immunoSEQ®

TECHNICAL NOTE

immunoSEQ Analyzer: Understanding Clonality

Overview

In the field of immunosequencing, the diverse nature of T- or B-cell populations can provide insights for a wide range of applications including transplantation, infectious disease, oncology, inflammation, and allergy as well as autoimmunity.[1-5] One of the metrics used to understand and interpret the diversity of these cells is clonality. Clonality measures how evenly receptor sequences (rearrangements) are distributed amongst a set of T or B cells. This can quantify how focused the immune repertoire is on a particular set of antigens. In this tech note, we discuss the Simpson clonality metric, which is a more stable measurement of immune repertoire focus and is less affected by variations in sample input material or T-cell fraction.

What does T- or B-cell repertoire clonality indicate?

Diversity metrics, like clonality, can provide insight into how disease or intervention mechanisms affect the immune system in various applications. Such metrics also have the potential to be a predictive or prognostic biomarker in research settings such as transplantation, oncology, and immune disease.[1,6-12]

A diversity metric is a single number that describes the characteristic shape of a sample repertoire. When applied to the immune repertoire it is a powerful overall indicator of the state of the adaptive immune system. It can demonstrate an immune response to a specific threat or signify whether the system is healthy enough to respond to a variety of pathogens.

Diversity metrics can often be deconvolved into measures of richness and evenness. Evenness metrics describe the extent to which one or a few clones dominate the sample repertoire. Measures of repertoire clonality are evenness metrics. Richness measures focus on how many clones are present in the sample repertoire.

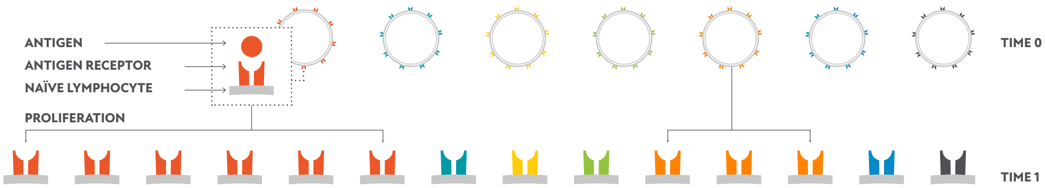

Figure 1. (A) Time 0: A perfectly even group of naïve lymphocytes, with a richness of 7 unique receptor rearrangements. **(B)** Time 1: An uneven group of lymphocytes post-proliferation, but with the same richness of 7 unique receptor rearrangements.

In Figure 1, time 0 shows a perfectly even group of naïve lymphocytes. Each receptor color represents a unique receptor rearrangement. Time 0 has a richness of 7 unique receptor sequences and a perfectly even distribution of one lymphocyte per receptor sequence. Time 1 shows an uneven group of lymphocytes. It still has a richness of 7 unique receptor sequences. However, one rearrangement appears on six lymphocytes, another on three, and the remaining five have one each. This is an uneven distribution.

TERMINOLOGY

Rearrangement:
Unique rearranged receptor sequence = clone.

Template:
A single input molecule that is the input template for PCR. When starting with gDNA this is one rearranged chromosome from a T or B cell.

Repertoire:
The set of all rearrangements and the abundance of each rearrangement in a sample identified by the immunoSEQ Assay.

Productive:
Only those rearrangements that are able to encode a functional protein. This includes only in-frame rearrangements and excludes those containing a stop codon or a frameshift mutation (termed "out of frame").

How is clonality calculated?

Multiple functions to calculate clonality and diversity exist in the literature. Many of these functions are derived from the fields of ecology, microbial biology, and immunology. Two functions for clonality are included in the immunoSEQ Analyzer and are derived from Shannon's Entropy and Simpson's Diversity metrics.

$$\text{SIMPSON CLONALITY} = \sqrt{\sum_{i=1}^{R} P_i^2} \qquad \text{SHANNON CLONALITY} = 1 - \frac{-\sum_{i=1}^{R} P_i \log_2 [P_i]}{\log_2 R}$$

Both Shannon and Simpson Clonality range from 0 to 1, where values approaching:

- 0 represents a completely even sample
- 1 represents a monoclonal sample

We recommend the use of Simpson Clonality as it is less sensitive to differences in sample size as seen in Figure 3. All analyses in this tech note use Simpson Clonality.

