**Editorial Board of *Nature Protocols***

Zurich, 5 August 2025

**Re: Presubmission enquiry to *Nature Protocols***

Dear Dr. Clyne, dear Editors,

We are writing to inquire about the suitability of our computational protocol for publication in *Nature Protocols*. This protocol represents the computational counterpart to our previously published wet lab methodology for affinity-based selections of DNA-encoded libraries (DELs),1 thereby providing the research community with a complete end-to-end solution for DEL-based drug discovery.

DELs are collections of organic compounds, individually attached to oligonucleotides, which serve as amplifiable identification barcodes.2-4 The aim of DEL technology is to identify small molecule ligands to target proteins of choice.5,6DELs are increasingly used for drug discovery, both by the pharmaceutical industry and - due to the inexpensive setup - by academic groups worldwide, which is reflected by the increasing number of research papers published in the field.

As highlighted in the recent *Nature Reviews Methods Primers* on DEL technology,6 the computational analysis of DEL data presents significant challenges, with existing solutions being largely proprietary, fragmented, or requiring extensive bioinformatics expertise. Our proposed protocol offers an end-to-end framework from raw sequencing to hit identification that scales to libraries with hundreds of millions of compounds, and handles both single and dual display architectures.

The computational bottleneck remains a significant barrier to broader adoption, particularly in academic settings. Our protocol addresses this critical unmet need by providing the first comprehensive, user-friendly computational framework specifically designed for large-scale DEL analysis.

The protocol has been extensively validated across multiple library architectures and screening campaigns, demonstrating robust performance with libraries containing up to hundreds of millions of compounds. We believe that it will be of broad interest to the chemical biology, drug discovery, and computational chemistry communities.

We would welcome the opportunity to submit a full protocol if you consider this topic suitable for  
*Nature Protocols*:

**Proposed Title:  
DELT-Hit: An end-to-end computational framework for DNA-encoded chemical library analysis**

**Authors:**Adriano Martinelli, Alice Lessing, Gary Hoppeler, Andreas Gloger, Jörg Scheuermann

**Abstract**

DNA-encoded chemical libraries (DELs) have revolutionized drug discovery by enabling the simultaneous screening of millions to billions of small molecules through DNA-tag identification via high-throughput sequencing. As outlined in the comprehensive Nature Reviews Methods Primers on this technology (Satz et al., 2022), DELs are now employed by numerous pharmaceutical companies and academic laboratories worldwide. However, the computational analysis of DEL screening data remains a critical bottleneck, requiring sophisticated integration of genomics, cheminformatics, and statistical analysis workflows that are currently accessible only through proprietary or highly specialized software solutions.

This protocol presents DELT-Hit (DNA-Encoded Library Technology Hit), a comprehensive open-source computational framework that makes DEL data analysis accessible through an intuitive command-line interface to both computational and experimental researchers. DELT-Hit is specifically designed to handle the scale and complexity of modern industrial DEL campaigns, supporting libraries containing hundreds of millions of compounds while maintaining computational efficiency and user accessibility.

DELT-Hit offers a complete pipeline that converts raw FASTQ reads into machine learning ready chemical information through five connected modules: (1) adaptive sequence demultiplexing using optimized RNA-seq algorithms with DEL-specific error correction and flexible barcode handling, (2) automated chemical structure reconstruction from building block libraries using reaction SMARTS templates with support for both single and dual display architectures, (3) comprehensive molecular property calculation and descriptor generation using established cheminformatics libraries, (4) statistical analysis and hit ranking with multiple normalization strategies adapted from proven RNA-seq methodologies, and (5) integrated quality control and visualization tools specifically designed for DEL data interpretation.

The modular architecture allows researchers to customize workflows while maintaining reproducibility through configuration files and standardized output formats. We demonstrate the protocol's effectiveness using representative single and dual display DEL screening datasets, showcasing the complete analysis pipeline from raw sequencing reads to ranked lists of chemical hits with computed chemical properties and representations for downstream machine learning tasks. The entire analysis, including quality control and visualization, can be completed within 4-6 hours on standard computational hardware for typical datasets, making it accessible to laboratories without specialized computing infrastructure.

