nature portfolio

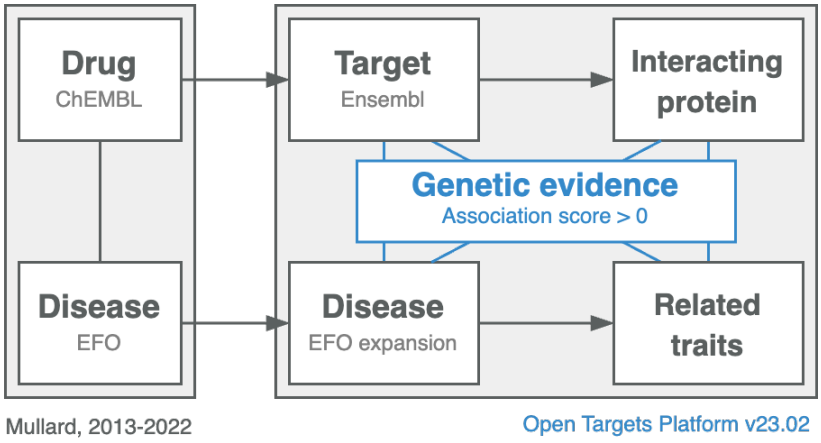
https://doi.org/10.1038/d41573-023-00158-x

Supplementary information

Genetic support for FDA-approved drugs over the past decade

In the format provided by the authors

Data curation and genetic evidence search
The list of 428 FDA drug approvals from 2013 to 2022 was manually curated from annual articles by Asher Mullard in *Nature Reviews Drug Discovery*[1–10]. The methodology for data curation and the search for genetic evidence, primarily adapted from a previous publication by Ochoa et al.[11], is illustrated in Supplementary Figure 1. The code and data necessary to reproduce the analysis and figures can be found at https://github.com/opentargets/approvalsSupport.



Supplementary Figure 1 | Data curation and genetic evidence search for drugs approved by FDA from 2013 to 2022.

Once the approved drug–indication pairs were curated, the on-target gene products were mapped to their corresponding Ensembl genes using the ChEMBL database[12]. An additional 32 mechanisms of action (MoAs) not originally documented in ChEMBL were incorporated after cross-referencing the DrugBank platform for pharmacological targets of the corresponding drugs. For enzyme therapies, the targets were designated as the enzymes themselves.

In addition, the network of targets was expanded by incorporating physically interacting proteins, a methodology systematically employed in previous studies [13]. A stringent criterion was adopted, requiring strong support (MI>0.42) for physical (not functional) interactions, as reported by the IntAct database [14].

The indications for the approved drugs were manually mapped to the Experimental Factor Ontology (EFO)[15] or Monarch Merged Disease Ontology (MONDO)[16] identifiers for the exact or related conditions, using the EMBL-EBI Ontology Lookup Service[17]. Related conditions were incorporated for 54 original indications, considering the pathogenesis and phenotypic similarity of the traits.

To retrieve the supporting genetic evidence for the individual approvals, we leveraged evidence integrated by the Open Targets Platform (https://platform.opentargets.org/)[18]. Among all the data sources in the platform, we considered a subset of 17 resources with a clear genetic basis as reported in the 23.02 release:

- Somatic: CGC (COSMIC) [19], IntOgen [20], Cancer Biomarkers (CGI) [21], ClinVar (Somatic) [22]
- Functional genomics (cancer): Project Score [23], SlapEnrich [24], Progeny [25]
- Common disease: OT Genetics Portal [26], Gene burden [18]

- Rare mendelian: ClinVar [22], Clingen [27], GEL PanelApp [28], Orphanet, gene2phenotype [29], Uniprot (gene-disease) and Uniprot (variants) [30]
- Mouse model: IMPC [31]

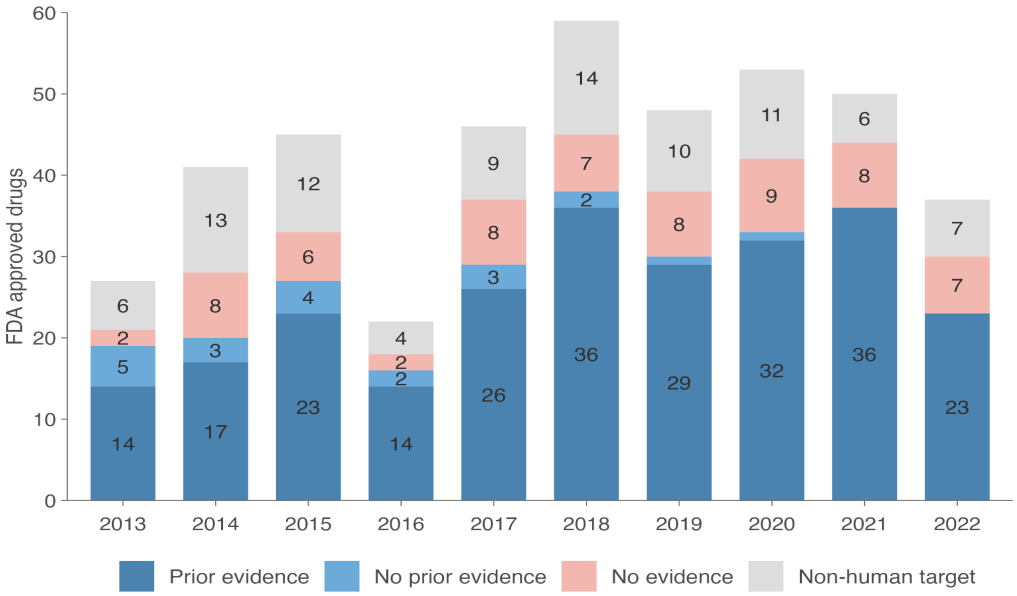
To maximise the phenotypic overlap between the drug indication and the genetic evidence trait, we propagated the genetic evidence to more specific phenotypes using the indirect associations from the platform https://platform-docs.opentargets.org/associations.

In order to estimate the earliest available date at which genetic support was available for the given target–disease pair (including interactors and related traits), we used as proxy the publication date associated with the evidence according to the Open Targets Platform. Drugs with prior genetic evidence were defined as those approved in years after the earliest genetic evidence was published. When the year of genetic evidence was the same as the year of approval, it was recognized as support for prior genetic evidence.

A drug–disease pair was considered to have genetic evidence support if any data source presented an Association Score greater than 0 for the pair, which included both original target–disease pairs and pairs involving interacting proteins or related conditions (Supplementary Figure 1). When a drug can target more than one gene according to the mechanism of action, genetic evidence in any of the targets was used to qualify as genetically supported.

