SNPrint: Customizable Genotyper for Clinical Sample Tracking and Disease Progression Monitoring for NGS-Based Targeted Gene Panels

Joshua Wemmer[1], Paulina Sanchez[2], and Julian D'Angelo[1] | [1]Invivoscribe Inc., [2]LabPMM LLC, San Diego, CA, USA

Introduction

- Clinical labs performing regular genetic testing to monitor and evaluate for refractory and relapsed disease rely on accurate sample labeling for quality assurance. However, sample swapping may occur at any step between sample collection and sequencing.
- For next generation sequencing (NGS) data, one strategy to ensure that data corresponds to the expected individual, is to genotype based on single nucleotide polymorphisms (SNPs). For NGS-based targeted gene panels (TGPs), specific SNPs used for genotyping are constrained by the genes targeted in the assay.
- During hematological testing in our clinical lab, LabPMM®, we have created SNPrint™, a customizable genotyper developed for our NGS-based TGPs, MyAML® and MyMRD®.

Methods

- SNPrint genotypes 47 SNPs shared across two of Invivoscribe's heme-oncology TGPs, MyAML® and MyMRD®. However, any SNPs can be used for universal implementation across TGPs.
- SNPrint takes in binary alignment map (BAM) files as input, determines the base call and variant read frequency (VRF) for each SNP and generates a genetic barcode for each sample.
- An additional quality check is performed by ensuring each SNP meets a minimum read depth threshold.
- To ensure that a individual SNPrint belongs to an expected patient, they are searched against an internal database of SNPrints to match those that are identical.
- Additional ad-hoc analyses may be required to resolve complex sample-to-patient discrepancies.

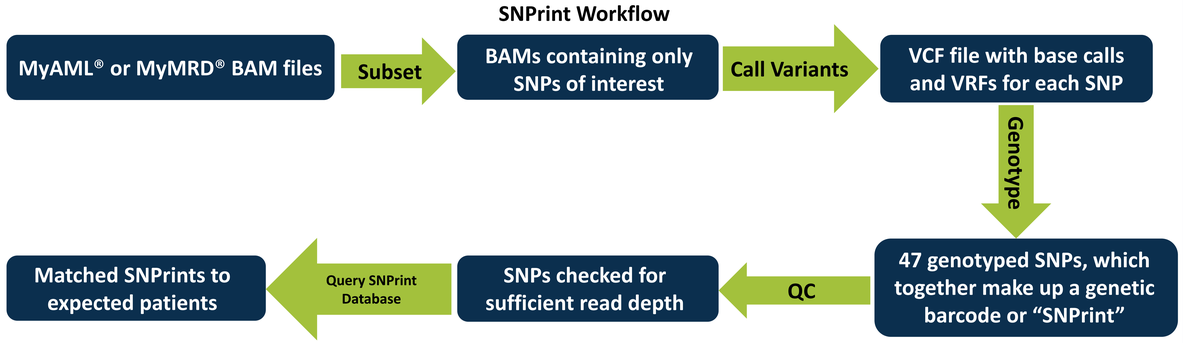

Figure 1. SNPrint Workflow. SNPrint takes in BAM files as input, subsets them to retain only SNPs of interest, and calls variants to obtain a VCF file with base calls for each SNP. The collective base calls together make up a genetic barcode, or "SNPrint".

What is a SNPrint?

A SNPrint is a genetic barcode that consists of 47 genotyped SNPs which together can be used to search against other SNPrints and determine sample integrity.

Comparing two SNPrints to identify mismatching and ambiguous SNP genotypes:

AA TT CC GG CC GG GG CC AA AA AA TT TG TT CT GG TT CC GG TT TT AC GA GG CC GG …
AA TT CC GG CC GG GG CC CA AA AA TT TG TT CX GG TT CC GG TT TT AC NN GG CC GG …

Mismatching SNPs, due to samples possibly not belonging to the same patient ID

X= Ambiguous Genotype (due to possible sample contamination or bone-marrow transplant)

N= Insufficient read depth at position (common with low QC spec samples)

Figure 2. Mismatching, Ambiguous and low read depth SNPs. Given two samples with the same patient ID, we expect both SNPrints to be identical. However, in the case of a sample-swap, contamination or possible bone marrow transplant we can expect to see discordant SNP genotypes. The above SNPrints show examples of when a genotyped SNP from two samples are mismatching (red arrow), ambiguous due to sample contamination or possible bone marrow transplant (purple arrow, marked by "X"), or undetermined due to insufficient read depth (green arrow, marked by "N").

Results

- SNPrint has helped resolve issues around sample swapping, but also has the added benefit of identifying whether a patient has recently undergone a possible bone marrow transplant.
- If a patient has not undergone bone marrow transplant and no sample swapping has occurred, we expect two samples with the same unique patient ID to have identical barcodes.
- However, these aforementioned cases can result in discordant barcodes containing multiple ambiguous SNP genotypes.
- In these ambiguous cases it is possible to see SNPs with no distinct genotype called, due to VRFs falling between established thresholds (i.e. homo/heterozygous frequencies).
- Further analysis of each VRF is then performed to determine the true genotypes for ambiguous SNPs.

