Wholistic approach: Transcriptomic analysis and beyond using archival material for molecular diagnosis

Nicolas Macagno, Daniel Pissaloux, Arnaud de La Fouchardiere, Marie Karanian, Sylvie Lantuejoul, Francoise Galateau Salle, Alexandra Meurgey, Catherine Chassagne-Clement, Isabelle Treilleux, Caroline Renard, et al.

Wholistic approach: Transcriptomic analysis and beyond using archival material for molecular diagnosis

Nicolas Macagno[1,2,3,4] | Daniel Pissaloux[1,5] | Arnaud de la Fouchardière[1,5] | Marie Karanian[1,3,5] | Sylvie Lantuejoul[1,5,7,8] | Françoise Galateau Salle[1,8] | Alexandra Meurgey[1,3] | Catherine Chassagne-Clement[1] | Isabelle Treilleux[1] | Caroline Renard[1] | Juliette Roussel[1] | Julie Gervasoni[1] | Vincent Cockenpot[1] | Carole Crozes[1] | Aline Baltres[1] | Aurélie Houlier[1] | Sandrine Paindavoine[1] | Laurent Alberti[5] | Adeline Duc[5] | Francois Le Loarer[3,6] | Armelle Dufresne[3,5,9] | Mehdi Brahmi[3,5,9] | Nadège Corradini[3,5,10] | Jean-Yves Blay[3,9,11,12] | Franck Tirode[1,11,5]

[1]Department of Biopathology, Centre Léon Bérard, Lyon, France

[2]Marmara Institute, INSERM, MMG, DOD-CET, Aix-Marseille University, Marseille, France

[3]NETSARC+, French Sarcoma Group (GSF-GETO) Network, France

[4]CARADERM, French Network of Rare Skin Cancers, France

[5]INSERM 1052, CNRS 5286, Cancer Research Center of Lyon (CRCL), Lyon, France

[6]Department of Biopathology, UNICANCER, Bergonié Institute, Bordeaux, France

[7]Grenoble Alpes University, Grenoble, France

[8]MESOPATH, MESOBANK, French Network of Mesothelioma, France

[9]Department of Oncology, UNICANCER, Centre Léon Bérard, Lyon, France

[10]Institute of Hematology and Pediatric Oncology (HOPe), UNICANCER, Centre Léon Bérard, Lyon, France

[11]Université Claude Bernard Lyon I, Univ Lyon, Lyon, France

[12]Headquarters, UNICANCER, Paris, France

Correspondence
Franck Tirode, INSERM 1052, CNRS 5286, Cancer Research Center of Lyon (CRCL), Lyon, France.
Email: franck.tirode@lyon.unicancer.fr

Funding information
Université de Lyon; Agence Nationale de la Recherche, Grant/Award Number: DEvweCAN #ANR-10-LABX-0061; Institut National du Cancer, Grant/Award Numbers: LYRICAN #INCa-DGOS-INSERM_12563, IMAPs #INCa-DGOS_13219; Institut National de la Santé et de la Recherche Médicale; Centre National de la Recherche Scientifique; Centre Léon Bérard

Abstract

Many neoplasms remain unclassified after histopathological examination, which requires further molecular analysis. To this regard, mesenchymal neoplasms are particularly challenging due to the combination of their rarity and the large number of subtypes, and many entities still lack robust diagnostic hallmarks. RNA transcriptomic profiles have proven to be a reliable basis for the classification of previously unclassified tumors and notably for mesenchymal neoplasms. Using exome-based RNA capture sequencing on more than 5000 samples of archival material (formalin-fixed, paraffin-embedded), the combination of expression profiles analyzes (including several clustering methods), fusion genes, and small nucleotide variations has been developed at the Centre Léon Bérard (CLB) in Lyon for the molecular diagnosis of challenging neoplasms and the discovery of new entities. The molecular basis of the technique, the protocol, and the bioinformatics algorithms used are described herein, as well as its advantages and limitations.

1 | INTRODUCTION

Unclassified tumors are challenging to handle, often needing a second opinion by an expert pathology center, resulting in a frequently delayed diagnosis. As an example, mesenchymal tumors including sarcomas are prone to these difficulties, owing to the combination of rarity and diversity, as they represent less than 2% of cancers with more than 100 subtypes.[1,2] Despite continuous advances in their diagnosis, many tumors remain a diagnostic challenge for pathologists, due to the current lack of immunohistochemical or molecular markers.

When confronted with a rare neoplasm that cannot be diagnosed solely on morphological and immunohistochemical criteria, pathologists turn to molecular techniques to identify potential oncogenic alterations that may guide their diagnosis. Several molecular tools can be successfully used to investigate DNA alterations, including karyotype, fluorescence in situ hybridization (FISH), array-comparative genomic hybridization (CGH), targeted Sanger sequencing, and more recently exome sequencing. RNA-based investigations for molecular diagnosis have expanded upon the development of reverse transcription techniques and mainly allowed detection of gene fusion by PCR or targeted RNA sequencing. At the turn of the millennium, gene expression profiling using DNA-based arrays gained reputation with the subclassification of acute leukemias[3,4] but was more prone to be used in research than in routine diagnostic procedures. More recently, with the advent of massively parallel sequencing (next-generation sequencing [NGS]) either on DNA and RNA, molecular diagnostics gained in precision and accuracy. Targeted DNA sequencing and whole-exome DNA sequencing (WES) now allow investigating mutations or variations as well as copy number alterations and loss of heterozygosity throughout the coding genome, increasing the power of detection of oncogenic alterations. Targeted RNA sequencing is now often used in routine diagnosis for the detection of gene fusion in contrast to whole transcriptome sequencing (WTS) or whole exome RNA sequencing (WERS) that are still not commonly used because of their higher costs and their demanding bioinformatics analyzes. WTS comprises different techniques to investigate either all RNAs or only messengers, depending on the RNA preparation. WERS is based on a capture of the RNA on a DNA library composed of (almost) all known gene exons. While WTS is generally used on good quality RNA samples composed of whole RNA molecules, WERS is adapted for recovering fragmented RNAs and therefore well suited for formalin-fixed, paraffin-embedded (FFPE) samples in which fixation procedures break RNA molecules in small fragments. Because of this fragmentation and of the formalin-induced chemical nucleotide modifications, RNA sequencing of FFPE material still remains scarcely used in routine molecular diagnostics, despite the demonstration of its utility.[5–7] Nonetheless, FFPE specimens and slides are widely available while fresh frozen (FF) tumor samples are generally more difficult to obtain, requiring specific costly storage and dedicated platforms.

From a single experiment and with the use of either WTS or WERS, most oncogenic genomic alterations can be detected together with gene expression analyzes, including fusion transcripts, internal-tandem duplications, small insertions/deletions, exonic small nucleotide variations, and aberrant splicing variants. We present here the experience of the Centre Léon Bérard on using WERS on FFPE archival material in molecular diagnostic routine. Each transcriptome data are meticulously annotated and added to a curated database. This ever-incrementing transcriptome database is used iteratively to refine diagnoses, compare cases, and identify novel homogeneous groups of tumors. Rather than presenting a full pipeline or detailed methods, we describe here the global functional scheme of our FFPE RNA-sequencing analysis, which we want to be more prone to be discussed, commented on, and enhanced than to be viewed as set in stone.

2 | OVERVIEW OF THE WORKFLOW

We describe our methods of extraction, quantification, and qualification of RNAs, sequencing, preanalytical quality assessment, methods to detect gene fusions and variations, methods to establish expression data and to validate their quality, clustering analyzes, application of transcriptomic signatures and the utility of a multidisciplinary board

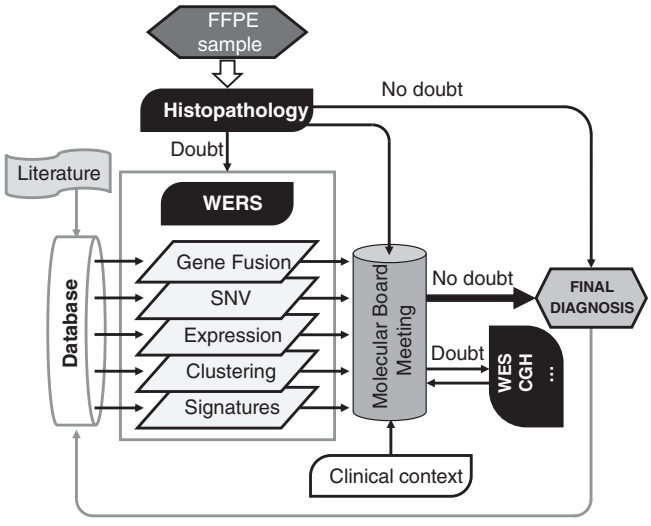

FIGURE 1 Overview of the workflow, with the generation of multiple layers of transcriptomic data and their integration with histopathological and clinical data during a multidisciplinary board for iterative annotation of the database

for the correct annotation of the diagnoses in the database (Figure 1). To date, this approach has been applied in our department to more than 5000 cases, consisting in a majority of cases for which pathologists remained with a doubt about the diagnosis after histopathology reviews, and a minority of control cases with typical morphologies/biomarkers to anchor clustering analyzes.