Available genetic support for the approved drugs

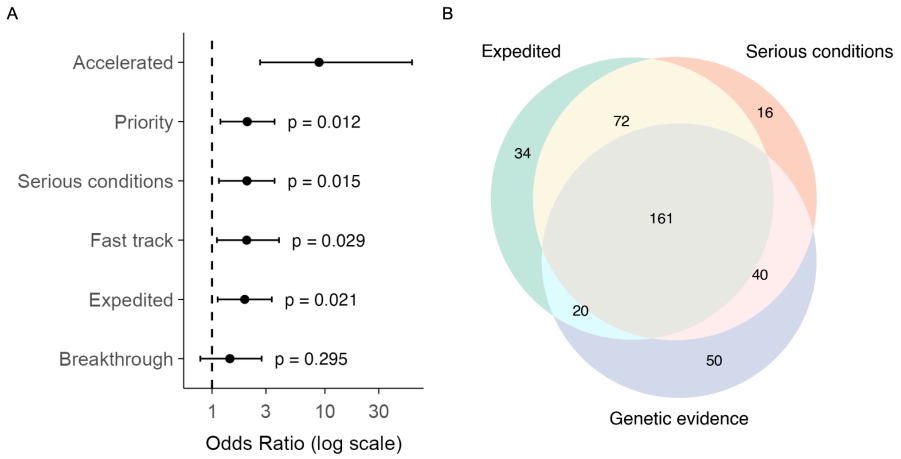
Out of the 428 approvals, 271 (63%) have support from genetic evidence, and we could link 250 of these to evidence publicly available before the approval date (Supplementary Figure 2). Alternatively, 65 out of 428 approvals (15%) lack support from genetic evidence and 84 (20%) have no human target (for example, the approval of remdesivir for COVID-19 treatment in 2020). A significant portion of these approvals without genetic evidence were observed to be for symptomatic treatments. This could be attributed either to the inherent symptomatic nature of the condition (such as chemotherapy-related nausea) or to the symptomatic mechanism of action (MoA) of the drug itself, as exemplified by deflazacort (2017), an anti-inflammatory treatment approved for Duchenne muscular dystrophy (DMD), as discussed in the main text.



Supplementary Figure 2 | Annual breakdown of drugs approved by the FDA from 2013 to 2022 based on the presence and timing of genetic evidence support. The depicted categories represent the number drugs with prior genetic evidence (shown in dark blue), drugs with any genetic evidence (light blue), drugs without genetic evidence (pink) and drugs with non-human targets or unknown mechanisms of action (grey).

The annotated list of drug approval and their respective genetic support according to the curation is available as Supplementary Table 1. All genetic evidence support for FDA-approved drugs, categorised by sources and by years, is illustrated in Supplementary Figures 4–13 below.

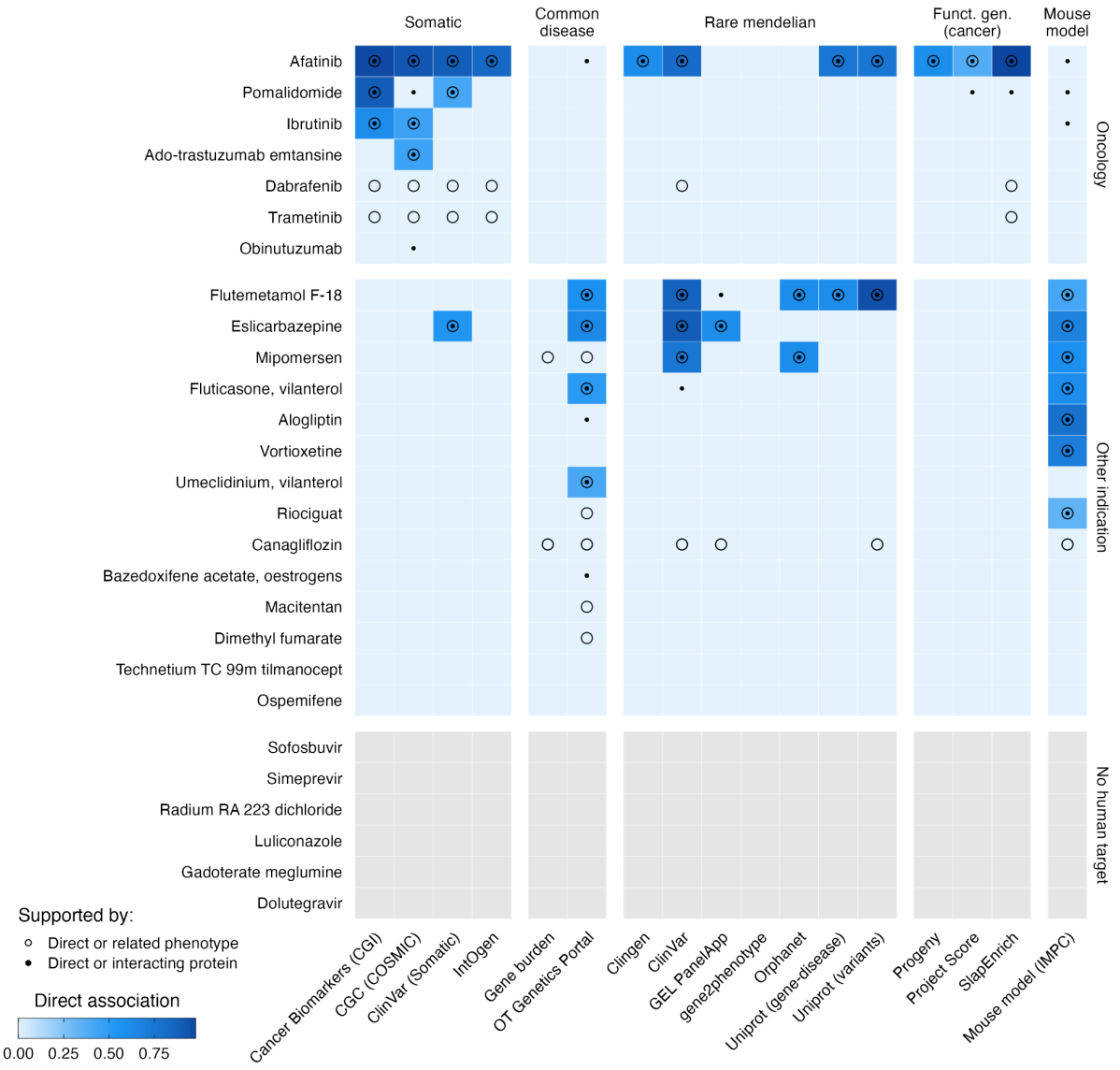
Genetic evidence presence for expedited approvals and serious diseases
Data concerning Accelerated (A), Priority (P), and Breakthrough (B) FDA review statuses were sourced from the publications by Asher Mullard. The information about the Fast Track (F) status for 2013–2022 drug approvals was manually retrieved from the official FDA website. The dataset of 2013–2022 approved drugs primarily consisted of expedited approvals (310 out of 428) including A, P, B, and F, in all possible combinations. To assess the variance in genetic evidence support for expedited approvals, odds ratios (ORs) were calculated (Supplementary Figure 3A).



