A Tale of Two Validations: Molecular Diagnostic Assays for the Detection of Clinically Significant Alterations

Ayako McGregor

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2019

Committee:

Christina Lockwood

Andrew Hoofnagle

Eric Konnick

Noah Hoffman

Program Authorized to Offer Degree:

Laboratory Medicine

University of Washington

Abstract

A Tale of Two Validations: Molecular Diagnostic Assays for the Detection of Clinically Significant Alterations

Ayako McGregor

Chair of the Supervisory Committee:
Christina Lockwood
Laboratory Medicine

Laboratory developed tests (LDTs) have become an integral part of laboratory medicine with applications to a variety of complex genetic testing and are at the frontlines of clinical medicine. Proper validation and implementation of LDTs is essential to ensure accurate and high quality results to aid in not only the diagnostic work-up but also for prognostic information and identification of therapeutic options for patients. With numerous techniques and technologies applied to LDTs, it is important to understand which ones are the best methods to apply to a given clinical problem. To best learn this, two validations were performed for two different genetic assays: first a real-time PCR SNP genotyping assay for Apolipoprotein L1 (*ApoL1*) variant alleles (G1 and G2 alleles) to assess the risk of developing chronic kidney disease (CKD) and secondly a comprehensive next generation sequencing (NGS) assay, OncoPlex version 6 (OncoPlex v6), to detect somatic mutations through an expanded panel of 340 genes.

Both validations proved excellent performance characteristics. The accuracy of the real-time PCR genotyping assay had an average of 99.4% and a perfect inter-run and intra-run reproducibility of 100%. Using a simple real-time PCR methodology could potentially determine the patient's risk for CKD before they progress to the advance stages of the disease. OncoPlex v6 demonstrated exceptional performance characteristics with approximately a 2.5-fold increase in average sequencing coverage and a 3.7-fold increase in percent on-target sequencing. The accuracy and precision of OncoPlex v6 across all classes of alterations was above 97% and 92% respectively, with the majority of variants not identified either present at low variant allele fractions or demonstrating low-level copy gains or losses. In addition to excellent performance characteristics, the modular capture chemistry of OncoPlex v6 demonstrates improvements over prior versions by reduced capture cost, as well as superior sequencing quality and faster turnaround time. Both assays are capable of providing critical information that can impact prognosis, diagnosis, and treatment options.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I could not have completed this undertaking without the support of my family, friends and everyone I have met through the Department of Laboratory Medicine. I'd like to express my appreciation particularly to the following:

First to my mentor, Dr. Tina Lockwood, for her continuous motivation and encouragement throughout my graduate school career. She gave me the confidence to become a better scientist and always pointed me in the right direction to the finish.

I am very fortunate to have been given the opportunity to work with the amazing members of the Genetics and Solid Tumors Laboratory. Special thanks to the research scientists, Mallory Beightol and Jennifer Hempelmann for all of the training and assistance they provided me.

Additionally, I would like to thank Dr. Josh Lieberman for his support in guiding me through my first research project as a graduate student, and Dr. Vera Paulson for all of her guidance and patience throughout the validation.

I would also like to thank my thesis committee, Drs. Eric Konnick, Andy Hoofnagle and Noah Hoffman for their time and valuable input to the thesis.

And last but not least a huge thanks to my husband, Frank, for supporting me through the entire journey.

INTRODUCTION

Genetic testing methods have become increasingly important and a preferred method of testing due to their varied and powerful applications in prevention, detection, diagnosis, prognosis, and treatment of various diseases. In oncology, the concept of "personalized medicine" has developed, where certain genetic characteristics help in the diagnosis and prognosis of the cancer and tumor type, and possible therapeutic options for a specific patient. Within the last decade the number of genetic tests has increased dramatically and there are now estimated to be greater than 55,000 tests for over 11,000 conditions both common and rare [1]. Although there are still numerous technical, clinical, operational, reimbursement, and ethical issues that surround genetic testing, it is without a doubt making an enormous impact in the health care industry.

Research and development of new and improved testing methods is an essential aspect of laboratories that offer clinical tests for patients. As discoveries are made and the findings are published and featured in the media, the clinical need for the testing rises. There are numerous different types of genetic testing available that include: 1) biochemical testing where the amount or activity level of proteins indicates changes in DNA, 2) chromosomal testing where the test interrogates long sections of DNA or detects whole chromosome changes, and lastly 3) molecular testing where the analyte is a single gene, short length of DNA, or a single base pair (bp) [2]. Whichever type of testing a laboratory chooses to conduct, to be able to offer it as a clinical assay, the methodology must go through a validation process to meet all regulatory requirements.

Validation is a process done to ensure that the assay will perform as expected and achieve the intended results. Most molecular diagnostic testing methods currently in clinical use for genetic testing and molecular oncology applications are Laboratory Developed Tests (LDTs) and are not FDA approved or cleared. The Clinical Laboratory Improvement Amendments (CLIA) requires LDTs to demonstrate the same performance characteristics as an FDA approved test: precision, accuracy, reportable range, and reference range. The verification process for laboratories that deploy FDA cleared tests should consist of a smaller number of evaluation experiments because the manufacturer has already established the performance characteristics. However, for LDTs all performance characteristics must be established internally and thus CLIA additionally requires LDTs to establish analytic sensitivity and specificity [3]. Molecular diagnostic testing encompasses different technologies including polymerase chain reaction (PCR), fluorescent in situ hybridization (FISH), microarrays, Sanger sequencing, and next generation sequencing (NGS). The specific validation approaches will vary between these different methodologies in order to meet the regulatory requirements. Table 0.1 shows the general requirements for validating molecular assays.

The validation of two different molecular assays, using different analytical approaches, is the focus of this thesis. The first validation is a real-time PCR assay utilized for genotyping risk allele status for the *ApoL1* gene. Secondly, the thesis describes the validation of a comprehensive NGS panel for the detection of 340 somatic cancer mutations.

Table 0.1 – Molecular Assay Validation Checklist [4], [5]

Checklist	Description
Clinical verification	Detailed description of the test purpose and how the test will be implemented in clinical practice. Comparison to current methods/standards, and correlation to disease.
Rationale for target gene(s)	Expert review of literature, gene-disease association
Acceptable samples	Information about sample type, preservation, rejection criteria, and storage requirements.
Testing design	Details about the panel, primers/probes/oligonucleotides utilized in the test.
Testing Method	Detailed step by step procedure with sufficient detail to reproduce the assay.
Detection method	Description of the molecular techniques used in the test.
Reagents/instruments	List of all components including consumables.
Bioinformatics	Comprehensive description to include all algorithms, software, scripts, reference sequences, and databases, whether in-house developed, or open source or vendor supplied.
Validation sample description	Description of sample (both subjects and controls) source and reason for using it in the validation.
Optimization and familiarization	Control experiments, determining correct conditions (e.g. sample input, incubation times, primer/probe concentrations, temperatures, balancing) and initial characteristics.
Performance characteristics	Information about assay sensitivity and reproducibility determined during the validation.
Validation results	Complete description of the results obtained including the unexpected/discordant results.
Assay acceptance/ rejection	Any actions to be implemented when an analytical process fails to meet acceptance criteria.
Assay limitations	Determined during the validation.
Quality control	Description of how the test quality will be monitored and maintained after clinical implementation, both internal and external proficiency testing.

Chapter 1. APOLIPOPROTEIN L1 & CHRONIC KIDNEY DISEASE

1.1 BACKGROUND

Molecular testing methods have grown substantially in the last decade as the ability to sequence the genome is continuously improving and sequencing expectations rise with NGS. Despite the spread and success of NGS methods for detection of molecular alterations, non-NGS methods still have important diagnostic power. Conventional molecular techniques such as PCR are accurate, cost-effective, and widely used.

PCR is a well-established, rapid, and sensitive technique that exponentially amplifies a single copy of DNA into millions of new copies. Real-time PCR, also called quantitative PCR (qPCR), monitors the amplification of the DNA target as it is being generated, which is an extremely useful tool in targeted assays. Real-time PCR allows for robust comparison of a particular gene or messenger RNA (mRNA) transcript abundance in a simple, reliable, and cost-effective way. Numerous applications exist for real-time PCR; here I describe its use for genotyping several single nucleotide polymorphisms (SNPs).

SNPs are the most common type of genetic variation; they represent a single nucleotide change in a DNA sequence that occur once in every 1000 nucleotides on average. SNPs can be unique to a specific population, or common to many individuals across multiple populations. The vast majority of SNPs do not have any clinical significance; however, some act as biological markers for certain genes associated with complex diseases such as heart or kidney disease [6].

Chronic Kidney Disease (CKD) is a prevalent public health problem worldwide. In the United States, approximately 14% of the general population falls in the spectrum of CKD but 96% of those who have early stages of CKD do not know that they have it [7]. CKD often has no symptoms until advanced stages and individuals are at higher risk for multiple additional diseases

including diabetes and cardiovascular disease. This increased risk is hypothesized to be due to metabolic disturbance, endothelial cell dysfunction, and an altered HDL composition and function originating from the impaired kidney function [8]. Compared to Caucasians, African-American populations have twice the risk of developing end-stage renal disease. The identification of the underlying cause of the difference in developing end-stage renal disease from CKD between the African-American and Caucasian populations is essential to reduce the burden and complications in African-American patients [9].

1.2 APOLIPOPROTEIN L1

Apolipoprotein L1 (ApoL1) is part of a relatively new class of lipoproteins that is encoded by the *ApoL1* gene. The protein is primarily created in the liver and when it is in the plasma it circulates on high-density lipoprotein (HDL) particles. There is on-going research about how CKD affects the composition of HDL and it is now known that genetic variants in *ApoL1* gene is common in the African-American population but rare in the Caucasian population [9]. Due to evolutionary pressure, human populations have maintained a number of variant alleles that in a heterozygous state confer an evolutionary advantage but may confer a disadvantage in the homozygous state. For example, sickle cell disease is a well-studied hemoglobinopathy associated with genetic resistance to malarial infection. Individuals with sickle cell trait (heterozygous for the sickle cell variant) generally exhibit no signs of sickle cell disease and are protected from malaria; however, those who are homozygous for the sickle cell variant, although still protected from malaria, will develop sickle cell disease [10]. A similar evolutionary advantage exists with variants in the *ApoL1* gene that confer resistance to trypanosomiasis in a heterozygous state, but increase risk of CKD in the homozygous state.

The *ApoL1* gene is presumed to play a role in innate immunity and is also known to be responsible for human resistance to *Trypanosoma brucei*. Trypanosomes possess a serum resistance-associated (*SRA*) gene that creates a variable surface glycoprotein (VSG). Trypanosomes are able to escape lysis by inactivating *ApoL1* through the trypanosome VSG binding an *ApoL1 SRA*-interacting region. There are two trypanosome subspecies, *T. b. rhodesiense* and *T. b. gambiense*, that express *SRA* and therefore produce VSG, which is believed to have evolved to develop resistance to *ApoL1* and is ultimately responsible for causing human African sleeping sickness across sub-Saharan Africa [11].

There are two common variant alleles in *ApoL1* associated with CKD risk that are commonly known as the G1 allele and G2 allele (Table 1.1 and Figure 1.1). *ApoL1* G1 has two missense variants (p.S358G and p.I400M) and *ApoL1* G2 has a six base pair deletion that removes two amino acids (p.N388_Y389del) [11]. The reference *ApoL1* allele is known as the G0 allele. The reference and variant DNA and protein sequences are listed in Appendix A and B.

Table 1.1 – *ApoL1* Variant Alleles Associated with CKD

	G1		G2
NCBI Transcript	NM003661.3	NM003661.3	NM_003661.3
c.	c.1024A>G	c.1152T>G	c.1164_1169delTTATAA
p.	p.S358G	p.I400M	p.N388_Y389del
dbSNP Identifier	rs73885319	rs60910145	rs71785313
Genomic Coordinates (GrCh38)	Chr22: 36265860	Chr22: 36265988	Chr22: 36266000-36266006

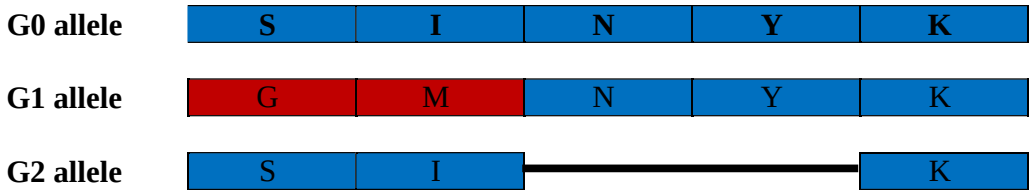

Figure 1.1 – *ApoL1* Reference G0 Allele and Variant G1 and G2 Alleles (modified from [12]). The rectangles in red indicate a change in amino acids and the black horizontal line indicates a deletion of amino acids.

Frequency of the *ApoL1* variant G1 and G2 alleles from the 1000 Genome public database is summarized in Table 1.2 [13]. These alleles are predominantly seen in African and Afro-descendant populations because of the evolutionary advantage of inhibiting the parasite infection and they are responsible for overriding the *SRA* resistance in *T. b. rhodesiense* and *T. b. gambiense.* A comprehensive description of the biological mechanism of the trypanolytic process will not be discussed here. In brief, Trypanosomes endocytose HDL particles, including the attached ApoL1 protein, in the bloodstream of the infected individual. Once ApoL1 reaches the lysosome, the endosome pH becomes acidic, allowing for ApoL1 to transition into the inner leaflet of the lysosome, insert into the membrane, and form a pore leading to trypanolysis [11].

