# Publication Guide

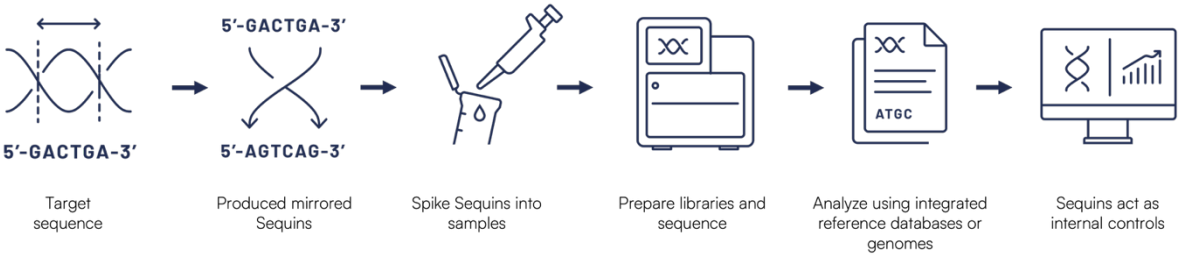
## The Future of Genomics Data Quality

## Background

Sequins™ is a privately held genomics technology company located in San Francisco, USA and Sydney, Australia. Born from a successful and widely published research program in Co-Founder Professor Tim Mercer's lab at the Garvan Institute in Sydney, Sequins has developed an innovative DNA and RNA sequencing control standards platform. Based on mirroring technology, Sequins controls can be spiked into individual samples without compromising sample integrity and results.

## Introduction

In the rapidly evolving landscape of Next-Generation Sequencing (NGS), achieving accurate, reliable, and reproducible results is paramount. Sequins (sequencing spike-ins) are synthetic nucleic acid controls that directly mirror naturally occurring sequences. Because Sequins retain the same nucleotide composition as natural sequences, they enable accurate representation of genomic complexity without compromising integrity of samples and results. Sequins perform equivalently throughout sequencing workflows, providing a true measure of control. Sequins' innovative design enables production of synthetic sequences that directly represent almost any genomic feature, in any organism with a reference genome. Products commercially available include Core Control Sets for Whole Genome Sequencing (WGS) and Metagenomics; and On-Demand custom designs. A circulating tumor DNA (ctDNA) control set is also currently in development.






SQNM_PG_2506_1

# Selected Publications

## General Sequencing

- Establish robust reference standards for NGS, ensuring accurate calibration and reliable assessment of sequencing performance.
- Measure error rates and improve the negative predictive value of sequencing results, minimizing false positives and negatives.
- Inform method development and performance characterisation with a trusted spike-in standard that complements regulatory considerations.

### Publication

Hardwick, S., Deveson, I. & Mercer, T. Reference standards for next-generation sequencing. *Nat Rev Genet* 18, 473—484 (2017).

**Summary**

Review publication that overviews the need for reference standards in NGS, covering various types of reference standard/controls and outlining key applications, use cases, and concepts behind the design and implementation.

**Abstract**

Next-generation sequencing (NGS) provides a broad investigation of the genome, and it is being readily applied for the diagnosis of disease-associated genetic features. However, the interpretation of NGS data remains challenging owing to the size and complexity of the genome and the technical errors that are introduced during sample preparation, sequencing and analysis. These errors can be understood and mitigated through the use of reference standards — well-characterized genetic materials or synthetic spike-in controls that help to calibrate NGS measurements and to evaluate diagnostic performance. The informed use of reference standards, and associated statistical principles, ensures rigorous analysis of NGS data and is essential for its future clinical use.

**Key Outcomes**

Reference standards must be commutable — perform comparably to samples undergoing testing — and can provide known truths with the difference between expected values and measured values yielding empirical estimate of variability. The abundance of a sequence in a patient sample can be determined by comparison to a reference sequence of known abundance, allowing standardization of measurements across multiple samples and diagnostic thresholds to be anchored to reference standards.

Detection of clinically relevant features in difficult regions of genome is limited and multiple technical variables are introduced during sample preparation, library construction, sequencing and analysis, that can lead to inaccurate analysis and misdiagnosis. Difficult regions often vary between individuals and host a range of clinically relevant mutations. Labs continue to validate variants with Sanger sequencing.

Common reference standards (e.g. cell lines) do not represent sufficient diversity and offer limited value as reference for most variants, and tumor samples can often contain multiple evolving clonal variants making them difficult to use as a reliable reference.

Spike-in controls are designed to be added to samples to undergo concurrent library prep, sequencing and analysis, acting as both internal qualitative and quantitative controls that are subject to the same downstream variables as the accompanying sample. They can be used as scaling factors for normalization between multiple samples in differential expression analysis for RNA-seq, assessing copy number variation between genomes, and normalizing microbial communities in environmental samples.