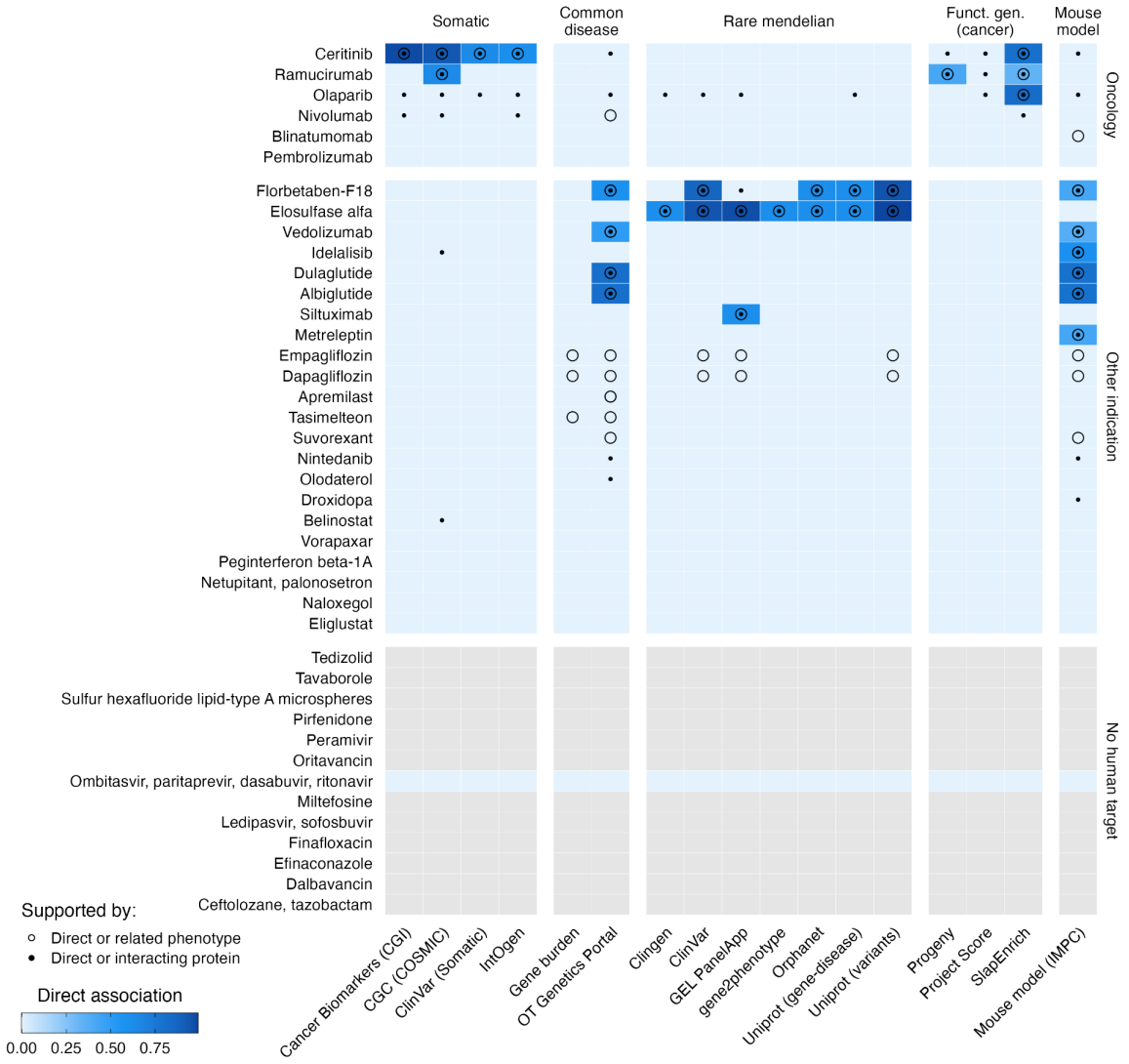
Supplementary Figure 3 | Association of genetic evidence with expedited approval statuses and serious conditions. A, Odds ratios comparing the approval type (either expedited review status or serious condition) with the presence of genetic evidence support. **B**, Overlap among approvals supported by genetic evidence, approvals with an expedited status, and approvals for serious conditions. The quantities denote the number of approvals in each category.

For the categorization of all conditions into serious and non-serious groups, ChatGPT-4 was utilised. A prompt with the definition of serious conditions, as per the FDA's official guidelines, was employed. Any conditions not aligning with this definition were labelled as either "Not serious" or "NA". To categorise all conditions from the approvals dataset spanning 2013–2022, ChatGPT-4 was run three times for each condition. If the results were concordant, the derived classification was accepted. In the case of incongruent results, the conditions were manually reassigned as either "Serious", "Not serious", or "Not concordant". Consequently, out of the 428 approvals, 289 (68%) were ascertained to be for serious conditions. ORs and p-values were computed using an exact Fisher test specifically for the "serious" category based on its presence or absence, given its more stringent definition (Supplementary Figure 3A).

