the normal 38–39 exon-exon junction (Figs. 1b, 2b). These aberrant splicing events are caused by a deep intronic c.5749 + 332 A > G change, that is visible in the RNAseq reads (Supplementary Fig. 7). The presence of this variant was confirmed with Sanger DNA sequencing as a de novo pathogenic splice variant by testing the patient and parents. This variant introduces a strong splice donor site at this position and consequently the activation of the above mentioned cryptic splice acceptor sites (Fig. 2b). Both aberrant transcripts contain a predicted premature stop codon and are consequently expected to result in loss-of-function through either nonsense-mediated mRNA decay (NMD), or an abnormal protein product (Fig. 2b). Comparing various heterozygous single nucleotide variants shows modest allelic imbalance in the RNA sequence versus DNA, which indicates some NMD (Supplementary Fig. 7). Of interest, the DNA variant is described multiple times and splicing was characterized by sequence analyses of cDNA clones with identical effect (both major and minor event) on RNA-splicing[33], or similar effect on RNA-splicing (only major event) using direct cDNA sequencing[13,28].

Table 3. Overview of molecular undiagnosed patients.

Sample	Inclusion criterion RNAseq^	Sample type	Description *NF1* top event(s) QURNAs^	Description *NF1* RNA splicing, effect on protein and ACMG classification^	*NF1* DNA splice variant	Allelic imbalance
1	Clinical NF1 diagnosis, DNA negative *NF1/SPRED1*	short term cultured blood lymphocytes + puromycin	Inclusion of 177 (ERS 51.3&20.4) bp of intron 38 Inclusion of 208 (ERS 1.7&20.4) bp of intron 38. Both encoding a PSC	r.5749_5750ins5749 + 155_5749 + 331 r.5749_5750ins5749 + 124_5749 + 331 PS2,PS3,PM2,PP5 =>P p.(Ser1916_Ser1917ins12*) p.(Ser1916_Ser1917ins10*)	c.5749 + 332 A>G (de novo)	modest
2	Clinical NF1 diagnosis, likely pathogenic variant c.6580–2 A>G, characterization splicing effect PVS1, PM2 =>LP	short term cultured blood lymphocytes	Exon 43 skipping (ERS 8.5) & activation cryptic acceptor; inclusion of 17nt (7.1). Frameshift	r.6580_6641del; r.6579_6580ins6580 −17_6580–1 PSV1,PS3,PM2 =>P p.(Ala2194Ilefs*6) p.(Ala2194Serfs*9)	c.6580−2 A>G	No imbalance detected
3	NF1 suspicion, DNA negative *NF1/SPRED1*	short term cultured blood lymphocytes (+/− puro)	ERS < 4	—	—	N/A (only 1 heterozygous SNV)
4	NF1 suspicion/segmental NF1, DNA negative *NF1/SPRED1*	short term cultured blood lymphocytes	ERS < 4	—	—	N/A (only 1 heterozygous SNV)
5	Clinical NF1 diagnosis, DNA negative *NF1/SPRED1*	short term cultured blood lymphocytes	inclusion of 42 bp (ERS 28.1&18.1) of intron 11. In frame but introduces PSC	r.1260_1261ins1260 + 1605_1260 + 1646 PS3, PM1,2,PP3,5=>P p.(Asn420_Ser421ins3*)	c.1260 + 1604A>G	modest
6	Clinical NF1 diagnosis, VUS c.2252 G>T (exon 19) PM1,2,PP3, BP1	Cultured fibroblasts	Exon 19 skipping (ERS 4.2) Frameshift	r.2252_2325del PS3,PM1,2,PP1,3,BP1 =>P p.(Arg752Leufs*17)	c.2252 G>T; p.(Gly751Val)	high
7 (WT)	WT	Cultured fibroblasts	ERS < 4	—	—	N/A
8	Clinical NF1 diagnosis, VUS *NF1* c.556 G>T PM2,PP3,BP1=>VUS	RNA directly isolated from pheripheral blood lymphocytes	inclusion of cryptic exon between exons 4 and 5, in combination with exon 5 skipping (ERS 33.9&18.8); Exon 5 skipping alone (ERS 3.6) Frameshift	r.480_586delins479 + 673_479 + 703 r.480_586del PS3,PM2,PP3,BP1 =>LP p.(Arg160Serfs*20) p.(Leu161Asnfs*4)	c.556 G>T; p.(Asp186Tyr)	No imbalance detected
9	NF1 suspicion, DNA negative *NF1/SPRED1*. Mother clinical diagnosis, segmental NF1, negative *NF1/SPRED1/GNAS*	short term cultured blood lymphocytes (+/− puro)	ERS < 4	—	—	N/A (no heterozygous SNPs)
10	NF1 suspicion, DNA negative *NF1/SPRED1*	short term cultured blood lymphocytes (+/− puro)	ERS < 4	—	—	No imbalance detected
11	NF1 suspicion, *NF1/SPRED1* and whole exome sequencing negative	short term cultured blood lymphocytes (+/− puro)	ERS < 4	—	—	N/A
12	Clinical NF1 diagnosis, DNA negative *NF1/SPRED1*	short term cultured blood lymphocytes	ERS < 4	—	—	No imbalance detected

These were suspected for, or having a clinical NF1 diagnosis without (clear) molecular diagnosis. Inclusion criteria and top event(s) determined with QURNAs are shown. RNA was enriched for *NF1* and *SPRED1* using hybridization based enrichment (Agilent SureSelect) for all except UND1. ^P= pathogenic; LP=likely pathogenic; VUS=variant of unknown significance; PSC=predicted stop codon. Variants are classified (using PVS1=pathogenic very strong 1, PS1–4=pathogenic strong 1–4, PM1–6=pathogenic moderate 1–6, PP1–5=pathogenic supportive 1–5 & BP1= benign supportive 1 scores) according to ACMG standards and guidelines for the interpretation of sequence variants[REF18]. Clinical NF1 diagnosis = according to NIH consensus criteria. NF1 suspicion = did not meet minimal NIH consensus criteria[REF4].