**Relevance to Sequins**

Sequins are commutable to the genome and as spike-in controls, provide a potentially better solution than parallel process controls, while also allowing for customizable content, like ladders or specified variants. Sequins can eliminate the biological noise of common reference materials.

## Publication

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

**Summary**

Publication outlines the concept of chiral sequences and mirroring — the underlying concept of Sequins — and introduces sequencing spike-ins (Sequins) as ideal controls based on the notion they act as a proxy for true human sequences. Demonstrates the equivalent performance of matched mirrored sequences with native sequences using PCR and NGS, and over a range of areas including whole genome, target enrichment, long read sequencing, clinical cancer samples, and their improvement of diagnostic specificity and performance.

**Abstract**

Chirality is a property describing any object that is inequivalent to its mirror image. Due to its 5'—3' directionality, a DNA sequence is distinct from a mirrored sequence arranged in reverse nucleotide-order, and is therefore chiral. A given sequence and its opposing chiral partner sequence share many properties, such as nucleotide composition and sequence entropy. Here we demonstrate that chiral DNA sequence pairs also perform equivalently during molecular and bioinformatic techniques that underpin genetic analysis, including PCR amplification, hybridization, whole-genome, target-enriched and nanopore sequencing, sequence alignment and variant detection. Given these shared properties, synthetic DNA sequences mirroring clinically relevant or analytically challenging regions of the human genome are ideal controls for clinical genomics. The addition of synthetic chiral sequences (sequins) to patient tumor samples can prevent false-positive and false-negative mutation detection to improve diagnosis. Accordingly, we propose that sequins can fulfill the need for commutable internal controls in precision medicine.

**Key Outcomes**

Mirrored Sequins perform equivalently during molecular and bioinformatic techniques including PCR, amplification, hybridization, whole genome, target enrichment, long read nanopore sequencing, sequence alignment, and variant detection.

Synthetic DNA sequences mirroring clinically relevant and analytically challenging regions of the genome are ideal controls for clinical genomics whereas, commonly used human reference cell lines cannot be added to patient samples without causing contamination. Common synthetic references do not recapitulate context and sequence-specific variables — whereas Sequins offers this “commutability”.

Using sequins, empirically determined sample-specific thresholds for variant filtering that would retain the maximum number of true positives while excluding all false positives, ultimately showing improvement in diagnostic specificity and performance.

Tested cancer driver mutations with sequins ladder using target enrichment and showed equivalent coverage of sequences and target regions; detected and quantified Sequins variants and known variants in the human sample with equivalent accuracy within the quantitative range of the ladder.

**Relevance to Sequins**

The concept of Sequins with mirroring/chirality preserves sequencing context to behave the same as target regions and act as a proxy source of truth. Sequins can be designed to target any region across a wide range of methods and applications, with the ability to increase confidence levels for variant calling.

## Publication

Reis, A.L.M., Deveson, I.W., Wong, T. et al. A universal and independent synthetic DNA ladder for the quantitative measurement of genomic features. Nat Commun 11, 3609 (2020).

**Summary**

Publication outlines the use of a synthetic DNA spike-in ladder that defines a quantitative standard encoded within a single molecule containing multiple unique artificial contiguous sub-sequences, each repeated at known copy numbers of 1x, 2x, 4x and 8x, to provide accurate quantitative scale across an 8-fold dynamic range. The ladder enables more accurate quantitative comparisons between samples and achieves best in class normalization with demonstrated examples across multiple applications (WGS, Targeted Enrichment, Shotgun Metagenomics), different sequencing technologies (ONT, Illumina), and multiple library prep methods (PCR-based, PCR-free, Transposon-based).

**Abstract**

Standard units of measurement are required for the quantitative description of nature; however, few standard units have been established for genomics to date. Here, we have developed a synthetic DNA ladder that defines a quantitative standard unit that can measure DNA sequence abundance within a next-generation sequencing library. The ladder can be spiked into a DNA sample, and act as an internal scale that measures quantitative

genetics features. Unlike previous spike-ins, the ladder is encoded within a single molecule and can be equivalently and independently synthesized by different laboratories. We show how the ladder can measure diverse quantitative features, including human genetic variation and microbial abundance, and also estimate uncertainty due to technical variation and improve normalization between libraries. This ladder provides an independent quantitative unit that can be used with any organism, application or technology, thereby providing a common metric by which genomes can be measured.

**Key Outcomes**

For WGS, the ladder was spiked into human NA12878 and calibrated such that the two copy number unit matched the median count of diploid human genome, enabling >90% of k-mers (genome wide) to be encompassed within the 8-fold dynamic range. The ladder accurately measured quantitative features in the human genome and trio analysis showed it could be used to identify significant biological differences between samples, with improvement in identification of quantitative fold differences.

Observed deviation from expected quantitative scale indicates scale of technical variation from various experimental sources and variability within the library. In metagenomics samples with mock community backgrounds, the ladder improved normalization across all methods tested for detection of fold differences between communities.

