CORRELATIONS OF ANTIBODY RESPONSE PHENOTYPE TO GENOTYPE REVEALED BY MOLECULAR AMPLIFICATION FINGERPRINTING

Sai T. Reddy, ETH Zurich, Department of Biosystems Science and Engineering, Switzerland
sai.reddy@ethz.ch
Tarik A. Khan, ETH Zurich, Department of Biosystems Science and Engineering, Switzerland
Simon Friedensohn, ETH Zurich, Department of Biosystems Science and Engineering, Switzerland
Arthur R. Gorter de Vries, ETH Zurich, Department of Biosystems Science and Engineering, Switzerland

Key Words: vaccine profiling, antibody repertoire, next-generation sequencing, systems immunology

It has long been possible to measure the phenotype of antibody responses (antigen-specific titers) through conventional serological assays (e.g., ELISA). In contrast, the ability to measure the genotype of antibody responses has only recently become possible through the advent of high-throughput antibody repertoire sequencing (Ig-seq), which provides quantitative molecular information on clonal expansion, diversity and somatic hypermutation. However, Ig-seq is compromised by the presence of bias and errors introduced during library preparation and sequencing and thus prevent reliable immunological conclusions from being made. By using synthetic antibody spike-in genes, we determined that Ig-seq data overestimated antibody diversity measurements by up to 5000-fold and was less than 60% accurate in clonal frequency measurements.

In order to overcome the widespread inaccuracies in Ig-seq, we developed a method known as molecular amplification fingerprinting (MAF). This consists of stepwise incorporation of unique molecular identifiers (UID), which starts with tagging first-strand cDNA during reverse transcription with a reverse-UID (RID), thus providing a unique tag to each transcript. Importantly MAF continues by tagging each DNA-RID molecule during multiplex-PCR amplification with a forward-UID (FID). This fingerprint of amplification for each molecule allowed us to implement an algorithm to normalize multiplex-amplification bias effects. We used several bioinformatic steps for error correction, this resulted in nearly absolute (98%) or absolute (100%) correction of intraclonal and clonal variants, respectively (Fig. 1a, b). Furthermore, by tagging molecules during amplification we were able to develop a novel algorithm for bias correction, this resulted in 98% accuracy of antibody clonal frequencies (Fig. 6c, d).

In order to probe the relationship between antibody genotype and phenotype, we have performed extensive *in vivo* experiments from immunized mice. Specifically, we modulated titer (phenotype) by varying the number of immunizations (boosts) and in parallel performed MAF Ig-seq to assess genotype through a series of quantitative metrics. These metrics describe clonal selection, expansion, and somatic hypermutation. Importantly, we also used different protein antigens, this allowed us to determine how epitope complexity impacted the genotype of antibody repertoires. Finally, by applying multivariate modeling using MAF-corrected data, we were able to predict the immune status of individual antibody clones. This extensive systems-based analysis demonstrates how accurate Ig-seq provides insight on the genotype and phenotype of humoral immunity.

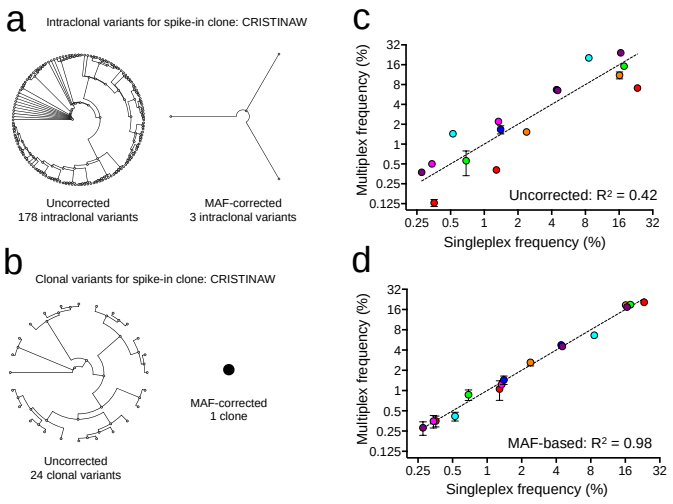

Figure 1. MAF error and bias correction validation. (a) Phylogenetic trees before and after MAF-error correction of intraclonal variants for a single spike-in example clone. (b) Phylogenetic trees before and after MAF-error correction of clonal variants (CDR3 a.a.) for a single spike-in example clone. (c) Correlation of uncorrected spike-in clonal frequencies from multiplex-PCR versus singleplex-PCR results in an $R^2 = 0.42$. (d) Correlation of MAF bias corrected spike-in clonal frequencies from multiplex-PCR versus singleplex-PCR results in a significantly improved $R^2 = 0.98$.