Deep Next-Generation Sequencing Enables Mutational Profiling of Classical Hodgkin Lymphoma

Deep Next-Generation Sequencing Enables Mutational Profiling of Classical Hodgkin Lymphoma

Hayley Robinson, MLS(ASCP)

A Thesis in the Field of Biotechnology

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

March 2019

Abstract

Effective treatments are available for classical Hodgkin Lymphoma (cHL); however, for many patients the success of initial treatment is often followed by a high rate of treatment-related conditions including infertility and secondary malignancies. Due to the rarity of the malignant Hodgkin and Reed/Sternberg (HRS) cells in cHL, molecular genotyping has not been successfully implemented. The inability to obtain the mutational profile currently precludes genotyping of cHL which is informative for clinical management in numerous other cancers. The genetic pathways implicated in the pathogenesis and oncogenic activity of cHL have been described and leveraged for the selection of first-line therapeutics. However, a comprehensive evaluation of relevant mutations in the *SOCS1* gene remain elusive despite its role in cHL being verified in several studies. To assess the feasibility of next-generation sequencing (NGS) for genotyping in cHL, an NGS panel employing Anchored Multiplex PCR (AMP) was designed to target and sequence three relevant genes: *B2M*, *JAK2*, and *SOCS1*. The primary endpoint was suitable detection of low-lying mutations in the rare HRS cells which could be adequately detected by a clinical-grade laboratory assay. Secondary endpoints were the prevalence of mutations and establishing the limit of detection (LOD). Sufficient average sequencing depth for *SOCS1* was achieved in normal feasibility samples at 236 total reads. In cHL cases sequencing coverage over genes *B2M*, *JAK2* and *SOCS1* was achieved at an average of 503 reads, 462 reads, and 191 reads respectively, which demonstrated the technical performance of the method. Diagnostic performance was satisfactory with mutations successfully detected in each of the three targeted genes

for all sequenced cHL cases, with an established LOD of 5% allelic fraction establishing the analytical performance. Despite the technical barriers imposed by the rarity of tumor cells and the necessity to test DNA extracted from tissue biopsies with variable quality, the application of this method can reliably detect somatic mutations with a sufficient level of high quality sequencing coverage across the coding regions of *B2M*, *JAK2*, and *SOCS1*. The demonstrated levels of technical, diagnostic, and analytical performance are adequate to move this approach forward towards implementation of a clinical-grade laboratory assay.

Acknowledgements

I express sincere gratitude to my thesis director Dr. Jochen Lennerz, M.D., Ph.D at the Massachusetts General Hospital's Center for Integrated Diagnostics who provided the inspiration for this work and who thoughtfully guided me throughout the entirety of the thesis process with unwavering support.

Sincere appreciation goes to my colleagues at the MGH without whom this work would not have been possible. Thank you to the Center for Integrated Diagnostics for help with DNA extraction and assisting in the acquisition of discarded samples. Thank you to the Computational Pathology group for bioinformatics analysis, and to Lena Stuart, M.D. of MGH Pathology who assisted in sample selection and who also provided professional interpretation of sequencing results.

I would also like to acknowledge Dr. A. John Iafrate, M.D., Ph.D at the MGH Department of Pathology for his continuous mentorship throughout the course of my graduate studies.

Finally, I express love and appreciation for the endless encouragement from my family.

Table of Contents

List of Tables

List of Figures and Graphs

Chapter 1

Introduction

The genetic pathways implicated in the pathogenesis and oncogenic activity of classical Hodgkin lymphoma (cHL) have been described and leveraged for the selection of first-line therapeutic selection, however a comprehensive evaluation of relevant *SOCS1* gene mutations remains elusive despite its role in cHL being verified in multiple studies. The goal of this work was to investigate whether a next-generation sequencing (NGS), anchored-multiplex PCR (AMP)-based assay could provide a robust assay with which to overcome existing technical obstacles in the genetic characterization of cHL.

Current Outlook for Classical Hodgkin Lymphoma

In 2018 there will be an estimated 8,500 new cases of classical Hodgkin Lymphoma (cHL) diagnosed in the United States (Table 1). Fortunately, with a combination of chemotherapy and radiation therapies remission is attainable. Current treatment options are usually effective - even in cases when cHL is diagnosed in advanced stages. The median age of onset for cHL is 20 to 30 years. It is this considerably young age group combined with a high rate of remissions that generally lends a very optimistic perspective to cHL and a significant proportion of patients can live for many years after initial remission. Specifically, 88% of cHL patients survive five years or greater which qualifies them as long-term survivors (Figure 1).

Table 1: Estimated Number of New Cancer Cases and Deaths, US, 2018 (American Cancer Society, Inc., Surveillance Research)

	Estimated New Cases	Estimated Deaths
All malignancies	1,735,350	609,640
Lymphoma (all)	83,180	20,960
Hodgkin Lymphoma	8,500	1,050
Non-Hodgkin Lymphoma	74,680	19,910

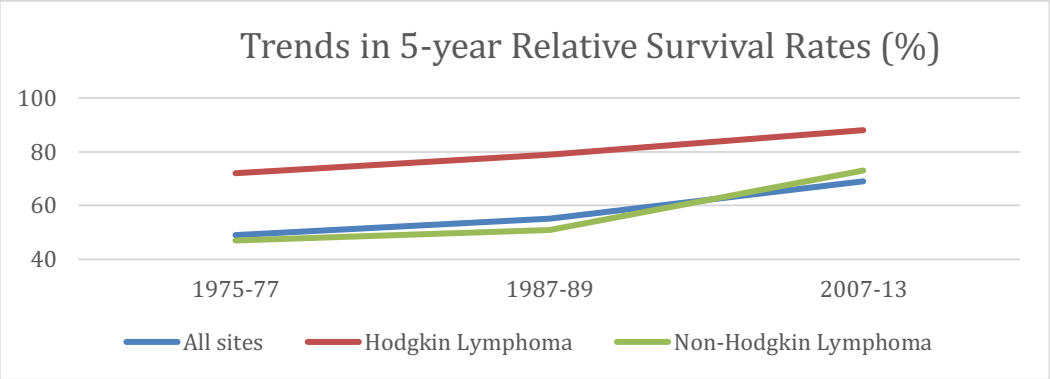

Figure 1. Trends in 5-year Relative Survival Rates (%), US, 1975-2013 (American Cancer Society, Inc., Surveillance Research, 2018).
Note: Non-Hodgkin Lymphoma trend provided for comparison.

Despite the optimistic long-term survival rates, a remaining key challenge is how to effectively identify the subset of patients who will suffer disease recurrence and relapse despite initial successful treatment. This diagnostic challenge persists and needs to be addressed since there is a lack of established clinical standards with which to stratify patient and disease characteristics as related to the efficacy of treatment selection. Of note, 12% of patients will die within the first five-years post-treatment as shown in Figure 1. This subset of patients requires further investigation to determine why they ultimately relapse. An additional unfortunate reality is that the toxic treatment courses of chemotherapy and radiation often cause damage to normal cellular DNA. Given the

typically long period of survival post-therapy, these DNA aberrations can propagate over time and manifest in an increased risk of developing a multitude of health problems including secondary malignancies (observed in around 19% of survivors), and a range of fertility issues (Biasoli, 2017). The majority of patients will successfully go into remission after treatment but remain extremely vulnerable to these off-target treatment effects. These effects are important to prevent; a greater ability to stratify patients by underlying mutational signature prior to treatment selection may help minimize deleterious side effects including secondary malignancies.

cHL, like many cancers, is in essence a genetic disease. One can argue, again in analogy to other malignancies, that molecular genetic characteristics might be an effective way to identify relevant patient subsets. Much progress has been made in this area in the field of cancer and the molecular characterization of tumors. While this approach has become the clinical standard and facilitates more personalized and tailored treatments with both efficacy and optimized long-term outcomes in many solid tumor diagnoses, routine genotyping has not been accomplished in cHL due to a series of technical, biological and practical barriers.

Technical Obstacles Unique to the Genotyping of cHL

cHL is a hematologic malignancy that presents multiple challenges in its genetic characterization. These obstacles hinder the ability to identify mutations in relevant genes that could be utilized for the development of targeted therapies and advance the improvement of diagnostic and prognostic treatment decision-making.

The current clinical standard for the evaluation of therapeutic options in cancer calls for molecular genotyping to uncover genetic mutations within the DNA sequence derived from a tumor cell. The goal is to uncover the mutations involved in the development and progression of the disease to leverage the application of targeted therapies; agents which are designed to specifically combat the mechanisms of the tumor and ideally reduce or eradicate their propagation. This process relies on the relative abundance and availability of DNA extracted from tumor cells contained in a biopsy or a blood sample. For most solid tissue cancers or leukemia this is not problematic because it is relatively easy to extract cellular tumor DNA. In contrast, there are several technical obstacles that make this problematic in cHL.

Scarcity of Tumor Cells

A critical part of genotyping is to specifically interrogate the genome of tumor cells contributing to the cancer. Only the cancer's tumor genome will hold the mutational signature of the malignancy. Therefore, tumor cells must be enriched prior to the extraction of nucleic acids, DNA and RNA, for this type of analysis. This process has been highly standardized for solid tumor malignancies. Typically, a biopsy will aim to remove a piece of tissue containing the tumor. With very few exceptions, every tumor contains non-tumor cells (e.g. the vessels supporting the tumor). A pathologist can select and prepare regions of tumor tissue that are enriched for tumor cells and minimize the presence of non-tumor cells. Of note, it is this variable combination of both tumor and non-tumor cells and subsequent nucleic acids which present a consistent limitation for

establishing sensitive and specific genotyping methods with which to determine mutation status.

The tumor cells in cHL, named Hodgkin and Reed/Sternberg (HRS) cells are the diagnostic hallmark of the disease. These cells are of B-cell origin which is why cHL is classified as a hematological disease. However, the HRS cells are extremely rare and are typically outnumbered by thousands of surrounding non-tumor cells (the so-called inflammatory infiltrate). Figure 2 depicts the typical ratio of HRS tumor cells to the non-tumor cells infiltrating the surrounding environment. The tumor microenvironment in cHL makes this disease unique and unlike any other hematologic malignancy.