**Relevance to Sequins**

A synthetic DNA ladder with no homology to known sequences can be used across various applications to evaluate different sequencing technologies, measure genetic variation between samples, assess technical variation, compare composition of different microbial communities, and improve overall quantitative comparisons. Sequins can produce synthetic ladders tailored to user needs through our On-Demand program.

## Whole-Genome Sequencing (WGS)

- Enhance the accuracy of genetic variant detection with qualitative and quantitative controls.
- Standardize variant detection, improve analytical accuracy, and assess bioinformatic pipelines with validated performance.
- Evaluate sequencing accuracy in difficult-to-sequence regions of the human genome, such as repeats and microsatellites, with specifically designed Sequins.

## Publication

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

### Summary

Publication introduces Sequins, a set of synthetic DNA standards designed to improve the accuracy of NGS by serving as qualitative and quantitative controls for genetic variant detection. Sequins mimic human genetic variation, including single nucleotide variants, structural rearrangements, and copy-number variations, while aligning exclusively to an artificial in silico reference chromosome. This separation enables parallel analysis and provides an internal benchmark for measuring sequencing performance.

### Abstract

The identification of genetic variation with next-generation sequencing is confounded by the complexity of the human genome sequence and by biases that arise during library preparation, sequencing and analysis. We have developed a set of synthetic DNA standards, termed 'sequins', that emulate human genetic features and constitute qualitative and quantitative spike-in controls for genome sequencing. Sequencing reads derived from sequins align exclusively to an artificial in silico reference chromosome, rather than the human reference genome, which allows them to be partitioned for parallel analysis. Here we use this approach to represent common and clinically relevant genetic variation, ranging from single nucleotide variants to large structural rearrangements and copy-number variation. We validate the design and performance of sequin standards by comparison to examples in the NA12878 reference genome, and we demonstrate their utility during the detection and quantification of variants. We provide sequins as a standardized, quantitative resource against which human genetic variation can be measured and diagnostic performance assessed.

### Key Outcomes

Sequins were designed to represent a broad spectrum of genetic variations — including SNPs, indels, and splice variants — that mimic the complexity of natural human genetic features. These synthetic constructs serve as robust spike-in controls for next-generation sequencing assays, enabling accurate benchmarking, normalization, and quality assurance in genomic, transcriptomic, and epigenomic analyses by facilitating precise quantification and variant detection.

### Relevance to Sequins

Validates Sequins against well-characterized genomic data, demonstrating their ability to standardize variant detection, improve diagnostic accuracy, and assess bioinformatic pipelines.

## Publication

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).

### Summary

Publication details the use of Sequins to enhance the accuracy and reliability of human genome sequencing. Sequins are synthetic DNA sequences designed to mimic human genetic variation while remaining bioinformatically distinguishable. They are added to DNA samples before library preparation, allowing for real-time internal quality control throughout the sequencing workflow. The study provides a step-by-step protocol for integrating Sequins into both whole genome and targeted sequencing, including their dilution, addition to DNA samples, and bioinformatic processing.

### Abstract

Next-generation sequencing (NGS) has been widely adopted to identify genetic variants and investigate their association with disease. However, the analysis of sequencing data remains challenging because of the complexity of human genetic variation and confounding errors introduced during library preparation, sequencing and analysis. We have developed a set of synthetic DNA spike-ins—termed 'sequins' (sequencing spike-ins)—that are directly added to DNA samples before library preparation. Sequins can be used to measure technical biases and to act as internal quantitative and qualitative controls throughout the sequencing workflow. This step-by-step protocol explains the use of sequins for both whole-genome and targeted sequencing of the human genome. This includes instructions regarding the dilution and addition of sequins to human DNA samples, followed by the bioinformatic steps required to separate sequin- and sample-derived sequencing reads and to evaluate the diagnostic performance of the assay. These practical guidelines are accompanied by a broader discussion of the conceptual and statistical principles that underpin the design of sequin standards. This protocol is suitable for users with standard laboratory and bioinformatic experience. The laboratory steps require ~1—4 d and the bioinformatic steps (which can be performed with the provided example data files) take an additional day.

### Key Outcomes

The key outcomes of the study include demonstrating how sequins can be effectively used to benchmark and validate the performance of sequencing technologies, variant calling, and genome assembly. The authors provide detailed guidance on the design, preparation, and application of sequins to simulate a broad spectrum of genomic features, enabling robust quality control and quantitative assessment. They highlight the utility of sequins in detecting biases, assessing coverage, and enhancing the reproducibility and interpretability of sequencing data across diverse platforms and experimental conditions. This protocol supports both whole genome and targeted sequencing applications, thereby reinforcing the value of sequins as essential tools for genomic assay calibration and standardization.

**Relevance to Sequins**

Sequins help detect technical biases, assess diagnostic sensitivity and precision, and serve as normalization controls across samples, and are ideal for workflow optimization. Their mirrored design ensures they behave similarly to human DNA in sequencing reactions, making them valuable for evaluating variant detection and sequencing performance while avoiding interference with the primary sample analysis.

