2016

Assessing clonal diversity in acute myeloid leukemia

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Thesis

ASSESSING CLONAL DIVERSITY IN

ACUTE MYELOID LEUKEMIA

By

WESTON CHRISTENSEN

B.S., Western Washington University, 2014

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2016

Approved by

First Reader _____
Vickery Trinkaus-Randall, Ph.D.
Professor of Biochemistry and Ophthalmology

Second Reader _____
Jerry Radich, M.D.
Member, Clinical Research Division, Fred Hutchinson Cancer Research
   Center
Professor, University of Washington School of Medicine
Director, Molecular Oncology Lab

ACKNOWLEDGMENTS

ASSESSING CLONAL DIVERSITY IN ACUTE MYELOID LEUKEMIA

WESTON CHRISTENSEN

ABSTRACT

Clonal diversity in cancer has been proposed as a mechanism underlying patient-to-patient variability in therapeutic response, as well as the variability in the likelihood of relapse and the time to relapse of acute myeloid leukemia (AML) as well as other cancers. As a neoplasm develops it often continues to mutate, diversifying into differing clonal populations. Darwinian evolutionary pressures such as inherent fitness imbalances, immune system interactions, and chemotherapy treatments target sensitive clones and drive competition between the clonal populations; selecting for dynamic and resistant cell lines. In this way clonal diversity is conceivable as an impediment to a complete remission with more populations offering more opportunities for therapy resistance. Bulk next generation sequencing (NGS) is currently used to assess clonal composition in leukemia but requires several broad assumptions be made, which can result in incorrect assessments of diversity. Factors such as differences in zygosity of mutations, convergent evolution, or contamination with wild-type/non-cancerous cells can artificially raise or lower reported variable allele frequencies (VAF), leading to errors in clonal assessments. To examine discrepancies between the actual clonal structure and the clonal structures determined through bulk sequencing we developed a novel method of sampling the cell population to identify concurrent mutations. We first created an *in silico* model which randomly draws cell samples from a simulated tumor multiple times and calculates the VAF for each mutant allele in each sample. By tracking the correlation of mutations

between sample replicates, a clonal composition that is not observable from the bulk NGS VAF becomes apparent. We then created *in vitro* model tumors from AML cell lines, isolated low cell number samples via flow cytometry, and applied a multiplex/nested PCR protocol with pyrosequencing to quantify VAFs in each sample. Again, by calculating the correlation of mutant alleles between replicates, previously unseen with NGS characteristics of the clonal structure becomes evident. Population sampling analysis may potentially offer a solution for clarifying how we can interpret NGS clonal analyses.

TABLE OF CONTENTS

LIST OF TABLES

<h1 style="text-align: center;">LIST OF FIGURES</h1>

LIST OF ABBREVIATIONS

ABI...Applied Biosystems Incorporated

AML.. Acute Myeloid Leukemia

APS ....Adenosine 5' phosphosulfate

ATP .... Adenosine Triphosphate

BLAST...Basic Local Alignment Search Tool

BSA.. Bovine Serum Albumin

dNTP...Deoxynucleoside Triphosphate

EtBr ....Ethidium Bromide

FBS ....Fetal Bovine Serum

FLT3 .... FMS-Like Tyrosine kinase 3

HSPC...Hematopoietic Stem/Progenitor Cells

ITD... Internal Tandem Duplication

NGS.. Next Generation Sequencing

NCBI...National Center for Biotechnology Information

NTC.. No Template Control

PBS ....Phosphate Buffered Saline

PCR..Polymerase Chain Reaction

PPi..Pyrophosphate

RPMI...Roswell Park Memorial Institute Medium

TKI...Tyrosine Kinase Inhibitors

INTRODUCTION

Acute Myeloid Leukemia

One of the most feared outcomes of any cancer treatment is relapse with a more aggressive and unmanageable cancer emerging from the post-treatment remission of the original. About one in five of all cancer survivors will experience a relapse within five years (Baker, Denniston, Smith, & West, 2005). Patients with acute myeloid leukemia (AML) face an even worse prognosis with much higher relapse and mortality rates (Szer, 2012).

In the United States the incidence of AML is 3.6 people per 100,000 per year, with over 11,000 new cases every year (Szer, 2012). While 20% of these cases are intractable to initial chemotherapy, the remainder will respond favorably and enter remission. Unfortunately most patients in remission will relapse, with an estimated incidence of relapse for patients that achieved a complete remission between 67% and 92% (Byrd et al., 2002). The basis of this relapse has been proposed to be related to clonal evolution, i.e. the continuous development of diseased cells with novel mutations generating new daughter colonies, increasing the range of targets that a regimen of chemotherapy must hit (Ding et al., 2012). Due to the role of clonal diversity in relapse assessment the clonal populations present in AML, and the relationship between clones present in neoplasms pre-treatment and post-treatment has become of interest to the scientific community. The high relapse rate of AML specifically makes it an excellent model for investigating clonality. Furthermore, successes in identifying impediments to full remissions would not only aide treatment of leukemia, but many cancers.

Clonality of Cancer

As modern cancer research has been evolving since its inception in the mid-1900s the direction of treatment has been shifting away from "cure-alls" and "wonder-drugs" to more precise and personalized therapies. This is due to both the understanding of the large variability that may be present in the genotypes of phenotypically identical cancers and the investigation into the potential clonal heterogeneity of individual cancers (Fernald, Capriotti, Daneshjou, Karczewski, & Altman, 2011; Paguirigan et al., 2015; Slamon et al., 2001). Previous models of clonality predicted a linear cancer model, beginning with an initial series of mutations that incrementally increase the fitness of a single clone which then expands to occupy the entire homogenous cancer population (Fearon, Hamilton, & Vogelstein, 1987). This has been shown to not be the whole situation. With the recent methodological advances in gene mutation detection, we can detect a surprising degree of heterogeneity between and within tumors (Navin et al., 2011). Distinct clonal differences can be observed inter-tumoral between primary tumors and subset metastasis in metastatic medulloblastoma as well as intra-tumoral in pancreatic and renal carcinomas (Gerlinger et al., 2012; Wu et al., 2012; Yachida et al., 2010). A complex network of evolutionary pressures creates genotypic "trees" for neoplasms with clonal populations experiencing competition within a tumor.

External and internal pressures from chemotherapy treatments, cell resilience, resource availability, and growth potential all play a role in the neoplasm's development, driving the success or failure of differing cell populations, or clonal populations, within a seemingly contiguous malignancy (**Figure 1**)(Greaves & Maley, 2012). In previous

cancer models relapse was explained by a failure of the intensity of the treatment to impact all effectively identical clones, driving physicians to extended and more rigorous treatment regimens. The new clonal model suggests that relapse is instead partially due to the medicinally based selection for therapy-resistant sub-clones (Turner & Reis-Filho, 2012). This has a particular relevance to leukemia; 30% of AML patients at initial diagnosis will contain some detectable allele load of the prognostically unfavorable FMS-like tyrosine kinase 3 (FLT3) mutations, such as internal tandem duplications (FLT3-ITD) or the rarer substitution mutations such as D835. Both of these mutations impart resistance to the typical first-round AML treatment, of tyrosine kinase inhibitors (TKIs) (Levis, 2013). Patients with FLT3 mutations who are treated with TKIs share similar remission rates of other AML subtypes; however, relapse is far more common and often --although not always-- associated with a higher FLT3-ITD associated mutation allele burden (Jan & Majeti, 2013). This pattern suggests that in these patients prior to treatment there existed a relatively small clonal population that contained a mutation on the FLT3 allele and was selected for by therapeutics which eradicated its competing colonies. With no other competition and a resistance to the therapeutics the clone could proliferate, thus increasing the number of cells that contain a mutation --and the FLT3-ITD variant allele frequency (VAF) -- until the patient was in full relapse with a now therapeutically resistant cancer. In this way, an increase in a neoplasms genetic clonal diversity can play a role in the deterioration of a patient's prognosis. Increasing clonal diversity has been revealed to predict malignancy in Barrett's esophagus and AML (Ding et al., 2012; Maley et al., 2006). By learning more about the impact of clonal diversity

and the relevance of specific recurrent mutations, physicians gain a powerful tool in

designing treatments; primarily targeting aggressive cell populations, predicting

imminent evolutions, and planning therapeutic regimens around probable remissions.
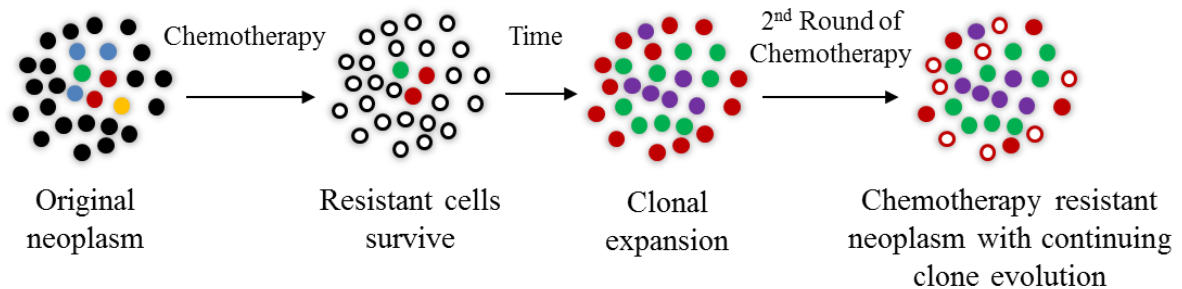
Figure 1. Clonal evolution and diversity. Initial regimens of chemotherapy, while often effective, may not eradicate every possible evolutionary product of a neoplasm. One or more surviving clones can grow to establish novel colonies and bring about a relapse. Enduring colonies also have the potential to further diversify with all subsequent colonies now at least partially resistant to therapy.

With an almost limitless permutation of potential clones that are possible and

identifiable it becomes necessary to classify them by their unique mutations that lead to

recognizable features, both phenotypic and genotypic. These mutations are often labeled

either as founder, driver, or passenger mutations. Founder mutations are the least

common and are responsible, in part, for the founding of a neoplasm colony. Generally at

least two founder mutations are needed for tumor's development from benign to

malignant; one in a tumor suppressor gene which permits the cell to embark in rampant

growth/expansion without self-destruction and one in a growth regulation gene which

provides the stimulus for the proliferation of the cell (Harada & Harada, 2015). Driver

mutations are the activating oncogenic changes that push a colony already containing a

founding clone toward ever more aggressive and resilient expansions. Passenger

mutations are relatively common background mutations and not constrained to cancerous

cells; most long-lived cells, specifically hematopoietic stem/progenitor cells (HSPC), will

collect passenger mutations throughout their life with little to no detriment, although

combinations can create effects similar to driver mutations (Corces-Zimmerman &

Majeti, 2014). When appropriate founder and driver mutations occur in a cell that

suitably increases its fitness, the colony expands and any passenger mutations present are

"captured" to be carried forward with the expanding clone (**Figure 2**). Subsequent driver

mutations that further increase the competitiveness of the host cell will "capture" the new

passenger mutations resulting in new and categorically distinct clonal populations (Welch

et al., 2012). Developing and diverging colonies with differing mutations can create a

challenging scenario for treatment in which a single clonal subpopulation with a

therapeutic-resistance granting mutation survives treatment and leads to relapse.
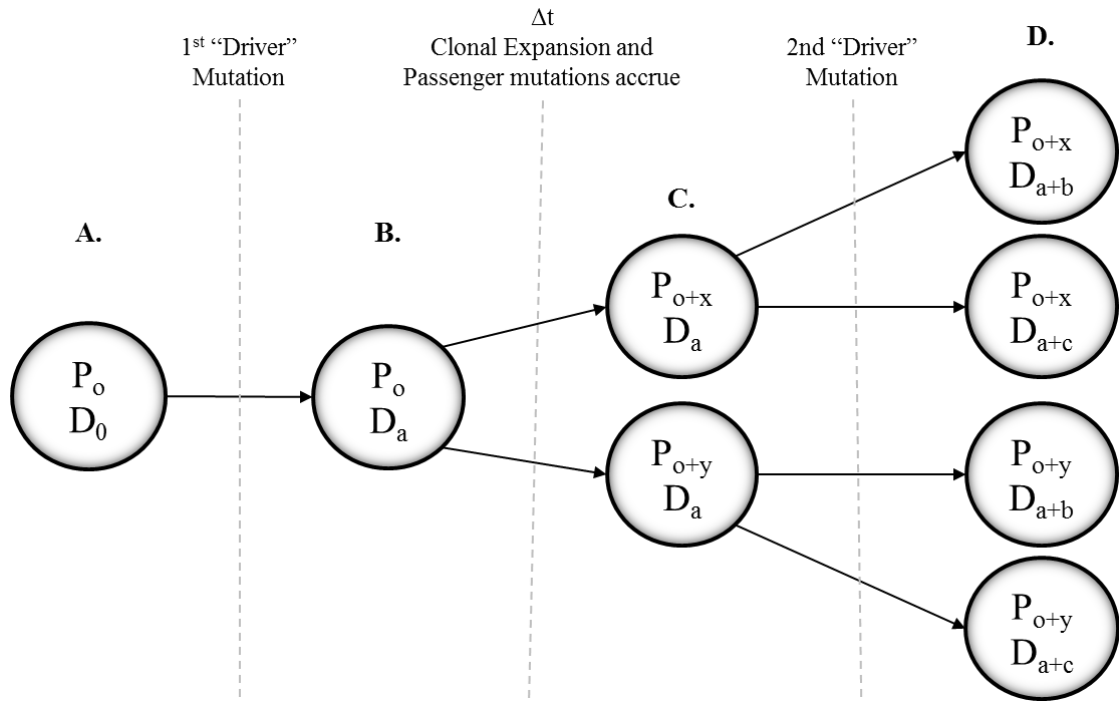
Figure 2. Driver and passenger mutations in clonal diversity. A) Healthy progenitor cells collect benign background mutations throughout their life (P_0). **B**) An initiating driver mutation (D_a) and **C**) clonal expansion "captures" the preexisting mutations and increases their frequency due to the colony's increased fitness with different passenger mutations continuing to accrue (x and y). **D**) Subsequent new driving mutations (b and c) "capture" new passenger mutations and clonal expansion occurs.

Sampling the Bulk Sample and Next Generation Sequencing

The current approach to genomic based cancer treatment has been the sequencing of the cancer genome to detect which mutations are present. This is an expensive and underpowered process. However, the difficulty and expense of next generation sequencing (NGS) has steadily been on the decline, prompting the emergence of new diagnostic genomic tests. Standard NGS testing methods often involve "bulk" input, where large amounts of neoplasm sample are processed as one homogenous set. Unfortunately this method requires the introduction of the following assumptions into the

6

model: **1)** All mutations that occur are heterozygous and the VAF of a mutation is one half of the cellular clonal population. **2)** All mutations that occur are singular. In other words, mutations in a given gene never arise separately in different clonal populations and mutations with higher allele frequencies come from "older" clones that are higher on the evolutionary tree and thus contain the same specific mutations. **3)** Mutations with similar VAF are from the same clonal population of cells (e.g. if FLT3-D835 and NRAS-Q61 both have a frequency of 25% then they are present *only* in the same clones). **4)** All cells of the sample are malignant and no healthy cells were processed. This last supposition can artificially lower mutant allele VAF via allele dilution. Such assumptions vastly simplify the process of modeling a tumor's development. By acknowledging that mutations can be either homozygous or wild-type, not only heterozygous, the range of possible clonal distributions significantly increases. Taken together these assumptions commonly lead to an underestimation of clonal complexity (Paguirigan et al., 2015) (**Figure 3**).