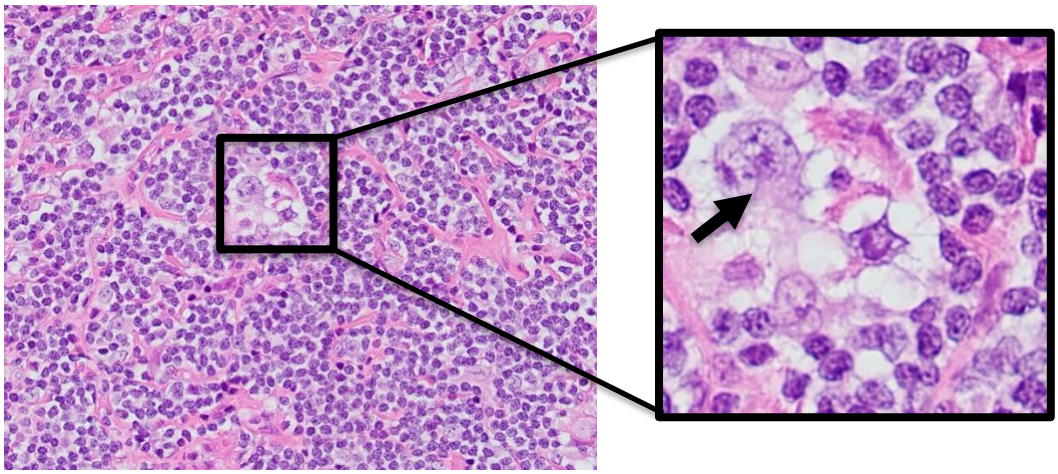

Figure 2: Diagnostic Hallmark of cHL: One HRS Tumor Cell, (inset, arrow) surrounded by thousands of inflammatory/non-tumor cells. Hematoxylin and Eosin (H&E) stain, 20x magnification.

To definitively diagnose cHL, it is essential to extract tissue to make a pathological assessment. HRS cells typically contribute to only 0.1 to 2% of the entire tumor mass; the remainder of the mass will be comprised of inflammatory immune cells (Aisenberg, 2017). This small fraction of tumor cells is a unique problem in the

diagnostic realm of cancer; comparatively in solid tumor malignancies like lung cancer, the fraction of tumor cells typically derived from a biopsy is usually at least 30% (Colemen & Tsongalis, 2006; Jennings et al., 2017). The HRS cells recruit surrounding immune cells to effectively cloak it from the immune system, thus escaping detection and destruction. This cloaking mechanism also complicates the histological identification and selection of these cells for molecular characterization since only HRS cells contain the mutated DNA required for genotyping. Any contributing DNA originating from the surrounding inflammatory cells collected during a biopsy and subsequent extraction of DNA will not provide a HRS-specific mutational signature. In other words, due to the small fraction of tumor to non-tumor DNA, one is effectively looking for a needle in a haystack. Practically speaking, the DNA extraction process begins with a section of formalin-fixed paraffin-embedded (FFPE) tissue being physically scraped and placed into a secondary container for enzymatic digestion. This physical step ultimately leads to DNA from both HRS and normal cells being combined. Therefore, both cell populations and contributing DNA profiles will be confounded in downstream sequencing analyses. For example, if a biopsy sample was comprised of a ratio of 1 tumor cell to 999 normal cells (1:1000); DNA from 999 non-tumor cells will be combined with DNA from just 1 tumor cell. By extension, and when considering an oncogenic driver mutation that generally affects a single allele, a biopsy slide containing 50% tumor cells and 50% normal cells would be expected to have 25% mutated DNA at the genomic location of a variant; or 25% of all total sequencing reads will show the mutation whereas 75% of reads (composed of 25% tumor wild-type reads and 50% non-tumor DNA) will show a normal or wildtype sequence.

Recent studies have revealed a signature genotype of cHL by interrogating the DNA of HRS cells (Schif et al., 2012; Reichel et al., 2014). However, due to the low HRS cell content within the larger inflammatory cell mass, DNA from surrounding cells essentially dilutes down the allelic frequency of mutations sought to be identified. Historically, molecular genotyping techniques have been unable to overcome this technical obstacle and genotyping requires laborious single cell isolation techniques like laser capture microdissection (Lennerz et al., 2015) to directly obtain HRS-derived nucleic acids.

FFPE Fixation of Biopsy Tissue

Another obstacle to the genotyping of cHL cases is the condition of the DNA acquired for analysis. Despite being classified a hematological malignancy, cHL typically affects lymph nodes and the histopathological diagnosis relies on a lymph node biopsy (Hudnall & Küppers, 2018). As with other histopathological diagnoses that cannot be ascertained by a blood draw, a biopsy of affected tissue is required to undergo subsequent chemical fixation which preserves the integrity of the cells. In order to examine the histology of the cells the fixed tissue is dehydrated and embedded in paraffin. The resulting tissue sample is typically referred to as Formalin-Fixed Paraffin-Embedded (FFPE) tissue. The method is widely-used and maintains cellular contents and morphology (Colemen & Tsongalis, 2006). While this process is compatible with DNA or RNA extraction and most genotyping applications including next-generation sequencing (NGS) the fixation process is harsh, involves DNA crosslinking, and tends to impact the quality of DNA by fragmentation. Both the FFPE process and DNA extraction

require the use of stringent agents which tend to cause DNA strand breakage leading to a smaller-than-ideal DNA fragment size. Additionally, FFPE tissues – despite fixation - degrade over time which is particularly important in diseases with survival times greater than 5-10 years like in cHL (Colemen & Tsongalis, 2006).

Delineating Genetic Aberrations in cHL is Clinically Relevant

Two genes that carry prognostic relevance in cHL are *JAK2* and *B2M*. These genes have significant implications for therapy selection and outcomes in cHL, particularly when considered in conjunction with *SOCS1* mutations (Schif et al., 2012). Mutations in *B2M* have been linked to a cHL subtype occurring in cHL patients in the lowest age group and who tend to have a better prognosis as compared to those without *B2M* mutations (Camus et al., 2016). While *B2M* and *JAK2* have been widely studied allowing a well-characterized picture for their roles in cHL, establishing a reliable genetic signature for individual patients in these genes has not been implemented. It was recently reported that mutations in the *SOCS1* gene can predict outcomes in cHL patients (Lennerz et al., 2015). Survival outcomes and even relapse after treatment were found to be associated with subtypes of mutations in *SOCS1.* However, a routine method with which to establish a clinical-grade assay that can provide a comprehensive genotype for this gene has not yet been established. *SOCS1* does not have any known hotspot mutations, meaning there are no commonly-seen mutations at similar locations within the gene across multiple cHL samples. Instead, mutations observed in *SOCS1* are randomly distributed throughout the coding region of *SOCS1*. Briefly, *SOCS1* is classified as a tumor suppressor gene which encodes for a protein inhibiting cellular growth and

proliferation. It has been observed that when *SOCS1* is mutated, its usual role in inhibiting the JAK-STAT pathway via a negative feedback loop is disrupted and JAK-STAT signaling is stabilized (Sukka-Ganesh, 2016; Zhang et al., 2012). This sustained activation of STAT promotes neoplastic cell division and induces inflammatory responses facilitated by immune cells as observed in the clinical presentation of cHL (Tiacci et al., 2018). Functionally, *SOCS1* provides a checkpoint in the cell cycle to either permit a cell to divide after verifying there are no significant DNA replication errors, or to selectively halt a cell from further division (Figure 3). In the case of an oncogenic and malignant process, it would be the goal of the malignant cell to disrupt and mutate the sequence of its tumor suppressor genes to proliferate cells as quickly as possible; effectively creating a fast-track through cell cycle checkpoints.

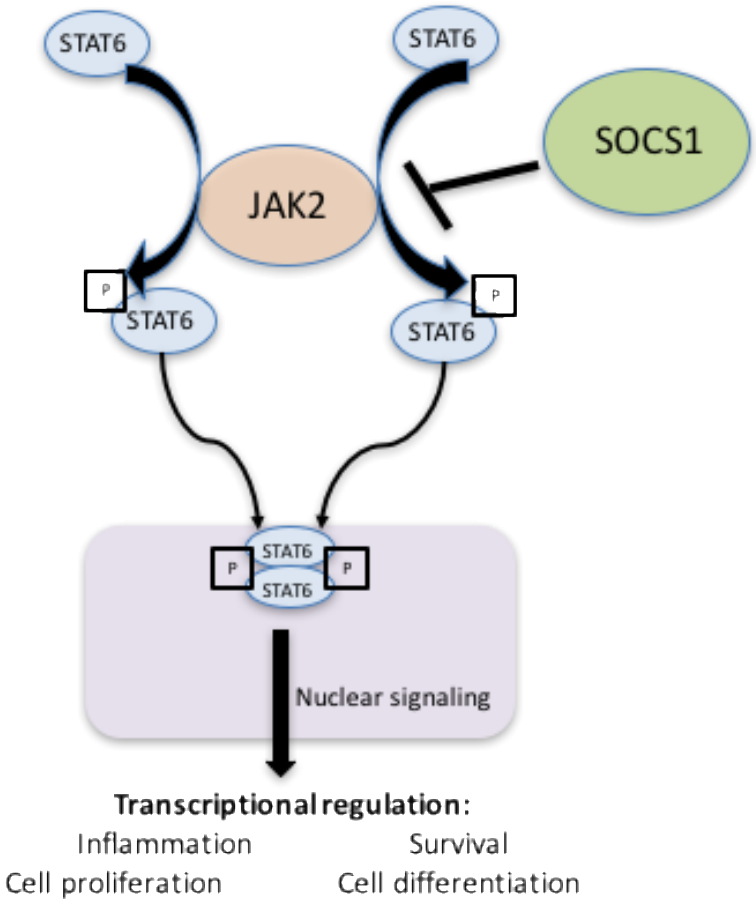

Figure 3. Role of *SOCS1* within the JAK-STAT Signaling Pathway of a Mutated Tumor/HRS Cell.
Note: If *SOCS1* is functionally impaired, there is a loss of inhibitory pathway activity which effectively results in hyperactive JAK/STAT signaling.

High Guanine-Cytosine Content of *SOCS1*

Certain regions of the genome are considered high guanine-cytosine (GC) areas, meaning that many of the DNA bases in the gene's sequence (at least 60%) are comprised of either guanine or cytosine as opposed to adenine or thymine (AT) (Jennings et al., 2017). *SOCS1* is 83% GC-rich (Lennerz et al., 2015) and this feature is an obstacle in molecular biology techniques. AT base pairs are bonded together with two hydrogen bonds whereas GC pairs are bonded with three hydrogen bonds. This additional hydrogen

bond leads to formation of a secondary structure that stabilizes the GC pair. Due to this increased stability of GC pairs, the standard temperatures used in PCR cycling protocols are generally not suited to target these genomic regions and will result in an over-representation of non-GC rich areas after sequencing analysis concludes. This is a major factor as to why *SOCS1* is severely under-represented in previous studies despite strong performance of other relevant genes in previously-published data analyses (Weniger et al., 2006, Reichel et al., 2014). The ability to compile *SOCS1* mutation characterization along with the existing body of knowledge for *B2M* and *JAK2* would provide a more comprehensive mechanism to biologically understand cHL.