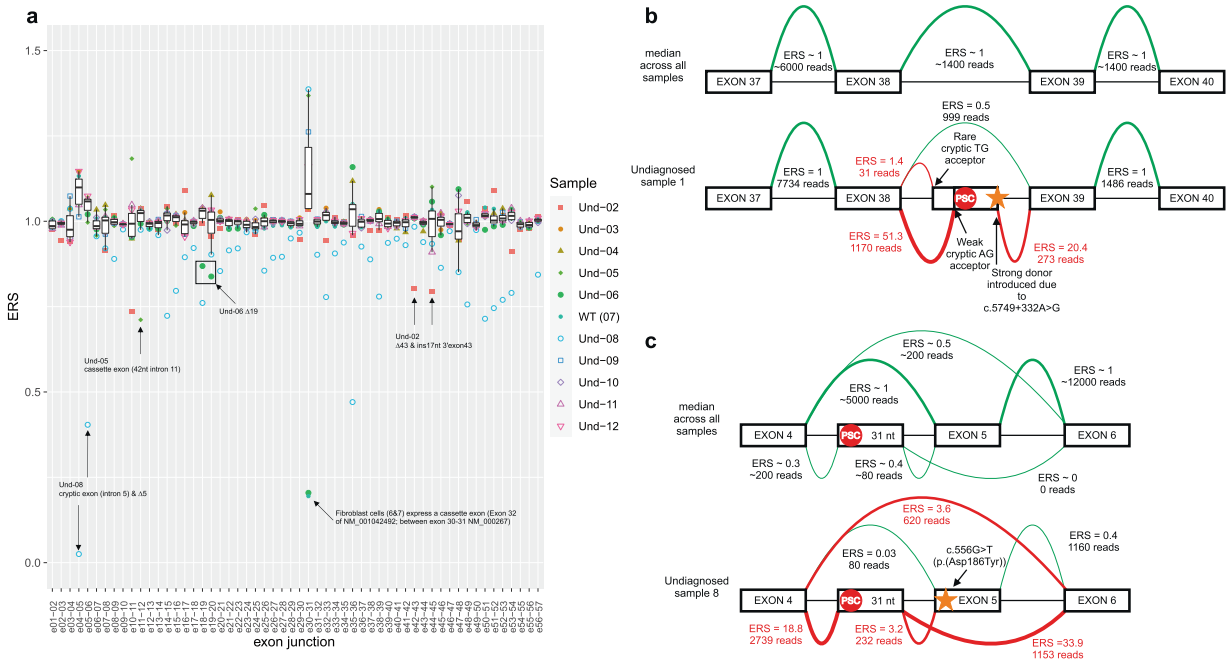

Fig. 2 For all patient derived samples with a pathogenic splicing event a concomitant drop of the normal exon junction in close proximity is visible. Enrichment score of normal exon-exon splice junctions of the undetermined cohort (**a**), annotated in the figure are the drop in ERS due to the detected pathogenic splicing event. Normal reference exon splice junctions are expected to have a median ERS ~1. center line of the boxplot represents the median value (Q2), bounds of box are Q1 and Q3 respectively and whiskers Q1–1.5xIQR and Q3 + 1.5xIQR respectively. Outliers are depicted with color and shape matching the specific sample (see inner figure legend). Splicing of undetermined 1 and 8 explained (**b**/**c**). Shown are the median ERS and reads for all samples within the same run vs the actual ERS and reads of the sample indicated; undiagnosed 1 (**b**) and undiagnosed 8 (**c**).

Patient 2 was included with a likely pathogenic (class 4) variant c.6580–2 A > G, which is predicted to show loss of the acceptor splice site that could lead to exon skipping. The splicing predictions however also indicate multiple cryptic acceptor sites upstream (Supplementary Fig. 8) that could be alternatively used. Experimental analysis showed exon 43 skipping (ERS 8.5, 136 reads) and insertion of 17nt intron sequence through use of the first upstream cryptic acceptor (ERS 7.1, 97 reads) was observed (Supplementary Fig. 8), leading to a drop in ERS of the normal exon-exon junction (Fig. 2a). The variant is visible in the RNAseq reads (Supplementary Fig. 8). No allelic imbalance was observed (data not shown) despite absence of an NMD inhibitor in the culture, suggesting that the truncated protein is synthesized and not degraded at the mRNA level by NMD. Based on these findings the c.6580–2 A > G variant is now considered pathogenic (class 5).