## Publication

Reis, A.L.M., Deveson, I.W., Madala, B.S. et al. Using synthetic chromosome controls to evaluate the sequencing of difficult regions within the human genome. *Genome Biol* 23, 19 (2022).

**Summary**

Publication evaluates the accuracy of NGS in analyzing challenging regions of the human genome using Sequins that have been designed to replicate complex genomic features such as repeats, microsatellites, HLA genes and immune receptors. Sequins were used to serve as benchmarks for assessing sequencing technologies and helped provide insights to enhance clinical diagnosis and treatment.

**Abstract**

*Background*. Next-generation sequencing (NGS) can identify mutations in the human genome that cause disease and has been widely adopted in clinical diagnosis. However, the human genome contains many polymorphic, low-complexity, and repetitive regions that are difficult to sequence and analyze. Despite their difficulty, these regions include many clinically important sequences that can inform the treatment of human diseases and improve the diagnostic yield of NGS.

*Results*. To evaluate the accuracy by which these difficult regions are analyzed with NGS, we built an in silico decoy chromosome, along with corresponding synthetic DNA reference controls, that encode difficult and clinically important human genome regions, including repeats, microsatellites, HLA genes, and immune receptors. These controls provide a known ground-truth reference against which to measure the performance of diverse sequencing technologies, reagents, and bioinformatic tools. Using this approach, we provide a comprehensive evaluation of short- and long-read sequencing instruments, library preparation methods, and software tools and identify the errors and systematic bias that confound our resolution of these remaining difficult regions.
*Conclusions*. This study provides an analytical validation of diagnosis using NGS in difficult regions of the human genome and highlights the challenges that remain to resolve these difficult regions.

**Key Outcomes**

Sequins are accurate references for evaluating complex genomic regions and can serve as ground-truth controls for pathogenic variants, filling gaps in existing references by representing rare and clinically relevant mutations.

While PCR amplification introduces errors, PCR-free approaches improve coverage, phasing, and structural variant detection, and therefore provide a more comprehensive library prep solution for NGS.

**Relevance to Sequins**

Sequins can be a valuable tool for comparison of library preparation methods and sequencing technologies. They provide the ability to assess rare mutations and can be used to address errors and biases in sequencing to improve diagnostic interpretation.

## Circulating tumor DNA (ctDNA)

- Evaluate the analytical performance of ctDNA sequencing assays for precision oncology, ensuring reliable detection of ctDNA.
- Monitor assay performance with Sequins controls that behave comparably to real cell free DNA, providing valuable insights into coverage and capture efficiency.
- Determine the sensitivity of ctDNA assays, particularly for low-abundance mutations, enabling accurate detection of rare variants.

### Publication

Deveson, I.W., Gong, B., Lai, K. et al. Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology. *Nat Biotechnol* 39, 1115—1128 (2021).

**Summary**

Publication systematically assesses the performance of ctDNA sequencing technologies using synthetic reference standards, including Sequins. The study benchmarks multiple commercial and academic ctDNA assays, focusing on their sensitivity, specificity, and quantitative accuracy for detecting low-frequency variants relevant to cancer diagnostics and treatment. By introducing controlled spike-in standards into plasma-like samples, the authors were able to highlight considerable variability in assay performance and identify methodological limitations affecting variant detection thresholds. The work underscores the need for standardized benchmarking tools and methods to support the clinical translation of ctDNA assays, with Sequins playing a critical role in ensuring analytical validity across diverse platforms and use cases in precision oncology.

**Abstract**

Circulating tumor DNA (ctDNA) sequencing is being rapidly adopted in precision oncology, but the accuracy, sensitivity and reproducibility of ctDNA assays is poorly understood. Here we report the findings of a multi-site, cross-platform evaluation of the analytical performance of five industry-leading ctDNA assays. We evaluated each stage of the ctDNA sequencing workflow with simulations, synthetic DNA spike-in experiments and proficiency testing on standardized, cell-line-derived reference samples. Above 0.5% variant allele frequency, ctDNA mutations were detected with high sensitivity, precision and reproducibility by all five assays, whereas, below this

limit, detection became unreliable and varied widely between assays, especially when input material was limited. Missed mutations (false negatives) were more common than erroneous candidates (false positives), indicating that the reliable sampling of rare ctDNA fragments is the key challenge for ctDNA assays. This comprehensive evaluation of the analytical performance of ctDNA assays serves to inform best practice guidelines and provides a resource for precision oncology.

**Key Outcomes**

Sequins were used to evaluate the analytical performance of ctDNA assays comparing multiple library prep, target enrichment and sequencing technologies.

Sensitivity in ctDNA assays is complex, with sequencing depth, coverage uniformity, and stringency (number of reads to make a call) interacting to determine overall performance and therefore, thresholds must be carefully set.