Anchored Multiplex PCR and Next-Generation Sequencing

A limited number of studies have incorporated the *SOCS1* gene into their genotyping panels but have been unable to provide sufficient sequencing coverage, and therefore lack satisfactory mutation characterization for cHL (Weniger et al., 2006, Reichel et al., 2014). Many studies have provided sequencing coverage over specific areas of the *SOCS1* gene, but not over the entire coding region. This goal is limited by traditional PCR methods in detecting mutations, because they require that PCR primers be designed specifically to amplify all possible regions of the gene. A reliable method of genotyping is necessary to obtain *B2M*, *JAK2* and *SOCS1* sequences for their inclusion within a clinical-grade genotyping assay. Given the multiple technical obstacles described above, this genotyping method will need to reliably provide clinical quality data despite (1) low tumor cellularity, (2) GC-rich DNA, and (3) DNA extracted from FFPE biopsies. A potential solution to the limitations in genotyping cHL samples is

Anchored Multiplex PCR (AMP); an NGS library preparation method which allows for the detection of mutations within genes of interest but without requiring precise design of two flanking primers as used in traditional PCR methods (Zheng et al., 2014).

Traditional DNA sequencing requires one pair of individual primers; short fragments of DNA between 18 and 22 bases long designed to match the sequence on either side of the DNA regions of interest. The primers are only able to attach or anneal if the sequence exactly matches the DNA region of interest; therefore, if a mutation exists in tumor DNA within the sequence where the primer is designed to anneal, the primer may not be able to attach and PCR enrichment over this area is unlikely to be successful. This creates an underrepresentation of this genomic region in downstream analyses. The goal of PCR enrichment is to synthesize a new copy of an original tumor DNA template molecule targeted by the primers. PCR enables this reaction to occur thousands of times resulting in thousands or millions of copies of the original genomic regions acquired from the biopsy's DNA.

After PCR amplification and sequencing, the sequence of thousands of reads can be mapped back to the genome. By utilizing this method, it may be feasible to uncover previously undescribed mutations using a sequencing panel comprised of genes known to be actionable and relevant for cHL.

Given the above barriers of limited tumor cellularity, FFPE-treated tissue and high GC-rich content for genes in the relevant pathways affected by somatic mutations in cHL, the aim of this research was to assess feasibility of an AMP-based NGS assay to enable direct genetic characterization of *B2M*, *JAK2* and *SOCS1* in DNA extracted from FFPE cHL samples. With an appropriate modality to perform direct molecular genotype

characterization, one can better understand the underlying pathway alterations, distinguish separate patient subsets, and possibly improve long-term outcomes. The concrete practical implications of this work are leveraging an AMP-based method to perform deep NGS sequencing and establish a clinical-grade assay for cHL.

Chapter II.

Materials and Methods

The assay design and sample processing techniques are outlined in this section. All experiments were performed in accordance with the Declaration of Helsinki.

Study Design and Institutional Approval

The study was designed as a retrospective biomarker study. The primary endpoint was the ability to detect DNA mutations in HRS cells. Secondary endpoints were the prevalence of mutations and establishing the limit of detection (LOD). The study was approved by the Institutional Review Board of the Human Research Protection Program at Harvard University (Protocol ID IRB18-0014).

Case Selection

Computer searches using a custom-made laboratory information system (Massachusetts General Hospital, Boston MA, USA) were conducted to identify excess tissue from Primary Mediastinal B-cell Lymphoma (PMBL) and cHL patients. At least two board-certified pathologists evaluated each case by examining hematoxylin and eosin (H&E)-stained tissue sections to determine the tumor cell-rich areas prior to DNA extraction. Due to the rarity of cHL, it was assumed that an ideal number of cases

available with sufficient excess tissue for evaluation would be limited. Therefore, information from previously published studies was employed to identify a high prevalence of *SOCS1* mutations in diseases that have higher tumor cell content and mutation prevalence and therefore greater availability of material to establish technical feasibility of the assay. Practically speaking, cHL test cases were supplemented with PMBL samples, a lymphoma that exhibits a similar mutational profile. Specifically, 45% of PMBL cases harbor *SOCS1* mutations (Weniger, 2006). Initial feasibility experiments employed two random non-tumor blood samples and three DNA samples derived from non-PMBL, non-cHL FFPE cases (feasibility samples).

DNA Extraction and Qualification of Samples

DNA was extracted from tumor tissue sections and blood samples. Feasibility samples and tumor tissue from PMBL and cHL cases was scraped from biopsy slides and cellular DNA was extracted using the Agencourt FormaPure reagent kit (Beckman Coulter catalog #A3348, Beckman Coulter, Brea CA, USA). Blood samples were extracted using the Maxwell RSC Blood DNA kit (Promega catalog #AS1400, Promega, Madison WI, USA). The DNA concentration for each nucleic acid extraction was quantified using a ThermoFisher Varioskan Flash instrument in conjunction with the Quant-IT Broad-Range dsDNA assay kit (Thermo Fisher Scientific catalog #Q33130, Waltham MA, USA).

Feasibility Assessment of AMP Panel

Samples originating from non-PMBL and non-cHL samples (from both blood and FFPE) were selected for feasibility testing. These samples served to establish that the primers were capable of accurately targeting the three genes and that they yielded accurate sequencing data in addition to showing that the genomic regions were in fact able to be sequenced (e.g., considering the high GC content of *SOCS1*).

AMP and NGS

NGS libraries were prepared from DNA extracted from feasibility, PMBL, and cHL cases utilizing an AMP approach. Using the ArcherDX Assay Designer tool (ArcherDX, Inc., Boulder CO, USA) and the University of California Santa Cruz (UCSC) Genome Browser (Casper et al., 2018), a primer panel was designed to specifically target and amplify the full coding regions of the *B2M*, *JAK2* and *SOCS1* genes (ArcherDX, Inc. catalog #CB0114, Boulder, CO, USA). Table 2 summarizes the content of the gene panel; a total of 91 primers were incorporated to target the three genes.

Table 2. Targeted Genes in the AMP Panel

Target name	NCBI reference sequence	Genomic location	Target exons
B2M	NM_004048	chr15:45003684-45010357	1 - 4
JAK2	NM_004972	chr9:4985244-5128183	3 - 25
SOCS1	NM_003745	chr16:11348273-11350039	1 - 2

Abbreviation: NCBI, National Center for Biotechnology Information

AMP Library Preparation of PMBL and cHL Cases

DNA samples were normalized to 200ng total DNA. NGS-compatible libraries were prepared using the AMP method. Specifically, DNA fragments underwent enzymatic fragmentation, end repair, A-tailing and adapter ligation with unique barcodes (ArcherDx, Inc. Custom VariantPlex Kit for Illumina, catalog #CB0114, Boulder CO, USA). The adapter-ligated DNA fragments were placed into the first of two rounds of PCR cycling. The first round involves the addition of gene-specific primers, designed to anneal to a region of interest in a sample's DNA fragment. The PCR cycling amplification step completes the sequence of this strand using the enzyme Taq polymerase. During the second round of amplification the second set of uni-directional gene-specific primers anneal to the same DNA fragments a few DNA bases away from where the first primer annealed. This step completes the template sequence in the same direction as the first primer. This technique allows for a nested, two-step amplification process in which the second amplification further enriches the first round and increases the abundance of sequence-specific library fragments. After two rounds of this nested PCR process, NGS libraries are quantified to ensure sufficient amplification (Universal Complete KAPA Library Quantification Kit, catalog #KK4824, Roche, Basel, Switzerland). DNA sequencing was performed on an Illumina NextSeq 500 instrument using a 300-cycle sequencing kit (NextSeq 500/550 High Output Kit-300 cycles, catalog #FC-404-2004, Illumina, Inc., San Diego CA, USA).

Bioinformatics Analysis

The raw data output from the NextSeq sequencer (.bcl files) were first de-multiplexed and consolidated into FASTQ files. This was achieved using individual barcodes ligated to each individual sample library. Using NovoAlign, FASTQ files were then aligned to the human reference genome version hg19. The NovoAlign tool (http://www.novocraft.com) was also utilized to tag duplicate reads. Briefly, duplicate reads are artifacts of the PCR process which represent exact copies of one template strand of DNA. Prior to duplicated reads being consolidated to remove redundancy (a process referred to as 'deduplication'), the total number of sequencing reads (typically provided as the mean absolute read coverage) is comprised of all sequencing reads mapping to a genomic region. After deduplication, the total number of unique sequencing reads is provided as the mean collapsed coverage (MCC) value. Single nucleotide variant (SNV) and insertion/deletion (InDel) detection was performed using MuTect1 (Cibulskis et al., 2013), LoFreq (Wilm et al., 2012), and GATK (Mckenna et al., 2010). The output file is referred to as a variant call file, or VCF. A BAM file was created for each library and detected variants were annotated using Variant Effect Predictor (VEP). VEP assigns a specific genetic consequence to a mutation based on the impact incurred on coding regions of exons, downstream transcripts, proteins, and regulatory regions of DNA. Integrated Genome Viewer (IGV, Broad Institute, Cambridge MA, USA) was used to visualize sequencing reads and variants. Quality metrics were generated with High-Throughput Quality Control (Yang et al., 2013) and SAMtools (Durbin et al., 2009).

Variant Calling

Each sequenced sample was reviewed by a board-certified pathologist. Variants were reviewed and qualified as true calls based on quality metrics. At least 5 bi-directional non-wildtype reads were required to confirm the presence of a variant. Data from the VCF files were extracted and summarized based on bioinformatics pipeline settings: variants were required to be non-synonymous, non-UTR, non-intronic, and not present in ExAC (Malmstroem, 2017).

Data Analysis and Statistics

Allelic frequency of a variant was derived by evaluating the total number of deduplicated reads mapped over a variant region including mutated or alternate (ALT) reads and wild-type or reference (REF) reads. The allelic fraction was calculated by dividing the number of ALT reads by the total number of ALT + REF reads. For example, for a *SOCS1* variant location with 262 total reads mapped, 10 of which contain a mutation call and 252 containing a non-mutated sequence, the allelic frequency of this *SOCS1* mutation would be derived as 3.8% (10/262). To assess sequencing coverage, findings were compared to previously-published work (Reichel et al., 2014). To derive the theoretical limit of detection, a binomial distribution was used to establish the probability of detecting a variant with a minimum of five alternate reads at different allelic fractions. Variants were classified into one of twelve tiers and plotted using ggplot2 per the number of variants observed at each genomic position. Statistical comparisons of means employed t-testing using Microsoft Excel and significance was defined as $p < 0.05$.

Chapter III.

Results

The assay was created and verified to successfully sequence the genes desired for targeted genotyping. Data collected from these experiments are described in this section.

Cases Selected for Study

A computer search identified a total of 56 cases for evaluation. This cohort of cases included 5 feasibility cases, 45 PMBL and 6 cHL cases. Fifteen PMBL cases were eliminated due to insufficient DNA yields resulting in a total of five feasibility, 30 PMBL and six cHL cases forming the study set. The five non-PMBL and non-cHL feasibility samples consisted of two blood and three FFPE samples, and all samples had sufficient DNA for library preparation. After feasibility testing was completed, processing of the PMBL and cHL samples commenced. After library preparation and quantification, a total of 28 PMBL cases and 6 cHL cases were deemed acceptable for sequencing based on their library yields being greater than 2000 pM. See Appendix 1 for a summary of sample attributes including sample age prior to extraction, DNA yield, and status of case after evaluation of quality.

