RNAsum: a tool for personalised genome and transcriptome interpretation for improved cancer diagnostics

Sehrish Kanwal[*,#,†,1,2], Jacek Marzec[*,#,1,3], Joseph H.A. Vissers[1,2], Peter Diakumis[1,2], Leila N. Varghese[1,2], Luke Tork[1], Kym Pham Stewart[1,2], Oliver Hofmann[1,2], Stephen J. Luen[4,5], Sean M. Grimmond[#,1,6]

* Equal contributions

Corresponding authors

† Lead author

1. **Collaborative Centre for Genomic Cancer Medicine, University of Melbourne, Melbourne, Victoria, Australia**
2. **Department of Clinical Pathology, University of Melbourne, Melbourne, Victoria, Australia**
3. **Collaborative Laboratory AccelBio, Cantanhede, Portugal**
4. **Department of Medical Oncology, Peter MacCallum Cancer Centre, Parkville, VIC 3000, Australia**
5. **Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, VIC 3010, Australia**
6. **Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia**

Author Information:

Sehrish Kanwal kanwals@unimelb.edu.au https://orcid.org/0000-0002-5044-4692
Jacek Marzec jacek.marzec@accelbio.pt https://orcid.org/0000-0001-6577-8866
Joseph H.A. Vissers joseph.vissers@unimelb.edu.au
Peter Diakumis peter.diakumis@unimelb.edu.au
Leila N. Varghese lvarghese@unimelb.edu.au
Luke Tork luke.tork@unimelb.edu.au
Kym Pham Stewart kym.phamstewart@unimelb.edu.au
Oliver Hofmann oliver.hofmann@unimelb.edu.au
Stephen J. Luen stephen.luen@petermac.org
Sean M. Grimmond sean.grimmond@unimelb.edu.au https://orcid.org/0000-0002-8102-7998

Abstract

Background The integration of whole-genome sequencing (WGS) and whole-transcriptome sequencing (WTS) has revolutionized cancer diagnostics by enabling comprehensive molecular profiling of tumours. WGS uncovers genomic alterations such as single nucleotide variants, structural variants, and copy number changes, whereas WTS reveals their functional consequences through expression profiles and fusion detection. Together, these technologies offer unparalleled potential to guide precision oncology by identifying actionable biomarkers and stratifying patients for targeted therapies or clinical trials. Recent studies have shown that nearly half of patients experience improved clinical outcomes when treatment is guided by combined WGTS analysis. However, automated and effective integration and visualisation of these complex data remain the major challenges for personalized cancer sample interpretation.

Results We developed RNAsum, an open-source tool for integrating and interpreting whole-genome sequencing (WGS) and whole-transcriptome sequencing (WTS) data from individual cancer patient samples. RNAsum compares patient data to The Cancer Genome Atlas (TCGA) cohorts, integrating quantitative expression data with genomic findings to validate and prioritise clinically relevant alterations and enhance diagnostic accuracy. Clinical applicability evaluation performed across 60 patients demonstrated that 68% (140/205) of the clinically reportable variants identified by WGS, spanning copy number changes, truncating mutations and gene fusions, were supported at the RNA level as detected by WTS and reported by RNAsum. Case studies further highlight the ability of RNAsum to confirm clinically reportable variants, refine diagnoses, and identify novel therapeutic targets, particularly in complex cases involving multiple genomic alterations and drug resistance mechanisms.

Conclusion RNAsum effectively bridges the gap between genome and transcriptome analyses, significantly advancing the integration of multiomics data in personalized cancer care. Its ability to validate clinically reportable variants at the RNA level and elucidate complex alterations, including those driving drug resistance, highlights RNAsum's potential to improve molecular tumour profiling and support clinical decision-making. Freely available as an R package on GitHub at https://github.com/umccr/RNAsum, RNAsum provides an accessible and scalable solution for researchers and clinicians, representing a significant step towards the routine application of integrated genomic and transcriptomic analyses in precision oncology.

Keywords Precision oncology; Whole-genome sequencing (WGS); Whole-transcriptome sequencing (WTS); Multiomics integration; Cancer genomics; Molecular profiling; Clinical variant interpretation; Bioinformatics tool; Open-source software

Background

High-throughput sequencing (HTS) technologies have revolutionized the field of precision oncology by enabling fast and accurate analysis of whole-genome and transcriptome data from tumours [1,2]. Owing to the decreasing sequencing cost, whole-genome sequencing (WGS) provides valuable insights into the molecular mechanisms underlying oncogenesis and genomic aberrations such as single nucleotide variants (SNVs), small insertions and deletions (indels), structural variants (SVs) and copy-number variants (CNVs) [3,4]. Furthermore, it has shown great utility in studying clinically actionable cancer genome alterations including tumour mutational burden (TMB) [5], homologous recombination deficiency (HRD) [6] and microsatellite instability (MSI) [7].

As the field of HTS continues to expand and support precision oncology initiatives globally, whole-transcriptome sequencing (WTS) has gained immense traction in cancer research and personalized medicine [8] due to its ability to provide comprehensive insights into the transcriptome, the complete set of gene readouts called RNA molecules in a cell or tissue. It allows quantification of expression levels to understand gene activation or silencing in a cancerous cell [9].

Combined whole-genome and transcriptome sequencing (WGTS) has demonstrated the ability to provide a nearly complete fingerprint of patients' tumour profiles and is increasingly being used to molecularly diagnose challenging malignancies [10–14]. While WGS provides insights into both coding and non-coding genomic alterations by detecting a range of variant types, including SNVs, indels, CNVs and SVs, integrating RNA sequencing (RNA-seq) with WGS enables confirmation of the functional consequences of these somatic changes. This comprehensive diagnostic approach can improve precision oncology practices by providing complete characterisation and molecular profiling for the identification of targetable alterations that stratify patients for eligible clinical trials and off-label drug treatment [15]. By understanding the unique genetic mutations and expression profile of a patient, clinicians can use these data for disease detection, prognosis and tailor therapies to be more effective and minimize side effects.

Recent studies have demonstrated the significant value of integrating RNA expression data with genomic profiling in precision oncology. [16] reported that nearly half of patients demonstrated positive clinical outcomes when treatment was guided by combined WGTS analysis. These findings highlighted how RNA data can confirm the functional consequences of genomic

mutations, either independently or in combination with DNA sequencing, thereby informing treatment decisions. This approach lays the groundwork for future precision oncology efforts, including our current research, which aims to incorporate WGTS data into routine clinical care for cancer patients.

While current clinical genomics tests focus primarily on WGS data, integrated transcriptome analysis has shown significant value in precision oncology. Studies have demonstrated the importance of cross-validating WGS findings with transcriptome-based results [17,18]. However, the field lacks a comprehensive bioinformatics approach for effectively combining and visualising WGS and WTS data for personalised cancer sample interpretation. To address this challenge, we developed RNAsum, a tool that streamlines the integration of WGS and WTS data from individual patient samples. RNAsum enables comprehensive analysis of the transcriptional consequences of somatic events and compares patient data against The Cancer Genome Atlas (TCGA) [19] reference cohort. RNAsum generates interactive reports summarising key findings and provides supplementary resources for further data exploration. We demonstrate the implementation and clinical utility of the RNAsum through case studies, discuss its features and limitations, and provide detailed usage information, thereby establishing its value in routine clinical practice for WGTS interpretation in precision oncology.

Implementation

RNAsum is an open-source R package that enables the integration and interpretation of WGS and WTS data from individual cancer patient samples while remaining easy to install and requiring minimal technical expertise or management overhead.

Reference data

To assist in the interpretation of gene expression data prospectively generated from patient cancer samples, we applied a solution that allows the comparison of single-subject tumour-only expression profiles to large cancer-matched expression landscapes taken from the TCGA. Read count data from 10,079 cancer patients, covering 33 cancer types (Table S1) were obtained from the National Cancer Institute (NCI) Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov) via the GDC Application Programming Interface (API, https://gdc.cancer.gov/developers/gdc-application-programming-interface-api) and data release 40.0 (March 29, 2024,

https://docs.gdc.cancer.gov/Data/Release_Notes/Data_Release_Notes/#data-release-400). These reference profiles were then harmonized with the individual sample to determine the relative gene expression level of any gene of interest in the individual, expressed as a percentile of the relative to the reference patient cohort expression spectrum, hereinafter referred to as "relative percentiles" (see the “WTS data integration” section for a detailed description).

Additionally, users can provide a customized list of genes of interest (GOIs) to the workflow, enabling exploration of their expression profiles in investigated patients relative to corresponding reference cohorts. If no custom gene set is supplied, the workflow defaults to analyse three preassembled gene sets: (1) “immune-related genes” (listed in the Immune_markers PanelApp gene panel, https://panelapp-aus.org/panels/243) and included to monitor potential cancer immune regulation), (2) “homologous recombination deficiency (HRD) genes” (listed in the Homologous_recombination_deficiency PanelApp gene panel, https://panelapp-aus.org/panels/242), and (3) “known cancer genes”, a curated set of 1,315 genes associated with tumorigenesis (defined at https://github.com/umccr/gene_panels/releases). These exemplar gene sets are routinely reviewed in individual patients to explore the expression of immune-related genes, genes associated with HRD, and key cancer-related genes.

