CDD Vault - A Collaborative Data Management Platform for DNA & RNA Drug Discovery

Introduction

Nucleic Acid (NA) Therapeutics are DNA and RNA molecules designed to act as therapeutic entities. They often target disease pathways that are traditionally considered undruggable and are promising vehicles for delivering personalized medicine.

Therapeutic DNA molecules largely constitute the growing number of approved cell and gene therapy drugs. The different types of cell and gene therapy molecules can be classified as circularized plasmids, viral vector based or CRISPR-Cas9 mediated molecules (reviewed by Damase et al 2021). The first DNA vaccine against COVID-19 was recently approved by drug regulators in India while around a dozen others are under clinical trials globally (Mallapaty 2021).

RNA therapies are regarded as safer due to their inability to alter the host genome. As a result, mRNA vaccines have been widely approved and delivered to global populations during the COVID-19 pandemic. RNA therapeutics may be generally classified into mRNA based cell therapies and mRNA vaccines, Antisense Oligomers (ASO), RNA interference (RNAi) strategies such as small interfering RNA (siRNA) and microRNA (miRNA) and RNA aptamers (reviewed by Damase et al 2021, Yu et al 2020).

Data Challenges for Developing NA Therapeutics

NA Therapeutics researchers have specific informatics requirements for their workflows. These include:

Oligo design and synthesis: Bioinformatics tools for designing oligomers and chemically modified RNA or DNA such as aptamers.

Visualization Analysis: Visualization tools for oligomer sequences, plasmids, viral vectors, sequence alignment and searching; visualization tools for cloning data such as gel imaging and qPCR plots, Western Blots, FACS data.

Analysis software: Various self built and/or enterprise software pipelines to import, normalise, analyse and interpret raw assay data.

Central database: Creating and maintaining a reliable ‘single source of truth’ in the form of QCed centralized repositories to store biological entities such as designed and synthesized constructs, ordered oligomers, donor and acceptor plasmids, delivery vectors, along with their associated characterisation data. Ability to record and visualize connections between the different entities or their modified versions and batches.

Collaboration: Ability to share data selectively with collaborators in a secure way without creating and holding multiple versions being sent back and forth via email.

Digitalization and secure access: An Electronic Lab Notebook (ELN) with inter-connectivity between all data, accessed securely by selected members in real time as and when data is uploaded.

The major challenge facing scientists working on NA therapeutics is the difficulty with cross-talk between tools. Scientists often use a variety of tools to fulfil their requirements, but struggle with establishing connections between the different tools. Cloning sequences for instance, may be designed and stored in the cloning software. Whereas the associated colony PCR and Western blots are stored elsewhere without any digital connection between the two.

Researchers also hold multiple databases of spreadsheets containing ordered oligo, donor and recipient plasmids, cell lines and intermediate bacterial clones showing properties of each of these. However, they find that the digital connections that would allow them to search and mine all constructs designed using a particular plasmid or particular oligo or a particular cell line for instance, are often missing.

In the absence of a central database, individual scientists often end up holding versions of self-generated constructs that do not exist outside their documentation. Moreover there is no way to connect individual databases and the ELN, creating additional data silos.

CDD Vault is a modern informatics platform that can resolve many of the challenges faced in the development of NA therapeutics.

Getting Started

1. **Register biological entities and calculated properties**

    - CDD Vault helps NA researchers register their nucleotide sequences of choice.
    - Researchers can register NA entities such as ASO, candidate mRNA or DNA vaccine molecules, siRNA, miRNA, plasmid or viral gene therapy constructs, commercial vectors, cell therapy lines, intermediate bacterial clones and so on.
    - Modified oligos can be captured in MarvinJS and as a unique sequence with user defined syntax.
    - Characteristics such as base pair length (BP), %GC/AT, Molecular Weight (MW) and composition can be automatically calculated in CDD Vault.

2. **Create batches and associate contextual information**
    - Each registered biological entity is assigned a unique identifier by CDD Vault which allows researchers to track batches explicitly via an entity-batch identifier.
    - Biologists can store metadata assigned to an entity or a specific batch. Users can associate data in the context of specified batches so it's easy to track batch-to-batch variations.
    - Researchers can also store and visualise all entity characterization in a single location. For example, when a biologist registers modified plasmids they can create links to its associated data characteristics such as nomenclature of the plasmid, modification method such as CRISPR Cas9, Viral vector used such as Lentiviral (LV), Adenovirus (AV) or Adeno-Associated Virus (AAV) and so on (See fig.1).