2.1 | Selection and sampling

Experienced pathologists select the most suitable FFPE block and identify a morphologically well-representative tumor area, with the highest ratio of tumor cells. Whenever possible, the pathologist attempts to exclude areas containing inflammatory infiltrate and any normal tissue to minimize the foreign transcriptomic background. A proportion of nontumoral cells up to 50% has a relatively low impact on the precision of biomarker detection, such as fusion transcripts and mutations. Maximizing this proportion is, however, mandatory to establish a specific expression dataset. The expression profile of a homogeneous tumor will be easier to classify according to the reference cohort. The selected areas of interest are then macrodissected by scraping directly the FFPE sample, or scraping 8 μm-thick unstained slides with a scalpel blade. It is worth noting that compared to frozen material, working on FFPE blocks allows to better target areas of interest that will be globally richer in tumor cells, following which expression profiles of the entire cohort will be more discriminative as a classifier.

2.2 | Methods of extraction and sequencing

For most samples, total RNAs are extracted from FFPE sections using the FormaPure RNA kit (Beckman Coulter, Brea, CA, USA), and the Ambion DNase I (Life Technologies, CA, USA) is used to remove DNA. A minority of samples are extracted using alternative methods (Qiagen RNeasy Micro kit, Hilden, Germany or Promega Maxwell 16 LEV RNA FFPE Purification kit, Madison, WI, USA). RNA quantification is assessed using NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA) measurement and RNA quality using the DV200 value (the proportion of the RNA fragments larger than 200 nt) is assessed by a TapeStation with Hs RNA ScreenTape (Agilent, Santa Clara, CA, USA). Samples with sufficient RNA quantity (>0.5 μg) and quality (DV200 > 30%) are considered qualified for sequencing. A 100 ng of total RNA is used to prepare each individual library with TruSeq RNA Exome (Illumina, San Diego, USA). The 12–32 libraries are pooled at a concentration of 4 nM each together with 1% PhiX. Sequencing is performed (paired end, 2 × 75 cycles) using either NextSeq 500/550 High Output kit on a NextSeq 500 machine or NovaSeq 6000 SP/S1 Reagent Kit on a NovaSeq 6000 machine (Illumina). The exact same method can also be used for frozen material, resulting in comparable transcriptomic data.

2.3 | Methods of bioinformatic analyzes

2.3.1 | Preanalytical: Assessment of the RNAseq quality

RNAseq quality is important for the interpretation of subsequent bioinformatic analyzes. First, samples that did not reach 10 million reads, as assessed using the fastQC tool, are discarded from further expression data analyzes, such as clustering. Then, after generating the BAM files by STAR,[8] using the MarkDuplicates tool from the GATK Picard suite, we also discard samples that did not reach 10 million uniquely mapped reads.

2.3.2 | Detection of gene fusion

A variety of fusion detection tools have been described,[9] all of them having their own specificities, but no fusion detection tool has yet been shown to be universal, fully accurate and without some degree of false negative and positive detections. We have made the assumption that a highly sensitive fusion detection pipeline was more important than a specific one to explore the genome of unclassified neoplasm in our diagnostic procedures, as fusions can be subsequently validated by other techniques such as FISH. Therefore, we have chosen to simultaneously use five orthogonal fusion gene discovery tools: STAR-Fusion,[10] FusionMap,[11] EricScript,[12] FusionCatcher,[13] and Arriba.[14] As a first layer of stratification, fusions predicted in-frame by at least two detection tools are prioritized. But considering the importance of having a high sensitivity for any known fusion gene, the second layer of stratification is the confrontation of the results of each tool with a list of published fusion genes.[15] Fusion genes from this list, even if detected by a single tool, are therefore further taken into account. Faced with undescribed gene fusions, high expression levels of the gene partners suggest a driver aberration and can help stratification, especially when the same fusion is encountered in several neoplasms with common clinicopathological features and/or similar expression profiles. When there is a doubt, gene fusion can be further confirmed by a second technique. Gene fusions that are not in-frame can also be biologically relevant: specifically, a frameshift fusion is a common mechanism that inactivates tumor suppressor genes by truncation,[16] while no-frame fusion genes (meaning that the phase could not be determined generally because of a breakpoint outside of exons boundaries) can induce overexpression of proto-oncogenes by promoter swapping[17] (see Supplementary Figure S1 for an example of *UPS6* fusion gene leading to its overexpression). The presence of numerous fusion transcripts located at the same chromosomal locus, easily visualized on a circos plot, suggests a probable chromothripsis mechanism. Such observations encourage the search for an associated amplification by correlating with expression data of the fusion genes.

2.3.3 | SNV detection

The detection of small nucleotide variations (SNV) may be challenging in archival samples, as formalin fixation damages DNA, not only by fragmenting it but also by inducing its oxidation, leading to C > T substitution. This modification then leads to transition (C > T or G > A) in the sequenced reads and may account for quite a number of false positive SNVs detection. To analyze SNVs from FFPE WERS, we follow RNAseq short variant discovery best practices (SNP + Indels) proposed by the Genome analysis toolkit[18] prior functional annotation using the ANNOVAR tool.[19]

To discriminate potential oncogenic variations within the thousands of SNVs detected per sample, the following filters are applied: (1) synonymous SNV; (2) SNVs with an allele frequency greater than 0.01 in the 1000 Genome Project, NHLBI-ESP 6500 exomes, and in the databases of the Exome Aggregation Consortium; (3) SNVs outside exonic regions or splicing sites; (4) SNVs not annotated as deleterious by at least one of the various mutation score assessment tools; (5) SNVs with a calculated alternative allele ratio less than 0.2 or with less than two supporting reads were discarded. SNVs are subsequently prioritized by their number of supporting reads, by their annotation as pathogenic or probably pathogenic in the CLINVAR database, by their annotation in the Cosmic database, and finally if they are present in a known tumor suppressor gene or oncogene, even if the SNV was not annotated by the previous databases. Applying these filters and focusing on a list of 247 Genes of Interest (GoI) with diagnostic, theranostic, and prognostic values reduced the number of SNV to a few dozens per sample (Supplementary Table S1). Any variation detected in genes that are not included on the current list of GoIs is accessible in the event of a new clinical indication.

2.3.4 | Expression profiles

Several tools to extract expression values from RNA sequencing have been reported, most of them using prior alignment on a reference genome. Recently, two highly similar methods that do not depend on prior alignment were described, namely Kallisto[20] and Salmon.[21] Both methods are based on a quick and efficient pseudo-alignment method with a significant gain in speed and computing resources. In our procedure, expression profiles are extracted from fastq files using the Kallisto tool with the Encode GRCh38 genome annotation. Transcripts from the same gene are summed, and gene expression values are aggregated to all other samples in a single expression matrix. We then apply a log2 transformation of the matrix values +1 and then a quantile normalization between samples using the limma package in the R environment.[22]

2.3.5 | Assessment of the RNAseq expression data quality

The quality of each tumor expression profile should be assessed prior to its use in clustering analyzes. Using the whole normalized expression matrix, a principal component analysis (PCA) is performed to look for unusual profiles. Due to the large collection and variety of tumor types in our dataset, a sample standing as an outlier in the PCA is considered to have abnormal features and is not considered further in the clustering analyzes. Most often, when resubmitted after a second RNA-sequencing starting with a new RNA extraction, expression profiles of these PCA outlier samples fall back within normal distribution, indicating a technical problem rather than a different tumor biology. Although this method performs well for homogeneous or large cohorts (i.e., composed of a single spectrum of tumor entities or large enough to contain various tumor entities), its application is not advised for smaller cohorts in which an outlier may represent a tumor demonstrating a different biology rather than a true technical artifact.