Table 1.2 – 1000 Genome Population Frequency for *ApoL1* Variant Alleles [13]

Population	Total	G1		G2
		A>G	T>G	delTTATAA
Global	5008	7.0%	6.9%	2.5%
African	1322	25.9%	25.9%	13.0%
East Asian	1008	0.0%	0.0%	0.0%
European	1006	0.0%	0.0%	0.0%
South Asian	978	0.0%	0.0%	0.0%
American	694	1.0%	1.0%	1.0%

Much of the increased CKD risk in Afro-descendant populations is explained by these two *ApoL1* allele variants. Similar to the relationship between sickle cell disease and malaria, having

a single *ApoL1* risk allele confers an advantage for the individual with protection from trypanosomiasis without risk for CKD. However, those with two *ApoL1* risk alleles have an increased risk for CKD although they are still protected from trypanosomiasis. *ApoL1* alleles for CKD risk are therefore inherited in an autosomal recessive pattern and the *ApoL1* high risk genotypes for CKD are defined as two risk alleles in any combination (homozygous G1/G1, homozygous G2/G2, or compound heterozygous G1/G2). Individuals with one of these three high risk genotypes have an increased risk of developing CKD, hypertension, focal segmental glomerulosclerosis, and HIV-associated nephropathy [8].

Since CKD is typically not diagnosed until advanced stages of the disease, early detection and prevention methods are highly desired. It would be very beneficial to have a molecular assay capable of determining the *ApoL1* genotype and potentially detecting those at high risk prior to development of CKD. In order to accomplish this goal, Josh Lieberman and I designed a new molecular assay using TaqMan SNP genotyping by real-time PCR to rapidly identify the *ApoL1* G0, G1, and G2 genotypes.

1.2.1 *Subject Demographics*

There are many cohorts of patients with CKD that are being studied to better understand the disease's natural history, one of which is the Seattle Kidney Study (SKS). Samples from this cohort have been previously used to study whether *ApoL1* risk alleles can be inferred from the circulating protein. Determining *ApoL1* genotype was of interest because the fraction of ApoL1 circulating in HDL has previously been shown to correlate with kidney function [14]. The characteristics of the subjects from the cohort are shown in Table 1.3. One-hundred four (104) African-American participants in the SKS were selected to perform TaqMan SNP Genotyping Real-Time PCR to determine their *ApoL1* G0, G1, and G2 genotype. These results were

compared to their associated proteomic data obtained by liquid chromatography mass spectrometry (LC-MS) performed at the Hoofnagle Lab at the University of Washington (UW) Medical Center to look for any association with circulating ApoL1.

Table 1.3 – Relevant Clinical Covariates

Clinical Covariate	Entire SKS cohort, n=507
Age, years	58.0 ± 13.9
Female	166 (32.7%)
White	334 (65.9%)
Black	122 (24.1%)
Other	51 (10.1%)
BMI	31.4 ± 7.9
Prevalent diabetes	254 (50.4%)
Current smoker	98 (19.8%)
Statin use	276 (54.4%)
CRP	2.1 mg/L (0.7, 5.8)
Estimated GFR	45.0 ± 26.0
Dialysis	
< 15 mL/min	42 (8.3%)
15-30 mL/min	133 (26.2%)
30-45 mL/min	117 (23.1%)
45-60 mL/min	92 (18.1%)
60+ mL/min	123 (24.3%)
HDL	41.6 ± 17.1 mg/dL
Triglyceride	162.2 ± 118.7 mg/dL
Total cholesterol	176.7 ± 59.6 mg/dL
Urine ACR	111.5 µg/mg (13.5, 750.8)

1.2.2 *TaqMan SNP Genotyping Assay*

The TaqMan SNP Genotyping Assay can determine targeted SNP alleles by amplifying and detecting specific alleles. The assay consists of a forward and reverse PCR primer to amplify the sequence of interest and two TaqMan Minor Groove Binder (MGB) probes to detect allele 1, allele 2, or both through fluorescent signals. Each probe is made up of target-specific oligonucleotides and is fluorescently labeled on the 5’ end with either VIC dye (G0 allele) or

FAM dye (G1 or G2 allele). Each probe also contains a non-fluorescent quencher (NFQ) at the 3' end (Figure 1.2) [15].

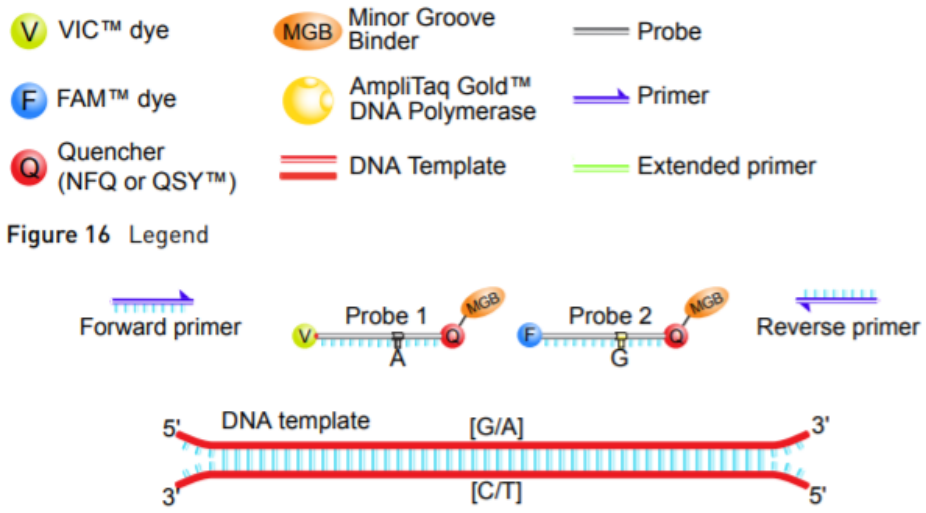

Figure 1.2 – TaqMan SNP Genotyping Components [15]

The G0 and either G1 or G2 polymorphisms of *ApoL1* were detected using two TaqMan allele-specific PCR assays, G0/G1 and G0/G2, respectively. The primer and probe specifications of the assay are listed in Table 1.4. When the probe is intact, the quencher dye suppresses the fluorescence of the reporter dyes, VIC and FAM. The real-time PCR is prepared by mixing genomic DNA with TaqMan Genotyping Master Mix (which contains AmpliTaq Gold DNA Polymerase UP (Ultra Pure), dNTPs without dUTP, ROX reference dye, and optimized mix components that helps with specificity for discrimination between alleles), the two PCR primers, and the two probes (either G0/G1 or G0/G2). If the genomic DNA contains the sequence of interest, the TaqMan probe anneals to its complementary strand on the target sequence. The DNA polymerase cleaves the probe to separate the reporter and quencher dye and ultimately release a fluorescent signal corresponding to VIC, FAM, or both. Increasing fluorescent signal indicates the allele of interest is present in the genomic DNA and is being amplified by PCR.

Table 1.4 – ApoL1 Primer Summary

	G1 Assay	G2 Assay
Primers	TaqMan Design (#4351379)	Custom Design
Forward primer	*	GCTCAGGAGCTGGAGGAGAA
Reverse primer	*	CCTGCCCTGTGGTCACA
Allele 1	G0	G0
Probe Seq	*	CCTGCAGAATCTTATAATT
Reporter Dye	VIC	VIC
Quencher	NFQ	NFQ
Allele 2	G1	G2
Probe Seq	*	CCTGCAGAATCTTATTG
Reporter Dye	FAM	FAM
Quencher	NFQ	NFQ

*Sequence information cannot be disclosed per Thermofisher. The context sequence is CAAGCTCACGGATGTGGCCCCTGTA[A/G]GCTTCTTTCTTGTGCTGGATGTAGT

1.2.3 *Methods*

Genotyping was attempted on 104 African-American subjects from the SKS cohort. Of these 104 subjects, 97 were successfully genotyped. Three samples were rejected due to insufficient DNA volume or concentration and four samples had indeterminate real-time PCR results, as discussed in detail in section 1.3 Genotyping Results and 1.4 Discussion.

A volume of 2 µL genomic DNA (3 ng/µL) was mixed with 12.5 µL 2X TaqMan Master Mix, 1.25 µL 20X Primer/Probe Assay Mix, and 9.25 µL of molecular-grade H_2O in a MicroAmp Optical 96-well reaction plate (Life Technologies Corporation, Carlsbad CA) (Table 1.5). PCR was performed on a ViiA 7 Real-Time PCR System (Applied Biosystems, Foster City CA) for 40 cycles including a pre-PCR read, 15 seconds denaturing at 95 °C, 30 seconds annealing at 62 °C, 30 seconds extension at 60 °C, and a post-PCR plate fluorescence acquisition.

Table 1.5 – TaqMan PCR Components per Single Reaction

Component	Volume (µL)
2X TaqMan Master Mix	12.5
20X SNP Genotyping Assay Mix	1.25
Nuclease-free H2O	9.25
DNA	2
Total Volume	25

Every real-time PCR run included three HapMap controls of known homozygous G0/G0, heterozygous G0/G1, and heterozygous G0/G2 genotypes (Coriell Institute, Camden NJ) from cell lines NA12878, NA20362, and NA20348, respectively. Both genotyping assays (G0/G1 and G0/G2) were evaluated for every sample with each sample tested at least in duplicate.

1.2.4 *Data Analysis*

The measured value from these end-point experiments is the normalized reported dye fluorescence intensity (Rn) value, based on fluorescent signal from each well to determine which alleles are present in each sample [15]. The VIC dye labeled probe corresponded to the G0 allele and the FAM dye labeled probe corresponded to either the G1 or G2 allele, depending on which assay was being run. The Viia7 Software (Applied Biosystems, Foster City CA) automatically assigns a genotype and every sample's delta Rn value (fluorescence of the reporter dye divided by the fluorescence of ROX dye minus the baseline) for each allele is plotted onto an allelic discrimination plot. The software algorithmically clusters the sample data, typically into one of three groups, as illustrated in Figure 1.3 where there are two homozygotes clusters, a heterozygote cluster, and the no template control (NTC)/indeterminate cluster [16].

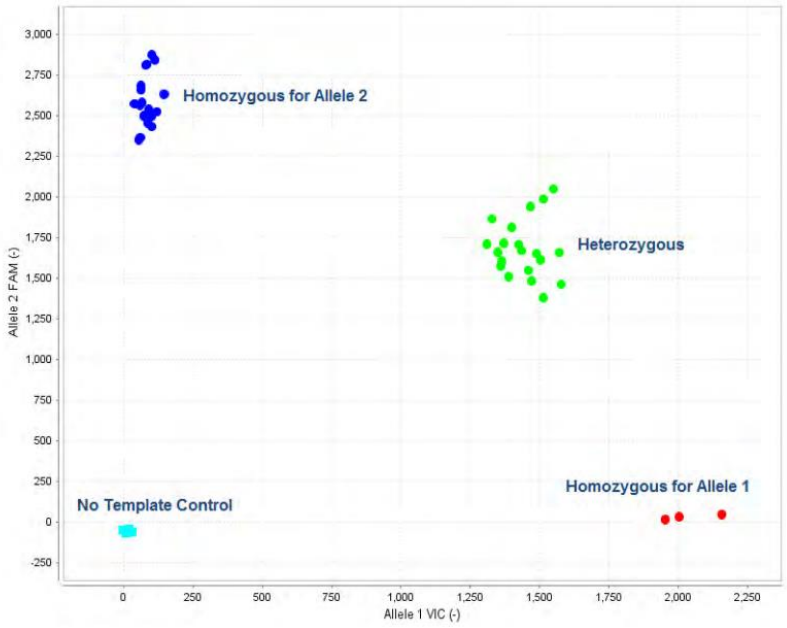

Figure 1.3 – Typical Clusters of TaqMan SNP Genotyping [16]

1.3 RESULTS

Genotyping results are summarized in Table 1.6. At the G1 locus there were 54/101 G0/G0 homozygotes, 33/101 G0/G1 heterozygotes, and 10/101 G1/G1 homozygotes. At the G2 locus there were 75/101 G0/G0 homozygotes, 19/101 G0/G2 heterozygotes, and 4/101 G2/G2 homozygotes. Three samples were indeterminate at both G1 and G2 loci. The allele discrimination plot of each assay is illustrated in Figure 1.4 and Figure 1.5.

