Original Article

Parallel DNA/RNA NGS Using an Identical Target Enrichment Panel in the Analysis of Hereditary Cancer Predisposition

(RNA / DNA / parallel / sequence capture / NGS / CZECANCA / hereditary cancer predisposition / germline genetic testing / alternative splicing / aberrant splicing / gene expression / reproducibility / deep intronic variant / *CHEK2* / *BRCA2* / *ATM* / *BRCA1* / *TSC2*)

PETRA KLEIBLOVÁ[1,2], MARTA ČERNÁ[1], PETRA ZEMÁNKOVÁ[1,3], KATEŘINA MATĚJKOVÁ[1,4], PETR NEHASIL[1,3,5], JAN HOJNÝ[6], KLÁRA HORÁČKOVÁ[1], MARKÉTA JANATOVÁ[1], JANA SOUKUPOVÁ[1], BARBORA ŠŤASTNÁ[1,7], ZDENĚK KLEIBL[1]

[1]Institute of Medical Biochemistry and Laboratory Diagnostics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic
[2]Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic
[3]Institute of Pathological Physiology, First Faculty of Medicine, Charles University, Prague, Czech Republic
[4]Department of Genetics and Microbiology, Faculty of Science, Charles University, Prague, Czech Republic
[5]Department of Paediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic
[6]Institute of Pathology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic
[7]Department of Biochemistry, Faculty of Science, Charles University, Prague, Czech Republic

Received March 16, 2024. Accepted March 25, 2024.

This study was supported by the Ministry of Health of the Czech Republic grant projects (NU23-03-00150 and RVO-VFN 64165); Charles University research projects (SVV 260516 and Cooperatio); and the Ministry of Education, Youth and Sports of the Czech Republic grant (Programme EXCELES, ID Project No. LX22NPO5102 – Funded by the European Union – Next Generation EU).

Corresponding author: Petra Kleiblová, Laboratory of Oncogenetics, Institute of Medical Biochemistry and Laboratory Diagnostics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Katerinska 1660/32, Prague 2, 128 00, Czech Republic. E-mail: pekleje@lf1.cuni.cz

Abbreviations: ACMG – American College of Medical Genetics and Genomics, AMP – Association for Molecular Pathology, BAM – binary alignment map, BCV – biological coefficient of variation, bp – base pair, CZECANCA – CZEch CAncer paNel for Clinical Application, GTEx – genotype-tissue expression, NGS – next-generation sequencing, NMD – nonsense-mediated decay, PCR – polymerase chain reaction, pre-mRNA – precursor mRNA, qCML – quantile-adjusted conditional maximum likelihood, RIN – RNA integrity number, RNA-seq – RNA sequencing, RT-PCR – reverse transcription PCR, TPM – transcripts per million, VCF – variant call format, VUS – variant of uncertain significance, WES – whole-exome sequencing.

Abbreviations for 226 genes listed in the manuscript are provided in Supplementary Table S1.

Abstract. Germline DNA testing using the next-generation sequencing (NGS) technology has become the analytical standard for the diagnostics of hereditary diseases, including cancer. Its increasing use places high demands on correct sample identification, independent confirmation of prioritized variants, and their functional and clinical interpretation. To streamline these processes, we introduced parallel DNA and RNA capture-based NGS using identical capture panel CZECANCA, which is routinely used for DNA analysis of hereditary cancer predisposition. Here, we present the analytical workflow for RNA sample processing and its analytical and diagnostic performance. Parallel DNA/RNA analysis allowed credible sample identification by calculating the kinship coefficient. The RNA capture-based approach enriched transcriptional targets for the majority of clinically relevant cancer predisposition genes to a degree that allowed analysis of the effect of identified DNA variants on mRNA processing. By comparing the panel and whole-exome RNA enrichment, we demonstrated that the tissue-specific gene expression pattern is independent of the capture panel. Moreover, technical replicates confirmed high reproducibility of the tested RNA analysis. We concluded that parallel DNA/RNA NGS using the identical gene panel is a

robust and cost-effective diagnostic strategy. In our setting, it allows routine analysis of 48 DNA/RNA pairs using NextSeq 500/550 Mid Output Kit v2.5 (150 cycles) in a single run with sufficient coverage to analyse 226 cancer predisposition and candidate genes. This approach can replace laborious Sanger confirmatory sequencing, increase testing turnaround, reduce analysis costs, and improve interpretation of the impact of variants by analysing their effect on mRNA processing.

Introduction

Cancer is one of the leading causes of death worldwide (Sung et al., 2021). It is caused by DNA alterations in genes that regulate cellular and tissue homeostasis. While the majority of malignancies arise from somatic DNA variants, hereditary predisposition, caused by germline DNA variants in cancer predisposition genes, drives cancer development in approximately 5–10 % of cancer patients (Rahman, 2014; Kleibl and Kristensen, 2016). Identification of the germline pathogenic variant is of paramount importance for tailored cancer prevention in its asymptomatic carriers and, in addition, for guiding personalized cancer treatment in hereditary cancer patients (Stadler et al., 2021).

Germline genetic testing for cancer predisposition is routinely based on DNA analysis, for which next-generation sequencing (NGS) has become the gold standard (Kuzbari et al., 2023). Although whole-genome sequencing (WGS) or whole-exome sequencing (WES) approaches are now affordable, sequencing of selected gene panels including only tens to hundreds of clinically relevant or candidate cancer predisposition genes remains the most common strategy (LaDuca et al., 2020). The CZECANCA (CZEch CAncer paNel for Clinical Application) panel targeting 226 cancer predisposition/candidate genes has been the most widely used for hereditary cancer diagnostics in the Czech Republic with over 30,000 probands analysed since 2016 (Soukupova et al., 2018; Lhotova et al., 2020; Wieme et al., 2021; Horackova et al., 2022; Kral et al., 2023).

In the past, RNA-based mutation analysis was mainly performed by reverse transcription polymerase chain reaction (RT-PCR) to analyse coding regions of selected genes (Pohlreich et al., 2003), to evaluate the effect of a variant of uncertain significance (VUS) on pre-mRNA splicing (Havranek et al., 2015; Lhota et al., 2016), or to estimate the biological consequence (impact on RNA) of DNA variants prior to their subsequent analysis by functional assays (Kleiblova et al., 2019). Analysis of RNA at the whole transcriptome or mRNA level using NGS-based RNA sequencing (RNA-seq) is a widely used approach allowing addressing all these questions at the global transcriptome level (Hong et al., 2020). Unfortunately, for many tissues, the mRNA expression level of many tumour suppressors (from which the majority of hereditary cancer predisposition genes recruit) falls below one transcript per million (TPM) in RNA-seq of total RNA or mRNA samples, limiting their use in routine diagnostics. For instance, the expression levels of *BRCA1*, *BRCA2*, and *CHEK2*, the most frequently altered hereditary cancer predisposition genes in breast cancer patients, are 0.88, 0.20, and 1.26 TPM in blood leukocytes according to the Genotype-Tissue Expression (GTEx) Portal (2023-10-10; https://gtexportal.org/). To overcome this issue, enhanced resolution of low expressed genes could be achieved by RNA capture. In diagnostics, most RNA capture-based NGS has been limited to analysing a few dozen genes (Davy et al., 2017; Karam et al., 2020; Horton et al., 2022, 2024). RNA capture-based NGS covering hundreds of genes is more commonly used for research purposes only (Hojny et al., 2022; Struzinska et al., 2023). It could not only precisely identify alternative splicing events in a tissue-specific manner but also assess the effect of DNA variants on the pre-mRNA splicing (Davy et al., 2017; Farber-Katz et al., 2018; Lattimore et al., 2018, 2019; Brandão et al., 2019; Lopez-Perolio et al., 2019; Walker et al., 2019). Several bioinformatic tools are available to quantify variant (alternative or aberrant) splicing events (Schafer et al., 2015; Leman et al., 2020). These are mainly based on counting the number of variant splicing events in comparison to all splicing events (reference and variant) in a given mRNA region. Consequently, the proportion of aberrant splicing events driven by a specific germline DNA variant is reflected in the classification of VUS, as specified in the ACMG/AMP guidelines (Walker et al., 2023). Simultaneous analysis of DNA and RNA samples thus improves the diagnostic potential of DNA germline genetic testing in hereditary cancer predisposition (Horton et al., 2022, 2024).