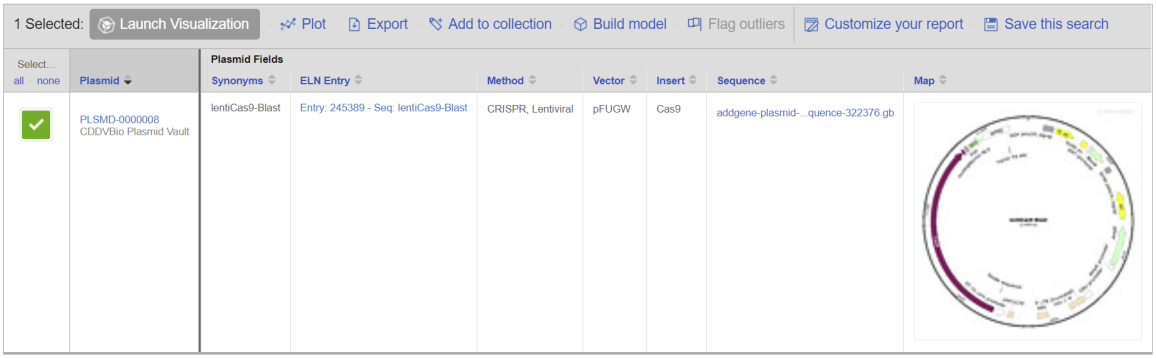

Fig 1: Screenshot of a registered plasmid used for generation of CRISPR constructs for gene therapy. Customized to show the modified plasmid (the registered entity), synonym, ELN entry associated with the entity, modification method as CRISPR Cas9, Lentiviral vector used, promoter, links to the sequence (Addgene) and the plasmid map.

3. Store and visualize assays

- CDD Vault's protocol feature is designed to allow users to associate each piece of assay data to its related registered entity. For example; entity characterization data such as on-target efficiency analysis using qPCR, FACS or Western Blot can be stored and visualized for its registered ASO, plasmid, viral vector construct or cell line depending on user application (See fig. 2).
- Data related to RNA screening plates, dose response assays, cell-based assays, ADME-toxicology studies, *in vivo* biology assays, PK/PD assays, etc. can be stored, visualized and analyzed.

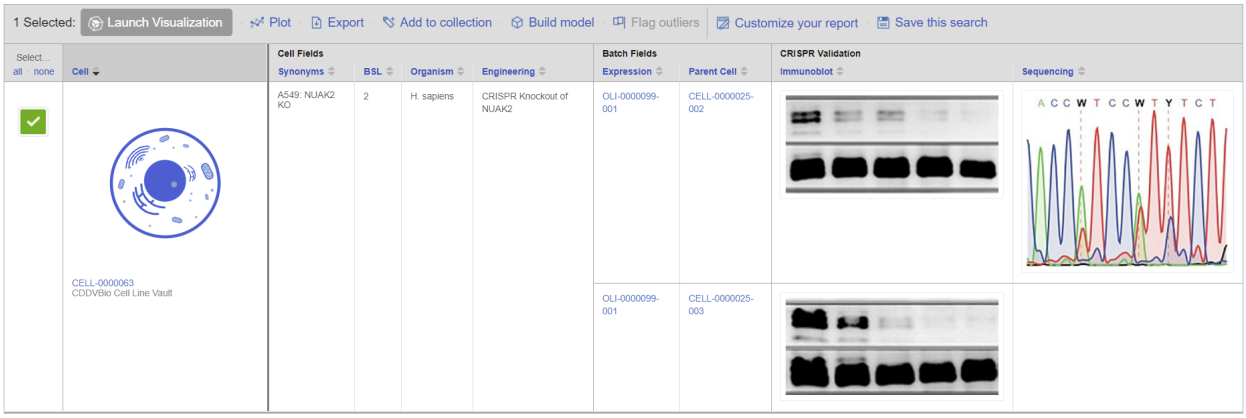

Fig 2: Screenshot of an engineered CRISPR Cas9 cell line. Customized to show the engineered cell line (the registered entity), synonym/nomenclature, biosafety level, organism, modification method such as CRISPR Cas9, oligo expressed for cell engineering, parent cell line used as well as validation assay data showing an immunoblot and a sequencing plot.

4. **Document experimental details in the ELN**
   - CDD's ELN can be used by researchers to store standardized SOPs and record all experimental details across departments (See fig. 3).
   - Full integration with Microsoft Office 365 allows scientists to attach, preview and edit MS office files within the ELN.
   - Image files can be displayed directly within the ELN entry in various sizes.
   - Any file format, including proprietary files which require their own application, can be stored in the ELN.
   - Users can link multiple entities in the registration system to an ELN entry. For example: the mRNA construct, its respective cell line clone and protein target can all be associated with a link in a single ELN entry, allowing easy reference of materials used in the experiment.