Table 1.6 – Genotype Frequency per Locus

	G1 Locus	%	G2 Locus	%
Homozygous G0/G0	54	53.5%	75	74.3%
Heterozygous	33	32.7%	19	18.8%
Homozygous	10	9.9%	4	3.9%
Indeterminate	4	3.9%	3	3.0%
Total	101		101	

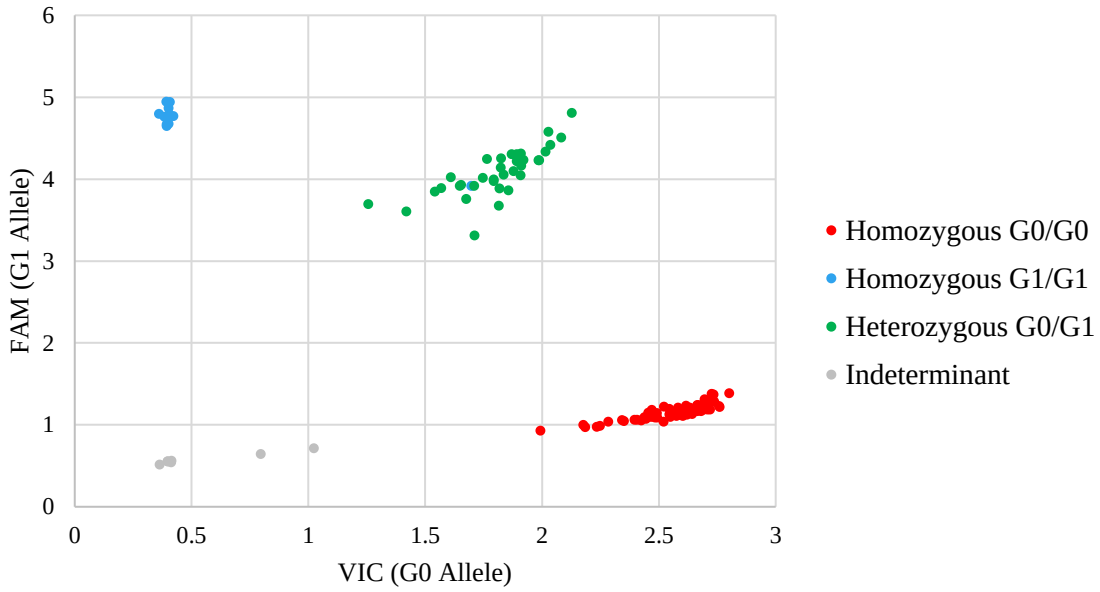

Figure 1.4 – G1 Assay Allele Discrimination Plot

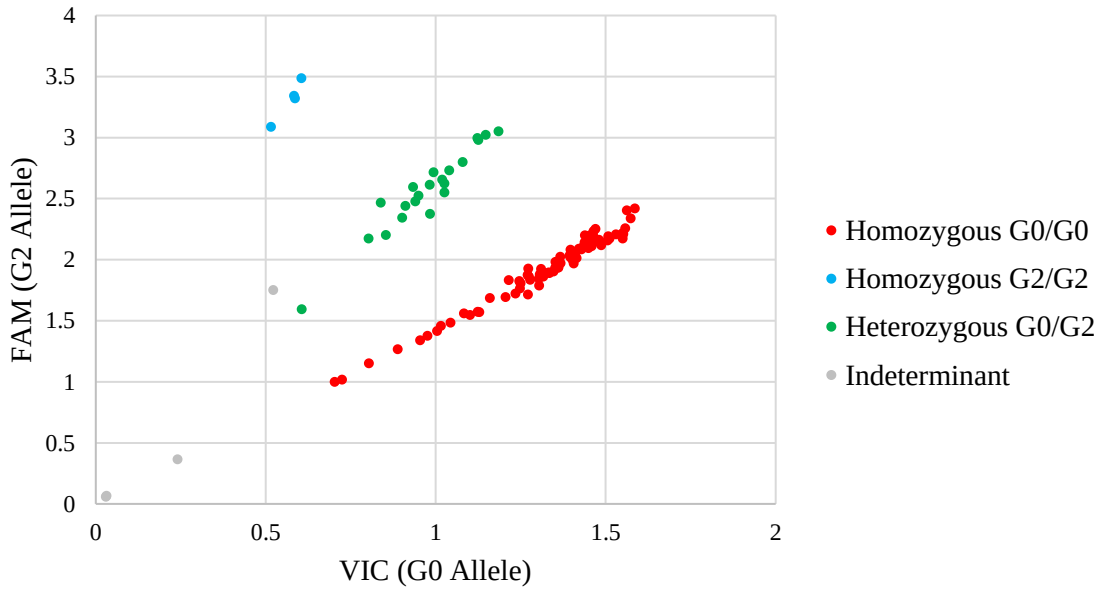

Figure 1.5 – G2 Assay Allele Discrimination Plot

The individual results for all alleles were compared to each other to determine the true genotype of the subjects (Table 1.7). As expected, the reference homozygous G0 is the most common genotype. Next most prevalent are the heterozygous genotypes G0/G1 and G0/G2, followed by the three high-risk genotypes (homozygous G1/G1, homozygous G2/G2, and compound heterozygous G1/G2).

Table 1.7 – Combined *ApoL1* Genotype

Genotype	Count	Percent
G0/G0	37	36.6%
G0/G1	27	26.7%
G1/G1	10	9.9%
G0/G2	13	12.9%
G2/G2	4	4.0%
G1/G2	6	5.9%
UND	4	4.0%
Total	101	100%

In order to assess the qualitative precision of the assay, all subject samples were run in duplicate within each individual experiment. Every experimental assay also included four controls that were processed across seven multiple, separate runs. Assay reproducibility was excellent with 100% concordance for both inter-run and intra-run repeats (Table 1.8 and Table 1.9).

Table 1.8 – Intra-Run Reproducibility

	G1 Locus	G2 Locus
Homozygous G0	54/54 (100%)	75/75 (100%)
Heterozygous	33/33 (100%)	19/19 (100%)
Homozygous	10/10 (100%)	4/4 (100%)
Indeterminate	4/4 (100%)	3/3 (100%)

Table 1.9 – Inter-Run Reproducibility

Control Sample	G1 Locus		G2 Locus	
	Expected	Concordance	Expected	Concordance
NA12878 (G0/G0)	G0/G0	7/7 (100%)	G0/G0	7/7 (100%)
NA20362 (G0/G1)	G0/G1	7/7 (100%)	G0/G0	7/7 (100%)
NA20348 (G0/G2)	G0/G0	7/7 (100%)	G0/G2	7/7 (100%)
NTC	NTC	7/7 (100%)	NTC	7/7 (100%)

There were four indeterminate samples in the G1 assay and three indeterminate samples in the G2 assay (Table 1.10). These samples failed to amplify in both assays as seen with very high Ct values at both alleles with the exception of one sample that was indeterminate on the G1 allele but heterozygous G0/G2 at the G2 allele. Samples 4 and 6 also had a low DNA concentration to begin with.

Table 1.10 – Summary of Indeterminate Samples

A) G1 Assay

Sample	Concentration (ng/µL)	Ct		Delta Rn		Genotype
		Allele1	Allele2	Allele1	Allele2	
4	2	37.45	UND	0.42	0.56	UND
6	4	36.43	35.92	1.02	0.71	UND
37	31	UND	UND	0.41	0.54	UND
55	44	38.34	UND	0.80	0.64	UND

B) G2 Assay

Sample	Concentration (ng/µL)	Ct		Delta Rn		Genotype
		Allele1	Allele2	Allele1	Allele2	
4	2	UND	UND	0.03	0.05	UND
6	4	36.25	37.02	0.30	0.41	UND
37	31	38.95	37.33	0.23	0.33	UND
55	44	31.28	31.29	1.16	1.64	G0/G2

In order to assess the accuracy of the genotyping assay, each subject genotype was compared to the proteomic data obtained from the Hoofnagle Lab using LC-MS (Table 1.11).

The four indeterminate samples listed in Table 1.10 were excluded from the analysis. Overall, the accuracy was excellent across all 6 possible genotypes, with only one discordant sample where the genotype was heterozygous G0/G1 and the phenotype was homozygous G0/G0.

Table 1.11 – Accuracy of TaqMan SNP Genotyping Assay

Result	real-time PCR	LC-MS	Accuracy
G0/G0	37	37	100%
G0/G1	27	26	96.3%
G1/G1	10	10	100%
G0/G2	13	13	100%
G2/G2	4	4	100%
G1/G2	6	6	100%

1.4 DISCUSSION

1.4.1 *Indeterminate and Discordant Results*

Although the DNA input for all samples were diluted to the same concentration of 3ng/µL, the dilution factors were based on concentration values provided by the cohort and was not confirmed internally. The actual DNA concentration may have been lower in the indeterminate samples and thus was below the minimum detection limit and failed to amplify. The limit of detection for the genotyping assay has yet to be established; however, the 3ng/uL used was likely at lower limit due to the consistently high Ct values observed throughout the experiments. The original DNA concentrations were re-measured using the Varioskan Lux Multimode Microplate Reader (ThermoFisher Scientific, Waltham MA) and majority of the samples had a decreased concentration, ranging from 3% to as much as 93% lower in some samples. With such a large variability, an important step that will be required for future experiments is to verify the DNA concentrations, especially when using samples from an outside location.

The reason for the one discordant sample between the real-time PCR and LC-MS method remains unclear as both methodologies evaluated the sample in duplicate and obtained the same, discordant result. Re-review of the raw genotypic data showed that the Rn values were 1.91 for VIC and 4.04 for FAM, both of which are well within the expected range for heterozygous G0/G1 and matching the known heterozygous G0/G1 control sample. The ApoL1 protein is primarily synthesized in the liver and thus patients who receive a liver transplant can potentially have a discordant genotype and phenotype. However in the SKS cohort, subjects who received a transplant is excluded. The samples could have been mislabeled and a repeat sample should be tested on both assays for confirmation.

1.4.2 *Genotype Interpretation Software*

The automatic genotyping function on the Viia7 Software (Applied Biosystems, Foster City CA) was used to assign a genotype to each sample. However, this software only worked when all three clusters were present in the allele discrimination plot. This limitation became apparent when the software from the G2 assay called an abnormally high number of samples that were erroneously interpreted by the software as homozygous G2/G2 and heterozygous G0/G2 with very few of the predominant homozygous G0/G0 genotype. Investigation of the allele discrimination plot per experiment showed that there were only two clear clusters generated.

In order to conclusively determine the genotype a second software algorithm from the probe manufacturer was used, TaqMan Genotyper Software (Applied Biosystems, Foster City CA). This software creates a “study” where multiple genotyping experiments can be overlaid and analyzed together. This was essential for obtaining accurate genotypes since not all of the experiments had a full complement of potential genotypes and a large number of samples was needed for reliable clustering [16].

Using the TaqMan Genotyper Software resolved previously incorrectly assigned genotype calls. However, this method of interpretation for a clinical assay is not acceptable since individual patient results should not be dependent on each other. A future consideration that would bypass this challenge would be to include control samples that are homozygous for G1/G1 and G2/G2 for each respective assay to ensure data points would be present in all potential genotype cluster.

1.4.3 *Workflow Troubleshooting*

After completing data analysis, several improvements were identified that could be incorporated into the wet lab processes in future experiments. First, for all of the experiments performed, the Ct values were very high ranging from 27 to 37 cycles. This suggests that the experimental DNA concentrations used were not optimal and they were lower than the intended 6 ng per reaction. Re-review of the DNA concentrations revealed that 95% of the samples had less than 6 ng used per reaction, ranging from 0.45 ng to 5.81 ng with a median of 2.87 ng. For this validation the balance between preservation of sample and having a straightforward procedure was taken into consideration and 3 ng/µL was chosen to accommodate a few samples in the cohort that had a low concentration. The same concentration was attempted to be used for all experiments to reduce any unexpected variation in the PCR efficiency. Majority of the samples were genotyped accurately even with the minimal DNA concentrations that were used. Re-analysis of the three samples that had an indeterminate genotype at both G1 and G2 loci showed that the true DNA concentration used for the experiments were below 1 ng. In order to offer this as a clinical assay, a range of acceptable DNA concentrations as well as the lower detection limit for genotyping must be established.

Secondly, only three known controls were included per run (G0/G0 homozygous, G0/G1, and G0/G2 heterozygous) for these experiments. Although specificity was assessed by including the heterozygous G0/G1 control in the G2 assay and the heterozygous G0/G2 control in the G1 assay, these were unnecessary. A more useful set of controls would have been to include a G0/G0 homozygote, G0 and either G1 or G2 heterozygote, and G1/G1 or G2/G2 homozygote for each respective assay. This may have helped with a more accurate automatic genotype calling on the Viia7 Software and obviate the need for use of the TaqMan Genotyper Software.

Lastly, when looking at the allelic discrimination plots for both assays, the G2 plot had more trailing clusters. This could be due to a number of reasons including insufficient DNA added to the wells, variable DNA concentration, insufficient mixing, or degraded DNA [16]. The two assays were performed separately, with G1 performed first followed by G2 a few weeks later. The same diluted DNA sample was used for all experiments as an attempt to reduce sample variability; however, by the time the G2 assay was performed the DNA quality may have decreased and thus caused the trailing clusters. Although the genotypes were accurately called, it is preferred as a clinical assay to have tighter clusters to increase the confidence in the genotype call being made.

1.4.4 *Expected versus Observed*

The overall frequency of each genotype from the real-time PCR assay for each locus was compared to a large public database, the 1000 Genome allele frequency data. Of the 101 African-American SKS cohort subjects genotyped, approximately 42% contained the G1 allele and approximately 22% had the G2 allele (Table 1.6). In the 1000 Genome public database, 26% of the African-American population in the study contained the G1 allele and 13% had the G2 allele (Table 1.2). Thus the ratio of G1 to G2 *ApoL1* alleles is 2:1 in this SKS cohort, comparable to the

1000 Genome frequency data. The percentage for each variant allele was higher in the observed SKS cohort, which was expected. This is likely due both to a small sample size as well as sample selection bias since only African-American subjects with CKD from the SKS cohort were used for genotyping *ApoL1*.

1.5 CONCLUSION

Using a simple real-time PCR method, I developed a potential clinical assay for CKD-risk assessment using *ApoL1* genotype status. Although optimization in the wet lab workflow would facilitate higher quality data, the accuracy and precision were excellent in both the *ApoL1* G1 and G2 assays. Future studies should correlate the subject's genotype to their stage of CKD. My hypothesis is that those who are homozygous G0/G0 would have an earlier stage 1 CKD, homozygous G1/G1, G2/G2 or compound heterozygous G1/G2 would have a later stage 4-5 CKD, and the heterozygous G0/G1 or G0/G2 genotypes would be at an intermediate stage.

Chapter 2. UW-ONCOPLEX

2.1 BACKGROUND

Within the last decade following the completion of the Human Genome Project, the ability to sequence a genome has advanced dramatically. Up until 2007 the cost of sequencing a genome closely followed Moore's Law, a prediction that the computing power doubles every two years; however, there was a significant decrease in sequencing cost around 2007 that has continued to decrease year after year [17]. The driving force behind this dramatic decrease in cost is the progression of next generation sequencing (NGS). Also called massively parallel sequencing (MPS), NGS is capable of sequencing millions to billions of short DNA fragments simultaneously and has revolutionized genomic research.