For patient 5, a cassette exon including 42 bp of intron 11 was detected (acceptor site ERS 18.1 and 1497 reads, donor site ERS 28 and 2278 reads), which introduces a premature stop codon. (Supplementary Fig. 9). This event was found based on high ERS (Table 3, Supplementary Data 5) and a drop in ERS of the normal exon-11–12 junction (Fig. 2a). The aberrant transcript is caused by a c.1260 + 1604A > G change that is visible in the RNAseq reads (Supplementary Fig. 9) and confirmed with Sanger DNA sequence analysis. Comparing various single nucleotide variants shows modest allelic imbalance in the RNA sequence versus DNA sequencing reads, which indicates some NMD (Supplementary Fig. 9). Interestingly, the pathogenic splicing variant c.1260 + 1604A > G is reported multiple times in the literature and found to have an identical effect on splicing[14,34,35].

Patient 6 has a clinical diagnosis of NF1 and a variant of unknown significance (VUS) c.2252 G > T (p.(Gly751Val)) in exon 19 of *NF1*. Her child also has the same variant and a clinical NF1 diagnosis. The glycine is conserved between species, the variant is not present in gnomAD and has, as annotated by Alamut strong disease causing in silico predictions. These predictions indicate partial loss of the (strong) acceptor site and introduction of a new donor site, potentially resulting in out-of-frame exon 19 skipping (Supplementary Fig. 10). For this patient, RNA from short term cultured fibroblast was used, without NMD inhibition. QURNAs indicated exon 19 skipping, which is a naturally occurring event, hence the ERS is only 4.2 (covered by 4155 reads). The drop in ERS of the normal exon 18–19 junction is prominent though (Fig. 2a, Supplementary Fig. 10). Of interest, comparing various single nucleotide variants, including c.2252 G > T, showed almost a complete loss of one allele in the RNA sequence versus DNA sequencing reads, suggesting NMD of the variant allele (Supplementary Fig. 10). This variant was previously also reported in another NF1-patient[36]. Together, these data support the classification of this variant as pathogenic.

For patient 8, RNA was directly isolated from peripheral blood lymphocytes, in contrast to the other samples in this run. This patient has a clinical diagnosis of NF1 and a VUS (class 3 variant) c.556 G > T (p.(Asp186Tyr)) in exon 5 of *NF1*. The aspartic acid is conserved between species and the variant is not present in gnomAD, as well as having strong disease causing in silico predictions (as annotated by Alamut). Splicing predictions show no effect on splicing through the introduction of a novel donor or acceptor site. Further in silico assessment showed effect on exonic splicing enhancer elements, and a slightly higher chance of exon skipping than the wild type (WT) allele were predicted (Supplementary Fig. 11). Interestingly, literature suggests that the variant c.557 A > T immediately next to the one identified in this patient (c.556 G > T) is disrupting an exonic splicing enhancer element, thereby causing exon skipping[37]. Therefore this sample was included in this RNAseq study and found indeed to cause skipping of exon 5 only (ERS 3.6; 620 reads), though predominantly in combination with the inclusion of a cryptic intron between exons 4 and 5 (acceptor ERS 33.9 and 1153 reads and donor ERS 18.8 and 2739 reads; see Fig. 2b). Both events are naturally occurring albeit that the drop in ERS of exon boundaries 4–5 and 5–6 for this patient is more prominent compared to the others in the same run (Fig. 2a).

The inclusion of the cryptic exon in combination with exon 5 skipping is most likely the consequence of this being an uncultured sample, since culturing of samples results in much lower levels of the cryptic exon (see Supplementary Data 2). Of interest, comparing the RNA versus the DNA sequencing reads showed an incomplete loss of the alternative T allele of the c.556 G > T variant (19% T allele present), caused by partial skipping of this exon. The other three informative single nucleotide variants do not indicate allelic imbalance (Supplementary Fig. 11), which suggests that the premature stop does not subject the transcript to strong NMD. In addition, residual presence of the missense variant p.(Asp186Tyr) could potentially have a functional effect. Together, these data support the classification of this variant as likely pathogenic. Further segregation and functional analysis should be performed to support pathogenicity.

DISCUSSION

Alternative splicing is an important mechanism wherein different mRNAs are generated from the same gene, thus increasing the coding capacity. It is often regulated in a tissue- or developmental-specific manner. Deregulation of this mechanism can be caused by DNA variants in regulatory regions of RNA splicing, resulting in aberrant splicing. Skipping of exons and/or the introduction of a premature stop codon can lead to loss-of-function or even complete loss of the encoded protein from the variant allele. Several naturally occurring alternative transcripts have been described for the *NF1* gene as well as aberrant splicing events due to DNA variants in this gene. In fact, a high percentage of up to ~30% of pathogenic *NF1* DNA variants have an effect on RNA splicing. Targeted RNA sequencing is currently more cost efficient compared to whole transcriptome for in depth analysis of the RNA splicing of a small set of genes. Previously, we used QURNAs to successfully detect *BRCA1* and *BRCA2* aberrant and normal splicing events in a small RNAseq data set[24]. Here we show that targeted RNAseq, in combination with QURNAs data-analysis (Supplementary Fig. 1), is a powerful approach to detect and distinguish normal and aberrant, pathogenic *NF1* splicing events.

Not only could targeted RNAseq confirm previously reported and/or predicted changes in RNA splicing for all validation and replication samples, additional splicing events caused by the same DNA variant, but not (easily) resolved using RT-PCR, could be detected (Tables 1, 2). In these samples, with RNA extracted from lymphocytes, all normal exon-exon junctions and rare normal events previously reported in blood derived RNA could also be detected, except the neuronal tissue specific inclusion of a cassette exon of 30nt intron 11 ('ins 9a/9br'). Additionally, the use of two different enrichment methods gave similar results (Table 2, Supplementary Fig. 5).

