Understanding Amplification Bias from the EOS Assay

Abbie Olson[1], Travers Ching[2], Bryan Howie[2]
Adaptive Biotechnologies, University of Oregon

Abstract

Named after the Greek goddess of the rising sun, EOS is Adaptive's preeminent assay for quantifying immune cell distributions, distinguishing cancer cells, and detecting minimal residual disease (MRD) in B-cell acute lymphoblastic leukemia (B-ALL) and multiple myeloma (MM). Via a multiplex polymerase chain reaction (PCR) assay, highly diverse B-cell V(D)J sequences can be tracked and quantified. Using a combination of Illumina primers, assay-specific molecular tags (moltags), and bioinformatic processes, EOS accurately quantifies cancer cell frequencies[1]. Here we delve into these aforementioned steps, though we specifically explore sources of amplification bias and computational methods for identifying and ultimately mitigating them.

Introduction

Figure 1: clonoSEQ corporate logo.

While our main object is to discuss amplification bias from both the *in vitro* and the *in silico* establishment of cancer cell lineages, MRD detection is one of this assay's cornerstones. Being an NGS-based assay, clonoSEQ really is a leader among alternative methods. In the past, Southern blot was often deployed as an immunoglobulin (Ig) quantification method, yet it was technically challenging and time-consuming—not to mention sensitivity limited. Nevertheless, this method was soon replaced by qualitative PCR (qPCR) and (even more commonly) multiparametric flow cytometry (MPFC) assays, the latter of which is still considered the standard in MRD detection. However, PCR-based, high-throughput sequencing (HTS) assays (like Adaptive's clonoSEQ assay) have proved worthy rivals, particularly within the realm of B-cell malignancies. These assays are frequently less time-consuming and more reproducible than qPCR, which requires patient-specific primers[3].

EOS in Action

Internally referred to as EOS, this assay is more colloquially known as the B-cell wing of Adaptive's clonoSEQ platform—a comprehensive, next-generation sequencing (NGS)-based, Food and Drug Administration (FDA)-approved assay for quantifying lymphoid malignancies[2]. The necessity for this assay can't be stressed enough, as Ig molecules are highly variable, and prior to NGS-based methods was remarkably difficult to quantify. All Ig molecules share a general structure with two heavy chains and two light chains, both consisting of amino-terminal variable and constant regions (with antigen recognition and effector functions, respectively). Ig molecule variability falls namely on the complimentary-determining region (CDR3). However, antigen receptor gene rearrangement is imperative in the development of diverse immune repertoires. The rearrangement is characterized by V(D)J recombination, which includes the rearrangement of Variable (V) and Joining (J) gene segments on the light chains, or V, Diversity (D), and J gene segments on the heavy chain[5]. The V(D)J recombination

site acts as a clone marker, thus allowing for the detection of cancerous clones among healthy lymphoid cells[2].

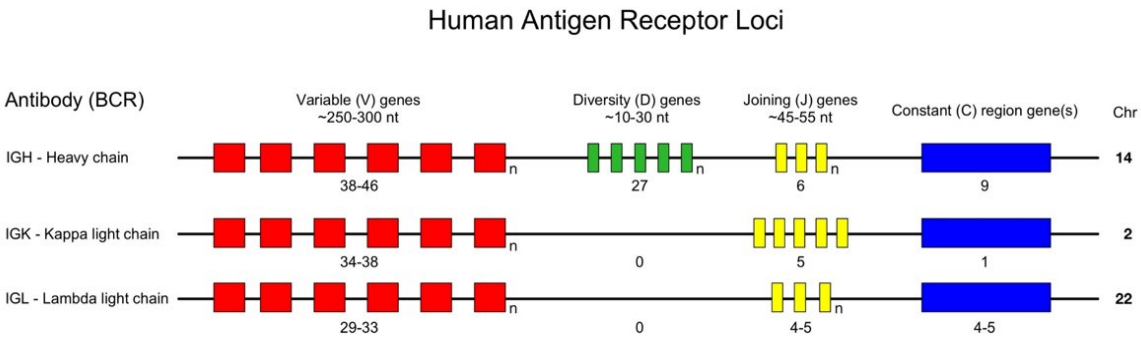

Figure 2: Each human antigen receptor locus of B and T cells.

