nature portfolio

https://doi.org/10.1038/d41573-022-00025-1

Supplementary information

AI in small-molecule drug discovery: a coming wave?

In the format provided by the authors

Supplementary information | Datasets and analysis approaches

AI-native drug discovery companies included in this analysis

Table S1 (below) shows details of the AI-native companies included in this analysis. Companies were identified based on the reported central role of artificial intelligence (AI) technologies in their strategy for small-molecule drug discovery and their development of their own in-house pipeline of small-molecule drugs.

To assess the number of assets of AI-native drug discovery companies, we used Citeline Pharmaprojects (https://citeline.informa.com) as an external source. For some companies, this may under-represent the size of the AI-enabled pipeline, but it allows our analysis to be consistent across companies and over time. For four of the AI-native drug discovery companies, historical information was not available; these were therefore not included in Figure 1 and are indicated in the table below with an asterisk; for these four companies, the numbers in Table S1 were determined using manual research from company websites, SEC filings, investor presentations etc.

Descriptions of the company's AI applications have been compiled in Table S1, with platform names in parentheses. Each company applies AI in different areas depending on their discovery context; for example, some focus on identifying novel targets or mechanisms of action (such as Relay Therapeutics and Turbine), others leverage AI approaches to pursue drug repurposing opportunities (such as Aria Pharmaceuticals, Collaborations Pharmaceuticals and Healx) and others are combining several AI technologies in order to address a number of challenges across the drug discovery pipeline (such as Exscientia and Insilico Medicine).

Table S1 | AI-native companies and their applications and in-house assets

AI-native small-molecule drug discovery companies		**In-house assets from Citeline Pharmaprojects (or from manual research*)**		
Company	**AI application**	**Discovery/ preclinical**	**Clinical**	**Comments and code names of clinical-stage assets (where available)**
A2A Pharmaceuticals	Molecule generation and optimization (Sculpt™)	6	0	
Aria Pharmaceuticals	Bioinformatics target discovery, polypharmacology	25	0	Previously twoXAR Pharmaceuticals
Atomwise	Virtual screening through molecular recognition, structure-guided molecule generation and optimization (AtomNet®)	1	0	
Auransa	Bioinformatics target discovery (SMarTR ENGINE®)	12	0	
BenevolentAI	Bioinformatics target discovery, Molecule generation and optimization (Benevolent Platform™)	2	1	Clinical-stage asset: BEN-2293
Berg	Bioinformatics target discovery (Interrogative Biology®)	1	1	Clinical-stage asset: BPM31510 (ubidecarenone)
C4X Discovery	Genomics-based bioinformatics target discovery (Taxonomy3®), Structure guided molecule generation & optimization (Conformetrix)	9	1	Clinical-stage asset: INDV-2000
Collaborations Pharmaceuticals	Bioinformatics drug repurposing (MegaPredict), molecule generation and optimization (MegaSyn), mechanism, ADMET prediction (AssayCentral, MegaTox, MegaTrans)	23	0	