After this thorough validation of our targeted RNAseq approach a total of 11 patients lacking a (clear) molecular diagnosis of NF1 was tested. Three patients had a clinical NF1 diagnosis and a *NF1* DNA variant suspected to have an effect on *NF1* splicing. The other eight patients showed some or clear NF1 symptoms, but lacked a molecular diagnosis through conventional DNA sequencing of *NF1* (and *SPRED1)*. (Likely) pathogenic *NF1* splicing events could be detected in five out of these 11 patients using RNAseq (45%). For three of them this was a confirmation of a (likely) pathogenic splicing event caused by a DNA variant suspected to have an effect on splicing. For two patients (of eight without a prior molecular diagnosis) the pathogenic splicing event was caused by a deep-intronic DNA variant (25%), that was previously not detected using standard DNA diagnostics.

For six patients, no aberrant, pathogenic *NF1*, nor *SPRED1* RNA splicing or expression could be detected. In most cases, bi-allelic gene expression could be confirmed, excluding the presence of a possible variant in the promoter region or transcription enhancer site that could result in strongly reduced or absent gene expression in the variant allele. To exclude other potential (rare) causes of NF1, the 5'UTR (up to c.−383) and the neuronal specific cassette exon of 30nt in intron 11 ('ins 9a/9br') of *NF1* were also sequenced complementary, since these regions are not covered by RNAseq, nor by our standard DNA diagnostics procedure. Variants in the 5'UTR of *NF1* creating or removing an upstream start codon (uAUG) and thus influencing translation can cause NF1[14,38]. However, no relevant variants were detected in these 2 regions.

In addition, retrotransposon insertions, that cause altered NF1 splicing, account for ~0.4% of all pathogenic *NF1* variants[17]. These are not targeted using conventional DNA diagnostics and not detected using RNAseq due to the exclusion of chimeric reads in in the RNAseq data analysis pipeline. However, these insertions can be picked up by Sanger based RT-PCR RNA sequencing, even though precise characterization of them requires additional customized efforts[17]. Another residual risk of missing variants in blood based DNA and RNA diagnostics is the presence of a mosaic postzygotic *NF1* pathogenic variant, in specific affected tissue, but absent in blood. This might be the case for patient 4 (Table 3) who presented with a segmental/mosaic NF1 phenotype restricted to the skin. Unfortunately, this affected tissue was not available for further DNA or RNA analysis. Finally, an NF1 (like) clinical phenotype, might also be explained by molecular defects in other genes, causing syndromes with overlapping phenotypes. These are, for example, Noonan syndrome (MIM PS163950; multiple genes/loci), Constitutional mismatch repair deficiency (MIM 276300; mismatch repair genes) and Proteus syndrome (MIM 176920; *AKT1*)[1,2,5].

The QURNAs algorithm facilitated the identification of significant changes in RNA splicing. It uses an unbiased approach to identify and characterize aberrant and novel splice junctions without prior assumptions about transcript isoforms. The pipeline identifies all splice events present in the data and calculates their enrichment score and statistical significance compared to other samples within the same run. Together with a generic targeted RNAseq procedure, the QURNAs method can be used in routine diagnostics to discover, confirm or exclude effect of genetic variants on RNA splicing in *NF1* or other genes of interest. RNAseq can potentially replace DNA based *NF1* diagnostics as a primary test to detect clinically relevant variant. However, this requires the availability of reliable SNV calling on RNAseq data in the laboratory. Also, in patients for whom SNV calling may be compromised by alternative splicing events, e.g. exon skipping resulting in lower variant allele coverage, or degradation of the variant allele by NMD, DNA based diagnostics should still be run complementary to detect the causative variant.

In conclusion, the presented targeted RNAseq approach in combination with the computational QURNAs pipeline successfully detects and quantifies pathogenic *NF1* splicing events driven by (deep) intronic or missense variants.

Equations and mathematical expressions

$$I^{spl} = \frac{r^{spl}}{s^{spl}} \tag{1}$$

$$m_i^{spl} = \frac{I_i^{spl}}{\sum_{j \neq i} \frac{I_j^{spl}}{n-1}} \tag{2}$$

$$p(spl) = \text{prop} - \text{test}\left(\frac{\sqrt{r_i^{spl}}}{\sqrt{s_i^{spl}}}\right) \tag{3}$$

$$p(spl_i) = \text{prop} - \text{test}^{\text{one-tail}}\left(\frac{\sqrt{r_i^{spl}}}{\sqrt{s_i^{spl}}}, \frac{\sqrt{\sum_{j \neq i} r_j^{spl}}}{\sqrt{\sum_{j \neq i} s_j^{spl}}}\right) \tag{4}$$

METHODS

Ethics approval and consent to participate

All patients provided written or oral consent for either coded use of their material to study the clinical utility of RNAseq according to their diagnostic question or anonymous use of their material for improving diagnostic testing in general. The institutional Medical Ethics Committee (medisch-ethische toetsingscommissie or METC) of the Maastricht University Medical Center confirmed that the Medical Research Involving Human Subjects Act (WMO) does not apply to our study (METC 2021–2666) and that an official approval of this study by the committee is not required.