2.3.6 | Clustering analyzes

Unsupervised hierarchical clustering analyzes are methods commonly used to investigate the relationship between samples and rely mainly on correlations between expression profiles. It is well known that depending on the samples and genes selected for these analyzes, the resulting clusters may be quite different. To alleviate this problem, we have chosen to perform a consensus clustering, summing 3000 different clusters (1500 using Pearson and 1500 using Spearman correlations) in which 20% of the samples and 20% of the genes are randomly ignored of the analyzes at each iteration. This method, while demanding large computing resources and processing time, yields robust results (Figure 2A–C). We also perform t-distributed stochastic neighbor embedding (tSNE) analyzes (Figure 2D–F). Both hierarchical clustering and tSNE analyzes are performed on the whole cohort or on the cohort specific to each type of disease (mesenchymal, melanocytic, mesothelial, epithelial, and hematological neoplasms).

2.3.7 | Expression of tumor suppressor genes, oncogenes, and biomarkers

Similarly to immunohistochemistry, relevant biomarkers can be assessed by gene expression, including relevant mutations of key oncogenes, tumor suppressor genes, and kinase-encoding genes, all of which can be detected on a transcriptomic level with diagnostic, prognostic, and theranostic implications.

The expression levels of 162 cancer-associated biomarkers (Supplementary Table S2) are measured and systematically reviewed to aid the diagnostic procedure. For each tumor investigated, the expression of these 162 genes is compared with their respective distribution among all samples, normalized by their median (Figure 3). An intertumoral score is calculated depending on whether their expression values are considered as outliers (over or underexpressed), above or below standard deviation, or within normal deviation. Furthermore, genes are also ranked according to the values of the 31 345 other genes evaluated within the same sample (intratumor comparison). To help uncover potential new biomarkers, the expression value of any

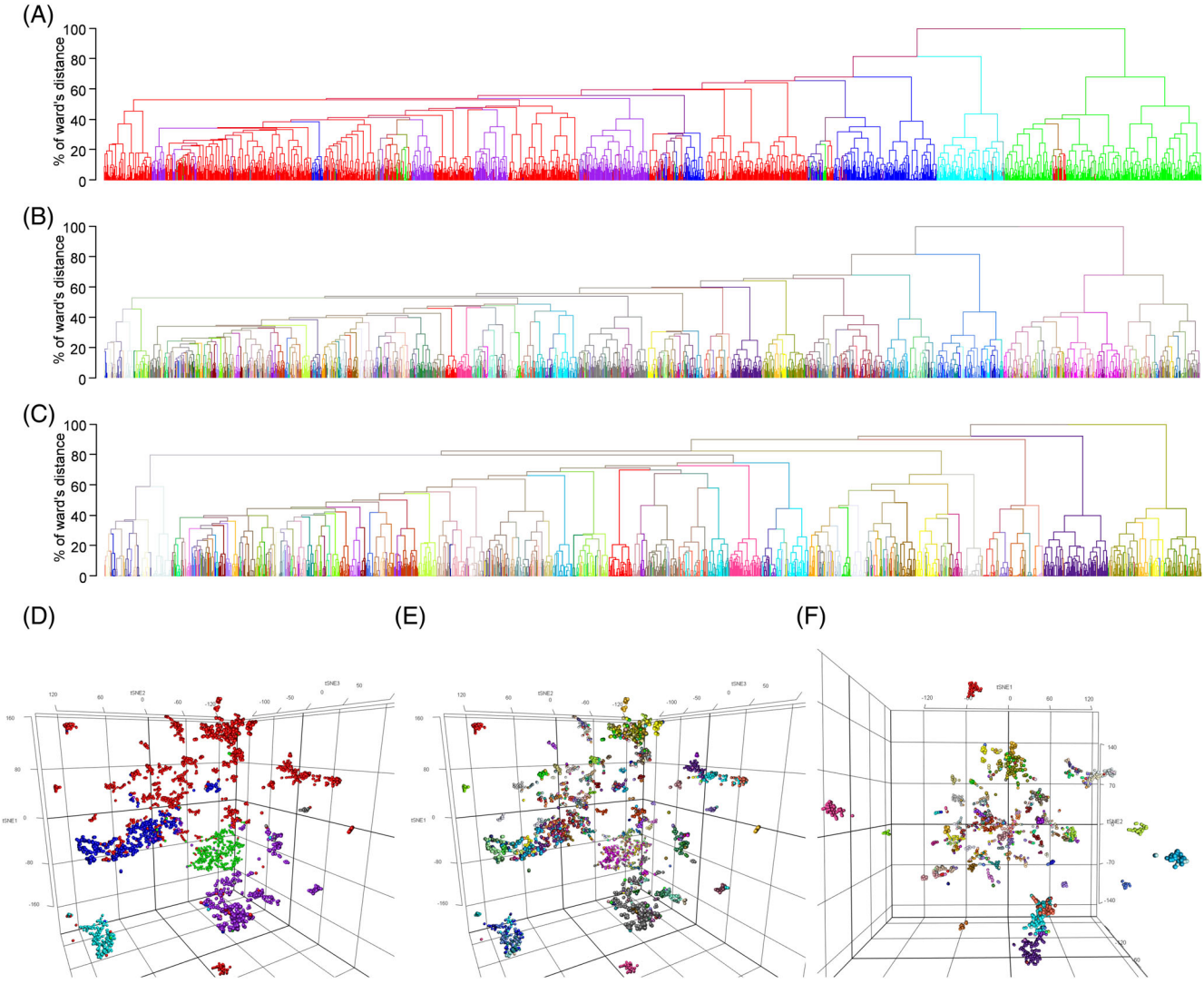

FIGURE 2 Examples of hierarchical clustering (A–C) and t-distributed stochastic neighbor embedding (tSNE) (D–F) analyses. (A) and (D) show 4020 samples of different types of tumors (mesenchymal neoplasms and sarcomas in red, melanocytic neoplasms in green, mesotheliomas in dark blue, epithelial tumors and carcinomas in purple, and hematologic proliferations in cyan). (B) and (E) are as (A) and (D), but each color now represents one of the 260 cancer subtypes. Hierarchical clustering and tSNE analyzes focusing on mesenchymal neoplasms and sarcomas (1894 samples, 153 molecular subtypes)

new GoI can be quickly explored in all samples stratified by tumor subtypes by biologists/pathologists using a simple R script.

The expression level of a fusion gene might also be empirically evaluated by counting the number of supporting reads divided by the number of million mapped reads. Of importance, this expression value could be compared only between samples presenting the exact same fusion point as the number of supporting reads might strongly differ depending on the mappability of the fusion point sequence.

2.3.8 | Gene signatures

From the expression profiles, several transcriptional signatures are analyzed, reflecting either tumor cell behaviors or the microenvironment composition. Single-sample gene set enrichment analysis (ssGSEA)[23,24] is performed on height hallmarks of cancer gene signatures from the Broad Institute Molecular Signature Database[25] (MSigDB): apoptosis, angiogenesis, epithelial mesenchymal transition, glycolysis, G2/M checkpoint, hypoxia, inflammatory response, and oxidative phosphorylation (Figure 4A). Similarly, ssGSEA scores are also calculated for each of the LM22 immune gene signatures[26] and the CINSARC signature.[27] The cell cycle G2/M transition score, compared to the reference cohort, is of particular diagnostic importance for borderline lesions without a clear benign versus malignant status. This score can be used as a surrogate marker to determine whether the lesion is indolent or highly proliferating, in addition to mitotic activity and Ki-67 labeling index. Furthermore, the LM22 signature can be used to basically identify or confirm the presence of a strong inflammatory infiltrate, which is likely to lead to a deviation of the global expression profile (Figure 4B).