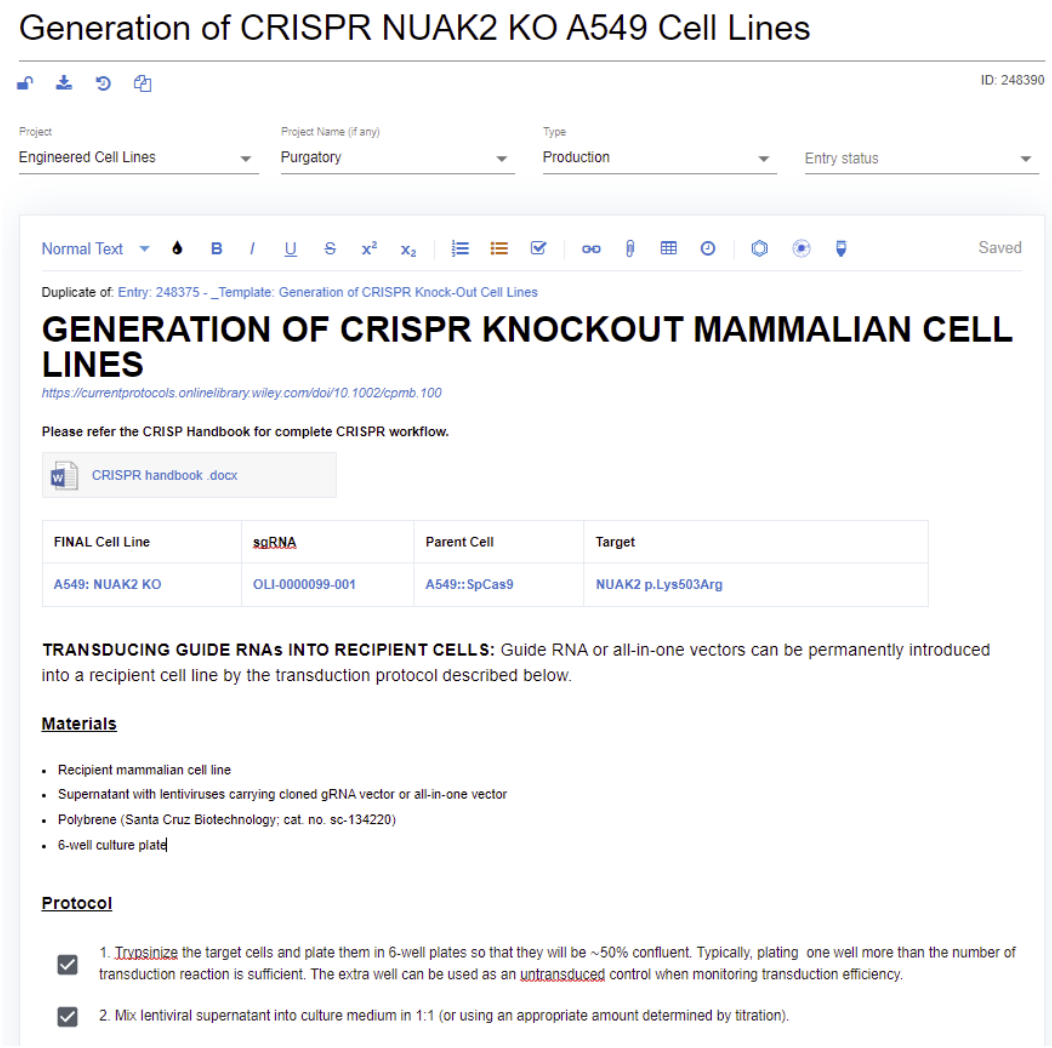

Fig 3: Partial screenshot of an ELN entry showing an experiment for the generation of a CRISPR NUAK2 knockout mammalian cell line.

5. **Mine data**
   - Researchers can search their data by sequence and easily locate records by sub-sequences. For example, searching for a given ASO will find all associated characterization and assay data related to that entity. Likewise, one can mine all constructs generated using a chosen guide RNA or a particular cell line.
   - One can also search for keywords in ELN entry bodies and attachments.

6. **Track location and quantities of entities and reagents**
   - Researchers can capture the location and quantities of their entities and other reagents including features such as date, vendor and batch using the inventory module (See fig 4).
   - Users can capture and track containers/ vials/ dewars/ flasks.
   - Scientists can use barcodes to manage their inventory. One can export identifiers and key metrics to their barcode template and print labels. Users can also locate samples or pull up a list using a barcode scanner.
   - A debit history provides a full record of the date, person, amount and/or location of each update.
   - Users can see inventory data alongside other batch characteristics and assay data associated with the batch.