Although whole genomes can be sequenced through NGS, the cost for whole genome sequencing remains prohibitive. Several techniques exist to enrich regions of the genome that are of interest for a specific application. By targeting a limited area of the genome, NGS can be used for "deep sequencing" specific DNA fragments hundreds to thousands of times in order to assess actionable or potentially actionable mutations that otherwise could not have been detected. The ability to interrogate relevant genes in cancer has paralleled the development of treatments that target molecules or pathways that are aberrant in cancer, which is often termed "precision medicine" or "precision oncology". Cancer is a heterogeneous disease with numerous potential genomic alterations and this variability makes NGS an extremely powerful tool to identify and improve therapies tailored for each individual based on their tumor's molecular characteristics.

Many genetic abnormalities are observed in more than one type of cancer. For example, *EGFR, KRAS, ALK* and *ROS1* are recurrently mutated in non-small cell lung cancers (NSCLC), *KIT* and *PDGRFA* mutations are prevalent in gastrointestinal stromal tumors (GISTs), and *BRAF,*

KIT and *NRAS* are commonly mutated in melanomas [18]. However, testing for all of these genes individually is time-consuming, would quickly deplete the limited tissue that is commonly obtained for diagnostic purposes, and further increase healthcare costs. Single gene testing may also miss any novel mutations in genes that are not commonly mutated in a specific cancer or have not been previously identified in cancer tissue. Thus, methods to assess for a panel of genes in a comprehensive manner have been in high demand.

UW-OncoPlex ™ (OncoPlex) is a comprehensive, targeted NGS assay designed to detect somatic mutations in neoplastic tissues for diagnosis, prognosis, and treatment of cancer developed by faculty at University of Washington, the Seattle Cancer Care Alliance, and the Seattle Children's Hospital. Since the test was initially offered clinically in August 2012, there have been five iterations of the assay, with increasing numbers of targeted genes with each version update as more genes become of clinical interest. However, the rapid discovery of clinically significant variants translates to large NGS panels such as OncoPlex to be out-of-date by the time the gene panel is designed, validated, and implemented. Additionally, although the sequencing technology and the bioinformatics pipelines that support all the sequencing data have evolved throughout the seven years of this assay, the method for genomic DNA enrichment for targeted genes of interest has remained static.

Here the validation of OncoPlex version 6 (OncoPlex v6) will be discussed, in which the assay has adopted a modular panel structure capable of rapid redesign and expansion of target genes as well as a completely new chemistry for the targeted enrichment portion of the library preparation. This newly validated version accommodates for the consistent evolution of the molecular oncology field and facilitates the generation and interrogation of higher quality sequencing data more quickly and cost-effectively than in the previous versions of the assay.

2.2 METHODS

2.2.1 *OncoPlex version 6 Panel Design*

The assay uses next-generation "deep" sequencing to detect the following five different classes of cancer mutations: 1. Single Nucleotide Variants (SNVs) in which one base pair is substituted for another base pair, 2. Insertions and Deletions (indels) of varying size, 3. Copy Number Variants (CNVs) where large areas of a gene or a chromosome are amplified or deleted, 4. Structural variants (SVs) including translocations and inversions, and 5. Microsatellite instability (MSI) in which the number of repeating microsatellite units is different from the germline, suggesting a defect in DNA repair.

The genes included in UW-OncoPlex were chosen to meet criteria for one of three tiers corresponding to current clinical significance or 'actionability'. Tier 1 includes genes with mutations that are currently clinically actionable by providing prognostic information or information about sensitivity or resistance to specific targeted therapies. Tier 2 genes are those expected to be actionable in the near future and has ongoing clinical trials for potential targeted therapies. Tier 3 genes are identified to be recurrently mutated in cancers and are supportive for overall diagnosis, but prognostic and therapeutic information is not yet available [19]. Additionally, the panel will be used to assess for the presence of microsatellite instability (MSI) status and determination of the overall tumor mutational burden (TMB), which will be completed in a future validation.

Unlike prior versions of OncoPlex that utilized targeted hybridization-based capture from a custom Agilent SureSelect design (Agilent Technologies, Santa Clara CA), OncoPlex v6 employs a custom modular panel of IDT xGen Lockdown Probes with corresponding reagents (Integrated DNA Technologies, Coralville IA). The new panel consists of 340 genes listed in

Appendix C. Six of these genes include complete intron and exon coverage (*BRCA1, BRCA2, MLH1, MSH2, MSH6,* and *PMS2*), and 22 genes also include select intronic regions recurrently involved in fusions.

2.2.2 *Library Preparation*

After DNA extraction, 250 ng of the purified DNA is sheared to a length of about 250 bp by sonication on a Covaris instrument (Covaris, Woburn, MA). The DNA fragments are purified with AMPure XP beads (Beckman Coulter, Brea CA) and then undergo a series of enzymatic steps: end-repair, addition of a poly-A tail to the 3'end, and ligation to xGen Dual Index Adaptors (Integrated DNA Technologies, Coralville IA) where each sample is barcoded with a unique 8 bp index on each end of the DNA fragment. Ligated DNA is purified by AMPure XP beads and amplified by PCR for 9 cycles to incorporate Illumina P5 and P7 adaptors on a C1000 Touch Thermal Cycler (Bio-Rad Laboratories, Hercules, CA). The amplified DNA with Illumina sequencing adaptors is purified by AMPure XP beads again and quantified. Before undergoing hybridization, the ligated DNA samples are combined or "pooled" at a concentration of 250 ng per sample with xGen Blocking Oligos (Integrated DNA Technologies, Coralville IA). The blocking oligos are universal blockers that are 10 bp in length that bind to the adaptors ligated to the DNA fragments [20]. Since the dual index adaptors ligate to both on-target and off-target DNA sequences within the library, during hybridization some of the adaptors will bind to each other and inevitably capture off-target fragments that are bound to the on-target fragment. The blockers help to reduce the non-specific binding between the adaptors and reduce off-target capture during hybridization by binding to the inverse strand of the adaptors [21]. Human concentration over time (Cot-1) DNA (Integrated DNA Technologies, Coralville IA) is also added into the pool to increase hybridization efficiency. Human Cot-1 DNA is a placental DNA

enriched for repetitive DNA sequences that increase hybridization efficiency by blocking repetitive DNA sequences in the library [22].

This pooled, dual-indexed library is hybridized at 65 °C for 4 to 18 hours to a custom designed panel of xGen Lockdown Probes (Integrated DNA Technologies, Coralville IA). These probes are complementary to the sequences of interest and consist of DNA biotinylated oligonucleotides that target 340 genes plus custom identity SNPs and MSI loci. The hybridized products are purified through Dynabeads M-270 Streptavidin (Integrated DNA Technologies, Coralville IA) and washed with a series of buffers to disrupt non-specific binding. The captured library is amplified by 11 cycles of PCR on a C1000 Touch Thermal Cycler and prepared for sequencing [20].

2.2.3 *Illumina Sequencing*

The University of Washington Genetics and Solid Tumor Laboratory uses Illumina sequencing instruments for all NGS assays, including a HiSeq 2500, NextSeq 500, and MiSeq instruments. There are two types of oligonucleotides bound on a flow cell that act as primers, enabling the prepared DNA fragments to go through solid-phase amplification by bridge PCR (Figure 2.1). Millions of clonal PCR clusters are formed for individual fragments on the flow cell that subsequently undergo sequencing. DNA sequencing is accomplished through sequencing-by-synthesis (SBS) using cyclic reversible dye-terminator chemistry (Figure 2.2). Multiple sequencing cycles occur in which each cycle consists of nucleotide incorporation, fluorescence imaging, and cleavage of the 3’blocking group before the cycle is repeated [23].

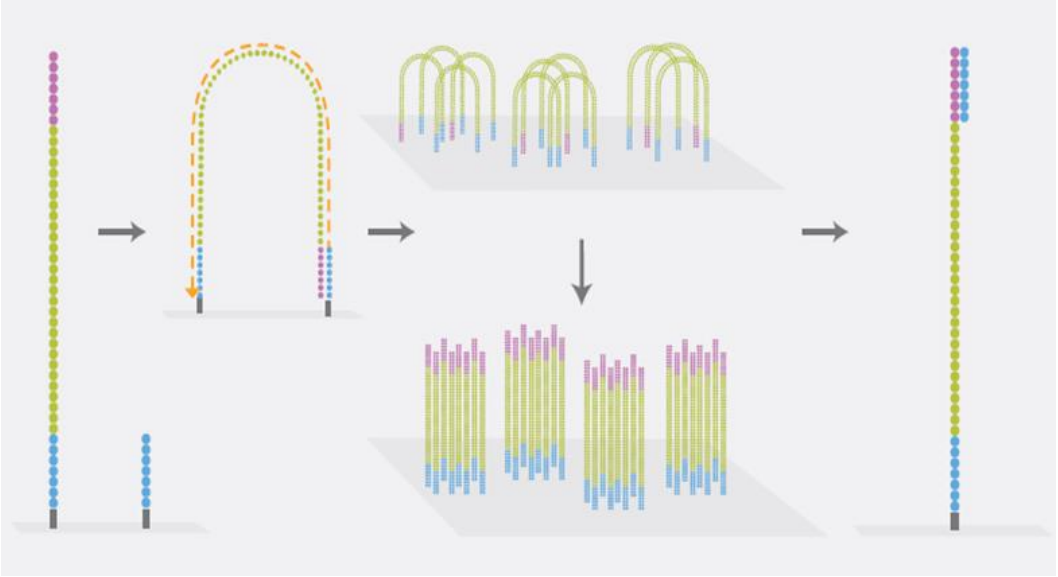

Figure 2.1 – Illumina Bridge Amplification PCR [24]

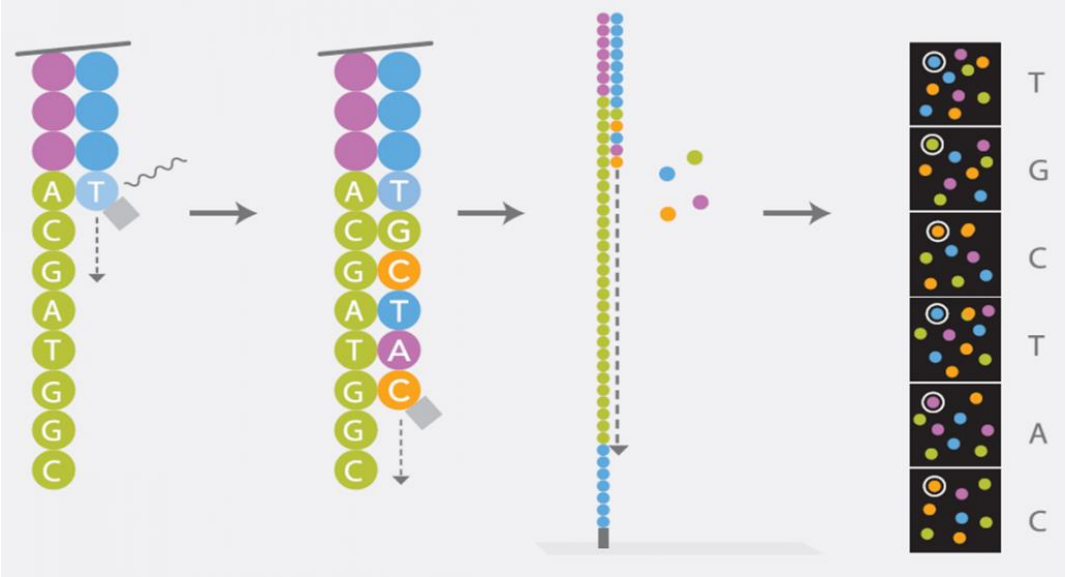

Figure 2.2 – Illumina Sequencing By Synthesis [24]

The termination of DNA synthesis after the addition of a single nucleotide is an important feature. A key advantage to the SBS method is the reversible dideoxynucleotide terminator. Blocking groups, such as 3′-O-allyl-2′-deoxyribonucleoside triphosphates (dNTPs) and 3′-O-azidomethyl-dNTPs may be used. Following fluorescent imaging, 3′-blocked terminators require

the cleavage of two chemical bonds to remove the fluorophore from the nucleobase and restore the 3′-OH group in order for chain elongation to occur in the next cycle [23]. The number of cycles determines the length of the sequencing reads (~100 bp currently on HiSeq and ~150 bp currently on NextSeq). The fluorescent image files are converted into the corresponding nucleotide sequences for analysis.

2.2.4 *Bioinformatics and Data Analysis*

Sequences generated by the above SBS method are processed by an automated, custom-designed Laboratory Medicine bioinformatics pipeline developed by the UW NGS Analytics Laboratory. Data are demultiplexed using bcl2fastq (Illumina, San Diego CA) and initial read mapping against the human reference genome (hg19/GRCh37) and alignment processing is performed using BWA and SAMtools, respectively. A FASTQ file is generated per patient and sample-level, fully local indel realignment is then performed using GATK Universal Genotyper. Duplicate reads are removed using PICARD and quality score recalibration is then performed using GATK. This recalibrated alignment is used for all subsequent analyses [19].

Sequence files are subsequently analyzed by variant calling algorithms comprised of: GRIDSS and Breakdancer for structural variants, the structural variant and indel algorithm Pindel, CONTRA for CNVs, and SNV algorithms VarScan2 and GATK (Figure 2.3) [19]. Evaluation of MSI status was done using mSINGS which calculates the fraction of unstable microsatellite loci [25]. Variant calls are collated into a single analysis document and analyzed by a Genetics and Solid Tumors Lab director for every case, including review by multiple directors at a consensus conference.

