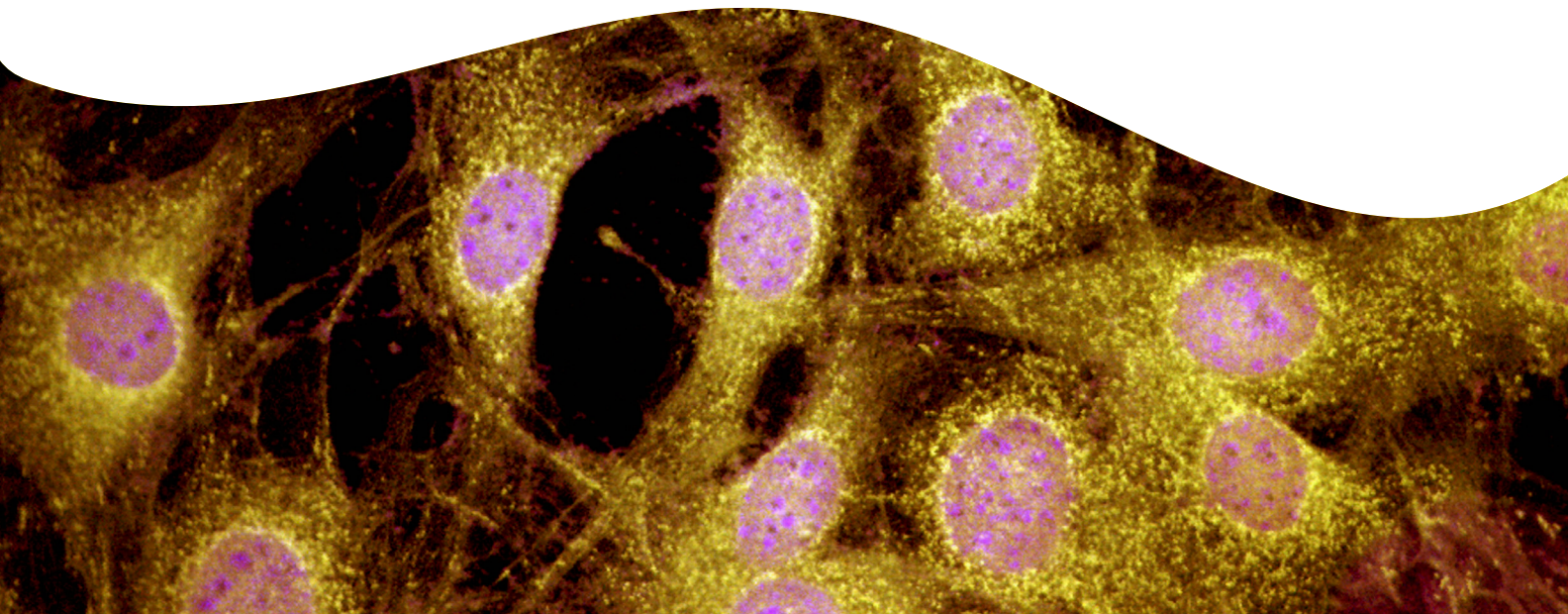


RNA-Seq characterization of model cell lines for drug discovery.

Introduction

Modern drug discovery increasingly relies on a systems-level understanding of disease biology, and RNA sequencing (RNA-seq) has become central to this shift. DNA sequencing provides a static map of potential cellular function, while RNA-seq delivers a dynamic and quantitative readout of gene activity, alternative splicing and pathway engagement, which are directly linked to phenotype, drug response and toxicity¹⁻³. Large-scale resources such as the Cancer Cell Line Encyclopedia and related pharmacogenomic efforts highlight how deep transcriptomic and genomic characterization of preclinical models can uncover actionable biomarkers and lineage-specific dependencies^{4,5}.

From a preclinical services perspective, comprehensive RNA-seq characterization enables more rational model selection by aligning projects with the most biologically relevant cell lines. For antibody-drug conjugates (ADCs), this is critical because the therapeutic index (the balance between effective tumor killing and toxicity) depends on selective, high antigen expression. RNA-seq reveals both target abundance and off-tumor risks, guiding the choice of appropriate models for ADC evaluation⁶. In addition, integrating RNA-seq data with whole-genome or exome data enhances interpretation of cell panel screening results. Transcriptomic signatures can distinguish sensitive from resistant models, support patient-stratification hypotheses, and reveal combination strategies^{3,7}.



