Molecular Pathological Analysis of Sarcomas Using Paraffin-Embedded Tissue: Current Limitations and Future Possibilities

Matt van de Rijn¹, Xiangqian Guo¹, Robert T Sweeney¹, Andrew H Beck² and Robert B West³

¹Department of Pathology, Stanford University Medical Center, Stanford, CA, USA
²Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA

Abstract
Sarcomas of soft tissue and bone are rare neoplasms that can be separated into a large number of different diagnostic entities. Over the past decades, a number of diagnostic markers have been developed that aid pathologists in reaching the appropriate diagnoses. Many of these markers are sarcoma-specific proteins that can be detected by immunohistochemistry in formalin-fixed, paraffin-embedded (FFPE) sections. In addition, a wide range of molecular studies have been developed that can detect gene mutations, gene amplifications or chromosomal translocations in FFPE material. Until recently most sequencing based approaches relied on the availability of fresh-frozen tissue. However, with the advent of next generation sequencing technologies FFPE material is increasingly used as a tool to identify novel immunohistochemistry markers, gene mutations and chromosomal translocations and to develop diagnostic tests.

Introduction
Sarcomas are malignant tumors that develop from connective tissues such as muscle, fat, bone and cartilage. A wide range of distinct sarcomas can be recognized and many

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/his.12290
This article is protected by copyright. All rights reserved.
differ from each other in their clinical behavior. The malignant tumors also need to be distinguished from the many benign lesions. Furthermore, sarcomas are rare with an estimated number of 11,000 new cases per year for soft tissue sarcoma and approximately 3,000 cases for bone sarcomas in the United States (http://www.seer.cancer.gov/). Due to the wide range of diagnostic entities and the rarity of the tumors, most pathologists do not have extensive experience in recognizing and classifying these lesions. Fortunately, the diagnosis of many tumor types is facilitated by the use of biomarkers. In most surgical pathology laboratories the markers used consist of monoclonal antibodies that are able to detect tumor-specific antigens in formalin-fixed, paraffin-embedded (FFPE) tissue. More specialized laboratories use a range of molecular tests to aid the diagnostic surgical pathologist. These tests can detect tumor-specific chromosomal translocations by fluorescence in situ hybridization (FISH) or RT-PCR. Other tests identify gene-specific mutations, such as for the KIT gene in gastrointestinal stromal tumors (GIST). The number of tumor-specific chromosomal translocations has dramatically increased over the past fifteen years and can only be expected to increase further in the next years. Recently Neuville et al. performed a large population based prospective study on the use of molecular testing in 1484 sarcomas. The sarcomas included GIST, well differentiated liposarcoma and 7 sarcomas associated with unique chromosomal translocations. Prior to molecular testing the diagnoses were classified as certain, probable or possible. It was shown that in a variable proportion of the cases (depending on the type of sarcoma) the molecular testing contributed significantly by either confirming a probable diagnosis or by providing a diagnosis where conventional analysis was inconclusive.