Finally, we collected information from public knowledge bases, including FusionGDB [20], Cancer Genome Interpreter (CGI) [21], Clinical Interpretation of Variants in Cancer (CIViC) [22], Variant Interpretation for Cancer Consortium (VICC, https://cancervariants.org) and OncoKB [23,24], to annotate WGTS findings and to determine the clinical significance of the identified alterations.

Workflow overview

The RNAsum workflow includes four data components (data from patient samples are summarized in Table 1) and consists of two integration and one annotation steps (Fig. 1). First, mRNA profiles from the patient and external (TCGA) reference cohort are combined at the WTS level (components 1 and 2). In the second step, the expression data and detected gene fusions are overlaid with genomic alterations at the WGTS level (component 3). Finally, clinically relevant annotations from public knowledge bases are added to facilitate interpretation (component 4).

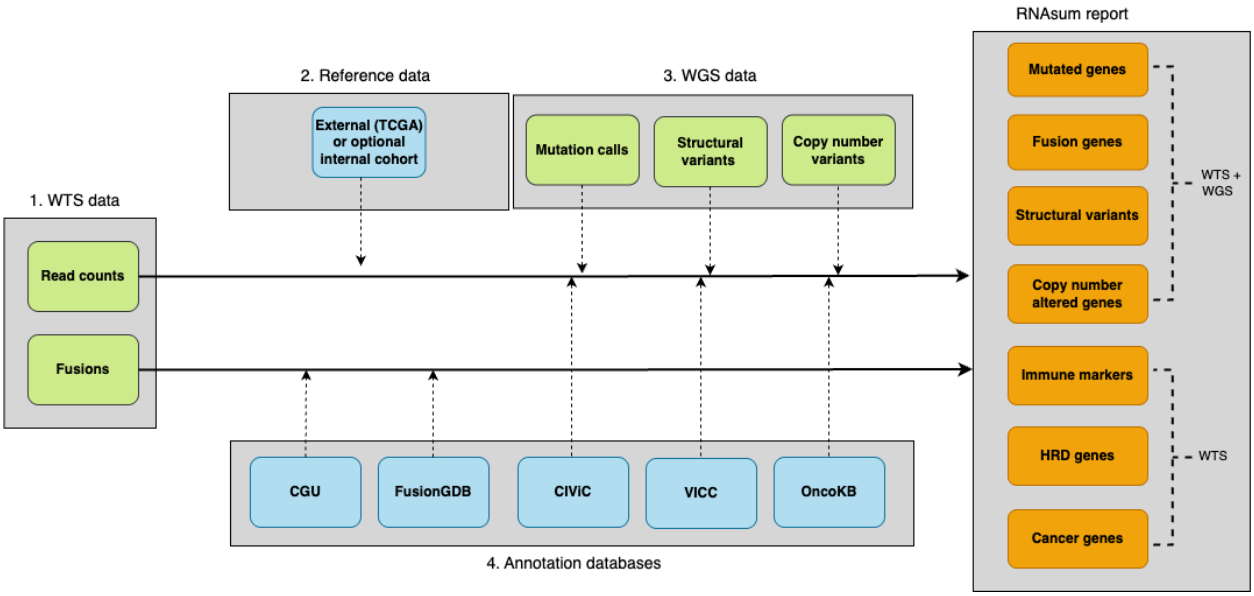

Fig. 1 Overview of the integrated genome and transcriptome analysis workflow. The process is divided into four main stages: (1) read count and gene fusion data collection, (2) reference data processing, (3) integration with WGS-based data, and (4) results annotation. Colour coding is used to distinguish different elements: green represents inputs from a patient, blue indicates reference data, and orange denotes outputs. This workflow enables comprehensive analysis of genomic and transcriptomic data by combining multiple data types and reference sources.

Table 1 Summary of input data from a patient (components in green in Fig. 1).

Workflow component (platform)	Input data	Data source	Required	Example tool/database
1 (WTS)	Transcript/gene abundances	Patient	Yes	Salmon [25], Kallisto [26]
1 (WTS)	Fusion genes	Patient	No	Arriba [27], DRAGEN [28]
3 (WGS)	SNVs/indels	Patient	No	PCGR [29]
3 (WGS)	CNVs	Patient	No	PURPLE [30]
3 (WGS)	SVs	Patient	No	Manta [31]

WTS data integration

The read count data from two sources are used as an inputs for the integration at the WTS level: (1) the patient sample and (2) an external reference cohort, which may either match the patient's

cancer type (tumour-specific) or consist of all patients available within TCGA (pan-cancer) (see the “Reference data collection” section and Table S1). Tumour-specific cohorts offer more context-relevant comparisons, whereas the pan-cancer cohort provides larger sample sizes and broader expression ranges, improving percentile resolution, especially for genes expressed across multiple tumour types. The pan-cancer cohort is especially useful in cases where the tumour origin is unknown, the histology is rare or ambiguous, or where broader benchmarking across cancer types is of interest.

Users can also choose between two TCGA reference cohort modes: the “full” or a memory-efficient “representative” subset. While the “full” mode improves precision, the default “representative” mode supports better scalability and integration into routine workflows, particularly when a memory-demanding pan-cancer reference cohort is used. To evaluate the impact of the two modes, we analysed 10 of the 60 patient cases used in the RNAsum clinical utility evaluation (see “Clinical applicability evaluation”). Specifically, we examined five cases combined with the tumour-matched cohort and five cases combined with the pan-cancer TCGA reference cohort. The benchmarking results revealed that 99% of the “known cancer genes” (see “Reference data collection”) presented <10 percentile point differences between the two modes (Fig. S1), with discrepancies affecting mainly genes with low or undetectable expression. This reflects typical behaviour in transcriptomics data, where low-abundance transcripts are more susceptible to variation after normalization.

To account for potential batch effects, an internal reference cohort assembled using in-house WTS samples, which are representative of the patient’s cancer type and processed through the same pipeline, can be incorporated into the workflow. We further explored 10 patient cases (five combined with tumour-matched and five with pan-cancer external reference cohorts) to evaluate the impact of batch-effect correction using an internal reference cohort (Fig. S2). For this benchmarking, we considered all clinically reported variants predicted to impact RNA expression, including truncating SNVs and small indels, activating and inactivating rearrangements, CN losses and amplifications. Our in-house data indicate that this step has little effect on the expression levels of genes involved in clinically relevant genomic alterations. Given this limited effect, creating an internal reference cohort is not essential.

The input read count data are combined and transformed to counts per million (CPM) values, genes with low counts are filtered out, and then the data are normalized to account for sample-

specific effects (Fig. S3). Gene expression levels in the patient sample are reported as percentiles relative to the expression spectrum of the reference patient cohort. Additionally, WTS data are standardized using Z-scores, which express deviations from the mean in terms of standard deviations. This transformation enables direct comparison across samples and facilitates the interpretation of expression values (Fig. S4).

WGTS data integration

WTS data for each patient are integrated with matched WGS data by overlaying gene expression levels onto genes harbouring genomic alterations. Various tools can provide the necessary input files in compatible formats, including SNVs and indels from tools such as the Personal Cancer Genome Reporter (PCGR) [29], SVs from callers like MANTA [31], and CNVs from tools such as PURPLE purity ploidy estimator [30] (Fig. 1). Moreover, gene fusions identified by Arriba [27] or DRAGEN [28] and supporting WGS-derived SVs are prioritised, as this concordant evidence strengthens their functional relevance. To enhance usability, RNAsum accepts simplified TSV input files containing gene-level SV or CN information. This flexibility enables users to integrate results generated by their preferred analysis tools. The content presented in the output tables is drawn directly from the user-provided TSVs, ensuring compatibility with diverse analytical workflows and eliminating dependence on any specific software.

Genome-based findings, if provided, serve as the primary source for prioritising expression profiles, ensuring that the differences in expression levels between patient sample and the corresponding reference cohort for key genes are evaluated first. The integrated WTS and WGS data are then ranked by the observed difference, highlighting genes with the greatest potential clinical impact. All reported alterations are annotated with interpretation-relevant information from public knowledge bases, including CGI [21], VICC (https://cancervariants.org), OncoKB [23,24] and CIViC [22]. Additionally, detected gene fusions are functionally annotated with information from FusionGDB [20].

WGTS variant interpretation

The outputs of the pipeline are an interactive report that prioritises clinically relevant information to facilitate results interpretation, as well as supplementary resources that can be used programmatically for further data exploration and comparisons against other patient cohorts. This report includes searchable tables and plots organised into several sections detailed below and

illustrated in Fig. 2 (see the Case study #1 section for more details). Each section features an interactive “summary” table and “expression profiles” plots.