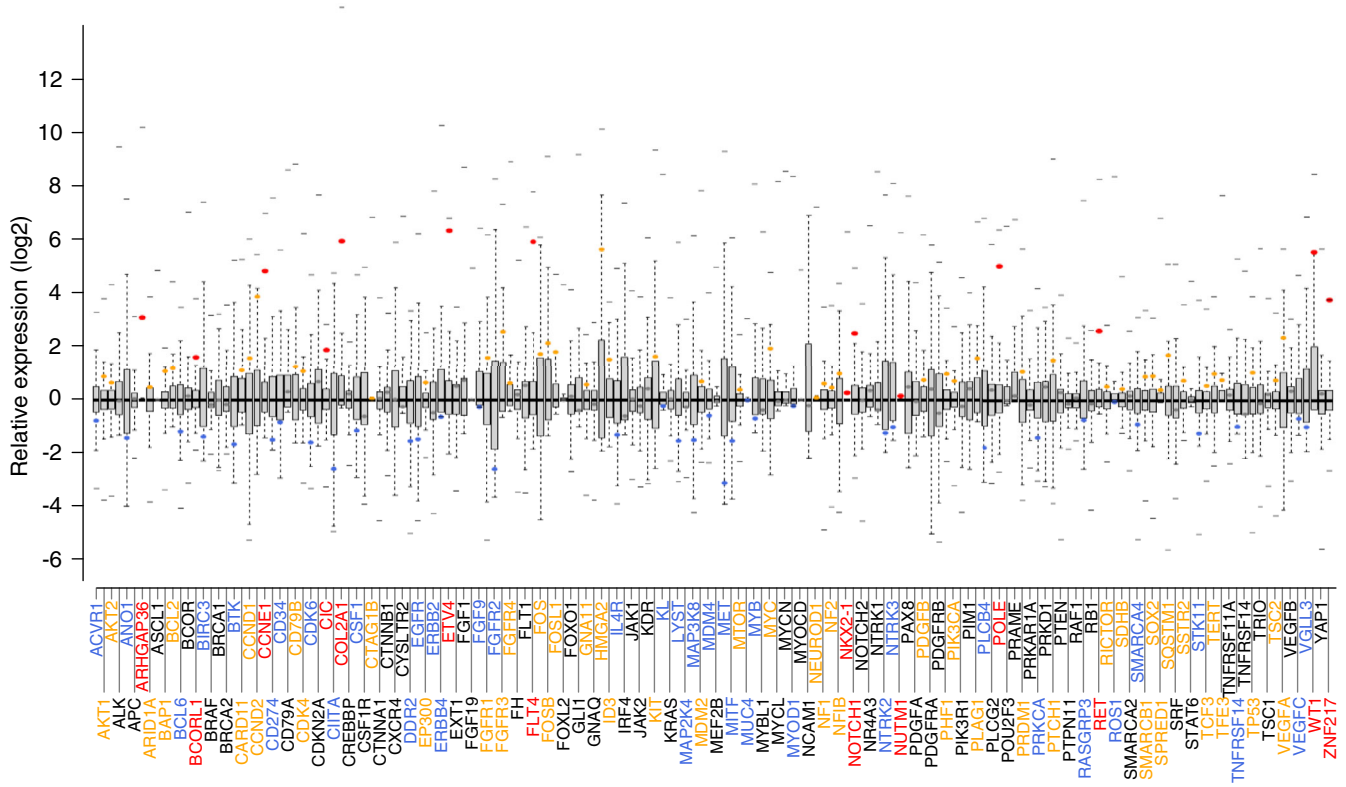

FIGURE 3 Boxplots representing the distribution (through all the samples present at the time of the analysis) of the expression of 162 cancer-associated biomarkers, normalized by their median, and their expression values (circles) in a CIC::DUX4 sarcoma sample. Genes with expression considered as outliers are marked in red (overexpressed) or green (underexpressed), above or below the standard deviation in orange (overexpressed) or blue (underexpressed) or within the normal deviation (in gray). The small horizontal bars represent the maximum and minimum gene expression values observed in our cohort. In this example, overexpression of ETV4 is indicative of a CIC::DUX4 sarcoma

2.4 | Clinicopathological-molecular integration: The board

Many molecular alterations are not specific, and their systematic integration with clinical, radiological, and histopathological data is recommended. This integration aims to achieve a definitive diagnosis and is performed during board meetings with referring pathologists, oncologists, molecular biologists, but also scientists and bioinformaticians strongly involved in translational research. During these sessions, clustering data, gene fusions, mutations, and transcriptomic signatures (notably proliferation score) are reviewed, integrated, confronted with the latest bibliography, and compared to similar cases (similar being either having the same expression profile or the same clinicopathological features) to refine diagnoses or discover new entities. If needed, ancillary techniques such as immunohistochemistry, FISH, WES, aCGH, and high-coverage targeted gene sequencing can be requested to complete the molecular characterization of a novel alteration or reveal alterations that RNA sequencing could have missed. The systematic annotation of the transcriptomic data for each sample, with an integrated, consensus and expert diagnosis, improves the accuracy of the database for cluster analysis over time. Some lesions lack a definitive diagnosis even after the board discussion, but database expansion can help identify incomplete diagnoses afterward (cold case solving). The sensitivity of this approach parallels the exhaustivity of the neoplastic transcriptomic profiles in the database: as more cases are added, robust transcriptomic clusters can emerge and help redefine previously unclassified cases over time.

2.5 | Multicenter generation of data and centralization of bioinformatics analyzes

The generation of transcriptomic data can be technically performed by different centers and subsequently analyzed by the same bioinformatics pipelines, independently of the type of source material (FFPE and FF), but requiring only the use of the same library prep kit (the same set of capture probes in our case) to ensure the production of comparable expression profiles for clustering analyzes. Remote data analysis was performed successfully between the Bergonié Institute (Bordeaux, France) and the Centre Léon Bérard (Lyon, France). Once analyzed, the cases followed the same workflow and were discussed using videoconferencing for proper annotation of a common database. These multicentric management of transcriptomic data make it possible to give even more strength to clustering analyzes, especially for extremely rare entities.

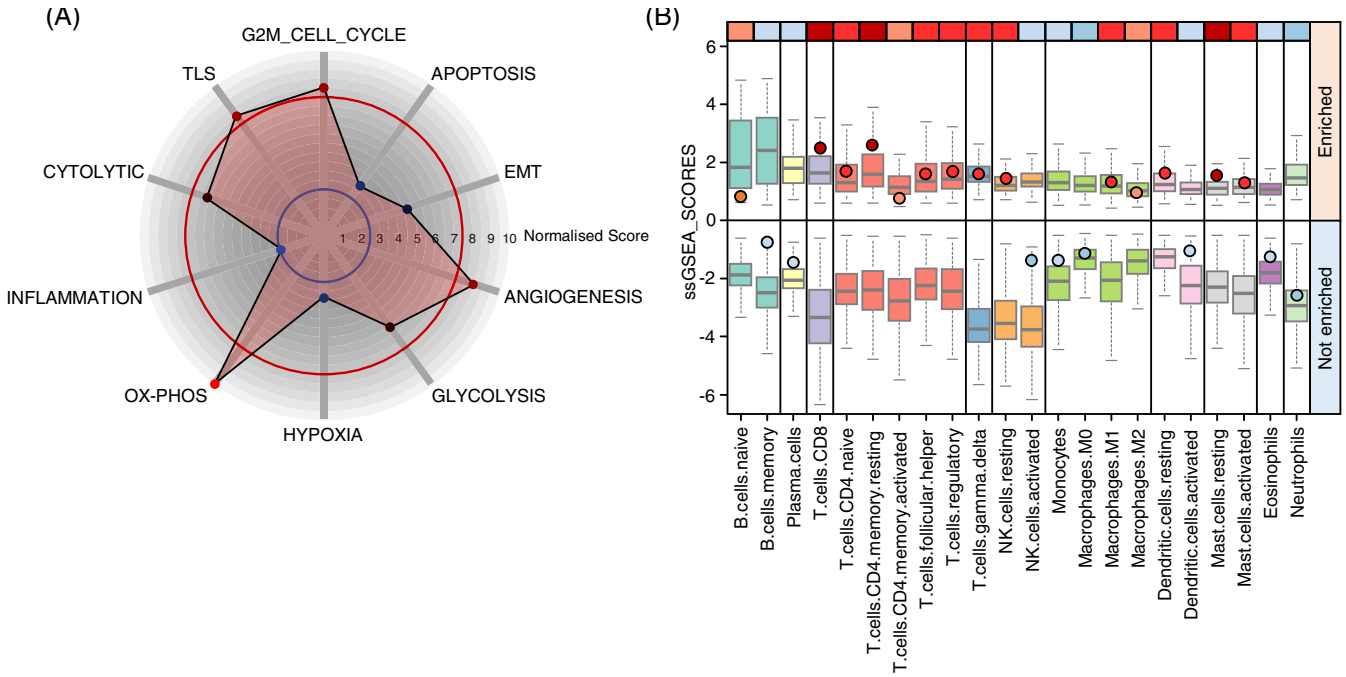

FIGURE 4 Example of gene signature analyses in an unclassified sarcoma. (A) Quantile normalized single-sample gene set enrichment analysis (ssGSEA) score for eight hallmarks of cancers together with the cytolytic and TLS score are presented; 25% (blue circle) and 75% (red circle) of the quantile distribution are shown. This sample presents a clear activation of the oxidative phosphorylation pathway. (B) ssGSEA score for the 22 immune populations of the LM22 signature. Boxplots represent the distribution of each population in all samples when enriched (ssGSEA score above 0) or not (below 0). ssGSEA scores for this unclassified sarcoma are represented by the colored (shades of red or blue for enriched or not enriched scores, respectively) small closed circles. This sample presents a strong immune infiltrate