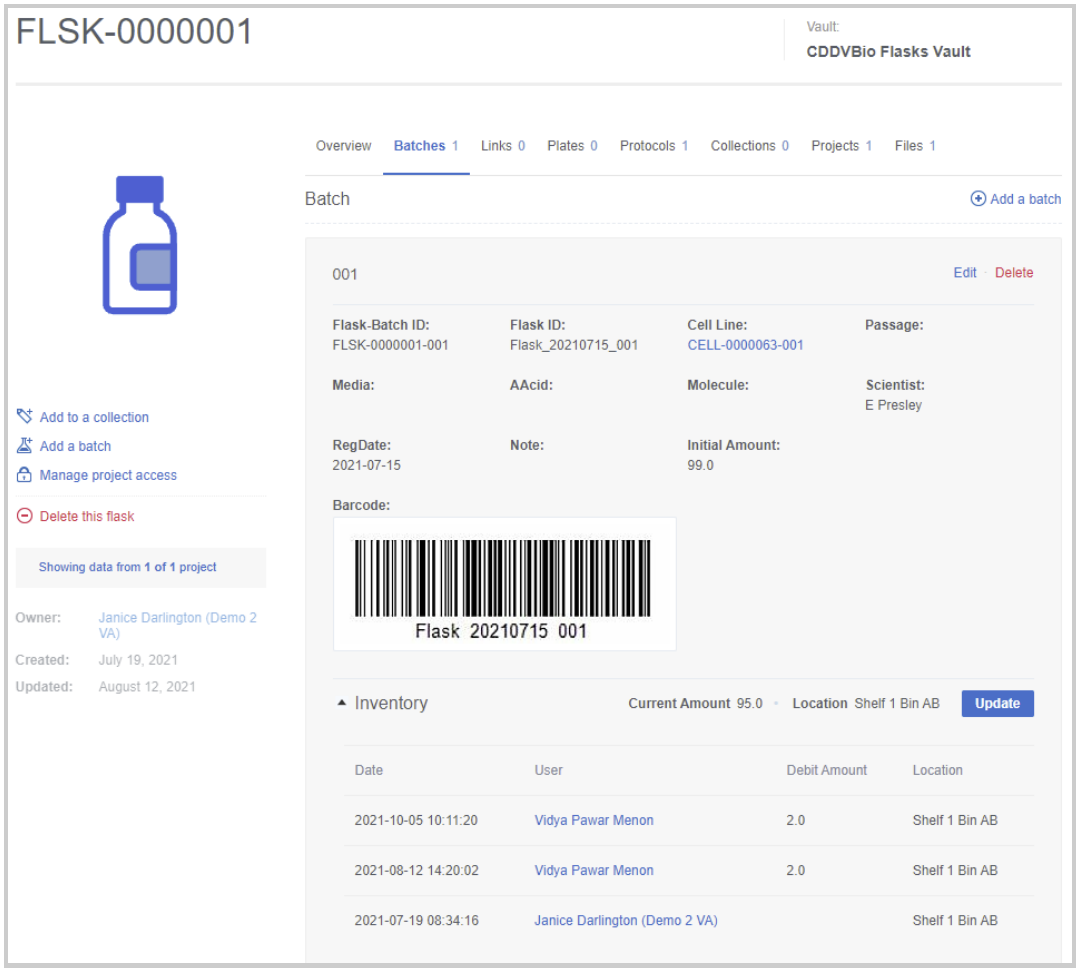

Fig 4: Snapshot of batch level inventory fields from a Vault customized to register cell line culture flasks.

7. **Link to other tools**
    - CDD Vault's RESTful Application Programming Interface (API) can be used to link to external databases and analysis tools.
    - Scientists can easily export data from CDD Vault in Excel, CSV or SDF file formats.
    - Likewise, users can use CDD's import data wizard to load data into their Vault.

8. **Collaborate and share data securely**
    - Data in CDD Vault can be securely shared with colleagues and collaborators.
    - Researchers can share data in real time and avoid sending data over email.
    - Data in CDD Vault is organized by projects. Members are assigned to projects to easily control data access and visibility.
    - Each member can also be allocated further roles where roles act as project permissions that control how each member interacts with the data that they can access.

CDD Vault is a modern data management platform with an integrated ELN and inventory that allows NA researchers to bring together data from various experiments, instruments and analysis software. Both Biologists and Chemists can use CDD Vault not only as a 'single source of truth' but also for efficient collaboration and data sharing.

To get started, please contact us at info@collaborativedrug.com to set up a free trial of CDD Vault.

References:

Damase, T. R., Sukhovershin, R., Boada, C., Taraballi, F., Pettigrew, R. I., & Cooke, J. P. (2021). The Limitless Future of RNA Therapeutics. *Frontiers in bioengineering and biotechnology*, *9*, 628137. https://doi.org/10.3389/fbioe.2021.628137

Mallapaty S. (2021).India's DNA COVID vaccine is a world first – more are coming. *Nature* 597, 161-162. https://doi.org/10.1038/d41586-021-02385-x

Yu, A. M., Choi, Y. H., & Tu, M. J. (2020). RNA Drugs and RNA Targets for Small Molecules: Principles, Progress, and Challenges. *Pharmacological reviews*, 72(4), 862–898. https://doi.org/10.1124/pr.120.019554