

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Practical Laboratory Medicine

journal homepage: www.elsevier.com/locate/plabm

Adopting solutions for annotation and reporting of next generation sequencing in clinical practice

Jinming Song, Mohammad Hussaini^{*}

Department of Hematopathology and Lab Medicine, Moffitt Cancer Center, Tampa, FL, 33612, USA

ABSTRACT

With advancements in the understanding of human cancers (carcinomas, sarcomas, and hematopoietic malignancies), molecular profiling, especially by Next Generation Sequencing (NGS), is playing an increasingly important role in the diagnosis, prognostication, and therapeutic management of cancer patients. The final and critical step in NGS is the annotation of detected variants and reporting of their clinical significance. Automated bioinformatics tools are available to assist with annotation, but the final responsibility for interpretation and validation of the annotation rests with the pathologist who may be constrained by the pressures of clinical sign-out and limited training in NGS. In this manuscript, we detail our experience in outsourcing variant annotation to a high-quality vendor to improve quality, standardize reporting, and decrease turn-around time of NGS reporting in clinical practice. We describe the composition of the evaluation team, steps that should be taken to evaluate potential annotation vendors, and detailed parameters that should be addressed before contracting with a vendor to guarantee the clinical reliability of the reported annotations.

1. Introduction

Human cancers, including carcinoma, sarcoma, and hematopoietic malignancies, have been traditionally diagnosed using morphology, immunohistochemistry, and cytogenetics studies. However, with the increasing availability of genomic profiling by Next Generation Sequencing (NGS), mutational status has become important element in the diagnosis, prognostic stratification, selection of targeted therapy, and clinical trial enrollment for many cancer types [1–5]. This is especially true in the case of hematopoietic malignancies [6–11]. For example, the detection of *SF3B1* mutation and *JAK2* mutations have diagnostic utility in the setting of myelodysplastic syndrome or myeloproliferative neoplasms (such as essential thrombocythaemia, polycythaemia vera, or primary myelofibrosis), respectively [12–14]. The detection of *TP53* mutations is associated with poor prognosis in most carcinomas, sarcomas, and hematopoietic malignancies [15]. With advances in NGS technology, large numbers of genes can now be tested efficiently in a single experiment making clinical mutational profiling feasible to facilitate cancer diagnosis and treatment decisions.

NGS typically involves the following steps: collection of tumor specimens, extraction of DNA from the tumor cells, amplification of the DNA, primer hybridization or target amplification to create library, DNA sequencing, comparison with reference human genome to detect variants, and finally annotation of the detected variants to identify clinically significant and relevant variants. Many of the steps in processing NGS samples and resultant data have been automated using bioinformatics tools, but the final step of annotation still requires manual curation. Multiple systems and resources are available to automatically pull available information from variable databases or sources to assist in annotation [16,17]. However, manual review, interpretation, and validation is still necessary, particularly given that the classifications of some variants are not consistent across available databases [18] such as ClinVar [19], CIViC [20], COSMIC [21] and dbSNP [22,23]. Final reported variant annotations ideally require input from trained scientists and/or clinicians to help distinguish between clinically irrelevant variants, variants of unknown significance, and clinically important pathogenic variant. Care must be taken

^{*} Corresponding author.

E-mail address: mohammad.hussaini@moffitt.org (M. Hussaini).

<https://doi.org/10.1016/j.plabm.2020.e00154>

Received 29 July 2019; Received in revised form 8 January 2020; Accepted 31 January 2020

2352-5517/© 2020 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

to ensure that variants reported as clinically pathogenic are not be germline mutations or benign single nucleotide polymorphisms (SNPs) that are present in normal, healthy cells.

NGS testing is performed in a variety of settings ranging from clinical molecular labs of pathology departments at academic centers to private for-profit pathology reference laboratories. NGS cases are signed out by pathologists (with or without the assistance of residents and fellows), clinical genomicists, and PhD level scientists. Many traditional pathologists are heavily trained in the assessments of tumor morphology, immunohistochemistry, and cytogenetics (karyotyping or fluorescence in-situ hybridization FISH) studies. However, those trained prior to the advent of NGS may lack a deep understanding of NGS technology. Consequently, they may have limited knowledge and/or experience in the annotation of detected variants. Furthermore, the heavy load of daily clinical surgical pathology service also allows only limited amount of time to research and draft interpretations of detected variants. Some pathologists undergo a fellowship training in molecular genetic pathology to gain a better understanding of NGS techniques as well as annotation. However, even for trained molecular pathologists, accurate and detailed annotation of the variants is still time-consuming and challenging task.

