Automated rarefaction analysis for precision B and T cell receptor repertoire profiling from peripheral blood and FFPE-preserved tumor

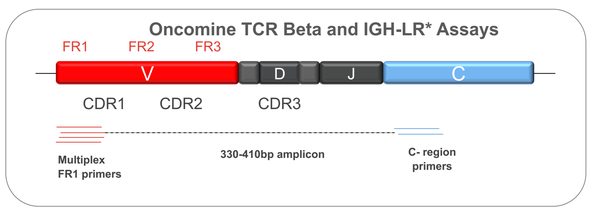
T Looney[1], M Toro[1], G Lowman[1], D Topacio-Hall[1], J Chang[1], L Pickle[1], F Hyland[1]
(1) Thermo Fisher Scientific, Austin, TX and Carlsbad, CA, USA
Correspondence to: Timothy Looney, PhD timothy.looney@thermofisher.com

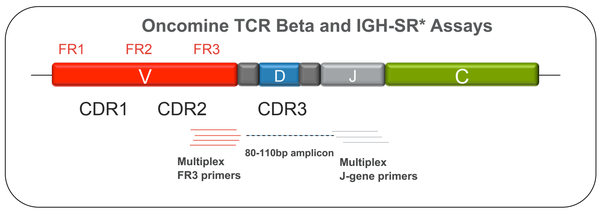
ABSTRACT

Identifying the optimal input amount and sequencing depth for B and T cell receptor repertoire profiling is challenging owing to variation in material quality and lymphocyte diversity in blood and FFPE preserved tumor specimens. Rarefaction analysis has emerged as a potential approach for assessing whether immune repertoire libraries have been sequenced to saturation. Here we present a novel automated method for saturation analysis of IGH and TCRB chain libraries derived from sequencing of peripheral blood leukocytes (PBL) and FFPE-preserved tumor RNA and DNA.

INTRODUCTION – Oncomine™ TCRB and IGH Assays

TCR Evenness is a measure of the similarity of clone frequencies in a TCR repertoire. It is also referred to as the normalized Shannon Entropy and is equivalent to 1 - “clonality”. Evenness nearing 1 indicates that all clones are found at similar frequencies in a sample.





INTRODUCTION – The Thermo Fisher Immune Repertoire Assay Portfolio

	TCRB Solutions	IGH Solutions
Human	Oncomine™ TCR Beta-LR Assay	Oncomine™ IGH-LR Assay
	Oncomine™ TCR Beta-SR Assay	Oncomine™ IGH-SR Assay
Mouse	Ion AmpliSeq Mouse TCR Beta-SR Assay	Ion AmpliSeq™ Mouse IGH-SR Assay

METHODS –Informatics Workflow

FR1-C or FR3-J multiplex PCR

↓

Annotate the V, D and J gene for each rearrangement by comparing to IMGT database

↓

Eliminate PCR and sequencing errors

↓

Report clones

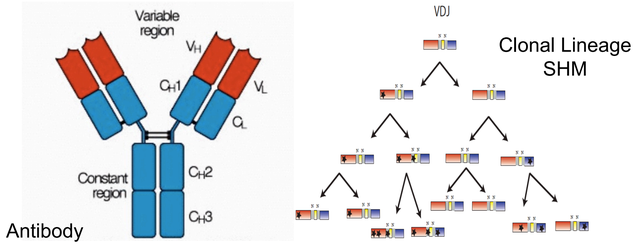
↓

Report clonal lineages and secondary repertoire features

Figure 3. Analysis Pipeline. Total RNA or gDNA is amplified by AmpliSeq multiplex PCR followed by sequencing on the Gene Studio S5 platform. VDJ rearrangements are annotated by comparison to the gold-standard IMGT database, sequencing and PCR errors are removed, then clones (unique VDJ rearrangements) and repertoire features are reported. For IGH workflows, a final step is performed to identify B cell clonal lineages.

METHODS – IGH Clonal Lineage Analysis

Clonal lineages represent B cells that derive from the same VDJ recombination event but differ at a sequence level owing to somatic hypermutation and class switch recombination. The software automatically identifies B cell clonal lineages using established methods such as described in (1).



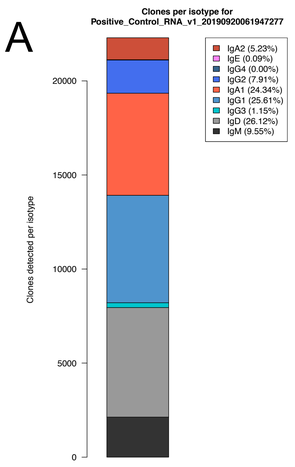
Images adapted from Janeway 2016

FIGURE LEGENDS (Right)