To completely avoid the myriad problems put forth by bulk sequencing and obtain a total view of the clonal diversity present in a patient case it would be necessary to take nigh-infinite single-cell samples, sequence them individually, and plot VAFs for every gene. Such a project would take such an excess of time and resources as to render it infeasible.
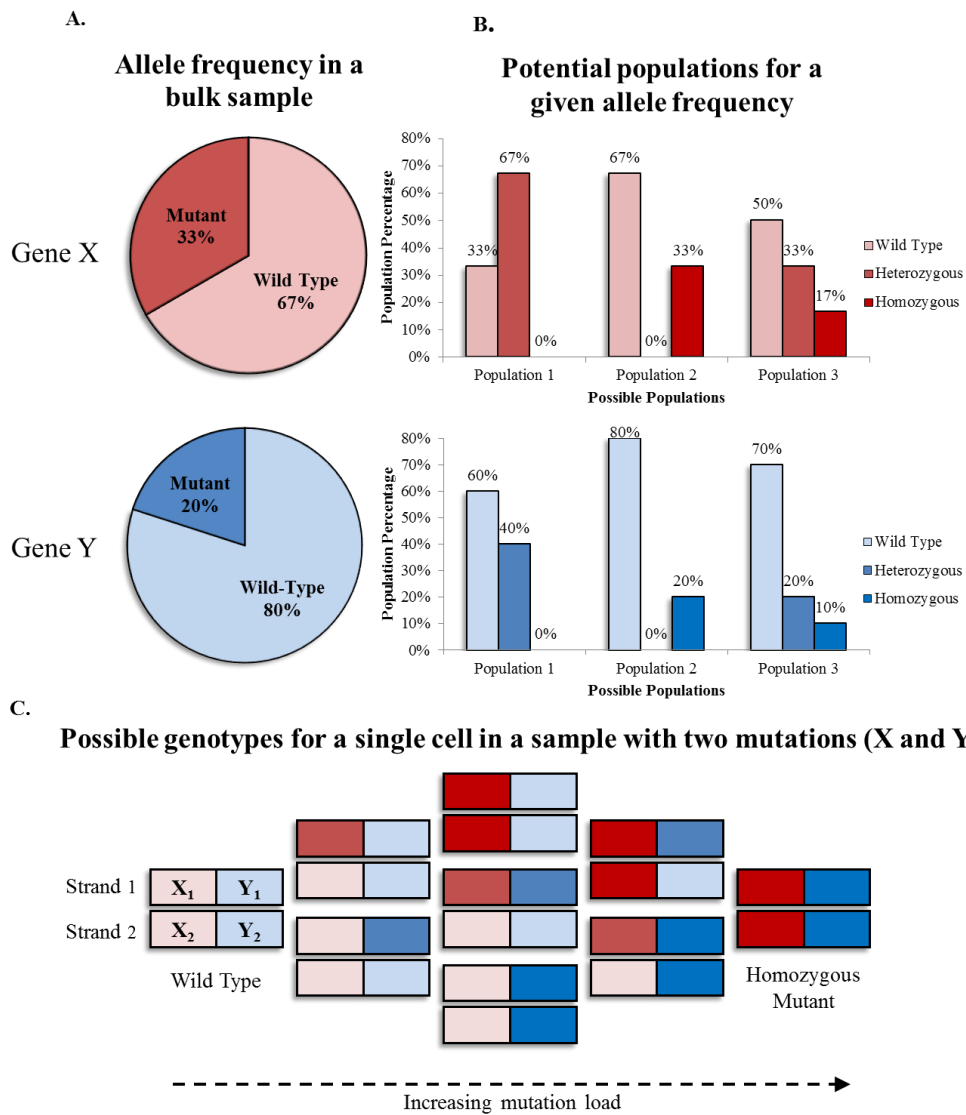
Figure 3. Allele frequency determinations through bulk sequencing coupled with bulk sequencing assumptions can obscure underlying clonal diversity. A) Allele frequencies (AF) determined through bulk sequencing for genes X and Y. The clonal population is assumed to be twice the mutant AF, providing that the mutation is heterozygous. **B)** For any given AF multiple clonal populations are possible when solely considering the potential effects of zygosity on a sample. If the standard assumption that all mutations are heterozygous unless the AF exceeds 50% then population 1 would be assumed, however an infinite potential of population distributions are possible if homozygous mutations are considered. **C)** For any sample with non-zero AF for mutations in X and Y any individual cell may have one of nine mutational states, any or all of which may be a unique clonal populations. These figures are all representative of a simple 2 mutation system; increases in the number of mutations or the presence of replicated genes can complicate the model significantly.

8

Sampling the Clonal Structure

To address bulk and single cell assay limitations, we devised an *in vitro* method which quantitates the allele frequencies of some of AML's "high-profile" gene targets from low cell count, or low-cellular-throughput, samples. By tracking relative allele frequencies between multiple samples and comparing the correlation between mutant alleles one can hypothetically get an accurate representation of the clonal population. To prove the efficacy of this clonality assay's multi-sample/low-throughput approach we show that **1)** it is feasible and **2)** more representative of the real clonal populations than bulk sequencing. Simulating a bulk sample we mixed several established AML cell lines at known ratios and, using a multiplex polymerase chain reaction (PCR) design coupled with pyrosequencing, determined the VAF of specific targets in replicated low-input samples. Through tracking the VAF of the targets through the multiple samples it is possible to determine which mutations are linked together and assess, in part, the clonal populations (**Figure 4**).
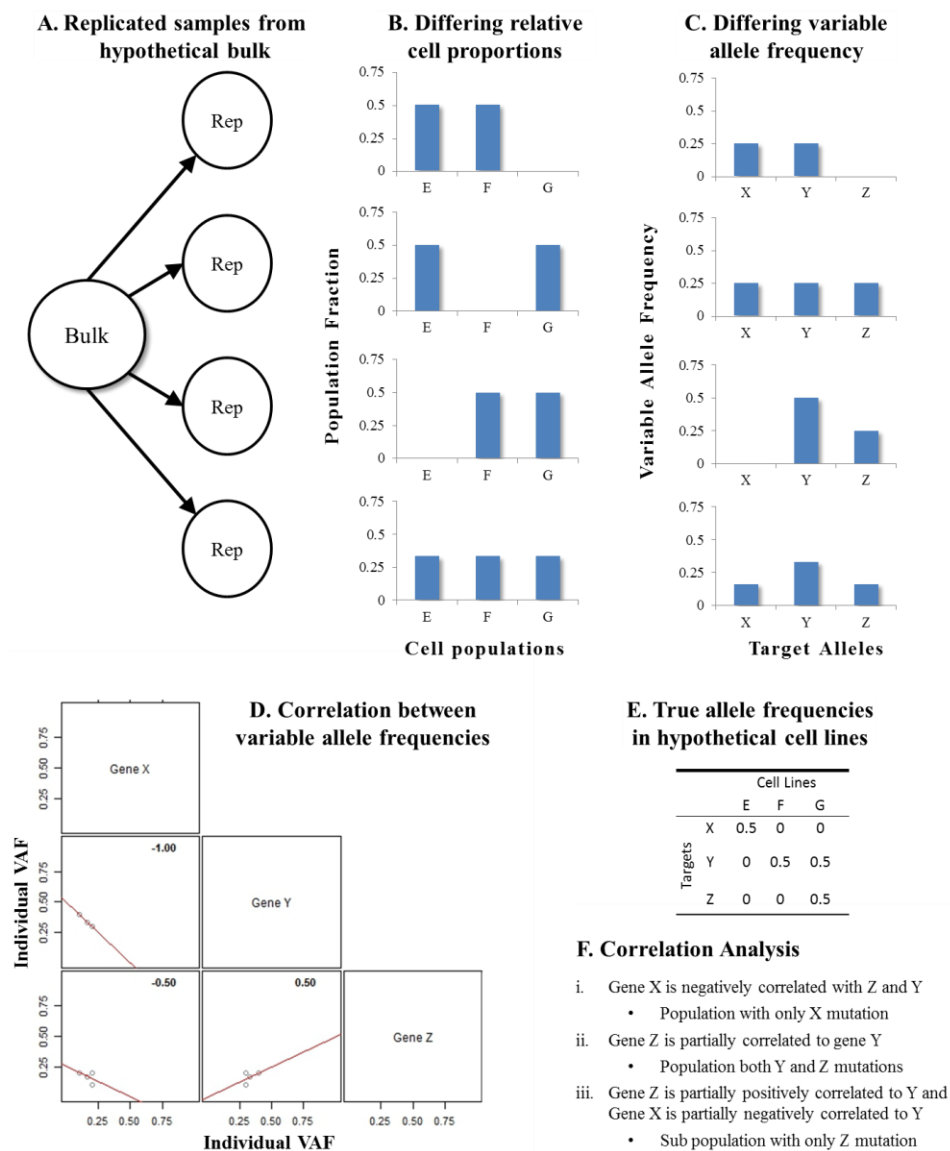
Figure 4. Determination of VAF in multiple low-throughput replicated samples can allow the determination of composition in a hypothetical cellular population. A) Multiple samples with approximately the same cell number are drawn from the bulk. **B)** Each sample will have slightly different proportions of cell lines, or clonal populations. **C)** Differing target allele frequencies of each cell line result in differing variable allele frequencies for each sample. **D)** Plotting variable allele frequencies for every replication against each other yields a correlation between each gene. **E)** The true allele frequencies of the hypothetical cell lines for comparison against the analytically derived clonality structure. **F)** Analysis of correlations reveals 3 different clonal populations, matching the hypothetical bulk. Increasing replications increases the accuracy of clonal population determination by adding more data points while increasing the number of cellular populations will decrease it.

Computer Modeling

As a proof-of-concept and a potential future tool we developed an *in silico*

computer model which simulates a multi-sample/low-throughput cell draw from a bulk

sample. The program, named Cell Sampling, was used analyze the successfulness of the

low-throughput clonality assay by comparing the VAF correlation results of the assay to

the results seen in the program (**Figure 5**).
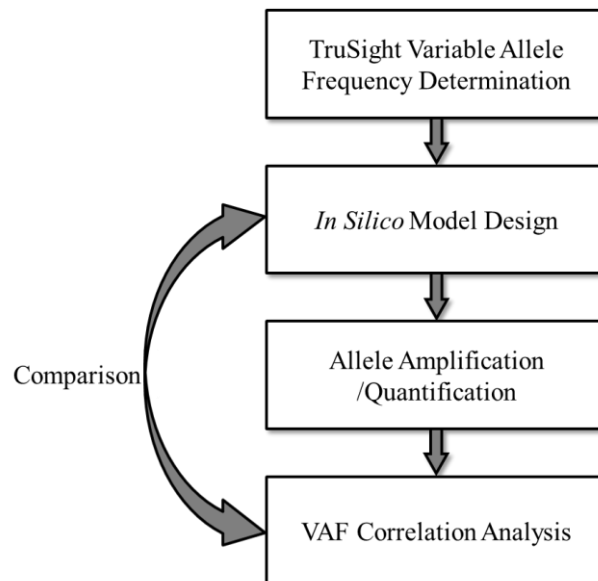
Figure 5. Flow chart of experimental design and implementation. Determination of
allele frequencies in the individual cell lines provides a basis for creating heterogeneous
tumor models with predictable VAFs. Computer modeling of the customized tumors
allow for the optimization of the experimental parameters such as cell line proportions,
sample size and replication number. After testing the computer model and the *in vitro*
assay against the clonal populations and mutant allele frequencies known to be present
the efficacy of both the assay and program in correctly predicting clonal populations and
diversity can be evaluated.

PCR Techniques

Multiplex PCR is a useful method to create numerous amplicons simultaneously

from one source, allowing the selection of several target genes to be amplified for a

relative VAF (Henegariu, Heerema, Dlouhy, Vance, & Vogt, 1997) (**Figure 6**). It ensures

that all amplicons are coming from one homogenous source and the VAF will be

consistent from sample to sample rather than several separate PCR reactions where the

VAF is dependent on the potential random composition differences.
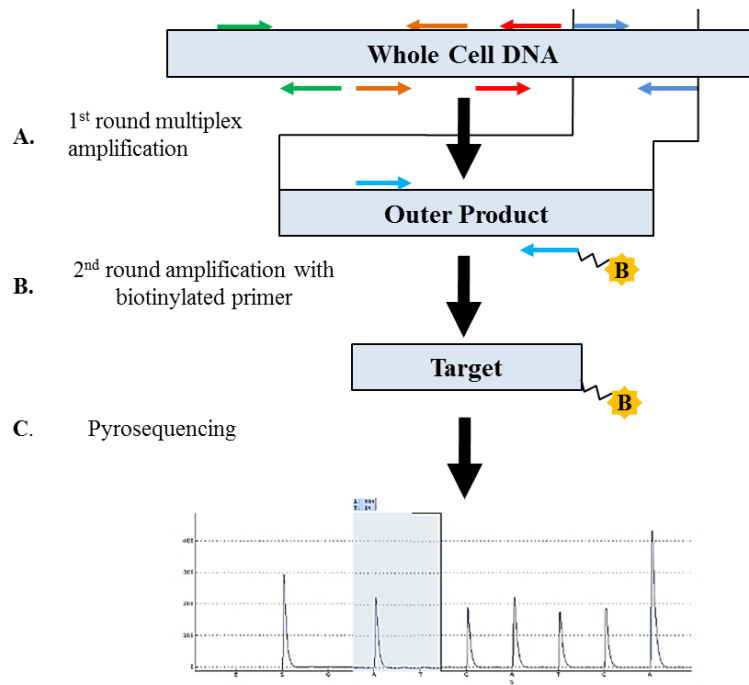
Figure 6. Multiplex/nested/pyrosequencing protocol. A) Whole cellular DNA with
distinct six targets is placed into the multiplex to amplify all targets simultaneously. **B)**
1st round product is separated into 6 different reactions with biotinylated primers for the
nested product. **C)** Nested product with biotin beads are submitted for pyrosequencing.

Nested PCR is a two-step process that can be used to amplify picogram amounts

of DNA; the first-round amplifies multiple large outer targets which are then put into the

second round reaction that individually amplifies smaller inner targets, the compounding

amplification creates sufficient amounts of specific DNA product for genomic analysis.

Through the coupling of multiplex PCR with nested PCR we can separate several targets

from low input concentrations of DNA and amplify the inner targets to sufficient levels

for sequencing. The nested protocol also offers a source for error as well though;

irregular interactions between multiplex amplicons can result in unbalanced target

amplification and non-proportional VAF determination.

Pyrosequencing

Once the nested amplicon products of the target alleles have been produced with

the hybrid multiplex/nested PCR, the sequence can be quantitatively analyzed with

pyrosequencing (Md Fakruddin, 2012). Biotin labels on the inner amplification primers

enable immobilization of the amplicons on sepharose beads that are subsequently

incubated with sequencing primers and a preset sequential series of dNTPs. Nucleotides

which are complementary to the template strand are incorporated to the sequencing strand

in a 5' to 3' direction via Taq polymerase, releasing inorganic pyrophosphate (PPi) from

dNTP. Enzymes present in solution (ATP sulfurylase, luciferase, and apyrase) along with

substrates adenosine 5´ phosphosulfate (APS) and luciferin, generate visible light

proportional to the amount of nucleotides incorporated. Any dNTPs not incorporated are

rapidly degraded by apyrase (**Figure 7**). The proportional nature of the signal observed

allows the calculation of VAF of specific mutations in the target sequence for each

sample.
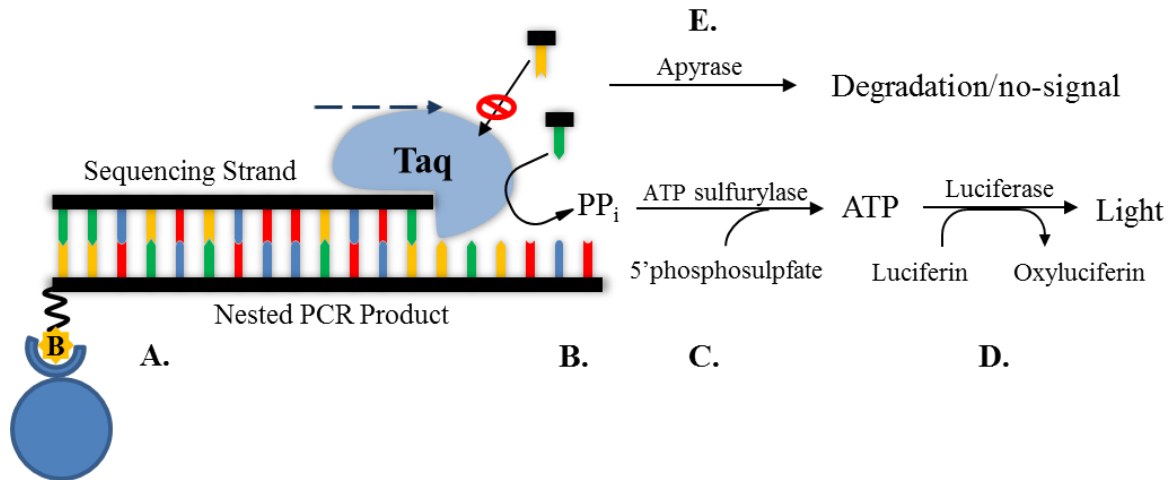
Figure 7. Pyrosequencing. A) Nested PCR product binds immobilizing sepharose beads and sequencing strands bind to the product. **B**) Solution is sequentially incubated with a single dNTP types and Taq polymerase; Taq releases PP_i with the addition of the nucleotide. **C**) ATP sulfurylase incorporates PP_i into 5' phosphosulfate to create ATP. **D**) ATP *in vitro* allows Luciferase to turn Luciferin into Oxyluciferin and light which is observed. **E**) dNTPs that do not match the template strand are not incorporated, degraded by an Apyrase enzyme, and no signal is observed.

