"CDD allows us to analyze our data using different algorithms, parameters, and for different targets or compounds, which can be automated with its API."

- Hanjo Kim, Ph.D., Deputy Head of Research, Standigm

Drug Hunting with AI: A Novel Drug Discovery Paradigm

Standigm's revolutionary approach to drug discovery is aided by CDD Vault informatics platform

The biggest asset for any drug discovery organization is the ability to access and analyze insightful, contextual, and actionable data, which then informs and underpins critical decision making. New technologies that embrace the 'omics' era – genomics, proteomics, transcriptomics, metabolomics, etc. – and high-throughput, high-content screening and assay techniques are generating vast amounts of data on disease mechanisms, drug targets, and drug compounds. But trying to analyze all this data 'by hand', using traditional techniques, results in considerable lost intelligence.

The advent of new analytical capabilities in both the biological and chemical space, and the requirement to handle big data, spurred Seoul-based Standigm, an artificial intelligence (AI)-driven drug discovery company, to develop a suite of innovative machine learning algorithms and AI-based workflows that take much of the guesswork out of drug discovery process and accelerate outcomes. To effectively handle the vast amount of data required for this approach, Standigm has implemented the CDD Vault® informatics platform as a central data management infrastructure component in this discovery engine.

CDD Vault is a comprehensive data management solution that can be implemented as a single repository for seamlessly storing, managing, and mining diverse chemistry, biology and assay-related data. Accessed from any modern web browser, the hosted platform offers Standigm a way to access their harmonized data without loss of context.

1

Cutting-Edge AI Technology

Standigm's automated, full-stack drug discovery workflows are founded on its key AI platforms: Standigm ASK™ (ASK) for target identification, Standigm BEST™ (BEST) for lead generation, and Standigm Insight™ (Insight) for drug repurposing. The complete automation of molecular design is enabled through DarkMolFactory™, a truly industrialized, AI-enabled workflow that requires no human intervention to create molecular structures as pipeline assets.

ASK is as an AI-aided interactive platform for explainable disease target identification (Figure 1). "It has been developed by training a deep neural network to prioritize targets for any query disease," explained Hanjo Kim, Ph.D., deputy head of research at Standigm, "We now offer industry a free, basic version of Standigm ASK as an open platform."

"From the ASK system, we draw a target protein that we want to modulate to treat a disease of interest. It facilitates generation of first-in-class

drug compounds," Kim continued, "We can harness our collection of public databases to effectively explore the Standigm chemical space." Used for both drug discovery and optimization, the BEST platform is an AI-generated chemical space, which is used to map and interrogate and modulate chemical structures, develop analogues of, or optimize existing structures. "BEST is our novel compound generation technology, which allows us to identify compounds against any defined target." BEST leverages AI techniques to analyze, modify and fine-tune chemical moieties within the molecule, such as the scaffold, or linking fragments, and so optimize the compound structure.