While one of the major benefits of EOS is its ability to track clonal malignancies, the fast-recombining nature of these cells may lead to amplification bias, a potentially detrimental source of error when the goal of this assay is accurate malignancy quantitation. Much of the diversity seen in immune cell repertoires comes from a process called somatic hypermutation (SHM), which can be thought of as a point mutation that helps diversify B-cell receptors (BCRs). It increases binding affinity for target antigen by largely diversifying the V gene of each Ig molecule, though it notably has more of an effect on the V gene of the heavy chain (IgH). Immune cells are "repertoired" with primers that amplify specific genomic regions present as diploid copies in normal genomic DNA (gDNA), which allows for the determination of total nucleated cell content[2]. That being said, the possibility of SHM occurring on a primer-binding site is highly possible and could lead to the prevention of an Illumina primer from binding during the assay's multiplex PCR step. In multiplex PCR, both the V and the J genes of the BCR get at least one Illumina primer. However, amplification bias can be reduced further by adding more primer binding sites to each segment. The J segment gets two primer binding sites, while the V gene gets three. Recall that the V gene of the IgH locus generally experiences more SHM than the other segments, which warrants an additional primer binding site. Once primer-bound, loci are sequenced by NGS, usually by an Illumina HiSeq 500 or 550.

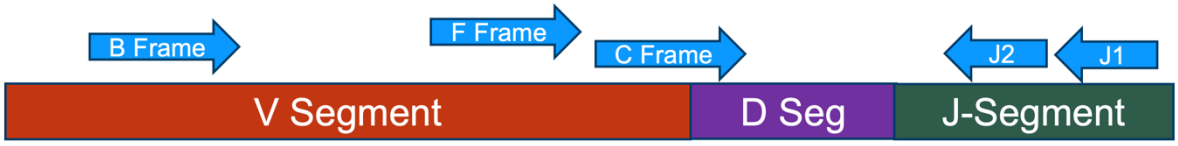

Figure 3: Multiple primer-binding sites on the human antigen receptor of an IgH locus.

One of the main goals after sequencing is the identification of how many templates (these are molecules of input material, post-extraction and pre-amplification) per clonotype are present in each sample. In the original, unmodified sample, one DNA molecule is added per biological receptor sequence. Four rounds of pre-amplification are then performed, at which point moltags are attached, meaning there are multiple moltags per biological sequence. Next, the sequences undergo twenty rounds of amplification, yielding many reads per moltag, as well as numerous universal Illumina primers. An Illumina sequencer (likely a HiSeq 500 or 550) will produce raw sequences, but the overarching issue is that there's no way to tell from which templates they originated. Luckily, template origin can be

identified via a series of bioinformatic steps derived from a proprietary probabilistic model that accurately identifies dominant clonotypes[4]. One of the best methods for identifying dominant clonotypes is actually to just "eyeball it" by plotting template counts. Figure 4 displays this method in both IgH and in the light chains, IgK and IgL. By establishing a cancer cutoff, it's easy to see which cells are likely malignant.

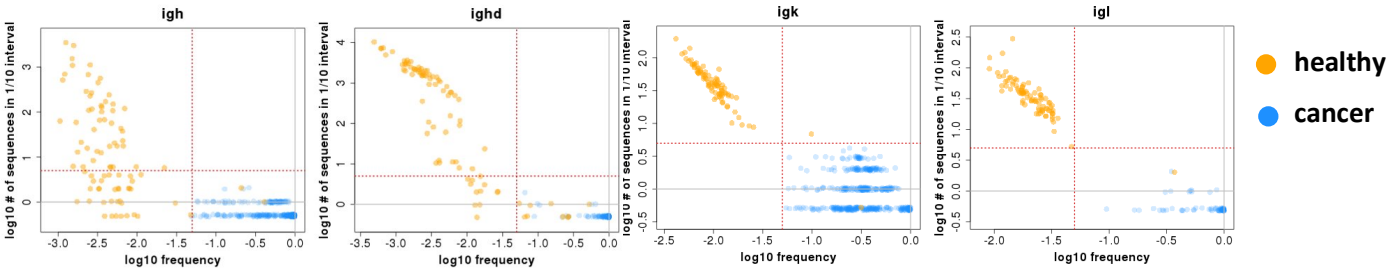

Figure 4: Malignancy cutoffs for the identification of dominant clonotypes.