The “summary” table contains comprehensive information about individual genes' mRNA expression levels in a patient sample, the average mRNA expression across samples from TCGA cancer patients, and the difference between these two measurements. Notably, mRNA measurements are presented in both percentiles, to facilitate intuitive interpretation, and Z-scores. Genes in these tables are sorted based on section-specific criteria and then by the decreasing absolute values representing the difference between patient and TCGA expression levels. The “summary” table also includes hyperlinks to relevant knowledge bases, such as VICC (https://cancervariants.org), OncoKB [23,24] and CIViC [22], providing additional context, as well as information specific to each report section. The interactive plots comprise two main components: a distribution panel showing percentile values as a function of expression levels and a box plot panel displaying expression levels for each gene in the patient sample compared with the reference cancer cohort. Additionally, we included a bar plot illustrating the read counts of individual genes across all samples.

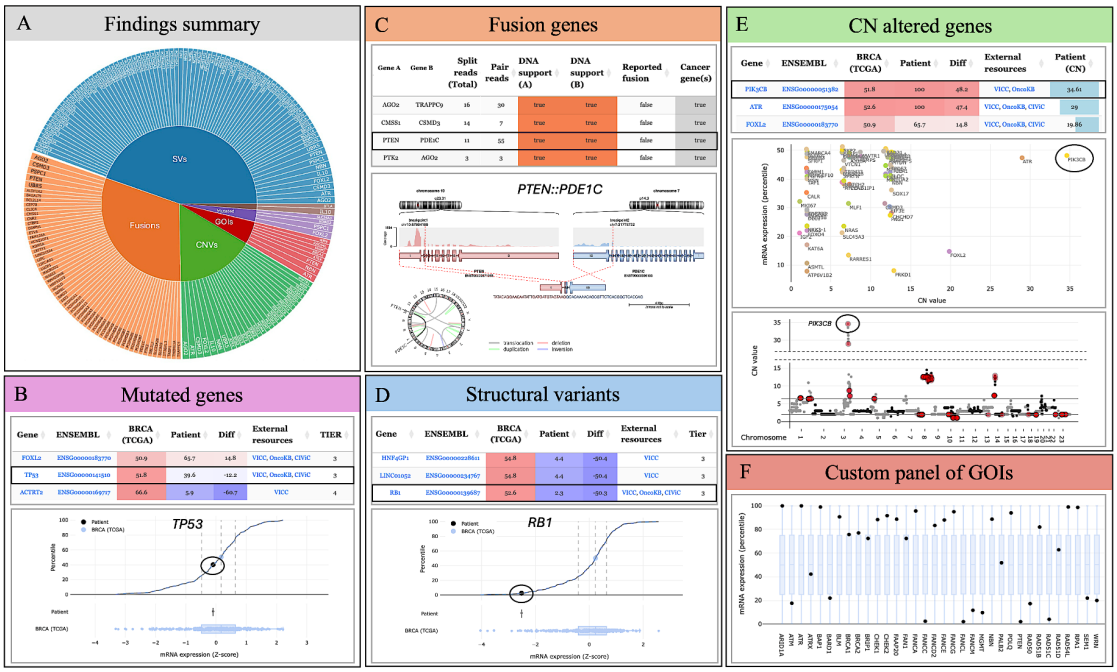

Fig. 2 Overview of RNAsum interactive report structure and visualisation features. The diagram illustrates the organisation of the report into key sections, including the input data summary, findings summary (**A**), mutated genes (**B**), fusion genes (**C**), structural variants (**D**), CN altered genes (**E**), and a custom panel of genes of interest (**F**), each described in detail in the main text. Example plots and tables demonstrate how RNAsum prioritises and visualises clinically relevant findings to support interpretation The expression values presented in the "summary" tables are presented as percentiles, and "expression profiles" plots are shown as Z-scores. *GOIs — genes of interest.*

Input data summary

The WTS input data are summarized at the top of the report, providing a comprehensive overview of the analysis. This section includes information about reference cohorts and patient samples used in the analysis, details about included and excluded genes along with the reasoning behind these decisions, visual representations of the initial library size generated across all samples, read count data filtering, transformation and normalization. Moreover, it contains interactive exploratory data analysis plots, including principal component analysis (PCA) and relative log expression (RLE) plots, which facilitate the identification of key factors affecting variability in the expression data.

Findings summary

The findings from the WGTS data integration are summarized in the form of an interactive doughnut plot and table presenting genes with detected alterations, which are further detailed in the corresponding report sections (Fig. 2A). The genes are ordered by the number of report sections in which they appear, highlighting those of particular interest owing to evidence from multiple sources. Moreover, all affected genes are hyperlinked to databases, including VICC (https://cancervariants.org), OncoKB [23,24] and CIViC [22], providing additional evidence for their clinical significance.

Mutated genes

The integrated WGTS data for genes with detected SNVs or indels are presented in an interactive "summary" table and "expression profiles" plots, as described previously (Fig. 2B). The "summary" table provides detailed information on mRNA levels for individual genes, along with detected variant tiers, variant characteristics (such as type, class, and allele frequency), and associated proteomic changes. Genes in the table are primarily ordered by variant tier, reflecting their classification and clinical relevance, and secondarily by the percentile difference between the patient's gene expression levels and those observed in the TCGA dataset.

Fusion genes

Gene fusions detected in WTS are summarized in an interactive table that provides comprehensive information, including the number of supporting split reads, the genomic coordinates of the fused genes, the locations of the breakpoints, and whether they overlap with structural events identified in WGS data (Fig. 2C). Gene fusions that support WGS-based SVs or are reported in FusionGDB [20] or CGI [21] databases are prioritised, given the additional evidence for their functional relevance and potential clinical significance in the investigated patient. These fusions are presented in a genomic context via a Circos plot, followed by an interactive “summary” table and “expression profiles” plots that demonstrate the mRNA levels of individual fusion genes observed in both the patient and reference cohorts. Additionally, this section provides visualisations (generated by Arriba [27]) that offer detailed information about the transcripts involved in the predicted fusions. These visualisations include the orientation of the transcripts, the retained exons in the fusion transcript, and statistics about the number of supporting reads.

Structural variants

The integrated WGTS data for genes affected by other SVs detected in the WGS are presented in the “summary” table ordered by the SV score and the difference between the investigated patient and TCGA expression levels (Fig. 2D). In addition to the genomic information associated with detected SVs, such as genomic location and CN changes, the table provides details about affected transcripts and potential overlapping gene fusions. The accompanying “expression profiles” plots display the expression levels of the affected genes as measured in the patient samples and reference cohort samples.

Copy number altered genes

The "CN altered genes" section overlays mRNA expression data with per-gene somatic CN data and information about SNVs and indels (Fig. 2E). It begins with an interactive Manhattan-like plot that presents CN values for genes located within gained or lost regions. Additionally, CNs and expression levels of affected genes are displayed in a scatterplot, which shows the per-gene difference in mRNA expression between the patient sample and the reference cohort as a function of the corresponding CN values. Following these visualisations, the section includes a "summary" table, ordered by the CN value and the difference between the investigated patient and TCGA expression levels, complemented by information derived from the “Mutated genes” section. The "expression profiles" plots depict the expression levels of the affected genes, as measured in the

patient and reference cohort samples. These combined visualisations and data tables allow for the correlation of gene expression with CN alterations, providing insights into their potential clinical significance.

Custom panel of genes of interest

The final section of the report presents results derived exclusively from WTS data and focuses on the expression profiles of a customizable list of GOIs provided by the user or three specific sets of genes: (1) “immune-related genes”, (2) “HRD genes” and (3) “known cancer genes” (Fig. 2F). The mRNA levels of genes involved in these gene sets are presented in the corresponding “summary” tables and “expression overview” box plots.

Results

Clinical applicability evaluation

To assess the clinical utility of RNAsum in interpreting aberrant events detected in WGS, we reviewed 60 patient cases based on the following criteria: 1) availability of matched WGS and WTS data, and 2) RNAsum corroborated the predicted impact on RNA expression of ≥1 clinically reported genetic alteration detected by WGS. All clinically reported variants predicted to impact RNA expression were aggregated. These included truncating SNVs and small indels, inactivating rearrangements, CN losses (low expression), amplifications (high expression) and fusions (fusion mRNA expression).

Across all clinically reportable variants identified by WGS, 205 were classified as variant types potentially amenable to qualitative or quantitative support from WTS data. Among these, 140 (68%) variants were supported at the RNA level, as detected by WTS and reported by RNAsum (Fig. 3). WTS support was determined based on stringent criteria, including activating variants and CN gains showing high relative expression (top quartile compared with the TCGA reference cohort), inactivating variants and CN losses showing low relative expression (bottom quartile compared with the TCGA reference cohort), the detection of gene rearrangements and aberrant transcripts such as fusions. According to the evaluated reports, WTS supported more than half of the clinically reportable variants in 52 out of 60 cases and all reportable variants in 31 cases, highlighting its complementary value to WGS (Table S2 and Fig. S5).