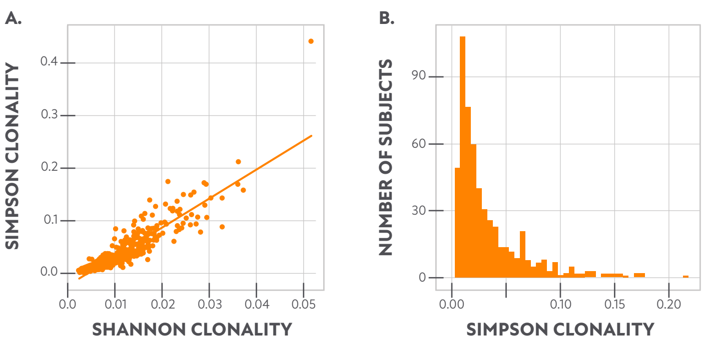

Figure 2. (A) Comparison of Shannon Clonality and Simpson Clonality across 555 healthy normal subjects from Emerson et al.[17] **(B)** Histogram of the distribution of Simpson Clonality values across the 555 subjects.

Across a healthy population, Shannon Clonality and Simpson Clonality have a strong linear correlation with an R^2 value of 0.82 (Figure 2A). Simpson clonality has a median value of 0.021 (Figure 2B).

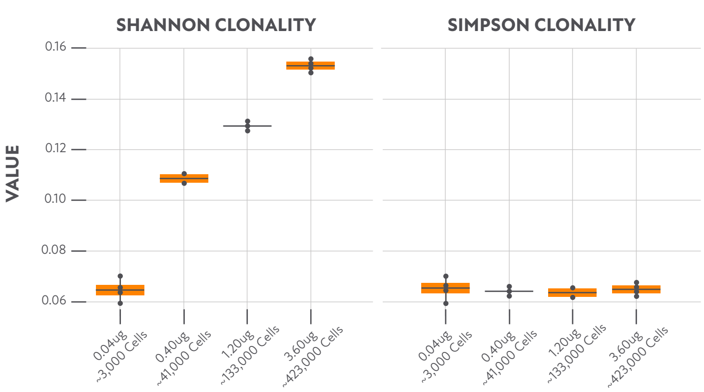

Figure 3. Comparison of technical replicates across 4 different loadings on the immunoSEQ Assay.

Across a range of input amounts Simpson Clonality yields a more stable value and is less influenced by variation due to changes in input material or T-cell fraction (Figure 3).

CLONALITY EQUATION

R = the total number of rearrangements; i = each rearrangement; $\mathbf{P_i}$ = productive frequency of rearrangement i

Shannon Clonality is 1-normalized entropy and also known as 1-Pielou's Evenness Index.

Simpson Clonality is the square root of Simpson's Index. Simpson's Index is 1—Simpson's Diversity Index.

In the immunoSEQ Analyzer output, Productive Entropy, Clonality [Shannon Clonality] and Simpson Clonality, only in-frame Productive rearrangements are included in R, and P_i is Productive frequency. For Sample Entropy, Sample Clonality [Shannon Clonality], and Sample Simpson Clonality, all rearrangements are included (productive, out-of-frame, and stop) and P_i is Frequency (fraction of specific rearrangement among all rearrangements).

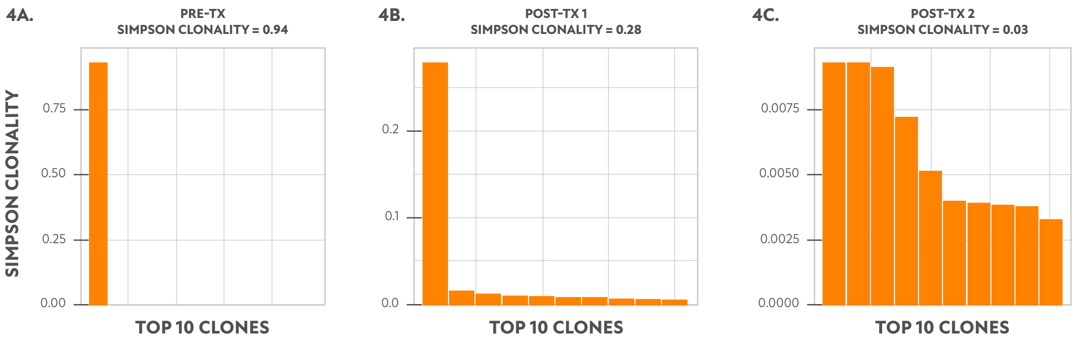

Figure 4. T-ALL Patients with 3 different clonalities. Data from Wu et al.[14]

Figure 4 compares representative samples with three different clonalities. Pre-treatment patients with T-cell acute lymphoblastic leukemia (T-ALL) have very high clonalities, where the repertoire is largely dominated by a single clone (Figure 4A). In this example, the top clone represents 93% of the repertoire. After treatment, depending on the response, a more even repertoire may be seen. In one example, the top clone has decreased to 28% (Figure 4B), while in the second case, the top clone represents <1% of the total repertoire (Figure 4C). When repertoires are dominated by a single clone, Simpson Clonality will converge on the max clone frequency.