AMP Assay Performance for Feasibility Cases

Feasibility testing to assess significant quality differences between FFPE and non-FFPE cases achieved an MCC of at least 200 reads over the *SOCS1* genomic region for DNA samples extracted from both blood and FFPE. Figure 4 shows a representation of where the primers designed to target *SOCS1* are located and intended to anneal during the library preparation protocol. In blood samples, the average MCC across *B2M*, *JAK2* and *SOCS1* was 577.3, 585.4, and 339.8 respectively, whereas in the FFPE-based samples the average MCC across *B2M*, *JAK2* and *SOCS1* was 346.9, 401.6, and 236.2 respectively. A minimum of 200X MCC was achieved for all feasibility samples, as shown in Figure 5. Figure 6 shows one exemplary FFPE sample's pileup of aligned sequencing reads visualized in IGV. This visualization shows that the coding regions of *SOCS1* can be successfully sequenced and that primer design was correct.

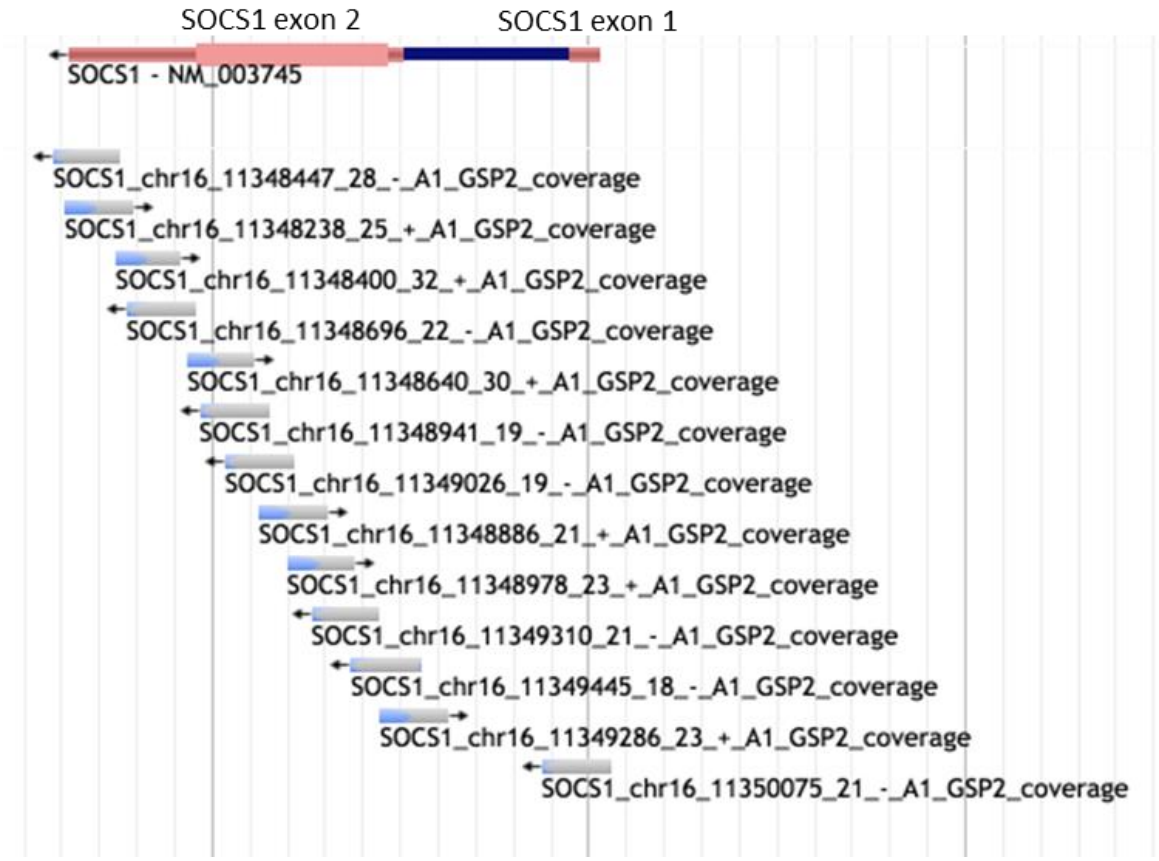

Figure 4: Gene-Specific Primers Designed to Target Coding Regions of *SOCS1*
Note: SOCS1 exon 1 is a non-coding region.

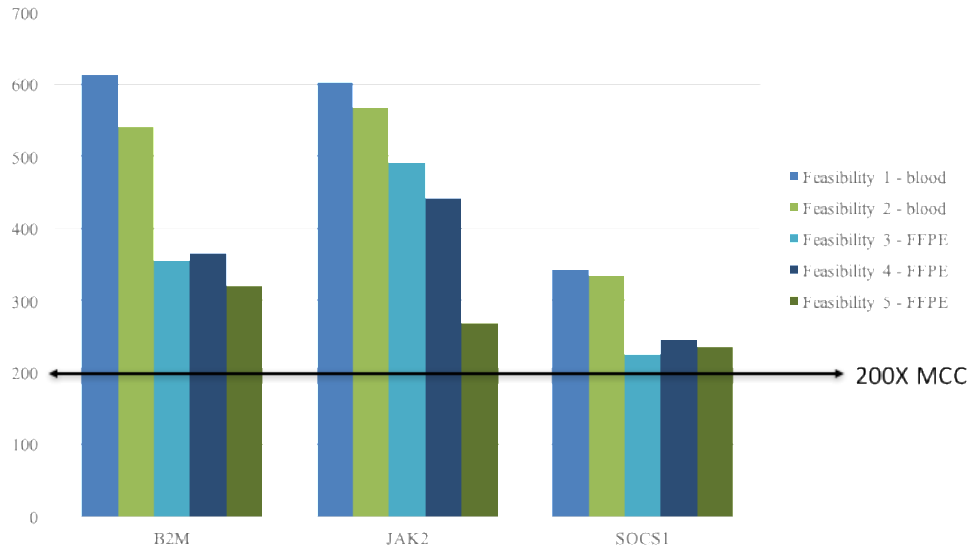

Figure 5: MCC Across Panel Genes for Feasibility Cases

Abbreviations: MCC: Mean Collapsed Coverage. Note: Ideal threshold of 200X MCC is marked.

AMP Assay Performance for PMBL and cHL Cases

Sufficient NGS library concentrations were achieved for 28 PMBL cases and 6 cHL cases. A minimum library yield of 2000 pM was needed for sufficient sequencing reads and two libraries had to be eliminated due to low library yields (10 pM and 366 pM, respectively). Each library was normalized prior to sequencing to achieve 4.4 Gb of sequencing data.

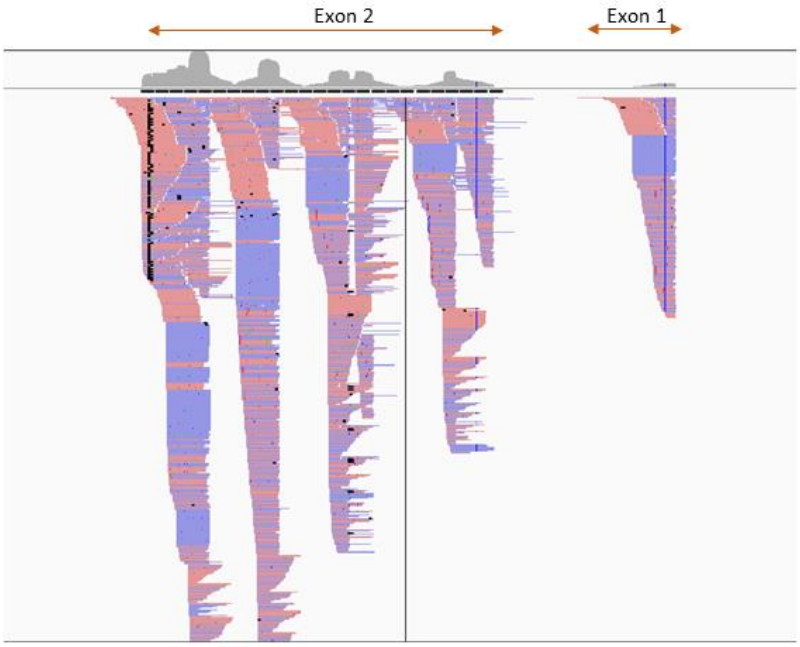

Figure 6. *SOCS1* Sequencing Coverage for One PMBL Sample in IGV
Abbreviations: IGV: Integrated Genome Viewer

Sufficient Sequencing Coverage Achieved for *B2M* and *JAK2*

Sufficient sequencing coverage was achieved in the PMBL cases over the three genes sequenced in the panel. Despite the suboptimal DNA yield and quality likely due to

the FFPE storage time of the PMBL cases, greater than 200 MCC was reached over *B2M* and *JAK2* in 75% and 82% of the PMBL cases, respectively (Figure 7). Comparison of the MCC for both PMBL and cHL cases showed that the performance is similar (Figure 8) with all six cHL cases reaching a minimum of 200X MCC over *SOCS1* exon 2. Of note, the cHL cases achieved significantly higher MCC over the four exons of *B2M*. The cHL cases had significantly shorter FFPE storage times (no longer than 2 years) which likely increased the DNA fragment size quality. By contrast, the storage time of the 28 sequenced PMBL cases ranged from 1 to 18 years with a median of 11 years.