With a candidate molecule in hand, the Insight platform then uses machine learning algorithms to identify additional, otherwise hidden disease indications for that compound, and predict molecular pathways and potential targets on which it may impact. "Insight also allows us to predict new indications and mechanisms of action, and highlight potentially new targets for approved drugs," Kim explained.
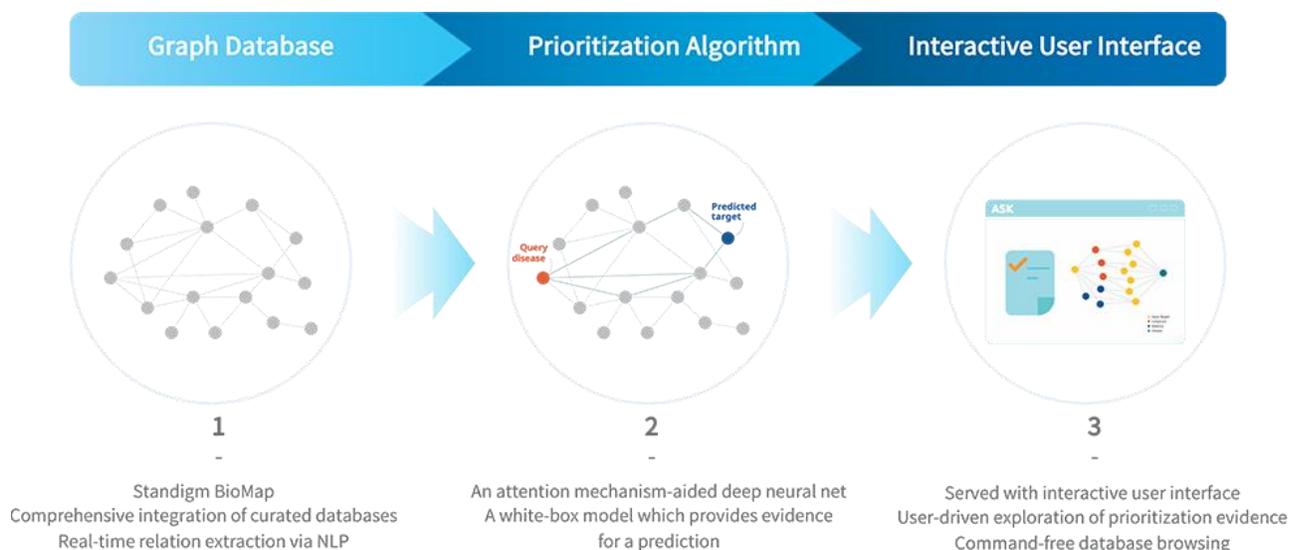

Figure 1: Standigm ASK™

Collaborative Drug Discovery – Standigm Case Study

Critically, the AI algorithms allow Standigm to build predictive models for activity, toxicity, ADME-related properties of specific molecules. "We can then use the results of our interrogations and design to filter and rank molecules, and so predict the best structure that will modulate the target in the desired way," Kim added.

Clean Data

Standigm has created vast, curated repositories of publicly available data on diseases, disease mechanisms, known drug targets and drug compounds, and has applied natural language processing (NLP) techniques to extract contextual data from published papers. These chemical and biological data repositories, known as Standigm ChemMap, and Standigm BioMap, can then be interrogated using the firm's state-of-the-art AI tools to identify new drug targets and compounds, and uncover previously unknown relationships and networks between diseases, biological mechanisms and potentially druggable pathways and proteins.

Importantly, the firm has dedicated considerable time and effort into making sure that the data in its repositories is 'clean' and highly curated. "Publicly available data, especially, are heterogeneous and often 'dirty,' so we have invested a lot of time to build a robust and trustful set of rules," Kim said. "The adage 'garbage in, garbage out' is really true, and data curation is very important so that we can understand the context of these data."

That is where good data management practice comes into play. Having CDD Vault as the infrastructure for managing vast amounts of data that must be interrogated alongside those repositories, is critical to the success of this approach, Kim suggested.
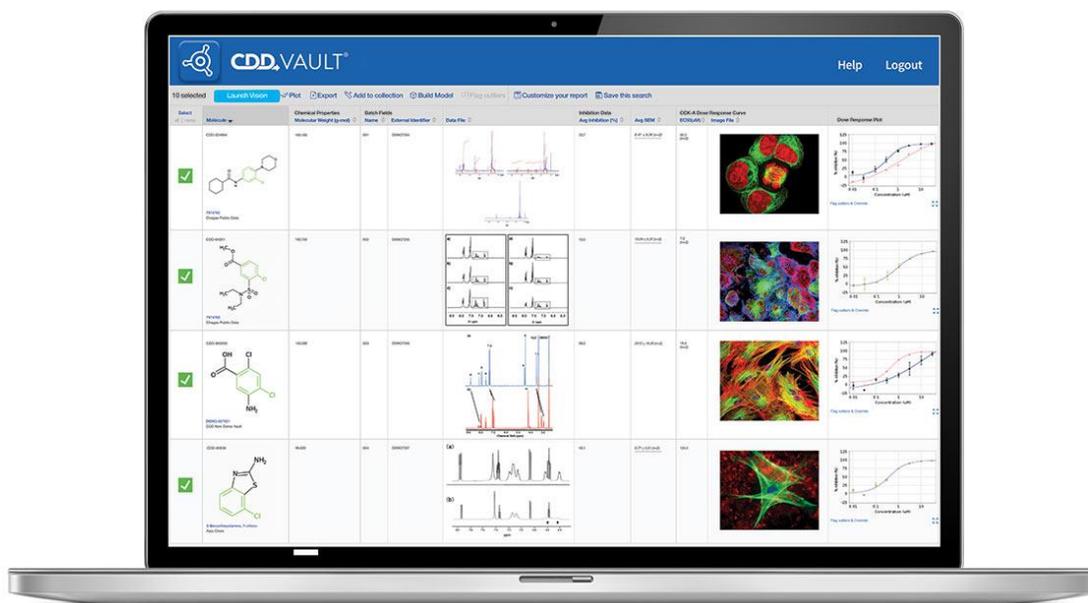