RNA-seq is also a powerful tool in functional genomics and mechanism-of-action studies. When coupled with CRISPR-based screening or chemical treatment, transcriptomic readouts map each perturbation to its downstream effects, clarifying pathway relationships, synthetic-lethal interactions, resistance mechanisms and off-target effects. Likewise, the comparison of expression before and after compound treatment continues to be widely used to define drug-responsive signatures, evaluate pathway modulation, and differentiate on-target effects from off-target transcriptional perturbations⁷⁻⁹.

In this application note, we present RNA-seq-based transcriptome profiling across a small panel of drug discovery-relevant cell lines, illustrating how these data support model selection, mechanistic interpretation, and functional genomics workflows. This data is just a subset of a much larger initiative led by Revvity's OncoSignature™ Cell Panel Services team, where ~300 cancer-related cell-lines are being thoroughly characterized with RNA-seq. Transcriptomic baselines models can be integrated with cell line panels and functional genomics to enable more confident interpretation of drug screening results.

Methods

Five tumor-derived cell lines commonly used as preclinical model in drug discovery workflow were selected (Table 1). Cell lines were revived and cultured in vendor-recommended media. Pellets of one million cells were collected between passage 3 to 15. RNA was extracted using the RNeasy Mini Kit (Qiagen) following manufacturer's protocol and quantified on the TapeStation 4200 (Agilent). One microgram (1µg) of input

RNA was used for RNA library prep. Samples were enriched for mRNA using NEXTFLEX™ Poly(A) Beads 2.0 (Revvity) and further processed using NEXTFLEX™ Rapid Directional RNA-seq Kit 2.0 (Revvity) following the standard protocol. Two biological replicates of each cell line were used. Samples were sequenced by Revvity Omics on AVITI™ sequencer (Element Biosciences), at the read depth of ~50 million 2 x 150 bp paired-end reads per sample.

RNA-seq quality control and quantification analysis were performed using the Sequana Snakemake¹⁰ pipeline (v0.19.2), with the sequana-rnaseq module (v0.20.0) running on Snakemake (v7.32.4). Raw sequencing data underwent quality assessment with FastQC (v0.19.2), followed by preprocessing with fastp (v0.23.2) for adapter trimming and quality filtering. Clean reads were then aligned to the human reference genome (GRCh38/hg38) using Bowtie2 (v2.4.4) in paired-end mode. Post-alignment processing included tagging aligned reads with sample identification information and marking potential PCR duplicates using Picard tools (v2.27.5). Gene-level quantification was performed with featureCounts (Subread v2.0.3) using GENCODE v44 annotations, with parameters optimized for forward-stranded libraries. Comprehensive quality metrics were generated using the RSeQC (v5.0.4) suite and aggregated with MultiQC (v1.27) for final reporting.

Custom Python scripts were used post- Sequana Snakemake pipeline. Gene detection coverage was benchmarked against DepMap Public 25Q3 reference data. Data quality and biological patterns were assessed using heatmaps, principal component analysis, and gene body coverage plots to evaluate transcriptional profiles, sample clustering/batch effects, and read distribution uniformity, respectively.

Table 1: Cell lines analyzed in this study and their major cancer-relevant mutations, which provide a molecular context for interpreting transcriptomic profiles and supporting model selection. Mutational status is based on Horizon/DepMap/COSMIC annotations.

Cell line	Tumor type	Mutations
A375	Skin	BRAF V600E
A549	Lung	KRAS G12S; KEAP1 G333C
Farage	B cell lymphoma	TP53 R248W
HCT116	Colon	KRAS G13D; CTNNB1 S45del; PIK3CA H1074R
HT29	Colon	APC E853; TP53 R273H; PIK3CA P449T; BRAF V600E

