Supplementary information

Herding in the drug development pipeline

In the format provided by the authors

Data sources and analysis

Figure 1a
Using PharmaProjects Trends data (that is time series/year-by-year, where data is available from 1995 onward), we identified all targets for which the first approval for that target occurred after 1995 and for which two or more drugs have been launched. This yielded a total of 124 targets assessed (data extracted in June 2022). For every year, we assessed the number of drugs being actively investigated for those targets (phase I–registration) divided by the number of targets being actively investigated in that year. Targets were organized into oncology versus non-oncology related categories using PharmaProjects classification data. Total assets per target does not represent the average of oncology and non-oncology, as targets may be pursued in both categories and total therefore additive (for example, if drug A and drug B both pursue the same target, with drug A in oncology and drug B outside of oncology, the total assets per target will be 2, while oncology and non-oncology assets per target will each be 1).

Figure 1b
The targets identified in Figure 1a were rank ordered year-by-year by number of assets pursuing the target. The number of assets for each target over time were evaluated.
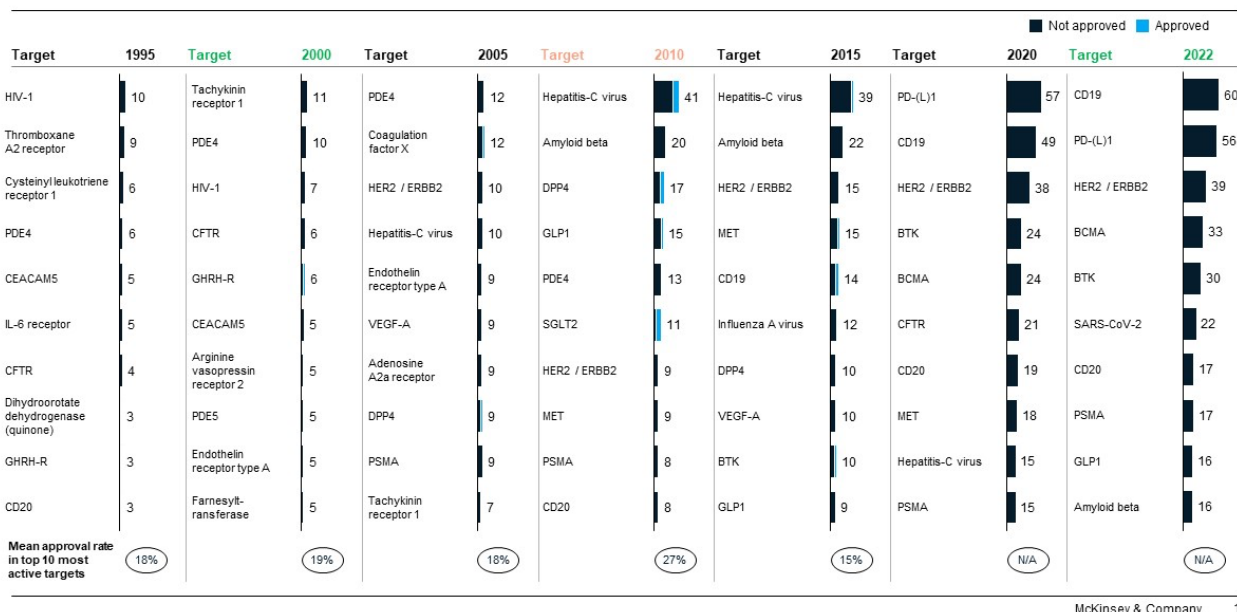
Figure 1c
The rank-ordered targets from Figure 1b were listed individually for selected years and for the top 10 most active targets in a given year.

Figure 2a
We identified the top 10 pharma companies in 2000 and 2020 based on publicly available revenue data. Utilizing PharmaProjects Trends data as described for Figure 1a, we identified targets within each companies' pipeline and the number of drugs being actively investigated for those targets (phase I–registration).

Figure 2b
We queried a PharmaProjects snapshot of all drugs currently in clinical development (phase I–phase III) for monogenic diseases and identified diseases with greater than 10 assets in active clinical development. For each identified disease, we determined the number of patients needed to enroll all active clinical trials for assets, based on average clinical trial sizes for that disease (from ClinicalTrials.gov). We compared this with the identified US total prevalent population based on searches of applicable epidemiological literature.

Target	1995	Target	2000	Target	2005	Target	2010	Target	2015	Target	2020	Target	2022
HIV-1	10	Tachykinin receptor 1	11	PDE4	12	Hepatitis-C virus	41	Hepatitis-C virus	39	PD-(L)1	57	CD19	60
Thromboxane A2 receptor	9	PDE4	10	Coagulation factor X	12	Amyloid beta	20	Amyloid beta	22	CD19	49	PD-(L)1	56
Cysteinyl leukotriene receptor 1	6	HIV-1	7	HER2 / ERBB2	10	DPP4	17	HER2 / ERBB2	15	HER2 / ERBB2	38	HER2 / ERBB2	39
PDE4	6	CFTR	6	Hepatitis-C virus	10	GLP1	15	MET	15	BTK	24	BCMA	33
CEACAM5	5	GHRH-R	6	Endothelin receptor type A	9	PDE4	13	CD19	14	BCMA	24	BTK	30
IL-6 receptor	5	CEACAM5	5	VEGF-A	9	SGLT2	11	Influenza A virus	12	CFTR	21	SARS-CoV-2	22
CFTR	4	Arginine vasopressin receptor 2	5	Adenosine A2a receptor	9	HER2 / ERBB2	9	DPP4	10	CD20	19	CD20	17
Dihydroorotate dehydrogenase (quinone)	3	PDE5	5	DPP4	9	MET	9	VEGF-A	10	MET	18	PSMA	17
GHRH-R	3	Endothelin receptor type A	5	PSMA	9	PSMA	8	BTK	10	Hepatitis-C virus	15	GLP1	16
CD20	3	Farnesyl-transferase	5	Tachykinin receptor 1	7	CD20	8	GLP1	9	PSMA	15	Amyloid beta	16
Mean approval rate in top 10 most active targets	18%		19%		18%		27%		15%		N/A		N/A

Legend: ■ Not approved ■ Approved

Supplementary Figure 1

The rank ordered targets from Figure 1b were listed individually for selected years and for the top 10 most active targets in a given year. The proportion of drugs approved was determined based on June 2022 PharmaProjects data (analysis not carried out for the 2015 and 2020 development pipelines to avoid a bias from insufficient time for follow-up).