In our approach, we implemented an identical target capture panel for parallel NGS testing of both DNA and RNA samples. We summarized the reproducibility and variability of parallel DNA/RNA testing and demonstrated practical utilization of our approach within hereditary cancer predisposition testing.

Material and Methods

Patients and samples

Pairs of DNA and total RNA samples isolated from peripheral blood leukocytes were obtained from 832 individuals undergoing germline genetic testing for cancer predisposition. Nasopharyngeal, skin and blood samples were obtained from healthy volunteers to test tissue-specific gene expression and technical reproducibility of the analysis. All participants provided written informed consent with genetic testing approved by the Ethics Committee of the General University Hospital in Prague and the study was performed in accordance with the Declaration of Helsinki. All participants were Europeans of the Czech origin.

DNA and RNA isolation

For DNA isolation, peripheral blood was drawn to K_2EDTA blood collection tubes (Becton Dickinson, Franklin Lakes, NJ), and the DNA was isolated from a 400 µl blood sample using a MagCore Genomic DNA Whole Blood Kit No. 101 in a MagCore Plus II instrument (RBC Biosciences, New Taipei City, Taiwan) according to manufacturer's instructions. For RNA isolation, peripheral blood was drawn to Tempus Blood RNA Tubes (ThermoFisher Scientific, Waltham, MA) and isolated using a Tempus Spin RNA Isolation Kit (ThermoFisher Scientific) according to manufacturer's instructions. Minor modifications included extended centrifugation time (50 min), increased centrifugation force (4,000 *g*) and extended elution time (10 min). For RNA isolation from nasopharyngeal epithelial cells, a nasopharyngeal swab (Copan, Murrieta, CA) was inserted to the nasopharynx and the tissue sample was collected by gentle rotation. The swab was immediately immersed in 600 µl of RLT buffer (Qiagen, Hilden, Germany) and the supernatant was collected by centrifugation at 400 *g* for 10 min. Total RNA was isolated using an RNeasy Mini Kit (Qiagen) according to manufacturer's instructions for purification of total RNA from cultured cells. RNA from the skin was isolated from small (< 100 µg) pendular skin tags cut by sterile scissors and immediately immersed in RNA later (Qiagen) and stored until processing according to the manufacturer. Approximately 30 µg of the tissue sample was homogenized by MagNA Lyser Green Beads (Roche, Basel, Switzerland) using a MagNa Lyser Benchtop Homogenizer (Roche), stabilized in QIAzol Lysis Reagent (Qiagen), and isolated using an RNeasy Mini Kit (Qiagen) according to manufacturer's instructions for purification of total RNA from the tissue. The concentration and quality of isolated RNA was measured in NanoDrop One (ThermoFisher Scientific), RIN was estimated in an Agilent 2100 instrument using an RNA 6000 Nano Kit for the selected subset of samples (Agilent, Santa Clara, CA).

NGS library preparation and sequencing

Targeted DNA library preparation was performed as described previously (Kral et al., 2023) with minor modifications that included the use of an updated KAPA EvoPlus Kit (Roche, Basel, Switzerland).

The targeted RNA library was prepared using the KAPA RNA HyperPrep Kit (Roche) as we described previously (Walker et al., 2019) with minor modifications that included reduction of the reaction volume to a half in the first steps of the library preparation (until the end of the second post-ligation clean-up). Briefly, 5 µl of total RNA (150–800 ng) was fragmented (6 minutes at 94 °C; targeted mean fragment length 180–200 bases). PCR amplifications were run for 9 and 11 cycles for pre- and post-hybridization PCR, respectively. A total of 40–48 barcoded samples were equimolarly pooled. Three µg of the pooled DNA pre-library sample was hybridized with custom-designed KAPA HyperChoice probes (CZECANCA v1.2; 226 cancer predisposition/candidate genes, 0.63 Mbp, Roche; Supplementary Table S1) (Soukupova et al., 2018) using a KAPA HyperCapture Reagent kit (Roche) and captured using a KAPA HyperCapture Bead kit (Roche).

To test the RNA capture-based approach with a much larger capture panel, we hybridized 24 barcoded pre-library samples (2 µg of total pooled DNA) with KAPA HyperExome Probes (Roche) targeting the whole exome (43 Mbp), followed by only eight cycles of post-hybridization PCR.

The final DNA and RNA libraries were paired-end sequenced in NextSeq 500 (for the CZECANCA panel) with a NextSeq 500/550 Mid Output Kit v2.5 (150 cycles, Illumina, San Diego, CA) or NovaSeq 6000 (for the CZECANCA panel or HyperExome panel) with a NovaSeq 6000 S1 Reagent Kit v1.5 (200 cycles, Illumina).

Bioinformatic analysis

For DNA NGS, fastq data were analysed as we described previously (Soukupova et al., 2018). For RNA NGS, fastq files were mapped to the human reference genome (hg19) by STAR aligner 2.5.2 in default settings (sjdbOverhang = 100) to generate BAM files (Dobin et al., 2013). PCR duplicates were removed using Picard tools v1.129. The BAM files were subsequently analysed by regtools (Cotto et al., 2023). For splice junction annotations, the results were analysed by the R package: SCANVIS (Agius et al., 2019) in R version 4.2. Splice junctions for an alternative or aberrant splicing analysis were included if they were identified in at least three unique reads per sample. BAM files were further processed by GATK 3.8 (McKenna et al., 2010) with the SplitNCigarReads tool to call variants with HaplotypeCaller to the variant call format (VCF). A kinship coefficient was computed with R package SNPRelate from final DNA/RNA VCF files (only single nucleotide substitutions in coding regions were considered) to identify paired DNA/RNA samples (Zheng et al., 2012).

Gene expression analysis

The annotated splice junction counts were used for the gene expression analysis of each sample. The highest count of wild-type (wt) junctions for the particular gene represented the number of its transcripts. The sum of all transcripts in a sample represented the input quantity. Expression of three monoexonic genes was assessed by the average coverage of the exon separately and was not used for input quantity estimation. To normalize different input quantities across the analysed samples, we normalized the absolute number of transcripts of particular genes to one million transcripts of 226 targeted genes.

Biological variability, reproducibility and statistical analysis

The biological variability of blood RNA samples and technical reproducibility of our analytical approach was based on the results of the gene expression analysis.

Biological variability was assessed by the analysis of gene expression considering all 226 targeted genes from 48 RNA samples analysed in a single sequencing run. Clustering analysis was conducted using the R package pheatmap (https://CRAN.R-project.org/package=heatmap) with the clustering method set to Ward D2 and distance function set to Manhattan. The biological variability of gene expression was analysed using gene-specific Tagwise dispersion estimated for each gene using R package edgeR (Chen et al., 2016). The dispersion is the biological coefficient of variation (BCV) squared (for example, if the gene expression typically varies by 20 % from replicate to replicate, its BCV is 0.2 and its dispersion is 0.04). EdgeR estimated dispersions from replicates using the quantile-adjusted conditional maximum likelihood method (qCML). Tagwise dispersion was plotted against logarithmized means of the number of transcripts across samples.

Inter-platform reproducibility was assessed by a comparison of sets with 24 identical sample libraries analysed in either NextSeq or NovaSeq instruments.

Inter-run reproducibility was assessed using a set of seven RNA samples analysed in duplicates (two independently prepared and sequenced libraries using the CZECANCA panel).

Inter-capture reproducibility compared NGS results obtained from sequencing of identical pre-libraries prepared from 17 RNA samples that were hybridized with either the CZECANCA capture panel (226 genes, 0.63 Mb) or HyperExome Probes (43 Mb). For direct comparison, the CZECANCA target was subtracted as a "virtual panel" from WES (HyperExome panel) data. The correlation was estimated using Spearman's rank correlation (r), since the distribution of all the datasets was non-normal (Shapiro and Anderson-Darling normality tests were used). The results were plotted in R using the ggpubr package (https://CRAN.R-project.org/package=ggpubr).