2.6 | Impact on the discovery of new entities

This integrative management of transcriptomic data allowed the first description and refinement of several neoplastic entities in the field of mesenchymal, melanocytic, mesothelial, and rare tumors, using almost exclusively archived FFPE material: alternative rearrangement of *PDGFD* in dermatofibrosarcoma *protuberans*,[28] endometrial stromal sarcoma in general,[6] sarcoma with *CIC::NUTM1* rearrangement,[29] perivascular myoid neoplasm with *SRF fusion*,[30] well-differentiated rhabdomyosarcomas with SRF fusion,[31] heterogeneity of rhabdomyosarcomas,[32] undifferentiated round cell sarcoma with *CRTC1::SS18* fusion,[5] superficial pleomorphic tumors with *PRDM10* fusion,[33] giant cell tumor with *NCOR2* fusion,[34] acral fibrochondromyxoid tumor with *THBS1* fusion,[35] unclassified sclerosing malignant melanomas with *AKAP9::BRAF* gene fusion,[36] cutaneous melanocytoma with *CRTC::TRIM11* fusion,[37,38] melanocytic myxoid spindle cell tumor with *ALK* rearrangement (MMySTAR),[39] melanocytoma with concomitant mutation of *IDH1* and *NRAS*,[40] *CYSLTR2*-mutant cutaneous melanocytic neoplasms,[41] agminated melanocytic Spitz nevi arising in a giant congenital hyperpigmented macule with three-way complex rearrangement of *TRPM1::PUM1::LCK*,[42] agminated Spitz nevi with *GOPC::ROS1* mosaicism,[43] primary melanoma of the lung with *FNBP1::BRAF* fusion,[44] melanocytic tumors with either *RASGRF1*[45] or *RASGRF2* fusion,[46] melanocytic Spitz neoplasms with *MAP3K8* fusion,[47] melanocytic spitz tumors in general,[43,48–50] clear cell tumor with melanocytic differentiation and *MITF::CREM*[51] and *ACTIN::MITF* fusion,[52] and solid papillary mesothelial tumor.[53]

3 | ADVANTAGES AND LIMITATIONS

3.1 | RNA-sequencing for diagnosis

Whole transcriptome and whole exome RNA sequencing are extremely powerful diagnostic techniques, as they allow exploring gene expression and comparing expression profiles. Several publications now describe RNA sequencing to improve the diagnosis of cancer in general,[54] glioma,[55] glioblastoma,[56] renal cell carcinoma,[57] breast carcinoma,[58–62] ovarian serous carcinoma,[63] lung idiopathic pulmonary fibrosis,[64] hematological malignancies,[65] and sarcoma[66] among others. WTS and WERS combine expression profiles and high-throughput detection of molecular biomarkers, including the detection of variations or small insertion/deletions. However, these techniques may have difficulties in identifying certain particular alterations, such as truncating mutations, as they may lead to mRNA decay. Although promoter swapping could be detected by aberrant expression of a gene, detection of such an event might remain challenging. Similarly, the use of alternative transcription initiation sites, as in the *ALK* gene,[67] may be missed unless transcript isoforms are also investigated.

3.2 | Clustering

Clustering of the transcriptomic profiles shows great performance and hints to diagnoses even when a specific molecular hallmark is lacking,

but depends on the exhaustivity and correct annotation of reference cases. If enough cases are present, clustering approaches are sufficient to identify tumors with similar expression profiles, similar oncogenic alterations and/or similar phenotypes. This is clearly demonstrated for CIC::DUX4 sarcomas, *PLAG1*-rearranged tumors, and *BCOR*-ITD tumors, all aggregating in robust and specific clusters and for which fusion detection tools sometimes miss detection of genomic rearrangements. For example, more than a third of the cases of lipoblastoma and fibrous lipoblastoma lacked detectable *PLAG1* fusion (RNA-seq and FISH), but their diagnosis was supported by the combination of lipoblastoma morphology, close clustering with *PLAG1*-rearranged lipoblastomas, and high levels of PLAG1 expression. Similarly, a quarter of the sporadic cases of malignant peripheral nerve sheath tumor (MPNST) in our cohort, tested due to pathological ambiguity with melanoma or to rule out rhabdomyosarcoma, lacked specific fusion, but their close clustering with other MPNSTs favored their diagnosis. In this aspect, clustering analysis has higher sensitivity than fusion detection alone. Similarly, it may be able to separate tumors with similar fusion genes as demonstrated for the *YWHAE*::*NUTM2B* fusion gene that could be identified in two distinct groups corresponding to high- and low-grade endometrial stromal sarcoma.[6] The same approach was successfully used when the histopathological differential diagnosis on microbiopsy was difficult to distinguish between clear cell sarcoma and angiomatoid fibrous histiocytoma, due to a common *EWSR1*::*ATF1* fusion.

It is very important to note that clustering does not resolve all diagnostic conundrums, notably for some unclassified sarcomas, undifferentiated pleomorphic sarcomas, sarcomas with complex genomics, and dedifferentiated (sarcomatoid) carcinomas. In contrast, transcriptomic profiles and clustering are useful in suggesting the diagnosis of undifferentiated melanoma and mesothelioma. The major flaw in clustering analyzes is the strong dependence on the cell content of the sample and its microenvironment. In fact, the presence of cells with specific expression signatures, such as inflammatory cells or highly differentiated normal cells from certain organs, could profoundly impact the correlations involved in the cluster algorithms, resulting in "organ clusters." Therefore, the pathologist is of utmost importance in selecting sample areas that contain as many tumor cells as possible and as few normal surrounding cells as possible. Alternatively, genes from specific cellular compartments could be omitted from the clustering analyzes. As an example, we found some cutaneous sarcomas that clustered together with melanomas because of the presence of some epidermal tissue. Mining different databases such as GTEX or msigDB, we identify a set of genes highly expressed in normal skin, which, when removed from the clustering analyses, often lead to a relocalization of these tumors into the sarcoma subgroup.

3.3 | The caveats of FFPE

Because formalin fixation in archival material is prone to artifacts and because of the reputation of uncertainty regarding the quality of FFPE-derived RNA due to their degradation, RNA-seq techniques were initially based on nucleic acids extracted from FF samples. Pre-, per- and postfixation conditions impact molecular techniques and should be carefully controlled.[68] Unfortunately, before awareness of the dramatic impact of preanalytical conditions on the performance of sequencing techniques, samples were embedded and stored under varying conditions without specific care. Formalin fixation and paraffin embedding induce a systematic chemical change in RNAs, including fragmentation, crosslinking, and loss of poly-A tails, which limits the effectiveness of oligo-dT primers for reverse transcription.[68] Several studies have reported the negative impact of archiving time on the RNA quality of FFPE samples over a long period.[69,70] A negative impact on sequencing success rates has been reported for surgical samples storage at 4°C and fixation without slicing,[71] highlighting the need for proper processing even before fixation.

Despite these disadvantages and when analytical and preanalytical conditions were optimized, many studies have reported a high correlation of gene expression between FFPE and FF samples,[54,57,59,72–82] although the RNA extracted from the FFPE samples was most often degraded and continued to degrade with age. Some studies have also investigated different extraction techniques or sequencing methods to optimize the workflow for archival materials.[83,84] Our experience is consistent with these results and demonstrates that FFPE samples can be used for RNA sequencing with sufficient reliability for diagnosis when sample handling conditions were carefully controlled.

Beyond the advantage of being widely available, FFPE samples can be stored at room temperature and easily transported, thus requiring significantly less storage resources and logistics compared to FF samples, with a significant impact on the costs. More importantly, most FFPE samples have been morphologically evaluated by a pathologist prior to sequencing, which is not always the case for fresh frozen samples, although performing such a control is of utmost importance for any molecular analyzes, especially for clustering approaches. We believe that systematic pathological review of samples prior to analysis and its correlation with transcriptomic data afterward are two fundamental steps of the whole process. As stated earlier, unless generated via the same library used for FFPE, the expression profiles generated from FF cannot be mixed with FFPE profiles for clustering analyses. In this context, we believe that precious FF should be prioritized for research purposes, and harmonization of preanalytical procedures regarding RNA sequencing from FFPE should be encouraged to ensure comparability of the expression profiles and decentralized analysis.