Another key challenge is the need for standardized variant annotation. Due to differences in knowledge-matter expertise and in the understanding of the association between a variant and disease, discrepancies may exist between the annotations of different pathologists for the same variant. This problem is compounded by the presence of conflicting reports in the literature regarding the clinical significance of some variants. For yet other variants, there are rare or no reports in the scientific cannon. Depending on which studies a pathologist refers to and his or her subjective assessment of the credibility of those studies, assessment of the clinical significance of a variant could vary significantly between pathologists.

The assignment of clinical significance of a variant in a disease also requires an understanding of the pathogenesis and context of the disease. For example, the clinical interpretation of *JAK2* mutations in myeloproliferative neoplasms is significantly different from the interpretation of the same variant in myelodysplastic syndromes or acute myeloid leukemia. Different variants in the same gene also sometimes carry different clinical significance. Thus, variant-specific annotation is of key importance. Lack of appreciation for this point and lack of familiarity with functional assessment of DNA mutations, may result in some pathologists reporting a variant as pathogenic while others not.

Finally, the rapidly changing nature of science poses a formidable challenge. There are scores of new cancer research articles published each month, and it is often not feasible for a busy pathologist to keep abreast with every new publication in the literature. This may result in interpretation of variants based on outdated knowledge. Since many annotators use a local knowledge base repository to populate annotations of previously identified variants, this serves as a timely reminder that these knowledge bases should be regularly updated.

Thus, there are various factors that may lead to the lack of uniformity in interpreting and reporting the significance of variants detected by NGS. In turn, this can be a source of significant frustration and confusion for the treating clinician. This is particularly true for community oncologists whose knowledge about gene mutations may be more limited. Therefore, it is critical to standardize the interpretation and the reporting of NGS studies. This can be done by implementing internal training, frequent consensus meetings, and quality control peer review of reports between the pathologists at the same institute. These are methodologies that we have implemented at our institution. However, in our experience, these processes do not obviate all attendant challenges. For example, it is difficult to bring every pathologist to the same level of knowledge-matter expertise or to achieve complete agreement regarding annotations due to differences in personal experience or habit, heavy clinical service load, and other responsibilities. There is also a challenge of pathologist turn-over which requires education and calibration of new staff members to institute-specific consensus interpretations.

Thus, although there was significant experience (with several MGP-boarded faculty) as a whole in our clinical genomics group, based on various factors we concluded it may be a better approach to outsource NGS annotation to a centralized institute. There were several potential advantages to this decision. First, it could be cost effective, particularly with regards to complex and/or highly mutated cases, which could consume hours of pathologist time with little reimbursement for the professional component. Second, it would help with standardization of annotations and NGS reporting. Third, it would allow us to keep and control the raw data for research and clinical care purposes. Fourth, we could use the annotation service on an “as-needed” basis. Fifth, 3rd party vendors often employ a group of annotation experts with comprehensive knowledge and experience in NGS, relevance in disease, and application in clinical practice. They also have automated algorithms that assist in timely population and generation of annotations. Moreover, third-party vendors often have large pre-existing databases that allow for more rapid searching of germline and somatic mutation databases and literature, facilitating pre-population of reports, leveraging of large knowledgebases developed either internally or by virtue of client samples over years, and decreasing turn-around time. They may also have integrated programmatic mechanisms to facilitate clinical trial matching.

In addition, as discussed above, the landscape of genomics in neoplastic disease is rapidly changing with 80 or more new publications in the field every week. This prompted us to consider the need to include artificial intelligence-driven or other automated solutions to facilitate annotations and to potentially address the issue of re-annotation of historical reports based on the emergence of new science, technologies, trials, and therapies. All these could be better carried out in a standardized fashion by a specialized and expert NGS annotation provider.

Accordingly, in this manuscript, we described our experience in selecting an NGS vendor and propose recommendations for the objective evaluation, identification, and adoption of a high-quality 3rd party NGS annotation provider.

2. Evaluation of the potential vendors

To evaluate a list of potential vendors, an evaluation team has to be assembled at the home institution. Then, organized step-by-step meetings and discussions should be arranged with each of the vendors to facilitate an efficient and productive evaluation. Many details and parameters need to be evaluated to guarantee the quality and the clinical reliability of the annotations from the vendors.

