WGS Clinical Control Set

v1.0

Maximize insights from every sample with Sequins™ internal standards

Background

The American College of Medical Genetics and Genomics (ACMG) recommends reporting secondary findings in Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES) because certain variants are highly clinically actionable. The current ACMG list (v3.3[1] as of August 2025, updated from v3.2 June 2023 which contained 81 genes) includes 84 genes associated with serious but preventable or treatable conditions, spanning hereditary cancers, cardiovascular disorders and metabolic diseases. As such, detection of these variants during sequencing for other indications, enables preventative care and health benefits for patients and families.

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 (Figure 1). 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.

1. Lee, Kristy et al. ACMG SF v3.3 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). Genetics in Medicine, Volume 27, Issue 8, 101454

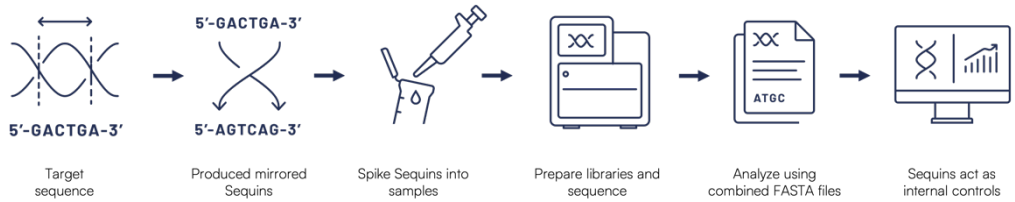

Figure 1. Schematic showing the design and use of Sequins in an NGS workflow.

Control for Secondary Findings using WGS

The Sequins WGS Clinical Control Set v1.0 is a spike-in control set that was designed based on the ACMG Secondary Findings v3.3 list, covering 80 genes and 96 conditions. The control set represents a range of variant classes, enabling robust assessment of variant detection, coverage and analytical performance across clinically important genes. This synthetic control set provides multiple benchmarking functions:

Postive Controls	Specific variants included in Sequins molecules emulate postive controls for specific clinically relevant variants.
Performance Proxies	Sequins variants act as proxies for other regions within the same gene, particularly those with similar coverage profiles or sequence complexity, and broader benchmarking for regions or genes with comparable genomic features such as GC content or variant type.
Dynamic Assay	Sequins enable dynamic control across a sequencing workflow, providing comprehensive QC performance in every sample.
Operational Control and Interopterability	Reliable and accurate controls, ensures consistent performance, regulatory readiness and data confidence across instruments, pipelines and workflows.

Developed to complement the Sequins WGS Core Control Set, the WGS Clinical Control Set v1.0 enables laboratories to align with clinical reporting standards for secondary findings in WGS with unparalleled accuracy, reproducibility, and ease of use. By incorporating engineered synthetic DNA standards directly into sequencing samples, Sequins provide in-sample validation of clinically important genes, supporting robust and reproducible detection and interpretation of secondary findings.

Integrated Controls for WGS

While the WGS Clinical Control Set can be used as a standalone control set, it is designed to complement the WGS Core Control Set. This integrated approach offers clinical focus and technical breadth, with the WGS Clinical Control Set focused on ACMG Secondary Findings genes, and the WGS Core Control Set providing a broad coverage of variant classes across the genome (including technically challenging regions such as high- and low-GC, repeats, homopolymers, repeats, microsatellites and structural variants). Used together, the two sets provide a comprehensive integrated control system, which offers:

- Clinical benchmarking for detection, interpretation and reporting of actionable variants
- Technical benchmarking for sensitivity, specificity and reproducibility across difficult regions of the genome
- Combined evaluation of assay performance, pipeline robustness and reporting accuracy

WGS Clinical Control Set v1.0 Content

Key Features

- **Comprehensive Coverage**: Sequins representing 80 of the 84 clinically actionable ACMG Secondary Findings (SF) genes across cancer, cardiovascular, metabolic, and rare diseases.
- **Flexible and Compatible**: Available as a standalone kit or in combination with the WGS Core Control Set for an integrated control system.
- **Extensive Variant Representation:** Covers pathogenic/likely pathogenic, VUS and benign/likely benign variants in ACMG-listed secondary findings genes, enabling robust benchmarking of variant detection and reporting pipelines.