While identifying templates that are measured at higher frequencies than others is a valid method, not all malignant clones can be quantified this way. Lineage establishment can be deployed via an assay-specific metric called *MaxAdjustedMutations*, which determines whether clonotypes are from the same lineage, and is critical to take into account, as cancer lineages may evolve. *MaxAdjustedMutations* is calculated by taking the Hamming distance between sequences and applying a threshold. For example, a very unique IgH sequence with lots of SHM will have a high *MaxAdjustedMutations*, while a simple IgK sequence with little to no SHM is practically germline, and therefore has a *MaxAdjustedMutations* score of 0. After lineages are established, more thresholds can be applied to determine malignancy. These thresholds: the clonal sequence is present in at least 3% of all sequences in the locus, the sample must have a nucleated cell frequency of at least 0.2%, the clone must be present in more than 40 templates, the clone must have more than 50 moltags, and there can be no more than 5 sequences within a decade below another clonotype.

After a sequence is determined to be part of cancer, another metric is used to adjust for unaccounted amplification bias. Called *DiseaseLoadMultiplier* (DLM), this metric is an equation for quantifying and correcting underamplification. For example, given two sequences that have passed all of the calibration thresholds, yet have very different abundances, this could signify some level of underamplification. If given a sequence count of 10,000, and one BCR is identified 8,000 times, it's much more likely to be correctly amplified, versus another BCR that is only identified 100 times when the sequence count is 10,000. Therefore, the second BCR is likely heavily underamplified. The importance of proper amplification cannot be stressed enough when clonal malignancies are involved.

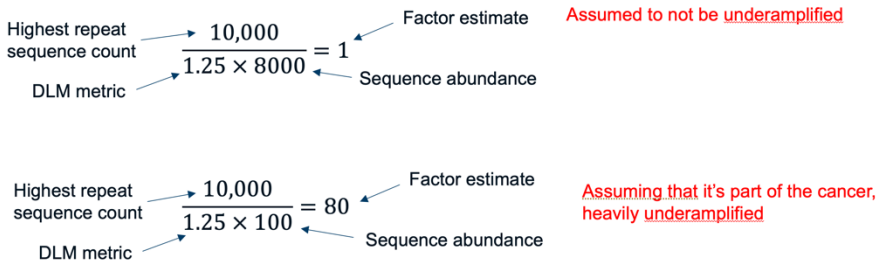

Figure 5: DLM equation examples.

While DLM is considered viable enough to yield accurate results, empirical evidence shows that the assay is still affected by low levels of amplification bias. Our recent efforts have attempted to identify features that likely contribute to amplification bias, specifically via machine learning and ensemble learning methods. Keeping in mind that some features are effectively known sources of amplification bias, we considered whether there may be a difference in sequence distributions between underamplifiers and non-underamplifiers. We were particularly interested in the length of a BCR's CDR3 region, V and J gene families, GC content in the BCR, and homopolymer distributions (repeating regions of nucleotides).

We started with raw EOS data from various sources (Table 1) and performed standard data preparations using a combination of Python's pandas library and scikit-learn, a popular machine learning library for Python. We first identified the target feature for our classifier, an adaptation of the EOS DLM output. After log transforming the output, anything above 0 would be considered "underamplified," and anything at 0 would be considering "not underamplified." Additional data wrangling consisted namely of dropping features that were largely comprised of NaNs, or were directly related to the target feature, which we labeled log10_DLM. We hypothesized that shorter amplicons are better amplifiers, which is reflected by the left-hand tail in Figure 5.

Sample Origin	Sample Count	Cancer Type
Dana Farber Cancer Institute	365	Multiple Myeloma (MM)
Child's Oncology Group	280	Acute Lymphocytic Leukemia (ALL)
MD Anderson	222	Chronic Lymphocytic Leukemia (CLL)
Commercially Purchased	66	ALL/MM/CLL

Table 1: Sample origins.