2.1. Composition of the vendor evaluation team

To ensure high quality reporting to meet clinical needs, we developed a multi-departmental collaborative taskforce, chaired by pathology, to come up with a solution to satisfy all stake holder requirements at our institute. The task force included representation from the following sectors:

- Pathologist with separate representation from hematopathology and anatomic pathology, given the different genomic needs of each program.
- Hematology/Oncology clinical faculty with interest and background in next generation sequencing in order to meet the practical needs of clinical practice.
- Bioinformatics core, to ensure that the clinical solution that would be brought in would be compatible with the data warehousing models at our institute.
- Information technology team, to ensure that solutions are compatible and can be integrated into the existing electronic medical record systems in place at our institute.
- Research support group, to ensure that our solution would be retrievable and be in a format that could potentially facilitate our research mission.

2.2. Steps of evaluation

Followings are the steps that could be taken for the evaluation of potential vendors.

- Identification of all potential vendors that provide annotation solutions for NGS. At the time of our initial search, the vendors that provided annotation services included but were not limited to the following: N-of-One, GenomOncology, IBM Watson Genomics, Tute Genome Oncology, Foundation One Cheetah, Jackson Laboratory, CollabRx, Qiagen Ingenuity, PierianDx.
- Preliminary meetings to understand the functionality, architecture, and capabilities of each vendor. These include researching vendor background, composition of personnel, their experience and expertise, their databases, bioinformatics pipelines, front-end interfaces, and customer support.
- Development of an internal team and algorithm to fairly and objectively evaluate between the various vendors to see if they are able to satisfy the variegated needs of the different stake holders involved. These include evaluation of the turnaround time of the annotation, quality of the annotation, contents of the annotation, if they meet internal clinical practice requirements, and reporting format. An ideal vendor should have a platform that can be easily integrated with the internal workflow and the existing information systems.
- Clearing of intellectual property and evaluation agreements by internal legal department to avoid any potential legal or intellectual property issues.

Table 1

Comparison amongst finalist vendors across selected key parameters.

Vendor	Cost Structure	Turnaround Time	Qualification	Work flow	Clinical Review by MD?	Frequency of Update	Experience
A	\$100 per case 54 gene Myeloid Panel for 3–6 variants average. \$75 per Solid Tumor case \$25 reflex per case [if the variant is not found in their database it is reflexed for further annotation] \$5000 implementation	1 business day for the 15 Gene Solid Tumor Panel 2 business days for the 26 Gene Solid Tumor Panel and the 54 Gene Myeloid Pane 3 business days for re-interpretation of reports from other diagnostic companies. Rapid Insights- Minutes	PhD/MD	Manual with backend automation	Yes	3 months	7 years 18 K cases 22 K annotations over 700 genes and 400 cancer subtypes
B	\$65/case but could be less or more depending on complexity	2–20 min	PhD/MD	Manual with backend automation	No	Continuously	18 years 340 K genomes, 30 K publications 10 million clinical findings
C	\$50-\$200 per report depending on complexity.	<24 h	PhD	Manual with backend automation	No	6 months, aim for 3 months	2 years 100s of cases
D	\$75/report for up to 6 variants	72 h	graduate (PhD/MS)/post-doctoral trained experts	Manual with backend automation	No	Weekly	5 years 15,000 cases

5. Testing of vendor annotation qualities. This step includes preparation of previous challenging cases, both myeloid and solid tumors, to test the vendor's ability to annotate not only variants in the coding regions, but also splice site variants, known artifacts, or benign SNPs etc. The raw data, typically in VCF format, are transferred to the vendor for annotation as test cases.
6. Comparison of the annotation results from each vendor with each other and with internal annotation data to see which vendor has the best quality. An evaluation team, including pathologists, clinicians, and molecular scientists would discuss and come to a consensus about the quality and the ranking of the vendors. For a sample comparison of key metrics, please refer to [Table 1](#) below.
7. Presentation of the consensus to the molecular pathology executive committee for final approval to choose and sign the contract with the proposed best vendor.
8. Finalize legal terms and sign the contract with the selected vendor.