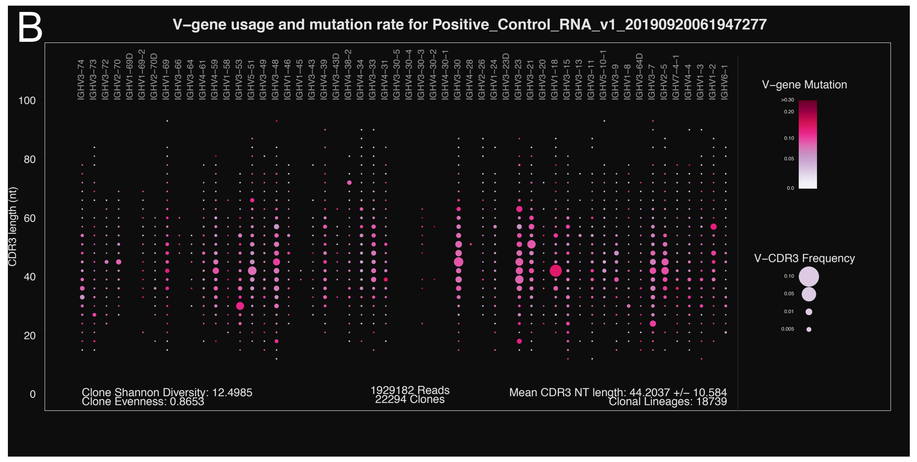
(A) Reads per isotype following sequencing of healthy donor peripheral blood leukocytes via the IGH-LR assay. The assay can distinguish all nine isotypes from one another **(B)** Spectratyping plot highlighting somatic hypermutation for the same sample. **(C)** Histogram of variable gene somatic hypermutation (SHM) for IgD expressing B cells and **(D)** IgA2 expressing B cells from the same sample. **(E)** The software reports the features of each clone and assigns clones to clonal lineages. Properties of detected clonal lineages are presented in a lineage summary table **(F)**. **(G)** Clone and lineage detection is repeated following automatic downsampling to a range of fixed read depths. **(G)** The number of clones and lineages detected following downsampling. **(H)** Clone and lineage normalized Shannon entropy (evenness) following downsampling. Evenness measurements are robust to sequencing depth.

RESULTS – Rarefaction Analysis of a Healthy Donor Peripheral Blood IGH Chain Repertoire via the IGH-LR Assay

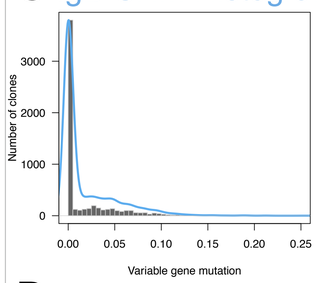
A



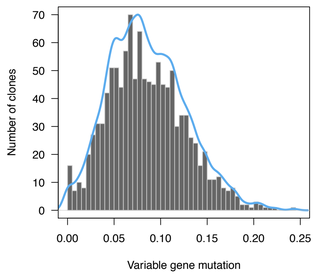
B



C IgD SHM Histogram



D IgA2 SHM Histogram



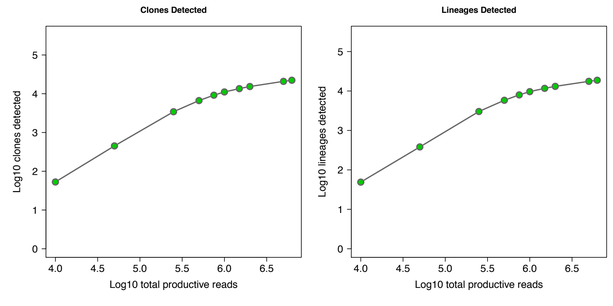
E Clone Summary Table

Lineage ID	Variable	Joining	CDR3 AA	CDR3 NT	Variable Mutation	Count	Frequency	Rank	Isotype
4	IGHV3-53	IGHJ5	TCTGTSPCTA	ACTTGTACTGGTAC…	0.121	24716	0.0128117	1	IGHG2
7	IGHV2-5	IGHJ4	AHITYYIDLYYFDF	GCGCACATTACATA…	0.091	16032	0.0083103	2	IGHG1
5	IGHV1-18	IGHJ4	ARDLDASNWYGIDY	GCGAGAGACCTCG…	0.155	14254	0.0073886	3	IGHA2
3	IGHV3-21	IGHJ4	VTEEATGWGLGVNYFDH	GTGACAGAGGAAG…	0.078	13594	0.0070465	4	IGHG1
1	IGHV1-18	IGHJ4	ARDMDRSNWYGTDY	GCGAGAGACATGG…	0.168	12756	0.0066121	5	IGHA1

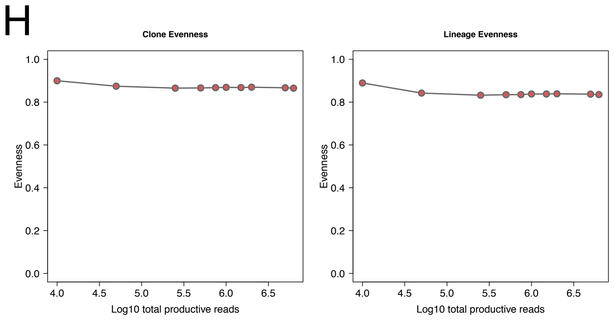
F Lineage Summary Table

Lineage ID	Variable	Top CDR3AA	Lineage Frequency	Number of Clones	Isotypes	Minimum V-gene SHM	Maximum V-gene SHM
1	IGHV1-18	ARDMDRSNWYGTDY	0.0158647	27	IGHA1;IGHA2;IGHG2	0.103	0.194
2	IGHV3-30	ARDSAGSGLLWLLDF	0.0150727	26	IGHA1;IGHA2;IGHD;IGHG2	0.065	0.135
3	IGHV3-21	VTEEATGWGLGVNYFDH	0.0141241	7	IGHA2;IGHG1;IGHG2	0.074	0.136
4	IGHV3-53	TCTGTSPCTA	0.0128117	1	IGHG2	0.121	0.121
5	IGHV1-18	ARDLDASNWYGIDY	0.0102914	8	IGHA1;IGHA2	0.089	0.181

G Automated Downsampling Analysis



H



CONCLUSIONS

- We observed an asymptotic relationship between the sequencing depth and the number of B and T cell clones detected, clone Shannon diversity, and B cell clonal lineage richness and diversity.
- Automated downsampling analysis may serve as a convenient tool for optimizing sequencing depth and input amount for B and T cell repertoire sequencing studies.

REFERENCES

1. Looney et al. Human B-cell isotype switching origins of IgE. Journal of Allergy and Clinical Immunology (2016). 137(2):579-586.e7

ThermoFisher SCIENTIFIC