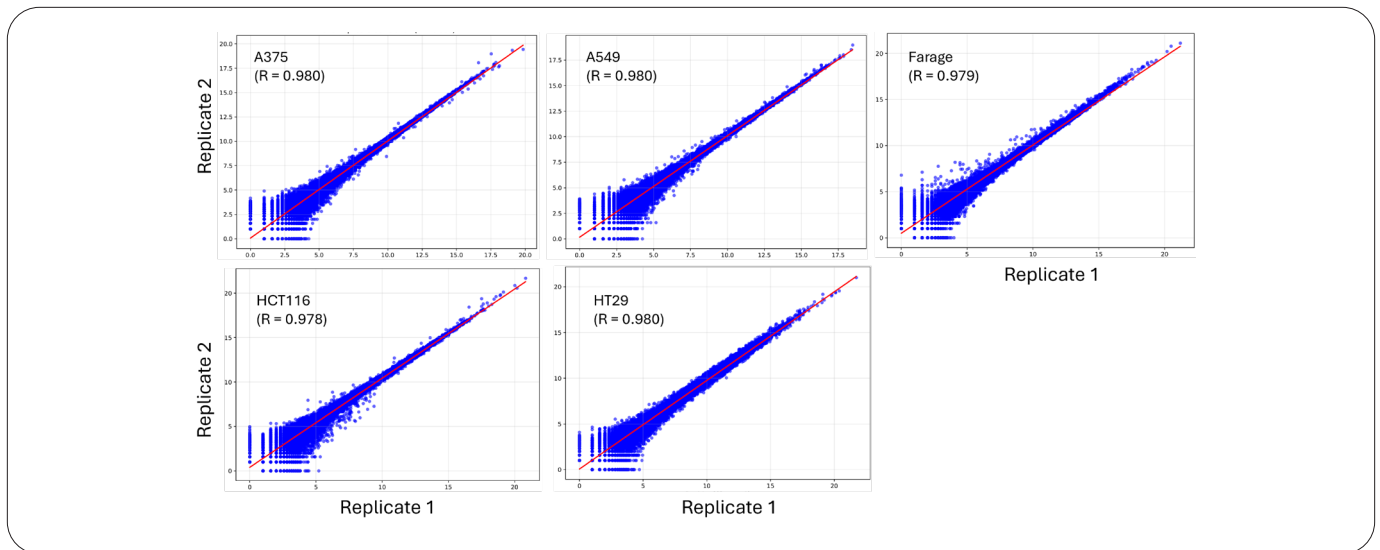


Figure 1: Scatter plots show high reproducibility ($r \geq 0.978$) between biological replicates across all cell lines. Each axis represents $\log_2(\text{counts} + 1)$ for a replicate.

Results

Sequencing quality and reproducibility

High sequencing quality was observed across all samples, with mean Q20 and Q30 scores of 97.9% and 94.7%, respectively. Filter pass rates exceeded 99%, and ribosomal RNA contamination was 1.19% on average. Duplication rates ranged from 3-11%, well below concerning thresholds.

Correlation analysis between biological replicates demonstrated high reproducibility across all samples (Figure 1). Volcano plots comparing replicates (Figure 2) confirmed minimal differential expression, further validating experimental reproducibility.