2.3. Parameters to evaluate

The standardized parameters that we used to evaluate each of the vendors include the following:

1. Qualifications of the personnel doing annotations

An ideal annotation vendor should have adequately trained and experienced personnel to evaluate the biological and clinical data for clinical annotation. Commercially available solutions have variable staffing combinations that may include:

- a. Ph.D. level scientists with experience in next generation sequencing technology, research background (e.g. oncology pathways), and hands-on clinical training (e.g. knowledge of therapeutic options and prognostic significance and knowledge of active clinical trials), to perform PhD level analysis.
- b. Pathologists and oncologists to validate or modify the annotations.
- c. Bioinformatics and information technologists with the capability to integrate guidelines from FDA, NCCN, ASCO, etc. automatically and ability to combine artificial intelligence analysis.

2. Are annotations automatically generated or manually curated or both?

Both automated and manual annotations approaches are necessary. Automated annotation, possibly incorporating artificial intelligence, is often the first and efficient step to pull in all the information that is available in different publicly available and federated databases. This job would be time consuming, labor intensive, and almost impossible for human beings considering the large number of databases and diversity of sources. However, automated annotation usually pulls in irrelevant or even incorrect information. Automated annotation is also likely to miss some relevant or important information. Therefore, automated annotation should always be reviewed manually by scientists and ideally by both pathologists and clinicians. The clinician review is important to ensure clinical relevance and appropriateness. However, software platforms that aid in the collection and categorization of information from different databases like COSMIC and cBioPortal greatly facilitate manual annotations.

3. What are your criteria for evidence to support an assertion of clinical actionability for a given variant? (Level of Evidence)

Some annotation vendors have their own levels of evidence, but all should reference FDA, ASCO, NCCN guidelines in final adjudication. For example, PierianDx has a five levels of evidence classification: level 1 for variants with direct clinical evidences for their involvement in the prognosis or FDA approved treatment of the patients' diseases, level 2 for variants with clinical evidence in other diseases but not the stated diagnosis, level 3 for variants reported in cancer, level 4 for variants of unknown significance, and level 5 for SNPs. We recommend, however, that CAP/ASCP and/or ACMG/AMP guidelines for annotation should be adopted to facilitate standardized practice in the precision medicine space.

4. What databases/resources are queried to formulate annotations?

While algorithms are proprietary, almost all vendors include interrogation of publicly available databases such as ExAC, NHLBI, ESP, dbSNP, COSMIC, TCGA, NCCN, FDA, ASCO etc. Some vendors allow crowdsourcing of annotations from other institutions that "opt in." In principle, the more databases or resources included, the more comprehensive the annotations will be. Missing information from some important databases will make the annotation incomplete and potentially misleading. For example, some clinically irrelevant variants might be reported as clinically important due to missed SNP referencing of germline databases. Some variants with poor prognosis might be reported as clinically neutral or good if a seminal study is omitted. However, this can be a two-edged sword. When more resources are used, this results in more information that needs to be sorted through and vetted. Too much information, especially irrelevant information in the annotations, will make it hard for clinicians to read through and challenging to capture the most important information for clinical practice. Sorting and cleaning algorithms are therefore necessary to handle the input and balance that with the information included in the annotations.

5. Is there a clinical review before release of annotation?

This is a key factor in determining the quality of the annotation service, because the genomic interpretation needs to make sense in the context of the patient and disease. Ideally, there should be a manual clinical review step before the annotations either going to client or initially getting populated in vendor knowledge databases. Some vendors have clinical review, while some rely on medical directors at the client facility to determine clinical implications. Either way, it is the responsibility of medical directors at the client facility to validate and proofread the annotations before reporting.

6. How are variants of unknown significance (VUS) handled?

VUS are variants for which clinical significance has not been clearly established. These could comprise a significant proportion of the variants reported. However, due to the rapid advancement and intensive work in medical and clinical research, the clinical significance of more and more VUS will be gradually deciphered. Before that, the annotations might have to resort to a general annotation for that gene (instead of a variant specific annotation), but this could be misleading given that not all the variants in a gene will have the same clinical significance. Therefore, efforts should be taken to reduce the number of VUS in a report. To this end, deep dives to look at known functional domains or similar mutations and creation of inferences may be useful. This is an important area for the clients because it is a time-consuming process and outside the expertise of many pathologists. Therefore, dedicated efforts from the vendors on this aspect are of great utility to clients.

7. Are annotations variant specific (as oppose to general about a gene)?

Annotations could be very generic at the gene level or be variant specific. Since not all the variants in a gene will have the same effects on the function of the gene and the same clinical significance, annotations should be variant specific. Almost all the vendors will promise to offer variant specific annotations; however, this needs to be confirmed.