Compared to DNA-based approaches, RNA-seq has a lower sensitivity considering the lower number of reads compared to massively parallel sequencing on DNA, since the latter is less prone to high degradation. In this context, canonical mutations involving key oncogenes and kinase encoding genes can be overlooked as a consequence of quality parameters (filtering and duplicate), while other inactivating alterations inducing dramatically reduced expression of tumor suppressor gene mRNA cannot be detected at all. The systematic discussion during the multidisciplinary board can hint at the need for a second look at the sequences or performing an orthogonal approach to reveal cryptic biomarkers.

3.4 | Alternative techniques and perspectives

Sometimes, high (as for *MDM2*) or low (as for *CDKN2A*) gene expression levels are suggestive of a gene amplification or deletion, but their detections by RNA-seq are highly dependent on the fraction of tumor cells in the sample. When WERS is not able to identify a probable oncogenic event or is not informative enough to classify a given sample, other techniques such as WES or CGH could then be used to identify unseen mutations or copy number variations (CNVs).

Improvements in our techniques and algorithms could involve the implementation of specific transcriptomic signatures to predict recurrence and/or metastatic risk, response to chemotherapy and immunotherapy to guide therapeutic decisions. Furthermore, the exponential data generated by WERS could benefit from deep learning algorithms to merge and annotate transcriptomes with CNV (aCGH or WES), methylation profiling, radiological images, histological whole slide images, and clinical data in order to find relevant patterns or correlations.

Array-based DNA methylation profiling has been proven to be a powerful tool for the diagnosis of sarcoma and has been successfully applied to show proximity[85] or to discriminate between histologically overlapping neoplasms,[85–87] with a diagnostic prediction for 55% of cases in a recent validation study.[88] This method also provides useful information on CNV. Automatically combining and integrating different levels of omics data, such as the methylome, genome, and transcriptome, could reveal new nosological entities.

The future development of the molecular diagnostic routine could involve the use of single-cell RNA or DNA sequencing. This technique could have real potential to precise the diagnosis of unclassified pleomorphic tumors or tumors with complex genomic alterations. Indeed, investigation of the gene expression of each cell of such a tumor fragment could allow the detection of subclones carrying specific alterations that may be of prime importance for the development (and/or evolution) of the tumor. One step further is the association of this single cell knowledge with information on cell localization. In the last decade, spatial transcriptomics, which allows the spatial distribution of mRNA molecules in histological sections to be defined, as well as their subcellular localization, has been developed. A wide range of methods have been described to reveal the cellular and spatial organization of the transcriptome of neoplastic cells in relation to the surrounding or infiltrating microenvironment, some of which can be performed in FFPE. Such techniques include, among others, in situ sequencing, cyclic-ouroboros smFISH (osmFISH), multiplexed error-robust fluorescence in situ hybridization, single-molecule FISH, spatially resolved transcript amplicon readout mapping (STARmap), sequential FISH, and Visium, Slide-seq, Expansion sequencing (ExSeq), and High Definition Spatial transcriptomics.[89–95] These methods are not yet suitable for the diagnostic routine, but the constant development of more specialized techniques could already be used if not for their cost.

Although RNA sequencing represents the closest reflection of RNA expression, it does not reflect either actual protein expression or cellular localization. The probable ultimate analysis to be performed, somehow mimicking histopathological morphology investigation, would be to explore the entire protein content of cells, tumoral or from their environment. With the development of ever-sensitive mass spectrometry techniques, quantitative proteomic approaches at the scale of a single cell are being developed[96] and enable one to foresee a future where it may be possible to access the cellular localization of at least thousands of expressed proteins. Similarly, the continuous development of highly multiplexed immunofluorescence techniques, which already allow the investigation of several tens of proteins at the same time, might provide a wonderful diagnostic tool to investigate the subcellular and cellular localization of all important biomarkers in a single experiment.

Finally, with the increasing amount of acquired data, either from RNA sequencing, DNA sequencing, methylation profiling, and proteomics, there is a tremendous need for novel bioinformatics integrative tools. Current developments include deep learning approaches to align transcriptomic and spatial data[93] to further improve spatial resolution and sensitivity. Therefore, the future of molecular pathology will more than ever depend on the expertise of bioinformaticians to reduce the enormous amount of raw acquired data to essential information, allowing for rapid interpretation by pathologists and biologists.

4 | CONCLUSION

Gene expression profiling can be successfully applied to archival material to diagnose challenging unclassified neoplasms, discover new entities, pinpoint aggressiveness, and refine the tumor microenvironment. Involvement of a multidisciplinary board with molecular biologists, physicians, scientists, bioinformaticians, and pathologists is a sine qua non condition for the correct integration and annotation of all generated datasets. Future directions are represented by spatial transcriptomics, in which morphological and molecular correlation will play a pivotal role.

ACKNOWLEDGMENTS

The authors wish to thank all the pathologists who provided material tumor. This work was partly supported by the Centre Léon Bérard, Centre National de la Recherche Scientifique (CNRS), Institut National de la Santé et de la Recherche Médicale (INSERM), Institut National du Cancer (INCa) (IMAPs #INCa-DGOS_13219 and LYRICAN #INCa-DGOS-INSERM_12563), the Direction Générale de l'Offre de Soins (DGOS) (IMAPs #INCa-DGOS_13219 and LYRICAN #INCa-DGOS-INSERM_12563) and from the Agence Nationale de la Recherche (DEvweCAN #ANR-10-LABX-0061) and the Université de Lyon (DEvweCAN #ANR-10-LABX-0061).

CONFLICT OF INTEREST

All authors declare no conflict of interest to disclose.

ORCID

Nicolas Macagno https://orcid.org/0000-0002-9882-2162
Daniel Pissaloux https://orcid.org/0000-0003-1118-950X
Arnaud de la Fouchardière https://orcid.org/0000-0003-2251-8241
Marie Karanian https://orcid.org/0000-0003-0434-636X
Sylvie Lantuejoul https://orcid.org/0000-0003-4310-6975
Françoise Galateau Salle https://orcid.org/0000-0002-2814-1644
Isabelle Treilleux https://orcid.org/0000-0003-3919-5506

Vincent Cockenpot https://orcid.org/0000-0002-7470-4436
Adeline Duc https://orcid.org/0000-0003-4581-7949
Francois Le Loarer https://orcid.org/0000-0001-8582-9819
Mehdi Brahmi https://orcid.org/0000-0001-9904-166X
Nadège Corradini https://orcid.org/0000-0002-1942-1222
Jean-Yves Blay https://orcid.org/0000-0001-7190-120X
Franck Tirode https://orcid.org/0000-0003-4731-7817