Figure 2: CDD Vault facilitates data capture, analysis, and sharing

Collaborative Drug Discovery – Standigm Case Study

"One notable advantage is that CDD hosts publicly available data sets," he added. "We purchase a lot of molecules, and can easily find those that we buy in the CDD public data source, together with related data. This combined data enriches our projects by adding additional data points. An on-premise solution would not have this feature, so it is specific to CDD Vault."

CDD Vault automates the management of all of this data without loss of context, but also gives Standigm the freedom to effectively access, search, mine, and manipulate diverse data in multiple ways, and using different approaches.

Data Security for Partnerships

Standigm is leveraging its AI technologies both in house, and through partnerships with the global pharma industry, to put greater intelligence into identifying drug targets, designing drug molecules and finding new indications for existing drug compounds.

For Standigm, CDD Vault represents a harmonized data repository platform to manage the huge volumes of external chemical and biological data coming in from diverse sources, including drug discovery partners and synthetic chemistry providers, as well as data derived in house. The vault offers a pivotal, holistic environment that supports how scientists can manage, search, and analyze cleanly curated data.

"We did consider building our own platform," said Kim, "But after discussions with technical experts, we decided not to reinvent the wheel. Standigm reviewed a number of platforms, but selected CDD Vault as the best solution that would enable it to route and analyze all of our externally derived data from CROs and other

partners, in parallel with our in-house data. Importantly, CDD Vault provides a suite of ready APIs so that our developers can play with our data in very different ways. Using CDD Vault we can integrate all of the different functionalities into our overall system."

Having CDD Vault as a central repository for data from partners, CROs and other providers, effectively gives Standigm the ability to effortlessly add new disease and compound structural data to its data repositories, automate data analysis through its AI system, and combine and configure data through workflows that can generate best insight on the relationship between diseases, targets and drug compound activity. "So CDD is key to that ability, and allows us to analyze those data again and again, using our different algorithms, different parameters and for different targets or different compounds. This process can be automated thanks to the CDD Vault API."