It is not cost effective and also labor intensive to maintain a wide range of individual CLIA-approved diagnostic tests for the many mutations and translocations that occur in the various sarcoma subtypes. Application of next generation sequencing technology to this field opens possibilities for a more efficient and ultimately more cost effective way to detect these genetic abnormalities. In the past, conventional technologies for genomic studies relied on the availability of fresh frozen tissue but next generation sequencing can be applied to FFPE material. Therefore, our understanding of the genomic landscape of sarcomas, will greatly improve with these new methods as we can now study the wide range of entities and subtypes widely available as FFPE tissue. Conventional DNA sequencing required high quality long length DNA as the goal was to sequence segments of DNA or cDNA that were as long as possible using relatively few sequencing reactions. The initial sequencing of the human genome was started in 1990, a rough draft was reported in 2000, the project was considered completed in 2003. The cost was estimated at more than three billion U.S. dollars. Next generation sequencing technology represents a dramatic shift in the approach to conventional DNA sequencing. Rather than performing a few reactions on long fragments of DNA, millions of sequencing reactions are performed simultaneously (“massively parallel”) on much shorter stretches of DNA. Powerful computer algorithms are then used to piece together the genome or the transcriptome. This next generation sequencing approach coupled with a dramatic decrease in the cost of these studies has led to many discoveries in cancer biology.
The DNA or RNA isolated from FFPE tissue is fragmented because formalin fixation and room temperature storage without special protection induces modifications and fragmentation of nucleic acids such that only relatively short fragments of DNA and RNA strands are left that cannot be used for conventional sequencing. However, these short fragments are eminently suitable for analysis by next generation sequencing. Thus, the maximization of data of next generation sequencing rests on the number of simultaneous sequencing reactions that occur on unique DNA molecules (commonly referred to as “depth”), rather than the length of the DNA (or quality of the DNA) for Sanger DNA sequencing. The effect of formalin fixation on the nucleic acids in paraffin blocks is thought to be an ongoing process with the result that older paraffin blocks usually yield progressively more degraded DNA and RNA. Fortunately, for clinical diagnostic purposes, most specimens will be analyzed within 6 months after paraffin embedding. A recent study by Spencer et al. showed that measurable differences do exist between data obtained through next generation sequencing from frozen tissue samples versus FFPE material. However, the differences found between FFPE material and matched frozen samples were minor and did not affect clinical interpretation. In the same study the investigators showed that variations in the length of formalin fixation also did not significantly affect sequencing results. To overcome age degradation, deeper sequencing can be applied to increase the abundance (although not the percentage) of high quality, usable reads. The protocols for extraction of nucleic acids from paraffin have improved dramatically and the chemical modification of nucleic acids of FFPE processing does not appear to interfere with next generation sequencing reactions. While the fragmentation of nucleic acids, that occurs as a result of formalin fixation, is actually a necessary step in next generation sequencing library preparation, one major limitation in standard processing, decalcification, is worth mentioning. Tissues that are exposed to decalcification are no longer amenable to sequencing while other studies such as FISH and even immunohistochemistry are also compromised; it is therefore important to save some tumor from exposure to decalcification solutions.

Molecular Aspects of Sarcomas

On a molecular basis, soft tissue sarcomas can be subdivided into two groups. In the first group a large number of chromosomal abnormalities co-exist simultaneously; these can include multiple chromosomal translocations, gene mutations, amplifications and deletions. The tumors in this group often display a highly aggressive histologic phenotype with significant nuclear pleomorphism and a high mitotic rate. To this group of tumors belong sarcomas such as undifferentiated pleomorphic sarcoma, pleomorphic liposarcoma, leiomyosarcoma, and others. The second group of sarcomas is characterized by having a pathognomonic genetic alteration in which three types can be discerned. The first are gene point mutations such as those that occur in the KIT or PDGFRA genes in GIST. Another genomic abnormality is that of gene amplification and an example of this is the amplification of the MDM2 gene that is seen in atypical lipomatous tumor/well differentiated liposarcoma and the de-differentiated form of this tumor. Last but certainly not least are the numerous gene translocations that have been identified in soft tissue sarcomas over the past decades. In a 2011 review by Demicco and Lazar, twenty-one soft tissue and bone tumors with diagnostic chromosomal
translocations were reported \(^4\). This represented a dramatic increase in the number of sarcomas with associated translocations since in a review in 1999 only nine sarcomas with chromosomal translocations were discussed \(^5\). Since the publication of the review by Demicco and Lazar in 2011 several additional chromosomal translocations have been reported such as the t(10;17) that characterizes the novel subset of high grade endometrial stromal sarcomas \(^6,7\), the t(5;8) that is found in angiofibroma \(^8\) and the gene fusion between NAB2 and STAT6 on chromosome 12 in solitary fibrous tumors \(^9\). Interestingly, chromosomal translocations have now been reported not only in malignant tumors (sarcomas) but also in benign soft tissue lesions and bone lesions. Some of these were not clearly understood to be even neoplastic. Examples are the t(17;22) found in nodular fasciitis \(^10\), the t(1;2) in the lesion formerly known as pigmented villonodular synovitis and now termed tenosynovial giant cell tumor \(^11,12\), the t(16;17) that is detected in aneurysmal bone cysts \(^13,14\) and the complex chromosomal rearrangements recently demonstrated in uterine leiomyomas \(^15\).

With the increase in the number of tumor specific chromosomal translocations there has also been growing complexity of these genetic lesions as reviewed by Ordoñez \(^16\). First, some sarcoma diagnostic entities can have more than one chromosomal translocation type. The most dramatic example of this is Ewings sarcoma where a large number of genes can be fused to EWSR1 (or less commonly to FUS and others genes) to result in the EWS phenotype. Second, a single gene can be involved in different translocations that are distinctive for multiple tumors. The EWSR1 gene can be found in at least ten different tumors consisting of nine soft tissue sarcomas and one leukemia as it is joined with a range of other genes. Finally, identical chromosomal translocations can be found in diverse tumor types. A recent example of this is the finding that YWHAE-FAM22 previously reported in clear cell sarcoma of the kidney also is present in the recently described high grade endometrial stromal sarcoma variant \(^6\).