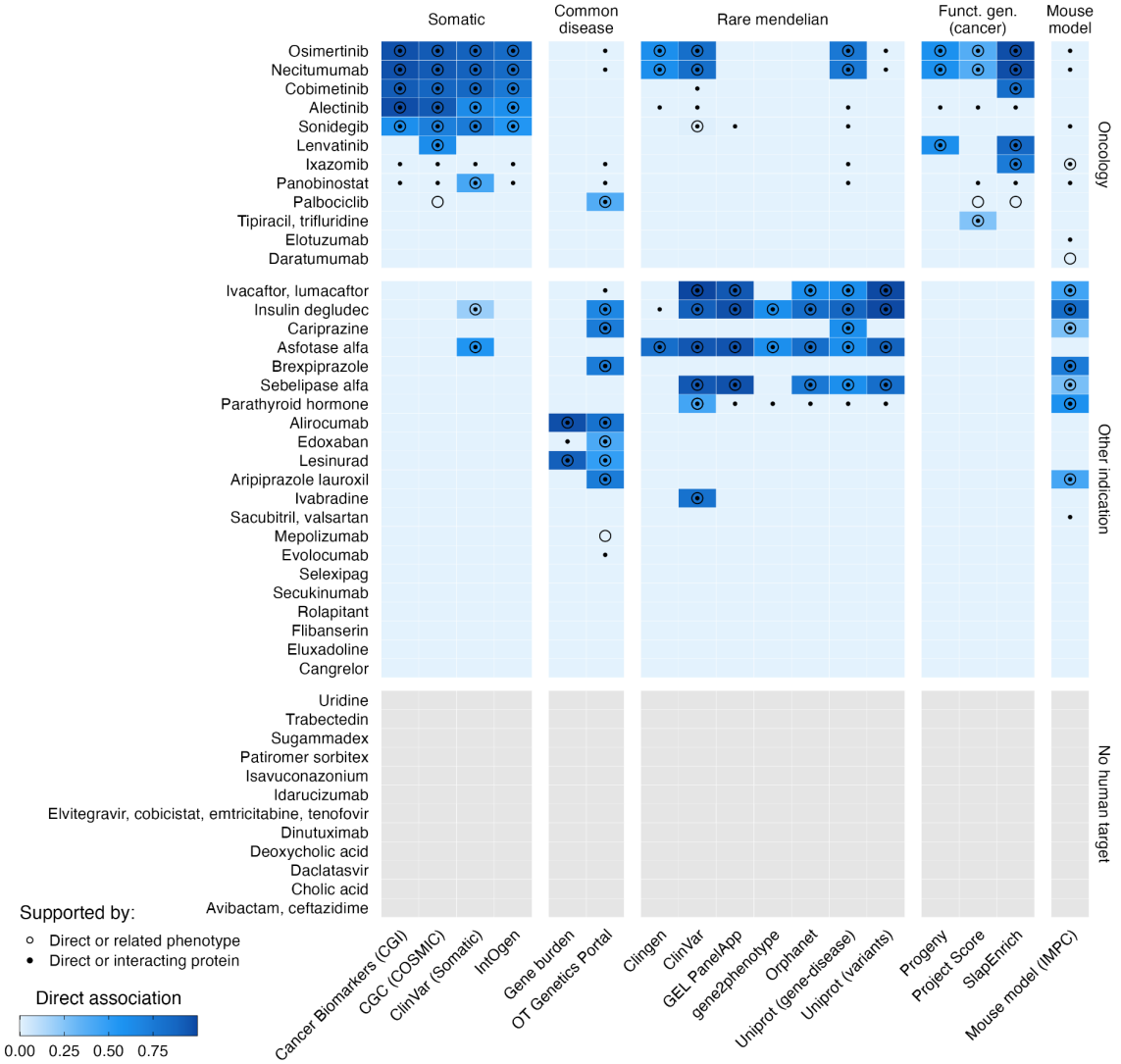
Expedited drug approvals are found to have twice as much genetic evidence support (OR = 1.95 [1.12–3.38], p = 0.021). However, it remains challenging to establish whether the availability of genetic evidence directly accelerates the approval process. One hypothesis could be that genetic evidence may contribute to a more robust therapeutic hypothesis, identifying targets causally linked with the disease, therefore increasing the regulators confidence on the drug efficacy. However, it might be that serious conditions — which constitute the majority of expedited approvals — have more detailed genetic characterization due to extensive research, which implies they tend to be better supported by genetic evidence (OR = 2.03 [1.15–3.56], p = 0.015). The intricacies of this issue are visualised in Supplementary Figure 3B.



Supplementary Figure 4 | Supporting genetic evidence for 27 drugs approved by the FDA in 2013. Evidence for each drug target (T) – disease (D) indication is divided by oncology drugs (top panel), drugs for other indications with human targets (middle panel) and drugs without a human target (bottom panel). The x-axis is grouped by the predominant genetic data in the respective resources. Support by data source is displayed as: a) evidence for T and D or any other descendant in the EFO ontology aggregated using the Open Targets Platform association score (blue heatmap colours); b) evidence for T and D or closely related phenotypes to D added by manual curation (open circles); and/or c) evidence for D and T or any protein physically interacting with T (black dots).