As a result of the deep sequencing approach, OncoPlex is an extremely sensitive assay and is able to detect rare variants that are at variant allele frequencies (VAFs) as low as 5% for

samples with 10% tumor fraction. However, the limitation of such a high sensitivity assay is a decreased specificity, such that the data must be analyzed manually to differentiate analytically real variants from technical artifact. Before variant analysis, the quality of sequencing libraries is evaluated based on metrics from library complexity, sequence read depth, and the fraction of reference variants recovered in control NA12878.

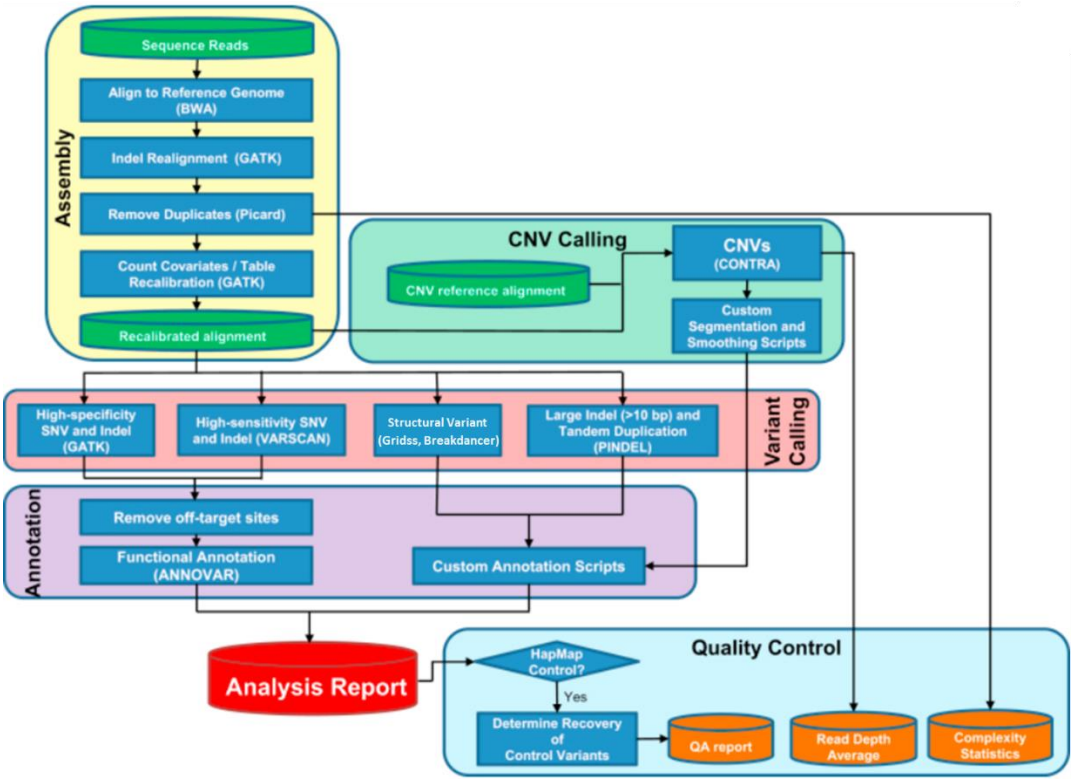

Figure 2.3 – Custom Bioinformatics Pipeline (modified from [19])

The sequencing quality is assessed at the run level to ensure optimal quality for assessment by evaluating the recovery of 565 credible SNVs in the NA12878 HapMap control. The sequencing run is considered valid for missed detection of up to 5 variants out of 565 SNVs. At the sample level the average read depth, percentage of usable bases on-target, and coverage across each gene must be evaluated. Quality metrics are also performed on gene and exon basis,

in which any regions below 50X are flagged as failed. For every variant class, the variants are distinguished between analytically real and technical artifact.

2.3 VALIDATION

Validation of OncoPlex v6 was performed using 90 unique samples chosen based on previous orthogonal clinical test results from several next generation sequencing assays, including prior hybrid capture assays (previous OncoPlex versions, UW BROCA, and Brigham and Women's Hospital's OncoPanel), and amplicon assays (UW Solid Tumor Hotspot), as well as the UW Cytogenomics Lab RNA sequencing solid tumor panel (FusionPlex, ArcherDx). Currently there are five different types of acceptable specimens including tissue, purified DNA, blood, bone marrow, and cell-free DNA. Approximately 27 different tumor types were tested. The breakdown of the sample types tested is listed in Table 2.1. All five classes of mutations were tested to ensure the assay's detection capability.

Table 2.1 – Validation Sample Breakdown

Sample Type	Count
FFPE	77
Blood	3
Bone Marrow	3
Cell-free DNA	2
Saliva	1
HapMap	4
Total	90

Four HapMap samples were also used to evaluate the performance characteristics of OncoPlex v6. NA12878 was used as a positive control for every sequencing run. Three other HapMap samples were also sequenced, NA12891, NA12892, and NA18517.

Determining the performance characteristic for each individual gene on the panel for every sample and tumor type is not feasible, but rather an error-based approach must be taken in order to validate a large panel NGS assay. The question to focus on is: "to what extent can the performance of the test for a given sample type, variant type, genomic region, or allele burden be extrapolated to other sample types, variant types, genomic regions, and allele burdens?" [4]. Validation samples were chosen to not only show the assay's ability to detect the common clinically significant mutations, but also the rare and challenging mutations.

A total of 11 sequencing runs (eight NextSeq runs and three HiSeq runs) were performed across 12 weeks and three technologists. Eleven samples plus a control were processed for NextSeq runs, and 23 samples plus a control were processed for HiSeq runs.

2.4 RESULTS

2.4.1 *Sequencing Quality*

Across all 11 runs, the average sequencing quality was excellent; the NextSeq runs had an average of 541X coverage and 60.1% on-target sequencing reads and the HiSeq runs had an average of 747X coverage and 61.6% on-target reads (Table 2.2).

Table 2.2 – Sequencing Quality Metrics

Instrument	Metric	OPX v5	OPX v6	Fold Change
NextSeq	Avg Coverage	291X	541X	1.9
	% On-target	15%	60%	3.9
	% Duplicate	23%	27%	1.2
	Total Human Read Pairs	35079820	16786623	0.48
HiSeq	Avg Coverage	255X	747X	2.9
	% On-target	18%	62%	3.4
	% Duplicate	18%	32%	1.8
	Total Human Read Pairs	25141535	22060520	0.87

Samples pass the quality control threshold if average read depth across greater than 90% of the genes is at least 150X. This threshold was determined by plotting the fraction of genes that met or exceeded the minimum coverage per sample against the sample's respective average coverage across all 340 genes. Four minimum coverage thresholds were investigated: 100X, 150X, 200X, and 250X (Figure 2.4). To achieve a minimum of 100X coverage of 100% of the genes (n=340) in the capture, 150X average coverage is required. To achieve a minimum of 150X coverage for 90% of the 340 genes, between 350X and 400X average coverage is required. To achieve a minimum of 200X coverage for 90% of the 340 genes, over 450X average coverage is required. And lastly, to achieve a minimum of 250X coverage for 90% of the 340 genes, 600X average coverage is required.

As 150X coverage is routinely capable of identifying SNVs with a 5% VAF, this threshold was deemed the optimal QC standard. However, these numbers can vary on a case by case basis and factors such as diminishing DNA quality and large deletion mutations will contribute to lower coverage. Additionally, samples with less robust coverage can be interpreted at a director's discretion.

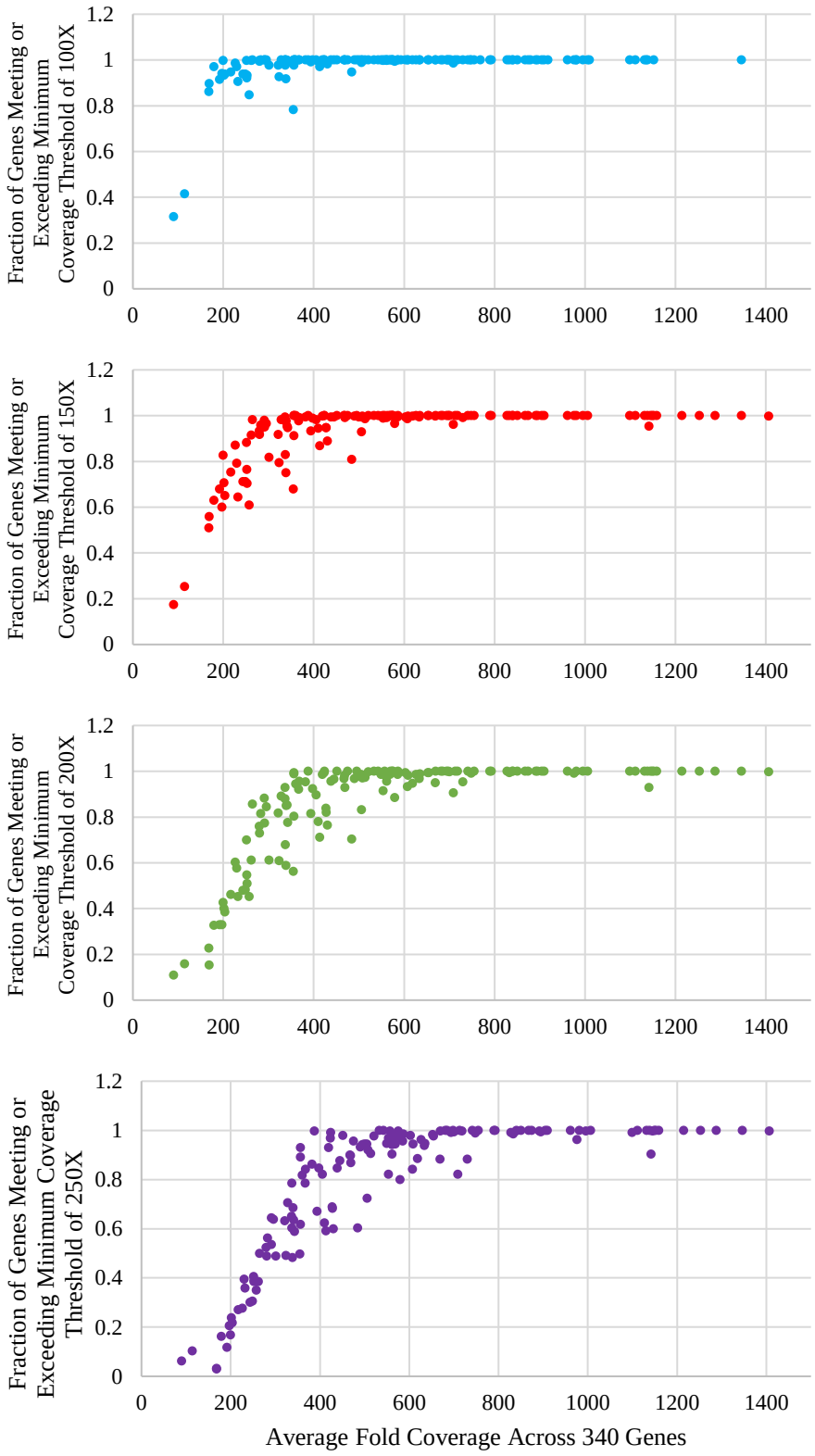

Figure 2.4 – Fraction of Genes Meeting or Exceeding Different Minimum Coverage Thresholds. Each dot represents a sample. Blue = 100X, red = 150X, green = 200X, and purple = 250X.

HapMap sample NA12878 was processed in every experiment and was used as the run-specific control. There were 565 SNVs in common between OncoPlex v6 and the HapMap SNP array from 1000 Genomes. For the sequencing run to pass QC and be considered as valid, 99% of the SNVs (560/565) must have been detected. This 99% threshold was chosen based on previous OncoPlex versions. Regardless of the technologist performing the assay and the sequencer that was used, 100% of the 565 SNVs were recovered from all 11 runs except for one NextSeq run that missed one intronic variant (Table 2.3).

Table 2.3 – Accuracy of OncoPlex v6 compared to 1000 Genome SNP Array

Experiment	Tech	Total SNVs	Detected	Not Detected	Accuracy
NextSeq Runs					
N1 (205)	Tech 1	565	565	0	100%
N2 (208)	Tech 1	565	565	0	100%
N3 (211)	Tech 2 and 3	565	565	0	100%
N4 (212)	Tech 2 and 3	565	565	0	100%
N5 (213)	Tech 1 and 2	565	564	1	99.8%
N6 (215)	Tech 3	565	565	0	100%
N7 (221)	Tech 3	565	565	0	100%
N8 (222)	Tech 3	565	565	0	100%
HiSeq Runs					
H1 (210)	Tech 1	565	565	0	100%
H2 (214)	Tech 2 and 3	565	565	0	100%
H3 (216)	Tech 3	565	565	0	100%

2.4.2 *Single Nucleotide Variants*

For all validation samples, a list of clinically reported SNVs was created to compare variant identification and allele fraction between previous clinical results and OncoPlex v6. A total of 348 SNVs were reported in previous clinical results across 62 samples and OncoPlex v6 detected 343/348, a sensitivity of 98.9%. Of the 348 SNVs, 324 had a VAF greater than or equal to 10% and when only looking at this subset, OncoPlex v6 detected 323/324, a sensitivity of 99.7%. As the VAF fell below 5% the sensitivity fell to 71.4%.

Reproducibility of the assay was excellent with 220/221 SNVs detected between at least two separate sequencing runs, which equates to an inter-run reproducibility of 99.5%. Inter-run reproducibility is at 100% when SNVs with VAF greater than or equal to 10% are observed. 39/40 SNVs were detected across six samples that were repeated (up to four repeats) within their respective runs for an intra-run reproducibility of 97.5% (Table 2.4).