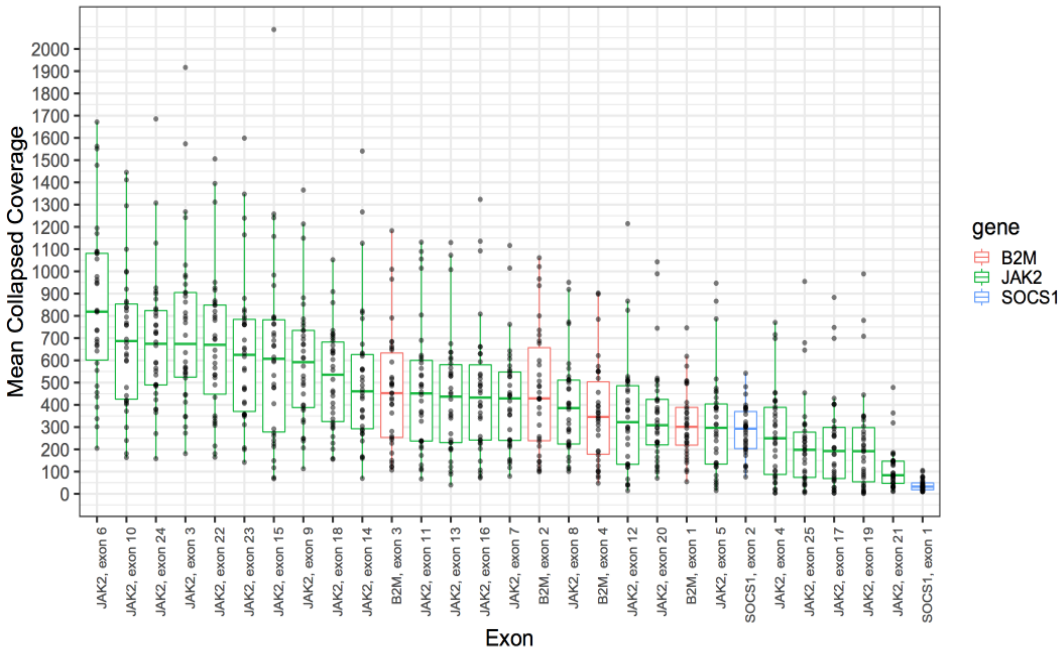

Figure 7. MCC for PMBL Cases Across All Exons Targeted by Panel
Abbreviations: MCC: Mean Collapsed Coverage

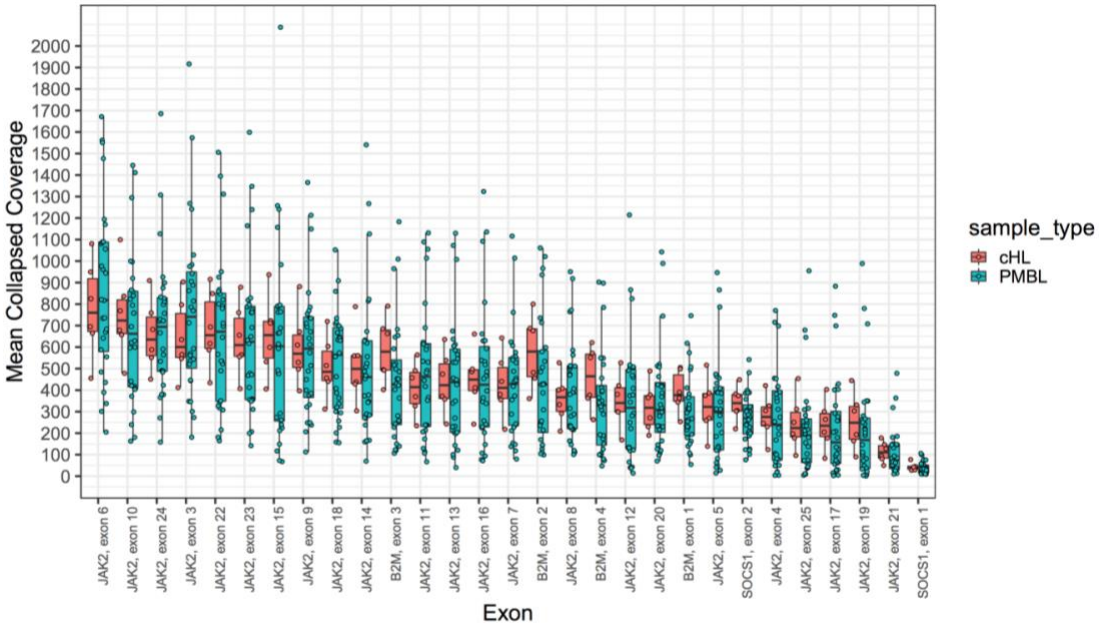

Figure 8. MCC Across All Exons for cHL and PMBL Cases
Abbreviations: MCC: Mean Collapsed Coverage

Sufficient Sequencing Coverage Achieved for *SOCS1*

The non-coding exon 1 region of *SOCS1* proved to be difficult in producing enough sequencing coverage (Figure 9) likely due to its higher GC rich content. Only two PMBL samples reached 100 MCC for exon 1. Similar performance was observed for the cHL cases (Figure 8). A sufficient amount of unique sequencing reads were produced over exon 2 of *SOCS1*: 75% of the PMBL samples had an MCC greater than 200 (Figure 10). This establishes that DNA derived from FFPE can show reproducible and sufficient sequencing coverage for variant calling over the *SOCS1* exon 2 coding regions.

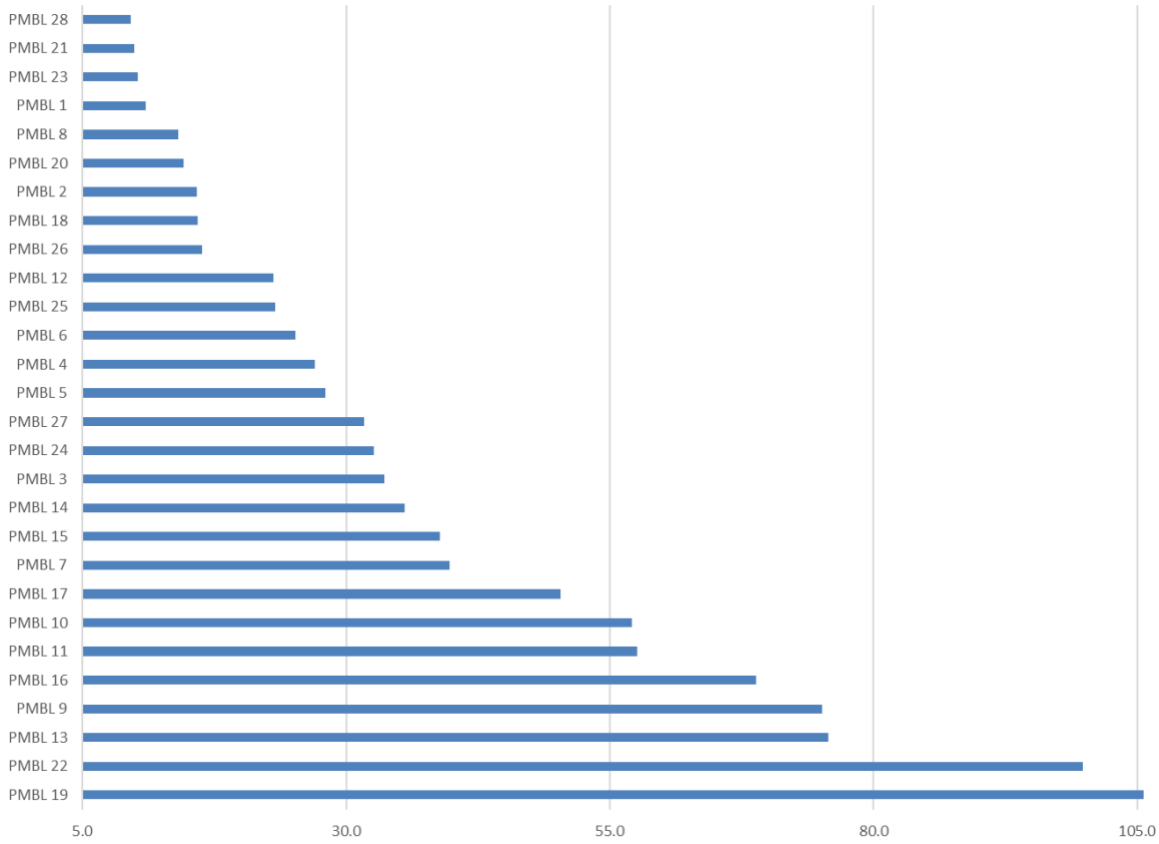

Figure 9. MCC Across PMBL Samples for *SOCS1* exon 1
Abbreviations: MCC: Mean Collapsed Coverage

Despite the suboptimal sequencing coverage over exon 1 of *SOCS1*, this result is encouraging since exon 1 is a non-coding region – the most biologically-relevant mutations affect the coding region of exon 2. The age of the tissue is likely an additional contributing factor in samples that did not perform optimally.

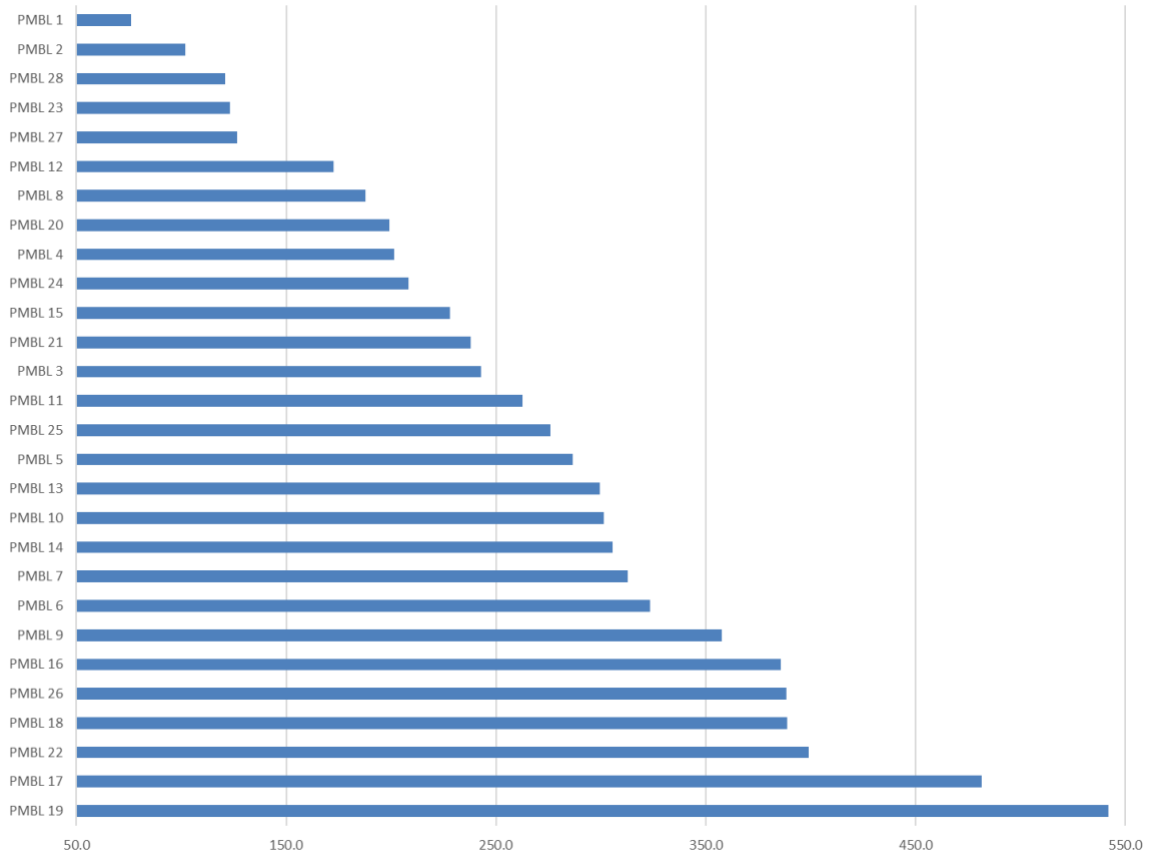

Figure 10. MCC Across PMBL Samples for *SOCS1* exon 2
Abbreviations: MCC: Mean Collapsed Coverage

Establishing the Limit of Detection

The limit of detection (LOD) based on the above sequencing results was established assuming a variant would be present with a minimum of 5 alternate sequencing reads featuring a genetic mutation; this minimum threshold is a quality metric currently employed to qualify a mutation call as being reliable (Center for Integrated Diagnostics, Massachusetts General Hospital, Boston MA, USA). It was determined that the LOD based on the data accumulated is 5% allelic fraction (99.99% sensitivity) as shown in Figure 11 with an acceptable level of sensitivity down to 2% allelic fraction (93.90% sensitivity) and a limited ability to detect variants at a 1% allelic fraction (44.73%

sensitivity). Alternatively, there is a 95% probability that with an MCC of at least 150X, variants can be detected at a 5% allelic fraction.