Sequins show equivalent performance to real genomic features with coverage near identical between the native sequence and synthetic Sequins. Capture of Sequins with probes for targeted sequencing showed the same variation observed in the "real" genome.

**Relevance to Sequins**

Sequins played a pivotal role as synthetic spike-in controls used to rigorously evaluate the analytical validity of circulating tumor DNA (ctDNA) sequencing assays. By incorporating Sequins that mimic clinically relevant mutations at known allelic frequencies into plasma-like samples, the study enables precise benchmarking of assay performance across multiple platforms. This approach provides a robust framework to assess sensitivity, specificity, and quantitative accuracy, especially for detecting low-frequency variants—crucial for early cancer detection and monitoring. The relevance of Sequins lies in their ability to simulate a wide range of genomic features and mutation types within a controlled and reproducible context, thus supporting standardization and regulatory evaluation of ctDNA assays for clinical use in precision oncology.

## Metagenomics

- Mimic microbial genomes and serve as reference standards for metagenomic analysis, enabling accurate characterization of complex microbial communities.
- Accurately measure fold-change differences and normalize data between samples, facilitating meaningful comparisons of microbial abundance.
- Benchmark and optimize new sequencing technologies, including nanopore long-read sequencing, for metagenomic applications.
- Support validation of new metagenomic workflows by improving quantitative accuracy and assay sensitivity.

## Publication

Hardwick, S.A., Chen, W.Y., Wong, T. et al. Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. *Nat Commun* 9, 3096 (2018).

### Summary

Publication introduces Sequins, which mimic microbial genomes and serve as reference standards for metagenomic analysis. Integrating Sequins into DNA samples before sequencing allows users to evaluate performance, compare samples, normalize data and assess sequencing technologies in the context of metagenomics sequencing. This approach addresses challenges posed by microbial community complexity and sequencing biases, enhancing the accuracy and reproducibility of metagenomic studies.

### Abstract

The complexity of microbial communities, combined with technical biases in next-generation sequencing, pose a challenge to metagenomic analysis. Here, we develop a set of internal DNA standards, termed “sequins” (sequencing spike-ins), that together constitute a synthetic community of artificial microbial genomes. Sequins are added to environmental DNA samples prior to library preparation and undergo concurrent sequencing with the accompanying sample. We validate the performance of sequins by comparison to mock microbial communities and demonstrate their use in the analysis of real metagenome samples. We show how sequins can be used to measure fold change differences in the size and structure of accompanying microbial communities and perform quantitative normalization between samples. We further illustrate how sequins can be used to benchmark and optimize new methods, including nanopore long-read sequencing technology. We provide metagenome sequins, along with associated data sets, protocols, and an accompanying software toolkit, as reference standards to aid in metagenomic studies.

### Key Outcomes

A Sequins ladder was specifically designed for metagenomic applications to mimic microbial community complexity and sequencing biases. performance was validated by comparing to mock microbial communities and real metagenome samples. The ladder enabled accurate measurement of fold change differences and quantitative normalization between samples. Sequins’ performance were also demonstrated with long-read (ONT) sequencing.

### Relevance to Sequins

Sequins designed to mimic microbial genomes can used as internal controls in metagenomic studies and enable accurate measurement of fold changes and normalization between samples, addressing potential biases in sequencing data. Sequins help evaluate and optimize sequencing technologies, including short- and long-read methods, and enhance microbial community analysis by providing a consistent benchmark for complex datasets.

## Publication

Davis, B.C, Vikesland P.J., Pruden A. Evaluating Quantitative Metagenomics for Environmental Monitoring of Antibiotic Resistance and Establishing Detection Limits. *Environmental Science & Technology* 59(12), 6192-6202 (2025).

### Summary

This study evaluates the use of quantitative metagenomics as a high-throughput, non-targeted method for monitoring antibiotic resistance genes (ARGs) in wastewater environments. They benchmarked this approach using meta sequins, which were spiked into wastewater samples to assess detection limits and quantify gene concentrations.

### Abstract

Metagenomics holds promise as a comprehensive, nontargeted tool for environmental monitoring. However, one key limitation is that the quantitative capacity of metagenomics is not well-defined. Here, we demonstrated a quantitative metagenomic technique and benchmarked the approach for wastewater-based surveillance of antibiotic resistance genes. To assess the variability of low-abundance oligonucleotide detection across sample matrices, we spiked DNA reference standards (meta sequins) into replicate wastewater DNA extracts at logarithmically decreasing mass-to-mass percentages (m/m%). Meta sequin ladders exhibited strong linearity at input concentrations as low as $2 \times 10^{-3}$ m/m% ($R^2 > 0.95$), with little to no reference length or GC bias. At a mean sequencing depth of 94 Gb, the limits of quantification (LoQ) and detection were calculated to be $1.3 \times 10^3$ and 1 gene copy per µL DNA extract, respectively. In wastewater influent, activated sludge, and secondary effluent samples, 27.3, 47.7, and 44.3% of detected genes were ≤LoQ, respectively. Volumetric gene concentrations and log removal values were statistically equivalent between quantitative metagenomics and ddPCR for 16S rRNA, *intI*1, *sul*1, CTX-M-1, and *van*A. The quantitative metagenomics benchmark here is a key step toward establishing metagenomics for high-throughput, nontargeted, and quantitative environmental monitoring.