Table 2.4 – SNVs Accuracy and Qualitative Precision

VAFs	Accuracy	Inter-Run	Intra-Run
All	343/349 (98.3%)	220/221 (99.5%)	39/40 (97.5%)
≥ 10%	323/325 (99.4%)	207/207 (100%)	36/37 (97.3%)
< 10%	20/24 (83.3%)	13/14 (92.9%)	3/3 (100%)
< 5%	5/7 (71.4%)	1/1 (100%)	1/1 (100%)

2.4.3 *Insertions and Deletions*

Depending on the size of the indel, the variant can be picked up by VarScan2, Pindel, or Breakdancer. Across all three algorithms, 112 unique indels were detected in all previous clinical results and OncoPlex v6 detected 110 of them, for a sensitivity of 98.2%. Reproducibility of both inter and intra-run was 100% (Table 2.5).

Table 2.5 – Indels Accuracy and Qualitative Precision

Size	Accuracy	Inter-Run	Intra-Run
All	110/112 (98.2%)	89/89 (100%)	10/10 (100%)
≥ 12bp	17/17 (100%)	8/8 (100%)	1/1 (100%)
< 12bp	93/95 (97.9%)	81/81 (100%)	9/9 (100%)

2.4.4 *Structural Variants*

Structural variants (SVs) were analyzed through three different algorithms, Pindel, Breakdancer and Gridss. There were 30 SVs targeted including 23 fusions across 36 fusion partners, 1 intragenic tandem duplication, 3 inversions, and 4 deletions. The sensitivity of SV detection is 97.1% overall (34/35 detected). The one SV that was not detected was a *STAMBP1*-

MSH2 fusion present in the raw data file but below the detection cutoff of the Gridss filter. A previously missed *CCDC6-RET* fusion was successfully detected by OncoPlex v6. With regards to reproducibility of the assay, 12/13 SVs were detected inter-run, and 5/6 SVs were detected intra-run (Table 2.6).

Table 2.6 – SVs Accuracy and Qualitative Precision

SV	Accuracy	Inter-Run	Intra-Run
All SVs	34/35 (97.1%)	12/13 (92.3%)	5/6 (83.3%)
Excluding indels	29/30 (96.7%)	11/12 (91.7%)	5/6 (83.3%)

2.4.5 *Copy Number Variants*

Clinically reported CNVs were assessed by comparing the average adjusted log ratio at the gene level. Gene amplifications and homozygous deletions were detected with 100% sensitivity and precision both inter-run and intra-run. Less robust single gene copy gain or copy loss demonstrated decreased sensitivity, as low as 80% (Table 2.7). These findings are not surprising as low copy number gains and losses (both focal and sub-chromosomal and chromosomal alternations) are challenging to detect in the presence of FFPE artifacts and they are of lower clinical significance.

Table 2.7 – CNVs Accuracy and Qualitative Precision

CNV	Accuracy	Inter-run	Intra-run
All CNVs	163/165 (98.8%)	54/55 (98.2%)	13/14 (92.9%)
Amplification	17/17 (100%)	8/8 (100%)	1/1 (100%)
Homozygous Del	16/16 (100%)	7/7 (100%)	3/3 (100%)
Focal Gain	15/15 (100%)	5/6 (83.3%)	1/1 (100%)
Focal Loss	13/13 (100%)	6/6 (100%)	NA
Sub/Chr Gain	43/44 (97.7%)	2/2 (100%)	1/1 (100%)
Sub/Chr Loss	59/60 (98.3%)	26/26 (100%)	7/8 (87.5%)

2.4.6 *Effect of Varying DNA Input*

Intra-run reproducibility was also measured while simultaneously experimenting with varying DNA input for library preparation. Previous OncoPlex versions ideally require 500 ng of DNA input. For OncoPlex v6, 250 ng versus 100 ng was compared with the same sample within the same sequencing run to analyze the performance with decreasing DNA mass. For all alteration types, the assay was able to detect 100% of the clinically reported mutations regardless of DNA input (Table 2.8). For the SVs that were analyzed, it is important to note that although they were detected at both input levels, the Gridss quality score of the variant was significantly lower for the 100 ng input compared to the 250 ng input (Table 2.9).

Table 2.8 – Reproducibility by Alteration Comparing DNA Input

Alteration	# of Samples	OncoPlex v6 (250ng)	OncoPlex v6 (100ng)
SNVs	2	23/23 (100%)	23/23 (100%)
Indels	2	8/8 (100%)	8/8 (100%)
SVs	2	2/2 (100%)	2/2 (100%)
CNVs	1	10/10 (100%)	10/10 (100%)

Table 2.9 – Comparison of Gridss Quality Score by DNA Input

Alteration	QUAL (250 ng)	QUAL (100 ng)	Fold Difference
ETV6-NTRK3 fusion	2951	1388	2.12
MSH2 inversion	547	261	2.09

DNA input also had an effect on the average coverage of the samples, with an average of 1.7-fold decrease when lowering the DNA input to 100 ng. However, there was no significant effect seen on the percent on-target sequence (Table 2.10).

Table 2.10 – Comparison of Average Coverage and Percent On-Target by DNA Input

Sample	Average Coverage		Fold Change	Percent On-Target		Fold Change
	250 ng	100 ng		250 ng	100 ng	
1	336.7	262.2	1.28	59.2%	58.7%	1.00
2	607.5	337.4	1.80	62.4%	55.3%	1.13
3	506.1	257.2	1.97	53.8%	56.2%	0.96
4	427.2	232.4	1.84	57.9%	52.9%	1.09

2.5 DISCUSSION

OncoPlex v6 offers several improvements from prior versions and is a highly anticipated update. Not only is the wet lab workflow 40% faster than prior versions (3 days versus 5 days for library preparation), but is also more cost effective in the long term and provides superior sequencing results.

2.5.1 *Workflow Improvements*

The xGen IDT library preparation workflow that was adopted for OncoPlex v6 is two days faster than prior versions. The ligation of adaptors on the first day of library preparation allows for the samples to be pooled for multiplex hybridization capture in a single Eppendorf tube. By being able to work with a single tube, the amount of reagents used and the amount of pipetting both decrease and hence reduces the amount of potential cross contamination or mix up of samples during the workflow. Not only did the capture switch from singleplex to multiplex but the hybridization time also decreased from 24 to 4 hours. The subsequent wash steps also shortened from 3 to 1.5 hours and the whole library preparation protocol can be completed in 3 days to be ready for sequencing.

However, there have been some limitations with the multiplex capture. There were some challenges when circulating tumor DNA (ctDNA) samples were included into the pool for

library preparation. At the quantification step after the samples were pooled together, library pools containing ctDNA had extra bands in the electrophoresis gel image (Figure 2.5). When this library pool was carried through sequencing, the coverage of the ctDNA samples were extremely high and impacted the data quality of the other samples. It is hypothesized that the probes bind much more efficiently to this specimen type, potentially due to the presence of shorter fragments in ctDNA, or because the DNA quality from ctDNA is higher since it is not as damaged as the DNA from other sample types such as FFPE. The fact that the probes from IDT are DNA oligos instead of RNA may also be contributing to the increased efficiency. Adjustments to how ctDNA samples will get pooled is still in progress, but these sample types will potentially be pooled at 0.5X to 1X concentration instead of at 2X which was the protocol for OncoPlex v5.

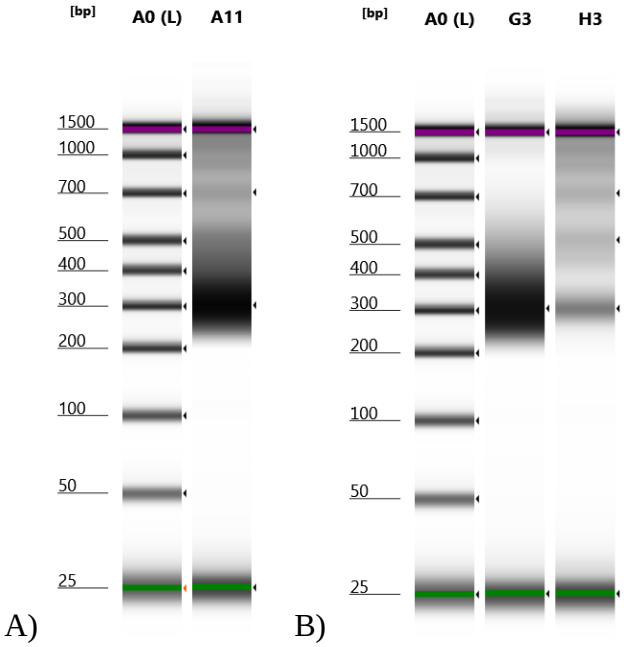

Figure 2.5 – Tapestation Gel Image

A) Final library pool containing ctDNA

B) Left lane = same final pool as image A without ctDNA, right lane = only ctDNA

Up until OncoPlex v5, library preparation was performed once a week, starting on Monday and loading the sequencer on Friday. But with the new version it may be possible to

process two OncoPlex runs within a week which would decrease the turnaround time; however, there will still be a bottleneck through the bioinformatics pipeline and data analysis. A large challenge for NGS testing is the high demand for bioinformaticians to support and process gigabytes of sequencing data.

2.5.2 *Panel Design*

The composition of the gene panel has also changed. The most noticeable change is the increase in the number of genes targeted from 262 genes to 340. The IDT gene panel is also a modular design in which new genes of interest can be added onto an existing panel without having to start the design process from scratch. This is a significant advantage, where new genes can be added to the capture panel with relatively little effort, allowing rapid adoption of genes that are found to be clinically relevant.

Within the last decade the molecular landscape of many cancers have evolved drastically as new genes are found to be clinically significant. With this increase in the number of genes that need to be tested, we have moved away from single gene testing since otherwise the other genes that also drive the tumors will be missed. In lung cancer alone there are over ten different genes that are known to be actionable; however, not all of them are common (Figure 2.6) [26]. For example, *RET* fusions are one of the more rare mutations observed in lung cancers. *RET* fusions are more common among the younger patients (<60 years old) who are non-smokers with adenocarcinoma histologic type. Patients positive for *RET* have significantly poorly differentiated tumors compared with patients with *ALK* fusions or *EGFR* mutations, but have several highly active therapies in clinical trials [27]. Despite *RET* mutations only making up a small percentage of lung cancer cases, being able to detect this fusion is life changing for the patient since they are eligible for a targeted drug.

Although *RET* fusions are rare in lung cancers, it is a common mutation found in Papillary Thyroid Carcinoma (PTC) [28]. There was a case of PTC that was tested on OncoPlex v5; however, no fusions were detected. Further testing was done at an outside location where a novel mutation *SNRNP70-RET* fusion was found. With the prior version of OncoPlex coverage of RET fusions was suboptimal and thus changes were made to ensure adequate coverage on OncoPlex v6.

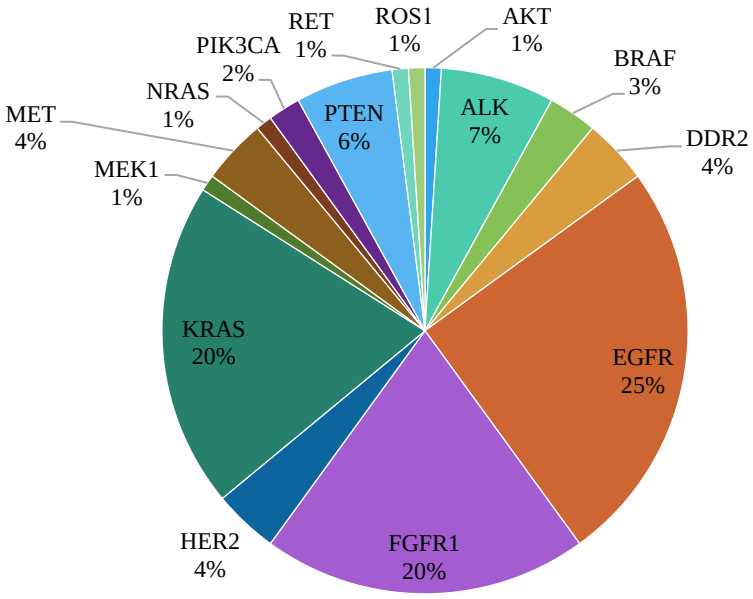

Figure 2.6 – Molecular Landscape of Lung Cancer (modified from [26])

As well 103 new genes, there are 22 intronic regions to genes in addition to the exonic regions for detection of structural variants that are difficult to detect without introns. Six of the genes in the OncoPlex v6 panel are complete gene regions (*BRCA1*, *BRCA2*, *MLH1*, *MSH2*, *MSH6*, and *PMS2*) meaning all non-repetitive sequence of the exons and introns are captured. Being able to detect whole genes is important so that we can detect fusions and single exon deletion mutations which is difficult to identify without intronic coverage.

The complete non-repetitive sequence for the mismatch repair (MMR) genes was incorporated into the new panel, which includes *MLH1, MSH2, MSH6* and *PMS2*. MMR genes are responsible for correcting any errors that occur when base pairs are incorporated by the DNA polymerase, especially in regions where there are repeats of nucleotide sequences. Mutations in these genes cause numerous mismatch and frameshift mutations across the genome, especially in the repetitive homopolymer regions. This leads to a characteristic genetic instability in cancer called microsatellite instability (MSI).

Many chemotherapeutic agents require a functional MMR system to initiate tumor damage, thus MMR deficient cells are frequently resistant to chemotherapy. It is essential to know whether a patient has mutated MMR genes because their response to chemotherapies may be limited. Using therapies that are unlikely to benefit the patient are contraindicated not only because of adverse side effects, but may also result in additional somatic mutations [29]. Although some traditional chemotherapies are not effective in MMR deficient (dMMR) tumors, such tumors are more likely to respond to immune checkpoint inhibitors, such as pembrolizumab, which was approved by the U.S. Food and Drug Administration (FDA) in May 2017 as the first tumor-agnostic, biomarker-driven drug approval [30].