Targeted Alleles

The AML cell lines KG1α, OCI-AML3, and MOLM14-L1 were used for this experiment. Circulating, non-adherent AML neoplasm cells provide an easy sampling model system for proof-of-concept research and the immortalized cell lines are generally robust. The selected targets for this clonality from these cell lines were the genes/codons DNMT3α-R882, EZH2-G628, FLT3-D835, FLT3-ITD, NRAS-Q61, STAG2-L526, and TET2-V218. The VAF for all targets with the exception of FLT3-ITD was determined through the multiplex/nested reaction coupled with pyrosequencing. FLT3-ITD was unique among these targets--it is a duplication mutation instead of a single nucleotide polymorphism. Tandem duplications of a FLT3 codon increase the length of the

14

amplicon produced in the PCR. Quantitative determination of the ITD VAF was calculated by comparing the proportion of tandem duplication FLT3 to wild-type FLT3.

Summary

Bulk analysis is currently one of the most used methods of genetic analysis in AML treatment and oncology today. And while next generation sequencing of the bulk sampling may underestimate clonal complexity, strictly single cell analysis raises several concerns that must be addressed prior to widespread clinical use (such as time, resources and efficacy). We hypothesize that our strategy of low cell count input, DNA amplification, and analysis of the variable allele frequency will be capable of detecting clonal structure, which cannot be determined from bulk analysis of a simulated AML patient. If successful, this strategy would not only support caution in the evaluation bulk analysis results, but also potentially offer a third option of clonality analysis for future clinicians.

METHODS

Cell Lines

Three AML myeloblast cell lines were selected to simulate a bulk sample; KG1α, MOLM14-L1, and OCI-AML3. KG1α is a bone marrow myeloblast of a patient with AML, MOLM14 was derived from the peripheral blood of a patient who initially had myelodysplastic syndrome and the relapsed with full AML, and OCI-AML3 is a derivative from peripheral blood in an AML patient. The cells were separately cultured in Roswell Park Memorial Institute medium (RPMI) with 20% fetal bovine serum (FBS).

Primer Design

Sequences for the outer primers were found with the NCBI's Primer-BLAST (Ye et al., 2012) and optimized for consistent and ideal annealing temperatures using OligoCalc (Kibbe, 2007). The resulting primers were run in NCBI's nucleotide BLAST program to verify that the primers targeted the specified genes (Altschul, Gish, Miller, Myers, & Lipman, 1990). Finally all seven primer pairs were put through simultaneous *in silico* PCR with MFE-Primer 2.0 to ensure no cross amplification or dimer formation (Qu et al., 2012). The primers were was then ordered through Integrated DNA Technologies (IDT) (**Appendix 1**). Inner biotinylated primer and pyrosequencing primer designs were generated seperatelywith Pyromark by Qiagen and also ordered through IDT (**Appendices 2 and 3**).

16

Flow Cytometry

Cell lines were previously analyzed using multiparameter flow cytometry to identify distinguishing cell surface markers. We identified that KG1α is CD34+/CD14-, MOLM14-L1 is CD34-/CD14- and OCI-AML3 is CD34-/CD14+. Cell lines were counted with a hemacytometer and pooled at an estimated fraction of 1/3 of each cell line with a total population of 3 million cells. The mixture was spun down at 200rcf for 5 minutes and the supernatant replaced with 1x PBS+0.1% BSA and incubated with antibodies from BD Biosciences (10uL/million cells) for CD14 and CD34 for 15 minutes at room temperature (**Table 4**). The cells were then washed with 1x PBS from Life Technologies with 0.1% bovine serum albumin (BSA) from Life Technologies, pelleted at 300rcf and resuspended in 1x PBS + 0.1% BSA along with calcein violet (1ul/1ml PBS) and incubated for 20 minutes. Viable cells were sorted into 384 well plates at 120 cells per well, and the true population frequency of the three cell lines was quantified via the cell surface marker gating and the plates were frozen at -20°C.

Multiplex PCR

First-round outer product was generated with 1x Platinum Taq and Platinum Taq PCR Buffer from Invitrogen, 2.5mM $MgCl_2$, 400nM dNTP, and 50nM of each individual primer. PCR conditions were 95°C for 5 minutes followed by 30 cycles of 95°C for 20s, 57°C for 45s and 72°C for 45s followed by a termination extension of 72°C for 5 minutes. The presence of bands was initially confirmed with ethidium bromide (EtBr) on a 2% agarose gel at 45 cycles before lowering the number of cycles to 30 for the two-step nested procedure.

Nested PCR

Inner round product was created with either biotinylated forward or reverse primers and a non-labeled complement. 2μL of first round product was added to each separated inner target reaction which consisted of 1x Platinum Taq and Platinum Taq PCR buffer, 1.5mM $MgCl_2$, 200nm dNTP, and 200nM of the labeled primer pair. Inner product PCR conditions were 95°C for 5 minutes followed by 35 cycles of 95°C at 15s, 57°C for 30s and 72°C for 30s. The presence of bands was again confirmed on a 2% agarose gel and EtBr.

Pyrosequencing

Product of the nested PCR reaction was submitted to in-house pyrosequencing on a PyroMark Q96 instrument from Qiagen present at the Fred Hutchinson Cancer Research Center along with 10mM aliquots of the pyrosequencing primers. Pyrosequencing results were analyzed using a tailored R-Script program related to the Cell Sampling program.

Whole-Cell Input

Whole cell PCR was tested prior to the pooling of the cell lines to ensure effectiveness of the primers in a whole cell setting. MOLM14-L1 cell lines were grown in FBS and diluted with PBS to form stocks of 500, 50, and 5 cells per microliter. 2μL of the cell stocks were then added to 13μL molecular biology grade H_2O (Mediatech), frozen at -80° C, heated to 95°C for 5 min, and then put on ice immediately prior to PCR. 10μL of Master Mix (1x PCR buffer, 2.5mM $MgCl_2$, 400nM dNTP, 0.05μM

multiplex primers, and 1x Platinum Taq) was added to the lysate. First-step multiplex

PCR was done at 95°C for 5 min, 40 cycles of 95°C for 30s, 57°C for 45s, 72°C for 45s

and a final extension at 72°C for 5min. The product was run on a 2% agarose gel with

EtBr.

Sanger Sequencing

Samples of the inner targets were prepared for Sanger sequencing with PCR

amplification from each cell line and purified using the ExoSAP protocol by

ThermoFisher Scientific. The product was submitted to be sequenced in-house on an

Applied Biosystems (ABI) 3730xl DNA Analyzer at the Fred Hutchinson Cancer

Research Center. The results were viewed using FinchTV by Geospiza.

Fragment Analysis

To analyze the relative level of FLT3-ITD in each replication a sample of each

replication was analyzed with DNA fragment analysis by capillary electrophoresis.

Samples from individual replications were input into a PCR reaction (1x Platinum Taq

and Platinum Taq Master mix, 50mM $MgCl_2$, and 10mM dNTP) with 10μM FLT3

reverse primer and 10μM FLT3 forward primer labeled with a FAM fluorescent tag

which absorbs at 492 nm and emits at 517 nm. The PCR product was diluted 1:10,000

and submitted to in-house genomics analysis via ABI 3730xl DNA Analyzers.

Cell Line DNA Titrations

Titrations of cell line DNA were produced by diluting that DNA (KG1α,

MOLM14, or OCI-AML3) in a differing cell line DNA (KG1α, MOLM14, or OCI-

AML3) which did not contain a specific mutation while maintaining a constant DNA input of 40ng.

Data Analysis

The program Cell Sampling was created in R-script for this project and utilizes the user-defined cell line parameters along with the number of replication of cell draws designated to simulate both bulk allele frequency and allele frequency correlation coefficients, identifying mutations which track together (**Appendix 4**). The program first creates a bulk cell count that simulates a whole tumor, randomly selects a set number of cells, categorizes the cell types, uses the known VAF for each cell line to calculate the total number of each mutant allele, and finally divides it by the total number of alleles (**Figure 8**). The VAF of every target mutation is then plotted against every other mutation for each individual replication. The correlation between each two mutations across replications is analyzed with the Pearson Pairwise method to determine the level of association between mutant alleles. The cell sampling program also repeats the experiment for a range of cell inputs to identify the ideal cell count that minimizes the mean correlation between unlinked loci in order to maximize observable correlations between linked loci. Finally Cell Sampling also sums all replicates' VAFs to determine the bulk VAF for comparison (**Appendices 5 and 6**).

Due to the conflicting considerations of the restricted capacity of the PyroMark pyrosequencing instrument and the objective of maximizing correlation trends a median of twenty replicates were performed. Cellular populations were approximately 33.3% of each cell line (34.3% KG1α, 38.4% MOLM14-L1 and 27.3% OCI-AML3, reflecting true

proportions as counted by flow cytometry). A total population of 3 million cells was used

and the VAFs were set equivalent to values found with Illumia's TruSight (**Table 4**).

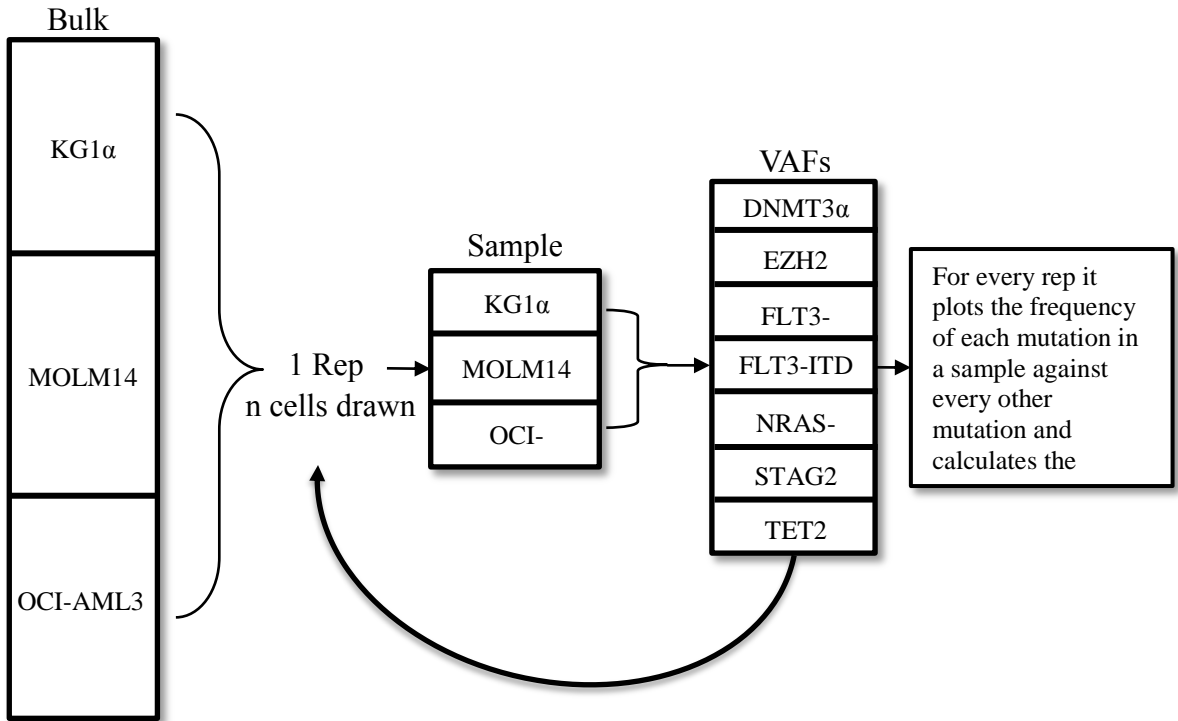**Figure 8. Diagram of program function-Cell Sampling.** n cells are drawn from a
simulated tumor and counted. A VAF for each target is determined from the number of
each cell line for that sample. Replicate samples are pulled from the bulk and VAFs are
determined for each sample. Cell Sampling then plots the allele frequencies against each
other for every replicate and calculates the correlation between every pair.

RESULTS

Establishment of VAF in Cell Lines

Mutant allele frequencies of each cell line (or shared in two) were previously

established through next generation sequencing via Illumina's TruSight Myeloid

Sequencing (**Table 1**) ("TruSight Myeloid Sequencing Panel," 2015). The sequencing

revealed that the KG1α line contains the mutation TET2-V218M (GTG>ATG) at 50%

VAF, EZH2-G628 (GGC>AGC) at 100% VAF. MOLM-14-L1 cells contain TET2-

V218M (GTG>ATG) at 50% VAF, the FLT3-D835 mutation at 50% VAF as well as a

21bp long FLT3-ITD. Finally, OCI-AML3 contains the DNMT3α-R882 mutation at 50%

VAF, the NRAS-Q61 mutation (CAA>CTA) at 100% VAF and STAG2-L526F

(CTT>TTT) at 25% VAF.

Table 1. Variant allele frequencies of seven targets present in our three cell lines and the antibody markers used for identification. VAFs were found through next gen sequencing (Illumina).

Cell Line	DNMT3α	EZH2	FLT3-D835	FLT3-ITD	NRAS-Q61	STAG2	TET2	Marker Tags
KG1α	0%	100%	0%	0%	0%	0%	50%	CD34+/CD14-
MOLM14	0%	0%	50%	50%	0%	0%	50%	CD34-/CD14-
OCI-AML3	50%	0%	0%	0%	100%	25%	0%	CD34-/CD14+

Optimization of Multiplex PCR

We first designed and tested assays which would selectively amplify target genetic

products and exhibit no unintended products. Multiplex primers designed with

PrimerBLAST were then analyzed *in silico* for their specificity in a multiplex setting

using MFEprimer-2.0; no extraneous amplicons were observed and bands were seen to be

discrete enough to proceed to testing (**Figure 9**). The primers were validated in

singleplex PCR to generate a single, specific band (**Figure 10**), and the multiplex

displayed all seven target amplicon bands for both DNA input and whole cell input

(**Figure 11** and **Figure 12**).
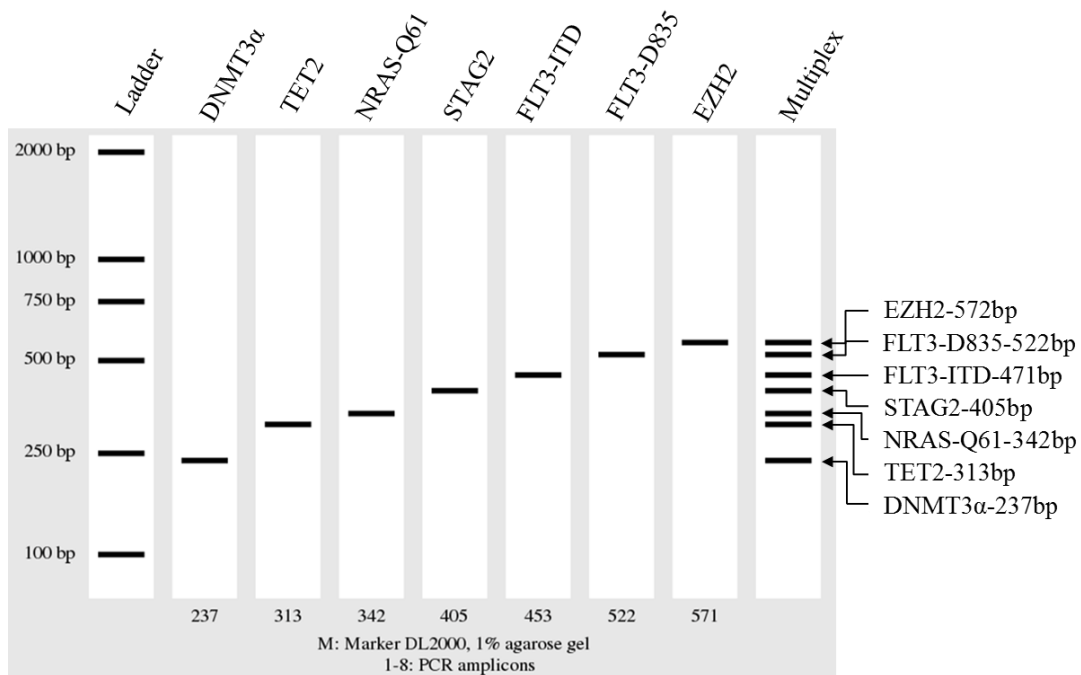
Figure 9. *In silico* **PCR of multiplex PCR reaction, generated with MFE-Primer 2.0.** Primer sequences from **Appendix 1** were input into the internet-based application with PCR conditions identical to the *in vitro* multiplex reaction.