8. Are variant allele frequencies reported?

Variant Allele frequency (VAF) is very important information that needs to be reported along with the variants. It sometimes helps to distinguish clinically significant variants from germline variants. The VAFs of the variants are also of interest to the clinicians in their assessment of driver mutations, reliability of the variants, clonal architecture, and for the follow up of disease progression or clonal evolution. Therefore, it is important to confirm that VAF will be routinely reported and easily accessible to the clinicians.

9. How are Single nucleotide polymorphism (SNPs) handled?

SNPs are variants observed in a portion of the general population and are usually deemed to be clinically insignificant. We used a cutoff of 1% allele frequency in the general population to define SNP and to filter them out for reporting. However, some SNPs have later been found to have some clinical significance and some consider them to be worth reporting. Different institutes might use different cutoffs, and the same institute might change their cutoff as their knowledge and preferences change. Therefore, it would be ideal for the vendors to offer user-definable cut-offs for SNP reporting to be flexible to the evolving needs in this space.

10. Does the report summarize major findings up front?

As mentioned previously, a simple and straightforward report is particularly important for the busy clinician. Up-front listing of important mutations with only brief annotations with detailed annotations provided on subsequent pages is one way to balance efficiency with comprehensiveness.

11. Does the report provide more detailed discussion if desired?

It is desirable to have the ability to ramp up the detail level of the annotations for some clinicians that are more at the cutting edge of clinical research and patient care. Sometimes these are even ahead of the advancements in the field of precision oncology and detailed information can provide them insights for patient management that may not initially be apparent to the annotating pathologist.

12. How standardized are the annotations, i.e. will the same mutation be reported the same way each time?

This is a key metric for the client and a key factor in quality control. The annotation for the same variant and same disease should be standardized. However, as more information becomes available, the annotations should be regularly updated and re-standardized. Different annotations for the same variant are confusing to clinicians and a source of annoyance.

13. How large is your existing annotation database?

Since it is a time-consuming process to accumulate high quality annotations, the bigger the existing annotation database the better. The size of the database can range from hundreds of variants to as large as 10 million annotations regarding biological, scientific and clinical findings. However, the annotations can differ significantly in terms of content and focus; thus, bigger is not always better.

Different institutes also have different patient populations and different disease focuses. It would be ideal if the existing focus of the annotations warehoused by the vendor match the disease areas of the client.

14. What is the frequency of updating the annotations and how are annotations kept up-to-date with the literature?

The annotation should be comprehensive at the time of report and regularly updated to reflect advancements in basic and clinical research as well as disease treatments. Frequency of updating of the database range from daily to weekly to every 120 days, depending on the vendor and on the commonness or rarity of the type of cancer of interest. Update of the annotations should use a combination of approaches, including structured and reproducible automated search and manual review. Automated annotation is an efficient way to pull in new information and new sources to update the databases. However, manual annotation is necessary to remove irrelevant or incorrect information. Updates beyond every 6 months would be deemed unfavorable.

15. Are FDA approved drugs listed?

The information about FDA approved drugs for a gene under annotation is very important for patient management and targeted therapy, and therefore, is mandatory for an annotation service.

16. Do annotations address interactions between multiple genes detected in any given case? How is this determined?

This was a key factor for us, particularly in myeloid malignancies, given that annotation of comutations is less well-defined in this space as opposed to solid tumors (e.g., *KRAS* in *EGFR*-mutated lung cancer). Since many patients carry mutations in more than one gene, annotations about co-mutated genes is important for personalized medicine, in which each patient is treated differently according to their genetic profile. There may be predefined rule engines integrated to flag potential interactions. Others look to FDA, NCCN, or ASCO for guidance. Yet others may be dependent on manual curation of the literature facilitated by computational methods and manual/automated assessment of affected pathways.

17. Do annotations facilitate triage to clinical trials?

For many patients, especially for refractory patients who failed traditional therapies, clinical trials are their last hope. Most large cancer centers have many clinical trials for novel drugs and are actively recruiting patients. It is therefore ideal for annotations to help to identify patients for appropriate clinical trials, especially those that are active at the sponsoring institution. The more detailed the annotation and the more closely aligned to clinical trials available, the more helpful the annotation will be to triage patients for clinical trial accrual.