### Key Outcomes

Meta sequins were spiked into wastewater influent, activated sludge, and secondary effluent DNA extracts at varying mass-to-mass percentages (m/m%), demonstrating strong linearity at input concentrations as low as $2 \times 10^{-3}$ m/m% ($R^2 > 0.95$), indicating reliable quantification even at low abundance levels. At a sequencing depth of approximately 94 Gb, the limit of quantification (LoQ) was established at $1.3 \times 10^3$ gene copies per µL DNA extract, and the limit of detection (LoD) was 1 gene copy per µL. Furthermore, gene concentrations of 16S rRNA, *intI1*, *sul1*, $bla_{\text{CTX-M-1}}$, and *vanA* measured using quantitative metagenomics were statistically equivalent to those obtained via droplet digital PCR (ddPCR), validating the accuracy and robustness of the quantitative metagenomics approach.

### Relevance to Sequins

Use of meta sequins allows for accurate calibration and normalization, and this was demonstrated through comprehensive evaluation across multiple wastewater matrices and sequencing depths, and with direct comparison with ddPCR.

## Publication

Langenfeld, K., Hegarty, B. Vidaurri, S. et al. Development of a quantitative metagenomic approach to establish quantitative limits and its application to viruses. *Nucleic Acids Research*, Volume 53, Issue 5 (2025)

### Summary

This study describes quantitative limits and its application to viruses in addition to developing QuantMeta for improving the accuracy and reliability of absolute abundance quantification in metagenomic datasets, using synthetic DNA standards.

### Abstract

Quantitative metagenomic methods are maturing but continue to lack clearly-defined analytical limits. Here, we developed a computational tool, QuantMeta, to determine the absolute abundance of targets in metagenomes spiked with synthetic DNA standards. The tool establishes (i) entropy-based detection thresholds to confidently determine the presence of targets, and (ii) an approach to identify and correct read mapping or assembly errors and thus improve the quantification accuracy. Together this allows for an approach to confidently quantify absolute abundance of targets, be they microbial populations, genes, contigs, or metagenome-assembled genomes. We applied the approach to quantify single- and double-stranded DNA viruses in wastewater viral metagenomes, including pathogens and bacteriophages. Concentrations of total DNA viruses in wastewater influent and effluent were >$10^8$ copies/ml using QuantMeta. Human-associated DNA viruses were detected and quantifiable with QuantMeta thresholds, including polyomavirus, papillomavirus, and crAss-like phages, at concentrations similar to previous reports that utilized quantitative polymerase chain reaction (PCR)-based assays. Our results highlight the higher detection thresholds of quantitative metagenomics (approximately 500 copies/µl) as compared to PCR-based quantification (approximately 10 copies/µl) despite a sequencing depth of 200 million reads per sample. The QuantMeta approach, applicable to both viral and cellular metagenomes, advances quantitative metagenomics by improving the accuracy of measured target absolute abundances.

### Key Outcomes

The QuantMeta pipeline enables absolute abundance quantification by using meta-sequins of known concentrations to convert relative read depths into copies per microliter. This approach was applied to measure DNA viruses in municipal wastewater influent and effluent, including human-associated pathogens such as polyomaviruses and papillomaviruses, as well as bacteriophages like crAss-like phages, with results aligning well with qPCR assay. Detection limits were assessed using both reference-based and contig-based methods, which, while higher than those of qPCR (~10 copies/µL), enable multiplex quantification of numerous targets simultaneously, with post-correction improving accuracy. Additionally, QuantMeta quantification of spiked marine phage HM1 closely matched droplet digital PCR (ddPCR) measurements, further validating the method.

### Relevance to Sequins

Meta sequins offer a standardized and reproducible approach for absolute quantification of microbial and viral targets. They can also be used to benchmark new pipeline for interpreting metagenomic data, which are valuable for environmental monitoring, wastewater epidemiology and virome studies.

## RNA Sequencing (RNA-seq)

- Establish quantitative thresholds for accurate transcriptome analysis, ensuring reliable measurement of gene expression.
- Determine thresholds for precise assessment of gene and isoform transcription levels, ensuring accurate isoform assembly and quantification.
- Define cut-offs for reliable fold-change evaluation and establish sensitivity thresholds for accurate detection of fusion transcripts, improving the accuracy of differential expression analysis.
- Establish detection thresholds for fusion transcripts in targeted RNA sequencing, enabling accurate identification of gene fusions.