23
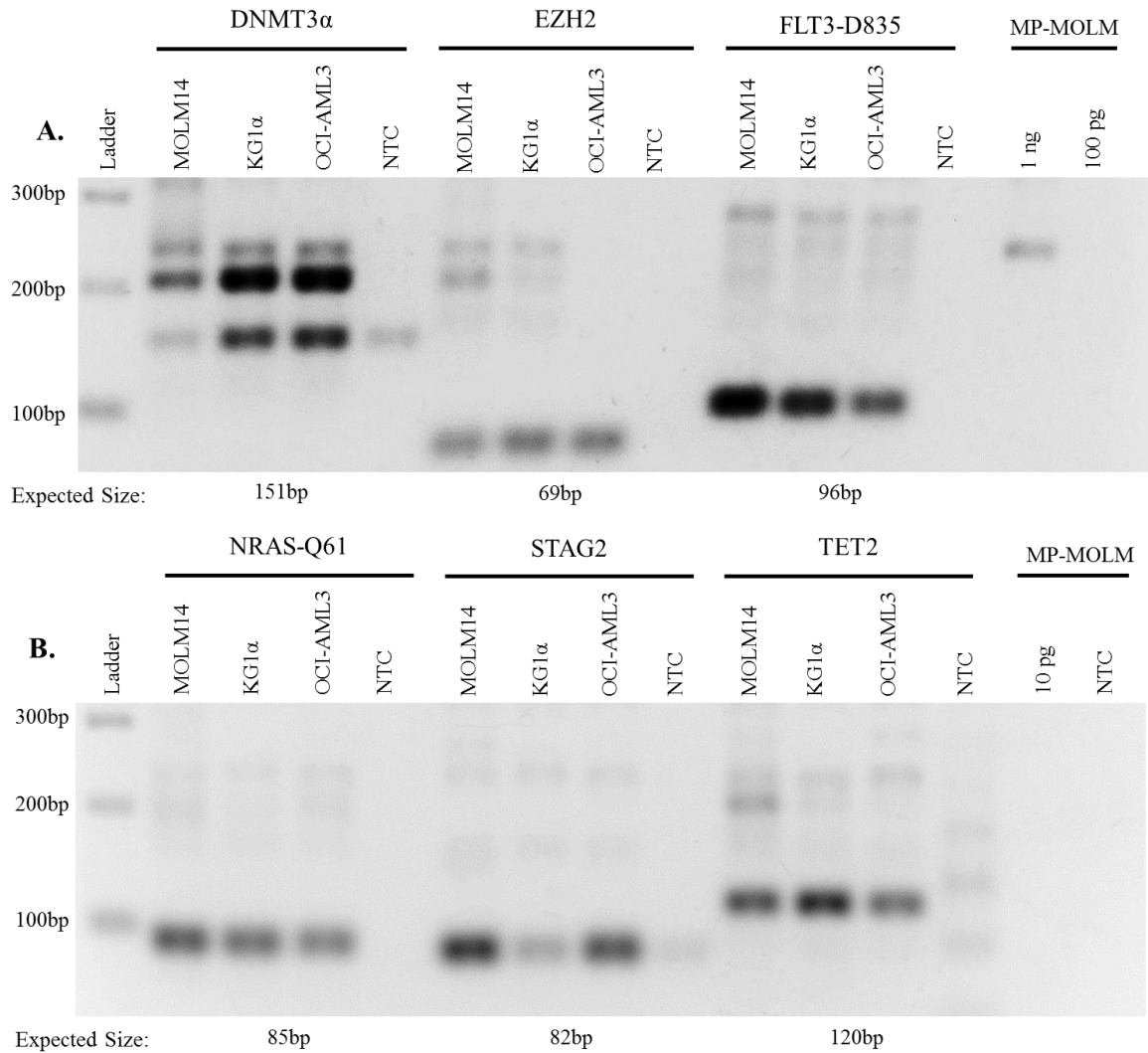
Figure 10. Nested product from second reaction produces anticipated bands. All primer sets were sufficiently specific, except for the DNMT3α primer pair, which exhibited large amounts non-specific amplification along with the expected.
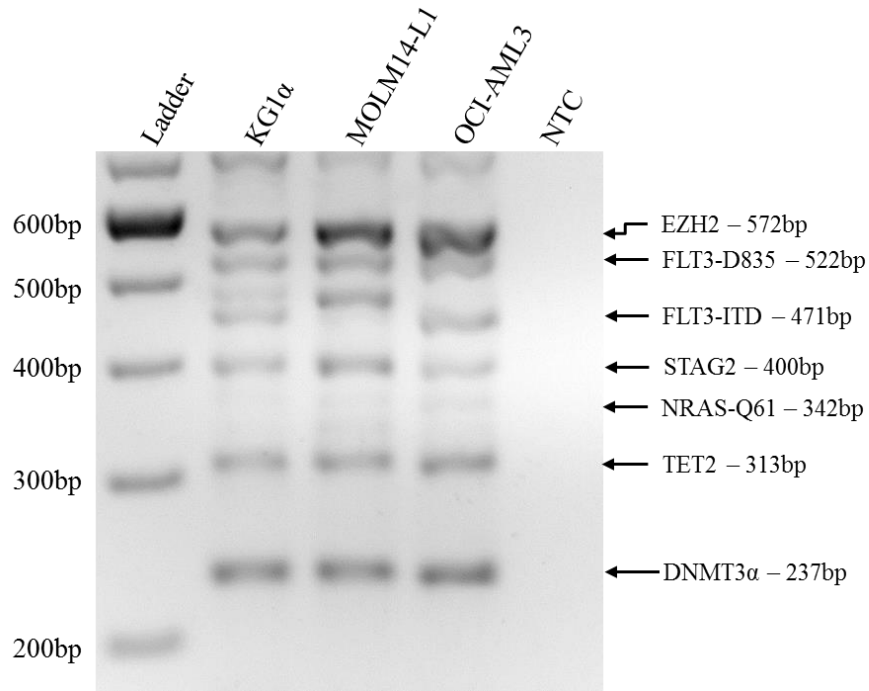
Figure 11. Multiplex of all three cell lines each produce seven target amplicons *in vitro,* matching *in silico* prediction. MOLM14 has more ITD duplications than KG1a or OCI-AML3 which increases its product size. NRAS-Q61 displays an evident but faint band.
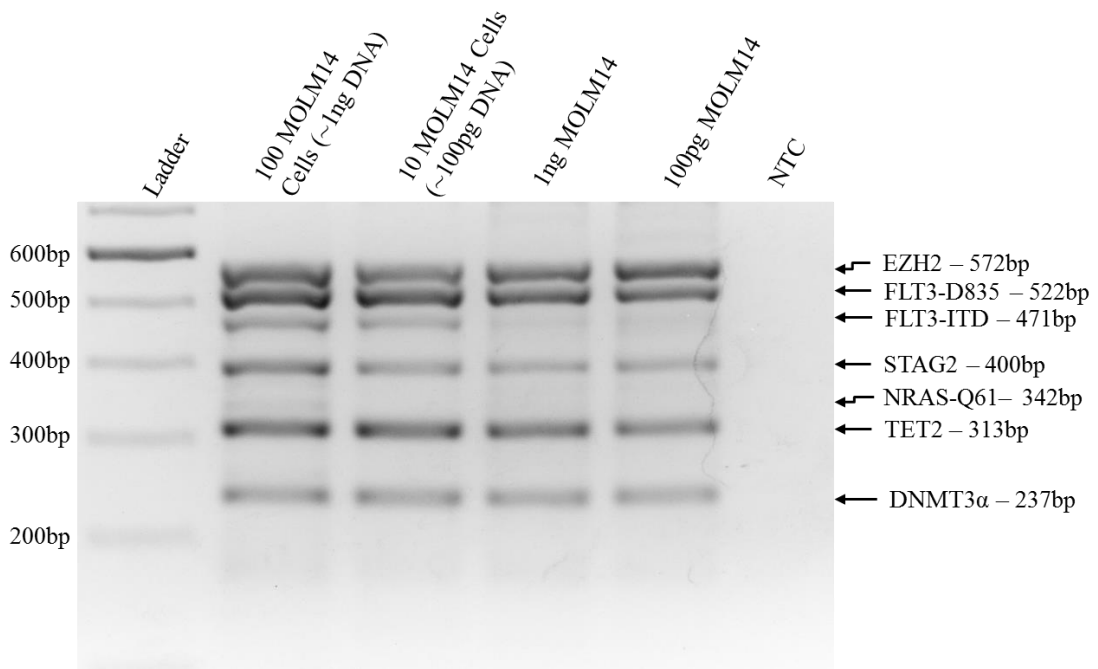
Figure 12. Whole cell input to multiplex PCR yields amplicons identical to DNA input.

Validation of Nested Primers and Targets

We next confirmed that the inner/nested primer pairs amplified the correctly sized product for both DNA-only and first-round reaction input for all targets in all three cell lines (**Figure 13**). Nested reactions with first-round input did yield a product that contained slightly more undesired amplicons, namely DNMT3α which showed distinct signs of cross amplification between unused first-round primers and second-round inner primers, but this was expected with nested reactions and minimized by decreasing the first-round cycle number and diluting the product transferred from first to second rounds.
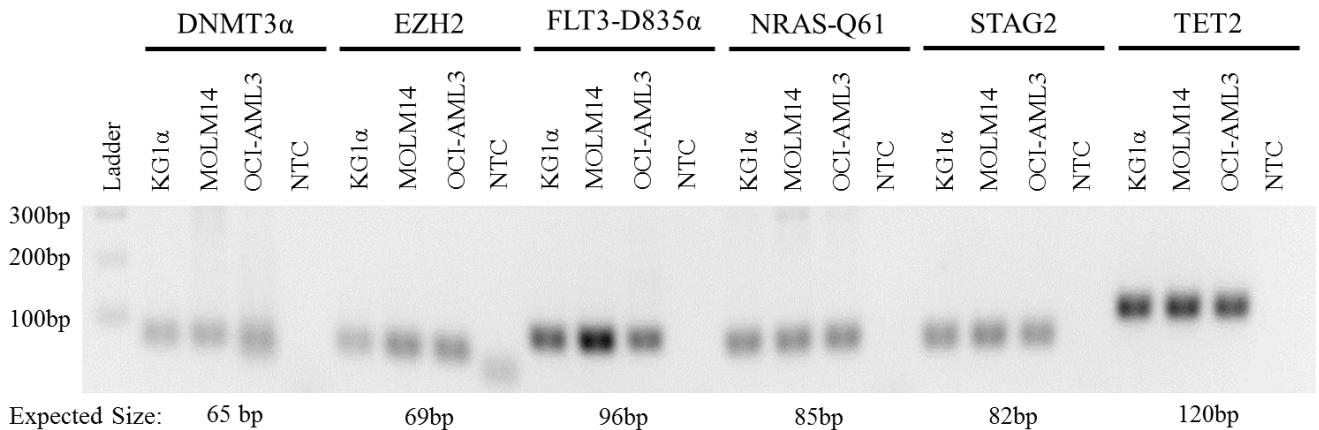
**Figure 13. Non-biotinylated pyrosequencing primers produce anticipated bands with whole DNA input.**

Product from the direct DNA nested primer validation was submitted for in-house pyrosequencing and displayed anticipated cell line specific allele frequencies with the exception of EZH2 and STAG2 which displayed no mutations at all. To verify the mutational status of all seven genes in the cell lines the targets were sent in for Sanger sequencing (**Figure 14**). All anticipated mutations were present in expected cell lines but

26

again with the exception of EZH2 and STAG2, which were subsequently dropped from the multiplex and continuing experiments.
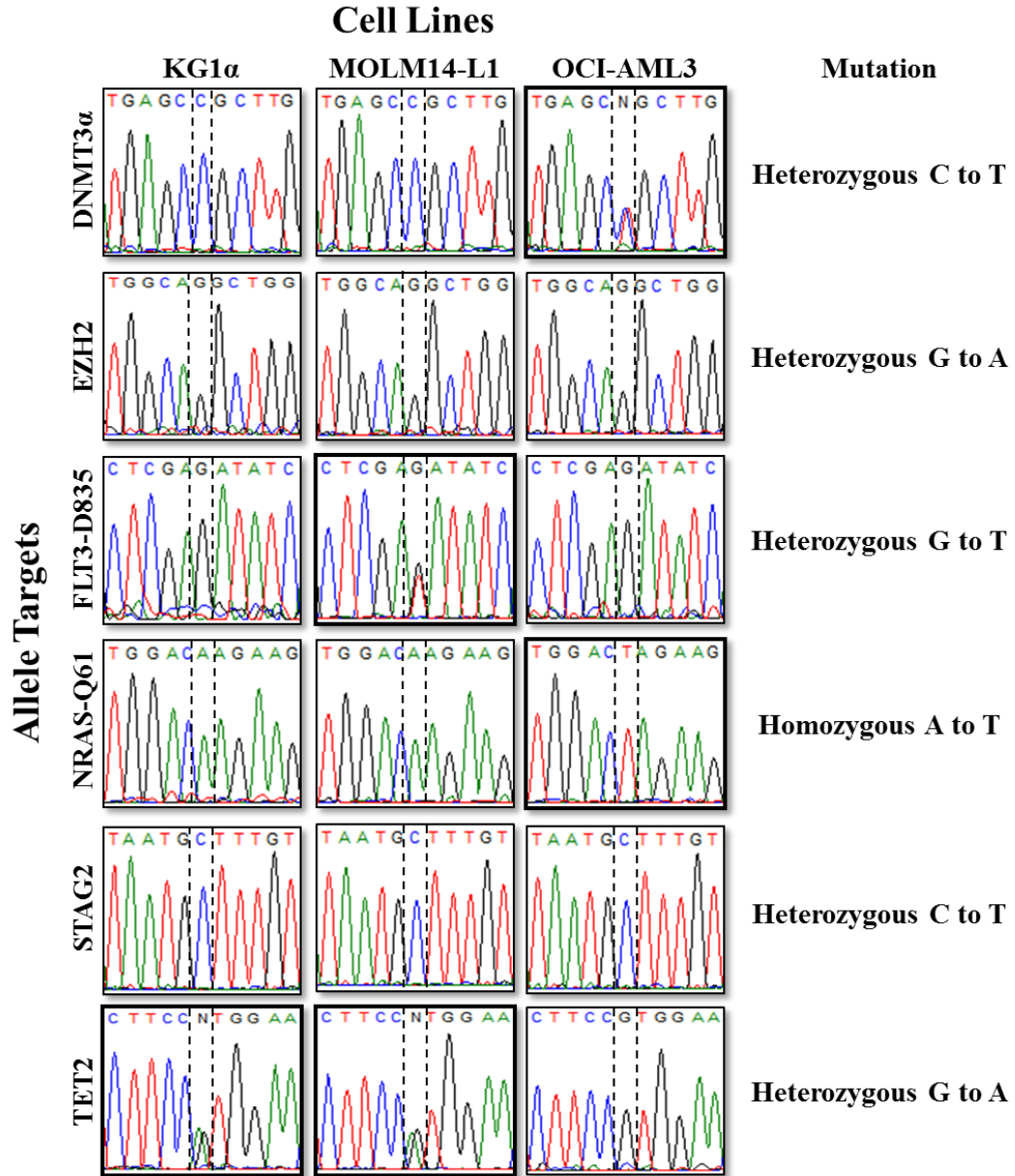
Figure 14. Sequences of target codons for the three model cell lines found through Sanger sequencing. The outlined nucleotide in each sequence represents individual target mutations. Bolded boxes represent mutant alleles. All possible mutations in each allele are not indicated, just mutations present through previous NGS.