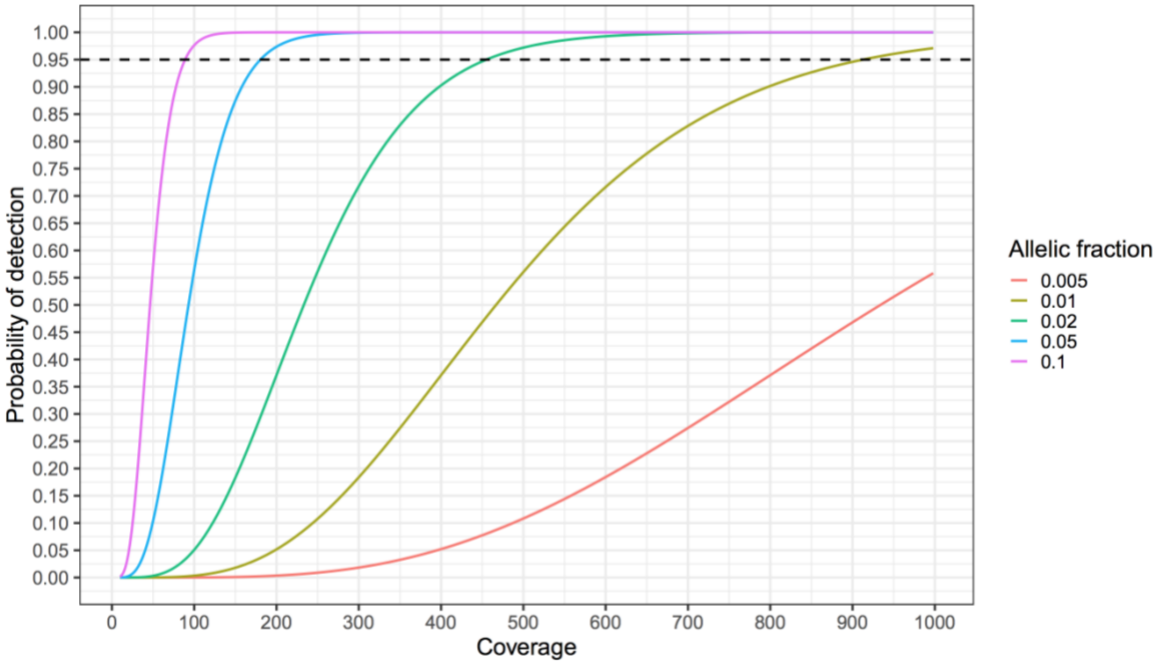

Figure 11. Limit of Detection Analysis
Mean Collapsed Coverage level required to call variants at various allelic fractions at a 95% confidence level.
Note: a binomial distribution was used to establish the probability of detecting a variant assuming a minimum of five alternate reads at different allelic fractions.

Sequencing Coverage Analysis

To keep these results in context, a summary of sequencing depth achieved for *SOCS1* from a recent study is summarized in Table 3 (Reichel et al., 2014). This study utilized whole exome sequencing to characterize the genetic signature of HRS cells. The average sequencing depth achieved over *SOCS1* was (95 ± 7 reads) and fell short of 200X, which would argue against feasibility in clinical practice. In comparison, AMP sequencing achieved (>853.8 ± 161.2 reads) which was significantly higher (p=<0.05, t-

test). These results present a contrast to the ability of this AMP-based assay to successfully sequence the *SOCS1* region of the exome.

Table 3. Comparison to Study by Reichel et al. (2014)

Gene	Chromosome position	Total Sequencing Depth (reads)	
		Reichel et al. (cases 1,2,3,4,7)	AMP Method (Representative PMBL case)
SOCS1	11348948	93	749
SOCS1	11348656	102	1476
SOCS1	11348989	102	911
SOCS1	11348766	120	241
SOCS1	11348907	70	900
SOCS1	11348956	84	846

Abbreviation: PMBL, Primary mediastinal B-cell lymphoma
Note: Total sequencing reads for one PMBL case are included for comparison. Reichel et al., *2014* data standard error of the mean (SEM) = 7.0, AMP data SEM = 161.2

Detection of Variants

Figures 12 and 13 show the distribution of detectable variants across the three genes targeted in the AMP assay for the sequenced PMBL and cHL cases, respectively. The distribution of mutations noted across the genomic regions of each gene for the PMBL cohort further illustrates the technical feasibility achieved by this method. While the number of cHL cases is limited, coding mutations in *SOCS1* were indeed detected. Appendix 2 indicates how many mutations were detected in each case across the three targeted genes. Additional detailed analysis of the specific variants is ongoing.

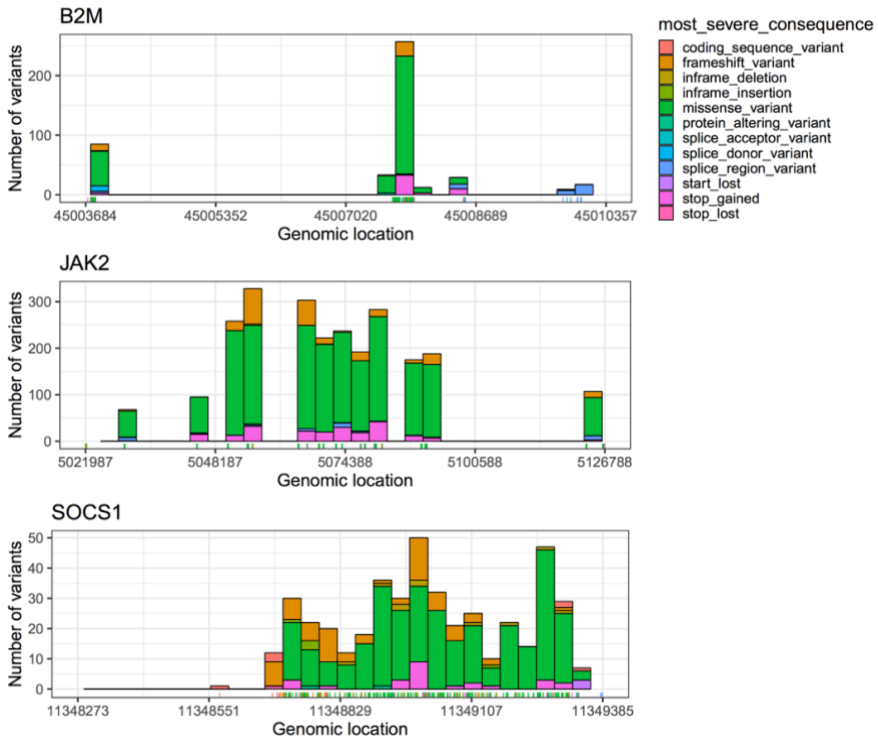

Figure 12. Total Number of Variants Detected in PMBL Cases

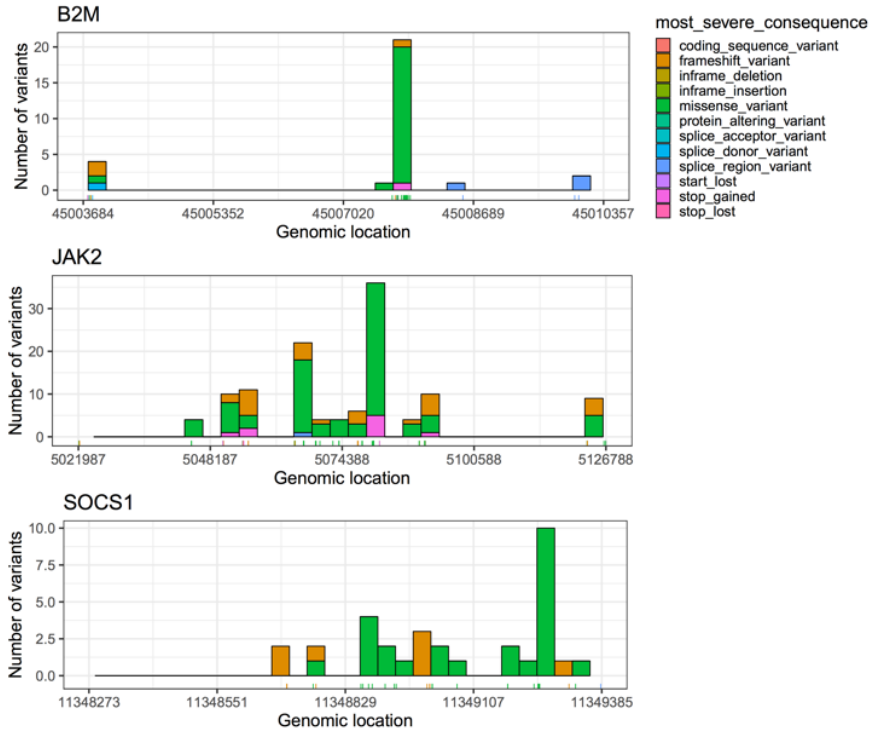

Figure 13. Total Number of Variants Detected in cHL Cases

Chapter IV.

Discussion

This study evaluated the performance of an AMP-based NGS panel and its ability to derive the mutational signature of cHL in the genes *B2M*, *JAK2*, and *SOCS1*. The key result of this study is that deep NGS sequencing of AMP libraries enables mutational profiling of cHL in three genes targeted. The method evaluated here in 28 PMBL, 6 cHL cases and derived results offer a direct genotyping approach that can overcome several significant technical obstacles that have thus far limited the ability of molecular diagnostic laboratories to directly obtain genetic information from HRS cells in cHL. This work shows it is possible to reliably detect and classify variants in the targeted genes including problematic coding regions of *SOCS1*. Despite multiple challenges presented by the disease characteristics of cHL, the technical, analytical and diagnostic performance of this assay argue for adoption of this method in the clinical setting.