## Publication

Hardwick, S., Chen, W., Wong, T. et al. Spliced synthetic genes as internal controls in RNA sequencing experiments. *Nat Methods* 13, 792—798 (2016).

### Summary

This study explores the applications of an RNA prototype Sequins control set in establishing quantitative thresholds for accurate transcriptome analysis. It demonstrates their utility in determining thresholds for precise assessment of gene and isoform transcription levels, ensuring accurate isoform assembly, defining cut-offs for reliable fold-change evaluation, and establishing sensitivity thresholds for the accurate detection of fusion transcripts.

### Abstract

RNA sequencing (RNA-seq) can be used to assemble spliced isoforms, quantify expressed genes and provide a global profile of the transcriptome. However, the size and diversity of the transcriptome, the wide dynamic range in gene expression and inherent technical biases confound RNA-seq analysis. We have developed a set of spike-in RNA standards, termed 'sequins' (sequencing spike-ins), that represent full-length spliced mRNA isoforms. Sequins have an entirely artificial sequence with no homology to natural reference genomes, but they align to gene loci encoded on an artificial in silico chromosome. The combination of multiple sequins across a range of concentrations emulates alternative splicing and differential gene expression, and it provides scaling factors for normalization between samples. We demonstrate the use of sequins in RNA-seq experiments to measure sample-specific biases and determine the limits of reliable transcript assembly and quantification in accompanying human RNA samples. In addition, we have designed a complementary set of sequins that represent fusion genes arising from rearrangements of the in silico chromosome to aid in cancer diagnosis. RNA sequins provide a qualitative and quantitative reference with which to navigate the complexity of the human transcriptome.

### Key Outcomes

In any RNA-seq experiment, plotting the fraction of assembled isoforms against their input concentration generates a dose-response curve, which can be used to determine the threshold for accurate isoform assembly. This study included two distinct Sequin mixtures—Mix A and Mix B—which contain the same set of synthetic genes and isoforms, but at predefined and differing relative abundances. These abundance differences simulate

differential expression and isoform-level variation in a controlled manner. By constructing the synthetic transcript mixes (Mix A and Mix B) with known fold-change differences across a broad dynamic range, they enabled precise benchmarking of RNA-seq sensitivity, quantification accuracy, and differential expression detection. To quantify technical performance, Limit of Detection Ratio (LODR - the smallest fold-change at which a transcript can be detected as differentially expressed at a given statistical confidence); Limit of Quantification (LOQ - The minimum expression level at which transcript quantification is accurate (i.e., observed counts are consistent with expected input across replicates); and Limit of Sensitivity (LOS - the lowest transcript abundance at which a transcript can be reliably detected (not necessarily quantified accurately) were assessed. LODR varied with transcript abundance: highly expressed Sequins could be reliably detected with small fold changes, whereas lower abundance transcripts required larger fold changes to be distinguished confidently. LOQ analysis demonstrated that transcripts below ~1 TPM (transcripts per million) exhibited significant variability and should be interpreted with caution. LOS was pipeline- and library-dependent, and correlated strongly with sequencing depth and aligner performance.

**Relevance to Sequins**

Sequins emulate both simple and complex splicing patterns, making them ideal for assessing isoform-level expression. Sequins revealed biases and errors in quantification and transcript assembly that were not evident using ERCC spike-ins. The study showed that using Sequins improves detection of technical variation, normalization issues, and tool-specific artifacts.

## Publication

Heyer, E.E., Deveson, I.W., Wooi, D. et al. Diagnosis of fusion genes using targeted RNA sequencing. *Nat Commun* 10, 1388 (2019).

**Summary**

Publication presents a targeted RNA sequencing approach for the accurate detection of gene fusions in clinical cancer samples. Their method uses a hybrid-capture panel targeting over 500 cancer-associated genes and incorporates synthetic spike-in RNA controls (Sequins) to assess assay performance, including sensitivity, dynamic range, and technical reproducibility. These internal controls enable benchmarking of fusion detection thresholds and quantification accuracy. The approach demonstrated high sensitivity and specificity across diverse cancer types, including sarcomas and leukemias, and supports the clinical utility of targeted RNA-seq for reliable fusion gene diagnosis and precision oncology applications.

**Abstract**

Fusion genes are a major cause of cancer. Their rapid and accurate diagnosis can inform clinical action, but current molecular diagnostic assays are restricted in resolution and throughput. Here, we show that targeted RNA sequencing (RNAseq) can overcome these limitations. First, we establish that fusion gene detection with targeted RNAseq is both sensitive and quantitative by optimising laboratory and bioinformatic variables using spike-in standards and cell lines. Next, we analyse a clinical patient cohort and improve the overall fusion gene diagnostic rate from 63% with conventional approaches to 76% with targeted RNAseq while demonstrating high

concordance for patient samples with previous diagnoses. Finally, we show that targeted RNAseq offers additional advantages by simultaneously measuring gene expression levels and profiling the immune-receptor repertoire. We anticipate that targeted RNAseq will improve clinical fusion gene detection, and its increasing use will provide a deeper understanding of fusion gene biology.