Cyclica*	Polypharmacology (Ligand Express®), molecule generation and optimization (Ligand Design™), off-target interaction prediction (MatchMaker™), ADMET predictions (POEM™)	2*	0*	
Denovicon Therapeutics	Bioinformatics target discovery, virtual screening, molecular optimization and ADMET predictions	2	0	
Exscientia	Bioinformatic target discovery, Phenotypic screening, molecule generation and optimization, ADMET prediction, clinical prediction based on patient tissue (all supported by CentaurAI systems)	7	3	Clinical-stage assets: DSP-0038, DSP-1181, EXS-21546
Frontier Medicines	Protein hotspot mapping, generation of compound libraries, molecule optimization	3	0	
Healx	Bioinformatics to discover novel drug–target relationships/repurposing (Healnet)	21	0	
Insilico Medicine*	Bioinformatics in novel target discovery (PandaOmics®), molecule generation and optimization with ADMET prediction (Chemistry42®), clinical trial prediction (InClinico®)	7*	0*	
Insitro*	Disease modelling, deconvolution of *in vitro* phenotypic disease models	2*	0*	
Nimbus Therapeutics	Molecular dynamics, ADME predictions	6	2	Clinical-stage assets: NDI-034858, NDI-101150
Pharos iBio	Bioinformatic target discovery, Protein structure characterization (e.g. binding site ID), molecule optimization and ADMET prediction (Chemiverse)	4	1	Clinical-stage asset: PHI-101
Recursion Pharmaceuticals	Bioinformatic target discovery, experimental target validation and hit identification with phenotypic screening (Recursion Operating System)	11	3	Clinical-stage assets: REC-2282, REC-4881, REC-994
Relay Therapeutics	Molecular dynamics and physics including protein motion (Dynamo Platform), molecule identification (from acquisition of ZebiAI)	3	2	Clinical-stage assets: RLY-1971, RLY-4008 Recent acquisition of ZebiAI (AI-enabled DEL discovery company)
Roivant Sciences	Molecular dynamics and physics including protein motion (Silicon Therapeutics), protein degrader deep learning models (VantAI)	6	1	Clinical stage asset: SNX-281 Discovery assets from acquisition of Silicon Therapeutics only
Schrödinger Inc	Molecular dynamics, molecule generation optimization, property prediction	7	0	
Turbine*	Predictive cell behaviour modelling to interrogate mechanism of action (Simulated Cell™)	3*	0*	
Valo Health	Bioinformatic target–biomarker linkages (leveraging patient data), virtual screening and molecule optimization, clinical simulations (Opal Platform)	4	0	
Verge Genomics	Genomic-based bioinformatics in novel target discovery	5	0	

Historical pipeline growth analysis
We used Citeline Pharmaprojects (https://citeline.informa.com) to aggregate historical data from our list of AI-native small-molecule drug discovery companies where available (and so only for 20 of the 24 companies above), as well as the top 20 pharmaceutical companies. Data was aggregated annually and reflects a snapshot of the respective year. Potential changes in program status (for example, progression into clinical stages or discontinuation of a program) are reflected by assigning programs to their new status (for example, from phase I to phase II) in the following year. Discontinued programs were removed from the analysis for the following years.

For the AI companies, inclusion criteria were that the asset appeared as part of the respective company or its associated subsidiaries and had an active status (for example, no discontinued programs). Additionally, since Citeline curates historical data only once per year in May, we added missing data points from the individual entries for the current year.

For the top 20 pharma companies, inclusion criteria were that the asset appeared as originated, not licensed, within the respective company or its subsidiaries and had an active status. Potential internal AI-driven discovery efforts from top 20 pharma companies were not separated from the analysis, due to limited publicly available information. Top 20 pharma companies were defined by their total revenues in 2020 (FiercePharma).

Pipeline composition analysis
For the top 20 pharma companies, we used Evaluate (https://www.evaluate.com/) to extract the pipeline assets, associated therapeutic areas and target classes.

Since not all AI companies are included in commonly used databases, we aggregated an as-complete-as-possible current pipeline of AI companies manually, leveraging information from their website, SEC filings, annual reports, press releases, and where possible cross-referenced with commonly used data bases (for example, Evaluate, Citeline). The resulting pipeline is larger than the historical pipeline of Figure 1, mostly due to the addition of missing companies and early discovery programs.

Reported examples of AI-derived compounds

SHP2
US Clinical Trial Registry: Identifier NCT04252339. RLY-1971 in Subjects With Advanced or Metastatic Solid Tumors: https://clinicaltrials.gov/ct2/show/NCT04252339

WRN
Roivant Sciences company website: https://discovery.roivant.com/therapeutics-and-pipeline/

Nimbus Therapeutics Announces Expansion of its Drug Discovery Pipeline Across Oncology, Immunology and Metabolism: https://www.nimbustx.com/2020/06/08/nimbus-therapeutics-announces-expansion-of-its-drug-discovery-pipeline-across-oncology-immunology-and-metabolism/

MALT1
Yin W. et al. Identification of Potent Paracaspase MALT1 Inhibitors for Hematological Malignancies. *Blood* **136** (Supplement 1), 30 (2020).