DNA from two blood and three FFPE cases were selected and processed to confirm the assay was correctly designed and that sequencing coverage mapped to the desired genomic areas correctly. Due to the rarity and limited availability of excess tissue deriving from cHL cases, DNA extracted from PMBL cases was used to supplement the cohort evaluated for this study. PMBL shares common genetic mutation signatures to cHL and both diseases share several pathobiological features. The sequencing analysis of the PMBL and cHL cohorts established that deep NGS sequencing with this AMP-based

panel established an LOD of somatic mutations in *B2M*, *JAK2*, and *SOCS1* at an allelic fraction of 5%. A lower LOD may have been established if a larger number of PMBL and/or cHL cases could have been obtained, in addition to ensuring that FFPE biopsies only be included if they were created relatively recently, within the past year. This would maximize the availability of DNA and provide DNA with a larger starting fragment size; these two characteristics being the most ideal for NGS library preparation. Given the rarity of the two diagnoses being sought, this was not a possibility - however reasonable feasibility for the assay was established with the cases tested. In addition to this first limitation of having a sparse amount of cases available to evaluate, a second limitation to this study was that the extraction of tissue from biopsy slides was done manually as opposed to the HRS cells being laser micro-dissected. Laser microdissection can be more precise as it can isolate the HRS cells while minimizing the contaminating DNA from surrounding non-tumor cells. This is an expensive and time-consuming workflow which is not suitable for many clinical laboratory operations. However, deep NGS sequencing combined with the molecular genotyping technique overcame the obstacle of sequencing the GC-rich areas of *SOCS1* to facilitate variant calling at an acceptable limit of detection without the use of laser microdissection.

The AMP-based panel used in this study was designed with prognostically-significant genes in mind and focused on those genes which have previously been difficult to genotype due to GC-rich content. This is particularly important when seeking to detect mutations in *SOCS1* since while the gene is relatively small in terms of its coding regions the high GC-rich nature of the gene resists traditional genotyping techniques. While the panel content was limited (*B2M*, *JAK2* and *SOCS1*), the aim was to

establish a minimum level of deep sequencing for variants to be detected. While additional genes are known to be relevant in the stratification of cHL patients for treatment purposes, they do not share the limiting technical barrier of GC-content as compared to *B2M*, *JAK2* and *SOCS1*. Given that more genes would be relatively easier to genotype it would be possible to supplement the panel designed here with those additional genes and utilize AMP with NGS to offer a clinical grade-assay capable of providing full mutation characterization of cHL for robust biomarker detection. The cost of sequencing using NGS methods has dropped considerably over the past decade. However, many disease cohorts (e.g., cHL) have not had sustainable genotyping panels established. In many cases, medical centers will offer very large genotyping panels consisting of hundreds of genes or will sequence the entire exome to provide full coverage, but as shown, this will not necessarily have any regard for a minimum sequencing coverage required over disease-relevant genes. An expanded sequencing capacity over many genes – revealing mutations in genes of both acknowledged significance and unknown significance - can be helpful in some instances if consequential mutations remain to be identified for a condition. With this information, retrospective analyses can be performed to determine the biological significance of a mutation when stratifying its frequency with various disease outcomes. For cHL however, given the current knowledge of which genes are likely implicated diagnostically and prognostically, a robust genotyping panel focusing on the JAK/STAT pathway would be very powerful. Furthermore, a limited panel consisting only of those consequential genes will have a relatively small economic burden for laboratory reagent and sequencing costs as opposed to a larger panel. Given the technical feasibility established here for the

sequencing of *B2M*, *JAK2* and *SOCS1*, it is likely that primers for these genes can be combined into a larger AMP-based panel to provide a more comprehensive genotyping snapshot of cHL patients.

The sequencing output achieved by utilizing this AMP-based panel successfully facilitates variant classification for the three targeted genes. This is especially significant in the aim of sequencing the entire coding region of *SOCS1* since this gene currently does not have any known, well-characterized hotspot mutations with which to establish relevant biomarkers to be leveraged for treatment course optimization. The results shown here are contributory towards the goal of moving forward in establishing a clinical-grade genotyping assay for cHL. A minimum of 200X MCC was reached in 100% of cHL cases evaluated, and in 75% of the PMBL cases (Figure 8). A majority of the cHL cases reached 300X MCC over this important exon which is very encouraging considering the goal of including this chief target in a clinical assay for cHL. At this level of sequencing depth, the LOD for a mutation with an allelic fraction of 2% would have a 70% probability of being detected, and the chance of detection reaches a 99.9% probability for mutations with allelic fractions going down to 5% (Figure 11). In comparing this analysis to recently published sequencing results (Reichel et al., 2014) in which the majority of *SOCS1* regions did not reach 200X coverage, the AMP-based panel here offers a significant benefit in the goal of establishing a clinical-grade assay for cHL, inclusive of the important *SOCS1* exon 2 region. The VEP bioinformatics tool classified detected variants based on the genetic consequence of each mutation. Here, the majority of variants detected in cHL cases within *SOCS1* were frameshift and missense mutations which is compatible with results from prior studies (Lennerz *et. al.* 2015).

Despite a positive outlook for cHL patients given the high rates of disease remission after treatment, many obstacles still remain for clinical management including long-term health problems with secondary malignancies and infertility. The current standard of care for cHL, a combination of radiation and chemotherapy, is associated with a 12.3-fold increased risk of developing premature menopause as opposed to receiving radiotherapy alone. Another common long-term problem due to treatment toxicity is permanent amenorrhea which manifests in 51% of women who received eight cycles of BEACOPP, a common chemotherapy regimen (De Bruin, 2008). Due to the characteristic of randomly-distributed mutations particularly in *SOCS1*, the technical limits of existing genotyping methods have resulted in an absence of a comprehensive summary of possible mutations. The lack of inclusive molecular classification required to investigate and stratify cHL patients will be necessary to include this disease on the list of cancers for which this is already possible, to improve long-term health outcomes. Tiacci et al. reported in 2018 that 87% of cHL cases will feature mutations in the JAK/STAT pathway. Of these cases, 59% will harbor *SOCS1* mutations and 33% will feature mutations in *JAK2*. The presented panel offers a reliable genotyping mechanism for the two key genes *SOCS1* and *JAK2* which comprise 92% of mutations in cHL. Additionally, the ability to establish a genotype in a previously uncharacterized disease may offer novel therapeutic options for cHL. Cell lines modeled after the cHL process are notoriously difficult to maintain (Tiacci, 2018) therefore the capability to more accurately construct a cHL-cell line to test various drugs and their efficacies would be greatly beneficial. By identifying the specific mutations and their allelic frequencies, cell lines could be designed to mirror the genetic signature of a patient's disease with which to test the

efficacy of multiple therapies in vitro before one is ultimately selected for a patient. Improved and more complete mutation characterization across *B2M*, *JAK2* and *SOCS1* could be leveraged to better stratify patients in choosing more appropriate therapy courses and potentially mitigate unnecessary toxic effects.

SOCS1 has been implicated in diseases other than lymphomas including Systemic Lupus Erythematosus and diabetes-related conditions like renal oxidative stress injury and vascular problems. Studies published by Ganesh (2016) and Lopez-Sanz (2018) show that *SOCS1* is not only an attractive target to repair therapeutically, but that it is possible to restore its function and diminish the inflammatory effects associated with the gene's dysfunction. Furthermore, a major healthcare breakthrough which recently emerged in the realm of cancer management is the ability to use blood draws, known as liquid biopsies, to detect mutated DNA shed from tumors within the body which effectively provides an early detection system in addition to disease monitoring. It is possible that liquid biopsies may not be effective for cHL due to the inflammatory process that accompanies the condition (Buedts, Lieselot, & Vandenberghe, 2016). However, if it can be established that a liquid biopsy utilizing AMP can reliably detect relevant mutations in cHL, the AMP method established here could be of benefit. This method would allow for monitoring of disease recurrence or relapse, if assessing a patient's historical mutation data with the latest allelic frequency of mutations being detected.

Due to the potential clinical utility of an NGS genotyping panel specific for cHL patients, it is justified to attempt to establish a diagnostic and prognostic panel despite the multiple limitations that exist. Clinical laboratories seeking to reach this objective must consider that genotyping is an expensive pursuit, particularly for diseases whose relevant

genetic targets have a less than ideal sequencing success rate. To justify the establishment of a new genotyping method, a clinical validation must ensure that the method is reliable in its ability to produce consistent results and that the method can provide results which can be clinically reported within a reasonable timeframe. There are multiple technical limitations of establishing a clinical-grade assay for the genetic characterization of cHL. Without enough high-quality DNA from fresh biopsy material sequencing coverage will plateau and limit the ability to provide variant calling over important genomic areas. Despite the average age of 11 years for the PMBL biopsies tested here, a sufficient amount of DNA was extracted from the majority of them and for most cases there was sufficient sequencing coverage over *SOCS1* exon 2. Suboptimal DNA quality due to the prolonged storage time of FFPE samples acquired for analysis likely played a role in the suboptimal sequencing results over *SOCS1* exon 1. A minimum starting DNA fragment size will be required in order to allow the targeted primers to anneal and provide sufficient enrichment of the region. AMP allows for full sequencing coverage of *SOCS1* by utilizing uni-directional primers and the nested PCR enrichment method – this technique can help minimize the gaps in sequencing coverage observed over coding regions despite having smaller than ideal DNA fragment sizes with which to generate NGS libraries. It is expected that not all biopsies will be successfully enriched with this technique as seen in two of the PMBL cases. An additional limitation is that for any cancer-related genotyping assay, it is unlikely that the tissue being tested will contain 100% malignant tumor cells. Therefore, there is always a level of heterogeneity that comes into play – downstream results will contain sequencing reads representing the tumor cell population while a different population of reads will represent non-tumor cells.

This is a particularly significant challenge to the molecular characterization of cHL due to the scarceness of HRS tumor cells. The last obstacle in the goal of genotyping cHL biopsies reliably is the high GC content of *SOCS1*. The optimized AMP library preparation method was able to overcome this barrier with the ability to leverage the unidirectional primer approach; one genomic region can be targeted by multiple primers, designed to anneal within a few bases of each other and increasing the chances of successful enrichment. This technique would be limited by a sample with insufficient starting DNA template input.

This work assessed the performance of an AMP-based NGS assay in overcoming technical hurdles that presently impede the establishment of a reliable genotyping assay for cHL. Sequencing coverage was achieved for 28 PMBL and 6 cHL cases over three GC-rich, diagnostically-relevant genes from FFPE-derived DNA of variable quality. Mutations were adequately detected with a limit of detection established at 5% allelic fraction despite the lack of HRS tumor cells. In summary, the performance of the NGS method evaluated here supports the inclusion of the three targeted genes, *B2M*, *JAK2*, and *SOCS1* into a clinical-grade laboratory assay for the direct molecular characterization of cHL.