Standigm doesn't operate its own wet labs, and so partners with a number of CROs around the world to carry out experimental work, such as in vitro and in vivo assays, and compound synthesis. "Each of our CROs has its own way of reporting and recording data, but we must have a seamless capacity to integrate those data with our internal processes," Kim pointed out. "CDD Vault allows us to manage the huge amounts of data, and data types, coming in from all of our partners, automatically, and securely."
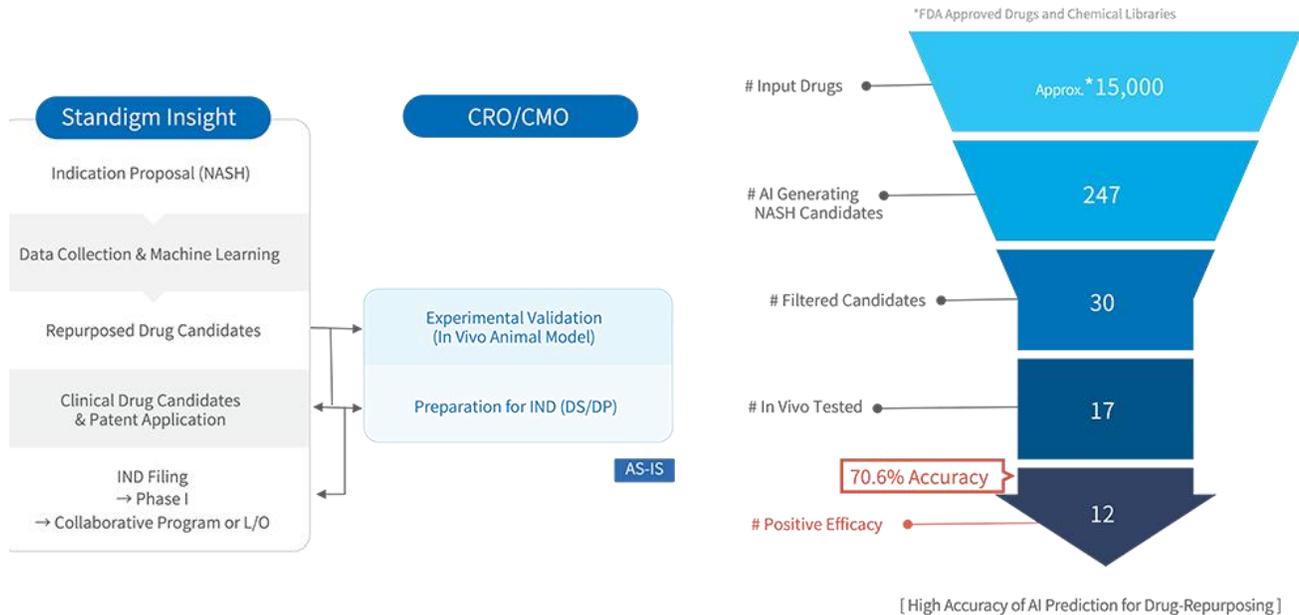
Collaborative Drug Discovery – Standigm Case Study

Figure 3: Standigm Insight™ program for NASH

This ability to automate the acquisition and storage of data from so many disparate sources becomes even more important as the number of projects Standigm undertakes grows, Kim continued. "When we are dealing with 20 or 30 different projects at a time, we have to have a comprehensive approach to managing that data, which gives us cohesive access to all of our data. And when we apply our own drug repurposing platform, Standigm Insight™, these growing contexts can make our products (drug candidates) more versatile."

A Strong Pipeline

Projects for new target identification, drug discovery and drug repurposing span a range of disease indications, including non-alcoholic steatohepatitis (NASH), fibrotic disease, cancer, and COVID-19, rheumatoid arthritis and some neurodegenerative and skin disorders.

Standigm announced a key project milestone in January 2021. The company, and partner SK Chemicals Co., Ltd, reported that through their open innovation agreement, signed in 2019, they had leveraged the Insight platform to identify a new rheumatoid arthritis indication for an already approved FDA drug. At the time of announcement, a use patent on the molecule had been filed.

The company also hopes clinical trials will also start during 2022 with clinical candidates identified though an in-house Insight program focused on NASH (Figure 3). The drug repurposing project initially evaluated more than 15,000 existing drug compounds, and yielded a dozen candidate molecules, from which three have been selected to progress to the next stage, Kim noted. "We now have use patents for three of these molecules, and are preparing IND applications for clinical development to start."

5

Collaborative Drug Discovery – Standigm Case Study

Moving forward, CDD Vault will continue to capture the full history and context of every compound in the Standigm database. "When we buy or synthesize a molecule, we collect all the relevant data from different sources in CDD Vault, allowing us to reuse the data at will. It may seem quite simple, but in fact, achieving this so seamlessly is a not an easy task."

About Collaborative Drug Discovery

Collaborative Drug Discovery provides a modern approach to drug discovery informatics that is trusted globally by thousands of leading researchers. Our CDD Vault is a hosted informatics platform that securely manages both private and external biological and chemical data. It provides core functionality including chemical registration, structure activity relationship, inventory, visualization, and electronic lab notebook capabilities. For more information, visit us at www.collaborativedrug.com.