BRCA1 and *BRCA2* are the other two genes where the complete non-repetitive sequence region is covered with OncoPlex v6. These genes code for tumor suppressor genes that are involved in homologous recombination DNA repair and maintain genomic stability in normal cells. Mutations in either gene is associated with defects in DNA damage response and thus individuals who carry a germline mutation in one of these genes have an increased risk for developing breast, ovarian, and other cancers. Individuals with mutations in these genes are

potentially eligible for receiving platinum therapies and/or poly ADO-ribose polymerase (PARP) inhibitors to treat their tumors [31].

Diffuse Intrinsic Pontine Glioma (DIPG) is a brainstem cancer that makes up about 10-20% of all childhood brain cancers and is the leading cause of death from solid tumors in children. The molecular landscape of high-grade gliomas differ slightly depending on the region of the brain affected, but approximately 60% of DIPG have a mutation in the *H3F3A* gene p.K27M (H3.3) (Figure 2.7). Immunohistochemistry (IHC) is non-specific for the histone proteins. One patient from our validation set had a positive IHC result for H3.3, and was tested on OncoPlex v5, but was negative for the expected *H3F3A* mutation. Approximately 15% of DIPG tumors are observed to have mutations in a related gene, *HIST1H3B,* which is included on OncoPlex v6 and was positive in this patient. Interestingly, H3.3 IHC cross reacts with the *HIST1H3B* gene product, and the IHC cannot differentiate between the two proteins [32].

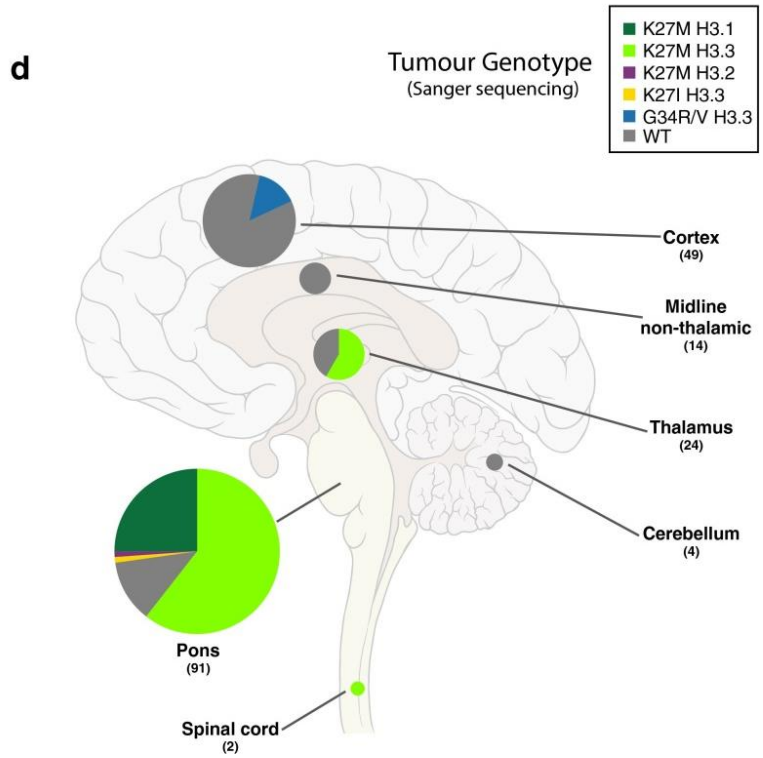

Figure 2.7 – Anatomical Distribution of H3 Family Mutations [28]

After careful considerations some gene targets were removed from the panel because there are more appropriate methods to detect them. For example, for patients with Chronic Myelogenous Leukemia (CML), the standard of care is to quantitatively measure the *BCR-ABL* fusion which is the underlying genetic defect responsible for the Philadelphia chromosome. The preferred method for *BCR-ABL* detection is either by Reverse Transcriptase PCR or by FISH, both of which are faster and cheaper than NGS. Another example is the *PML-RARA* fusion which is detected in the majority of patients with patients with Acute Promyelocytic Leukemia (APL) and requires rapid identification to allow for appropriate therapy induction on a timescale that is faster than what NGS methods can offer. In both of these hematologic malignancies, the fusions are the driver mutations and although the patients may have other somatic mutations that can be detected with OncoPlex, there are limited therapeutic benefits in knowing what they are.

2.5.3 *Sequencing Data Quality*

The average coverage and percent of on-target sequencing was drastically improved between version 5 and 6. The sample read depth had a large improvement from prior versions, approximately 1.9-fold increase with the NextSeq and 2.9-fold increase with the HiSeq. The average percent on-target sequencing reads improved even more with 3.9-fold and 3.4-fold increase with the NextSeq and HiSeq respectively (Table 2.2).

The increase in percent on-target sequencing indicates that a larger total number of bases output by the sequencer within the target region has increased. This demonstrates the efficiency of the probe hybridization to the target sequence more effectively and not wasted on off-target sequences allowing for more efficient use of sequencing space. Another contributing factor to the improved sequencing data quality is the use of xGen Blocking Oligos and the Human Cot-1 DNA. The increase in coverage had significant implications in cases where mutations were

missed. There was a sample from a patient with papillary thyroid carcinoma (PTC) that was tested on OncoPlex v5; however, the results were indeterminate. Many PTC cases commonly have fusions that are drivers of the tumor, therefore the tissue was sent to an outside hospital for additional testing. As suspected this patient had a *TPM3-NTRK1* fusion and was also detected by OncoPlex v6 (Figure 2.8).

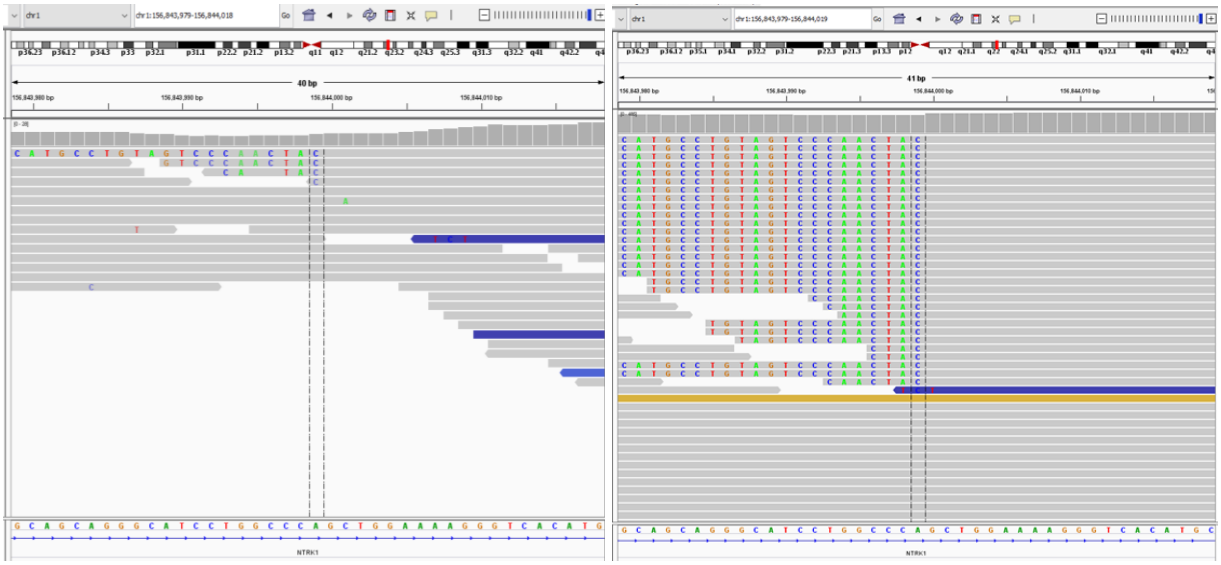

Figure 2.8 – Images of the *TPM3-NTRK1* Fusion on Integrative Genomics Viewer at *NTRK1*. The nucleotide sequence shown within the gray bars indicates that the sequence does not match the *NTRK1* gene. The discordant sequence matches with the *TPM3* gene, indicating a fusion between *NTRK1* and *TPM3*. The *TPM3-NTRK1* fusion was missed on OncoPlex v5 (left) and detected on OncoPlex v6 (right).

Although coverage did improve with the initial OncoPlex v6 gene panel, the balance of the individual gene probes needed adjustments due to lack of adequate coverage of some targets. As well as adding new genes to an existing panel, the concentration of low-performing probes can be increased within the panel to improve the coverage. Although the initial phase of testing demonstrated only rare failing genes for the majority of tested samples, there were 32 genes that contained clinically significant regions and required coverage improvement.

Table 2.11 – Average Coverage of the 32 Genes Before and After the 2X Spike-in.

Gene	Before Spike in Average Reads	After Spike in Average Reads	Fold Increase
BTK	557	939	1.7
CD33	368	674	1.8
CDC27	320	692	2.2
FANCB	592	1202	2
FOXR2	545	1201	2.2
H3F3A	381	835	2.2
IDH1	423	940	2.2
JAK1	362	906	2.5
JAK2	441	888	2
KDM6A	373	866	2.3
KDR	370	1026	2.8
KIF5B	560	773	1.4
KIT	408	857	2.1
KRAS	401	811	2
MAP2K4	401	833	2.1
NAT2	372	996	2.7
NRAS	339	700	2.1
NUDT15	392	851	2.2
PDCD1LG2	436	909	2.1
PHF6	364	921	2.5
PMS2	387	539	1.4
PRPS1	425	744	1.8
PTEN	450	966	2.1
QKI	368	400	1.1
RAD21	404	1023	2.5
RECQL	440	761	1.7
RPL10	407	836	2.1
RPS14	348	647	1.9
SDHB	382	768	2
SDHC	360	758	2.1
SMAD2	447	1015	2.3
STAG2	519	1242	2.4

These 32 genes were spiked into the prior 1X gene panel, previously containing all 340 genes in equal concentration, at 2X to achieve a combined 3X concentration. This enrichment was expected to double the coverage of these genes. All targets demonstrated clinically significant improvement in coverage ($p < 0.0001$), with the exception of *QKI* (Table 2.11). To confidently call variants, especially those that are present at low VAFs, it is essential that all

genes are uniformly covered. Any gaps in coverage or any regions where sequencing depth is lower compared to the average depth of coverage can result in false negative variant calls.

2.6 CONCLUSION

OncoPlex v6 has significant improvements from prior versions. First, the library preparation can now be completed in three days instead of five days and thus decreasing turnaround time. The sequencing data quality is superior in terms of both average coverage and percent on-target sequences and the performance characteristics is excellent across all alteration types. Not only are all of the clinically significant mutations detected but OncoPlex v6 was also able to detect mutations that were previously missed due to lack of coverage of certain regions in previous OncoPlex versions. The ability to identify and detect a wide array of genes is essential and is one of the largest benefits of a multiplex panel. Panel size will only continue to grow as new clinically relevant findings are discovered on many different genes. Most importantly, its modular design guarantees a nimble clinical laboratory response in addressing the expansions necessary to meet the needs of a continuously evolving molecular oncology field.

CONCLUSION

Molecular diagnostic assays are being applied to many disciplines in laboratory medicine and expectations are rising as it transforms the medical practice. Given this, more and more clinical laboratories are developing complex LDTs in order to continue providing the best patient care. This thesis described the validations of two separate molecular diagnostic assays which used different molecular techniques.

The real-time PCR SNP genotyping assay was developed to assess potential CKD risk by looking at the *ApoL1* G0, G1, and G2 genotype. An immense clinical utility for genotyping exists especially for the African-American population who have a higher risk of developing CKD. This in part is due to the inheritance of the *ApoL1* variant allele (G1 and G2 allele) which is more prevalent in the Afro-descendent population due to its ability to provide resistance to trypanosomiasis. The accuracy of this real-time PCR genotyping assay was an average of 99.4% and both the inter-run and intra-run reproducibility was 100%. Although optimization in the wet lab workflow is still needed, this real-time PCR genotyping assay is a fast, simple method that can be implemented easily in a clinical laboratory.

A comprehensive NGS assay that detects 340 somatic cancer mutations, OncoPlex v6, was validated and provides significant improvements from prior versions. Accuracy and precision across all alteration classes exceeded 97% and 92% respectively, with the majority of variants not identified either present at low VAFs or demonstrating low-level copy gains or losses. Compared to prior OncoPlex versions, OncoPlex v6 demonstrated an increase in sequencing data quality with a 2.5-fold increase in average sequencing coverage and a 3.7-fold increase in percent on-target sequencing. Additionally, there was a 40% reduction in library preparation turnaround time (three days versus five days) at an overall decreased cost per library.

Most importantly, the gene panel has a modular design and thus the clinical laboratory is able to integrate future changes as needed as the molecular oncology field evolves.

Both assays are capable of providing critical information that can impact prognosis, diagnosis, and treatment options. Further understanding and improvements within the molecular diagnostics field will drive the development of additional molecular assays and provide real benefits for patient care.

BIBLIOGRAPHY

[1] Luh, F., Yen, Y. (October 2018). FDA guidance for next generation sequencing-based testing: balancing regulation and innovation in precision medicine. *NPJ Genomic Medicine.* 3(28). DOI: 10.1038/s41525-018-0067-2

[2] What is genetic testing? (April 2018). *Genetics Home Reference*. NIH. Retrieved from https://ghr.nlm.nih.gov/primer/testing/genetictesting. Accessed May 5th 2019.

[3] Castellani, W. (April 2010). Laboratory Developed Tests. *CAP.* Retrieved from http://webapps.cap.org/apps/docs/education/lapaudio/pdf/042110_Presentation.pdf

[4] Jennings, L. J., Arcila, M. E., Corless, C., et al. (2017). Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels. *Journal of Molecular Diagnostics: JMD*, 19:341-36. DOI: http://dx.doi.org/10.1016/j.jmoldx.2017.01.011.