Design Rationale

- **Clinically Anchored**: Variants from ACMG SF genes (pathogenic, benign, likely benign, VUS, synonymous) reflect real-world diagnostic practice.
- **Population Representation**: Includes 140 common GRCh38 variants to emulate realistic variant density.
- **Variant Diversity**: Covers SNVs, indels (including large indels) and a range of molecular consequences.
- **Strategic Selection**: Variants chosen across nine genomic contexts - ancestry, coding regions, GC content, structural complexity, repeats, mappability, segmental duplications, difficult regions defined according to the Genome in a Bottle (GIAB) benchmarks and chromosomal context for clinical relevance and technical challenge.
- **Broad Scope**: Non-disease-specific, representing a wide spectrum of clinically relevant genes and mutations.

Cardiovascular		Cancer		Cardiovascular Metabolic
ACTA2	*MYH11*	*APC*	*RET*	*GLA*
ACTC1	*MYH7*	*BMPR1A*	*SDHAF2*	*PRKAG2*
APOB	*MYL2*	*BRCA1*	*SDHB*	
BAG3	*MYL3*	*BRCA2*	*SDHC*	**Metabolic**
CALM1	*PCSK9*	*MAX*	*SDHD*	*BTD*
CALM2	*PKP2*	*MEN1*	*SMAD4*	*GAA*
CALM3	*RBM20*	*MLH1*	*STK11*	*OTC*
CASQ2	*RYR2*	*MSH2*	*TMEM127*	
DES	*SCN5A*	*MSH6*	*TP53*	**Rare Diseases**
DSC2	*SMAD3*	*MUTYH*	*TSC1*	*ACVRL1*
DSG2	*TGFBR1*	*NF2*	*TSC2*	*ATP7B*
DSP	*TGFBR2*	*PALB2*	*VHL*	*CACNA1S*
FBN1	*TMEM43*	*PMS2*	*WT1*	*ENG*
FLNC	*TNNC1*	*PTEN*		*HFE*
KCNH2	*TNNI3*	*RB1*		*HNF1A*
KCNQ1	*TNNT2*			*RPE65*
LDLR	*TPM1*			*RYR1*
LMNA	*TRDN*			*SMAD4*
MYBPC3	*TTN*			*TTR*

Sequins are manufactured and configured in a preset wildtype to variant allele ratio representing the heterozygous state. Sequins are spiked into the samples at 1% (mass/mass) of the total gDNA input for 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.

Sequins are compatible with most short- and long-read sequencing technologies using both PCR- and PCR-free based library preparation methods. A Sequins-specific FASTA file is provided for straightforward integration into reference genomes enabling alignment and analysis using standard pipelines and tools.

WGS Clinical Control Set v1.0 Performance

To assess the performance of Sequins relative to native human genomic DNA (hgDNA — Coriell NA12878), we compared read coverage across matched Sequins target regions after calibration. Sequins from the WGS Clinical Control Set v1.0 demonstrated strong concordance with Coriell NA12878, and this relationship was preserved when combined with the WGS Core Control Set (Figure 2). These results confirm that Sequins

recapitulate native sequence behaviour and can be integrated across control modules without compromising coverage.

