WGS Core Control Set

Maximize insights from every sample with Sequins™ internal standards

Background

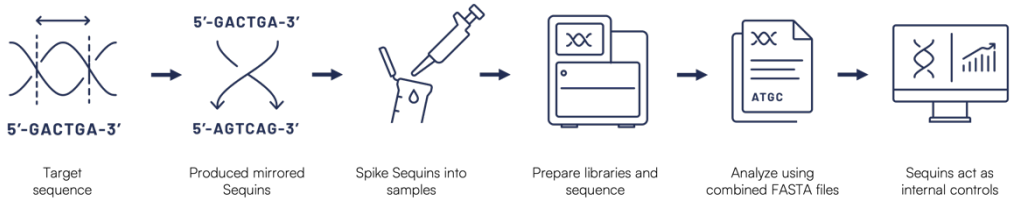
Next-generation sequencing (NGS) can be used to identify genetic variation and disease-associated mutations and has become a principal tool in biomedical research and clinical testing. However, numerous factors influence the accuracy of variant detection using NGS including sequencing depth, read length, sequencing errors, and PCR amplification biases introduced during library preparation. It is therefore imperative that a system of control standards is incorporated to account for accumulated errors, and improve data quality and interpretation, maximizing genomic insights.

Introduction to Sequins

Sequins (sequencing spike-ins) are synthetic nucleic acid controls that directly mirror naturally occurring sequences. Because Sequins retain the same nucleotide composition as the natural sequence, they enable accurate representation of genomic complexity without compromising the integrity of the sample and results. Sequins perform equivalently throughout sequencing workflows, providing a true measure of control.

Sequins' innovative design enables the production of synthetic mirrored sequences that directly represent almost any genomic feature, in any organism with a reference genome. This includes common and clinically relevant variants and analytically challenging regions of the genome. By combining sequins in precise ratios, quantitative features of genome biology, such as variant allele frequencies or copy-number variation, can also be emulated.

Sequins are simply 'spiked-in' to a sample prior to library preparation and progressed together through a workflow. Sequins controls can then be distinguished from the native sample in the output library by virtue of their synthetic sequence enabling standardization and comparison between samples, runs, laboratories, chemistries, and sequencers.



Schematic showing the design and use of Sequins in an NGS workflow.

SQNM_PSWCCS_2509_1

Sequins for WGS

Whole Genome Sequencing (WGS) has become a powerful tool for identifying disease-causing genetic variants. From newborn screening to cancer genome profiling and national health databasing programs, WGS has seen a rapid uptake globally in centralized laboratories in government, clinical, and research settings. Given the varied source of samples across multiple laboratories and cohorts, the need to standardize WGS is paramount. In research settings, cohorts will also span significant timeframes ensuring the need for longitudinal comparability.

The Sequins WGS Core Control Set is a comprehensive, easy-to-use, pre-configured control set comprising a comprehensive range of variant classes, with a focus on difficult regions of the genome to sequence. Variant classes include simple repeats; homopolymers; structural variants; common genetic variants; microsatellites; and mitochondrial DNA.

Sequins are compatible with most short- and long-read sequencing technologies using both PCR- and PCR-free based library preparation methods. In PCR-based workflows, Sequins can provide insights for introduced biases.

A Sequins-specific FASTA file is provided for simple concatenation with a reference genome for alignment and further analysis using standard pipelines and tools.

Sequins WGS Core Control Set Content

Variant Class	Count	Description
Difficult variants	11	Germline variants at simple repeats: – Homopolymers (mono-nucleotide repeats) – Di-nucleotide repeats – Tri-nucleotide repeats – Quad-nucleotide repeats – Low and high GC% regions
Structural variants	9	Tandem duplications, deletions, and inversions >50bp in size, long tandem repeats
Microsatellites	1	Stable microsatellite sequences
Mitochondrial DNA	4	Reference sequences representing the entire mitochondrial genome
Common genetic variants	58	Representative of general background genetic variation

Sequins are manufactured and configured in a preset ratio of wildtype to variant alleles to reflect the heterozygous state and are spiked into the sample at approximately 1% of the total calculated gDNA input to a 250ng library. Because Sequins are subject to the same technical variables as the accompanying biological sample, they can be used to assess the impact of laboratory and bioinformatic variables at any stage of a WGS workflow. Sequins can measure the performance (e.g. accuracy and precision) of a given WGS assay, enable rapid troubleshooting and operational quality control, and act as reference factors by which to standardize between multiple samples.

Benefits of Sequins for WGS

Standardization within and between samples, users, equipment and locations	The use of Sequins controls for standardization mitigates the heterogeneity of samples to enable unprecedented interoperability.
Workflow monitoring and optimization	Sequins are subjected to the same technical influences and errors as the samples they are combined with, enabling the evaluation of workflow performance.
Enhanced data insights	Sequins are uniquely designed to represent specific genomic features for accurate determination of data quality, enhancing the ability to confidently and reliably call variants.

Cross-Platform Confidence: Illumina and ONT Aligned

The Sequins WGS Core Control Set has been validated for compatibility with Oxford Nanopore Technologies (ONT) whole genome sequencing using the Genome in a Bottle (HG002) reference standard. Sequins were spiked into human genomic DNA and processed with the EPI2ME wf-human-variation workflow and maintained variant calling accuracy (0.9635 and 0.8974 for the HG002 genome; and 1.0000 and 0.9516 for the Sequins target regions, for precision and sensitivity, respectively). All expected variants within the Sequins regions were detected, with only three false negatives observed in homopolymer regions. Similarly, allele frequencies for sentinel single nucleotide variants were consistent with Illumina short-read data (0.9987 and 0.9940 for the HG002 genome; and 1.0000 and 1.0000 for the Sequins target regions, for precision and sensitivity, respectively). Together, Illumina and ONT results demonstrate that Sequins deliver cross-platform confidence, ensuring reliable variant detection and benchmarking across technologies.

Product Ordering Information

Product	Catalog Number	Description
WGS Core Control Set	PN-10004	WGS Core Control Set (24 samples*)
WGS Core Control Set	PN-10005	WGS Core Control Set (96 samples*)

**based on a 1% spike-in; 250 ng library input*

Contact Information

Ordering and Support

enquiries@sequins.bio
support@sequins.bio

Key Publications

Deveson, I.. Chen, W., Wong. T. et al. (2016) Representing genetic variation with synthetic DNA standards. *Nat Methods.* 13. 784- 791.

Hardwick SA, Deveson IW. Mercer TR. (2017) Reference standards for next-generation sequencing. *Nat Rev Genet.* 18(8):473-484.

Deveson, I.W., Madala, B.S., Blackburn, J. et al. (2019) Chiral DNA sequences as commutable controls for clinical genomics. *Nat Commun.* 10, 1342.

Blackburn, J., Wong. T., Madala, B.S. et al. (2019) Use of synthetic DNA spike-in controls (sequins) for human genome sequencing. *Nat Protoc.* 14, 2119-2151.