Cell Sampling

R studio program Cell Sampling returned an optimum cell draw number of about 120 cells (**Figure 15**). A mean correlation between VAFs of unlinked loci over 20 reps was -0.36 when 120 cells were then drawn randomly from the bulk *in silico*, the allele frequencies for each allele in each cell line were set, and the VAF for each locus was determined for each replicate (**Figure 16**, **Appendix 4**).
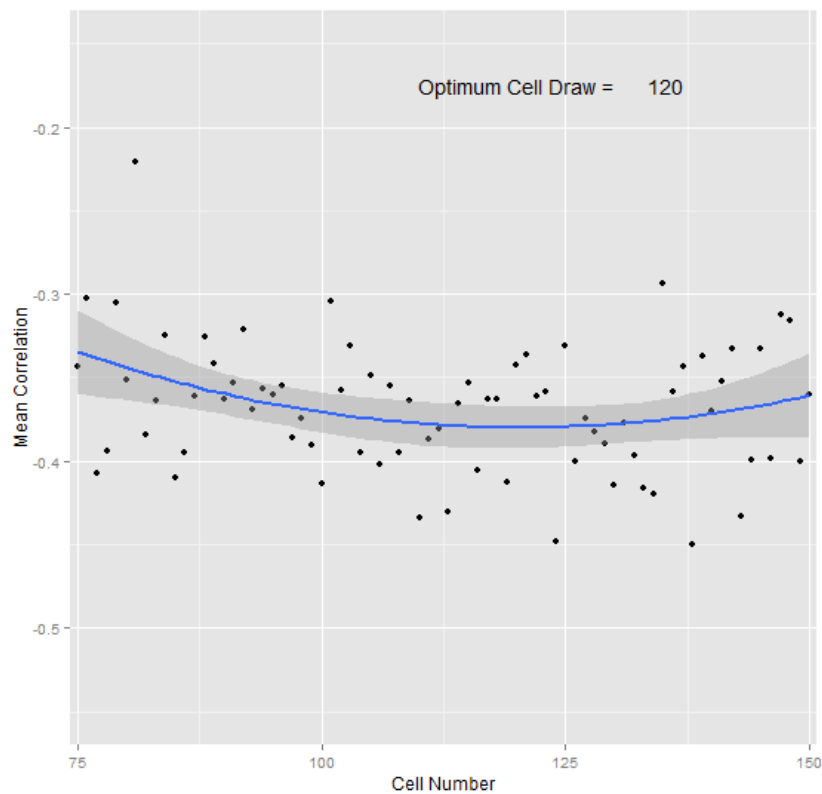
Figure 15. Determination of the optimum cell draw number via *in silico* cell draws. Mean correlation between unlinked loci was calculated with incremental numbers of cell drawn from the bulk to determine the cell number with the lowest correlation --and conversely the highest correlation between linked loci--.
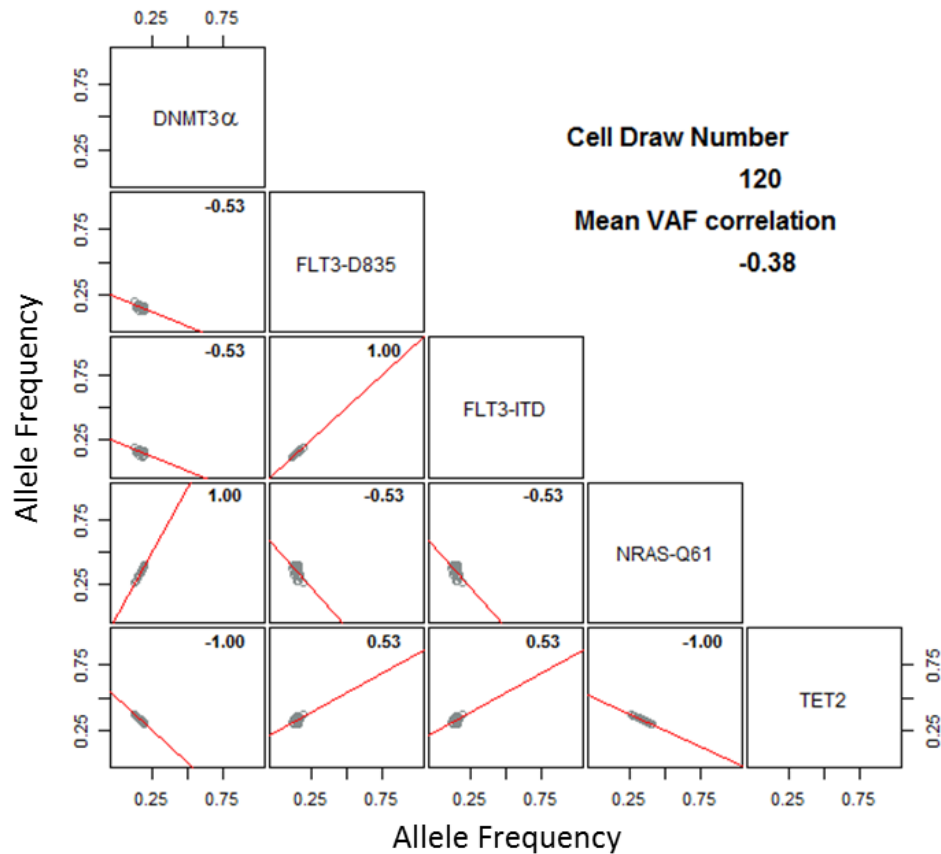
Figure 16. VAFs plotted *in silico* against each other for every replication with a calculated correlation. "Perfectly" linked or unlinked loci have a theoretical correlation of ±1 while partially linked or unlinked loci will have a partial correlation of less than |1|. The VAFs were determined from sampling a simulated cell population of three million cells with approximately one million cells from each of the three cell lines.

Fragment Analysis

A VAF for FLT3-ITD was determined by dividing the relative fluorescent intensity of the FLT3 peak with the ITD mutation (present at 351bp) by the intensity of the total FLT3 peaks (peaks at 330 bp and 351 bp) (**Figure 17**) (**Table S4**).
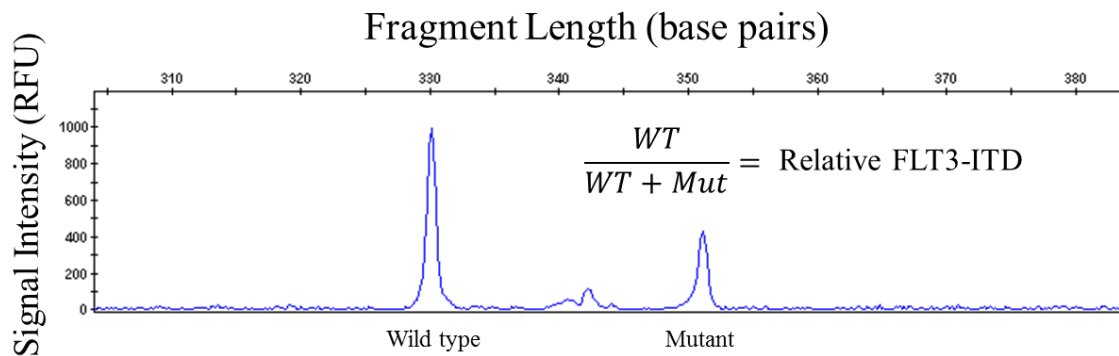
Figure 17. Example of VAF determination for FLT3-ITD. The FLT3-ITD mutation incorporates a seven times codon duplication, lengthening the allele by 21 base pairs, or from 330 bp to 351bp.

Pyrosequencing

To verify that the pyrosequencing assays are quantitative, mixtures of DNA from cell lines were sequenced for specific mutations. DNMT3α and NRAS-Q61 were verified by decreasing OCI-AML3 and increasing MOLM14-L1, FLT3-D835 was tested through increasing MOLM14 while decreasing KG1α and TET2 was verified through increasing KG1α while decreasing OCI-AML3 (**Figure 18)**.
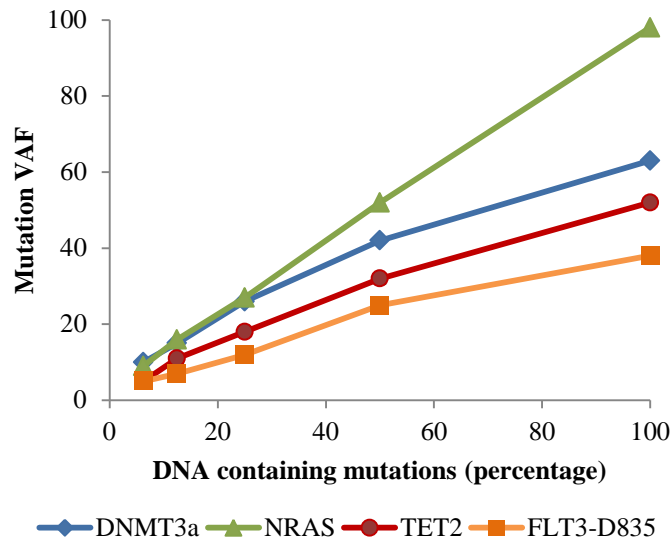
Figure 18. Quantifiable nature of pyrosequencing. Increasing the proportion of input DNA which contains designated mutations correspondingly increases the detected variable allele frequency. NRAS-Q61, a homologous mutation, approaches a 100% VAF while DNMT3α, FLT3-D835, and TET2 are heterozygous and approach 50% VAF. A constant DNA input of 40ng into each 2nd round (inner) reaction was used with titrations of increasing cell line DNA (OCI-AML3 for DNMT3α and NRAS-Q61, MOLM14 for FLT3-D835 and KG1α for TET2 all of which were diluted with differing cell line DNA which did not contain the mutation.

The simulated tumor population generated through cell line mixing was analyzed for the target allele frequencies. The twenty replications of the cell line pool were partitioned through flow sorting and the cell line proportions were determined to be 34.2% KG1α, 38.4% MOLM14-L1, and 27.3% OCI-AML3, the percentages were used to update the program Cell Sampling for more accurate predictions *in silico*. Five of replications failed pyrosequencing due to signal errors but the remaining 15 passed (**Figure 19** and **Appendix 7**). The aggregate allele frequencies in the pooled cell lines -

similar to what would be seen in a bulk sample- can be estimated *in silico* based on the

established VAFs and the proportions of the cellular populations.
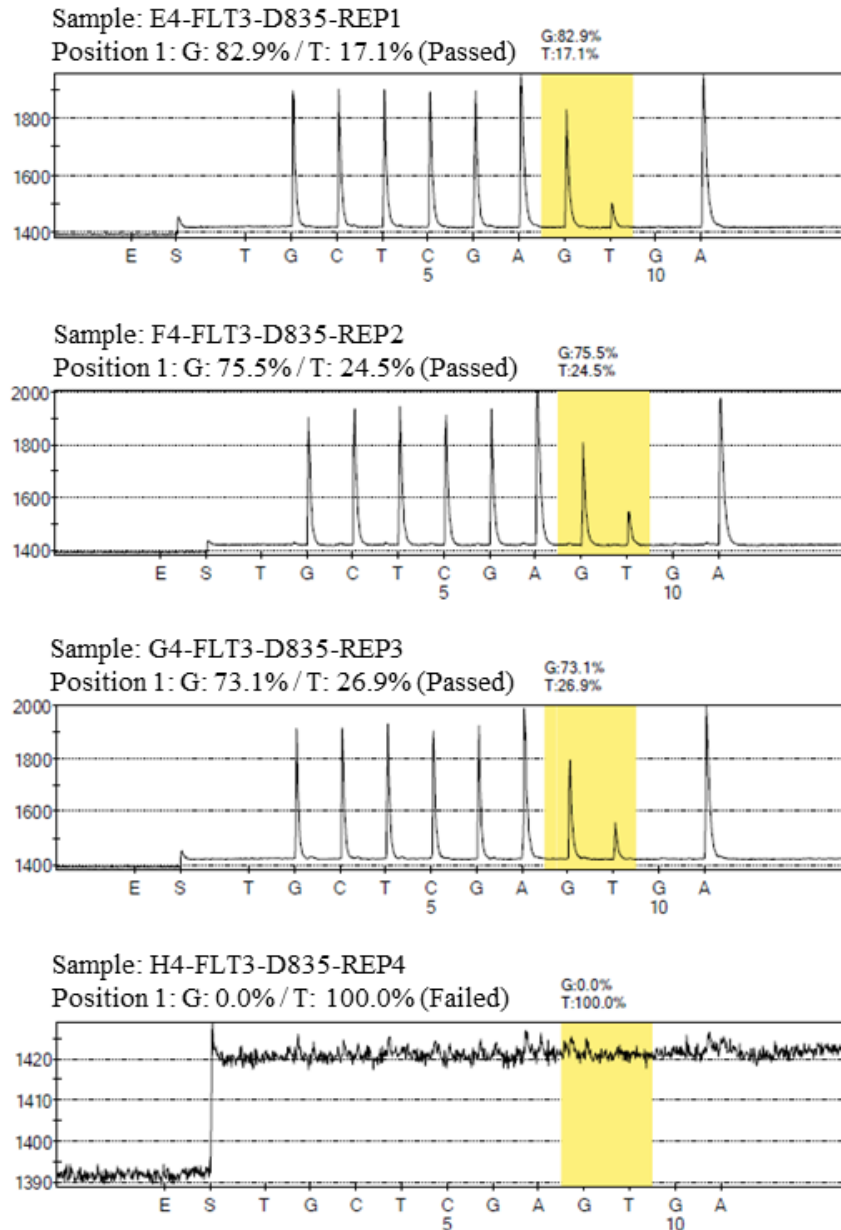
Figure 19. FLT3-D835 Pyrographs of replicates 1-4 out of 20 total. VAFs determined through pyrosequencing, in these examples guanine addition to the pyrosequencing primer denotes wild type alleles while thymine additions implies mutant alleles. The percentage denoted at Position 1 represents the VAF for a specific allele in a sample. Replicate 4 failed due to signal error.

Correlation Analysis

The VAFs determined from the successful pyrosequencing were then plotted against each other for each replicate and the correlations between each allele were determined using the Pearson-Pairwise method (**Figure 20**).
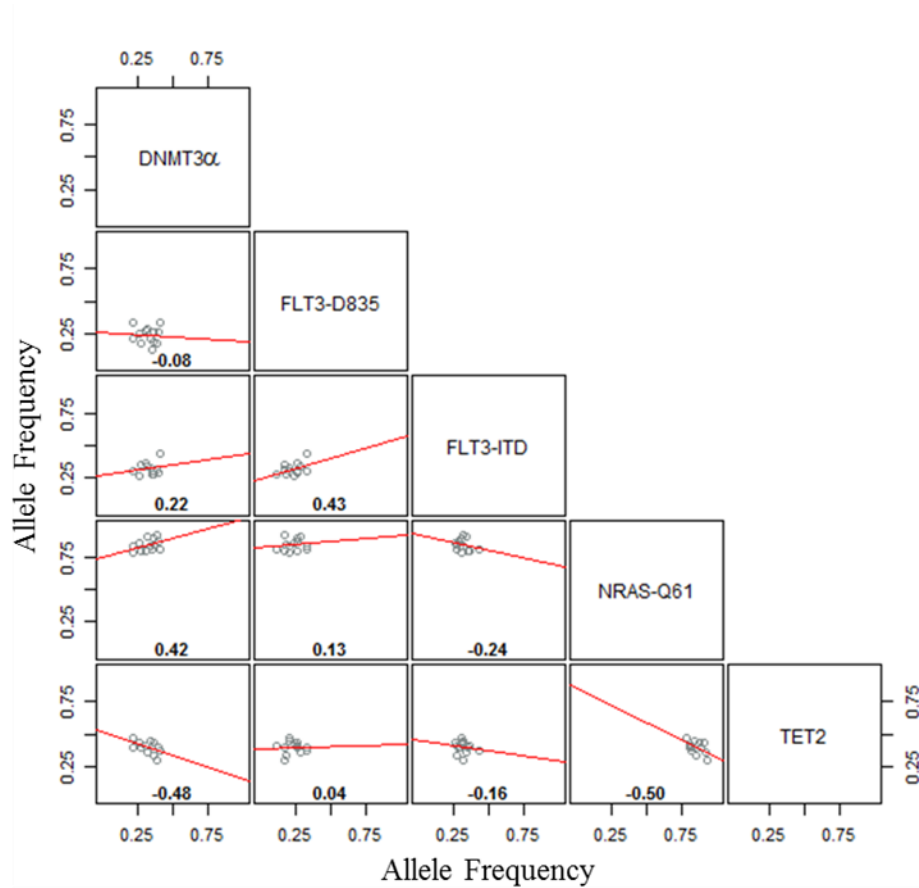
Figure 20. Experimentally derived mutant allele frequencies across replications. Correlation values between replicates were determined through the Pearson-Pairwise method.