Patient derived cells

White blood cells were isolated using ficoll (Sigma-Aldrich) from fresh whole blood (collected in EDTA tubes) and used either fresh or frozen in FCS with 10% of DMSO for subsequent short term-culture in a complete medium consisting of RPMI 1640 supplemented with 12.5% FCS, 1x L-glutamine (Gibco), 0.8 mM sodiumpyruvate (Gibco), 17 mM Hepes buffer (Gibco), $4.2 \times 10{-}2$ mM 2-mercaptoethanol (Gibco), 42 units/ml penicillin–streptomycin and 0.21 g/ml amphotericin B solution (Sigma-Aldrich). Lymphocyte growth was stimulated with 50 µl/ml PHA (Gibco) and 10 units/ml of IL-2 (Roche). At day 7, 4–6 h before harvesting the cells, each culture was split evenly and one part was treated with 200 µg/ml of puromycin (Sigma-Aldrich) as an inhibitor of nonsense mediated RNA decay (NMD)[39].

For the initial validation of the targeted RNAseq procedure, RNA was isolated from uncultured peripheral blood lymphocytes of nine patients from the pre-implantation genetic testing (PGT) program at our institution, who had a pathogenic *NF1* variant that leads to aberrant splicing (Table 1). Pheriphal blood lymphocytes were isolated and stored in DMSO in liquid nitrogen within 24 h after collecting the EDTA blood. Two additional samples from anonymous patients without a clinical diagnosis of NF1 were included as wild type controls. These validation samples were enriched for *NF1* RNA using NuGEN SPET for RNA (see RNA isolation and Target enrichment/Library preparation).

The diagnostic performance of targeted RNAseq was further assessed using RNA samples shipped from the Centre for Medical Genetics, Ghent University Hospital, Ghent, Belgium (Table 2). RNA was isolated from blood lymphocytes (collected in EDTA tubes) that were short-term cultured. The molecular results of these samples were initially blinded, to test the ability to detect pathogenic *NF1* splicing events without prior knowledge, comparable to the diagnostic setting. These replication cohort samples were enriched for *NF1* using both Nugen SPET for RNA and for *NF1*/*SPRED1* using Agilent SureSelect RNA Target Enrichment (see RNA isolation and Target enrichment/Library preparation). This allowed a direct performance comparison of these two different commercial platforms.

To further validate the targeted RNAseq approach and show its added value for the detection of pathogenic splicing events complementary to DNA diagnostics, an additional set of twelve patients was selected (Table 3). Eight patients were included because of having (or suspected of having) a clinical NF1 diagnosis without molecular confirmation. While two patients had a *NF1* DNA variant with unclear effect on RNA splicing (variant of unknown significance – VUS [class 3]) and one patient had an *NF1* DNA variant likely to affect splicing (likely pathogenic – LP [class 4]). One additional sample from an anonymous patient without a clinical diagnosis of NF1 was included. Available material from these patients, i.e. either lymphocytes directly isolated from blood or short-term cultured cells with or without NMD inhibition, was used to isolate RNA. Blood was collected in EDTA tubes and shipped and processed within a maximum of 48 h in our lab, except und8 that came into the lab > 48 h after blood draw. These samples were enriched for *NF1*/*SPRED1* using SureSelect RNA Target Enrichment (see RNA isolation and Target enrichment/Library preparation).

RNA isolation and target enrichment/library preparation

Total RNA was purified and DNAseI treated using RNeasy Plus micro columns (Qiagen). Purified RNA was used as input for the SureSelect RNA Target Enrichment (below) and for the Ovation cDNA module followed by custom Ovation target enrichment, based on single primer enrichment technology (SPET) for RNA, according to the protocol of the manufacturer (NuGEN). Primer design for *NF1* (NM_000267.3) was performed by the manufacturer for the coding regions and 5′/3′UTR with a probe spacing of 150 bp. Library preparation was according to the manufacturer instructions.

Additionally, the SureSelect RNA Target Enrichment for Illumina Paired-End Multiplexed Sequencing kit (Agilent; protocol version 2.2.1) was used to target the *NF1 (*NM_000267.3) and *SPRED1* (NM_152592.2) coding regions and 5′/3′UTR. Probe design was performed using the Agilent eArray online tool, setting the region of interest to minus 100nt of the normal 5′exon boundary and plus 150nt of the normal 3′exon boundary and 6x tiling of probes. Library preparation was according to the manufacturer's instructions. Primers (SPET) and/or probes (Sureselect) design is available upon request.

Paired-end sequencing (2x150bp) was performed on a NextSeq500 instrument (Illumina) using NextSeq500/550 high output kit v2 (Illumina).

Read alignment

RNAseq reads were mapped with the STAR mapper (Version 2.4.1d) and an index created from HS.GRCh37 with 92 bp overhang[40]. We performed unique mapping onto the genome without trimming and with max. 10 mismatches per read while no chimeric reads were allowed (default STAR settings). To elucidate the PCR amplification bias, de-duplication was performed either using a build in Unique Molecular Identifier (UMI) in the probe design (NuGen SPET) or by removing identical reads (Agilent Sureselect). Duplicate read pair detection was carried out with Picard tools (https://github.com/broadinstitute/picard). Start and end positions from STAR output refer to the first nucleotide in the intron (AG | **g**u) and last nucleotide of the intron (a**g** | G), respectively.