Schrödinger to Present Data from Its MALT1 Inhibitor Program at the American Society of Hematology (ASH) 2020 Annual Meeting: https://ir.schrodinger.com/news-releases/news-release-details/schrodinger-present-data-its-malt1-inhibitor-program-american/

TYK2

Nimbus Therapeutics Announces First Patient Dosed in Phase 2b Study of Oral Allosteric TYK2 Inhibitor in Patients with Moderate to Severe Psoriasis; 2021 Sep 14; Available from https://www.nimbustx.com/2021/09/14/nimbus-therapeutics-announces-first-patient-dosed-in-phase-2b-study-of-oral-allosteric-tyk2-inhibitor-in-patients-with-moderate-to-severe-psoriasis/

Masse C. et al. Identification of highly potent and selective Tyk2 inhibitors for the treatment of autoimmune diseases through structure-based drug design. *J Immunol* **194** (1 Supplement) 67.12 (2015).

Masse C. et al. TYK2 inhibitors and uses thereof. US Patent US9630970B2. (2017)

Serotonin receptors

Exscientia announces second molecule created using AI from Sumitomo Dainippon Pharma collaboration to enter Phase 1 clinical trial; 2021 May 13; Available from https://investors.exscientia.ai/press-releases/press-release-details/2021/exscientia-announces-second-molecule-created-using-ai-from-sumitomo-dainippon-pharma-collaboration-to-enter-phase-1-clinical-trial/Default.aspx

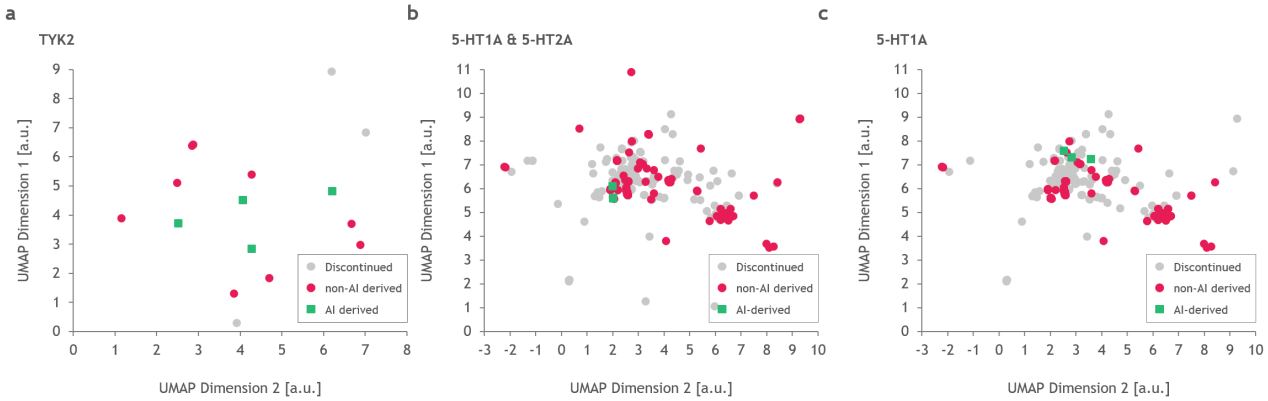
Sumitomo Dainippon Pharma and Exscientia Joint Development New Drug Candidate Created Using Artificial Intelligence (AI) Begins Clinical Study: https://www.ds-pharma.com/ir/news/2020/20200130.html

Yoshinaga H. et al. Fused Ring Lactam Derivatives. Japanese patent PCT/JP2019/028577. (2019)

Yoshinaga H. et al. 2,6-disubstituted pyridine derivative. Japanese patent PCT/JP2018/009418. (2018)

Chemical space analysis

To compare chemical space of AI-derived and classically discovered small molecules, we leveraged a method published by Reverie Labs (https://blog.reverielabs.com/mapping-chemical-space-with-umap/) to compute chemical space referenced to random sample of 100,000 compounds from CHEMBL (https://www.ebi.ac.uk/chembl/) and applied Uniform Manifold Approximation and Projection (UMAP), a non-linear dimensionality reduction technique, to plot a two-dimensional approximation of chemical space. The AI-derived molecules were selected from relevant patents and traditional small-molecule therapeutics were identified for the same target by leveraging Citeline Pharmaprojects (https://citeline.informa.com), where SMILES strings were available.