What are examples of expected clonalities?

- Naïve B cells in a healthy subject's peripheral blood constitute an example of a low-clonality population of lymphocytes and are expected to have a clonality close to 0.[13]
- Sample clonality for 43 T-cell acute lymphoblastic leukemia patients with malignant TCRB sequences ranged from 0.40 to 0.99.[14]
- For PBMCs from 555 healthy research volunteers, Simpson Clonality ranged from 0.002 to 0.44.[17] See Figure 2: note the strong peak, as well as the long tail.
- Age, along with other underlying features, including clinical features including chronic viral conditions can influence peripheral repertoire clonality. Over time there is a trend towards larger numbers of individuals exhibiting higher clonality, resulting in much greater variance in clonality over time.[15,16] See Figure 5.

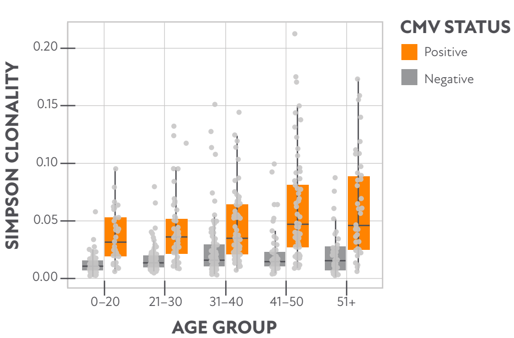

Figure 5. TCRB Simpson Clonality vs. age and CMV status. Data from Emerson et al.[17]

For Research Use Only. Not for use in diagnostic procedures.

How can clonality be used as a metric?

- Clonality is a measurement of the sample evenness, not the evenness of the underlying pool from which the sample was taken. As with all sampling regimes, larger samples more accurately reflect the diversity of the underlying populations.
- Clonality can be used to compare samples with different amounts of input material. Simpson Clonality is robust for comparing peripheral repertoires down to approximately 5,000 total T cells. For the comparison of tumor data, lower numbers of input cells can be used, although very small samples may show a bias towards increased Simpson Clonality.
- **Pro tip: Higher clonality is often seen in older individuals.[17, 18] If you are attempting to correlate clonality to a clinical or experimental factor in a population with a wide age range, control for age. One way is to do this is to use a mixed-effects model.**

REFERENCES

1. Wong, et al. J Immunol. 2016;197(5):1642–9.
2. Schneider-Hohendorf, et al. Nat Commun. 2016;7:11153.
3. Weir, et al. J Immunother Cancer. 2016;4:68.
4. Nunes-Alves, et al. PLoS Pathog. 2015;11(5):e1004849.
5.* Suessmuth, et al. Blood. 2015;125(25):3835–50.
6. Mahalingam, et al. Clin Can Res. 2020;26(1):71–81.
7. Morris, et al. Sci Transl Med. 2015;7(272):272ra10.
8. Roh, et al. Sci Transl Med. 2017;9(379).
9. Zhu, et al. OncoImmunology. 2015;4(12):e1051922.
10.* Tumeh, et al. Nature. 2014;515(7528):568–71.
11.* Keane, et al. Clin Cancer Res. 2017;23(7):1820–1828.
12. Kirsch, et al. Sci Transl Med. 2015;7(308):308ra158.
13.* Hershberg, et al. Phil. Trans. R. Soc. B. 2015;370(1676)20140239.
14.* Wu, et al. Sci Transl Med. 2012;4(134) L 134ra63.
15. Carey, et al. J Immunol. 2016;196(6):2602–13.
16* Seay, et al. JCI Insight. 2016;1(20):e88242.
17.* Emerson, et al. Nat Genet. 2017;49(5):659–665.
18.* Lindau et al. J Immunol. 2018; ji1800217

*Data available for exploration through immuneACCESS at immuneaccess.com

For Research Use Only. Not for use in diagnostic procedures.