REFERENCES

1. de Pinieux G, Karanian M, Le Loarer F, et al. Nationwide incidence of sarcomas and connective tissue tumors of intermediate malignancy over four years using an expert pathology review network. *PLoS One*. 2021;16(2):e0246958.
2. WHO Classification of Tumours Editorial Board. *Soft Tissue and Bone Tumours*. Vol 3. 5th ed. International Agency for Research on Cancer; 2020:607.
3. Ben-Neriah Y, Daley GQ, Mes-Masson AM, Witte ON, Baltimore D. The chronic myelogenous leukemia-specific P210 protein is the product of the bcr/abl hybrid gene. *Science*. 1986;233(4760):212-214.
4. Ilaria RL, Van Etten RA. P210 and P190(BCR/ABL) induce the tyrosine phosphorylation and DNA binding activity of multiple specific STAT family members. *J Biol Chem*. 1996;271(49):31704-31710.
5. Alholle A, Karanian M, Brini AT, et al. Genetic analyses of undifferentiated small round cell sarcoma identifies a novel sarcoma subtype with a recurrent CRTC1-SS18 gene fusion. *J Pathol*. 2018; 245(2):186-196.
6. Brahmi M, Franceschi T, Treilleux I, et al. Molecular classification of endometrial stromal sarcomas using RNA sequencing defines nosological and prognostic subgroups with different natural history. *Cancers (Basel)*. 2020;12(9):E2604.
7. Perret R, Escuriol J, Velasco V, et al. NFATc2-rearranged sarcomas: clinicopathologic, molecular, and cytogenetic study of 7 cases with evidence of AGGRECAN as a novel diagnostic marker. *Mod Pathol*. 2020;33:1930-1944.
8. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
9. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol*. 2019; 20(1):213.
10. Haas BJ, Dobin A, Stransky N, et al. STAR-fusion: Fast and accurate fusion transcript detection from RNA-Seq. BioRxiv. 2017;120295. http://biorxiv.org/content/early/2017/03/24/120295.abstract
11. Ge H, Liu K, Juan T, Fang F, Newman M, Hoeck W. FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics*. 2011;27(14):1922-1928.
12. Benelli M, Pescucci C, Marseglia G, Severgnini M, Torricelli F, Magi A. Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics*. 2012;28(24):3232-3239.
13. Nicorici D, Satalan M, Edgren H, et al. FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data [internet]. *bioRxiv*. 2014;011650. http://biorxiv.org/lookup/doi/10.1101/011650
14. Uhrig S, Ellermann J, Walther T, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res*. 2021;31(3): 448-460.
15. Mitelman F, Johansson B, Mertens F. Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer [Internet]. 2021. https://mitelmandatabase.isb-cgc.org
16. Sato T, Sekido Y. NF2/Merlin inactivation and potential therapeutic targets in mesothelioma. *Int J Mol Sci*. 2018;19(4):E988.
17. Kas K, Voz ML, Röijer E, et al. Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12) translocations. *Nat Genet*. 1997;15(2):170-174.
18. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303.
19. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
20. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34(5):525-527.
21. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417-419.
22. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
23. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009;462(7269):108-112.
24. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43): 15545-15550.
25. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425.
26. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods*. 2015;12(5): 453-457. https://doi.org/10.1038/nmeth.3337
27. Chibon F, Lagarde P, Salas S, et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. *Nat Med*. 2010; 16(7):781-787.
28. Dadone-Montaudié B, Alberti L, Duc A, et al. Alternative PDGFD rearrangements in dermatofibrosarcomas protuberans without PDGFB fusions. *Mod Pathol*. 2018;31:1683-1693.
29. Le Loarer F, Pissaloux D, Watson S, et al. Clinicopathologic features of CIC-NUTM1 sarcomas, a new molecular variant of the family of CIC-fused sarcomas. *Am J Surg Pathol*. 2019;43(2):268-276.
30. Karanian M, Kelsey A, Paindavoine S, et al. SRF fusions other than with RELA expand the molecular definition of SRF-fused perivascular tumors. *Am J Surg Pathol*. 2020;44:1725-1735.
31. Karanian M, Pissaloux D, Gomez-Brouchet A, et al. SRF-FOXO1 and SRF-NCOA1 fusion genes delineate a distinctive subset of well-differentiated rhabdomyosarcoma. *Am J Surg Pathol*. 2020;44:607-616.
32. Butel T, Karanian M, Pierron G, et al. Integrative clinical and biopathology analyses to understand the clinical heterogeneity of infantile rhabdomyosarcoma: a report from the French MMT committee. *Cancer Med*. 2020;9(8):2698-2709.
33. Perret R, Michal M, Carr RA, et al. Superficial CD34-positive fibroblastic tumor and PRDM10-rearranged soft tissue tumor are overlapping entities: a comprehensive study of 20 cases. *Histopathology*. 2021;79(5):810-825.
34. Brahmi M, Alberti L, Tirode F, et al. Complete response to CSF1R inhibitor in a translocation variant of Teno-synovial giant cell tumor without genomic alteration of the CSF1 gene. *Ann Oncol*. 2018;29(6): 1488-1489.
35. Bouvier C, Le Loarer F, Macagno N, et al. Recurrent novel THBS1-ADGRF5 gene fusion in a new tumor subtype "Acral FibroChondroMyxoid tumors". *Mod Pathol*. 2020;33:1360-1368.
36. Perron E, Pissaloux D, Neub A, et al. Unclassified sclerosing malignant melanomas with AKAP9-BRAF gene fusion: a report of two cases and review of BRAF fusions in melanocytic tumors. *Virchows Arch*. 2018; 472(3):469-476.
37. Ko JS, Wang L, Billings SD, et al. CRTC1-TRIM11 fusion defined melanocytic tumors: a series of four cases. *J Cutan Pathol*. 2019;25: 810-818.