18. How robust is your ability to annotate for large panels (100+ genes, thousands of translocations, exomes, etc.)?

Currently, for hematopoietic malignancies, NGS often focuses on mutations in myeloid neoplasms with <100 gene panels with notable exceptions. The sequencing efficiency of NGS is rapidly growing and the genes of interest are rapidly expanding. It is deemed to be more efficient to include many genes in a large panel to cover many different diseases than performing sequencing on many different smaller panels, simultaneously or sequentially. Therefore, a foreseeable trend in NGS is bigger sequencing panels. It is important to make sure the collaborating vendor can handle expanding and large sequencing panels to avoid future issues and bottlenecks. It is also desirable for the vendor to be able to serve future clinical programmatic needs, such as translocations in myeloid neoplasm, lymphoid neoplasms, myeloma, acute lymphoblastic leukemia, and outreach efforts.

19. What is your turn around time?

Turnaround time is an important factor in our institute, where sometimes treatment decisions or clinical trial enrollment are time sensitive. The turnaround time ranges for vendor annotation ranges from real-time generation of automated reports (2–20 min) to 3 days for manually curated reports. Factors that affect this include whether or not it involves reinterpretation of reports from other diagnostic companies and the size/complexity of the panel. Depending on the clinical practice, different institutes might require different turn-around time.

20. How would you describe your capacity in terms of annotations/week that your team could handle?

This was not an issue typically. Most of the vendors can handle 120 cases per week with ability to quickly ramp up, should the need arise. But if the vendor has contracts with multiple institutes, the capability to handle your institute's case volume and potential growth should be taken into consideration.

21. How much experience do you have in annotating genomic data? The number of cases? For how many years?

As a cancer center that is involved in patient care, the quality of the annotation and the turn-around time are both critical. Therefore,

it is preferred not to contract with a startup biotechnological company without depth and expertise in this field. Vendors that are available range from a few years to more than 18 years of experience, and from hundreds of cases to several hundreds of thousands of cases. The more the experience and the better the track record, the more reliable will a vendor be.

22. Is the biology of the gene discussed in the annotations?

Annotations should begin with a short discussion of the biological function of the genes affected and their biological pathways to facilitate understanding of their role in disease and pathogenesis. This will also help to validate the interpretation of their role in disease diagnosis or prognosis.

23. Are possible targetable pathways suggested even if the mutation itself is not targetable?

This may be a default setting for some vendors while it is user defined for others. The degree of separation between a variant and pathway affected may also be defined by the user. Care should be taken in this arena as there is a fine line between mere theoretical benefit and scientific support.

24. How does the report look aesthetically? Is it easy to read?

This is also an important factor given that it would affect adoption and reception by our end users, namely the clinicians. Some clinicians have less knowledge and experience in understanding and using molecular information for clinical practice and therefore need the annotation to be simple and straightforward in its assertions of clinical relevance. Even for experienced clinicians, too verbose of a report is annoying and inefficient.

25. How flexible are the parameters in reporting/annotation output? Can the output be customizable?

This is important because there are different clinical needs for different clinical programs that may require adjustments of the way the report looks, what is reported, and the order that it is reported in. For example, some clinicians might be more interested in genes with clinical trials or targeted therapy, while other clinicians might be more interested in genes essential in diagnosis or prognosis. Some clinicians might only be interested in genes with clinical evidence of relevance, while some other clinicians might be interested in genes with non-clinical evidence of relevance. Vendors that offer considerable customizability in this regards with modifiable widgets are desirable.

26. In what format is the deliverable?

Annotations should be easily transferable and accessible. For daily sign out, the annotations should be integrated into the sign out pipeline of the client facility and easily accessible. For client side storage and data mining, the annotations should be transferable in batch. Possible formats include PDF, XML, HL7-compliant message, or Word files. XML files are most formatted for parsing and loading into internal database. The other file formats are also possible to parse but with more difficulty. The method and frequency of annotation transfer should also be discussed. The annotations can be transferred by data file bulk transfer or interface APIs. The client facility might have preference of the method depending on their IT support or other preferences. It is ideal to have the annotations as deliverables that can be searched for validation, quality control, data mining, patient management, or clinical research. The more formatted the deliverable, the easier it is to search. A deliverable in PDF file will be hard to search or to do data mining and is not recommended.

27. Can the annotations be integrated well into data warehousing needs of the client?

The annotation should be easily uploaded or integrated with the client slide database or warehouse. The client database could be Oracle database, SQL database, or REST-based API. The format of the annotation files or API should be compatible with in-house database or API and there should be expertise to facilitate the integration or uploading into in-house database for easy information integration, presentation, and data mining, either for patient management or for clinical research.