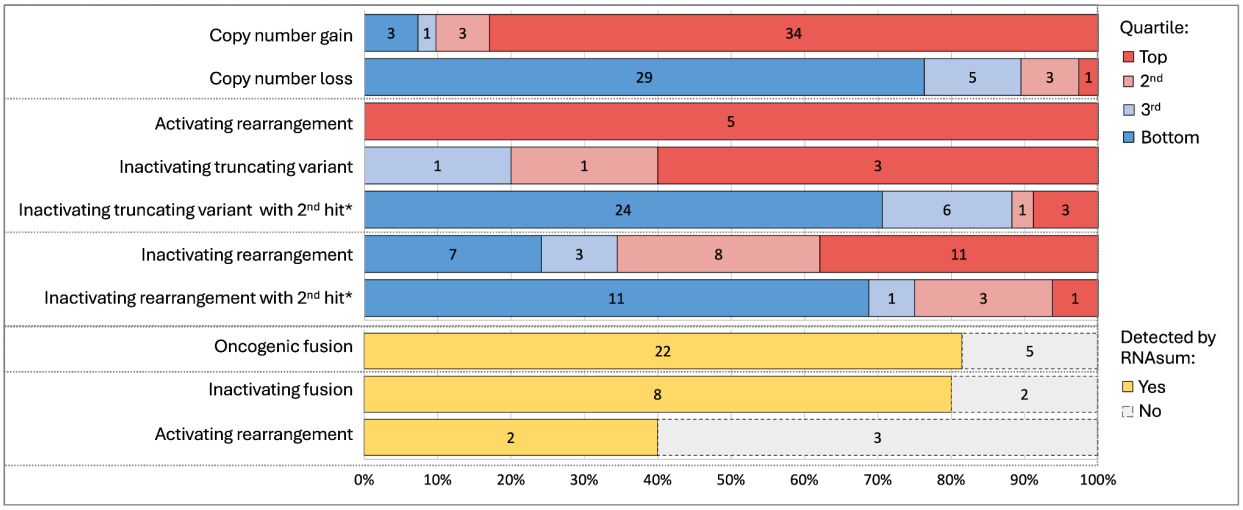

Fig. 3 Enhanced validation of clinically relevant genomic alterations through RNAsum. The bar plot summarises the genomic alterations reported across 60 cases, categorised by WTS-based evidence. The stacked segments within each bar indicate the type of WTS support: relative expression quartiles for copy number alterations, activating and inactivating variants (with or without a second hit*), and detection of aberrant transcripts for structural rearrangements. The numerical labels within the bars indicate the number of variants supported by each WTS evidence category. This overview highlights the value of integrating WGS and WTS data to validate clinically reportable genomic variants, demonstrating RNAsum's role in contextualizing expression changes and fusion events alongside genomic alterations. **Includes inactivating variants on the X chromosome in males.*

These events included truncating mutations (nonsense and frameshifts), CN gains and losses, structural rearrangements, and gene fusions. WTS revealed high relative expression in 83% of the CN gain variants and low relative expression in 76% of the CN loss variants (Fig. S6). Additionally, bottom quartile relative expression was observed in 71% of inactivating truncating variants with a second hit and 69% of inactivating rearrangements with a second hit, where a second hit was defined as inactivating mutation of the second allele or loss of heterozygosity. In the case of inactivating variants without a second hit, WTS provided supporting evidence for low relative expression in fewer instances. This pattern is consistent with residual expression from the remaining wild-type allele. In addition, 76% of the fusions and activating gene rearrangements were detected by RNAsum, including 81% of the oncogenic fusions. This comprehensive review of clinical cases underscores RNAsum's effectiveness in detecting and validating clinically relevant genomic alterations, reinforcing its utility as a robust tool for curating clinical data.

Case studies

To further demonstrate the utility of RNAsum in the clinical interpretation of somatic genomic alterations, we present two cases in which the data curated and summarized by RNAsum played a crucial role in guiding clinically significant decisions.

Case study #1

Whole-genome and whole-transcriptome sequencing was performed on a fresh metastatic triple-negative breast cancer (TNBC) biopsy sample. The tumour had a low tumour mutational burden (1.9 mutations/Mb) and contained somatic features typical of this disease [32]: a *TP53* truncating mutation (*TP53* c.637C>T p.R213*) accompanied by CN neutral loss of heterozygosity (LOH), chromosomal rearrangements of *PTEN* and *RB1,* and amplification of several genomic regions including chr8q.

RNAsum was run for this case using the TCGA breast cancer reference cohort. The expression of the *TP53* c.637C>T nonsense mutation was confirmed, with 47% of 217 reads carrying the mutation, and the expression level estimated in the 41st percentile relative to the TCGA cohort (Fig. 4A). These observations support the expression of a truncated p53 open reading frame and its interpretation as an LOF mutation [33,34]. The WGS-detected rearrangement in *PTEN* involved reciprocal translocation with *PDE1C*. RNAsum identified a *PTEN::PDE1C* fusion (*PTEN* exon 2 - *PDE1C* exon 17), which caused a frameshift and predicted an aberrant PTEN protein truncated early in the phosphatase domain. Coupled with its low expression (2nd percentile), these findings strongly supported the loss of function of *PTEN* (Fig. 4B).

The WGS-detected rearrangement in *RB1* involved a tandem duplication of exon 15. WTS confirmed the expression of transcripts carrying the duplication, resulting in frameshift and predicting truncation after Phe473, which is the N-terminus of the essential RB1 pocket domain. RNAsum also reported low *RB1* expression (Fig. 4C) in the 2nd percentile compared with the TCGA Breast Cancer cohort, adding confidence to the DNA-based interpretation. WGS also revealed amplification of the entire long arm of chromosome 8 (CN range between 10.8 and 14.5). This large amplicon contained 32 cancer-related genes (Fig. 4D), 3 of which are known oncogenes according to the OncoKB [23,24] database: *LYN* (Fig. 4E), *RUNX1T1* (Fig. 4F) and *MYC* (Fig. 4G). RNAsum identified only MYC as highly expressed (92nd percentile relative to the TCGA Breast Cancer cohort). *MYC* amplification and expression were reported as potential

biomarkers for inclusion in a clinical trial (Australian and New Zealand Clinical Trial Registry ACTRN12620001146987).

In addition, amplification of the chr3q22 region, encompassing *PIK3CB* (CN of 34.6), was detected. *PIK3CB* encodes p110β, an activating subunit of PI3 kinase (PI3K). In contrast to its paralogue *PIK3CA*, mutations of *PIK3CB* are rarely observed in TNBC [35–37]. RNAsum reported high *PIK3CB* expression (100th percentile when compared with the TCGA breast cancer cohort) Fig. 4H. This amplification was reported as a variant of unknown clinical importance due to the ability of p110β to drive PI3K signalling [38].