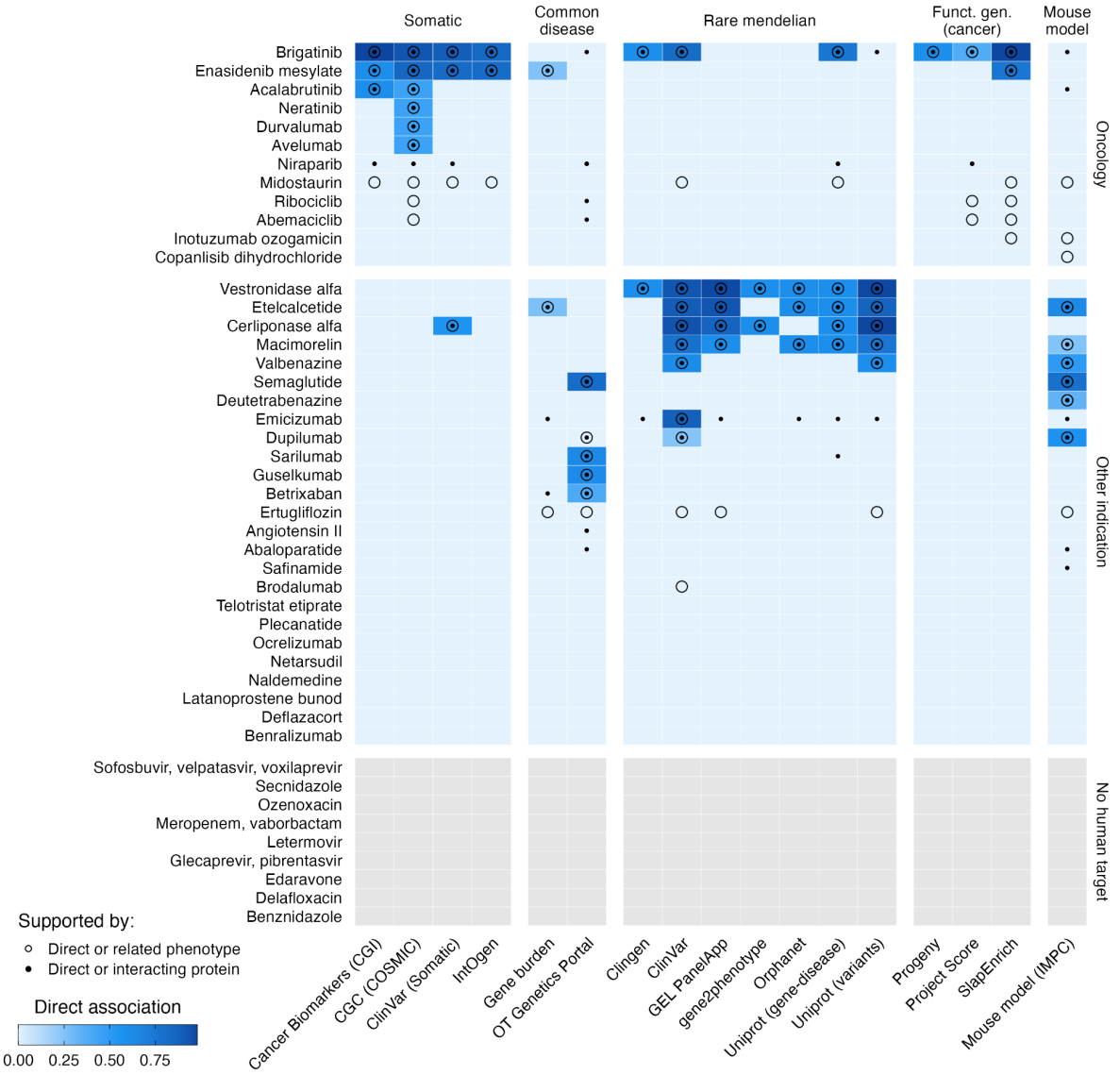


Supplementary Figure 5 | Supporting genetic evidence for 41 drugs approved by the FDA in 2014. Evidence for each drug target (T) – disease (D) indication is divided by oncology drugs (top panel), drugs for other indications with human targets (middle panel) and drugs without a human target (bottom panel). The x-axis is grouped by the predominant genetic data in the respective resources. Support by data source is displayed as: a) evidence for T and D or any other descendant in the EFO ontology aggregated using the Open Targets Platform association score (blue heatmap colours); b) evidence for T and D or closely related phenotypes to D added by manual curation (open circles); and/or c) evidence for D and T or any protein physically interacting with T (black dots). Viekira Pak (Ombitasvir, paritaprevir, dasabuvir, ritonavir), approved for chronic HCV genotype 1 infection, is considered to “no human target” due to its antiviral mechanism, with only ritonavir serving as a CYP3A inhibitor to boost paritaprevir concentrations, not as an active agent against HCV.

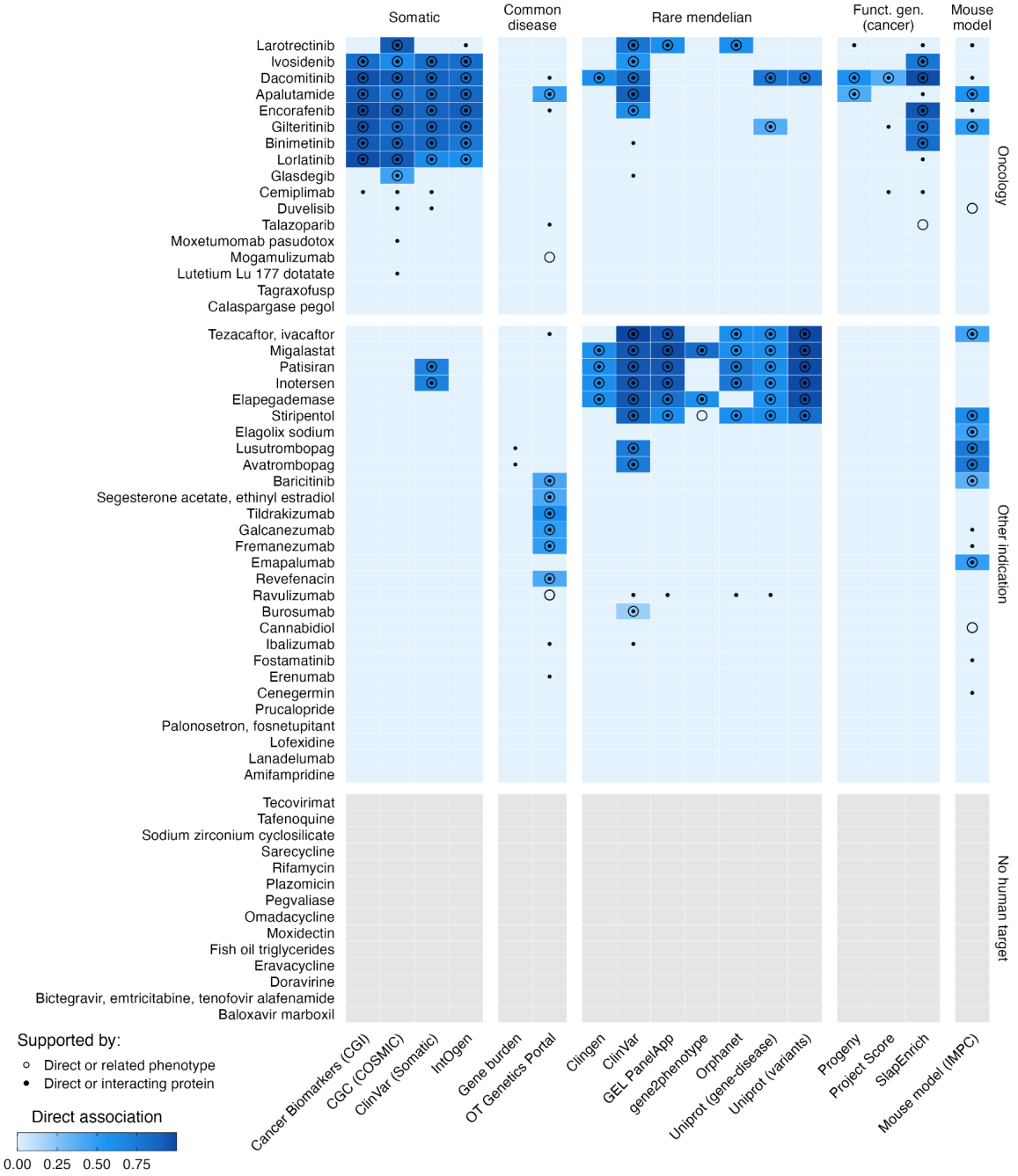


Supplementary Figure 6 | Supporting genetic evidence for 45 drugs approved by the FDA in 2015. Evidence for each drug target (T) – disease (D) indication is divided by oncology drugs (top panel), drugs for other indications with human targets (middle panel) and drugs without a human target (bottom panel). The x-axis is grouped by the predominant genetic data in the respective resources. Support by data source is displayed as: a) evidence for T and D or any other descendant in the EFO ontology aggregated using the Open Targets Platform association score (blue heatmap colours); b) evidence for T and D or closely related phenotypes to D added by manual curation (open circles); and/or c) evidence for D and T or any protein physically interacting with T (black dots).