Quantitative enrichment of aberrant splicing events in targeted RNAseq—QURNAs

The QURNAs computational pipeline was developed to identify sample-specific splicing events in targeted RNAseq data, including novel and possibly aberrant transcripts or increased abundance of normal isoforms (see Supplementary fig. 1). To minimize the number of possible artefacts or non-relevant events in the QURNAs data analysis, it is important to perform QURNAs analysis on a group of at least 4 samples that are equally processed. This way, events that are either induced by sample specific treatment, and/or are more or less equally present in all samples from the group, get a neutral enrichment score of ~1 in the QURNAs output after normalization and inter-sample comparison (see below for methodology). QURNAs is available at https://dataverse.nl/dataset.xhtml?persistentId=hdl:10411/LY8ZQ4. The mapping of short sequences to the reference genome without prior assumptions about transcript architecture, enables statistical evaluation of significance of observed RNA isoforms.

De novo prediction of splice junctions from RNAseq data

QURNAs first creates a collection of splice sites based on the reads mapped onto the genome (BAM file). For these splice sites, reads in BAM files that overlap with the splice site are counted. While iterating over all reads in the region, two additional features are counted to estimate the fraction of the reads in the region congruent with the splice junction. First, the number of reads is calculated that overlap with and contain the last nucleotide of the exon at the donor site. This includes, next to reads with a different acceptor site, "unbroken" reads corresponding to intron retention events. The second count consists of reads that "span" the splice site, thus start upstream of the nucleotide and end downstream, but do not necessarily contain the nucleotide next to the donor site in their sequence (*e.g.*, exon skipping events or splice events that use an alternative upstream donor site).

Enrichment score (ERS) calculation

Enrichment for a splicing event in a sample is calculated as follows. For each splice site, a relative read count I^{spl} is calculated that measures the fraction of aligned reads that are congruent with the splice site among all the reads that span the splice site. Formally, we count reads (r^{spl}) that are congruent with the splice *spl*=(start,end) *i.e.*, the genomic alignment of the read contains a gap between exact (start, end) positions, allowing for other insertions/deletions in the read. Reads for each of the splice r^{spl} are considered relatively to reads that *span* the splice site S^{spl} (spanning reads). These start their genomic alignment upstream of the donor site and end downstream of this site but do not necessarily share the same donor or acceptor sites. To minimize the impact of spurious reads and sequencing noise, pseudocounts of at least 10 reads and up to 1% of spanning reads (max(10,S^{spl} 1%)) are added to both r^{spl} and s^{spl}. The relative read count I^{spl} for each sample is then calculated using Eq. (1). In sample *i*, after

calculating the relative splice read count, enrichment of reads m_i^{spl} is calculated by comparing the relative read count to the average read counts from other samples using Eq. (2). To prevent samples with low sequencing coverage having inflated relative read counts due to pseudocounts (e.g. $l_i^{spl} = 1$ for the extreme case of absent coverage, i.e. $r_i^{spl} = s_i^{spl} = 0$), the enrichment m_i^{spl} is calculated using lr_i^{spl} without pseudocounts.

Statistical significance of enriched isoforms

To test whether splice isoforms are expressed at significantly higher levels in a sample, a two-tailed proportion test that tests the null hypothesis that proportions in several groups are the same is used. For each splice junction a group of trials as number of spanning reads (s^{spl}) that includes reads that support the junction r^{spl} (positive outcome of the trial) is defined. If significant ($p(spl) < 0.01$), the proportion test rejects the null hypothesis that the expression of the splice junction is the same in all samples. The test assumes independence between trials in the group and this is not the case for RNAseq due to read redundancy and overlap between 150-bp reads. The marginal information contributed by additional reads, drops with the number of reads mapped to a splice junction. To reflect lower marginal information, we use square root of RNAseq read numbers using Eq. (3) where i corresponds to subsequent samples. To determine the significance of a splice event in a specific sample i, the reads for the sample are compared to the background distribution of reads in remaining samples, testing the enrichment hypothesis (one-tailed test expecting enrichment) using Eq. (4). Note that in this statistical framework it is possible that $\boldsymbol{p}(spl) < \mathbf{0.01}$ but $\boldsymbol{p}(spl_i) > \mathbf{0.01}$ for all samples $\boldsymbol{i}$ (the distribution of reads is different between samples, but none are significant). This is, for example the case for WT splice isoforms in mutated samples where the WT isoform becomes depleted ($\boldsymbol{p}(spl_i)$ tests only for enrichment).