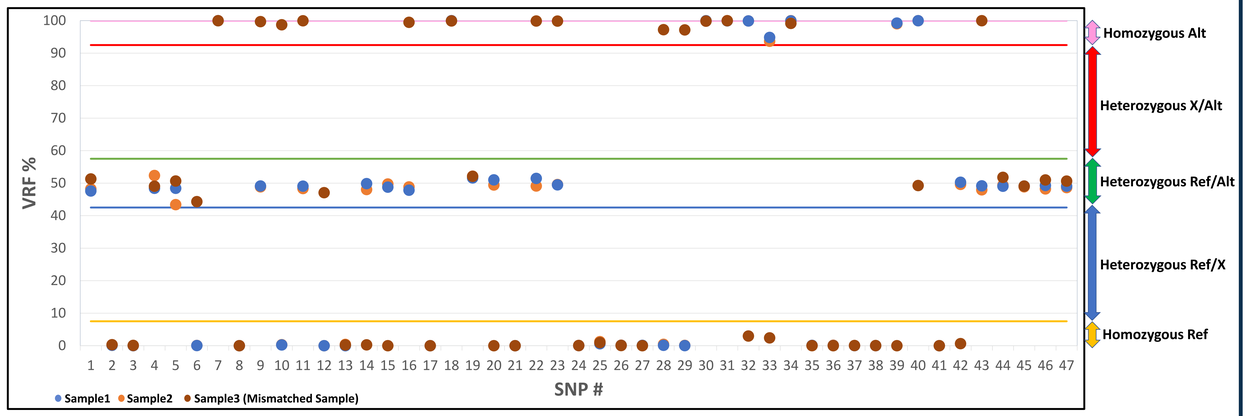

Figure 3. SNP VRFs for three samples with the same patient ID. For "Sample1" and "Sample2" (orange and blue data points), it is clear that these indeed belong to the same patient as the SNP genotypes are identical for all 47 SNPs. For "Sample3" (mahogany colored data points), it is clear that this sample does not belong to the same patient as many of the SNP genotypes fall in completely different established VRF thresholds.

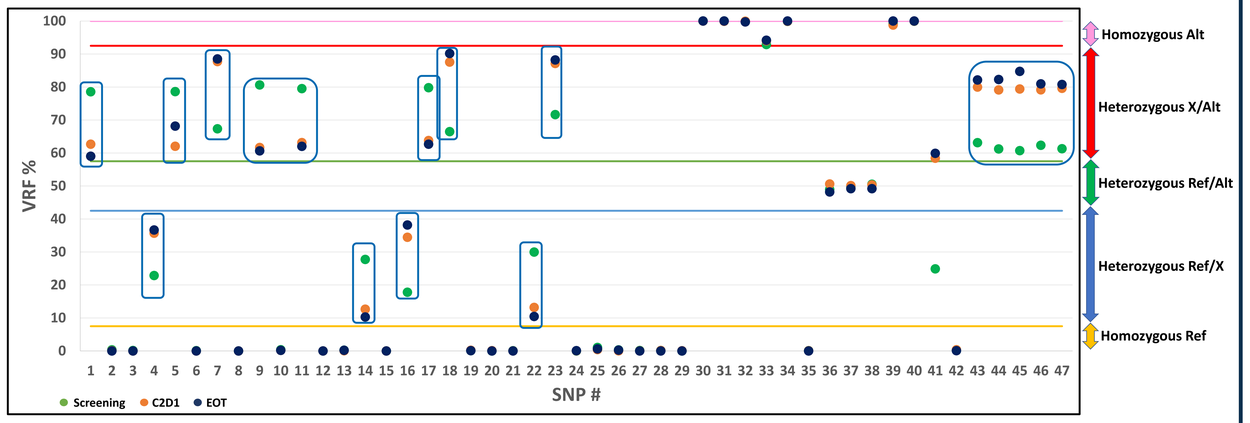

Figure 4. Three samples from the same patient across different timepoints during treatment (Initial screening [Screening], Cycle 2 Day 1 of treatment [C2D1] and End of Treatment [EOT]). For all three samples, we observe many Ambiguous "X" SNP genotypes which is indicative of either sample contamination or possible bone-marrow transplant which results in a "mixed" genotype. Consistent VRF differences are observed for the ambiguous "X" SNP genotypes (shown in blue boxes). From this ad-hoc analysis our team can infer that this patient likely underwent a bone marrow transplant thus resulting in "mixed" SNP genotypes that fall between homo/heterozygosity VRF thresholds for many of these SNPs.

Conclusions

- SNPrint has allowed for robust and streamlined QC of clinical samples, and has been proven to mitigate sample swap issues.
- SNPrint has the added benefit of being able to resolve sample-to-patient discrepancies that may arise due to bone-marrow transplants.
- Lastly, the customizability of SNPrint lends itself to ubiquitous implementation for any NGS-based TGP.