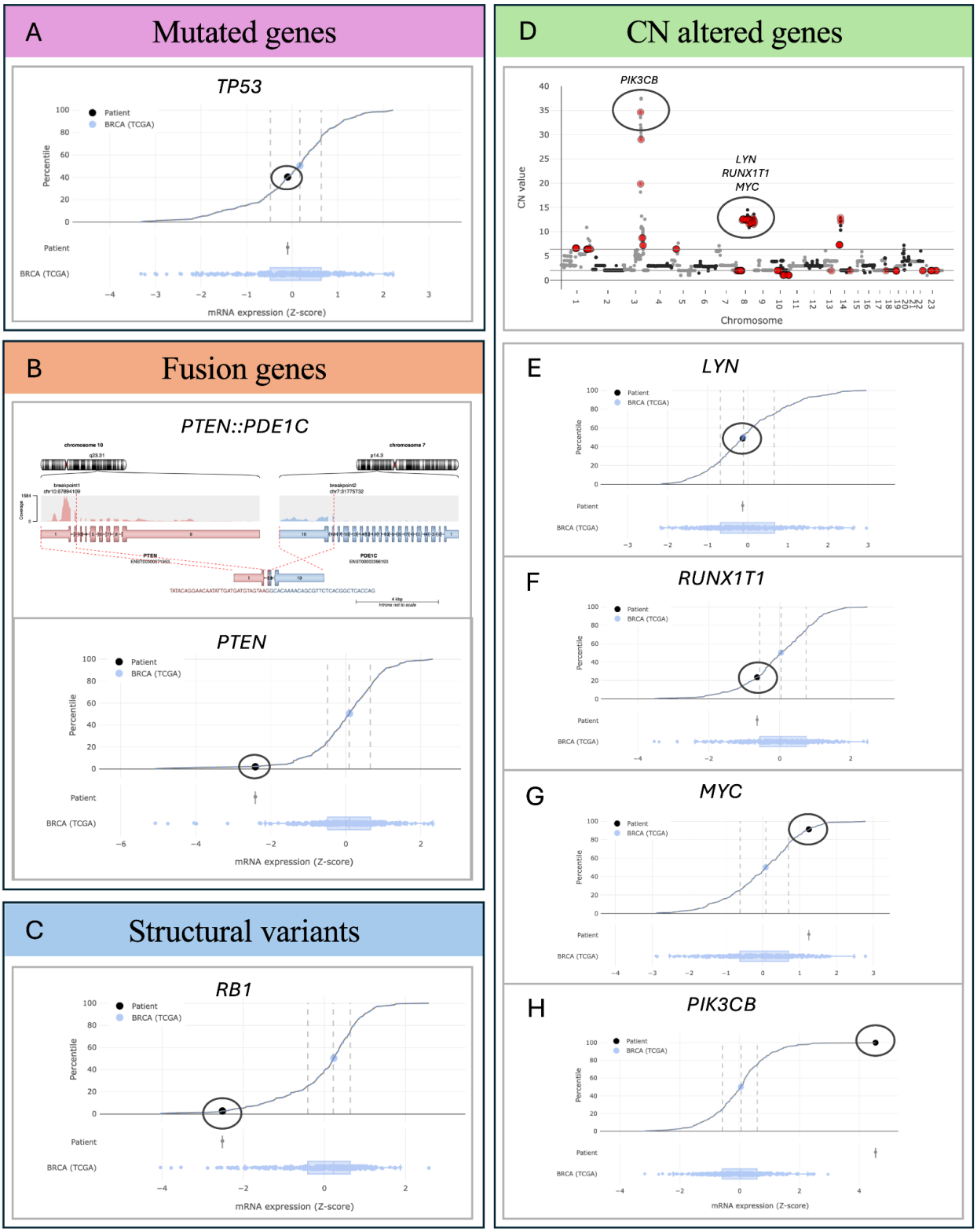
A
Mutated genes
TP53
Patient
BRCA (TCGA)
Percentile
mRNA expression (Z-score)
B
Fusion genes
PTEN::PDE1C
chromosome 10
chromosome 7
PTEN
Patient
BRCA (TCGA)
Percentile
mRNA expression (Z-score)
C
Structural variants
RB1
Patient
BRCA (TCGA)
Percentile
mRNA expression (Z-score)
D
CN altered genes
PIK3CB
LYN
RUNX1T1
MYC
CN value
Chromosome
E
LYN
F
RUNX1T1
G
MYC
H
PIK3CB

Fig. 4 RNAsum visualisations supporting the key genomic and transcriptomic findings from the analysis of TNBC patient. Panel **A** shows the *TP53* expression level at the 41st percentile relative to the TCGA breast cancer cohort. Panel **B** highlights the *PTEN*::*PDE1C* fusion transcript detected by WTS, indicating that an aberrant PTEN protein was truncated early in the *PTEN* phosphatase domain. This fusion is coupled with low *PTEN* expression (2nd percentile), which is consistent with a loss of *PTEN* function. Panel **C** depicts the low expression of *RB1* (2nd percentile compared with TCGA) supporting frameshift mutation and loss of function following tandem duplication involving exon 15, as confirmed by WTS. Panel **D** displays the amplification of chromosome 8q, including 32 cancer-related genes, as well as the chr3q22 region encompassing *PIK3CB*. Panels **E**-**G** show the expression levels of other genes within the 8q amplification, including *LYN* and *RUNX1T1*, which were not highly expressed, and *MYC*, which was highly expressed (92nd percentile). Panel **H** illustrates the amplification of *PIK3CB* in the chr3q22 region, with expression at the 100th percentile relative to the TCGA cohort. *PIK3CB* overexpression is noted as a potential driver of PI3K signalling but remains classified as a variant of uncertain significance.

Case study #2

We performed whole-genome and transcriptome sequencing on a fresh biopsy sample from a woman in her late 20's with hormone receptor-positive metastatic breast cancer who had progressed on aromatase inhibitor and gonadotropin-releasing hormone agonist therapy.

Amplification of *CCND1* (12 copies) and *FGFR1* (20 copies) was detected by WGS (Fig. 5A) and RNAsum reported high expression of both genes relative to the TCGA Breast Cancer cohort (98th and top percentiles, respectively) (Fig. 5B). These two genes are recurrently co-amplified in hormone receptor-positive breast cancer [39].

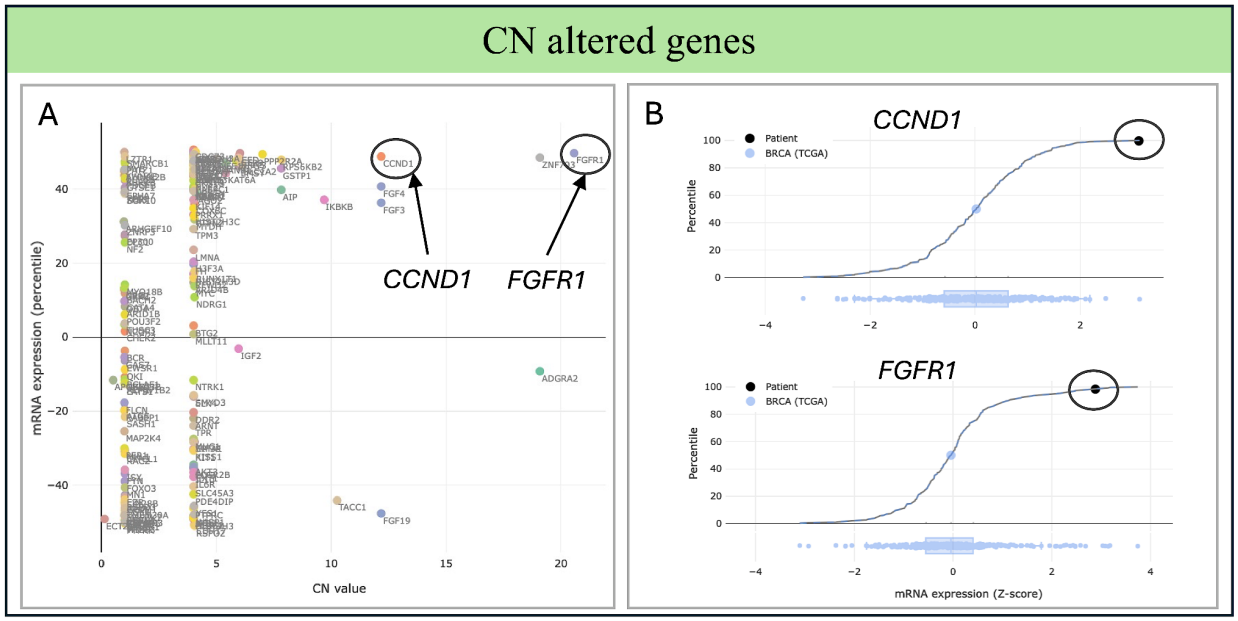

Fig. 5 RNAsum analysis and visualisation of *CCND1* and *FGFR1* co-amplification in metastatic breast cancer. Panel **A** presents a scatter plot of CNVs detected by WGS, plotted against corresponding gene expression levels from WTS. The recurrently co-amplified genes *CCND1* (12 copies) and *FGFR1* (20 copies) are highlighted, where both show markedly elevated expression. Panel **B** shows plots illustrating *CCND1* and *FGFR1* expression levels in the patient's tumour compared with TCGA, positioned at the 100th and 98th percentiles, respectively.

In addition, a hotspot missense mutation (c.1613A>G p.D538G) and complex structural rearrangement of ESR1, encoding oestrogen receptor alpha (ERα), were detected. Genetic alterations in ESR1 are commonly detected in hormone receptor-positive breast cancer patients with acquired resistance to prolonged endocrine therapy [40–45]. WGS identified interchromosomal translocations involving *ESR1* and *PLEKHG1* on chromosome 6, and *HNRNPM* on chromosome 19 (Fig. 6A). RNAsum reported robust expression of an in-frame fusion transcript coupling *ESR1* exon 4 to *PLEKHG1* exon 12, clarifying the interpretation of this complex rearrangement (Fig. 6B-D). The p.D538G substitution observed in this patient confers resistance to oestrogen deprivation therapies but predicts sensitivity to ER degraders such as elacestrant. On the other hand, few clinical studies suggest that fusion genes involving *ESR1* exons 1-6, which encode ERα proteins that lack binding sites for selective ER modulators and degraders, may predict resistance to ER degrader therapy [43,44], highlighting the complexity of acquired endocrine therapy resistance mechanisms. Thus, elucidation of a complex *ESR1* rearrangement in this case illustrates the potential of RNAsum in guiding clinical management.