Interpretation QURNAs output

The QURNAs output for all splicing events was reduced with an additional GAP filter of >3 ((pos end_STAR – pos start_STAR) + 1), to exclude artefacts that do not represent splicing since the intron size would be smaller than 4 nucleotides. The absolute read counts of all normal exon splice-junctions in the reference transcript were inspected to confirm good sequence coverage throughout the gene for all samples. The ERS values of these reference exon splice junctions were subsequently checked for a possible drop (normal reference exon splice junctions are expected to have a median ERS ~1). Highly expressed alternative splicing events usually result in lower expression of the normal exon splice junctions in the same region, reflected by such a drop (ERS <0.8 or the lowest value of the total group of samples). Subsequently, all splicing events for a specific sample were sorted high to low based on the ERS. All events with ERS > 5 were further evaluated to discriminate possible artefacts from true alternative splicing events. Importantly, relevant ERS values can be lower than 5 when; 1) the splice variant gives rise to multiple different splicing events, instead of one or two major events; 2) the splicing events concerns an upregulated normal splicing event which is also present in other samples in the run at lower expression levels or 3) when 2 samples are present in the same run with the same effect on splicing. Alamut v2.15 (Interactive Biosoftware) was used to identify true splice-donor and acceptor sites based on in silico prediction. Regions with alternative splicing were further inspected in the BAM file in a genome browser (IGV and Alamut) to check for possible causal sequence variants in close proximity. Allelic imbalance was assessed by comparing heterozygous variants in the DNA vs the RNA sequences in IGV. A minimum of at least 2 heterozygous variants in the coding regions and/or UTRs was considered informative to assess allelic imbalance. Consistent variant allele ratio's in the sequence from 60/40% to 70/30% were considered as modest imbalance, 70/30% to 80/20% as intermediate and >80/20% as strong. But these boundaries are set arbitrarily for now and need further validation with more samples, and are also under the assumption that the imbalance is for the same allele. Since short-read sequencing was performed, the SNP phase is not known.

Nomenclature

The description of genetic variants follows the Human Genetic Variation Society (HGVS) approved guidelines[41], where c.1 (and r.1) is the A of the ATG translation initiation codon. Alternative splicing events are those incorporating splice junctions not present in the reference transcript NM_000267. Exon numbering used by QURNAs and in the tables is systematically 1–57 using NM_000267, for conversion according to the NF1 best practice meeting[42] this numbering is +1 for all exons after exon 30 (LRG_214).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The authors confirm that all data, except aligned sequence reads, supporting the findings of this study are available within the article and/or its supplementary materials. Aligned sequences reads (BAM files) are deposited at the European Variant Archive under accession code EGAD00001007978.

CODE AVAILABILITY

The in-house developed tool QURNAs is available at https://dataverse.nl/dataset.xhtml?persistentId=hdl:10411/LY8ZQ4.

Received: 6 May 2021; Accepted: 15 October 2021;
Published online: 15 November 2021

REFERENCES

1. Friedman, J. M. in GeneReviews((R)) (eds M. P. Adam et al.) (1993).
2. Ferner, R. E. et al. Guidelines for the diagnosis and management of individuals with neurofibromatosis 1. *J. Med. Genet.* **44**, 81–88 (2007).
3. Friedman, J. M., Gutmann, D. H., MacCollin, M. & Riccardi, V. M. (The Johns Hopkins University Press, Baltimore, MD, 1999).
4. Stumpf, D. et al. Neurofibromatosis. Conference statement. National Institutes of Health Consensus Development Conference. *Arch. Neurol.* **45**, 575–578 (1988).
5. Perez-Valencia, J. A. et al. Constitutional mismatch repair deficiency is the diagnosis in 0.41% of pathogenic NF1/SPRED1 variant negative children suspected of sporadic neurofibromatosis type 1. *Genet. Med.* https://doi.org/10.1038/s41436-020-0925-z (2020).
6. Cawthon, R. M. et al. Identification and characterization of transcripts from the neurofibromatosis 1 region: the sequence and genomic structure of EVI2 and mapping of other transcripts. *Genomics* **7**, 555–565 (1990).
7. Viskochil, D. et al. Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell* **62**, 187–192 (1990).
8. Wallace, M. R. et al. Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science* **249**, 181–186 (1990).
9. Trovo-Marqui, A. B. & Tajara, E. H. Neurofibromin: a general outlook. *Clin. Genet.* **70**, 1–13 (2006).
10. Stenson, P. D. et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136**, 665–677 (2017).
11. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
12. van Minkelen, R. et al. A clinical and genetic overview of 18 years neurofibromatosis type 1 molecular diagnostics in the Netherlands. *Clin. Genet.* **85**, 318–327 (2014).
13. Wimmer, K. et al. Extensive in silico analysis of NF1 splicing defects uncovers determinants for splicing outcome upon 5′ splice-site disruption. *Hum. Mutat.* **28**, 599–612 (2007).
14. Evans, D. G. et al. Comprehensive RNA analysis of the NF1 gene in classically affected NF1 affected individuals meeting NIH criteria has high sensitivity and mutation negative testing is reassuring in isolated cases with pigmentary features only. *EBioMedicine* **7**, 212–220 (2016).
15. Messiaen, L. M. et al. Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Hum. Mutat.* **15**, 541–555 (2000).
16. Valero, M. C. et al. A highly sensitive genetic protocol to detect NF1 mutations. *J. Mol. Diagn.* **13**, 113–122 (2011).
17. Wimmer, K., Callens, T., Wernstedt, A. & Messiaen, L. The NF1 gene contains hotspots for L1 endonuclease-dependent de novo insertion. *PLoS Genet.* **7**, e1002371 (2011).
18. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).