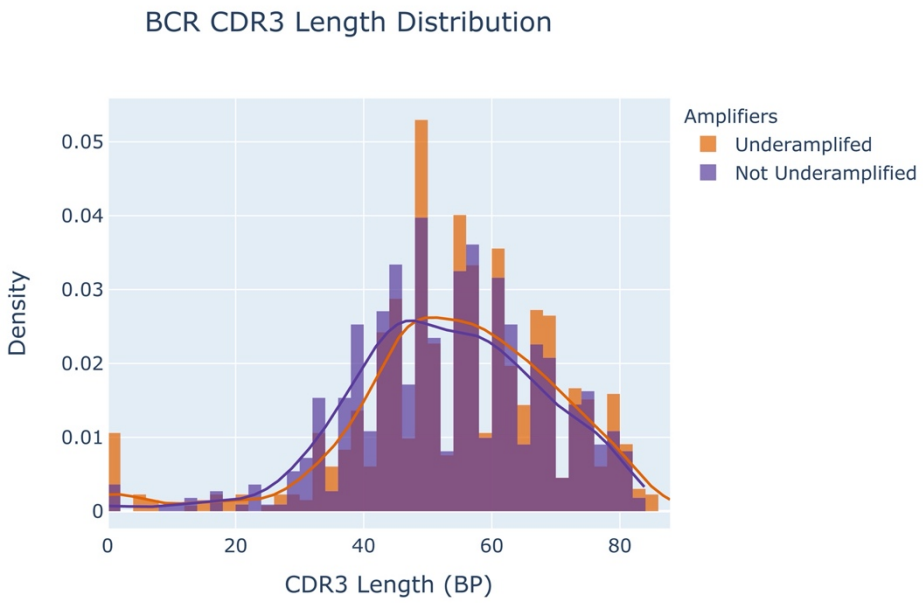

Figure 5: BCR CDR3 length distributions.

We further prepared the data by one-hot encoding categorical data and standardizing continuous data to a mean of 0 and standard deviation of 1 (Figures 6a and 6b, respectively). Using an exhaustive grid search method, we attempted the following classifiers: Decision Trees, Logistic Regression, Gradient Booster, and Random Forest. When the areas under the curve (AUCs) of each classifier were assessed, the most accurate model was the Random Forest model.

aminoAcid_CARPGLSGWYWTFGYW	aminoAcid_CARPLTSGTTPFHHW	aminoAcid_CARPRRRFLEWLL*GHYYYYYGMDVW
0	0	0
0	0	0
0	0	0
0	0	0
0	0	1

Figure 6a: One-hot encoding.

cdr3Length	vGeneAllele	dGeneAllele	jGeneAllele	vDeletion	d5Deletion	d3Deletion	jDeletion	n2Insertion
-2.093508	-0.301384	-0.370312	-0.068487	3.435688	0.007349	2.240055	0.508010	3.412419
0.993374	1.131146	-0.370312	-0.068487	-0.448023	-0.576407	2.079259	-0.368622	-0.105763
-1.321788	0.056748	-0.370312	-0.068487	-0.448023	-0.770992	-0.815070	0.654115	-1.007861
1.057684	-0.301384	-0.370312	-0.068487	-0.324731	0.007349	-0.815070	-0.953043	-0.195973
1.572164	-0.301384	-0.370312	-0.068487	-0.386377	0.201934	-0.493478	-1.099148	-0.376393

Figure 6b: Standardization.

We achieved our best Random Forest model via the hyperparameter tuning of various parameters within a grid search. Gini impurity was the chosen method for splitting the trees. This measure of misclassification applies to multiclass classifiers that contain both continuous and categorical data. Additionally, we used a stratified 10-fold cross-validation due to a slight imbalance in our log-transformed DLM class, and we wanted to maintain the percentage of class samples. The resulting AUC had 80% accuracy with a holdout of 30% (Figure 7a). When we looked at the top features associated with the target feature, we saw some of our hypothesized features, in addition to some that surprised us. We predicted that SHM, *MaxAdjustedMutations*, CDR3 length, and V genes would have an effect, and our results suggest they do. Interestingly, indels and indices from the V/NDN/J junction and specific gene families also had some effect. The presence of SHM as a main factor in predicting underamplification was expected, as we recall that SHM on priming sites may lead to underamplification. *MaxAdjustedMutations* is also another significant cause, as any large *MaxAdjustedMutations* factor estimates are likely to represent SHM-containing sequences. Regarding CDR3 length, it wasn't surprising to see CDR3 make the top features list, but the accompanying indels and indices from the V/NDN/J junction suggest that amplicon length really does have an effect. In the case of Gini impurity, when the decrease of impurity over trees is averaged, it is calculated as the probability of mislabeling an element, assuming the element is randomly labeled according to the distribution of all classes in the set. In other words, continuous data has many more split points, leading to the multiple testing problem. A downside of selecting this method is that it obviously favors continuous data. Future analyses would require looking at individual p-values and executing a Bonferroni correction.