**Key Outcomes**

Commonly used human reference cell lines cannot be added to patient samples without causing contamination. Additionally, common synthetic references do not recapitulate context and sequence-specific variables. This is in contrast to synthetic RNA sequences (i.e. RNA Sequins) which mirror clinically relevant and analytically challenging regions of the genome which offer this “commutability” and are thus ideal controls for clinical genomics.

**Relevance to Sequins**

Sequins played a critical role in this study by acting as internal synthetic controls that mimic fusion gene events. Their inclusion allowed for quantitative benchmarking of the assay’s performance and supported robust validation of fusion detection in clinical RNA samples. This underscores the value of Sequins in enhancing diagnostic confidence and reproducibility in RNA-seq—based fusion detection assays.

# Products

## Whole Genome Sequencing (WGS) Core Control Set

The Sequins WGS Core Control Set is a comprehensive, easy-to-use, pre-configured control set with built-in redundancy that ensures the integrity of results. Designed against the hg38 reference genome and compatible with hg37, it includes Sequins covering multiple regions in a single tube to be used with PCR-free library preparation methods for research use. Variant classes include difficult genetic variants; germline variants at simple repeats; homopolymers; structural variants; common genetic variants; microsatellites; and mitochondrial DNA.

## Metagenomics Core Control Set

The Sequins Metagenomics Core Control Set is a comprehensive, easy-to-use, pre-configured control set with built-in redundancy that ensures integrity of results. It contains synthetic sequences that collectively represent 52 microbial species at varying abundances to emulate the complexity found in a natural microbial community. Features include: quantitative ladders; wide representation of domains including bacteria, archaea and eukaryotes; GC% content range (27% to 71%); single tube blend; demonstrated on various samples using Illumina and Oxford Nanopore technologies.

## Circulating tumor DNA (ctDNA) Evaluation Set *(in development)*

The Sequins ctDNA Evaluation Set targets nine clinically important cancer gene variants at precise variant allele frequencies and is applicable to a range of cancers. Sequins are pre-mixed into a single tube and can be spiked-in to each sample prior to library preparation and sequencing.

## On-Demand

Sequins controls can be designed to any genomic region in any organism with a reference genome. Our On-Demand program and design capabilities enable customization for your assays to ensure implementation of fully optimized control sets.

# Conclusion

Sequins are a gold standard for NGS quality control, providing the accuracy, reliability, and reproducibility you need to unlock the full potential of your sequencing data.

# Contact Us

Visit our website (www.sequins.bio) or contact us at support@sequins.bio to learn more about how Sequins can transform your NGS workflows.

## References


1. 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.
2. Davis, B.C, Vikesland P.J., Pruden A. (2025). Evaluating Quantitative Metagenomics for Environmental Monitoring of Antibiotic Resistance and Establishing Detection Limits. Environmental Science & Technology 59(12), 6192-6202.
3. Deveson, I., Chen, W., Wong, T. et al. (2016). Representing genetic variation with synthetic DNA standards. Nat Methods 13, 784—791.
4. Deveson, I.W., Madala, B.S., Blackburn, J. et al. (2019). Chiral DNA sequences as commutable controls for clinical genomics. Nat Commun 10, 1342.
5. Deveson, I.W., Gong, B., Lai, K. et al. (2021). Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology. Nat Biotechnol 39, 1115—1128.
6. Hardwick, S., Chen, W., Wong, T. et al. (2016). Spliced synthetic genes as internal controls in RNA sequencing experiments. Nat Methods 13, 792—798.
7. Hardwick, S., Deveson, I. & Mercer, T. (2017). Reference standards for next-generation sequencing. Nat Rev Genet 18, 473—484.
8. Hardwick, S.A., Chen, W.Y., Wong, T. et al. (2018). Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. Nat Commun 9, 3096.
9. Heyer, E.E., Deveson, I.W., Wooi, D. et al. (2019). Diagnosis of fusion genes using targeted RNA sequencing. Nat Commun 10, 1388.
10. Langenfeld, K., Hegarty, B. Vidaurri, S. et al. (2025). Development of a quantitative metagenomic approach to establish quantitative limits and its application to viruses. Nucleic Acids Research, Volume 53, Issue 5.
11. Reis, A.L.M., Deveson, I.W., Wong, T. et al. (2020). A universal and independent synthetic DNA ladder for the quantitative measurement of genomic features. Nat Commun 11, 3609.
12. Reis, A.L.M., Deveson, I.W., Madala, B.S. et al. (2022) Using synthetic chromosome controls to evaluate the sequencing of difficult regions within the human genome. Genome Biol 23, 19.
13. Vyshenska D., Sampara P., Singh K., et al. (2023). A standardized quantitative analysis strategy for stable isotope probing metagenomics. mSystems 8:e01280-22.