Serimmune

SERA for Sample Profiling and Identification

Introduction

In laboratory settings where many clinical samples are managed, it is imperative to maintain the identity of a sample as it is processed to ensure that the correct results are returned to researchers or clinicians. Mislabeled samples can result in costly or harmful research and clinical errors that should be avoided. While DNA has been in some settings to track and identify samples, sample DNA requirements and costs make this form of identification often prohibitive.

SERA Technology

While antibody profiles have been used to identify samples, the number of antibodies may not be sufficiently large to positively match samples in large cohorts. In contrast, Serum Epitope Repertoire Analysis (SERA), identifies to 1-3 million peptides that form an immune signature for each sample. These peptides can be used to "fingerprint" both the sample as well as the subject, as they collectively represent enriched epitopes that are unique to each individual and are remarkably stable over extended time periods.

SERA for Sample Identification Method

We have found that each individual has an antibody epitope repertoire that represents an immune history of exposures and immune responses that is unique to that individual. The reproducibility of the SERA assay, combined with this stable repertoire, allows for the identification of samples that have been obtained from the same subject over multiple longitudinal draws. In a large cohort of hundreds of individuals, each individual immune signature is specific enough so that it can be used to positively match samples that have been drawn from the same subject.

How it works

Each sample is processed through SERA, resulting in 1-3 million unique peptides per sample. These peptides are the basis for identifying epitopes that are enriched in the sample. Using epitopes that are highly enriched in each individual, epitopes that are common in populations, and epitopes that are present in known conditions, we are able to match samples that are originate from the same individual. The extremely large number of epitopes used in the analysis ensures that high correlations between epitope signals represent true matches. For a given cohort, samples that originate from the same individual, such as those collected before and after vaccination or therapy, are found to have a high correlation. In contrast, samples from other subjects have low correlation and do not match (Figure 1).

Systematic analysis of samples in a large cohort: Mismatch and Duplicate Identification

In a large cohort, all samples can be compared against all other samples to easily validate that samples are labeled to originate from the correct subjects. This is easily visualized in an N x N matrix, where the diagonals represent the same sample matched against itself (identity), and samples are grouped together by subject. Samples from the same subject are seen to have very high correlation with each other but low correlation with samples from other subjects (Figure 2,3).