Appendices

Appendix 1. Sample Attributes of PMBL, cHL and Feasibility Cases

Case	Type	Age of sample prior to DNA extraction (years)	DNA (ng/uL)	Status
1	**Feasibility 1 - blood**	1	52.6	Sequenced for analysis
2	**Feasibility 2 - blood**	1	111.5	Sequenced for analysis
3	**Feasibility 3 - FFPE**	1	24.5	Sequenced for analysis
4	**Feasibility 4 - FFPE**	1	66.7	Sequenced for analysis
5	**Feasibility 5 - FFPE**	1	87.3	Sequenced for analysis
6	**PMBL 1**	18	11.0	Sequenced for analysis
7	**PMBL 2**	16	7.4	Sequenced for analysis
8	**PMBL 3**	16	10.8	Sequenced for analysis
9	**PMBL 4**	7	5.7	Sequenced for analysis
10	**PMBL 5**	6	10.2	Sequenced for analysis
11	**PMBL 6**	6	36.1	Sequenced for analysis
12	**PMBL 7**	13	13.4	Sequenced for analysis
13	**PMBL 8**	13	11.6	Sequenced for analysis
14	**PMBL 9**	12	15.0	Sequenced for analysis
15	**PMBL 10**	13	11.0	Sequenced for analysis
16	**PMBL 11**	14	8.8	Sequenced for analysis
17	**PMBL 12**	15	14.4	Sequenced for analysis
18	**PMBL 13**	11	4.9	Sequenced for analysis
19	**PMBL 14**	8	5.4	Sequenced for analysis
20	**PMBL 15**	6	3.3	Sequenced for analysis
21	**PMBL 16**	15	12.2	Sequenced for analysis
22	**PMBL 17**	3	310.2	Sequenced for analysis
23	**PMBL 18**	2	41.4	Sequenced for analysis
24	**PMBL 19**	1	33.7	Sequenced for analysis
25	**PMBL 20**	3	156.0	Sequenced for analysis
26	**PMBL 21**	5	263.6	Sequenced for analysis
27	**PMBL 22**	11	30.0	Sequenced for analysis
28	**PMBL 23**	11	38.5	Sequenced for analysis
29	**PMBL 24**	11	21.5	Sequenced for analysis
30	**PMBL 25**	11	281.4	Sequenced for analysis
31	**PMBL 26**	11	86.1	Sequenced for analysis
32	**PMBL 27**	10	36.8	Sequenced for analysis
33	**PMBL 28**	10	128.8	Sequenced for analysis

34	**PMBL 29**	22	3.6	Excluded (low library yield)
35	**PMBL 30**	13	371.0	Excluded (low library yield)
36	**PMBL 31**	16	0.3	Excluded (low DNA yield)
37	**PMBL 32**	16	0.7	Excluded (low DNA yield)
38	**PMBL 33**	15	1.1	Excluded (low DNA yield)
39	**PMBL 34**	15	1.0	Excluded (low DNA yield)
40	**PMBL 35**	16	1.9	Excluded (low DNA yield)
41	**PMBL 36**	16	2.5	Excluded (low DNA yield)
42	**PMBL 37**	8	1.6	Excluded (low DNA yield)
43	**PMBL 38**	7	2.1	Excluded (low DNA yield)
44	**PMBL 39**	12	0.3	Excluded (low DNA yield)
45	**PMBL 40**	12	2.2	Excluded (low DNA yield)
46	**PMBL 41**	11	3.3	Excluded (low DNA yield)
47	**PMBL 42**	14	2.2	Excluded (low DNA yield)
48	**PMBL 43**	14	0.6	Excluded (low DNA yield)
49	**PMBL 44**	15	0.5	Excluded (low DNA yield)
50	**PMBL 45**	11	2.2	Excluded (low DNA yield)
51	**cHL 1**	2	242.4	Sequenced for analysis
52	**cHL 2**	1	8.3	Sequenced for analysis
53	**cHL 3**	2	90.5	Sequenced for analysis
54	**cHL 4**	2	156.0	Sequenced for analysis
55	**cHL 5**	1	20.4	Sequenced for analysis
56	**cHL 6**	2	47.7	Sequenced for analysis

Abbreviations: PMBL, Primary Mediastinal B-cell Lymphoma; cHL, classical Hodgkin Lymphoma; DNA quantitated in nanograms per microliter

Appendix 2. Total Number of All Variants Detected in All Sequenced Cases

Case	Total Variants Detected		
	B2M	***JAK2***	***SOCS1***
PMBL 1	7	74	14
PMBL 2	21	152	16
PMBL 3	19	108	14
PMBL 4	3	88	17
PMBL 5	39	156	12
PMBL 6	22	129	22
PMBL 7	19	129	22
PMBL 8	10	142	3
PMBL 9	18	81	11
PMBL 10	19	148	12
PMBL 11	15	97	23
PMBL 12	17	83	17
PMBL 13	9	38	31
PMBL 14	19	120	16
PMBL 15	15	75	19
PMBL 16	20	125	16
PMBL 21	5	74	3
PMBL 17	19	114	8
PMBL 18	25	156	9
PMBL 19	16	94	13
PMBL 20	19	121	20
PMBL 24	13	122	30
PMBL 22	12	153	30
PMBL 23	6	39	9
PMBL 25	20	103	24
PMBL 26	23	99	13
PMBL 27	12	39	9
PMBL 28	6	77	9
cHL 1	8	27	4
cHL 2	2	26	5
cHL 3	5	20	3
cHL 4	11	18	8
cHL 5	2	18	4
cHL 6	3	27	9

Abbreviations: PMBL, Primary Mediastinal B-cell Lymphoma; cHL, classical Hodgkin Lymphoma

Bibliography

Aisenberg, Alan C. (1999) Problems in Hodgkin's Disease Management. *Blood*. 93(3): 761-779.

American Cancer Society. (2018, June) *Cancer Facts & Figures 2018*. Retrieved from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf

Biasoli, Irene., et al. (2012) Fertility in Female Survivors of Hodgkin's Lymphoma. *Revista Brasileira de Hematologia e Hemoterapia.* (34): 48–53.

Broad Institute. (2018) Integrated Genome Viewer [Software]. Available at http://software.broadinstitute.org/software/igv/

Buedts, Lieselot, and Peter Vandenberghe. (2016) Circulating Cell-Free DNA in Hematological Malignancies. *Haematologica* 101(9): 997–999.

Camus, V., et al. (2016) Detection and prognostic value of recurrent exportin 1 mutations in tumor and cell-free circulating DNA of patients with classical Hodgkin lymphoma. *Haematologica* 101(9): 1094-1101. doi:10.3324/haematol.2016.145102

Casper J, Zweig AS, Villarreal C, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Karolchik D, et al. (2018) The UCSC Genome Browser. *Nucleic Acids Res*. (46): D762–D769.

Cibulskis, Kristian, et al. (2013) Sensitive Detection of Somatic Point Mutations in Impure and Heterogeneous Cancer Samples. *Nature Biotechnology*. 31(3): 213–219. doi:10.1038/nbt.2514

Colemen, W. B., & Tsongalis, G. J. Molecular diagnostics: For the clinical laboratorian. New York, NY: Humana Press, 2006. Print.

Durbin, R., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25(15): 2078-2079. doi:10.1093/bioinformatics/btp352

Hudnall, S. D., & Küppers, R. Precision Molecular Pathology of Hodgkin Lymphoma. New York, NY: Springer International Publishing, 2018. Print.

Jennings, Lawrence J., et al. (2017) Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels. *The Journal of Molecular Diagnostics*. 19(3): 341-365.

Lennerz, Jochen K., et al. (2015) Suppressor of Cytokine Signaling 1 Gene Mutation Status as a Prognostic Biomarker in Classical Hodgkin Lymphoma. *Oncotarget.* 6(30): 29097–29110.

Lopez-Sanz, L., Bernal, S., Recio, C., Lazaro, I., Oguiza, A., Melgar, A., . . . Gomez-Guerrero, C. (2018) SOCS1-targeted therapy ameliorates renal and vascular oxidative stress in diabetes via STAT1 and PI3K inhibition. *Laboratory Investigation*. doi:10.1038/s41374-018-0043-6

Malmstroem, L. (2017) Faculty of 1000 evaluation for Analysis of protein-coding genetic variation in 60,706 humans. *Post-publication Peer Review of the Biomedical Literature*. doi:10.3410/f.726646180.793533989

McKenna, A., et al. (2010) The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Research*. 20(9): 1297–1303. doi:10.1101/gr.107524.110

Reichel, J., Chadburn, A., Rubinstein, P. G., Giulino-Roth, L., Tam, W., Liu, Y., . . . Roshal, M. (2014) Flow sorting and exome sequencing reveal the oncogenome of primary Hodgkin and Reed-Sternberg cells. *Blood. 125*(7): 1061-1072. doi:10.1182/blood-2014-11-610436

Schif, B., Lennerz, J. K., Kohler, C. W., Kreuz, M., Bentink, S., Melzner, I., . . . Möller, P. (2012) SOCS1 Mutation Subtypes Predict Divergent Outcomes in Diffuse Large B-Cell Lymphoma (DLBCL) Patients. *Oncotarget.* 4(1): 35-47. doi:10.18632/oncotarget.774

Sukka-Ganesh, B., & Larkin, J. (2016). Therapeutic Potential for Targeting the Suppressor of Cytokine Signaling-1 pathway for the treatment of SLE. *Scandinavian Journal of Immunology.* 84(5): 299–309.

Tiacci, E., et al. (2018) Pervasive mutations of JAK-STAT pathway genes in classical Hodgkin lymphoma. *Blood*, 131(22): 2454-2465. doi:10.1182/blood-2017-11-814913

Weniger, M. A., Melzner, I., Menz, C. K., Wegener, S., Bucur, A. J., Dorsch, K., . . . Möller, P. (2006) Mutations of the tumor suppressor gene SOCS-1 in classical Hodgkin lymphoma are frequent and associated with nuclear phospho-STAT5 accumulation. *Oncogene,* 25(18): 2679-2684. doi:10.1038/sj.onc.1209151

Wilm., et al. (2012) LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*. 40(22): 11189-201.

Yang, X., Liu, D., Liu, F., Wu, J., Zou, J., Xiao, X., . . . Zhu, B. (2013) HTQC: A fast quality control toolkit for Illumina sequencing data. *BMC Bioinformatics*. 14(33). doi:10.1186/1471-2105-14-33

Zhang, J., Li, H., Yu, J., Wang, S. E., & Ren, X. (2012) Role of SOCS1 in tumor progression and therapeutic application. *International Journal of Cancer*. 130(9): 1971-1980. doi:10.1002/ijc.27318

Zheng, Zongli., et al. (2014) Anchored multiplex PCR for targeted next-generation sequencing. *Nature Medicine*. 20(12): 1479–1484. doi:10.1038/nm.3729