Results

We adopted targeted RNA NGS for routine analysis of paired DNA and RNA samples using the identical custom capture panel CZECANCA (targeting 226 genes), which so far has only been used for DNA analysis in routine settings (Soukupova et al., 2018). Each NGS set for the analysis included 40–48 DNA with paired RNA samples processed in parallel and analysed in a single run in NextSeq 500 using the NextSeq 500/550 Mid Output Kit v2.5 (150 cycles). The technical parameters of a representative sequencing run (Supplementary Table S2) were as follows: Q30 bases was 94.66 %, pass filter was 91.66 %, and a sequencing output was 25.83 Gbp. Approximately half (48.6 %) of the sequencing output was attributed to the RNA analysis. After removal of PCR duplicates (median 32.6 %; range 25.0–43.5 %), the RNA sequencing yielded 29,129,631 unique reads mapped to exon-exon junction regions for 48 analysed RNA samples. The mean of the unique reads covering the exon-exon junction per sample was 606,867 (range 211,667–1,009,961) and the majority of them (mean 95.61 %) covered wild-type exon-exon junctions (range 93.17–96.15).

Kinship analysis of DNA and paired RNA samples was used to confirm matched DNA/RNA samples from the same individual. The kinship coefficients for paired samples fluctuated close to the maximal kinship coefficient value of 0.5, while these coefficients for unpaired samples were significantly lower (Fig. 1A). Thus, the kinship coefficient analysis provides a reliable method identifying matched DNA and RNA samples for their parallel processing.

The expression of a gene of interest in a tested tissue is always the limiting factor for RNA-based analysis. To assess the expression levels of 226 targeted genes, we analysed up to 48 blood RNA samples (processed within a single NGS run). The input quantity (sum of all transcripts) per sample in a representative sequencing run ranged between 30,445–129,797 across 48 blood RNA samples (mean 88,628; median 93,628; Supplementary Table S3) and was used to normalize the number of transcripts per million for each sample in the run (Supplementary Table S4).

We also compared the expression of the analysed genes between blood, nasopharyngeal and skin RNA samples. The targeted genes were divided into four categories according to their expression level in the blood, skin and nasopharyngeal tissues as: high, moderate, low, and very low/no (reflecting their TPM > 5000, 1500–5000, 500–1500, and < 500, respectively; Fig. 1B–C, Supplementary Table S4). Reliable assessment of alternative/aberrant splicing events was possible only in transcripts detectable at high and moderate levels.

While the group of 55 clinically relevant cancer predisposition genes (representing the genes of primary interest for diagnostic purposes), 37 (67.3 %) showed sufficient expression (high or moderate) in blood RNA samples (Fig. 1B). Seven out of the remaining 18 genes with low or very low/no expression in the blood were sufficiently expressed in nasopharyngeal RNA (high: *CDH1*, *BMPR1A*, *EPCAM*; moderate: *RAD51C*, *KIT*, *MET*) and four in skin RNA (*MEN1*, *PALB2*, *CDK4*, *SUFU*). Thus, 47/55 (85.5 %) of the clinically relevant cancer predisposition genes could be assessed by RNA analysis of easily accessible tissues. On the other hand, RNA analysis for other tissue types would be necessary for several clinically relevant cancer predisposition genes. *CLSPN* and *BRCA2* had low detection levels both in the blood and in nasopharyngeal and skin samples, *POLD1*, *RAD51D*, and *CDKN2A* had very low/no expression in the blood with low expression in nasopharyngeal and skin RNA, and *RET*, *WT1*, and *HOXB13* had very low/no expression in all analysed tissue types.

To analyse the biological variability of blood RNA samples, we performed cluster analysis of normalized expression data of 217 genes (rows) in 48 RNA samples (columns) from a representative sequencing run (Fig. 1D, Supplementary Table S4). Six genes (*HNF1A*, *HOXB13*, *KCNJ5*, *LMO1*, *PHOX2B, PLA2G2A*) were

