

Scalable end-to-end immune repertoire analysis: the ImmunoGenomiX Platform

Henk-Jan van den Ham, Alvise Trevisan, Jorrit Posthuma, Jan Blom, Jos Lunenberg, Nicola Bonzanni

ENPICOM B.V. – 's-Hertogenbosch, the Netherlands



16th CIMT Annual Meeting

PUSHING FRONTIERS IN CANCERIMMUNOTHERAPY
MAY 15-17, 2018

RHEINGOLDHALLE
CONGRESS CENTER MAINZ, GERMANY

Immune Repertoire Sequencing

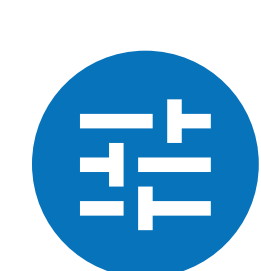
Lymphocytes form the core of the adaptive immune response. By expressing unique receptors that can recognize specific antigens, B and T cells raise specific immune responses that are subsequently remembered by memory responses. The diversity of B and T lymphocytes within a host is called the immune repertoire and represents the total set of receptors that can recognize antigens. Since these receptors are enormously diverse, the nucleotide sequence of the antigen-recognizing part of these receptors (the 'CDR3 region') can be used as a barcode to identify B and T cell clones (1,2). As a result of the high diversity of the immune repertoire, the limited output of Sanger sequencing provides only limited visualization of this variability.

Next generation sequencing (NGS) platforms are ideally suited to extensively characterize and visualize the complexity and plasticity of the TCR and BCR repertoires. By applying high throughput sequencing, sequences of millions of receptor can easily be obtained. This data can be used to monitor clones with specific sequences and to profile the immune repertoire as a whole (3,4).

The ImmunoGenomiX Platform

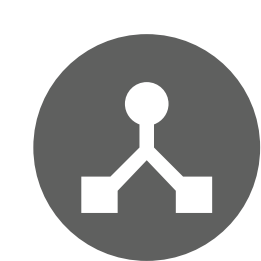
The ImmunoGenomiX Platform (IGX) is an end-to-end immunosequencing data analysis platform designed to analyze, monitor, and compare over time the immune repertoires at all stages of treatment. Starting from the high-throughput sequencing data, IGX Platform delivers an easy-to-read report to be used in research, diagnosis, patient stratification, and treatment monitoring.

The IGX Platform is founded on innovative bioinformatics methods and coded from scratch using rock-solid software engineering. It consists of a basic module for clonotype analysis and different optional modules for downstream analyses to answer all repertoire related questions will be added one-by-one.



Flexible at the frontend

allowing customers to use their own sample preparation protocols and the Next-Generation Sequencing (NGS) technology of choice



Integrative

with a full workflow management system, assuring full connectivity with LIMS and/or other data management environments at the customer sites



Easy-to-use

no skilled bioinformatics staff required, intuitive and simple graphical user interface



Versatile

the platform can be modularly expanded with additional functionality and applications



Open

allowing customers to connect it to tools that are already in use via API's and adapters



Applications

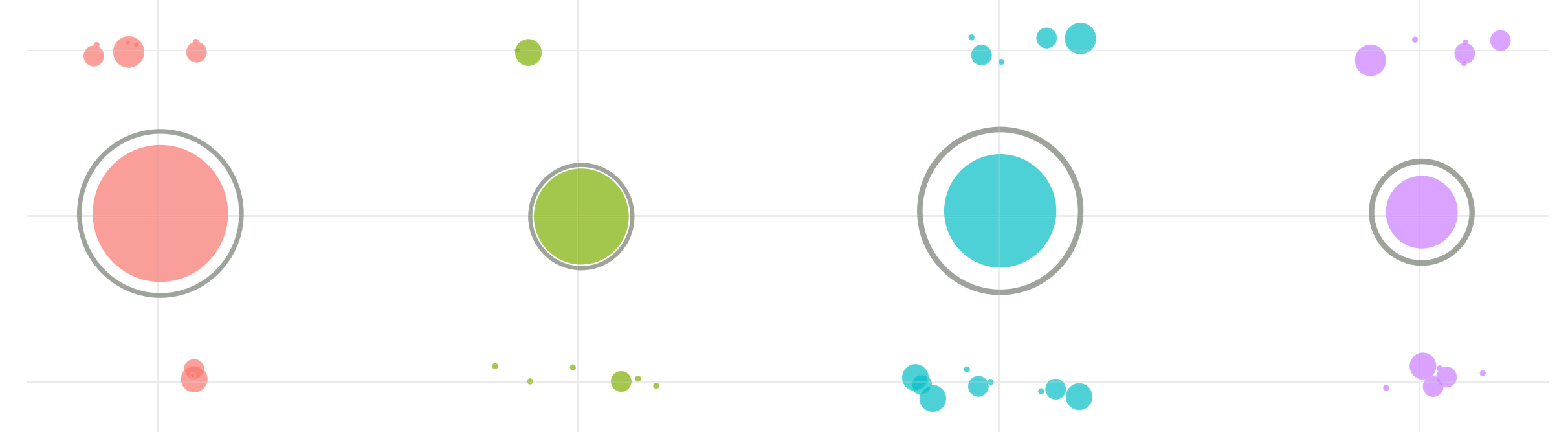
Immune repertoire profiling provides an excellent method for monitoring the adaptive immune response. By integrating the identity and magnitude of clones over time within an individual or across different cohorts, a detailed picture of repertoire dynamics can be obtained. Areas of application include:

- Minimal Residual Disease
Assessing MRD in hematological cancers
- Personalised repertoire monitoring
Quantify changes to an individual's repertoire as a result of disease, (immuno-)therapy or vaccination
- Patient stratification
Stratify patients for immunotherapy using repertoire biomarkers
- Immunotherapy development
Accelerate and optimize the development of new immunotherapies

Adaptive Error Correction

In order to perform repertoire clonality analysis, the sequencing errors that occur during high-throughput sequencing have to be addressed. By learning sample-specific mutation rates from the data, erroneous sequences can be corrected to provide accurate clone sequences, sizes, and frequencies (as shown in (5)).

Figure 1. Illustration of sequencing error in relation to clone sizes.



Benchmark

In order to test the accuracy of the ImmunoGenomiX Platform, we generated a synthetic dataset consisting of 26600 T cell clones. 10⁶ synthetic NGS reads were generated by simulating the sequencing process of an Illumina MiSeq sequencer with a reagent kit V3 and paired-end amplicon sequencing technology.

Table 1. Benchmark results using synthetic data.

Reads with not identifiable CDR3s	0.16%
In frame CDR3s matching the reference	95.77%
Total CDR3s matching the reference	95.12%
V Genes correctly annotated	99.31%
J Genes correctly annotated	99.98%
Clones correctly identified by CDR3	98.23%

References

- Liu et al, Cell 2017, 10.1016/j.cell.2017.01.014
- Liu et al, Cell Biol Tox 2018, 10.1007/s10565-018-9426-0
- Kurz et al, Blood 2015, 10.1182/blood-2015-03-635169
- Roschewsky et al, Lan Oncol 2015, 10.1016/s1470-2045(15)70106-3
- Gerritsen et al, Bioinf 2016, 10.1093/bioinformatics/btw339