Conclusion

Amplification bias—more specifically underamplification—is a serious cause for concern when the goal of the EOS assay is malignant clone identification. While the assay's current method for identifying amplification bias is sufficient, empirical evidence of amplification bias still remains. By exploring the potential sources of this bias with machine learning methods (in our case with a Random Forest classifier), we identified known sources of amplification bias, and uncovered new potential sources that may lead us to a richer understanding of the biological and biochemical complexities of B-cell receptor recombination and malignant clone identification.

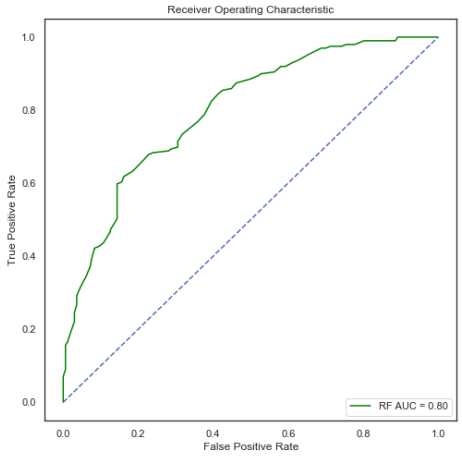

Figure 7a: AUROC curve of the Random Forest classifier.

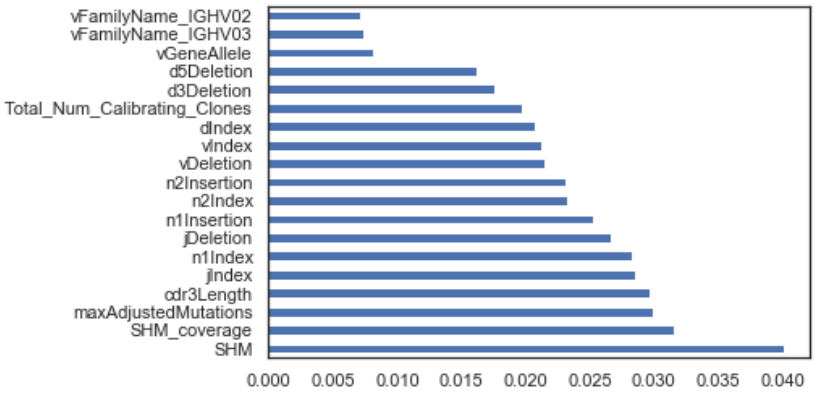

Figure 7b: Top features from the Random Forest classifier.

References

1. Brüggemann, Monika, and Michaela Kotrova. "Minimal residual disease in adult ALL: technical aspects and implications for correct clinical interpretation." *Hematology 2014, the American Society of Hematology Education Program Book* 2017.1 (2017): 13-21.
2. Caers, Jo, et al. "European Myeloma Network recommendations on tools for the diagnosis and monitoring of multiple myeloma: what to use and when." *haematologica* 103.11 (2018): 1772-1784.
3. Ching, Travers, et al. "Analytical evaluation of the clonoSEQ Assay for establishing measurable (minimal) residual disease in acute lymphoblastic leukemia, chronic lymphocytic leukemia, and multiple myeloma." (2020).
4. Monter, Anna, and Josep F. Nomdedéu. "ClonoSEQ assay for the detection of lymphoid malignancies." *Expert review of molecular diagnostics* 19.7 (2019): 571-578.
5. Yanamandra, Uday, and Shaji K. Kumar. "Minimal residual disease analysis in myeloma–when, why and where." *Leukemia & lymphoma* 59.8 (2018): 1772-1784.