28. How does the provided report fit into our LIS (lab information system) for reporting?

The report should be easily connected to and displayed in the in-house lab information system (LIS) and electronic health record (EHR). Ideally, the annotation should be integrated into the in-house database and easy to pull out with other medical results. However, with advances in NGS, more and more genes and variants are tested and consequently detected in a patient. The integration and display could be a challenge to the pathologists and also to the information technology personnel. At our institute, the final report is reported in the EMR as a clickable PDF file, which is easy to view but introduces extra clicks and steps to open the results in comparison with other routine lab results, which is less than ideal. Locations to store the PDF files also need to be maintained and managed and broken links are possible. A better method for data presentation and reporting remains an important topic of discussion.

29. Is the informatics infrastructure present for exchange of PHI in a secure manner?

The exchange of PHI between the vendor and the client should be HIPAA certified and compliant with cybersecurity requirements of the institution. An annual audit is important. To maintain PHI security and HIPAA compliance, a HIPAA business associate agreement should be in place between the covered entity and vendor.

30. What platforms will be used and how easy is it for the exchange of data between the vendor and the client.

For the collaboration to occur, large amounts of data will need to be exchanged between the client and the vendor. The specimen information, the patient information, and VCF files will need to be transferred to the vendor for annotation, while the annotation results will need to be transferred back to the clients. Without a robust IT platform to exchange the data safely and securely, the annotation service will unlikely to be incident free or occur in a timely fashion. This could affect the clinical service and the patient care and, at the very least, affect the turn-around time.

31. What is your cost structure?

Service fees may be charged per case, per case with tiers for number of variants (for example one price if < 6 and another if more than 6 variants), or based on an annual subscription. Other factors that could affect pricing include: pricing based on panel size, complexity, and volume tier (total cases/year). Custom configuration and integration will typically be an extra charge. Quoted rates could vary from \$25 to \$400 per report. Setup fee can run from \$5 K to \$10 K. License and maintenance fees should also be discussed.

32. Who are your other clients?

Information about other clients a vendor has will yield insights about the complexity of the information a vendor can handle. A vendor with mostly private practice clients and a vendor with many academic center clients are likely to have different capabilities and focus. If a vendor has prestigious academic institutes as its clients, it lends confidence that they have passed vetting by another high tier academic institution and that they carry the experience of practical implementation at such complex centers. In addition, one can request contact information of other clients to learn from their experiences with the vendor and collect more objective 'real-world' evaluation of the vendor under consideration.

33. How is medico-legal liability for annotations handled?

This is a concern for pathology team members. While vendors are intricately and extensively involved in report generation, the ultimate responsibility (medical and legal) for the report lies with the signing out pathologist. However, some vendors may carry admissions and error insurance, which is worth checking for. Some vendors may provide professional sign-out services shifting liability from the clients.

34. With regards to customer service and follow up support on questions regarding report, is there something in place?

Sometimes there needs to be further discussion about variants and their significance based on the clinical context or sequencing data (artifact, miscall, etc.). New annotation issues and new sequencing issues could occur anytime and interrupt the routine sign-out of the cases and clinical practice. A reliable vendor should have adequate personnel to support concerns with prompt and high-quality customer service.

3. Discussion and conclusions

Mutational profiling of tumor cells by NGS has become an important part of disease diagnosis, prognosis, and treatment in cancer care. The annotation of the variants detected is a time consuming but critical process of NGS. Due to cost, time and expertise constraints, pathologists may decide not to engage in high-volume NGS annotation or may not be able to provide annotations of the highest quality. A practical alternative is to outsource the annotations to a reliable vendor to reduce the clinical workload of the pathologists and to provide high quality, standardized reporting to the clinicians. In this manuscript, we outline our experience and recommendations concerning the composition of the evaluation team, the steps that should be taken to evaluate potential vendors, and describe the parameters and issues that should be discussed with the vendors or evaluated before a contract is finalized. Our personal final choice of vendor was based on satisfaction of the 34 parameters delineated above, favorable cost structure, ease of integration, high-quality of annotations, wide industry adoption, and high-level of customer service. We believe that this information will be greatly helpful to pathologists and institutes that are considering outsourcing of NGS annotations to improve their quality and turn-around time.