DELT-Hit addresses the critical computational gap identified in the DEL field and provides the standardization necessary for reproducible analysis across research groups. The protocol is accompanied by comprehensive documentation, tutorial datasets with both single and dual display examples, ensuring broad adoption and consistent implementation across the growing DEL community.

**Key Points**

* **Industrial-scale capabilities**: DELT-Hit is designed to handle the computational demands of modern pharmaceutical DEL campaigns, efficiently processing libraries containing hundreds of millions of compounds while maintaining user-friendly operation
* **Comprehensive dual architecture support**: The framework provides native support for both single and dual display DEL architectures, addressing the full spectrum of current library designs used in industry and academia
* **Validated algorithms**: Integrates proven bioinformatics tools (Cutadapt for sequence processing, edgeR for statistical analysis) with specialized DEL-specific optimizations, error handling, and quality control metrics developed through extensive validation studies
* **Flexible and robust design**: Modular architecture accommodates diverse library formats, custom reaction templates, and building block definitions
* **Research-grade quality assurance**: Built-in quality control metrics, automated validation checks, and standardized reporting ensure reliable results and facilitate systematic troubleshooting across different experimental conditions
* **Machine learning ecosystem integration**: Generates standardized, analysis-ready datasets fully compatible with downstream machine learning workflows for advanced hit prediction, structure-activity relationship analysis, and virtual screening applications

**Technical Overview**

DELT-Hit is implemented as a Python package organized into five modules:

**Core Analysis Modules:**

* init: Project initialization and configuration management
* demultiplex: Sequence processing and demultiplexing with adaptive error correction
* compute: Chemical structure reconstruction and molecular property calculation
* qc: Comprehensive quality control and validation metrics
* viz: Integrated data visualization and reporting

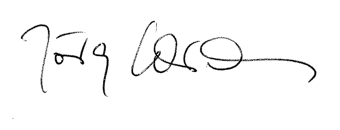
**Input Requirements:**

The framework processes three primary input types: (1) library definition files containing building block structures, reaction SMARTS, and constant sequences; (2) experimental metadata specifying selection conditions and sample identifiers; and (3) raw FASTQ files from high-throughput sequencing platforms.

We believe that this protocol may be of broad interest, and we would be happy to publish it in *Nature Protocols* if you were interested.

Should you require any additional information, please do not hesitate to contact me.

I thank you for your kind consideration,

Yours sincerely,

1 Decurtins, W. *et al.* Automated screening for small organic ligands using DNA-encoded chemical libraries. *Nat Protoc* **11**, 764-780 (2016). <https://doi.org/10.1038/nprot.2016.039>

2 Franzini, R. M., Neri, D. & Scheuermann, J. DNA-encoded chemical libraries: advancing beyond conventional small-molecule libraries. *Acc Chem Res* **47**, 1247-1255 (2014). <https://doi.org/10.1021/ar400284t>

3 Neri, D. & Lerner, R. A. DNA-Encoded Chemical Libraries: A Selection System Based on Endowing Organic Compounds with Amplifiable Information. *Annual Review of Biochemistry* **87**, 479-502 (2018). <https://doi.org/10.1146/annurev-biochem-062917-012550>

4 Peterson, A. A. & Liu, D. R. Small-molecule discovery through DNA-encoded libraries. *Nat Rev Drug Discov*  (2023). <https://doi.org/10.1038/s41573-023-00713-6>

5 Huang, Y., Li, Y. & Li, X. Strategies for developing DNA-encoded libraries beyond binding assays. *Nature Chemistry* **14**, 129-140 (2022). <https://doi.org/10.1038/s41557-021-00877-x>

6 Satz, A. L. *et al.* DNA-encoded chemical libraries. *Nature Reviews Methods Primers* **2** (2022). <https://doi.org/10.1038/s43586-021-00084-5>