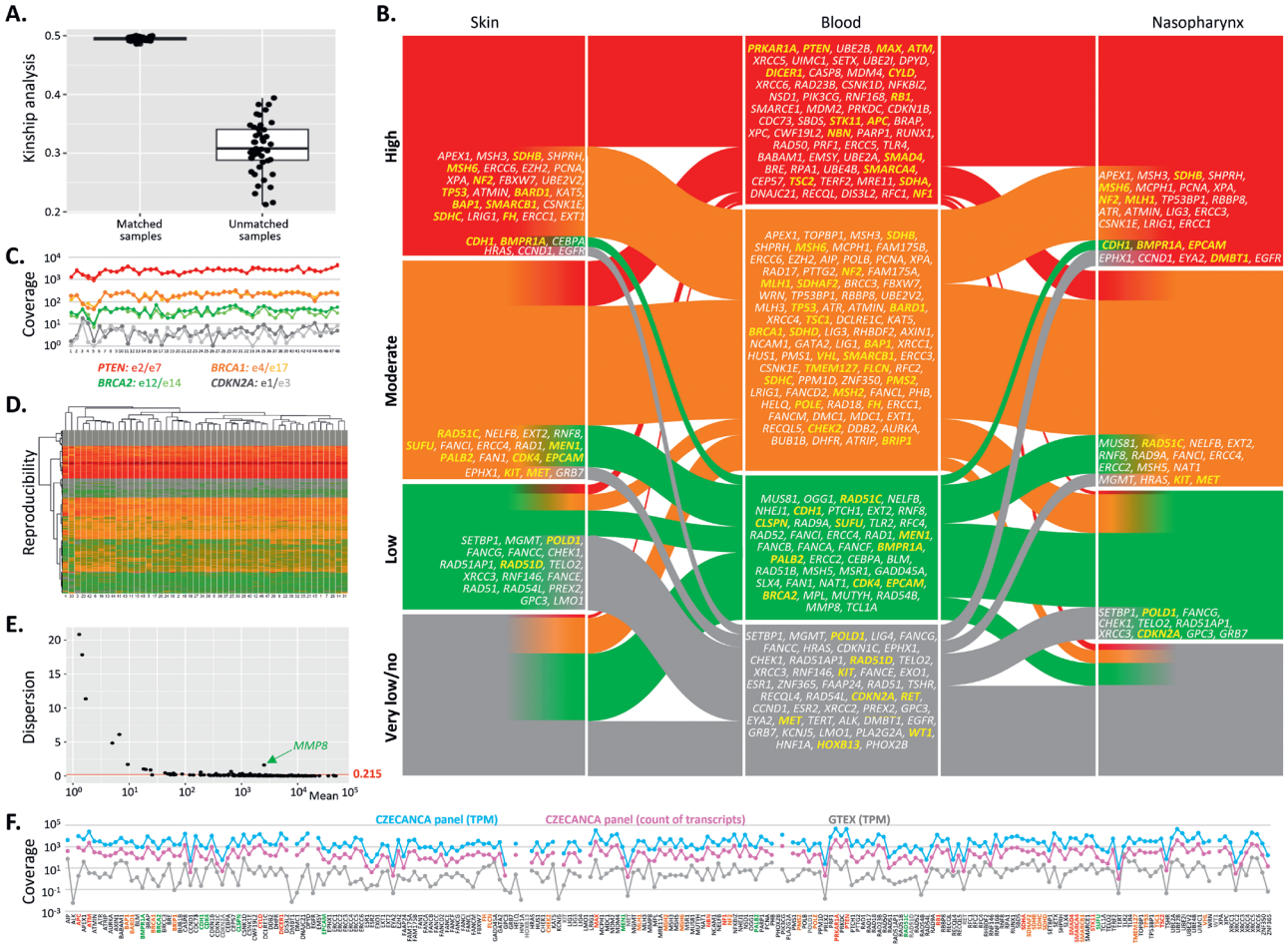

Fig. 1. (**A**) Sample pair identification by kinship coefficients for matched, paired DNA and RNA samples from 48 individuals analysed within a single sequencing run (left) and for unmatched DNA and RNA pairs (right) for comparison. (**B**) Sankey diagram depicting the expression levels of 226 genes targeted in the CZECANCA panel in peripheral blood (middle column) and their comparison to the skin (left column) and nasopharynx (right column). mRNA expression levels were categorized as high (red), moderate (orange), low (green), and very low/no (grey). All targeted genes are shown in the blood RNA (middle column). Only genes with increased expression in the skin and nasopharyngeal RNA (compared to blood) are shown in the corresponding columns. Yellow text highlights 55 clinically relevant cancer predisposition genes. (**C**) Mean absolute exon coverage of representative genes (two exons across 48 blood RNA samples from a single run) from high (*PTEN*), moderate (*BRCA1*), low (*BRCA2*), and very low/no (*CDKN2A*) expression categories (log mean coverage expresses logarithmized normalized coverage per million for selected exons). (**D–E**) Analysis of biological variability of 217 genes in 48 unrelated blood RNA samples analysed within a single representative sequencing run based on gene expression. Nine/226 genes were excluded from the analysis (three monoexonic and six unexpressed genes). (**D**) Unsupervised hierarchical clustering analysis of gene expression (rows) in individual RNA samples (columns). The colour scale represents the logarithmic number of gene transcripts per million (Supplementary Table S4). (**E**) Tagwise dispersion analysis of biological RNA replicates. The x-axis shows the gene expression (visualized as logarithmic mean number of transcripts for each of the 217 genes), the y-axis shows the dispersion coefficient. The red line indicates a dispersion of 0.215, a normal variability parameter for human RNA samples. (**F**) CZECANCA panel capture-based expression data (medians) in absolute (magenta) and normalized (blue) unique transcript counts compared to GTEx median expression in blood RNA (grey), predominantly generated by RNA-seq.

not uniformly expressed (expressed in 0–2 of 48 analysed samples) and were excluded from the analysis. Similarly, three monoexonic genes (*PTTG2*, *FANCF*, and *CEBPA*; missing exon-exon junctions required for the expression analysis) were also excluded from this analysis. Next, we used tagwise dispersion analysis (Fig. 1E) to quantify the biological variability. It showed that 193/217 (88.94 %) analysed genes did not exceed the biological variability threshold with a dispersion value < 0.215, which has been reported as a normal average value for non-cancer human samples (McCarthy et al., 2012; Yoon and Nam, 2017).

The majority of 25/217 (11.5 %) genes exceeding the normal variability belonged to the very low/no expres-

sion group (18 transcripts: *ALK*, *CCND1*, *CDKN1C*, *CDKN2A*, *DMBT1*, *EGFR*, *ESR2*, *EYA2*, *GPC3*, *GRB7*, *MET*, *MPL*, *TCL1A*, *TERT*, *TSHR*, *WT1*, *XRCC2*, *ZNF365*), and only seven were low (*CDH1*, *EPCAM*, *GADD45A*, *MMP8*, *MPL*, *MSR1*) and moderately (*DMC1*) expressed transcripts (Supplementary Table S5). Importantly, we did not identify significant biological variability in 53 highly expressed genes. The highest biological variability among the genes expressed in all 48 samples was observed for *MMP8* (Fig. 1E). Its expression was categorized as low, but in 7/48 (14.58 %) samples it surprisingly scored in a high expression category for an unknown underlying biological reason (Supplementary Table S5).

In conclusion, transcripts of clinically relevant cancer predisposition genes could be in the majority assessed from blood RNA samples and other easily accessible epithelial tissues (nasopharyngeal or skin), allowing evaluation of additional genes of interest. Moreover, RNA capture-based NGS enables evaluation of genes with low expression in a particular tissue and surpasses RNA-seq analyses of total RNA or mRNA for low expression transcript analysis in the order of magnitude (Fig. 1F).

Analysis of identical libraries (prepared from 24 RNA samples) in either NextSeq or NovaSeq instruments showed excellent inter-instrument reproducibility for mRNA level detection (r = 1; Fig. 2A). Further, we independently prepared an NGS library duplicate (from seven blood RNA samples) using the standard CZECANCA panel protocol that also demonstrated high inter-run reproducibility for the mRNA level detection (r = 1; Fig. 2B). High inter-capture reproducibility for mRNA level detection (r = 0.97) was demonstrated using 18 sample pre-libraries enriched with CZECANCA (226 genes, 0.63 Mb) or HyperExome probes (43 Mb), respectively (Fig. 2C). All tested technical parameters confirmed a high reproducibility and low variability of RNA-capture based NGS analysis.

Identification of splicing events is a major advantage of RNA-based analysis. Gene expression of a wild-type DNA allows identification of tissue-specific alternative splicing events (exon skipping or retention or splice donor/acceptor shifts). Therefore, a low coverage of alternatively skipped exons (in comparison to the reference transcript) may be observed when analysing RNA from a particular tissue type. Specifically, our analysis showed that in the case of *TSC2* gene, a complete in-frame skipping of the canonical exon 26 was observed in the blood RNA, whereas exon 26 was included in a considerable proportion of TSC2 transcripts from the nasopharynx and skin (Fig. 3). Alongside the alternative exon 26 skipping, an alternative splice acceptor shift in exon 27 (exclusion of three bases at a NAGNAG site in the 5' end of exon 27; del27p) was present in mRNA from all analysed tissues (Fig. 3, Supplementary Table S9A). Similar splice acceptor shifts in NAGNAG sites were identified as predominant alternative splicing events in BRCA1 mRNA del8p, del12p, and del13p (also known as del8p, del13p, and del14p by the still used historical nomenclature of *BRCA1* exons), present (mean) in 22.5 %, 7.5 %, and 28.1 % of BRCA1 blood RNA transcripts, respectively (Supplementary Table S9B).

In contrast to alternative splicing, aberrant splicing refers to a splicing event that occurs as a consequence of DNA sequence variant. DNA variants localized at the consensus splice sites (± 1, ± 2) are generally highly probably spliceogenic. However, exonic variants can also affect pre-mRNA splicing. It has been shown, e.g., for variants affecting the last/penultimate nucleotide in an exon. By RNA NGS, we correctly identified the aberrant splicing pattern in the presence of well-known spliceogenic *BRCA2* variant c.8486A>G (Fig. 4A), and abnormally increased proportion of an alternative splicing isoform in the presence of the *CHEK2* variant c.683G>A (Fig. 4B).

RNA NGS can also reveal aberrant splicing events resulting from the presence of deep intronic variants lo-

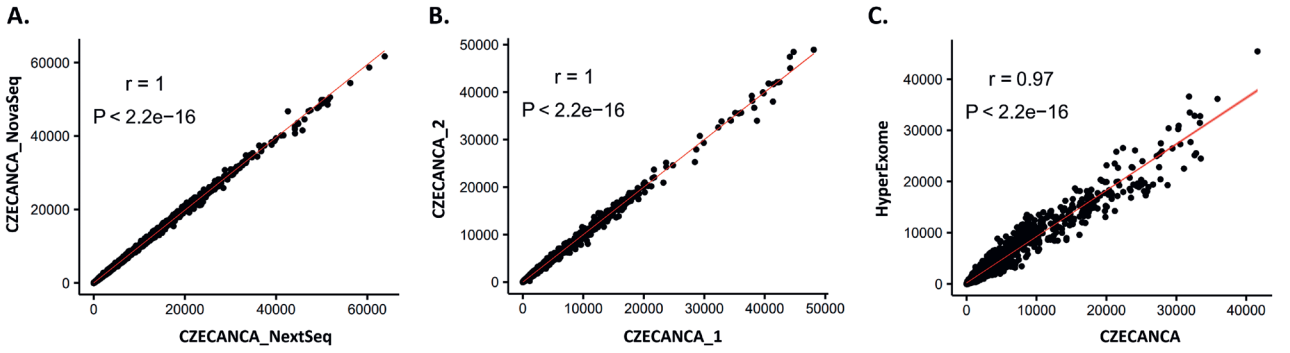

Fig. 2. Technical reproducibility of the RNA-targeted sequencing approach: inter-instrument (**A**), inter-run (**B**), and inter-capture (**C**). The x- and y-axes show the expression of particular genes (denoted as black dots), normalized to million transcripts.

Note: data for the correlations of individual samples are available as Supplementary Tables S6–8 and Supplementary Figures S1–2.