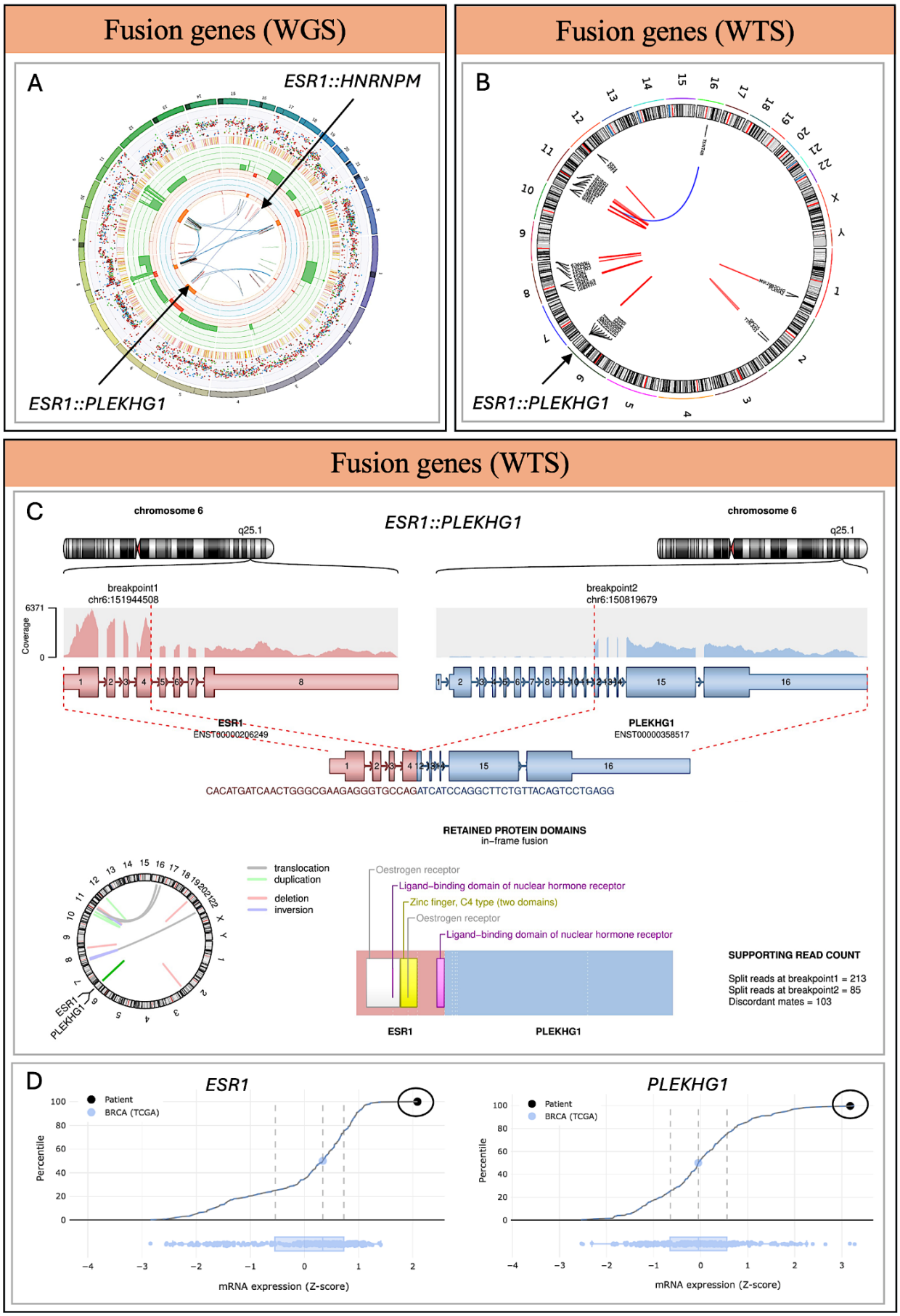
Fusion genes (WGS)
A
ESR1::HNRNPM
ESR1::PLEKHG1
Fusion genes (WTS)
B
ESR1::PLEKHG1
Fusion genes (WTS)
C
ESR1::PLEKHG1
chromosome 6
q25.1
chromosome 6
q25.1
breakpoint1
chr6:151944508
breakpoint2
chr6:150819679
Coverage
6371
0
ESR1
ENST00000206249
PLEKHG1
ENST00000358517
CACATGATCAACTGGGCGAAGAGGGTGCCAGATCATCCAGGCTTCTGTTACAGTCCTGAGG
RETAINED PROTEIN DOMAINS
in-frame fusion
translocation
duplication
deletion
inversion
Oestrogen receptor
Ligand-binding domain of nuclear hormone receptor
Zinc finger, C4 type (two domains)
Oestrogen receptor
Ligand-binding domain of nuclear hormone receptor
ESR1
PLEKHG1
SUPPORTING READ COUNT
Split reads at breakpoint1 = 213
Split reads at breakpoint2 = 85
Discordant mates = 103
D
ESR1
PLEKHG1
Patient
BRCA (TCGA)
Percentile
mRNA expression (Z-score)

Fig. 6 RNAsum visualisation of a complex *ESR1* structural rearrangement detected in hormone receptor–positive metastatic breast cancer. Panels **A**-**D** summarise genomic and transcriptomic evidence from WGS and WTS. **A** Circos plot based on WGS data showing genome-wide somatic alterations. Outer rings display chromosomes (dark regions = centromeres/heterochromatin), somatic variants (SNPs coloured according to the type of base change, e.g. C>T/G>A in red, and Indels in yellow/red), CN changes (red = losses, green = gains), and minor allele CN (orange = LOH, blue = gain). Inner arcs represent structural variants (blue = translocations, red = deletions, green = duplications, black = inversions). WGS revealed multiple *ESR1* rearrangements, including *ESR1*::*HNRNPM* (chr19) and *ESR1*::*PLEKHG1* (chr6), indicating a complex event not fully resolved by WGS alone. **B** WTS Circos plot showing high-confidence fusion transcripts (red lines), including *ESR1*::*PLEKHG1*, which clarified the functional consequence of the complex WGS rearrangement. **C** Schematic of the *ESR1*::*PLEKHG1* fusion detected by WTS, joining *ESR1* exon 4 with *PLEKHG1* exon 12 (chr6:151944508–150819679). The in-frame fusion lacks key ER modulator and degrader binding sites. RNA read support and domain structures are shown. WTS confirmed this as the expressed and functional rearrangement, distinguishing it from the additional *ESR1*::*HNRNPM* event. **D** Expression profiles for *ESR1* and *PLEKHG1*, both at the 100th percentile relative to the TCGA cohort, confirming robust transcription of the fusion partners and supporting the fusion's functional and clinical relevance.

Discussion

We present RNAsum, a robust open-source bioinformatics reporting tool designed to address the complex challenges of integrating and summarizing genomic and transcriptomic data in cancer research. RNAsum leverages RNA-seq data from cancer patients to either verify and complement whole-genome profiling results or provide comprehensive summary statistics from WTS data. This tool tackles several key obstacles in the field, including the tumour-only nature of transcriptome sequencing, the need to harmonize and interpret outputs from diverse analyses and the lack of cancer-specific reference data and complex integration tools. By offering a simple unified approach to data interpretation, RNAsum aims to enhance our understanding of cancer biology and support informed clinical decision-making in precision oncology.

RNAsum incorporates transcriptome profiling into a precision oncology framework by utilizing landmark TCGA RNA-seq data, thereby increasing confidence in the clinical utility of findings. RNAsum integrates data at multiple levels by combining mRNA profiles from the patient and external reference cohorts, overlaying expression data with genomic changes, and annotating findings with clinically relevant information from public knowledge databases. This comprehensive integration process enables the identification and interpretation of cancer-relevant molecular aberrations, facilitating precision oncology by streamlining genomics research, accelerating discoveries, and translating them into actionable clinical insights. Based on comprehensive benchmarking (Fig. S1), we recommend using the representative TCGA reference cohort mode

for routine workflows, particularly when dealing with memory-demanding pan-cancer analyses. This mode offers improved scalability with minimal loss of precision, as 99% of known cancer genes show less than a 10th percentile point difference compared with the full mode, with discrepancies primarily affecting low-expression genes.

RNAsum generates an interactive HTML-based report summarizing clinically relevant information, including mutated genes, fusion genes, SVs, CN alterations, as well as customizable gene sets (e.g. immune markers and DNA damage repair genes). The report provides comprehensive summaries of mRNA expression levels, prioritised findings based on whole-genome sequencing results and public databases, and visual data summaries for quality control and interpretation. Comprehensive reporting approaches often require setting up custom platforms with production-ready database instances [46], which demand continuous management and specialized domain expertise. RNAsum provides a complementary, flexible and platform-agnostic alternative, developed as an R package that can be easily installed and run with minimal technical expertise or management overhead.