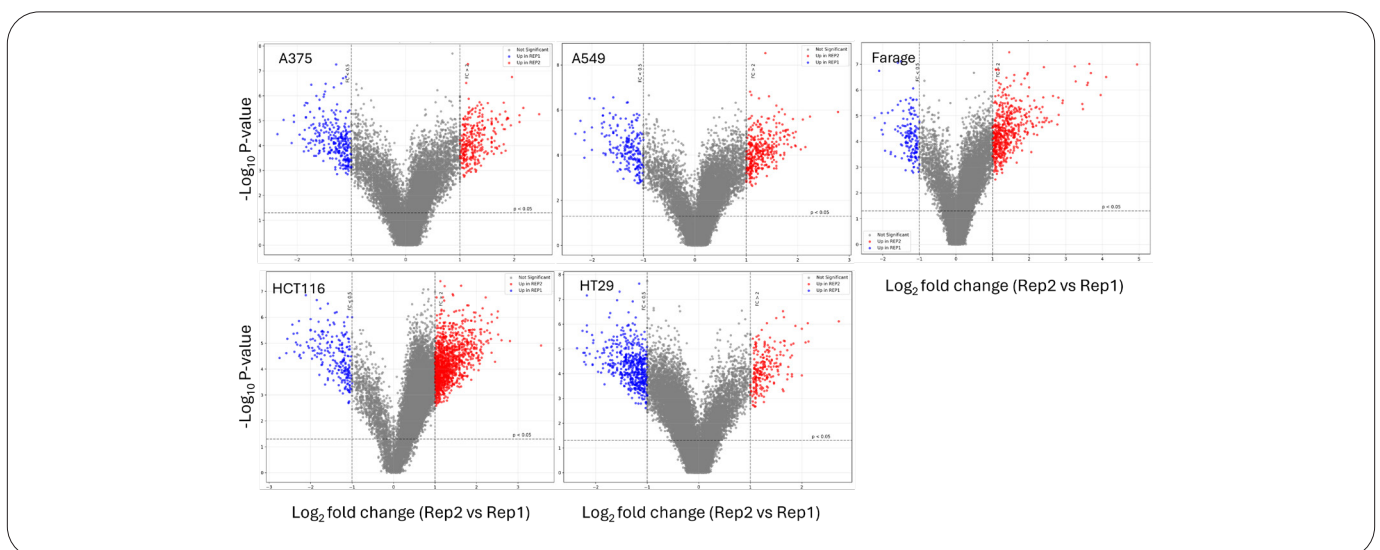


Figure 2: Volcano plots highlight genes with both large fold changes and strong statistical significance. Most genes do not show significant differential expression between biological replicates, for the cell lines investigated.

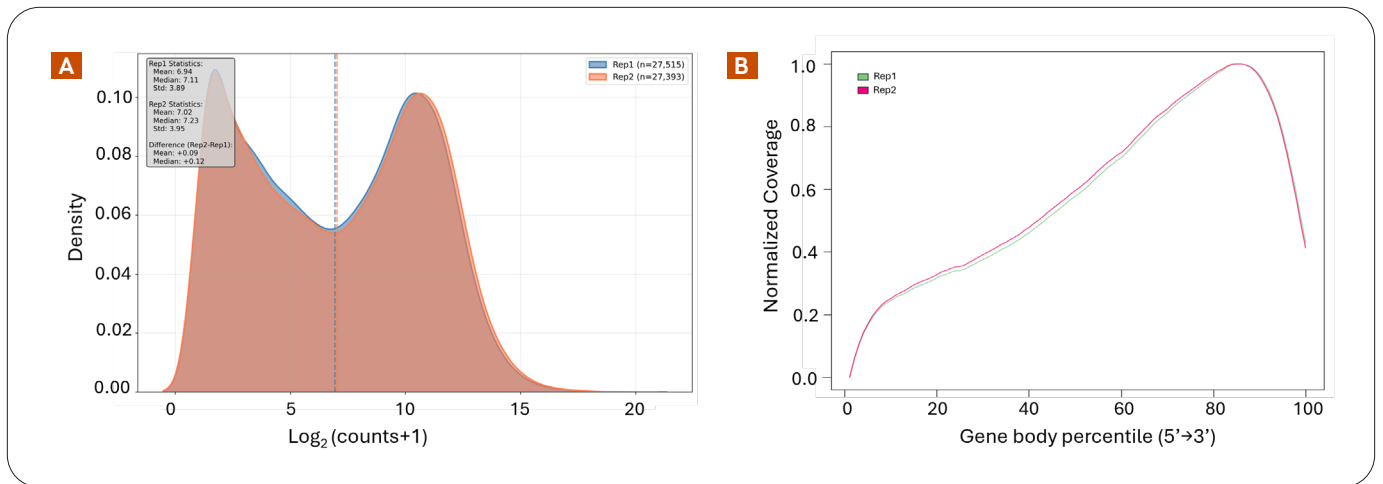


Figure 3: **A)** Density plot showing the distribution of gene expression values across all detected genes in cell line A375. The plot summarizes how many genes fall within low, medium, or high expression ranges, providing a global view of transcript abundance **B)** Normalized coverage profile across the gene body, from 5' to 3' ends, aggregated across all expressed transcripts. The decrease in coverage close to 5' indicates potential bias in fragmentation or reverse transcription.

Gene expression landscape

21,000-25,000 genes were detected per sample, which captures most annotated protein-coding genes (Figure 3A). Coverage averaged 1,417-2,546 reads per gene across samples, supporting robust differential expression analyses. Normalized coverage across transcript length was inspected. The full transcript was covered although with lower depth at 5' end (Figure 3B).

