GenoCells: individual somatic cell counts of dairy cows by sequencing tank milk

Coppieter Wouter[1], Latour Samuel[1], Georges Michel[1], Perrin Florent[2], Davière Jean-Bernard[3], Lallemand Tanguy*[3]

[1] Université de Liège, [2] Agreelia, [3] Seenovia

GenoCells® offer the possibility to determine somatic cells count for each cow by genotyping the bulk tank sample. GenoCells® technology is now accessible via sequencing, making it possible to target larger herds (>1000 cows).

We implemented a GenoCells® deconvolution workflow in a 500-cow herd using Illumina NovaSeq 6000 short-read WGS of tank milk to estimate cow-specific DNA fractions and infer individual SCC. Cows were genotyped with a 50K SNP array and imputed to WGS-density markers using Beagle against a phased reference panel. Tank reads (49x) were mapped with BWA and at each imputed SNP position, reference and alternate allele counts were extracted and converted to B-allele frequencies (BAF). We fitted a weighted least-squares mixture model in which each tank BAF is expressed as a linear combination of cow allele dosages, with per-SNP weights proportional to read depth to account for shallow coverage noise. DNA fractions were constrained to be non-negative and to sum to one. Cow SCC was reconstructed by combining each estimated DNA fraction with routine milk-yield records and the measured tank SCC ($SCC_i = SCC_{tank} * V_{tank} * f_i / V_i$). To quantify the sequencing-depth requirement for large tanks, the 49x dataset was downsampled in 5x steps and benchmarked against SCC reference values.

Predictions remained stable at 5x ($R^2=0.97$ comparable to 50x, demonstrating that moderate coverage is sufficient to resolve contributors even at 500 cows when dense cow genotypes are available. At 1x , accuracy dropped ($R^2=0.72$) and dispersion increased, particularly among low-SCC cows, consistent with higher sampling variance in allele-frequency estimates and reduced leverage of rare or low-coverage SNPs. Nevertheless, the 1x setting still retained useful signal for screening and could trigger targeted confirmatory sampling. Overall, GenoCells® deconvolution turns one tank sample into cow-level SCC phenotypes and ranked alerts for subclinical mastitis. For recording organizations, the approach leverages existing 50K genotyping programs and routine yield/tank SCC measurements to enable higher-frequency monitoring with minimal additional on-farm sampling. In our 500-cow context, a practical operating point is approximately 5x tank coverage on NovaSeq 6000, balancing cost and accuracy while remaining compatible with automated pipelines in routine dairy herds, now.