Our evaluation of RNAsum's clinical utility highlights its robust performance in supporting genomic findings at the transcriptomic level. In a comprehensive review of 60 cases with matched WGS and WTS data, RNAsum provided additional supporting information in 68% of the clinically reportable variant types where transcriptomic confirmation was deemed possible. Notably, in 87% of cases, at least half of the WGS-detected variants were supported by WTS data, with full support for all possible variants observed in 52% of cases. This high concordance across various mutation types underscores the effectiveness of RNAsum in enhancing the accuracy and reliability of clinical genomic interpretations, supporting its integration into precision oncology workflows and potentially improving patient management and health outcomes.

While traditional molecular pathology techniques such as FISH and IHC are widely used for identifying genomic changes, the integration of RNA-seq data with genomic analysis has proven invaluable for detecting and characterising complex genomic rearrangements, particularly gene fusions. This multiomic approach enables the identification of in-frame chimeric transcripts and the assessment of their functional consequences, which may not be apparent from genomic data alone. Our method has demonstrated the significant utility of WGTS analysis in refining cancer diagnoses and facilitating the discovery of potential therapeutic targets. Using RNAsum, we successfully characterised a complex *HERPUD1*::*RAF1* fusion gene in a case of pancreatic acinar

carcinoma [47] and a diagnostic *CIC*::*DUX4* fusion in a case of *CIC*-rearranged sarcoma [48]. WGTS analysis was instrumental in determining diagnoses for 32% of sarcoma patients in an adolescent and young adult cohort [49], and for 66% of patients in a Cancer of Unknown Primary (CUP) cohort [50]. Finally, RNAsum supported the validation of a novel *RET*::*SEPTIN9* fusion in a pheochromocytoma patient, leading to successful treatment with an *RET* inhibitor [51].

RNAsum offers great flexibility in data processing and parameter selection, but it requires users to provide specific WGTS analysis inputs in a compatible format for integration into the reporting. This presents opportunities for developing modules that can support outputs from various bioinformatics workflows and tools. Although TCGA provides a rich resource of cancer-specific datasets for gene expression comparisons, it has been shown to exhibit unwanted variation arising from library size, tumour purity, and batch effects [52]. Hence, the results should be interpreted within a specific biological context and corroborated through orthogonal evidence or supported by the literature.

To ensure efficient data and dependency management, the RNAsum codebase was developed as an R-package that can be easily accessed and downloaded from GitHub. It offers flexibility in execution, as it can be run on either an HPC or a cloud infrastructure, and is also available as a docker container. The TCGA reference data utilized by RNAsum are conveniently managed through an R data package [53], which is also accessible via GitHub. Moreover, RNAsum is a cost-effective solution that can be seamlessly adapted to run on outputs from WTS analysis exclusively or from both WGS and WTS analysis.

Conclusion

We developed RNAsum, a novel bioinformatics method that deploys a comprehensive approach to data preprocessing, quantification, and integration of multiomics datasets. RNAsum represents a significant advancement in RNA-seq data analysis, seamlessly integrating WGS and WTS data to provide critical insights for cancer patient genomes. Additionally, we demonstrate that by successfully achieving its aims of delivering clinically relevant information and prioritising results for therapeutic intervention for clinical cases, RNAsum enhances patient prognosis, management, and health outcomes in precision oncology settings. The tool's potential to improve data interpretation and accelerate discoveries in genomics research is substantial, paving the way for streamlining research efforts and driving personalised medicine initiatives. To enable its wider

applicability and usage, RNAsum is available as an R package that can be downloaded from GitHub https://github.com/umccr/RNAsum/tags.

Availability and requirements

Project name: RNAsum
Project home page: https://github.com/umccr/RNAsum
Operating system(s): Platform independent
Programming language: R
Other requirements: R 4.0.0 or higher
Licence: MIT

List of abbreviations

WGS: whole-genome sequencing
WTS: whole-transcriptome sequencing
TCGA: The Cancer Genome Atlas
HTS: High-throughput sequencing
SNV: single nucleotide variant
Indels: small insertions and deletions
SV: structural variant
CNV: copy-number variant
TMB: tumour mutational burden
HRD: homologous recombination deficiency
MSI: microsatellite instability
WGTS: whole-genome and transcriptome sequencing
RNA-seq: RNA sequencing
NCI: National Cancer Institute
GDC: Genomic Data Commons
API: Application Programming Interface
GOIs: genes of interest
CGI: Cancer Genome Interpreter
CIViC: Clinical Interpretation of Variants in Cancer
VICC: Variant Interpretation for Cancer Consortium
CPM: counts per million

PCGR: Personal Cancer Genome Reporter
PCA: principal component analysis
RLE: relative log expression
TNBC: triple-negative breast cancer
LOH: loss of heterozygosity
ERα: oestrogen receptor alpha
CUP: Cancer of Unknown Primary

Declarations

Ethics approval and consent to participate

The VCCC PRECISION program received ethics approval from the Peter MacCallum Cancer Centre Human Research Ethics Committee (HREC/48455/PMCC-2018).

Consent for publication

Informed consent was obtained from all subjects involved in the study.

Availability of data and materials

RNAsum is available as an R package distributed under the MIT Licence (https://opensource.org/licenses/MIT), allowing unrestricted reuse with appropriate attribution, through GitHub (https://github.com/umccr/RNAsum) and Anaconda (https://anaconda.org/umccr/r-rnasum) repositories.
The original TCGA [19] datasets used as reference cohorts are available from the NCI Genomic GDC Data Portal (https://portal.gdc.cancer.gov/) via the GDC API (https://gdc.cancer.gov/developers/gdc-application-programming-interface-api) and data release 40.0 (March 29, 2024, https://docs.gdc.cancer.gov/Data/Release_Notes/Data_Release_Notes/#data-release-400).
The clinical case data presented in this study are available upon request from the corresponding author and pending approval from the VCCC PRECISION program investigators. The data are not publicly available due to ethics restrictions.
RNAsum reports from these case studies are deposited in the Zenodo repository (DOI: 10.5281/zenodo.17353510).

Competing interests

SJL reports honoraria from AstraZeneca, Daiichi Sankyo, and Novartis; institutional research funding from Roche, BeiGene, Novartis, SpringWorks Therapeutics, and AstraZeneca; Travel, accommodation and expenses from AstraZeneca. All disclosures are outside the submitted work. Other authors declare that they have no competing interests.

Funding

Whole-genome and transcriptome sequencing were available through enrolment in the VCCC PRECISION program funded by the Victorian Comprehensive Cancer Centre (VCCC, https://vcccalliance.org.au/about-us/victorian-comprehensive-cancer-centre). SG is a recipient of Investigator Fellowship support from the National Health and Medical Research Council of Australia (Project Grant APP1178568).

Author Contributions

SMG, OH, SK and JM conceptualised the study. SK and JM implemented the software. KP was responsible for sequencing. LT, LNV and JHAV contributed to data curation and interpretation. SK and JM drafted the manuscript. JHAV and LNV proposed and wrote the case studies. SK and JM prepared the figures and data summaries. SMG and SJL acquired funding and ethics approval. All authors reviewed and approved the case study.

Acknowledgements

We would like to acknowledge the Collaborative Centre for Genomics Cancer Medicine and Genomics Platform for WGTS and bioinformatics analysis. We would also like to thank all patients, the clinicians and the pathologists who facilitated data acquisition and generation.

Supplementary Material

Provided in a separate file “RNAsum_MS_2025_11_06_suppl.docx” and “Table_S2.xlsx”.

References

1. Cuppen E, Elemento O, Rosenquist R, Nikic S, IJzerman M, Zaleski ID, et al. Implementation of whole-genome and transcriptome sequencing into clinical cancer care. JCO Precis Oncol. 2022;6:e2200245.

2. Kumar-Sinha C, Chinnaiyan AM. Precision oncology in the age of integrative genomics. Nat

Biotechnol. 2018;36:46–60.

3. Alioto TS, Buchhalter I, Derdak S, Hutter B, Eldridge MD, Hovig E, et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. Nature communications. 2015;6:1–13.

4. Rosenquist R, Cuppen E, Buettner R, Caldas C, Dreau H, Elemento O, et al. Clinical utility of whole-genome sequencing in precision oncology. Semin Cancer Biol. 2022;84:32–9.

5. Allgäuer M, Budczies J, Christopoulos P, Endris V, Lier A, Rempel E, et al. Implementing tumor mutational burden (TMB) analysis in routine diagnostics-a primer for molecular pathologists and clinicians. Transl Lung Cancer Res. 2018;7:703–15.

6. de Luca XM, Newell F, Kazakoff SH, Hartel G, McCart Reed AE, Holmes O, et al. Using whole-genome sequencing data to derive the homologous recombination deficiency scores. NPJ Breast Cancer. 2020;6:33.

7. Fujimoto A, Fujita M, Hasegawa T, Wong JH, Maejima K, Oku-Sasaki A, et al. Comprehensive analysis of indels in whole-genome microsatellite regions and microsatellite instability across 21 cancer types. Genome Res. 2020;30:334–46.