38. Cellier L, Perron E, Pissaloux D, et al. Cutaneous Melanocytoma with CRTC1-TRIM11 fusion: report of 5 cases resembling clear cell sarcoma. *Am J Surg Pathol*. 2018;42(3):382-391.
39. Perron E, Pissaloux D, Charon Barra C, et al. Melanocytic Myxoid spindle cell tumor with ALK rearrangement (MMySTAR): report of 4 cases of a nevus variant with potential diagnostic challenge. *Am J Surg Pathol*. 2018;42(5):595-603.
40. Macagno N, Pissaloux D, Etchevers H, et al. Cutaneous melanocytic tumors with concomitant NRASQ61R and IDH1R132C mutations: a report of 6 cases. *Am J Surg Pathol*. 2020;44:1398-1405.
41. Goto K, Pissaloux D, Paindavoine S, Tirode F, de la Fouchardière A. CYSLTR2-mutant cutaneous melanocytic neoplasms frequently simulate "pigmented epithelioid Melanocytoma," expanding the morphologic Spectrum of blue tumors: a Clinicopathologic study of 7 cases. *Am J Surg Pathol*. 2019;43:1368-1376.
42. Goto K, Pissaloux D, Durand L, Tirode F, Guillot B, de la Fouchardière A. Novel three-way complex rearrangement of TRPM1-PUM1-LCK in a case of agminated Spitz nevi arising in a giant congenital hyperpigmented macule. *Pigment Cell Melanoma Res*. 2020;33(5):767-772.
43. Goto K, Pissaloux D, Kauer F, Huriet V, Tirode F, de la Fouchardière A. GOPC-ROS1 mosaicism in agminated Spitz naevi: report of two cases. *Virchows Arch*. 2021;479:559-564.
44. Kervarrec T, Jacques BJ, Pissaloux D, Tirode F, de la Fouchardière A. FNBP1-BRAF fusion in a primary melanoma of the lung. *Pathology*. 2021;53:785-788.
45. Goto Keisuke, Pissaloux Daniel, Fraitag Sylvie, Amini Mona, Vaucher Richard, Tirode Franck, de la Fouchardière Arnaud. RASGRF1-rearranged cutaneous melanocytic neoplasms with spitzoid cytomorphology. *American Journal of Surgical Pathology*. 2021. https://doi.org/10.1097/pas.0000000000001839
46. Houlier A, Pissaloux D, Tirode F, et al. RASGRF2 gene fusions identified in a variety of melanocytic lesions with distinct morphological features. *Pigment Cell Melanoma Res*. 2021;34:1074-1083.
47. Houlier A, Pissaloux D, Masse I, et al. Melanocytic tumors with MAP3K8 fusions: report of 33 cases with morphological-genetic correlations. *Mod Pathol*. 2020;33(5):846-857.
48. Kervarrec T, Pissaloux D, Tirode F, et al. Morphologic features in a series of 352 Spitz melanocytic proliferations help predict their oncogenic drivers. *Virchows Arch*. 2021. http://doi.org/10.1007/s00428-021-03227-x
49. Goto K, Pissaloux D, Tirode F, de la Fouchardière A. Spitz nevus with a novel TFG-NTRK2 fusion: the first case report of NTRK2-rearranged Spitz/reed nevus. *J Cutan Pathol*. 2021;48:1193–1196.
50. de la Fouchardière A, Tee MK, Peternel S, et al. Fusion partners of NTRK3 affect subcellular localization of the fusion kinase and cytomorphology of melanocytes. *Mod Pathol*. 2021;34(4):735-747.
51. de la Fouchardiere A, Pissaloux D, Tirode F, Hanna J. Clear cell tumor with melanocytic differentiation and MITF-CREM translocation: a novel entity similar to clear cell sarcoma. *Virchows Arch*. 2021;479(4):841–846.
52. de la Fouchardiere A, Pissaloux D, Tirode F, Karanian M, Fletcher CDM, Hanna J. Clear cell tumor with melanocytic differentiation and Actin-MITF translocation: report of 7 cases of a novel entity. *Am J Surg Pathol*. 2020;45(7):962–968.
53. Churg A, Le Stang N, Dacic S, et al. Solid papillary mesothelial tumor. *Mod Pathol*. 2021;35:69-76.
54. Hedegaard J, Thorsen K, Lund MK, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS One*. 2014;9(5):e98187.
55. Alenda C, Rojas E, Valor LM. FFPE samples from cavitational ultrasonic surgical aspirates are suitable for RNA profiling of gliomas. *PLoS One*. 2021;16(7):e0255168.
56. Civita P, Franceschi S, Aretini P, et al. Laser capture microdissection and RNA-Seq analysis: high sensitivity approaches to explain histopathological heterogeneity in human glioblastoma FFPE archived tissues. *Front Oncol*. 2019;9:482.
57. Li P, Conley A, Zhang H, Kim HL. Whole-transcriptome profiling of formalin-fixed, paraffin-embedded renal cell carcinoma by RNA-seq. *BMC Genomics*. 2014;11(15):1087.
58. Ma Y, Ambannavar R, Stephans J, et al. Fusion transcript discovery in formalin-fixed paraffin-embedded human breast cancer tissues reveals a link to tumor progression. *PLoS One*. 2014;9(4):e94202.
59. Norton N, Sun Z, Asmann YW, et al. Gene expression, single nucleotide variant and fusion transcript discovery in archival material from breast tumors. *PLoS One*. 2013;8(11):e81925.
60. Pennock ND, Jindal S, Horton W, et al. RNA-seq from archival FFPE breast cancer samples: molecular pathway fidelity and novel discovery. *BMC Med Genomics*. 2019;12(1):195.
61. Plaska SW, Liu C-J, Lim JS, et al. Targeted RNAseq of formalin-fixed paraffin-embedded tissue to differentiate among benign and malignant adrenal cortical tumors. *Horm Metab Res*. 2020;52(8):607-613.
62. Sinicropi D, Qu K, Collin F, et al. Whole transcriptome RNA-Seq analysis of breast cancer recurrence risk using formalin-fixed paraffin-embedded tumor tissue. *PLoS One*. 2012;7(7):e40092.
63. Zhao Y, Mehta M, Walton A, et al. Robustness of RNA sequencing on older formalin-fixed paraffin-embedded tissue from high-grade ovarian serous adenocarcinomas. *PLoS One*. 2019;14(5):e0216050.
64. Vukmirovic M, Herazo-Maya JD, Blackmon J, et al. Identification and validation of differentially expressed transcripts by RNA-sequencing of formalin-fixed, paraffin-embedded (FFPE) lung tissue from patients with idiopathic pulmonary fibrosis. *BMC Pulm Med*. 2017;17(1):15.
65. Hayette S, Grange B, Vallee M, et al. Performances of targeted RNA sequencing for the analysis of fusion transcripts, gene mutation, and expression in hematological malignancies. *Hema*. 2021;5(2):e522.
66. Watson S, Perrin V, Guillemot D, et al. Transcriptomic definition of molecular subgroups of small round cell sarcomas. *J Pathol*. 2018;245(1):29-40.
67. Wiesner T, Lee W, Obenauf AC, et al. Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature*. 2015;526(7573):453-457.
68. Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol*. 2002;161(6):1961-1971.
69. Newton Y, Sedgewick AJ, Cisneros L, et al. Large scale, robust, and accurate whole transcriptome profiling from clinical formalin-fixed paraffin-embedded samples. *Sci Rep*. 2020;10(1):17597.
70. Jovanović B, Sheng Q, Seitz RS, et al. Comparison of triple-negative breast cancer molecular subtyping using RNA from matched fresh-frozen versus formalin-fixed paraffin-embedded tissue. *BMC Cancer*. 2017;17(1):241.
71. Choi Y, Kim A, Kim J, Lee J, Lee SY, Kim C. Optimization of RNA extraction from formalin-fixed paraffin-embedded blocks for targeted next-generation sequencing. *J Breast Cancer*. 2017;20(4):393-399.
72. Wehmas LC, Wood CE, Chorley BN, Yauk CL, Nelson GM, Hester SD. Enhanced quality metrics for assessing RNA derived from archival formalin-fixed paraffin-embedded tissue samples. *Toxicol Sci*. 2019;170(2):357-373.
73. Marín de Evsikova C, Raplee ID, Lockhart J, Jaimes G, Evsikov AV. The transcriptomic toolbox: resources for interpreting large gene expression data within a precision medicine context for metabolic disease atherosclerosis. *J Pers Med*. 2019;9:E21.
74. Kwong LN, De Macedo MP, Haydu L, et al. Biological validation of RNA sequencing data from formalin-fixed paraffin-embedded primary melanomas. JCO precis. *Oncologia*. 2018;2018:1-19.

75. Bossel Ben-Moshe N, Gilad S, Perry G, et al. mRNA-seq whole transcriptome profiling of fresh frozen versus archived fixed tissues. *BMC Genomics*. 2018;19(1):419.
76. Auerbach SS, Phadke DP, Mav D, et al. RNA-Seq-based toxicogenomic assessment of fresh frozen and formalin-fixed tissues yields similar mechanistic insights. *J Appl Toxicol*. 2015;35(7):766-780.
77. Penland SK, Keku TO, Torrice C, et al. RNA expression analysis of formalin-fixed paraffin-embedded tumors. *Lab Invest*. 2007;87(4):383-391.
78. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (a) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics*. 2014;2(15):419.
79. Morton ML, Bai X, Merry CR, et al. Identification of mRNAs and lincRNAs associated with lung cancer progression using next-generation RNA sequencing from laser micro-dissected archival FFPE tissue specimens. *Lung Cancer*. 2014;85(1):31-39.
80. Graw S, Meier R, Minn K, et al. Robust gene expression and mutation analyses of RNA-sequencing of formalin-fixed diagnostic tumor samples. *Sci Rep*. 2015;23(5):12335.
81. Peng J, Feng Y, Rinaldi G, et al. Profiling miRNAs in nasopharyngeal carcinoma FFPE tissue by microarray and next generation sequencing. *Genom Data*. 2014;2:285-289.
82. Erdem-Eraslan L, van den Bent MJ, Hoogstrate Y, et al. Identification of patients with recurrent glioblastoma who may benefit from combined bevacizumab and CCNU therapy: a report from the BELOB trial. *Cancer Res*. 2016;76(3):525-534.
83. Turnbull AK, Selli C, Martinez-Perez C, et al. Unlocking the transcriptomic potential of formalin-fixed paraffin embedded clinical tissues: comparison of gene expression profiling approaches. *BMC Bioinformatics*. 2020;21(1):30.
84. Trejo CL, Babić M, Imler E, et al. Extraction-free whole transcriptome gene expression analysis of FFPE sections and histology-directed subareas of tissue. *PLoS One*. 2019;14(2):e0212031.
85. Koelsche C, Stichel D, Griewank KG, et al. Genome-wide methylation profiling and copy number analysis in atypical fibroxanthomas and pleomorphic dermal sarcomas indicate a similar molecular phenotype. *Clin Sarcoma Res*. 2019;9:2.
86. Koelsche C, Hartmann W, Schrimpf D, et al. Array-based DNA-methylation profiling in sarcomas with small blue round cell histology provides valuable diagnostic information. *Mod Pathol*. 2018;31(8):1246-1256.
87. Koelsche C, Kriegsmann M, Kommoss FKF, et al. DNA methylation profiling distinguishes Ewing-like sarcoma with EWSR1-NFATc2 fusion from Ewing sarcoma. *J Cancer Res Clin Oncol*. 2019;145(5):1273-1281.
88. Lyskjaer I, De Noon S, Tirabosco R, et al. DNAmethylation-based profiling of bone and soft tissue tumours: a validation study of the 'DKFZSarcoma classifier'. *J Pathol Clin Res*. 2021;7(4):350-360.
89. Vickovic S, Eraslan G, Salmén F, et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods*. 2019;16:987-990.
90. Ståhl PL, Salmén F, Vickovic S, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*. 2016;353(6294):78-82.
91. Eng C-HL, Lawson M, Zhu Q, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature*. 2019;568(7751):235-239.
92. Codeluppi S, Borm LE, Zeisel A, et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat Methods*. 2018;15:932-935.
93. Biancalani T, Scalia G, Buffoni L, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nat Methods*. 2021;18:1352-1362.
94. Wang X, Allen WE, Wright MA, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*. 2018;361(6400):eaat5691.
95. Alon S, Goodwin DR, Sinha A, et al. Expansion sequencing: spatially precise in situ transcriptomics in intact biological systems. *Science*. 2021;371(6528):eaax2656.
96. Em S, Furtwängler B, Üresin N, et al. Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nat Commun*. 2021;12(1):3341-3355.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.