19. Vreeswijk, M. et al. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. *Hum. Mutat.* **30**, 107–114 (2008).
20. Brandão, R. D., van Roozendaal, K., Tserpelis, D., Gomez Garcia, E. & Blok, M. J. Characterisation of unclassified variants in the BRCA1/2 genes with a putative effect on splicing. *Breast Cancer Res. Treat.* **129**, 971–982 (2011).
21. Sangermano, R. et al. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res.* **28**, 100–110 (2018).
22. Houdayer, C. et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum. Mutat.* **33**, 1228–1238 (2012).
23. Taggart, A. J., DeSimone, A. M., Shih, J. S., Filloux, M. E. & Fairbrother, W. G. Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. *Nat. Struct. Mol. Biol.* **19**, 719–721 (2012).
24. Brandao, R. D. et al. Targeted RNA-seq successfully identifies normal and pathogenic splicing events in breast/ovarian cancer susceptibility and Lynch syndrome genes. *Int. J. Cancer* **145**, 401–414 (2019).
25. Danglot, G. et al. Neurofibromatosis 1 (NF1) mRNAs expressed in the central nervous system are differentially spliced in the 5′ part of the gene. *Hum. Mol. Genet.* **4**, 915–920 (1995).
26. Gutmann, D. H., Zhang, Y. & Hirbe, A. Developmental regulation of a neuron-specific neurofibromatosis 1 isoform. *Ann. Neurol.* **46**, 777–782 (1999).
27. Geist, R. T. & Gutmann, D. H. Expression of a developmentally-regulated neuron-specific isoform of the neurofibromatosis 1 (NF1) gene. *Neurosci. Lett.* **211**, 85–88 (1996).
28. Ars, E. et al. Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum. Mol. Genet.* **9**, 237–247 (2000).
29. Pros, E. et al. Nature and mRNA effect of 282 different NF1 point mutations: focus on splicing alterations. *Hum. Mutat.* **29**, E173–E193 (2008).
30. De Conti, L., Skoko, N., Buratti, E. & Baralle, M. Complexities of 5′splice site definition: implications in clinical analyses. *RNA Biol.* **9**, 911–923 (2012).
31. Vandenbroucke, I., Vandesompele, J., De Paepe, A. & Messiaen, L. Quantification of NF1 transcripts reveals novel highly expressed splice variants. *FEBS Lett.* **522**, 71–76 (2002).
32. Messiaen, L. M. et al. Exon 10b of the NF1 gene represents a mutational hotspot and harbors a recurrent missense mutation Y489C associated with aberrant splicing. *Genet. Med.* **1**, 248–253 (1999).
33. Perrin, G., Morris, M. A., Antonarakis, S. E., Boltshauser, E. & Hutter, P. Two novel mutations affecting mRNA splicing of the neurofibromatosis type 1 (NF1) gene. *Hum. Mutat.* **7**, 172–175 (1996).
34. Sabbagh, A. et al. NF1 molecular characterization and neurofibromatosis type I genotype-phenotype correlation: the French experience. *Hum. Mutat.* **34**, 1510–1518 (2013).
35. Serra, E. et al. Schwann cells harbor the somatic NF1 mutation in neurofibromas: evidence of two different Schwann cell subpopulations. *Hum. Mol. Genet.* **9**, 3055–3064 (2000).
36. Stella, A. et al. Accurate Classification of NF1 Gene Variants in 84 Italian Patients with Neurofibromatosis Type 1. *Genes* **9**, https://doi.org/10.3390/genes9040216 (2018).
37. Zatkova, A. et al. Disruption of exonic splicing enhancer elements is the principal cause of exon skipping associated with seven nonsense or missense alleles of NF1. *Hum. Mutat.* **24**, 491–501 (2004).
38. Whiffin, N. et al. Characterising the loss-of-function impact of 5′ untranslated region variants in 15,708 individuals. *Nat. Commun.* **11**, 2523 (2020).
39. Andreutti-Zaugg, C., Scott, R. J. & Iggo, R. Inhibition of nonsense-mediated messenger RNA decay in clinical samples facilitates detection of human MSH2 mutations with an in vivo fusion protein assay and conventional techniques. *Cancer Res.* **57**, 3288–3293 (1997).
40. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
41. den Dunnen, J. T. & Antonarakis, S. E. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum. Mutat.* **15**, 7–12 (2000).
42. Messiaen, L. in Multidisciplinary Approach to Neurofibromatosis Type 1 (eds G. Tadini, E. Legius, & H. Brems) (Springer International Publishing, 2020).

ACKNOWLEDGEMENTS

We thank Roel Brandts for excellent technical assistance in *NF1* DNA and RNA diagnostics.

AUTHOR CONTRIBUTIONS

Conceptualization: M.v.G., M.J.B.; Data curation: R.K., C.E.P.v.R., C.N.v.O.; Formal Analysis: D.T., R.K.; Funding acquisition: H.G.B., A.v.d.W., M.v.G., M.J.B.; Investigation: R.K., M.J.B.; Methodology: R.S., R.K., R.D.B., M.J.B.; Project administration: R.K., D.T., M.J.B.; Resources: K.B.M.C., A.D.C.P., M.S., M.V., V.v.d.S., C.T.R.M.S., M.P.G.B., L.S., M.C.J.J., S.A.J.L.O., A.S.P., M.M.-I., F.A.D., M.J.L.; Software: R.S., K.W.J.D.; Supervision: H.G.B., A.v.d.W., M.v.G., M.J.B.; Validation: K.B.M.C., R.K., M.J.B.; Visualization: R.K., M.J.B.; Writing—original draft: R.K., M.J.B.; Writing—review & editing: R.K., R.D.B., K.B.M.C., A.D.C.P., V.v.d.S., M.C.J.J., S.A.J.L.O., A.S.P., M.M.-I., F.A.D., M.J.L., R.S., K.W.J.D., H.G.B., A.v.d.W., M.v.G., M.J.B.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-021-00258-w.

Correspondence and requests for materials should be addressed to M. J. Blok.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.