**Next generation sequencing applied to RNA derived from FFPE material**

A range of molecular techniques that were initially developed on fresh frozen tissue or cell lines have been applied with variable success to DNA and/or RNA derived from FFPE tissue. Initially genome-wide approaches relied on gene microarrays to interrogate either gene expression levels or gene copy number changes. Gene expression profiling by gene microarrays showed moderate success when using cDNA derived from RNA isolated from FFPE material. More success was obtained by (RT-)PCR on FFPE derived material. An example of a commercial application of this approach can be found in the Oncotype test that prognosticates breast carcinoma based on the expression level of a number of genes determined by RT-PCR from paraffin material. Array-based comparative genomic hybridization to determine gene amplification or deletion was first described by Pollack et al. \(^17\) on frozen tissue and subsequently on paraffin material \(^18\).

Recent applications of genome-wide next generation sequencing approaches to FFPE derived material can be subdivided in several categories. The genome-wide approach should be distinguished from those techniques where a pre-defined number of genes can be analyzed. In contrast to these targeted applications where a limited number of

This article is protected by copyright. All rights reserved.
genes are interrogated, the genome-wide technologies look at essentially all DNA (whole genome sequencing) or mRNA in the specimens submitted and thus have a broader potential for discovery. In 2010 we developed “3SEQ”, a method to perform genome wide expression profiling of mRNAs derived from FFPE tissue. In this technique 3’ fragments of mRNA are isolated through hybridization of the mRNA’s poly-A tail to oligo-dT coated electromagnetic beads. Next, these fragments from the 3’ ends of mRNAs are parallel sequenced, generating short reads between 36 and 100 base pairs long. The identity of the gene from which they are derived is determined by mapping the sequencing reads to the human genome. The number of occurrences of fragments for a certain gene can then be measured, giving an accurate assessment of the level of transcription of that gene in the specimen examined. This approach does not allow sequencing of the entire length of the mRNA molecules as one would want to do for mutation analysis but the technology nevertheless allows for a good quantification of the number of mRNA molecules for each gene. In 2012 Lee et al. used next generation sequencing on frozen tissue to identify a new chromosomal translocation defining a novel subset of endometrial stromal sarcomas. 3SEQ gene expression profiling on FFPE-derived mRNA was used to confirm the uniformity of gene expression of this new subtype of tumor and led to identification of a diagnostic marker for this new entity.

The ability to perform 3SEQ gene expression profiling on FFPE tissue markedly increases the number of specimens and the kind of specimens available for research. No longer dependent on what is available as frozen material one can for example select only those cases for which clinical follow-up data is available, or analyze FFPE specimens that belong to a clinical trial. In addition since the mRNA is extracted from cores taken from FFPE blocks, very small lesions can be analyzed. An example of this application can be found in the study of rare fibroblastic lesions that are usually removed through small excisional biopsies and often entirely submitted for histologic analysis. For these lesions, that include giant cell tumor of tendon sheath and others, essentially no frozen tissue is available (Figure 1). Studies such as these can be used to identify novel diagnostic markers for these fibroblastic lesions and other tumor types but can also be used to define gene signatures for distinct subsets of fibroblastic tumor stroma that play a role in breast carcinoma behavior.

In addition to studying levels of mRNA expression for conventional genes, i.e. genes that encode for cellular proteins, the 3SEQ approach can also be used to study a relatively novel class of RNA molecules called long non-coding RNA (lncRNA). The function of lncRNAs is not yet fully defined but a major role of these molecules appears to be their ability to regulate gene transcription through their interaction with nuclear proteins. We recently reported the genome-wide expression profile of known lncRNAs and a significant number of newly discovered candidate lncRNAs on a diverse group of tumors representing 17 different diagnostic classes from 30 carcinomas and 36 sarcomas. The clinical relevance of this study remains to be determined but as lncRNAs appear to be relatively specific for different cell differentiation states they have potential as differential diagnostic markers to distinguish soft tissue tumors derived from different cell types.