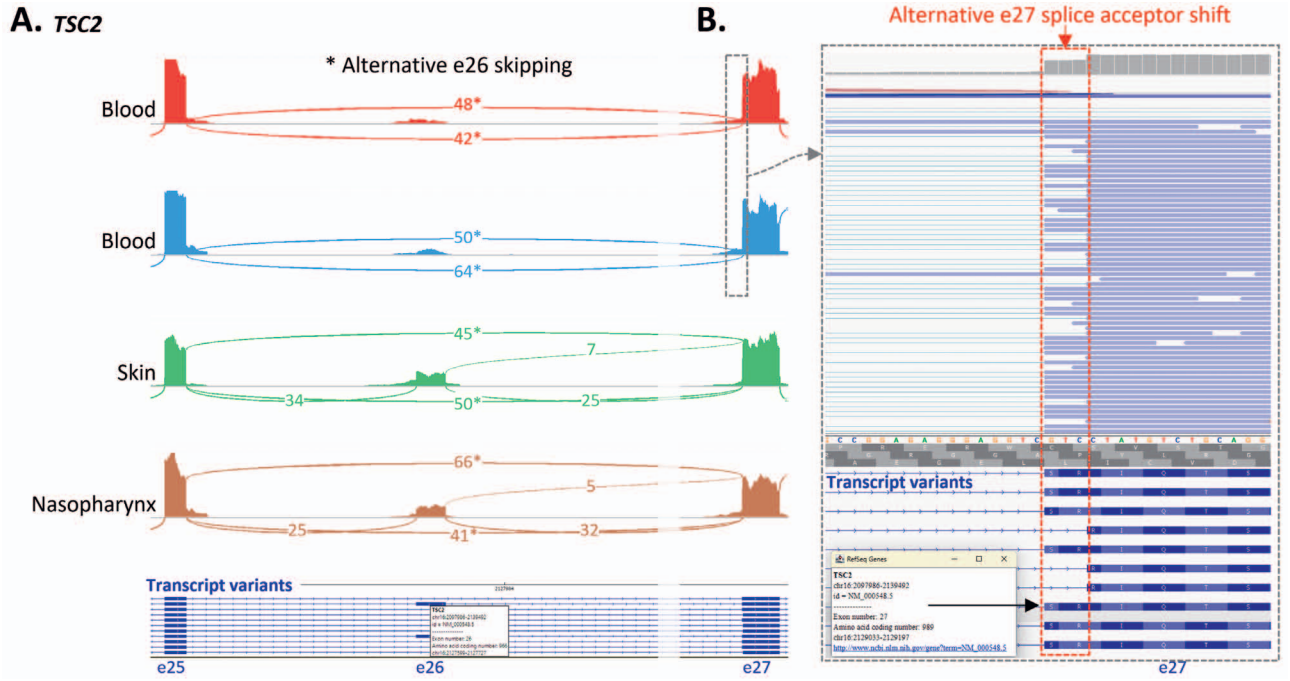

Fig. 3. Two concurrent alternative splicing events in TSC2 mRNA (NM_000548.5), including exon (e)26. (**A**) Alternative skipping of e26 was observed as a prevailing transcript isoform in the blood RNA, but was accompanied by a high proportion of the present wild-type (wt) reference isoform in the skin and nasopharynx. (**B**) Moreover, an alternative splice acceptor shift (3 bases) in e27 (del27p(3)) was present in mRNA from all analysed tissues in similar proportions.

calized outside the regions captured by NGS probes. As an example in our dataset, RNA capture-based NGS performed in a cancer patient who tested negatively by panel NGS from DNA identified aberrant splicing events in ATM mRNA that were reported by SCANVIS as novel exons. Detailed analysis revealed three different deep intron retentions resulting in the formation of transcripts containing premature termination codons. These were together present in 419/1351 (31 %) of the annotated splicing events (Fig. 5; Supplementary Table S9). In this patient, we further identified the presence of 109 unique unprocessed transcripts (82 % of unprocessed ATM pre-mRNA) with guanine at position c.1899-123 instead of the reference adenine nucleotide in the RNA sample. Detailed analysis of DNA NGS data from paired DNA samples identified spliceogenic ATM variant c.1899-123A>G in heterozygous state within the poorly covered intronic region, which was not annotated in the routine DNA analysis due to its low quality.

These examples demonstrated the utility of parallel DNA/RNA NGS for evaluating the biological significance of rare VUS, identified by DNA analysis, or for indicating the presence of deep intronic spliceogenic DNA variants beyond the exonic targets of panel NGS.

Discussion

NGS is now a routine method for germline DNA genetic testing. As the NGS technology continues to improve through simplification of sequencing library preparation, improved sequencing chemistry and reduced cost per Gbp, other procedures, such as confirmation of sample identity, confirmation of prioritized variants by Sanger sequencing, or interpretation of the clinical significance of detected VUS, become major constraints in the diagnostic pipeline. These time-consuming and labour-intensive activities can be significantly reduced by using RNA capture-based NGS performed in parallel with DNA analysis.

Confirmation of the germline variant or genotype from a second sample (usually DNA) reduces the risk of confusion, laboratory error, or sample identity mismatch (Foretova et al., 2016; Deans et al., 2022). It is also required by stakeholders as part of good laboratory practice in many countries including the Czech Republic. Parallel DNA/RNA NGS can verify both, correct sample identification and confirmation of variants of interest. Previously published methods suggested the use of internal standards or linkage disequilibrium analysis (Javed et al., 2020). Our solution, based on calculating the kinship coefficient for DNA-RNA pairs from the same individual, convincingly discriminated between matched and unmatched samples and represented a simple and fast approach that can be automated within the bioinformatic pipeline.

The above-mentioned advantages of the RNA capture-based NGS approach in VUS classification for hereditary cancer predisposition were recently demonstrated by Horton et al. (2022, 2024). The authors implemented a capture panel targeting transcripts of 18