DISCUSSION

While clonal diversity appears to be a prospective measurement of the malignancy of cancer, its evaluation presents difficulties in clinical implementation. Time and effort factors support high-throughput bulk sequencing but, while the assumptions are often superficially accurate and useful for a cursory evaluation, it ultimately mis-defines clonal diversity and the sub-clonal populations (Hughes et al., 2014). The goal of our project is to primarily illustrate the possible information that is overlooked by bulk NGS and secondarily to propose, through *in silico* and *in vitro* proof-of-concept models, a possible method of assessing the clonal diversity.

Next Generation Sequencing Anomalies

The seven mutant allele targets chosen for this low-throughput clonality assay were found with Illumina's TruSight Myeloid Sequencing Panel performed on each individual cell line. Subsequent Sanger and pyrosequencing revealed, however, that the mutations EZH2 and STAG2 were not present in the leukemia cell lines we used and thus were not appropriate markers for tracking cellular populations. The allele targets with their sequencing and amplifications primers were dropped from further experiments. Sequencing panels are useful for genetic analysis of a broad range of targets but Sanger sequencing offers a much more robust and target specific method of specific codon evaluation. With only five targets remaining in the low-throughput clonality assay, specificity, and the ability to pick out clonal populations, was decreased. This outcome

demonstrates that next generation sequencing data, even when it meets quality control criteria, may need to be viewed as preliminary and separately verified.

Bulk Variant Allele Frequency Evaluation

The "bulk" approach is to use the total allele frequencies of a sample to make statements on the clonal populations. However, these clonal assertions based on NGS data are one dimensional guesses and offer little opportunity for extrapolation. Seemingly identical frequencies which may imply common allele mutations within a clone may in fact just be happenstance. This effect can be seen in the approximated bulk VAFs, connections between various alleles can be guessed at but have no supporting evidence (**Appendix 6**). Trying to establish clonal populations through *in vivo* bulk NGS would complicate the task by likely adding more clones and incorporating experimental error. Multisampling adds another dimension to allele frequency analysis permitting that more reliable assertions of clonal structure be made.

Computational Simulation

To test the potential efficacy of multisampling we wrote an R-script program named Cell Sampling, which was successful as a preliminary proof-of-concept. Although this approach requires previously establishing and verifying the VAFs of the mutant alleles in the separate cell lines it proved an excellent starting model for predicting correlations and clones. By pulling replicated low-throughput count draws from a simulated bulk sample the program was able to correctly determine the correlations between every mutant allele that were known to occur in the same cell line. With simple

alterations the program could be used to predict correlations between different allele targets and cell lines. In time, the program could develop into a powerful model for future clinical low-throughput clonality assays where it could be used to determine which alleles to target in a patient to get the most definitive *in vivo* clonality evaluation. The program can also be "reverse engineered" into a new, separate program which takes correlations between target alleles and predicts the present clones. Much more work would be required to reach these stages but the Cell Sampling program, and *in silico* modeling with R-script, was an extremely useful and cost saving means to model an experiment and test the hypothesis.

Data Management

One of the goals of this project was to see what kind reliability was attainable from the pyrosequencing results and how the *in vitro* results compare to the *in silico*. What level of correlation is a true association and what is random background noise? By establishing the mixture of clonal populations ourselves, with the known mutational allele frequencies established though sequencing panels and Sanger sequencing, it is possible to know what correspondences are expected from the pyrosequencing results and interpret the correlations accordingly. With that knowledge we were able to determine that in our data set correlation values which were between 0.33 and -0.33 were the product of random association and equivalent with no allele linking. Conversely, mutant alleles with correlation values greater than 0.34 or less than -0.34 appear to be representative of either positive or negative correlation respectively. For this preliminary experiment it was necessary to know the actual clonal populations and set the correlation limits to prevent

36

false positives or false negative correlations. With more work increased specificity leading to limits closer to zero and correlation coefficients of linked genes closer to 1 or -1 might be attained, allowing for more accurate clonal discrimination and eventually the ability to determine the clonal structure without prior knowledge of its composition.

Several replications were removed from final correlation analysis. Replicates 4, 5, 12, 13 and 20 were reported to have as saturated signal errors by the pyrosequencing instrument for one or more of the targets and so the entire replicate was removed from correlation analysis and clonal extrapolation. Additionally, the presence of univariate outliers between the inter-replicate VAFs were identified through interquartile range (IQR) statistical analysis (Rousseeuw & Hubert, 2011). Replicates which contained a VAF value below or above the first and third quartile respectively by more than 1.5 times the IQR for that VAF were also removed from correlation analysis (replications 1 and 14). The reason for the presence of the outliers is suspected, but not proven, to be due to exponential error propagation originating from the nested PCR step (**Appendix 7**). While the proportional amount of data points removed from this study, seven out of twenty in total, is unfortunate, the reasons for the removals are explicit and seen to be valid. Increased expertise and practice with the low-throughput clonality assay should lower the amount of removals in the future.

Low-Throughput Clonality Assay

The correlation between any two alleles in a population may be in one of three states depending on common mutations between the clones: **1)** Uncorrelated, where

alleles are not wholly shared between clones, or for which experimental variation

significantly impacts correlations determined, **2)** Positive correlation, where alleles are

wholly shared between clones, **3)** Negative correlation, where mutations are mutually

exclusive i.e. if a clone does not contain mutation A then it must contain mutation B.

When limiting the scope of the evaluation of the pyrosequencing results to solely

DNMT3α associations it can been seen that the DNMT3α mutation has a positive

correlation to the NRAS-Q61 mutation that is also present in OCI-AML3 cells, a negative

correlation to the TET2 mutation which is not present in OCI-AML3 but present KG1α

and MOLM-14, and no correlation to the FLT3-D835 or FLT3-ITD mutations (**Figure

21**). Thus, while the DNMT3a mutation and the FLT3-D835 mutation occur at

approximately the same VAF in the bulk, via cell sampling, we can determine that they

are not actually present in the same clone. By identifying which variants do or do not

correlate with each other we can refine our proposed clonal structure identified via the

bulk data.
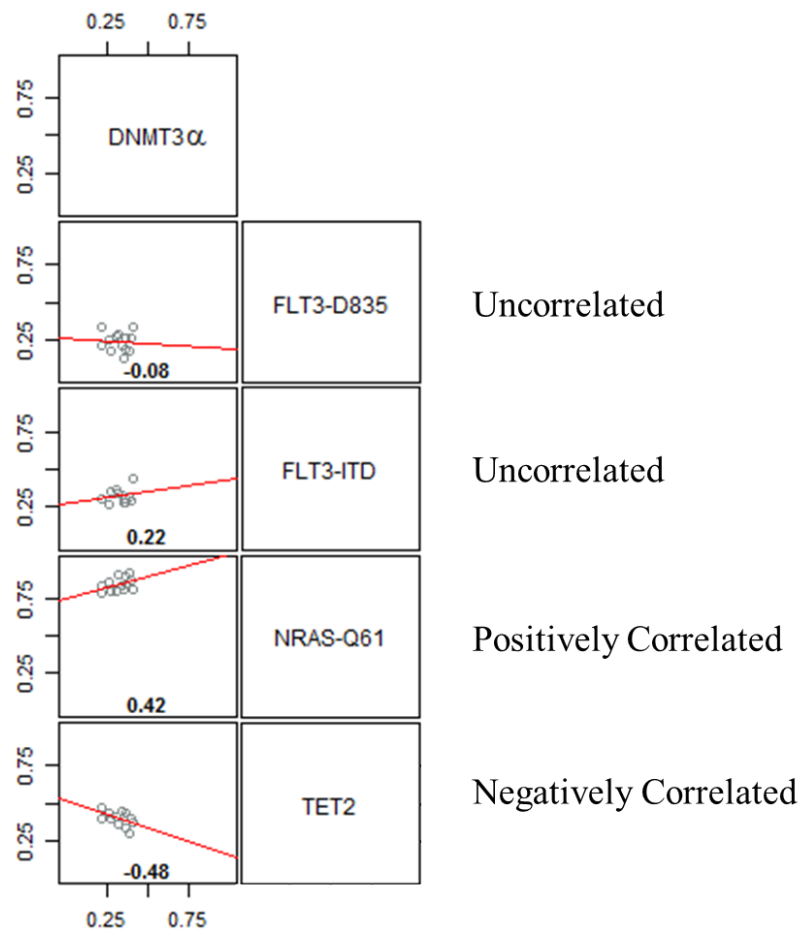
Figure 21. Correlations between DNMT3α and other target mutant alleles. Three possible correlation states are possible between any two alleles; uncorrelated, positively correlated or negatively correlated.

By expanding the scope of focus to include all five targets and their correlations to each other the clonal model can be further refined (**Figure 20**). The NRAS-Q61 mutation also negatively correlates with the TET2 mutation, supporting the previous assertion that the clones with the NRAS-Q61 and DNMT3α mutations form an opposing group of clones from those with the TET2 mutation. The FLT3 mutations are positively correlated, indicating that they are in the same clone, but are uncorrelated with anything else, suggesting that they comprise a subpopulation within either of groupings. In all, the

39

in vitro low-throughput clonality assay was able to correctly determine the minimum number of clones, the presence of the two groupings of clones and the positive correlations between uniquely shared mutations (DNMT3α with NRAS-Q61 and FLT3-ITD with FLT3-D835).

Despite what the low-throughput clonality assay was able to determine, several clonal characteristics known to be true were not seen in the results. The FLT3 mutations are known to inhabit a sub clonal population with the TET2 mutation but a positive correlation between the two was not observed (**Figure 22**). In fact, the FLT3 mutations--which were not whole shared with any other mutation besides each other--were not seen to correlate positively or negatively with any other mutant alleles and were responsible for all of the uncertainty in the assay. This implies that a mutation that is shared across two clones does not appear to correlate with a mutation that is in one of said clones in a way that is currently detectable with this system. This may occur because as cells are randomly sorted into replicates the VAF of the unshared mutation may fluctuate wildly as individual cell proportions change but the VAF of the shared mutation stays relatively constant as the overall cell proportions are constant. Increases in sensitivity through expertise and practice should lower the random correlations (i.e. noise) and permit detection of partially linked mutations. However, this factor represents a potential flaw in the multi-sampling model and requires further testing and modeling. Also unseen in this assay is the homozygous nature of the NRAS-Q61 mutation in the OCI-AML3 cells. By examining the mean VAF from all replicates it can be seen that the average VAF of

NRAS-Q61 is twice that of DNMT3α and so may be estimated to be homozygous but

also may simply be allele bias.
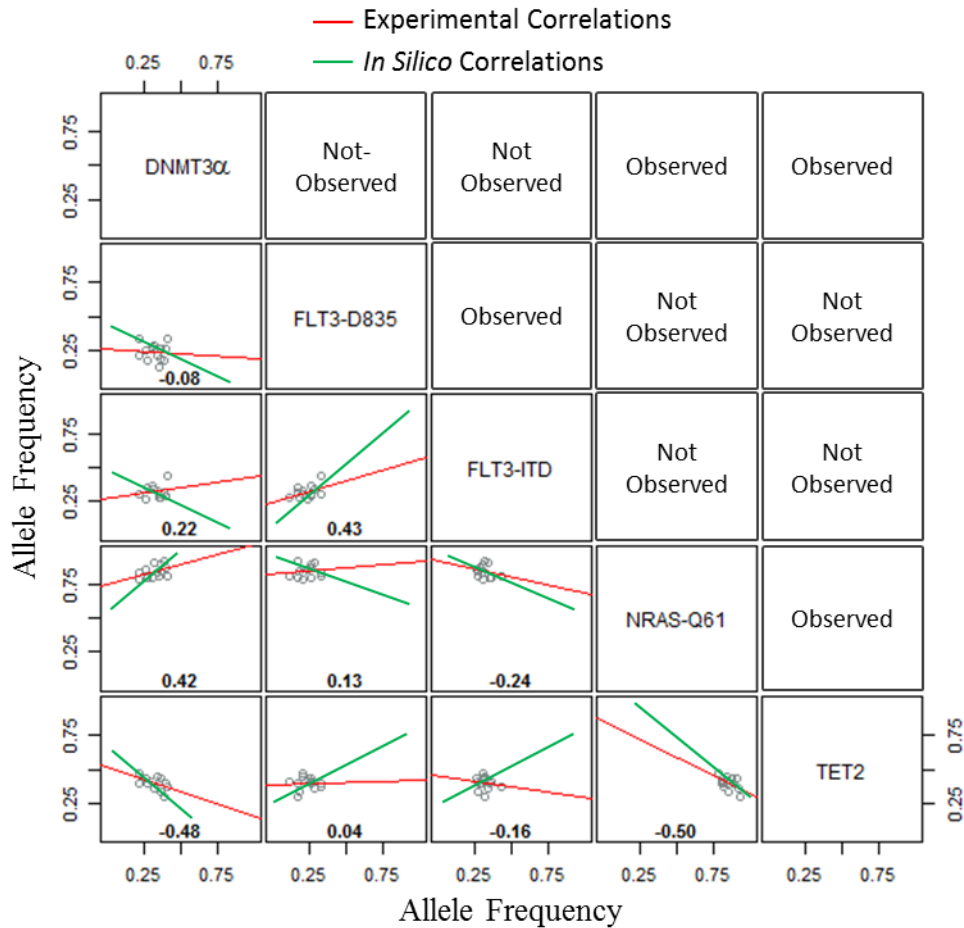
Figure 22. Predicted vs. experimental results. Lower Left: red lines indicate the experimentally derived correlations between VAFs and green lines indicate *in silico* correlation predictions. Upper Right: whether the correlations between the two mutant alleles that were predicted *in silico* was observed *in vitro*.

The low-throughput/multi-sampling clonality assay offers a more accurate picture

on the clonal populations and diversity present in a sample than simple bulk analysis. It

verified in part, the *in silico* modeling by Cell Sampling and supported the multisampling

practice. More time invested in the assay to increase proficiency should increase

specificity and more results may make any shortcomings or possible improvements of the model more evident. This data could then be implemented back into Cell Sampling to improve its prediction accuracy.

The low throughput clonality assay does entail several drawbacks. High-throughput bulk analysis may overlook rare sub-clones due to high allele frequency background noise, but low-throughput may miss rare sub-clones simply because fewer cells are targeted and the chosen target mutations must necessarily be common to many diverging cell lines. The low-throughput clonality assay protocol entails that for each unique set series of allele targets are appraised the multiplex primer mixture must be re-designed. The nested PCR portion of experimental method is sensitive to contamination and can result in exponential error propagation. Finally the assay has difficulty in determining the correlations between alleles where one allele is shared across multiple clones. Overall the assay did an excellent job of demonstrating the limitations of NGS bulk samples, supporting the *in silico* model and offer a potential alternative to bulk sampling.