[5] Raymaekers, M., Smets, R., Maes, B., Cartuyvels, R. (2009). Checklist for Optimization and Validation of Real-Time PCR Assays. *Journal of Clinical Laboratory Analysis.* 23:145-151. DOI: 10.1002/jcla.20307

[6] What are single nucleotide polymorphisms (SNPs)? (April 2019). Genetics Home Reference. NIH. Retrieved from https://ghr.nlm.nih.gov/primer/genomicresearch/snp. Accessed May 5th 2019.

[7] AKF Staff. (February 2017). Infographic: Kidney disease: A silent killer affecting African-Americans. American Kidney Fund. Retrieved from: http://www.kidneyfund.org/kidney-today/infographic-kidney-disease-african-americans.html. Accessed May 7th 2019.

[8] Lieberman, J. (October 2017). Of Parasites and Evolutionary Pressure: ApoL1 Risk Alleles and Chronic Kidney Disease. (PowerPoint Slides).

[9] Parsa, A., Kao, L., Xie, D., Astor, B. C., et al. (2013). APOL1 Risk Variants, Race, and Progression of Chronic Kidney Disease. *The New England Journal of Medicine*: 369:2183-2196. DOI: 10.1056/NEJMoa1310345

[10] Luzzatto, L. (October 2012). Sickle Cell Anemia and Malaria. *Mediterranean Journal of Hematology and Infectious Diseases.* 4:1. DOI: 10.4084/MJHID.2012.065

[11] Friedman, D. J., Pollak, M. R. (April 2016). Apolipoprotein L1 and Kidney Disease in African Americans. *Trends in Endocrinology & Metabolism*: 27:204-215. DOI:10.1016/j.tem.2016.02.002

[12] Limou, S., Nelson, G. W., Kopp, J. B., & Winkler, C. A. (2014). APOL1 kidney risk alleles: population genetics and disease associations. *Advances in chronic kidney disease*, *21*(5), 426–433. doi:10.1053/j.ackd.2014.06.005

[13] Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. dbSNP accessionrs73885319 and rs71785313 (dbSNP Build ID: 152). Available from: http://www.ncbi.nlm.nih.gov/SNP/

[14] Rubinow, K., Henderson, C., Robinson-Cohen, C. et al. (December 2017). Kidney function is associated with an altered protein composition of high-density lipoprotein. *Kidney international.* 92(6): 1526-1535. DOI:10.1016/j.kint.2017.05.020

[15] TaqMan® SNP Genotyping Assays USER GUIDE. (2017). *Life Technologies Corporation.* Retrieved from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0009593_TaqManSNP_UG.pdf

[16] TaqMan® Genotyper Software Getting Started Guide. (2010). *Life Technologies Corporation.* Retrieved from: http://tools.thermofisher.com/content/sfs/manuals/cms_081968.pdf

[17] Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Retrieved from: www.genome.gov/sequencingcostsdata.

[18] Pritchard, C. (November 2013). Precision Oncology: Experience at UW. (PowerPoint Slides).

[19] Pritchard, C. C., Salipante, S. J., Koehler, et al. (2014). Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene

rearrangement in clinical cancer specimens. The Journal of molecular diagnostics: JMD, 16(1), 56–67. DOI:10.1016/j.jmoldx.2013.08.004
[20] Integrated DNA Technologies, Inc. (2019). xGen hybridization capture of DNA libraries. Retrieved from: https://sfvideo.blob.core.windows.net/sitefinity/docs/default-source/protocol/xgen-hybridization-capture-of-dna-libraries.pdf
[21] xGen Blocking Oligos. *Integrated DNA Technologies.* https://www.idtdna.com/pages/products/next-generation-sequencing/hybridization-capture/blockers/blocking-oligos
[22] Integrated DNA Technologies, Inc. (2018). Human Cot DNA. Retrieved from: https://www.idtdna.com/pages/products/next-generation-sequencing/hybridization-capture/hybridization-reagents/human-cot-dna
[23] Illumina Sequencing Technology. (2010). *Illumina Inc.* Retrieved from: https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf
[24] Next Generation Sequencing. *CeGaT.* Retrieved from: https://www.cegat.de/en/services/next-generation-sequencing/
[25] Pritchard, C. C., Smith, C., Salipante, S. J., et al. (2012). ColoSeq Provides Comprehensive Lynch and Polyposis Syndrome Mutational Analysis Using Massively Parallel Sequencing. *The Journal of Molecular Diagnostics: JMD,* 14(4): 357–366. DOI: http://doi.org/10.1016/j.jmoldx.2012.03.002
[26] Lovly, C., L. Horn, W. Pao. 2018. Molecular Profiling of Lung Cancer. My Cancer Genome https://www.mycancergenome.org/content/disease/lung-cancer/
[27] Ferrara, R., Auger, N., Auclin, E., Besse, B. (January 2018). Clinical and Translational Implications of RET Rearrangements in Non-Small Cell Lung Cancer. *Journal of Thoracic Oncology.* 13(1):27-45. DOI: 10.1016/j.jtho.2017.10.021
[28] Cancer Genome Atlas Research Network. (October 2014). Integrated genomic characterization of papillary thyroid carcinoma. *Cell.* 159(3): 676-690. DOI: 10.1016/j.cell.2014.09.050.
[29] Abedalthagafi, M. (October 2018). Constitutional mismatch repair-deficiency: current problems and emerging therapeutic strategies. *Oncotarget.* 9(83): 35458-35469. DOI: 10.18632/oncotarget.26249.
[30] Hempelmann, J., Lockwood, C., Konnick, E., et al. (April 2018). Microsatellite instability in prostate cancer by PCR or next-generation sequencing. *Journal for ImmunoTherapy of Cancer.* 6(29). DOI: 10.1186/s40425-018-0341-y
[31] Cerrato, A., Morra, F., Celetti A. (November 2016). Use of poly ADP-ribose polymerase [PARP] inhibitors in cancer cells bearing DDR defects: the rationale for their inclusion in the clinic. *Journal of Experimental & Clinical Cancer Research.* 35(179). DOI: 10.1186/s13046-016-0456-2.
[32] Castel, D., Philippe C., Calmon. R, et al. (December 2015). Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes. *Acta Neuropathologica.* 130(6):815-27. DOI: 10.1007/s00401-015-1478-0.

APPENDIX A

Chromosome 22 genomic sequence corresponding to *ApoL1*

NC_000022.11:36253071-36267531

CTCAGACGAGCCAGAGCCAATCTTCAGTCAGTACCGCATGCCTCAGCCTCACGCCCC
CGGGTCACTGAGCCAATCTCAGCTGAAAGCGGTGAACAGGTGGAGAGGGTTAATGA
ACCCAGCATCCTGGAAATGAGCAGAGGAGTCAAGCTCACGGATGTGGCCCCTGTAA
GCTTCTTTCTTGTGCTGGATGTAGTCTACCTCGTGTACGAATCAAAGCACTTACATGA
GGGGGCAAAGTCAGAGACAGCTGAGGAGCTGAAGAAGGTGGCTCAGGAGCTGGAG
GAGAAGCTAAACATTCTCAACAATAATTATAAGATTCTGCAGGCGGACCAAGAA
CTGTGACCACAGGGCAGGGCAGCCACCAGGAGAGATATGCCTGGCAGGGGCCAGG
ACAAAATGCAAACTTTTTTTTTTTTCTGAGACAGAGTCTTGCTCTGTCGCCAAGTTGG
AGTGCAATGGTGCGATCTCAGCTCACTGCAAGCTCTGCCTCCCGTGTTCAAGCGATT
CTCCTGCCTTGGCCTCCCAAGTAGCTGGGACTACAGGCGCCTACCACCATGCCCAGC
TAATTTTTGTATTTTTAATAGAGATGGGGTTTCACCATGTTGGCCAGGATGGTCTCG

APPENDIX B

ApoL1 Reference Protein Sequence

MEGAALLRVSVLCIWMSALFLGVGVRAEEAGARVQQNVPSGTDTGDPQSKPLGDWAA
GTMDPESSIFIEDAIKYFKEKVSTQNLLLLLTDNEAWNGFVAAAELPRNEADELRKALD
NLARQMIMKDKNWHDKGQQYRNWFLKEFPRLKSELEDNIRRLRALADGVQKVHKGTT
IANVVSGSLSISSGILTLVGMGLAPFTEGGSLVLLEPGMELGITAALTGITSSTMDYGKKW
WTQAQAHDLVIKSLDKLKEVREFLGENISNFLSLAGNTYQLTRGIGKDIRALRRARANL
QSVPHASASRPRVTEPISAESGEQVERVNEPSILEMSRGVKLTDVAPVSFFLVLDVVYLV
YESKHLHEGAKSETAEELKKVAQELEEKLNILNNNYKILQADQEL

Phenotyping Peptides:

LNILNNNYK (G0)
LNMLNNNYK (G1)
LNILNNK (G2)

Quantification Peptides:

ALDNLAR
VTEPISAESGEQ

APPENDIX C

OncoPlex v6 Gene List

ABL1, ABL2, ACVR1, AKT1, AKT2, AKT3, ALK, ANGPTL1, ANKRD26, APC, AR, ARAF, ARID1A, ARID1B, ASXL1, ASXL2, ATM, ATR, ATRX, AURKA, AURKB, AXIN2, AXL, BABAM1, BAK1, BAP1, BARD1, BCL2, BCL2L11, BCOR, BCORL1, BCR, BIRC3, BMPR1A, BRAF, BRCA1, BRCA2, BRIP1, BTK, C11orf95, CALR, CARD11, CBL, CBLB, CBLC, CCND1, CCND2, CCNE1, CD19, CD274, CD33, CD74, CDC27, CDH1, CDK12, CDK4, CDK6, CDK8, CDK9, CDKN1A, CDKN2A, CDKN2B, CEBPA, CHD1, CHEK1, CHEK2, CREBBP, CRLF2, CRX, CSF1R, CSF3R, CTCF, CTNNA1, CTNNB1, CUX1, DAXX, DDR2, DDX41, DEPDC5, DICER1, DNAJB1, DNMT3A, DOCK7, DPYD, EBF1, EGFR, EIF3E, ELF1, EML4, EP300, EPAS1, EPCAM, EPHA3, EPHA5, EPHB2, EPHB6, ERBB2, ERBB3, ERBB4, ERCC2, ERG, ESR1, ESR2, ETV6, EZH2, FAM175A, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, FBXW7, FGFR1, FGFR2, FGFR3, FGFR4, FH, FKBP1A, FLT1, FLT3, FLT4, FOXA1, FOXR2, GAB2, GALNT12, GATA1, GATA2, GATA3, GEN1, GLI1, GLTSCR1, GLTSCR2, GNA11, GNAQ, GNAS, GREM1, GRIN2A, GRM3, H3F3A, H3F3B, HDAC4, HDAC9, HIF1A, HIST1H3B, HNF1A, HRAS, HSPH1, ID3, IDH1, IDH2, IGF1R, IKZF1, IL7R, JAK1, JAK2, JAK3, KDM6A, KDR, KIF5B, KIT, KLF4, KMT2A, KMT2C, KMT2D, KRAS, MAP2K1, MAP2K2, MAP2K4, MAPK1, MAX, MC1R, MCL1, MDM2, MDM4, MED12, MEGF6, MEN1, MET, MIOS, MITF, MLH1, MLH3, MN1, MPL, MRE11A, MSH2, MSH6, MSLN, MTAP, MTOR, MUTYH, MYB, MYC, MYCL1, MYCN, MYD88, MYOD1, NAB2, NAT2, NBN, NF1, NF2, NKX2-1, NOTCH1, NOTCH2, NOTCH3, NPM1, NPRL2, NPRL3, NR4A3, NRAS, NT5C2, NTHL1, NTRK1, NTRK2, NTRK3, NUDT15, PAK1, PALB2, PAX5, PBRM1, PDCD1LG2, PDGFRA, PDGFRB, PHF6, PHOX2B, PIK3CA, PIK3CB, PIK3R1, PLCG2, PLK1, PLK2, PLK3, PLK4, PML, PMS2, POLD1, POLE, PPM1D, PRKAR1A, PRPF40B, PRPS1, PTCH1, PTEN, PTPN11, PTPRD, QKI, RAC1, RAD21, RAD51B, RAD51C, RAD51D, RAF1, RARA, RB1, RECQL, RELA, RET, RHEB, RICTOR, RINT1, RIT1, ROR1, ROS1, RPL10, RPS14, RPS15, RPS20, RPTOR, RRM1, RRM2, RSPO2, RSPO3, RUNX1, SAMD9L, SDHA, SDHB, SDHC, SDHD, SETBP1, SETD2, SF1, SF3B1, SH2B3, SHH, SLX4, SMAD2, SMAD3, SMAD4, SMARCA4, SMARCB1, SMC1A, SMC3, SMO, SPOP, SPRY4, SRC, SRP72, SRSF2, STAG2, STAT5B, STAT6, STK11, SUFU, SUZ12, TACC3, TACSTD2, TCF3, TERC, TERT, TET1, TET2, TET3, TFE3, TFG, TGFBR2, TLX1, TMPRSS2, TP53, TP73, TRAF7, TRRAP, TSC1, TSC2, TTYH1, TYMS, U2AF1, U2AF2, VHL, WRN, WT1, XRCC2, YAP1, ZBTB16, ZRSR2

Fusions
ALK, BRAF, C11orf95, CD74, DNAJB1, EGFR, EIF3E, EML4, ETV6, FGFR1, FGFR2, FGFR3, KMT2A, MET, NTRK1, NTRK2, NTRK3 (detected with ETV6), RAF1, RET, ROS1, RSPO3, TMPRSS2, YAP1

MSI, SNPs