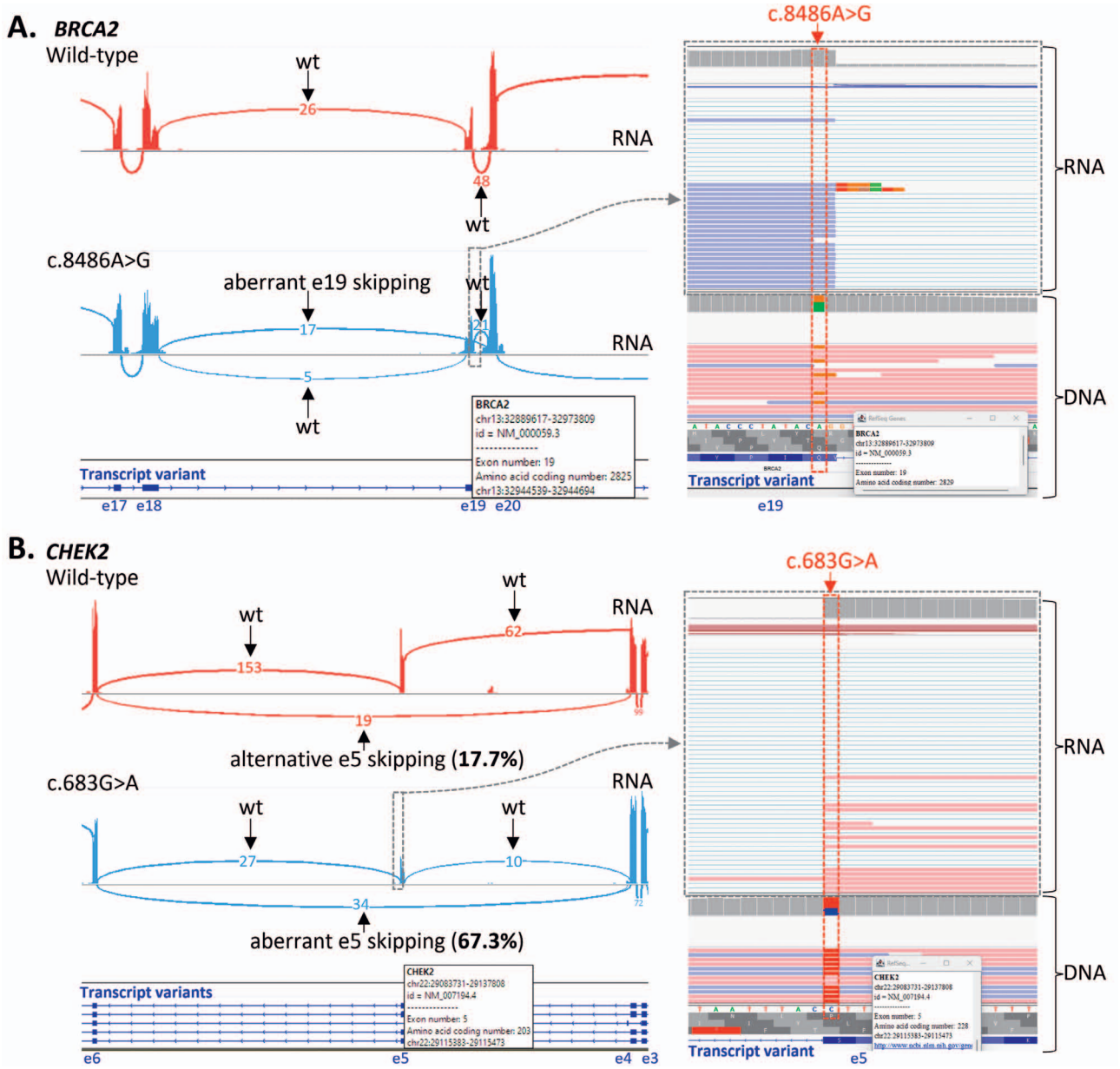

Fig. 4. Aberrant splicing events in mRNA samples from carriers of spliceogenic variants in *BRCA2* (NM_000059.3) and *CHEK2* (NM_007194.4). (**A**) The c.8486A>G variant in *BRCA2* (also known as p.Q2829R) localized to the penultimate nucleotide in exon 19 resulted in the loss of splice donor site, complete skipping of exon 19 (r.8332_8487del) and in-frame deletion of 52 amino acids (p.I2778_Q2829del) with complete absence of the variant nucleotide at position c.8486 within the transcript (right panel). (**B**) The c.683G>A variant in *CHEK2* (also known as p.S228N) localized to the last nucleotide in exon 5 resulted in the loss of splice donor site, complete skipping of exon 5 (r.592_683del92), frameshift and premature termination of translation (p.F199Vfs*6) with complete absence of the variant nucleotide at position c.683 within the transcript (right panel). Exon 5 skipping also occurred as alternative splicing and did not exceed 20 %, but was present in 67.3 % of the transcripts in the sample with a heterozygous c.683G>A variant.

clinically relevant cancer predisposition genes in a high-throughput diagnostic laboratory and showed that variant classification, based on parallel analysis of RNA samples, impacted 1.3 % of individuals being tested. Compared to the Horton studies, we decided to use a much larger panel for RNA analysis. For practical reasons, we chose the same panel CZECANCA as we routinely use for DNA analysis. Our approach reduced the cost and turnaround time and provided information about the transcripts for the full spectrum of analysed genes (if expressed in the analysed tissue type). As we have shown, the process is highly reproducible. It also allows generation of a catalogue of normal splicing patterns of genes of interest, which can be used to determine the natural splicing background. The described combination of the alternative exon 26 skipping and

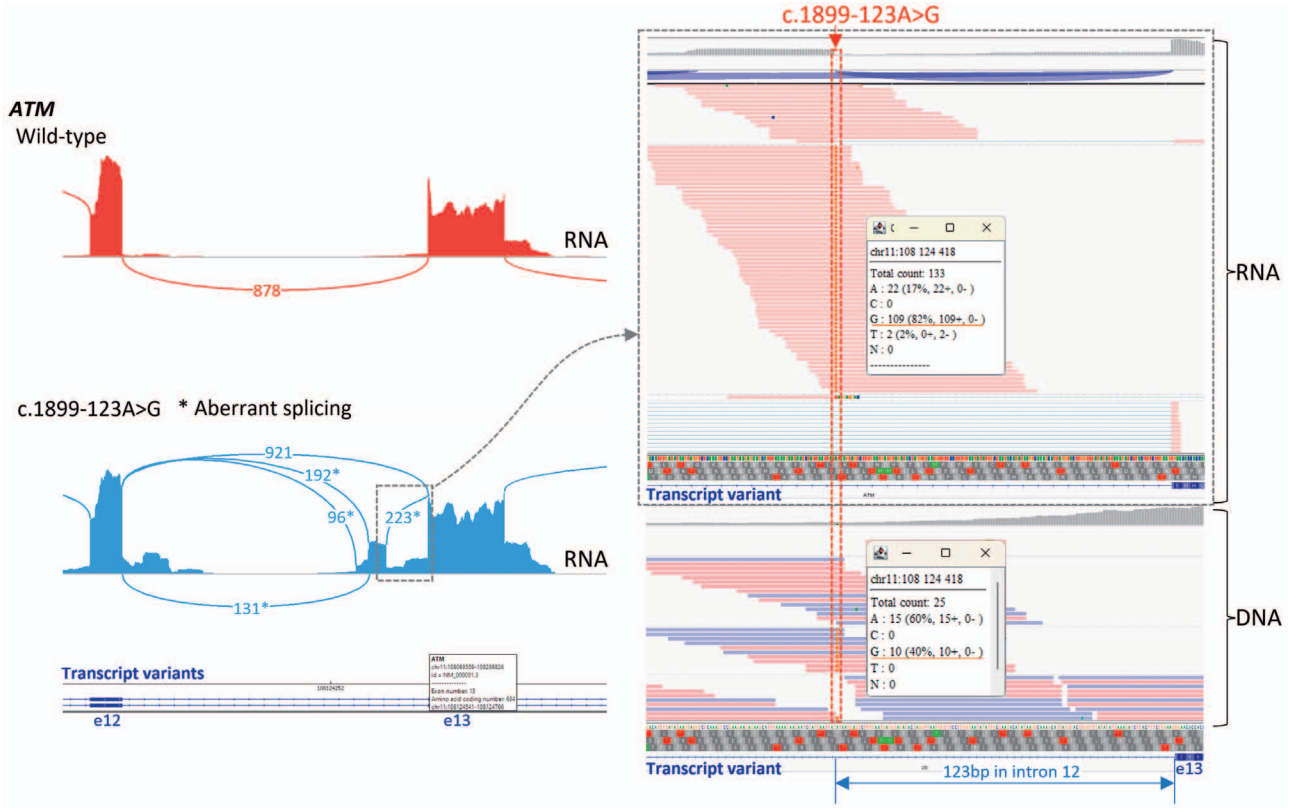

Fig. 5. Complex deep intronic retention in ATM mRNA (NM_000051.3) caused by the germline heterozygous variant c.1899-123A>G, visualized by Sashimi plot (left). Three new exons were retained in the mRNA transcripts in similar quantities: ins12A.1 (r.1899_1900ins1899-174_1899-124), ins12A.2 (r.1899_1900ins1899-177_1899-124), and ins12A.3 (r.1899_1900ins1899-213_1899-124). The right panel shows disproportionate enrichment of the variant allele (82 %) in unprocessed pre-mRNA transcripts and variant heterozygosity at the DNA level. Note: alternative splicing event. r.1899_1939del41 (present in 1.2 % of all transcripts in both control and mutant samples) has been hidden for better clarity.

splice acceptor shifts in TSC2, encoded by the *TSC2* gene (inactivated in patients with tuberous sclerosis 2; TSC2), can serve as an example. Skipping of exon 26 in the TSC2 mRNA has been previously reported as an alternative splicing event that does not cause the TSC2 phenotype (Ekong et al., 2016). The splice acceptor shifts in tandem NAGNAG sites, co-occurring in TSC2 mRNA alongside to exon 26 skipping, are common events with variable (but usually low) functional impact (Szafranski and Kramer, 2015; Hujova et al., 2021). Other examples of NAGNAG sites could be found in exons 8, 13, and 14 of *BRCA1*, where we also demonstrated that RNA capture-based analysis can be used in a quantitative manner. Our findings of splice acceptor shift events for del8p, del13p, and del14p (in 22.5 %, 7.5 %, and 28.1 % of transcripts, respectively; Supplementary Table S9B) were fully comparable to the frequencies determined by Colombo et al. (2014) using semi-quantitative capillary electrophoresis (~35 %, ~8 %, and ~30 %, respectively) (Colombo et al., 2014) and by Hojny et al. (2017) using multiplex PCR and NGS-based analysis (28.1 %, 11 %, and 26.6 %, respectively). Considering the low biological variability of independent blood RNA samples, we can conclude that our RNA capture-based NGS analysis provides quantitatively reproducible results.