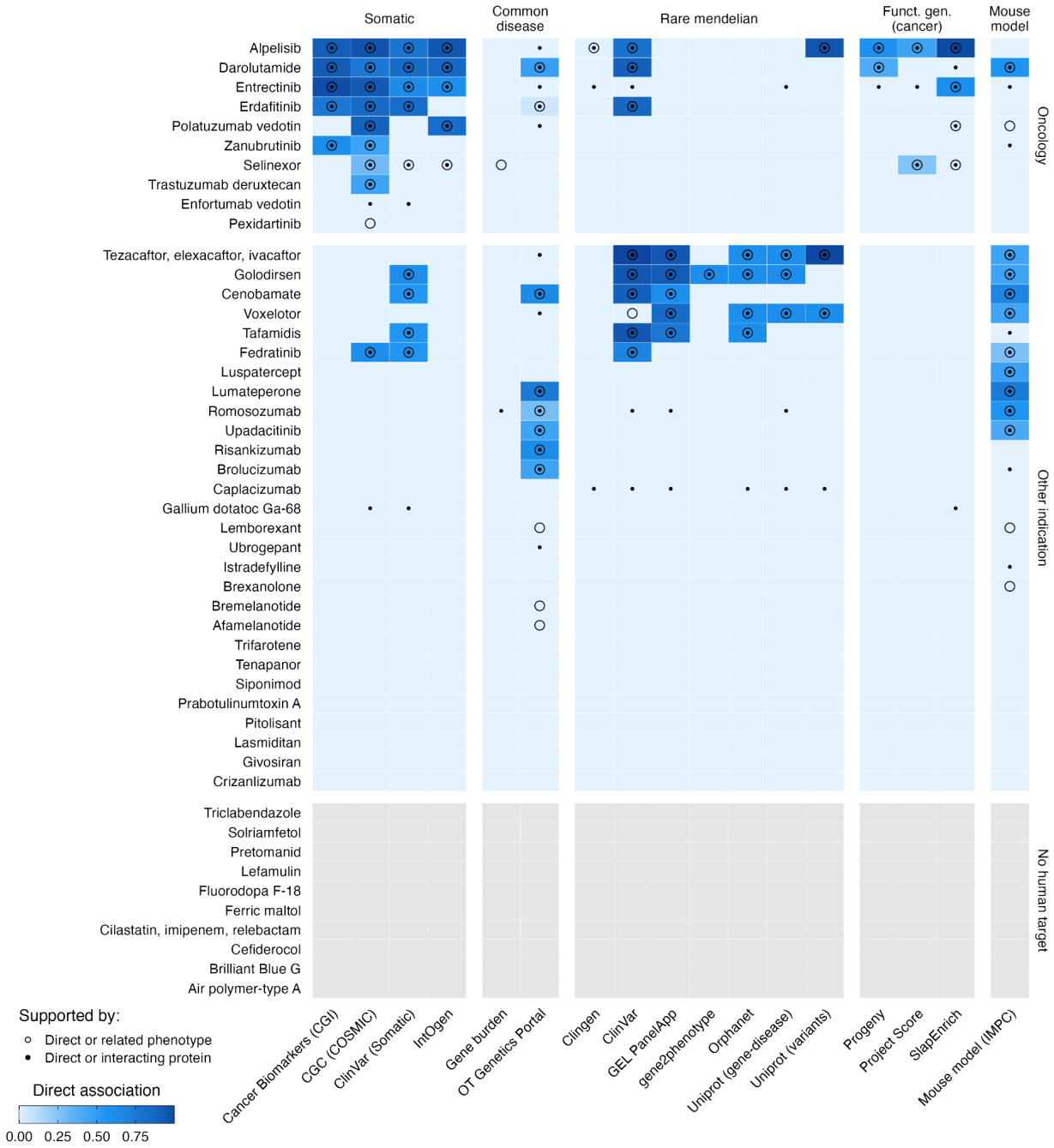
Supplementary Figure 7 | Supporting genetic evidence for 22 drugs approved by the FDA in 2016. Evidence for each drug target (T) – disease (D) indication is divided by oncology drugs (top panel), drugs for other indications with human targets (middle panel) and drugs without a human target (bottom panel). The x-axis is grouped by the predominant genetic data in the respective resources. Support by data source is displayed as: a) evidence for T and D or any other descendant in the EFO ontology aggregated using the Open Targets Platform association score (blue heatmap colours); b) evidence for T and D or closely related phenotypes to D added by manual curation (open circles); and/or c) evidence for D and T or any protein physically interacting with T (black dots).