As mentioned above, the past decades have seen a dramatic increase not only in the number of chromosomal translocations that have been discovered to be diagnostic for a
range of tumors but also in the complexity of these translocations. It is a major challenge for any molecular pathology laboratory to maintain CLIA-approved individual diagnostic tests for each of these soft tissue and bone sarcomas. This, coupled with the rarity of the disease, results in a very unsatisfactory cost-benefit ratio. It is probably for this reason that there are very few laboratories that offer more than the most common FISH and RT-PCR studies for diagnostic purposes. In addition most laboratories use simple split-apart FISH tests that will not identify all possible translocations that can occur in a particular sarcoma and that may also lack specificity if the gene tested is involved in different translocation in different tumor types. While significant degradation does occur in FFPE material, this happens in a time-dependent fashion and cases that have only recently been embedded in paraffin can still yield relatively long fragments of mRNA. Given this, we hypothesized that we could detect fusions in FFPE RNA through high throughput sequencing. The cost of the original Illumina sequencers put this approach out of reach of most clinical laboratories. However, recently a smaller version of the Illumina sequencer became available. MiSeq is a much smaller, desktop sized, instrument that is relative easy to operate. Its lower cost puts the instrument within reach for many molecular pathology laboratories. We performed next generation sequencing on mRNA derived from FFPE tissue with a protocol that allows for long surviving fragments of RNA in the FFPE material to be sequenced. cDNA fragments underwent paired-end sequencing for 150 cycles, such that each end of the fragments results in a 150 nucleotide sequence. High quality 50-mer sequences were identified in each fragment and these were mapped to a custom-made library (the “sarcomatome”) that incorporates all 83 genes known to be involved in chromosomal translocations and mutations that occur in human sarcomas. Fragments where the two 50-mer sequences mapped to different genes in the sarcomatome were identified and within this group fragments that spanned the actual breakpoint for each gene could be found (Figure 2) as a further confirmation of the translocation. The “sarcomatome” library against which the sequences obtained from the samples is matched can be easily modified to accommodate any novel translocations that may be discovered. With this approach we could identify translocations in synovial sarcomas (3/3 cases), myxoid liposarcoma (3/3 cases) and clear cell sarcoma (1/1 case). In 2 cases of Ewings sarcoma tested, the translocation could not be identified on the MiSeq instrument but was detectable using deeper sequencing on the original Illumina platform. Clearly more work needs to be done to see to what extent this approach can be used in a clinical setting on a much larger group of translocations and also to what extent this approach will compare favorably or not to other approaches that use targeted high throughput sequencing.

DNA analysis from FFPE
Methods used for RNA can equally apply to DNA. The lack of the 2' hydroxyl group means that DNA is considerably more stable than RNA; nevertheless, it does undergo extensive cleavage in archival conditions. To date several studies have been published on the use of genome-wide high throughput sequencing using archival DNA. Recently we reported the results of whole genomic DNA sequencing based on FFPE tissue. In this study of breast neoplasia progression to breast cancer, paired-end libraries were built and sequenced on the Illumina HiSeq platform with slight variations from conventional protocols. Analysis of the mapped reads demonstrated excellent library
quality that was indistinguishable from that of comparable libraries constructed from fresh DNA. This study demonstrated the advantages of working with FFPE material in that small samples (including ductal hyperplasias of 1-3 mm in size) and difficult to obtain samples (patient matched normal breast, hyperplasia, carcinoma in situ, and invasive carcinoma) could be analyzed. This study and those of others\(^\text{28-30}\) show that whole genome sequencing can be performed on FFPE tissue and it is expected that these studies will soon be extended to the study of sarcomas.

**Conclusion**

Traditional histology remains the most effective and cost-efficient technique to discriminate between the many different benign and malignant tumors that originate in the soft tissues and bone. The diagnosis of these lesions is however greatly helped by a range of immunohistochemical markers and, increasingly, by diagnostic test that detect gene abnormalities. The advent of next generation sequencing technology can be expected to lead to an expansion of the diagnostic tools currently available.

**Legends to Figures**

**Figure 1**

Heat map of differentially expressed genes in small fibroblastic tumors (DTF: desmoid type fibromatosis, PF: palmar fibromatosis, NF: nodular fasciitis, SFT: solitary fibrous tumor, DFSP: dermatofibrosarcoma protuberans, NPAF: nasopharyngeal angiofibroma) obtained through 3SEQ-based gene expression profiling from FFPE material.

**Figure 2**

Schematic representation of the use of desktop transcriptome sequencing to determine the presence of chromosomal translocations in FFPE material (figure modified from ref.\(^\text{26}\)).


This article is protected by copyright. All rights reserved.