With improvements the assay may also offer some clinical applicability as it has several benefits over bulk sequencing: **1)** A majority of the equipment used, with the possible exception of the PyroMark Q96 instrument, is commonly found in research labs **2)** Non-cancerous HSPC or other cell populations in the samples will not skew the correlations between two mutant alleles in the way that is found in bulk sequencing. **3)** The assay makes no assumptions of zygosity in its population determination and is thus

less likely to overestimate a clone's population size. **4)** Extremely small samples may be used –our experiment only used ~ 2400 cells in total--. The low-throughput clonality assay, with more work, may be a clinically viable option for assessing clonal diversity and populations.

Future Work

The program Cell Sampling and the low-throughput clonality assay both function well as proof-of-concepts but also leave room for development. Optimization of the PCR reaction's conditions (i.e. salt concentrations, stage times, or nucleotide concentrations) and the multiplex primers may increase sensitivity and should be verified. Testing more variations in cellular proportions and sample sizes would increase the number of data points for analysis, allowing for more accurate determination of correlations between linked and unlinked mutant alleles, theoretically approaching the values determined in Cell Sampling. Increasing the number of targets would also increase the number of data points and allow the differentiation of similar clones, with more targets increasing specificity. By incorporating distorting factors such as primer biases and potential amplicon interactions into the Cell Sampling computer model the program could more accurately predict pyrosequencing results, speeding development through *in silico* experimentation. Finally, a reverse direction computational program which distinguishes unique and unknown clones from a multi-sample's pyrosequencing results would be a large step toward clinical applicability for the multiplex/nested/pyrosequencing protocol.

Appendix 1. Multiplex primers for the first-round product found through PrimerBLAST and verified *in silico* with MFEPrimer 2.0.

Gene/Primer-Name	Product (bp)	Direction	Sequence (5' to 3')
DNMT3α-SP-F/R	237bp	F	GTGTGGTTAGACGGCTTCCGG
		R	CTCAGTTTGCCCCCATGTCCCTT
EZH2-SP-F/R	571bp	F	CTCGAGGAAATCAAGGGCTGAAGAT
		R	CATGCAGAAGTCCAGGCTGAAAAG
FLT3-ITD-SP-F/R	453bp	F	CTTTCCTCTATCTGCAGAACTGCCT
		R	GCATGGGTGGGAAACTGTGCCT
FLT3-D835-SP-F/R	522bp	F	CTCACGGCACAGCCCAGTAAAGA
		R	TTGCACTCAAAGGCCCCTAACTGAT
NRAS–Q61-SP-F/R	342bp	F	GGCAATAGCATTGCATTCCCTGTGGT
		R	CCTAGTGTGGTAACCTCATTTCCCCATA
STAG2-SP-F/R	400bp	F	GTGGCATATAGGGAGAAGAAATAAGCTAA
		R	CATGAAGGCAGGGACTGTCAATCA
TET2-SP-F/R	313bp	F	CACATAACTGCAGTGGGCCTGAAA
		R	CCTGAGGTATGCGATGGGTGAGT

Appendix 2. Inner primers for nested reaction, found through Qiagen Pyromark software. One of each paired primers were biotin tagged for pyrosequencing except for FLT3-ITD which was has a fluorescing FAM marker for fragment analysis.

Gene/Primer Name	Product (bp)	Direction	Sequence/Biotin (5' to 3')
DNMT3α-Pyro-F/R	65bp	F	CCCAGTCCACTATACTGACGTCTC-Biotin
		R	ACCGGCCCAGCAGTCTCT
EZH2-Pyro-F/R	69bp	F	CATCTATTGCTGGCACCATC-Biotin
		R	CATTTTTCTGCACAGGATCTTTG
FLT3-D835-Pyro-F/R	96bp	F	AGGAACGTGCTTGTCACCC
		R	GCCCCTGACAACATAGTTGGA-Biotin
NRAS-Q61-Pyro-F/R	85bp	F	CCTGTTTGTTGGACATACTGGATACAGC-Biotin
		R	AAGCCTTCGCCTGTCCTCATGTA
STAG2-Pyro-F/R	82bp	F	ACTAACAGATAGGCAAGAGAGT-Biotin
		R	CACGGGAGGATGACATTC
TET2-Pyro-F/R	120bp	F	GCTAATGCCTAATGGTGCTACA
		R	GTTTTCTGCACCGCAATGGAA-Biotin
FLT3-ITD	330bp	F	AGCAATTTAGGTATGAAAGCCAGC-FAM
		R	CTTTCAGCATTTTGACGGCAACC

Appendix 3. Sequencing primers for pyrosequencing, found with Qiagen Pyromark software.

Gene/Name	Direction	Sequence (5' to 3')
DNMT3α-seq	R	AGTCTCTGCCTCGCC
EZH2-seq	R	TGATAAAAATCCCCCA
FLT3-D835-seq	F	GATATGTGACTTTGGATTG
NRAS-Q61-seq	R	CATGGCACTGTACTCTTCT
STAG2-seq	R	GCTTGTCTAATGGTACAAA
TET2-seq	F	GCTACAGTTTCTGCCTCT

Appendix 4. Cell sampling program R-script program written to simulate an AML patient and calculate the correlations between mutant alleles with the goal of evaluating the clonal structure.

Weston Christensen
October 20, 2015

Initialization parameters

```
#install.packages("knitr", "ggplot2", "polynom")
library('knitr')
library('ggplot2')
library('polynom')

#cells<-120        #Cells drawn per rep (NULL when cell# is determined by correlation)
reps<-20          #Number of draws from bulk
totalcells<-300000  #Total cell count in bulk
KG1a<-0.343
MOLM14<-0.384       #Cell percentages
OCIAML3<-0.273
set.seed(0)

#Cell-line Matrix Generator
celllines<-matrix(c(0,0,0,0,.5,0,.5,.5,0,.5,.5,0,0,1,0), ncol =5, byrow = T)
colnames(celllines)<-c("DNMT3a","FLT3-D835","FLT3-ITD", "NRAS-Q61", "TET2")
rownames(celllines)<-c("KG1a","MOLM14", "OCI-AML3")
kable(celllines, caption = "Variable Allele Frequencies for 3 AML cell lines")
```

Variable Allele Frequencies for 3 AML cell lines (fraction).

Cell Lines	DNMT3α	FLT3-D835	FLT3-ITD	NRAS-Q61	TET2
KG1a	0.00	0.00	0.00	0.00	0.50
MOLM14	0.00	0.50	0.50	0.00	0.50
OCI-AML3	0.50	0.00	0.00	1.00	0.00

Sampling Function

```
#initiates Sampling function
Sampling<-function(cellsf, repsf, KG1af, MOLM14f, OCIAML3f, celllinesf, totalcellsf)
 {

#Creates and fills VAFtable (VAFt) with titles
VAFt<-matrix(data=NA, nrow=repsf, ncol=5)
colnames(VAFt)<-c("DNMT3a","FLT3-D835","FLT3-ITD", "NRAS-Q61", "TET2")
rownames(VAFt)<-c(paste("Rep",1:repsf))

#creates vector "bulk" and fills it with the total cell number(Sum of percentage*total)
bulk<-c(rep("KG1a", (KG1af*totalcellsf)), rep("MOLM14",(MOLM14f*totalcellsf)), rep("OCI-AML3", (
OCIAML3f*totalcellsf)))
```

```
#builds table "samples" and fills it with cell# of
rownames(samples)<-c(paste("Cell -",1:cellsf))

#For loop that fills VAFt from each rep
i<-0
for (i in 1:repsf)
  {

#Counts cell numbers in each rep and fills it into vector count
count<-c((sum(samples[,i] == "KG1a")),(sum(samples[,i] == "MOLM14")), (sum(samples[,i] == "OCI-A
ML3")))

#multiplies count of each cell type times 2 alleles, multiplied by VAF and sums, then divides by total allele
s
DNMT3a<-(sum(2*count[1]*celllinesf[1,1],2*count[2]*celllinesf[2,1], 2*count[3]*celllinesf[3,1])/(2*cells
f))
D835<-(sum(2*count[1]*celllinesf[1,2],2*count[2]*celllinesf[2,2],  2*count[3]*celllines[3,2])/(2*cellsf))
ITD<-(sum(2*count[1]*celllinesf[1,3],2*count[2]*celllinesf[2,3], 2*count[3]*celllinesf[3,3])/(2*cellsf))
NRAS<-(sum(2*count[1]*celllinesf[1,4],2*count[2]*celllinesf[2,4], 2*count[3]*celllinesf[3,4])/(2*cellsf))
TET2<-(sum(2*count[1]*celllinesf[1,5],2*count[2]*celllinesf[2,5], 2*count[3]*celllinesf[3,5])/(2*cellsf))

#fills VAF table with vectored VAF for each rep
VAFt[i,]<-c(DNMT3a,D835,ITD,NRAS,TET2)
}

#data frames VAFt into dat
dat.VAF<-data.frame(VAFt)

return(dat.VAF)
}
```

Calculates mean correlation for un-linked loci for based on cell number

```
#for loop establishing parameters, min/max cell number
min<-75
max<-150
b<-min

#Creates Correlation table "cor.table"
cor.table<-matrix(data=NA, ncol=2, nrow=max)

#for loop incrementing cells or reps
for (b in  min:max)
  {

#Number of cells drawn from bulk per rep, changing to reps will produce similar results but for reps instead
cells <- b

#calls sampling function
dat.VAF<-Sampling(cellsf=cells, repsf=reps, KG1af=KG1a, MOLM14f=MOLM14, OCIAML3f=OCIAM
L3, celllinesf=celllines, totalcellsf=totalcells)
```

```r
#averages the correlations for the un-linked locci
cor.vect<-mean(c(
  cor(dat.VAF$FLT3.D835, dat.VAF$DNMT3a),
  cor(dat.VAF$FLT3.D835, dat.VAF$NRAS.Q61),
  cor(dat.VAF$FLT3.D835, dat.VAF$TET2),
  cor(dat.VAF$FLT3.ITD, dat.VAF$DNMT3a),
  cor(dat.VAF$FLT3.ITD, dat.VAF$NRAS.Q61),
  cor(dat.VAF$FLT3.ITD, dat.VAF$TET2),
  cor(dat.VAF$TET2, dat.VAF$NRAS.Q61),
  cor(dat.VAF$TET2, dat.VAF$DNMT3a)
  ))

#Fills CoVar table with mean CoVar for each cell count
cor.table[b,]<-c(b,cor.vect)
}

#dataframe fixes cor.table and names
dat.cor<-na.omit(data.frame(cor.table))
colnames(dat.cor)<-c("Cell Number","Mean Correlation")

#plots cor.tablr with CI
g<-ggplot(data=dat.cor, aes(`Cell Number`,`Mean Correlation`))+ geom_point()+stat_smooth(method = "lm", formula = y ~ poly(x, 2), size = 1)+ylim(-.55,-.15)+scale_x_continuous(breaks=c(75,100,125,150), expand = c(0.01, 0.01))

#Extracts Coefficients and solves for min Covariance (as long as there are zeros)
Coeff<-coef(lm(formula = dat.cor$'Mean Correlation' ~ dat.cor$'Cell Number' + I(dat.cor$'Cell Number'^2)))
p<-polynomial(Coeff)
p<-solve(p)
OptCellDraw<-mean(p)
print(c("Optimum Cell Draw Count =",OptCellDraw))
## [1] "Optimum Cell Draw Count =" "118.043096549912"
g<-g+annotate("text", x =120, y = -.175, label = "Optimum Cell Draw =")
g<-g+annotate("text", x=135.2, y= -.175, label = round(OptCellDraw,0))
g
```
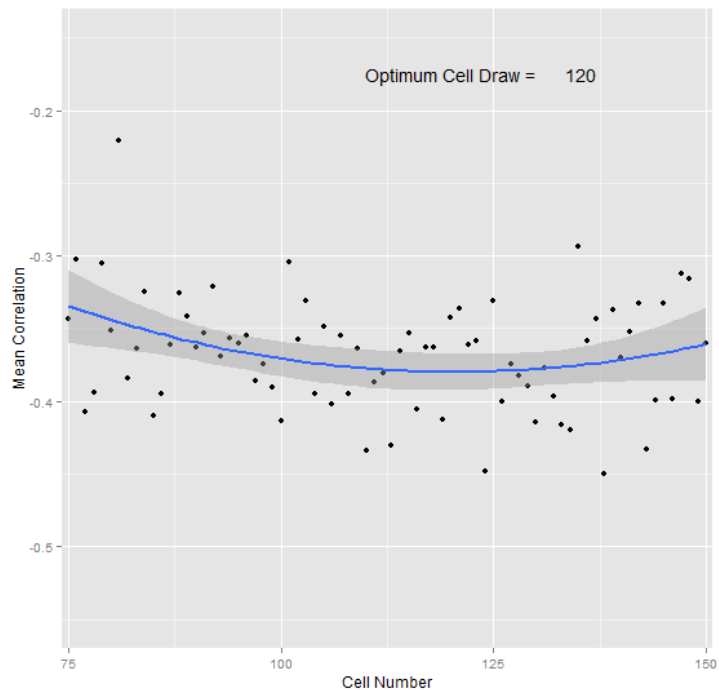
Determination of the optimum cell draw number via *in silico* cell draws.

Creates VAF figure Using Sampling Function and cell number from previous function

```
#cells<-ceiling(OptCellDraw)
cells<-OptCellDraw

#calls Sampling function to fill dat.VAF
dat.VAF<-Sampling(cellsf=cells, repsf=reps, KG1af=KG1a, MOLM14f=MOLM14, OCIAML3f=OCIAM
L3, celllinesf=celllines, totalcellsf=totalcells)

#Table of VAF means
mean.table<-matrix(round(c(mean(dat.VAF$DNMT3a),mean(dat.VAF$FLT3.D835),mean(dat.VAF$FLT3
.ITD),mean(dat.VAF$NRAS.Q61),mean(dat.VAF$TET2)),3),byrow=T)
 rownames(mean.table)<-c("DNMT3a","FLT3-D835","FLT3-ITD", "NRAS-Q61", "TET2")
 colnames(mean.table)<-"Mean VAF"
 kable(mean.table, caption="Mean Target VAFs, Representative of the Bulk data")
```

Mean Target VAFs, Representative of the Bulk data

Target	Mean VAF
DNMT3α	0.131
FLT3-D835	0.198
FLT3-ITD	0.198
NRAS-Q61	0.263
TET2	0.369

```r
#graphs dat.VAF with best fits and correlations
pairs(dat.VAF, diag.panel = NULL, upper.panel=NULL, lab=c("DNMT3a","FLT3-D835","FLT3-ITD","N
RAS-Q61", "TET2"), xaxs="r",  xaxp=c(.25,.75,2), xlim=c(0,1), ylim=c(0,1), yaxs="r", yaxp=c(.25,.75,2),
gap=.25, cex.labels=1.1, panel=function(x, y, digits=2, prefix="",cex.cor,... ){       #function creates best fit
 lines for scatterplots
  points(x,y, col='azure4')
  abline(lm(y~x), col='red')

 #Adds Pearson Correlation
  usr <- par("usr"); on.exit(par(usr))
    par(usr = c(0, 1, 0, 1), font=2)
    r  <- cor(x, y,use="pairwise", method="pearson")
    txt <- format(c(round(r,digits),
    0.123456789), digits=digits)[1]
    txt <- paste(prefix, txt, sep="")
    text(0.75, .92, txt,cex=1)
})

#averages the correlations of the unlinked locci
cor.vect<-mean(c(
 cor(dat.VAF$FLT3.D835, dat.VAF$DNMT3a),
 cor(dat.VAF$FLT3.D835, dat.VAF$NRAS.Q61),
 cor(dat.VAF$FLT3.D835, dat.VAF$TET2),
 cor(dat.VAF$FLT3.ITD, dat.VAF$DNMT3a),
 cor(dat.VAF$FLT3.ITD, dat.VAF$NRAS.Q61),
 cor(dat.VAF$FLT3.ITD, dat.VAF$TET2),
 cor(dat.VAF$TET2, dat.VAF$NRAS.Q61),
 cor(dat.VAF$TET2, dat.VAF$DNMT3a)
 ))


mtext("Mean VAF correlation",adj=.8,padj=5)
mtext(round(cor.vect,4), adj=.8,padj=7)
mtext("Cell Draw Number", adj=.76, padj=9)
mtext(round(cells,0), adj=.77, padj=11)
```
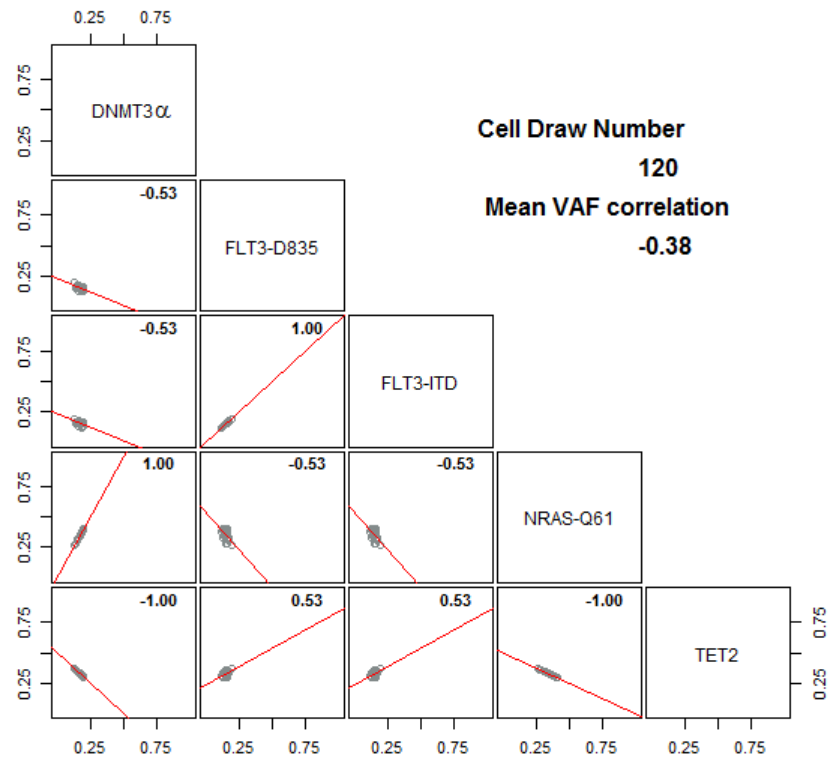
VAFs plotted *in silico* against each other for every rep with a calculated correlation.