Comparative transcriptomics

A heatmap was generated to compare expression profiles across cell-lines and their biological replicates. For this application note, we focus on the 7 mutated genes most relevant to these models. Remarkably, expression of just these genes is sufficient to clearly discriminate each cell line (Figure 4). Such expression-based separation is relevant for model selection, as it highlights lineage- and mutation-specific transcriptional states that determine how each cell line may respond to pathway-targeted compounds.

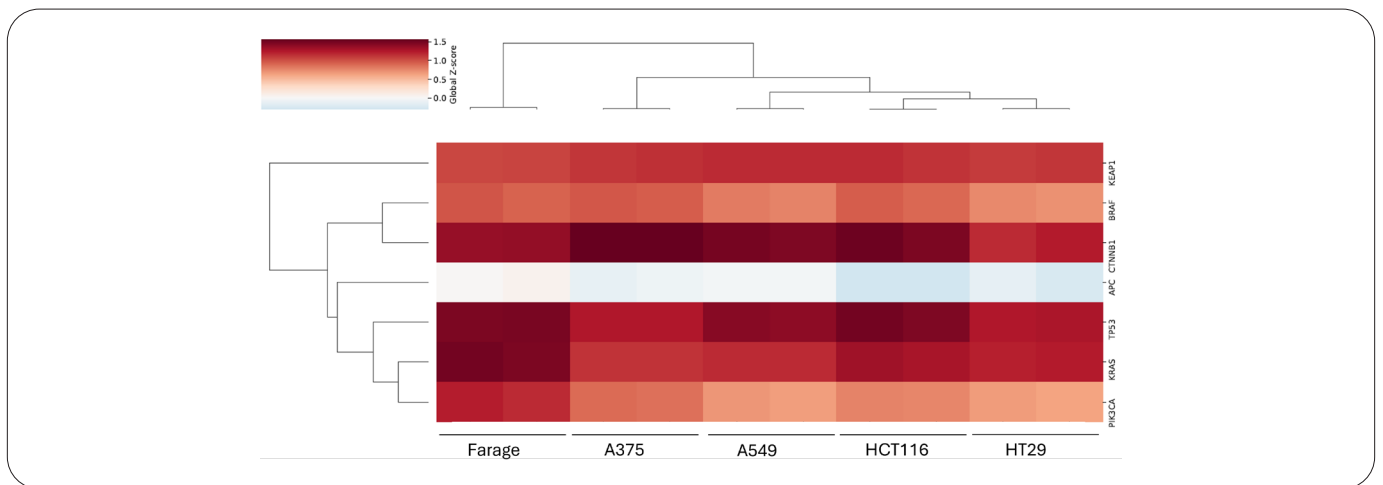


Figure 4: Multi-sample gene expression heatmap across the five cell lines and biological replicates, focused on the 7 genes containing known mutations. Red colors indicate genes with higher expression compared to the average of all expressed genes, while blue colors indicate genes with lower expression than the average. Hierarchical clustering of both genes and samples highlights lineage-specific signatures and confirms tight grouping of replicates for each model.

Principal Component Analysis (PCA) was performed on normalized RNA-seq expression values to assess global transcriptional differences between samples. The first two principal components (PC1 and PC2) capture the major sources of variation in the dataset. Biological replicates from each cell line clustered tightly together, indicating high intra-group reproducibility, while the five cell lines form distinct, well-separated clusters, reflecting strong transcriptomic differences between cell types.

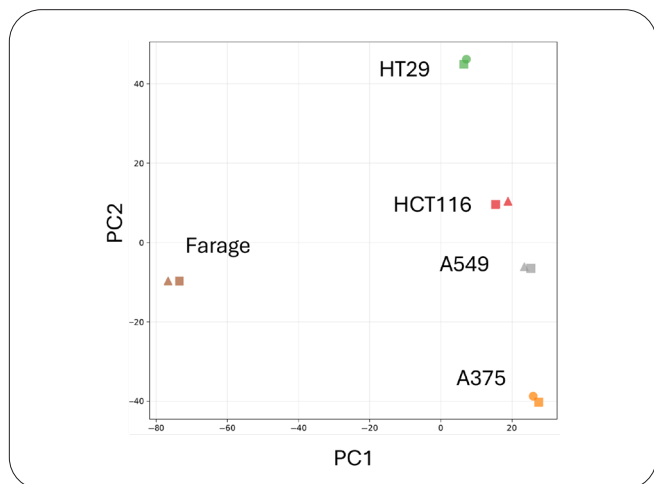


Figure 5: PCA shows tight clustering of biological replicates and clear separation of the five cell lines along PC1 and PC2, indicating high reproducibility and strong cell line-specific transcriptomic differences.