The RNA capture-based NGS approach allows for selective enrichment of transcripts of interest, enabling successful analysis of poorly expressed genes (including many cancer predisposition genes) or ultra-low input samples (Curion et al., 2020). Compared to RNA-seq, RNA capture-based NGS achieves much higher coverage of transcripts of interest at a fraction of the cost. Analysis of the c.8486A>G variant in the *BRCA2* gene (with median expression in blood total RNA 0.2 TPM according to GTEx Portal data; https://gtexportal.org) demonstrated the feasibility of RNA capture-based NGS to analyse splicing events in genes with low blood RNA expression. Identical aberrant splicing profiles associated with the c.8486A>G *BRCA2* variant resulting in deletion of exon 19 have been described in several other RNA studies (Houdayer et al., 2012; Acedo et al., 2015; Kraus et al., 2017; Machackova et al., 2019).

An example of a novel variant associated with aberrant splicing and increased proportion of an alternative splicing event is the c.683G>A variant (also known as p.S228N) in the *CHEK2* gene. This variant has been reported as a VUS in ClinVar (ID: 826658), not yet analysed

at the RNA level. The mentioned substitution affects the last nucleotide of exon 5 and we have shown that it causes complete skipping of exon 5 (r.592_683del92), frameshift, and premature termination of translation (p.F199Vfs*6). To further support that, the variant nucleotide at position c.683 was not present in any of canonical transcripts including exon 5. Our analysis, therefore, clearly indicates that this apparently missense *CHEK2* variant should be classified as a spliceogenic pathogenic variant.

The precision and quantitative reproducibility of RNA capture-based NGS were also demonstrated using *ATM* variant c.1899-123A>G as an example. Using RT-PCR with semi-quantitative capillary electrophoresis, Moles-Fernandez et al. (2021) observed the same pseudoexons (12A.1, 12A.2, and 12A.3) generated from a cryptic donor splice site (created by the c.1899-123A>G transition), which in turn activated three different cryptic splice acceptor sites (c.1899-213, c.1899-177, and c.1899-174). Moles-Fernandez et al. detected a proportional presence of exons 12A.1 and 12A.2 and a lower proportion of exon 12A.3, but did not find the impaired pre-mRNA processing observed in our analysis. Consistent with our findings, the expression level of aberrantly spliced mRNA was lower than that of the wild-type transcript. It raises the question whether c.1899-123A>G only partially affects pre-mRNA splicing or whether the aberrant transcripts containing premature termination codons are partially degraded by nonsense-mediated mRNA decay (NMD). This could be solved by NMD inhibition; however, the procedure requires cultivation of lymphocytes prior to RNA isolation, which is not feasible within the routine diagnostic setting.

The limitations of the RNA NGS approach mainly include the need for sufficient gene of interest expression in the analysed tissue. In addition, the absence of aberrant splicing events in the presence of a suspected spliceogenic DNA variant must be interpreted with caution due to the possibility of loss of aberrant transcripts by NMD. We also found that, in contrast to RNA-seq, the capture-based analysis using the CZECANCA panel has several poorly covered regions in 5' untranslated exons (e.g., in *APC*, *UIMC1*, *DICER1*, or *CYLD*) that were not targeted in the capture panel design.

In conclusion, we demonstrated the feasibility and effectiveness of the parallel DNA/RNA NGS approach using an identical capture panel for the analysis of hereditary cancer predisposition. Our approach allows confirmation of DNA variants from an independent RNA sample (reducing the laborious Sanger confirmatory sequencing and limiting the risk of sample identity mismatches), analysis of the impact of VUS on mRNA processing, and discovery of unidentified DNA variants with impact on the aberrant splicing pattern. Although our work describes the use of parallel DNA/RNA capture-based NGS specifically within diagnosis of cancer predisposition, this approach is universally applicable to any genetic testing in general.

Acknowledgements

We thank Pavel Pešek and Eva Turečková for excellent technical support and Lenka Steiner Mrázová for providing the protocol for RNA isolation from nasopharyngeal swabs. We would like to thank our colleagues Viktor Stránecký and Markéta Šafaříková for their technical help and support.

Conflict of interest

The authors declare no conflict of interest.

Supplementary Figures

Supplementary Fig. S1. Inter-instrument reproducibility of the RNA-targeted sequencing approach

Supplementary Fig. S2. Inter-run (**A**) and inter-capture (**B**) reproducibility of the RNA-targeted sequencing approach

Supplementary Tables

Supplementary Table S1. List of 226 CZECANCA genes and transcript variants

Supplementary Table S2. Absolute numbers of unique reads of individual wild-type exon-exon junctions in 48 blood RNA samples analysed in a single run

Supplementary Table S3. Gene expression – absolute numbers of maximal unique reads of wild-type exon-exon junctions for each of 226 targeted genes in 48 blood RNA samples analysed in a single run

Supplementary Table S4. Gene expression – numbers of maximal unique reads of wild-type exon-exon junctions normalized to one million of transcripts for each of 226 targeted genes in 48 blood RNA samples analysed in a single run

Supplementary Table S5. Biological variability of gene expression assessment using dispersion coefficient in 48 blood RNA samples analysed in a single run

Supplementary Table S6. Inter-instrument reproducibility

Supplementary Table S7. Inter-run reproducibility

Supplementary Table S8. Inter-capture reproducibility

Supplementary Table S9. Annotated numbers of unique junctions in examples of alternative and aberrant splicing

References

Acedo, A., Hernandez-Moro, C., Curiel-Garcia, A. et al. (2015) Functional classification of BRCA2 DNA variants by splicing assays in a large minigene with 9 exons. *Hum. Mutat.* **36**, 210-221.

Agius, P., Geiger, H., Robine, N. (2019) SCANVIS: a tool for SCoring, ANnotating and VISualizing splice junctions. *Bioinformatics* **35**, 4843-4845.

Brandão, R. D., Mensaert, K., López-Perolio, I. et al. (2019) 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.

Chen, Y., Lun, A. T., Smyth, G. K. (2016) From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res.* **5**, 1438.

Colombo, M., Blok, M. J., Whiley, P. et al. (2014) Comprehensive annotation of splice junctions supports pervasive alternative splicing at the BRCA1 locus: a report from the ENIGMA consortium. *Hum. Mol. Genet.* **23**, 3666-3680.

Cotto, K. C., Feng, Y. Y., Ramu, A. et al. (2023) Integrated analysis of genomic and transcriptomic data for the discovery of splice-associated variants in cancer. *Nat. Commun.* **14**, 1589.

Curion, F., Handel, A. E., Attar, M. et al. (2020) Targeted RNA sequencing enhances gene expression profiling of ultra-low input samples. *RNA Biol.* **17**, 1741-1753.

Davy, G., Rousselin, A., Goardon, N. et al. (2017) Detecting splicing patterns in genes involved in hereditary breast and ovarian cancer. *Eur. J. Hum. Genet.* **25**, 1147-1154.

Deans, Z. C., Ahn, J. W., Carreira, I. M. et al. (2022) Recommendations for reporting results of diagnostic genomic testing. *Eur. J. Hum. Genet.* **30**, 1011-1016.

Dobin, A., Davis, C. A., Schlesinger, F. et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21.

Ekong, R., Nellist, M., Hoogeveen-Westerveld, M. et al. (2016) Variants within TSC2 exons 25 and 31 are very unlikely to cause clinically diagnosable tuberous sclerosis. *Hum. Mutat.* **37**, 364-370.

Farber-Katz, S., Hsuan, V., Wu, S. et al. (2018) Quantitative analysis of BRCA1 and BRCA2 germline splicing variants using a novel RNA-massively parallel sequencing assay. *Front. Oncol.* **8**, 286.