Appendix 5. VAFs determined through *in silico* cell sampling of the simulated bulk and extrapolation of VAFs. Bulk contains all three cell lines and proportions are reflective of the true proportions determined through flow cytometry.

Rep	DNMT3α	FLT3-D835	FLT3-ITD	NRAS-Q61	TET2
Rep 1	0.12	0.23	0.23	0.24	0.38
Rep 2	0.09	0.19	0.19	0.18	0.41
Rep 3	0.14	0.16	0.16	0.29	0.36
Rep 4	0.12	0.20	0.20	0.25	0.38
Rep 5	0.13	0.17	0.17	0.25	0.37
Rep 6	0.14	0.19	0.19	0.28	0.36
Rep 7	0.11	0.20	0.20	0.23	0.39
Rep 8	0.16	0.18	0.18	0.31	0.34
Rep 9	0.17	0.16	0.16	0.33	0.33
Rep 10	0.13	0.21	0.21	0.25	0.37
Rep 11	0.13	0.21	0.21	0.25	0.37
Rep 12	0.14	0.23	0.23	0.28	0.36
Rep 13	0.13	0.19	0.19	0.25	0.37
Rep 14	0.13	0.19	0.19	0.26	0.37
Rep 15	0.13	0.20	0.20	0.25	0.37
Rep 16	0.13	0.20	0.20	0.25	0.37
Rep 17	0.11	0.23	0.23	0.21	0.39
Rep 18	0.13	0.23	0.23	0.26	0.37
Rep 19	0.15	0.18	0.18	0.30	0.35
Rep 20	0.16	0.20	0.20	0.31	0.34
Mean	0.13	0.20	0.20	0.26	0.37
STD Dev	0.02	0.02	0.02	0.04	0.02

Appendix 6. Allele frequencies calculated from actual population proportions. With the mutated allele frequencies for each cell line determined through sequencing and the composition of the cell line mixture determined though flow cytometry the total allele frequency for each target can be determined, simulating a bulk analysis. Identical frequencies between FLT3-ITD and FLT3-D835 indicate being in a similar clone. DNMT3α frequency is half of that NRAS-Q61 indicating either that they are in the same clones at a different zygosity or that a subpopulation division occurred. TET2 and DNMT3α have a combined allele frequency of exactly 50% indicating that the whole population shares either or the other mutated allele. It should be noted that increasing targets and cell populations will dramatically complicate the picture and lineages would be much harder to pick apart.

Target	Allele Frequency
DNMT3α	13.65%
FLT3-D835	19.2%
FLT3-ITD	19.2%
NRAS-Q61	27.3%
TET2	36.35%

Appendix 7. Pyrosequencing and fragment analysis allele VAF results in percentage from the low-throughput clonality analysis on cell line mixture. Rows highlighted in light red were removed from correlation calculations due to either pyrosequencing failure or the presence of outliers which were detected using the interquartile range [1.5IQR below first quartile or above 3rd quartile] highlighted in dark red. Mean values differ from the *in silico* simulation however, as long as the correlation between VAFs is constant the actual value may fluctuate.

Replication	DNMT3a	FLT3-D835	FLT3-ITD	NRAS-Q61	TET2
Rep 1	27.6	17.1	24.6	89.7	26.4
Rep 2	27.1	24.5	28.9	86.9	43.8
Rep 3	31.1	26.9	38.0	80.2	41.5
Rep 4	26.3	FAIL	34.0	FAIL	39.7
Rep 5	36.6	23.7	20.2	FAIL	29.9
Rep 6	40.2	26.3	33.3	87.9	40.2
Rep 7	36.0	12.3	35.4	81.8	40.7
Rep 8	39.4	17.4	35.7	92.1	29.5
Rep 9	37.3	19.0	26.7	84.8	33.9
Rep 10	22.7	20.6	32.3	78.5	47.0
Rep 11	27.6	17.4	37.4	80.0	39.9
Rep 12	30.8	14.1	30.3	FAIL	45.3
Rep 13	25.3	FAIL	33.1	93.0	57.6
Rep 14	38.5	19.2	36.6	65.3	43.4
Rep 15	34.9	21.3	35.3	83.4	45.2
Rep 16	22.5	33.1	31.0	83.5	39.3
Rep 17	36.5	26.3	34.6	90.1	43.4
Rep 18	32.4	29.0	39.1	91.4	35.4
Rep 19	41.6	33.1	48.0	81.8	37.3
Rep 20	FAIL	40.3	49.9	93.2	38.7
Mean	33.0	23.6	32.0	84.8	39.8
STD Dev	6.4	6.3	5.3	4.5	4.8
1st Quartile	26.30	17.40	30.35	80.05	39.30
3rd Quartile	37.30	25.40	35.71	87.65	45.20
IQR	11.00	8.00	5.36	7.60	5.90

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. http://doi.org/10.1016/S0022-2836(05)80360-2

Baker, F., Denniston, M., Smith, T., & West, M. M. (2005). Adult cancer survivors: How are they faring? *Cancer*, *104*(S11), 2565–2576. http://doi.org/10.1002/cncr.21488

Byrd, J. C., Mrózek, K., Dodge, R. K., Carroll, A. J., Edwards, C. G., Arthur, D. C., … Bloomfield, C. D. (2002). Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood*, *100*(13), 4325–4336. http://doi.org/10.1182/blood-2002-03-0772

Corces-Zimmerman, M. R., & Majeti, R. (2014). Pre-leukemic evolution of hematopoietic stem cells – the importance of early mutations in leukemogenesis. *Leukemia*, *28*(12), 2276–2282. http://doi.org/10.1038/leu.2014.211

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., … DiPersio, J. F. (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*, *481*(7382), 506–510. http://doi.org/10.1038/nature10738

Fearon, E. R., Hamilton, S. R., & Vogelstein, B. (1987). Clonal analysis of human colorectal tumors. *Science (New York, N.Y.)*, *238*(4824), 193–197.

Fernald, G. H., Capriotti, E., Daneshjou, R., Karczewski, K. J., & Altman, R. B. (2011). Bioinformatics challenges for personalized medicine. *Bioinformatics*, *27*(13), 1741–1748. http://doi.org/10.1093/bioinformatics/btr295

Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., … Swanton, C. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England Journal of Medicine*, *366*(10), 883–892. http://doi.org/10.1056/NEJMoa1113205

Greaves, M., & Maley, C. C. (2012). Clonal evolution in cancer. *Nature*, *481*(7381), 306–313. http://doi.org/10.1038/nature10762

Harada, H., & Harada, Y. (2015). Recent advances in myelodysplastic syndromes: Molecular pathogenesis and its implications for targeted therapies. *Cancer Science*, *106*(4), 329–336. http://doi.org/10.1111/cas.12614

Henegariu, O., Heerema, N. A., Dlouhy, S. R., Vance, G. H., & Vogt, P. H. (1997). Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques*, *23*(3), 504–511.

Hughes, A. E. O., Magrini, V., Demeter, R., Miller, C. A., Fulton, R., Fulton, L. L., … Graubert, T. A. (2014). Clonal Architecture of Secondary Acute Myeloid Leukemia Defined by Single-Cell Sequencing. *PLoS Genet*, *10*(7), e1004462. http://doi.org/10.1371/journal.pgen.1004462

Jan, M., & Majeti, R. (2013). Clonal evolution of acute leukemia genomes. *Oncogene*, *32*(2), 135–140. http://doi.org/10.1038/onc.2012.48

Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, *35*(suppl 2), W43–W46. http://doi.org/10.1093/nar/gkm234

Levis, M. (2013). FLT3 mutations in acute myeloid leukemia: what is the best approach in 2013? *ASH Education Program Book*, *2013*(1), 220–226. http://doi.org/10.1182/asheducation-2013.1.220

Maley, C. C., Galipeau, P. C., Finley, J. C., Wongsurawat, V. J., Li, X., Sanchez, C. A., … Reid, B. J. (2006). Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature Genetics*, *38*(4), 468–473. http://doi.org/10.1038/ng1768

Md Fakruddin, A. C. (2012). PYROSEQUENCING-PRINCIPLES AND APPLICATIONS. *International Journal of Life Science and Pharma Research*, *2*(2), 65–76.

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., … Wigler, & M. (2011). Tumour evolution inferred by single-cell sequencing. *Nature*, *472*(7341), 90+.

Paguirigan, A. L., Smith, J., Meshinchi, S., Carroll, M., Maley, C., & Radich, J. P. (2015). Single-cell genotyping demonstrates complex clonal diversity in acute myeloid leukemia. *Science Translational Medicine*, *7*(281), 281re2–281re2. http://doi.org/10.1126/scitranslmed.aaa0763

Qu, W., Zhou, Y., Zhang, Y., Lu, Y., Wang, X., Zhao, D., … Zhang, C. (2012). MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity. *Nucleic Acids Research*, *40*(W1), W205–W208. http://doi.org/10.1093/nar/gks552

Rousseeuw, P. J., & Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 73–79. http://doi.org/10.1002/widm.2

Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., … Norton, L. (2001). Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. *New England Journal of Medicine*, *344*(11), 783–792. http://doi.org/10.1056/NEJM200103153441101

Szer, J. (2012). The prevalent predicament of relapsed acute myeloid leukemia. *ASH Education Program Book*, *2012*(1), 43–48. http://doi.org/10.1182/asheducation-2012.1.43

TruSight Myeloid Sequencing Panel. (2015). [Manufacturer]. Retrieved January 4, 2016, from http://www.illumina.com/products/trusight-myeloid.html

Turner, N. C., & Reis-Filho, J. S. (2012). Genetic heterogeneity and cancer drug resistance. *The Lancet Oncology*, *13*(4), e178–e185. http://doi.org/10.1016/S1470-2045(11)70335-7

Welch, J. S., Ley, T. J., Link, D. C., Miller, C. A., Larson, D. E., Koboldt, D. C., … Wilson, R. K. (2012). The origin and evolution of mutations in acute myeloid leukemia. *Cell*, *150*(2), 264–278. http://doi.org/10.1016/j.cell.2012.06.023

Wu, X., Northcott, P. A., Dubuc, A., Dupuy, A. J., Shih, D. J. H., Witt, H., … Taylor, M. D. (2012). Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature*, *482*(7386), 529–533. http://doi.org/10.1038/nature10825

Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., … Iacobuzio-Donahue, C. A. (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*, *467*(7319), 1114–1117. http://doi.org/10.1038/nature09515

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*, 134. http://doi.org/10.1186/1471-2105-13-134

<h1 style="text-align:center">CURRICULUM VITAE</h1>

Weston Christensen
360.320.4900 • Weston.Christensen@yahoo.com • Jan, 1990

PERSONAL SUMMARY	Second-year student in the Masters of Medical Science Program at Boston University School of Medicine. As a component of my graduate program I studied different aspects of medicine including histology, biochemistry, anatomy, and physiology. I am also currently working at the Fred Hutchinson Cancer Research Center researching AML, which I have written a thesis on, as part of my graduation requirement. I am slated to finish my thesis mid-January and have my degree approved by May 2016.

EDUCATION

WESTERN WASHINGTON UNIVERSITY, BIOCHEMISTRY B.S. AND MATHEMATICS A.A

Sept 2009 – Graduated in March 2014

As an undergraduate student, I covered subjects including organic and inorganic chemistry, biology, microbiology, biochemistry, physical biochemistry, virology, genetics, and mathematics.

BOSTON UNIVERSITY OF MEDICAL SCIENCES, MASTER'S OF MEDICAL SCIENCES – *IN PROGRESS*

Sept 2014 – *Pending (paperwork stage)*

The Graduate Medical Sciences program at Boston University gives its students a strong foundation in the human body's anatomy and physiology. I covered System's Physiology, cellular differentiation and composition, biochemistry, pharmacology and ethical standards.

MCAT

Scored in the 95th percentile in the 2015 test.

EXPERIENCE

FRED HUTCHINSON CANCER RESEARCH CENTER

07/06/2015-*Present*

Research into the clonal diversity of Acute Myeloid leukemia culminating in a thesis paper key to completing my Master's degree in Medical Sciences.

ENDOSCOPY CLINIC – BOSTON MEDICAL CENTER

03/01/2015-05/31/15

In the endoscopy clinic I showed patients to their beds, ensured their comfort and assisted the doctors and nurses though setting up kits (for scopes and intravenous

catheters).

BOSTON CARES

08/25/14-05/31/15

Through the volunteer based non-profit organization Boston Cares I connected with multiple opportunities through which I donated my time.

CHILD FAMILY HEALTH INTERNATIONAL, QUITO, ECUADOR

08/01/2014-08/29/2014

Summer internship in the capital of Ecuador where I worked with underprivileged children in daycares, did rotations at a hospital with Hematology and Surgery, and took classes to develop my proficiency in Spanish.

UNDERGRADUATE RESEARCH, WESTERN WASHINGTON UNIVERSITY, SPIEGEL LAB

09/24/2012-05/14/2014

Two years of research, specifically focusing on protein refolding and disulfide bonds. I ran my own project concerning growth and refolding protocols of insoluble protein domains of Factor VIII.

PHYSICAL THERAPY AIDE, FAMILYCARE PHYSICAL THERAPY

06/29/2008-07/29/2014

Hands-on work with patients in a healthcare setting, particularly providing assistance with exercise form, maintaining a clean and orderly environment and providing necessary attention to detail in the day-to-day operation of the clinic.

COMMUNICATION **PROTEIN SOCIETY CONFERENCE, BOSTON, M.A.**

Jul 2013

I presented my undergraduate research: *Expression and Purification of the Factor VIII A2 Domain for Structural and Biochemical Studies*.

VOLCANO CONFERENCE, PARK FOREST, W.A.

Feb 2014

I presented my research to faculty and other students in the Pacific Northwest area.

REFERENCES **CLINT SPIEGEL, PHD**

 Associate Professor, Western Washington University

 Spiegep@chem.wwu.edu

VICKERY TRINKHAUS-RANDALL, PHD

 Director, Cell and Molecular Biology Graduate Program, Boston University

School of Medicine
Vickery@bu.edu

JERRY RADICH
Member, FHCRC
Professor, University of Washington School of Medicine
Director, Molecular Oncology Lab
Jradich@fredhutch.org