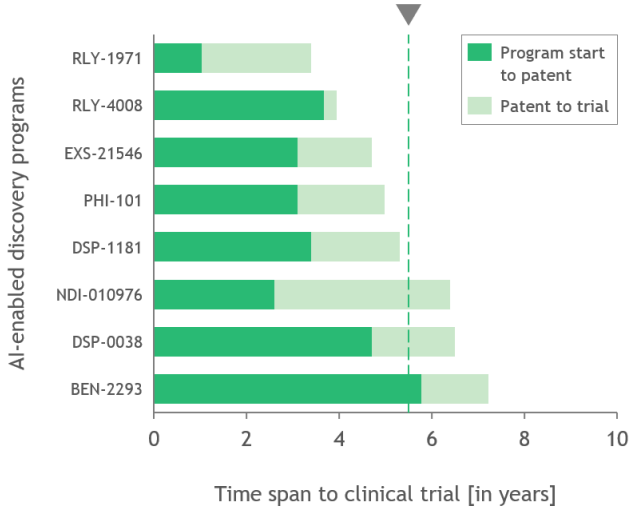
Supplementary Figure 1 | Chemical space analysis of selected AI-derived assets. a | Comparison of chemical space for different assets targeting TYK2. **b** | Comparison of chemical space for different assets targeting 5-HT1A and 5-HT2A. **c** | Comparison of chemical space for different assets targeting 5-HT1A. UMAP is a non-linear dimensionality reduction technique that allows to display chemical space, and thus differences between molecules, in low dimensions.

Timeline reconstruction
For a subset of clinical stage assets of AI companies, we were able to reconstruct the timeline from program start until entering the clinic through outside-in analysis. We excluded from the analysis assets for which we could not clearly assign a published patent, as well as assets which were repurposed.

Program start. Program start is defined as the timepoint at which the start of the program has been announced. Start points can be press releases, collaboration announcements, or minority stake announcements. In case of no explicit mentioning of the start date, we took the foundation date of the company as a start date.

Patent. Patent timepoint is defined as the date of the first patent application filing of the discovery program.

Clinical trial. Clinical trial start dates were reconstructed from clinical trial registries, or from company announcements and press releases.



Supplementary Figure 2 | Outside-in timeline analysis of selected AI-enabled discovery programs. The analysis is based on externally published time-points (collaboration start dates; patent applications and trial start times). Dotted line indicates industry average from target-to-hit until start of clinical trials, based on: Paul et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery* 9, 203–214 (2010). Further details on clinical stage assets included in this analysis can be found in Table S2 (below).

Please note that some AI companies have reported internal timelines for their discovery programs that are much faster than those shown in Supplementary Figure 2. For example, Exscientia reports that 7 programs took <18 months from target identification to candidate identification (Exscientia Corporate Presentation, January 11 2022). And Insilico Medicine reported the development of a preclinical candidate for IPF in <18 months, which subsequently entered first-in-human studies in 9 months; a second preclinical candidate for kidney fibrosis was reportedly developed in 6 months.

Table S2 | Details on clinical-stage assets included in Supplementary Figure 2

Company	Asset	Target	Disease/therapeutic area	Current phase
Relay Therapeutics	RLY-1971 / RG-6433	SHP2	Solid tumours	Phase I
	RLY-4008	FGFR2	FGFR2-driven cancers (ICC and other advanced solid tumors)	Phase I
Exscientia	EXS-21546	A2AR	Solid tumours	Phase I
	DSP-1181	5-HT1A	Obsessive compulsive disorders	Phase I
	DSP-0038	5-HT1A / 5-HT2A	AD psychosis	Phase I
Pharos iBio	PHI-101	FLT3	Acute myelogenous leukemia, platinum-resistant refractory ovarian cancer, other cancers	Phase I
Nimbus Therapeutics	NDI-010976 / GS-0976	ACC	NASH	Phase II
BenevolentAI	BEN-2293	Pan-Trk	Atopic dermatitis	Phase I