Discussion

Bulk RNA-seq performed with the NEXTFLEX Rapid Directional RNA-seq Kit 2.0 provides a reliable foundation for characterizing preclinical cell-line models, with high sequencing quality, low duplication and rRNA content, and strong concordance between replicates. The resulting gene-detection depth enables robust quantitative analyses and straightforward comparison with large public datasets such as DepMap or CCLE.

Global transcriptomic analyses clearly separate the five cell lines and tightly cluster replicates, underscoring lineage- and mutation-driven differences that are essential for rational model selection in drug discovery. Even simple signatures, such as MAPK-pathway gene expression in *BRAF*- or *KRAS*-mutant models, can guide selection of lines most relevant for targeted inhibitor testing. Likewise, for ADCs, RNA-seq expression of tumor antigens and related markers supports rapid triage of models into high- or low-relevance categories.

Importantly, the models and transcriptomic baselines presented here are part of Revvity's OncoSignature™ Cell Panel preclinical services and can be integrated with pharmacology assays and CRISPR functional genomics screens to create a unified, data-driven framework for model selection and mechanism-of-action interpretation.

Learn more about Revvity's solutions:

[OncoSignature™ Cell Panel Services team](#)

[Functional Genomics Screening Services](#)

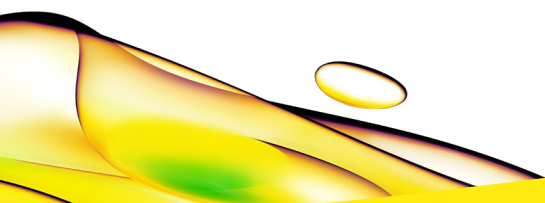
[Revvity Omics Services](#)

[NEXTFLEX™ Poly\(A\) Beads 2.0](#)

[NEXTFLEX™ Rapid Directional RNA-seq Kit 2.0](#)

References

1. Wang, Z., *et al.* (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 10(1):57-63. doi: 10.1038/nrg2484.
2. Sá, A.C.C., *et al.* (2018). Whole Transcriptome Profiling: An RNA-Seq Primer and Implications for Pharmacogenomics Research. *Clin Transl Sci.* 11(2):153-161. doi: 10.1111/cts.12511.
3. Deshpande, D., *et al.* (2023). RNA-seq data science: From raw data to effective interpretation. *Front Genet.* 14:997383. doi: 10.3389/fgene.2023.997383.
4. Barretina, J., *et al.* (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607. doi:10.1038/nature11003.
5. Carli, F., *et al.* (2025). Learning and actioning general principles of cancer cell drug sensitivity. *Nat Commun* 16, 1654. doi: 10.1038/s41467-025-56827-5.
6. Beck, A., *et al.* (2017). Strategies and challenges for the next generation of antibody–drug conjugates. *Nat Rev Drug Discov* 16, 315–337. doi:10.1038/nrd.2016.268.
7. Qi, X., *et al.* (2024). Predicting transcriptional responses to novel chemical perturbations using deep generative model for drug discovery. *Nat Commun* 15, 9256. doi:10.1038/s41467-024-53457-1.
8. Szalai, B., Veres, D.V. (2023). Application of perturbation gene expression profiles in drug discovery-From mechanism of action to quantitative modelling. *Front Syst Biol.* 3:1126044. doi: 10.3389/fsysb.2023.1126044.
9. Van de Sande, B., *et al.* (2023). Applications of single-cell RNA sequencing in drug discovery and development. *Nat Rev Drug Discov.* 22(6):496-520. doi: 10.1038/s41573-023-00688-4.
10. Cokelaer, T., *et al.* (2017), 'Sequana': a Set of Snakemake NGS pipelines, *Journal of Open Source Software*, 2(16), 352, doi:10.21105/joss.00352.



revvity