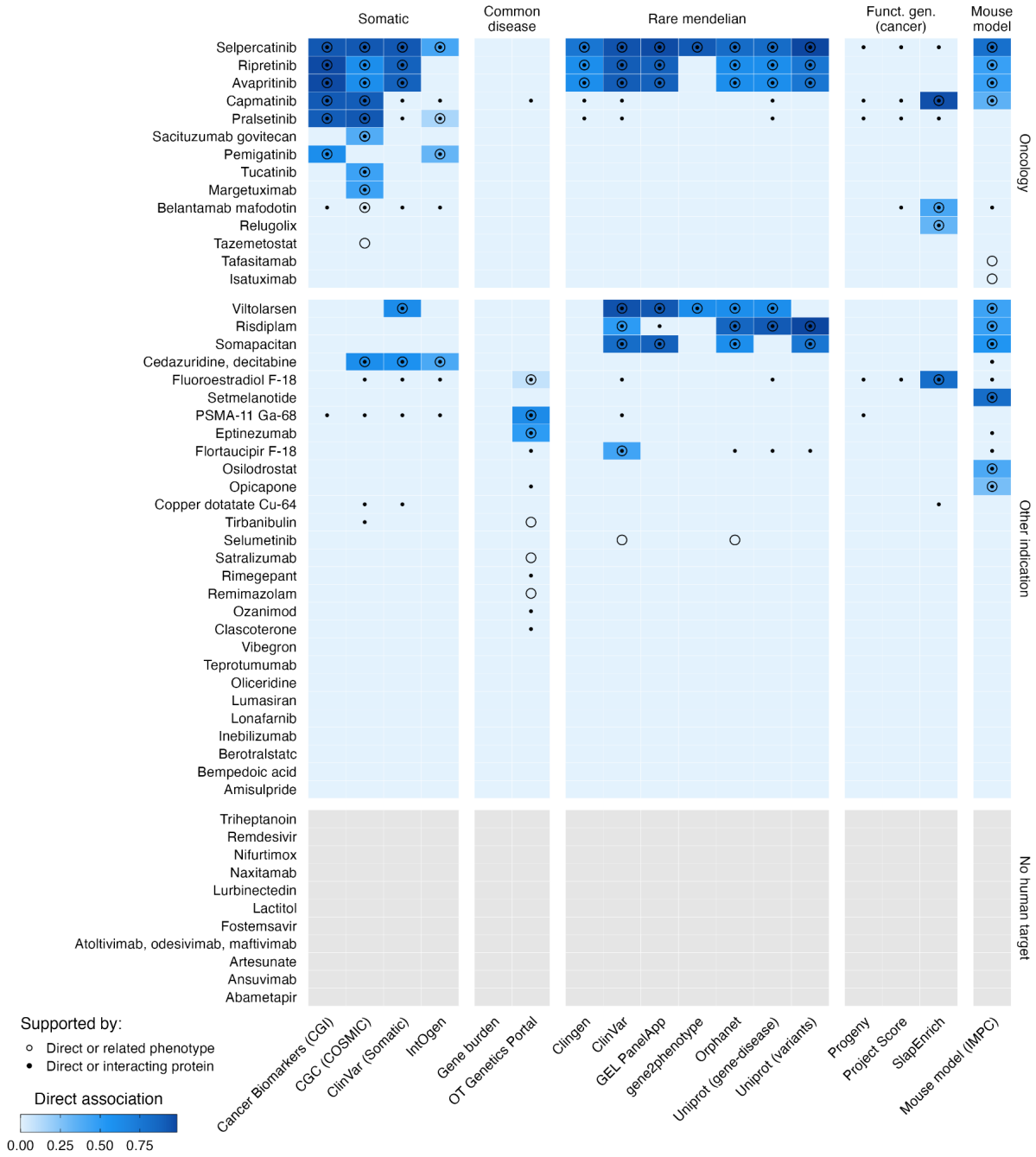
Supplementary Figure 8 | Supporting genetic evidence for 46 drugs approved by the FDA in 2017. Evidence for each drug target (T) – disease (D) indication is divided by oncology drugs (top panel), drugs for other indications with human targets (middle panel) and drugs without a human target (bottom panel). The x-axis is grouped by the predominant genetic data in the respective resources. Support by data source is displayed as: a) evidence for T and D or any other descendant in the EFO ontology aggregated using the Open Targets Platform association score (blue heatmap colours); b) evidence for T and D or closely related phenotypes to D added by manual curation (open circles); and/or c) evidence for D and T or any protein physically interacting with T (black dots).



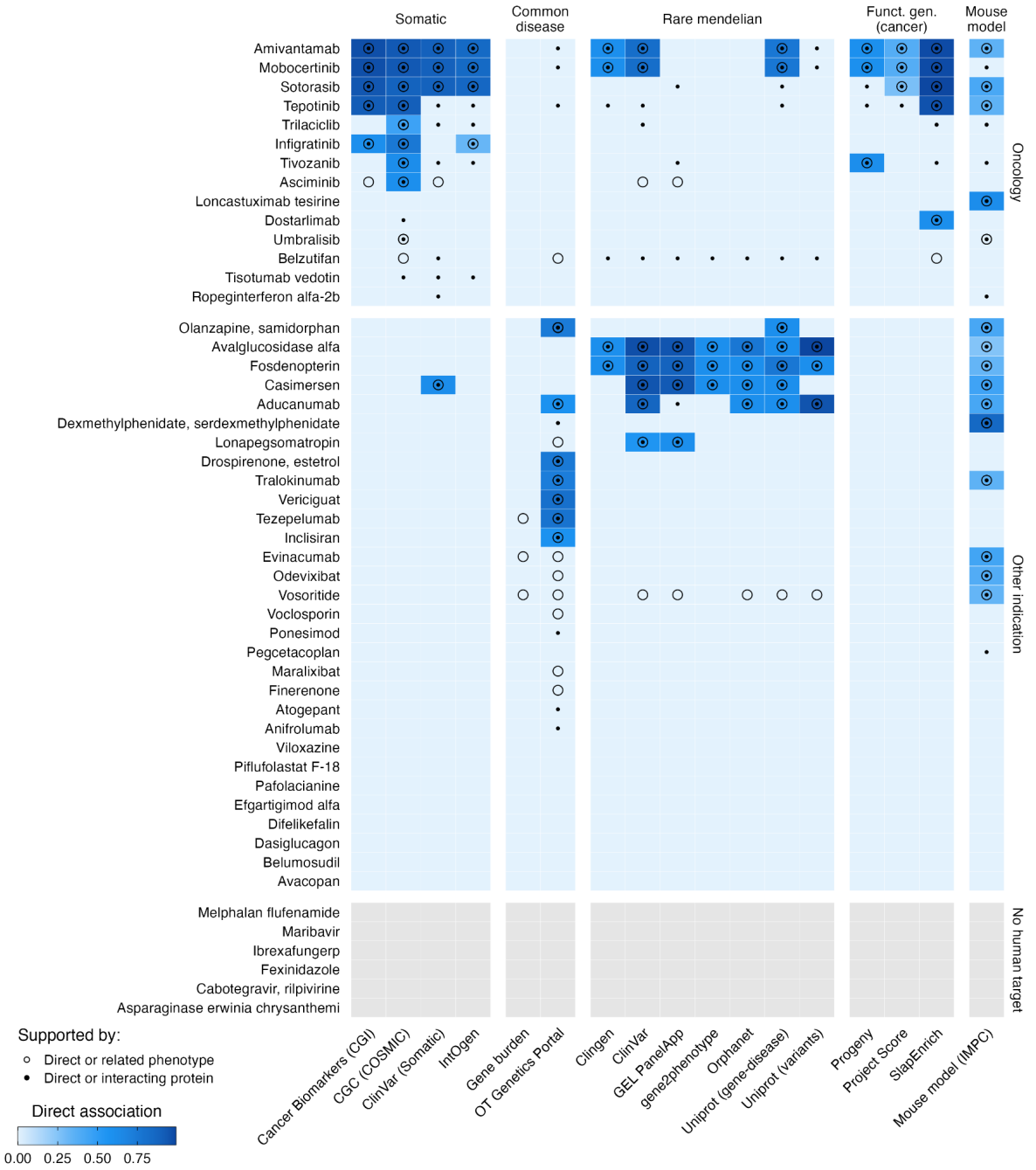
Supplementary Figure 9 | Supporting genetic evidence for 59 drugs approved by the FDA in 2018. Evidence for each drug target (T) – disease (D) indication is divided by oncology drugs (top panel), drugs for other indications with human targets (middle panel) and drugs without a human target (bottom panel). The x-axis is grouped by the predominant genetic data in the respective resources. Support by data source is displayed as: a) evidence for T and D or any other descendant in the EFO ontology aggregated using the Open Targets Platform association score (blue heatmap colours); b) evidence for T and D or closely related phenotypes to D added by manual curation (open circles); and/or c) evidence for D and T or any protein physically interacting with T (black dots).