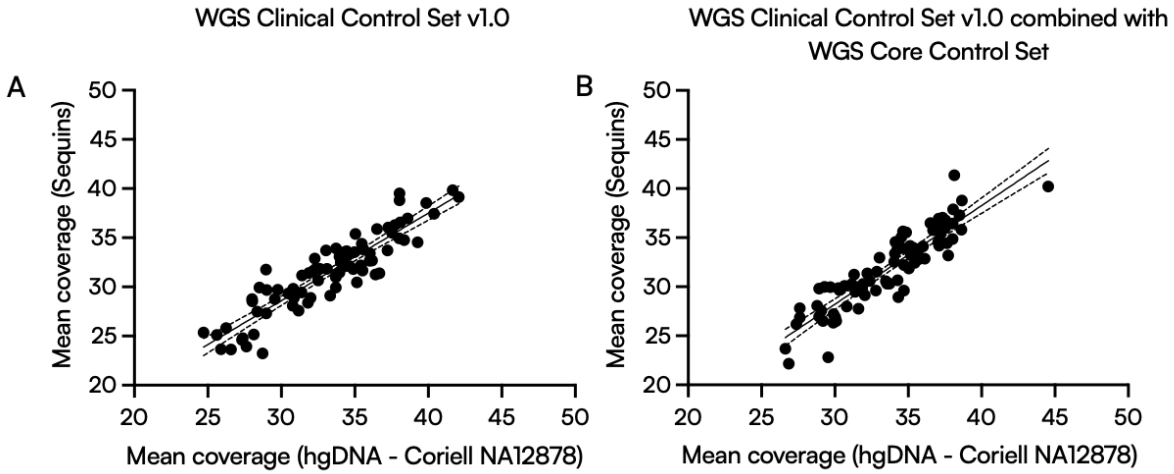

Figure 2. Mean coverage of Sequins and hgDNA (Coriell NA12878) at the same location. (A) Reads mapped to the Sequins WGS Clinical Control Set v1.0 strongly correlated with the same regions of Coriell NA12878 (slope 0.89, $R^2 = 0.83$). (B) This correlation was maintained when combined with the WGS Core Control Set (slope 1.00, $R^2 = 0.82$), confirming compatibility without loss of performance.

Variant calling across the full genome of Coriell NA12878 showed precision and sensitivity above 99% in all conditions. Importantly, combining the WGS Clinical Control Set v1.0 with the WGS Core Control Set did not alter the accuracy of Coriell NA12878 variant calls. The observed differences were within the expected background noise of large-scale variant calling pipelines.

Sample	TP	FP	FN	Precision	Sensitivity	F-measure
Coriell NA12878 + WGS Clinical Control Set v1.0	3,682,018	28,890	18,083	0.9922	0.9951	0.9937
Coriell NA12878 + WGS Clinical Control Set v1.0 + WGS Core Control Set	3,682,079	28,849	18,009	0.9922	0.9951	0.9937

TP — true positive, FP — false positive, FN — false negative

Within the Sequins target regions, variant calling performance was near-perfect. Precision reached 100%, sensitivity exceeded 99.7%, and the only false negative was an expected structural variant not detectable by the germline variant caller used. Results were identical when combined with the WGS Core Control Set.

Sample	TP	FP	FN*	Precision	Sensitivity	F-measure
Coriell NA12878 + WGS Clinical Control Set v1.0	385	0	1	1.0000	0.9974	0.9987
Coriell NA12878 + WGS Clinical Control Set v1.0 + WGS Core Control Set	385	0	1	1.0000	0.9974	0.9987

TP — true positive, FP — false positive, FN — false negative

*The single false negative was expected due to limitation of the variant caller for structural variant

Product Ordering Information

Product	Catalog Number	Description
WGS Clinical Control Set	PN-10024	WGS Clinical Control Set v1.0 (24 samples*)
WGS Clinical Control Set	PN-10025	WGS Clinical Control Set v1.0 (96 samples*)
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. Representing genetic variation with synthetic DNA standards. *Nat Methods*. 13. 784- 791 (2016)

Hardwick SA, Deveson IW. Mercer TR. Reference standards for next-generation sequencing. *Nat Rev Genet*. 2017 Aug:18(8):473-484. doi: 10.1038/nrg.2017.44. Epub 2017 Jun 19. PMID: 28626224

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

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

Lee, K., Abul-Husn, N. S., Amendola, L. M. et al. ACMG SF v3.3 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*, 27, 101454 (2025)