Figure 1

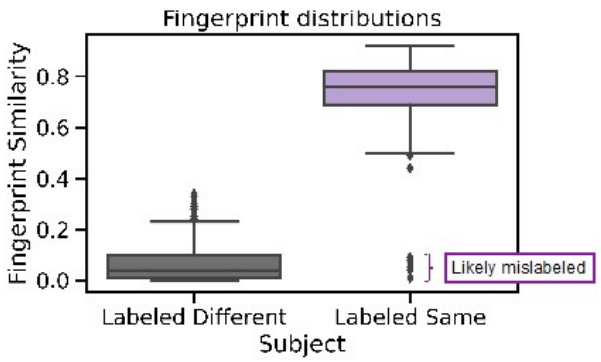


Figure 2a

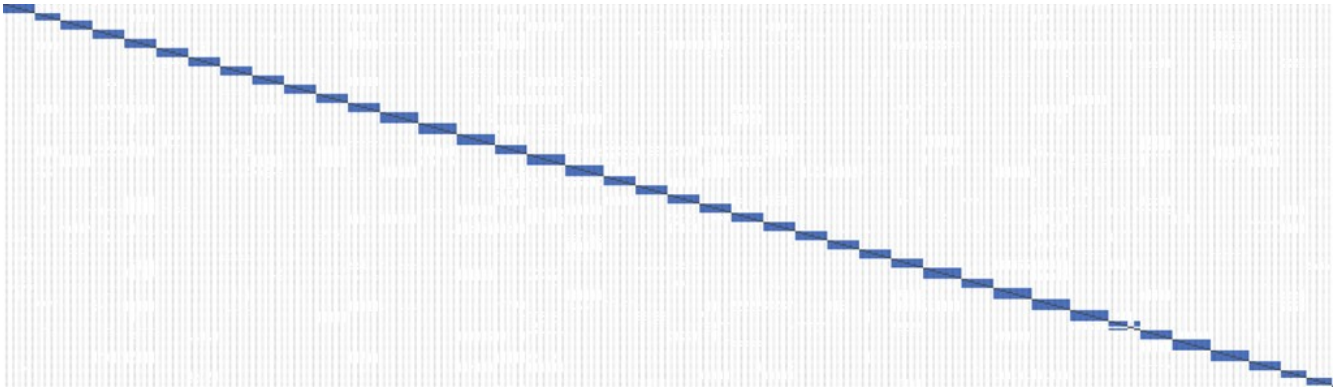


Figure 2b

Subject ID			Subject 1						Subject 2				
	Sample ID		Sample1	Sample2	Sample3	Sample4	Sample5	Sample6	Sample7	Sample8	Sample9	Sample10	Sample11
		Timepoint	1	2	3	4	5	6	1	2	3	4	5
	Sample1	1	1	0.868	0.834	0.822	0.794	0.8	0.103	0.123	0.138	0.101	0.111
	Sample2	2	0.868	1	0.848	0.827	0.821	0.8	0.103	0.127	0.143	0.11	0.118
	Sample3	3	0.834	0.848	1	0.809	0.828	0.817	0.101	0.123	0.135	0.098	0.11
	Sample4	4	0.822	0.827	0.809	1	0.872	0.76	0.099	0.121	0.134	0.097	0.106
	Sample5	5	0.794	0.821	0.828	0.872	1	0.793	0.098	0.124	0.134	0.094	0.104
Subject 1	Sample6	6	0.8	0.8	0.817	0.76	0.793	1	0.104	0.128	0.141	0.101	0.11
	Sample7	1	0.103	0.103	0.101	0.099	0.098	0.104	1	0.706	0.704	0.852	0.8
	Sample8	2	0.123	0.127	0.123	0.121	0.124	0.128	0.706	1	0.893	0.735	0.782
	Sample9	3	0.138	0.143	0.135	0.134	0.134	0.141	0.704	0.893	1	0.743	0.789
	Sample10	4	0.101	0.11	0.098	0.097	0.094	0.101	0.852	0.735	0.743	1	0.854
Subject 2	Sample11	5	0.111	0.118	0.11	0.106	0.104	0.11	0.8	0.782	0.789	0.854	1

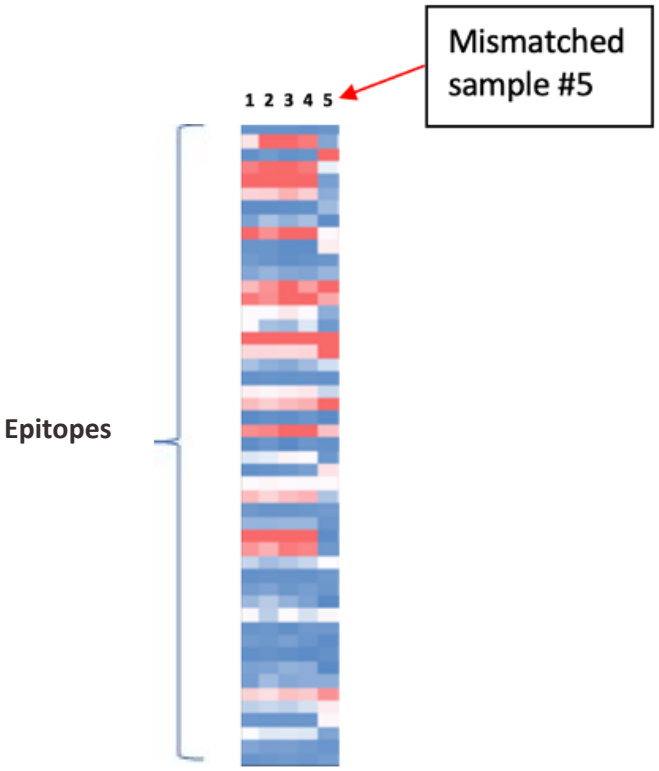
When samples that are labeled to originate from the same subject have low correlation and do not match, this can indicate that a sample has been mishandled or mislabeled (Figure 3). Conversely, when samples that are labeled to originate from different subjects show high correlation, this can indicate that a sample has been inadvertently mislabeled or duplicated.

Figure 2c

Subject ID			Subject71					Subject72				
	Sample ID		Sample343	Sample344	Sample345	Sample346	Sample347	Sample348	Sample349	Sample350	Sample351	Sample352
		Timepoint	1	2	3	4	5	1	2	3	4	5
	Sample343	1	1	0.771	0.752	0.054	0.776	0.043	0.054	0.049	0.044	0.046
	Sample344	2	0.771	1	0.746	0.063	0.836	0.092	0.098	0.082	0.096	0.08
	Sample345	3	0.752	0.746	1	0.051	0.758	0.047	0.056	0.05	0.044	0.046
	Sample346	4	0.054	0.063	0.051	1	0.058	0.02	0.039	0.034	0.018	0.03
Subject71	Sample347	5	0.776	0.836	0.758	0.058	1	0.091	0.098	0.08	0.089	0.077
	Sample348	1	0.043	0.092	0.047	0.02	0.091	1	0.785	0.75	0.778	0.722
	Sample349	2	0.054	0.098	0.056	0.039	0.098	0.785	1	0.854	0.769	0.791
	Sample350	3	0.049	0.082	0.05	0.034	0.08	0.75	0.854	1	0.788	0.846
	Sample351	4	0.044	0.096	0.044	0.018	0.089	0.778	0.769	0.788	1	0.842
Subject72	Sample352	5	0.046	0.08	0.046	0.03	0.077	0.722	0.791	0.846	0.842	1

To confirm a mismatch, specific epitope signals from the individuals can be compared to each other to confirm a mismatch as seen in figure 4 below.

Figure 3



Columns 1-5 in in the figure represent individual samples that are labeled to be from the same subject. Each row represents a unique epitope. Cells are colored according to strength of signal, with strong signals in red and weak signals in blue. Samples 1-4 are well correlated with most epitopes, whereas epitope signals for 5 are divergent from 1-4, indicating a mismatch.