CRedit authorship contribution statement

Jinming Song: Conceptualization, Data curation, Writing - original draft, Writing - review & editing. **Mohammad Hussaini:** Conceptualization, Data curation, Writing - original draft, Writing - review & editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.plabm.2020.e00154>.

References

- [1] H.D. Shukla, J. Mahmood, Z. Vujaskovic, Integrated proteo-genomic approach for early diagnosis and prognosis of cancer, *Canc. Lett.* 369 (1) (2015) 28–36.
- [2] R. Kamps, et al., Next-generation sequencing in oncology: genetic diagnosis, risk prediction and cancer classification, *Int. J. Mol. Sci.* 18 (2) (2017).
- [3] M. Bahassi el, P.J. Stambrook, Next-generation sequencing technologies: breaking the sound barrier of human genetics, *Mutagenesis* 29 (5) (2014) 303–310.
- [4] W.S. El-Deiry, et al., The current state of molecular testing in the treatment of patients with solid tumors, 2019, *Ca - Cancer J. Clin.* (2019).
- [5] S. Morganti, et al., Complexity of genome sequencing and reporting: next generation sequencing (NGS) technologies and implementation of precision medicine in real life, *Crit. Rev. Oncol. Hematol.* 133 (2019) 171–182.
- [6] N. Gangat, M.M. Patnaik, A. Tefferi, Myelodysplastic syndromes: contemporary review and how we treat, *Am. J. Hematol.* 91 (1) (2016) 76–89.
- [7] L. Martínez-Aviles, et al., TET2, ASXL1, IDH1, IDH2, and c-CBL genes in JAK2- and MPL-negative myeloproliferative neoplasms, *Ann. Hematol.* 91 (4) (2012) 533–541.
- [8] L. Malcovati, et al., Driver somatic mutations identify distinct disease entities within myeloid neoplasms with myelodysplasia, *Blood* 124 (9) (2014) 1513–1521.
- [9] R.C. Lindsley, B.L. Ebert, Molecular pathophysiology of myelodysplastic syndromes, *Annu. Rev. Pathol.* 8 (2013) 21–47.
- [10] E. Conway O'Brien, S. Prideaux, T. Chevassut, The epigenetic landscape of acute myeloid leukemia, *Adv Hematol* 2014 (2014), 103175.
- [11] Z.R. Zhao, et al., Mutation profile of resected EGFR-mutated lung adenocarcinoma by next-generation sequencing, *Oncol.* (2019).
- [12] E. Papaemmanuil, et al., Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts, *N. Engl. J. Med.* 365 (15) (2011) 1384–1395.
- [13] W. Vainchenker, R. Kralovics, Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms, *Blood* 129 (6) (2017) 667–679.
- [14] W. Vainchenker, et al., JAK inhibitors for the treatment of myeloproliferative neoplasms and other disorders, *F1000Res* 7 (2018) 82.
- [15] T. Barnoud, J.L.D. Parris, M.E. Murphy, Common genetic variants in the TP53 pathway and their impact on cancer, *J. Mol. Cell Biol.* (2019).
- [16] Y.J. Na, Y. Cho, J.H. Kim, AnSNs: an annotation system to sequence variations of next generation sequencing data for disease-related phenotypes, *Healthc Inform Res* 19 (1) (2013) 50–55.
- [17] J.P. Desvignes, et al., VarAFT: a variant annotation and filtration system for human next generation sequencing data, *Nucleic Acids Res.* 46 (W1) (2018) W545–W553.
- [18] J. Singer, et al., Bioinformatics for Precision Oncology, *Brief Bioinform.* 2017.
- [19] M.J. Landrum, et al., ClinVar: public archive of relationships among sequence variation and human phenotype, *Nucleic Acids Res.* 42 (2014) D980–D985. Database issue.
- [20] M. Griffith, et al., CIVIC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer, *Nat. Genet.* 49 (2) (2017) 170–174.
- [21] S.A. Forbes, et al., COSMIC: exploring the world's knowledge of somatic mutations in human cancer, *Nucleic Acids Res.* 43 (2015) D805–D811. Database issue.
- [22] S.T. Sherry, M. Ward, K. Sirotkin, dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation, *Genome Res.* 9 (8) (1999) 677–679.
- [23] S.T. Sherry, et al., dbSNP: the NCBI database of genetic variation, *Nucleic Acids Res.* 29 (1) (2001) 308–311.