Foretova, L., Machackova, E., Palacova, M. et al. (2016) Recommended extension of indication criteria for genetic testing of BRCA1 and BRCA2 mutations in hereditary breast and ovarian cancer syndrome. *Klin. Onkol.* **29 (Suppl. 1)**, S9-S13. (in Czech)

Havranek, O., Kleiblova, P., Hojny, J. et al. (2015) Association of germline CHEK2 gene variants with risk and prognosis of non-Hodgkin lymphoma. *PLoS One* **10**, e0140819.

Hojny, J., Michalkova, R., Krkavcova, E. et al. (2022) Comprehensive quantitative analysis of alternative splicing variants reveals the HNF1B mRNA splicing pattern in various tumour and non-tumour tissues. *Sci. Rep.* **12**, 199.

Hojny, J., Zemankova, P., Lhota, F. et al. (2017) Multiplex PCR and NGS-based identification of mRNA splicing variants: analysis of BRCA1 splicing pattern as a model. *Gene* **637**, 41-49.

Hong, M., Tao, S., Zhang, L. et al. (2020) RNA sequencing: new technologies and applications in cancer research. *J. Hematol. Oncol.* **13**, 166.

Horackova, K., Frankova, S., Zemankova, P. et al. (2022) Low frequency of cancer-predisposition gene mutations in liver transplant candidates with hepatocellular carcinoma. *Cancers (Basel)* **15**, 201.

Horton, C., Cass, A., Conner, B. R. et al. (2022) Mutational and splicing landscape in a cohort of 43,000 patients tested for hereditary cancer. *NPJ Genom. Med.* **7**, 49.

Horton, C., Hoang, L., Zimmermann, H. et al. (2024) Diagnostic outcomes of concurrent DNA and RNA sequencing in individuals undergoing hereditary cancer testing. *JAMA Oncol.* **10**, 212-219.

Houdayer, C., Caux-Moncoutier, V., Krieger, S. et al. (2012) 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.

Hujova, P., Soucek, P., Radova, L. et al. (2021) Nucleotides in both donor and acceptor splice sites are responsible for choice in NAGNAG tandem splice sites. *Cell. Mol. Life Sci.* **78**, 6979-6993.

Javed, N., Farjoun, Y., Fennell, T. J. et al. (2020) Detecting sample swaps in diverse NGS data types using linkage disequilibrium. *Nat. Commun.* **11**, 3697.

Karam, R., LaDuca, H., Richardson, M. E. et al. (2020) RNA-seq analysis is a useful tool in variant classification. *JCO Precis. Oncol.* **4**, 1226-1227.

Kleibl, Z., Kristensen, V. N. (2016) Women at high risk of breast cancer: molecular characteristics, clinical presentation and management. *Breast* **28**, 136-144.

Kleiblova, P., Stolarova, L., Krizova, K. et al. (2019) Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer. *Int. J. Cancer* **145**, 1782-1797.

Kral, J., Jelinkova, S., Zemankova, P. et al. (2023) Germline multigene panel testing of patients with endometrial cancer. *Oncol. Lett.* **25**, 216.

Kraus, C., Hoyer, J., Vasileiou, G. et al. (2017) Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. *Int. J. Cancer* **140**, 95-102.

Kuzbari, Z., Bandlamudi, C., Loveday, C. et al. (2023) Germline-focused analysis of tumour-detected variants in 49,264 cancer patients: ESMO Precision Medicine Working Group recommendations. *Ann. Oncol.* **34**, 215-227.

LaDuca, H., Polley, E. C., Yussuf, A. et al. (2020) A clinical guide to hereditary cancer panel testing: evaluation of gene-specific cancer associations and sensitivity of genetic testing criteria in a cohort of 165,000 high-risk patients. *Genet. Med.* **22**, 407-415.

Lattimore, V. L., Pearson, J. F., Currie, M. J. et al. (2018) Investigation of experimental factors that underlie BRCA1/2 mRNA isoform expression variation: recommendations for utilizing targeted RNA sequencing to evaluate potential spliceogenic variants. *Front. Oncol.* **8**, 140.

Lattimore, V. L., Pearson, J. F., Morley-Bunker, A. E. et al. (2019) Quantifying BRCA1 and BRCA2 mRNA isoform expression levels in single cells. *Int. J. Mol. Sci.* **20**, 693.

Leman, R., Harter, V., Atkinson, A. et al. (2020) SpliceLauncher: a tool for detection, annotation and relative quantification of alternative junctions from RNAseq data. *Bioinformatics* **36**, 1634-1636.

Lhota, F., Zemankova, P., Kleiblova, P. et al. (2016) Hereditary truncating mutations of DNA repair and other genes in BRCA1/BRCA2/PALB2-negatively tested breast cancer patients. *Clin. Genet.* **90**, 324-333.

Lhotova, K., Stolarova, L., Zemankova, P. et al. (2020) Multigene panel germline testing of 1333 Czech patients with ovarian cancer. *Cancers (Basel)* **12**, 956.

Lopez-Perolio, I., Leman, R., Behar, R. et al. (2019) Alternative splicing and ACMG-AMP-2015-based classification of PALB2 genetic variants: an ENIGMA report. *J. Med. Genet.* **56**, 453-460.

Machackova, E., Claes, K., Mikova, M. et al. (2019) Twenty years of BRCA1 and BRCA2 molecular analysis at MMCI – current developments for the classification of variants. *Klin. Onkol.* **32 (Suppl. 2)**, 51-71.

McCarthy, D. J., Chen, Y., Smyth, G. K. (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288-4297.

McKenna, A., Hanna, M., Banks, E. et al. (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303.

Moles-Fernandez, A., Domenech-Vivo, J., Tenes, A. et al. (2021) Role of splicing regulatory elements and in silico tools usage in the identification of deep intronic splicing variants in hereditary breast/ovarian cancer genes. *Cancers (Basel)* **13**, 3341.

Pohlreich, P., Stribrna, J., Kleibl, Z. et al. (2003) Mutations of the BRCA1 gene in hereditary breast and ovarian cancer in the Czech Republic. *Med. Princ. Pract.* **12**, 23-29.

Rahman, N. (2014) Realizing the promise of cancer predisposition genes. *Nature* **505**, 302-308.

Schafer, S., Miao, K., Benson, C. C. et al. (2015) Alternative splicing signatures in RNA-seq data: percent spliced in (PSI). *Curr. Protoc. Hum. Genet.* **87**, 11.16.1-11.16.14.

Soukupova, J., Zemankova, P., Lhotova, K. et al. (2018) Validation of CZECANCA (CZEch CAncer paNel for Clinical Application) for targeted NGS-based analysis of hereditary cancer syndromes. *PLoS One* **13**, e0195761.

Stadler, Z. K., Maio, A., Chakravarty, D. et al. (2021) Therapeutic implications of germline testing in patients with advanced cancers. *J. Clin. Oncol.* **39**, 2698-2709.

Struzinska, I., Hajkova, N., Hojny, J. et al. (2023) A comprehensive molecular analysis of 113 primary ovarian clear cell carcinomas reveals common therapeutically significant aberrations. *Diagn. Pathol.* **18**, 72.

Sung, H., Ferlay, J., Siegel, R. L. et al. (2021) Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209-249.

Szafranski, K., Kramer, M. (2015) It's a bit over, is that ok? The subtle surplus from tandem alternative splicing. *RNA Biol.* **12**, 115-122.

Walker, L. C., Hoya, M., Wiggins, G. A. R. et al. (2023) Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: recommendations from the ClinGen SVI splicing subgroup. *Am. J. Hum. Genet.* **110**, 1046-1067.

Walker, L. C., Lattimore, V. L., Kvist, A. et al. (2019) Comprehensive assessment of BARD1 messenger ribonucleic acid splicing with implications for variant classification. *Front. Genet.* **10**, 1139.

Wieme, G., Kral, J., Rosseel, T. et al. (2021) Prevalence of germline pathogenic variants in cancer predisposing genes in Czech and Belgian pancreatic cancer patients. *Cancers (Basel)* **13**, 4430.

Yoon, S., Nam, D. (2017) Gene dispersion is the key determinant of the read count bias in differential expression analysis of RNA-seq data. *BMC Genomics* **18**, 408.

Zheng, X., Levine, D., Shen, J. et al. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326-3328.