8. Roychowdhury S, Chinnaiyan AM. Translating cancer genomes and transcriptomes for precision oncology. CA A Cancer J Clinicians. 2016;66:75–88.

9. Nagy Á, Munkácsy G, Győrffy B. Pancancer survival analysis of cancer hallmark genes. Sci Rep. 2021;11:6047.

10. Jobanputra V, Wrzeszczynski KO, Buttner R, Caldas C, Cuppen E, Grimmond S, et al. Clinical interpretation of whole-genome and whole-transcriptome sequencing for precision oncology. Semin Cancer Biol. 2022;84:23–31.

11. Deyell RJ, Shen Y, Titmuss E, Dixon K, Williamson LM, Pleasance E, et al. Whole genome and transcriptome integrated analyses guide clinical care of pediatric poor prognosis cancers. Nat Commun. 2024;15:4165.

12. Tessier-Cloutier B, Grewal JK, Jones MR, Pleasance E, Shen Y, Cai E, et al. The impact of whole genome and transcriptome analysis (WGTA) on predictive biomarker discovery and diagnostic accuracy of advanced malignancies. J Pathol Clin Res. 2022;8:395–407.

13. Horak P, Heining C, Kreutzfeldt S, Hutter B, Mock A, Hüllein J, et al. Comprehensive genomic and transcriptomic analysis for guiding therapeutic decisions in patients with rare cancers. Cancer Discov. 2021;11:2780–95.

14. Vissers JHA, Mitchell C, Prall OWJ, Lo W-Y, Kanwal S, Luen SJ, et al. Pathologist-initiated whole genome and transcriptome sequencing demonstrates diagnostic utility in resolving difficult-to-diagnose tumors. Genome Med. 2025;17:107.

15. Gong Y, Fan L, Fei X, Zhu Y, Du X, He Y, et al. Targeted Next-Generation Sequencing Reveals Heterogenous Genomic Features in Viscerally Metastatic Prostate Cancer. Journal of Urology. 2021;206:279–88.

16. Pleasance E, Bohm A, Williamson LM, Nelson JMT, Shen Y, Bonakdar M, et al. Whole-genome and transcriptome analysis enhances precision cancer treatment options. Ann Oncol.

2022;33:939–49.

17. Zheng S, Cherniack AD, Dewal N, Moffitt RA, Danilova L, Murray BA, et al. Comprehensive pan-genomic characterization of adrenocortical carcinoma. Cancer Cell. 2016;29:723–36.

18. Kinnersley B, Sud A, Everall A, Cornish AJ, Chubb D, Culliford R, et al. Analysis of 10,478 cancer genomes identifies candidate driver genes and opportunities for precision oncology. Nat Genet. 2024;1–10.

19. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. Cell. 2018;173:371–85.e18.

20. Kim P, Tan H, Liu J, Lee H, Jung H, Kumar H, et al. FusionGDB 2.0: fusion gene annotation updates aided by deep learning. Nucleic Acids Res. 2022;50:D1221–30.

21. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med. 2018;10:25.

22. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet. 2017;49:170–4.

23. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: A precision oncology knowledge base. JCO Precis Oncol [Internet]. 2017;2017. Available from: http://dx.doi.org/10.1200/PO.17.00011

24. Suehnholz SP, Nissan MH, Zhang H, Kundra R, Nandakumar S, Lu C, et al. Quantifying the expanding landscape of clinical actionability for patients with cancer. Cancer Discov. 2024;14:49–65.

25. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14:417–9.

26. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525–7.

27. Uhrig S, Ellermann J, Walther T, Burkhardt P, Fröhlich M, Hutter B, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. Genome Res. 2021;31:448–60.

28. Behera S, Catreux S, Rossi M, Truong S, Huang Z, Ruehle M, et al. Comprehensive genome analysis and variant detection at scale using DRAGEN. Nat Biotechnol. 2025;43:1177–91.

29. Nakken S, Fournous G, Vodák D, Aasheim LB, Myklebost O, Hovig E. Personal Cancer Genome Reporter: variant interpretation report for precision oncology. Bioinformatics. 2018;34:1778–80.

30. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575:210–6.

31. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta:

rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32:1220–2.

32. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61–70.

33. Malcikova J, Tichy B, Damborsky J, Kabathova J, Trbusek M, Mayer J, et al. Analysis of the DNA-binding activity of p53 mutants using functional protein microarrays and its relationship to transcriptional activation. Biol Chem. 2010;391:197–205.

34. Giacomelli AO, Yang X, Lintner RE, McFarland JM, Duby M, Kim J, et al. Mutational processes shape the landscape of TP53 mutations in human cancer. Nat Genet. 2018;50:1381–7.

35. Jiang Y-Z, Ma D, Suo C, Shi J, Xue M, Hu X, et al. Genomic and transcriptomic landscape of triple-negative breast cancers: Subtypes and treatment strategies. Cancer Cell. 2019;35:428–40.e5.

36. Lehmann BD, Colaprico A, Silva TC, Chen J, An H, Ban Y, et al. Multi-omics analysis identifies therapeutic vulnerabilities in triple-negative breast cancer subtypes. Nat Commun. 2021;12:6276.

37. Angus L, Smid M, Wilting SM, van Riet J, Van Hoeck A, Nguyen L, et al. The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. Nat Genet. 2019;51:1450–8.

38. Costa C, Ebi H, Martini M, Beausoleil SA, Faber AC, Jakubik CT, et al. Measurement of PIP3 levels reveals an unexpected role for p110β in early adaptive responses to p110α-specific inhibitors in luminal breast cancer. Cancer Cell. 2015;27:97–108.

39. Bautista S, Theillet C. CCND1 and FGFR1 coamplification results in the colocalization of 11q13 and 8p12 sequences in breast tumor nuclei. Genes Chromosomes Cancer. 1998;22:268–77.

40. Toy W, Shen Y, Won H, Green B, Sakr RA, Will M, et al. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. Nat Genet. 2013;45:1439–45.

41. Robinson DR, Wu Y-M, Vats P, Su F, Lonigro RJ, Cao X, et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat Genet. 2013;45:1446–51.

42. Jeselsohn R, Yelensky R, Buchwalter G, Frampton G, Meric-Bernstam F, Gonzalez-Angulo AM, et al. Emergence of constitutively active estrogen receptor-α mutations in pretreated advanced estrogen receptor-positive breast cancer. Clin Cancer Res. 2014;20:1757–67.

43. Hartmaier RJ, Trabucco SE, Priedigkeit N, Chung JH, Parachoniak CA, Vanden Borre P, et al. Recurrent hyperactive ESR1 fusion proteins in endocrine therapy-resistant breast cancer. Ann Oncol. 2018;29:872–80.

44. Brett JO, Ritterhouse LL, Newman ET, Irwin KE, Dawson M, Ryan LY, et al. Clinical implications and treatment strategies for ESR1 fusions in hormone receptor-positive metastatic breast cancer: A case series. Oncologist. 2023;28:172–9.

45. Heeke AL, Elliott A, Feldman R, O'Connor HF, Pohlmann PR, Lynce F, et al. Molecular

characterization of ESR1 variants in breast cancer. Breast Cancer Res Treat. 2022;196:279–89.

46. Reisle C, Williamson LM, Pleasance E, Davies A, Pellegrini B, Bleile DW, et al. A platform for oncogenomic reporting and interpretation. Nat Commun. 2022;13:756.

47. Prall OWJ, Nastevski V, Xu H, McEvoy CRE, Vissers JHA, Byrne DJ, et al. RAF1 rearrangements are common in pancreatic acinar cell carcinomas. Mod Pathol. 2020;33:1811–21.

48. Aranza S, Roydhouse C, Mitchell C, Vissers JHA, Lo W-Y, Grimmond SM, et al. A rare diagnostically challenging case of CIC-DUX4 sarcoma arising in the neck. Pathology. 2023;55:568–71.

49. Andrew EC, Lewin J, Desai J, Orme L, Hamilton A, Bae S, et al. Clinical impact of comprehensive molecular profiling in adolescents and young adults with sarcoma. J Pers Med. 2024;14:128.

50. Rebello RJ, Posner A, Dong R, Prall OWJ, Sivakumaran T, Mitchell CB, et al. Whole genome sequencing improves tissue-of-origin diagnosis and treatment options for cancer of unknown primary. Nat Commun. 2025;16:4422.

51. Mweempwa A, Xu H, Vissers JHA, Tothill RW, Pattison AD, Fellowes AP, et al. Novel RET fusion RET-SEPTIN9 predicts response to selective RET inhibition with selpercatinib in malignant pheochromocytoma. JCO Precis Oncol. 2021;5:1160–5.

52. Molania R, Foroutan M, Gagnon-Bartsch JA, Gandolfo LC, Jain A, Sinha A, et al. Removing unwanted variation from large-scale RNA sequencing data with PRPS. Nat Biotechnol. 2023;41:82–95.

53. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic acids research. 2016;44:e71–e71.