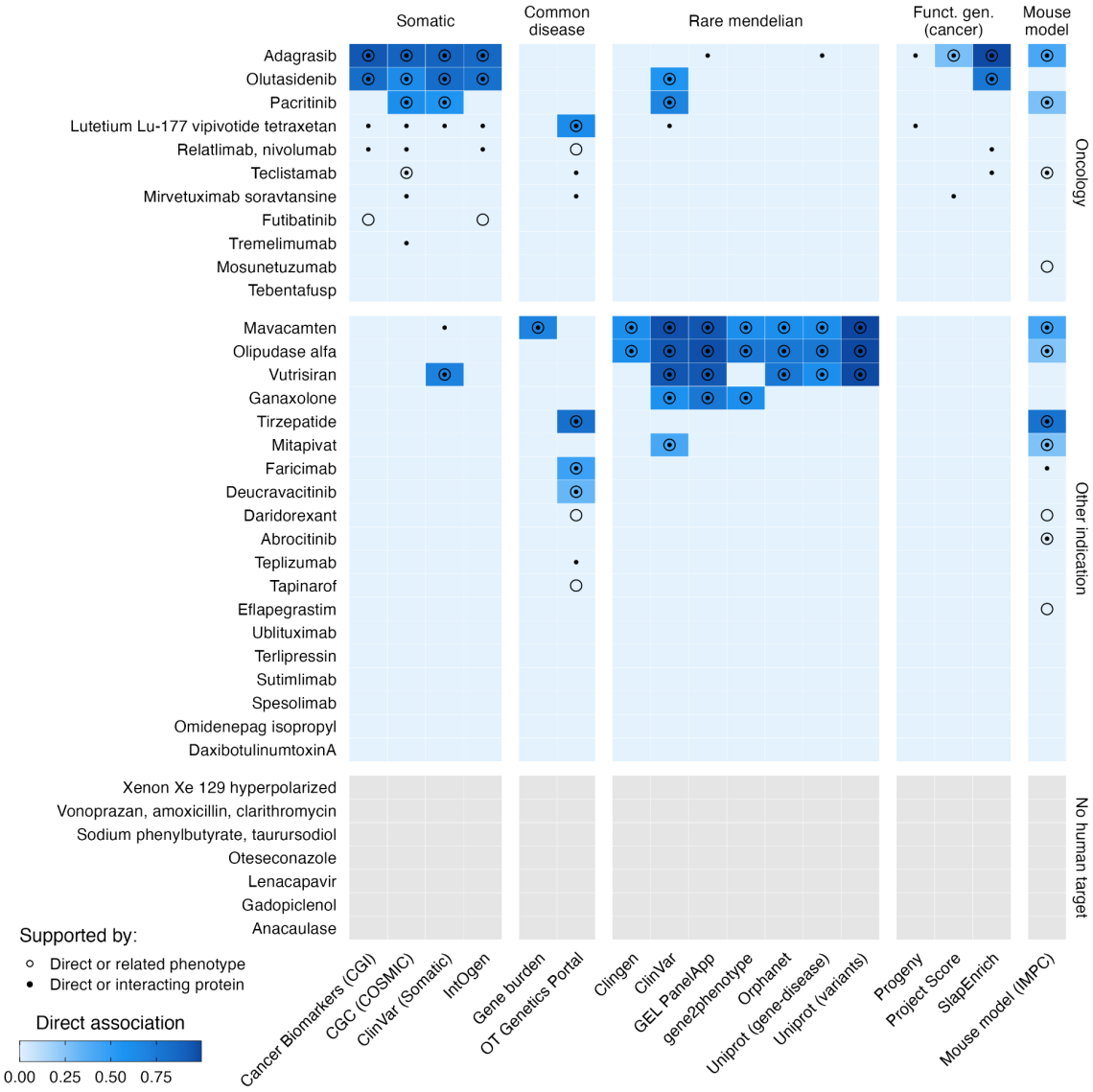
Supplementary Figure 10 | Supporting genetic evidence for 48 drugs approved by the FDA in 2019. Evidence for each drug target (T) – disease (D) indication is divided by oncology drugs (top panel), drugs for other indications with human targets (middle panel) and drugs without a human target (bottom panel). The x-axis is grouped by the predominant genetic data in the respective resources. Support by data source is displayed as: a) evidence for T and D or any other descendant in the EFO ontology aggregated using the Open Targets Platform association score (blue heatmap colours); b) evidence for T and D or closely related phenotypes to D added by manual curation (open circles); and/or c) evidence for D and T or any protein physically interacting with T (black dots).



Supplementary Figure 11 | Supporting genetic evidence for 53 drugs approved by the FDA in 2020. Evidence for each drug target (T) – disease (D) indication is divided by oncology drugs (top panel), drugs for other indications with human targets (middle panel) and drugs without a human target (bottom panel). The x-axis is grouped by the predominant genetic data in the respective resources. Support by data source is displayed as: a) evidence for T and D or any other descendant in the EFO ontology aggregated using the Open Targets Platform association score (blue heatmap colours); b) evidence for T and D or closely related phenotypes to D added by manual curation (open circles); and/or c) evidence for D and T or any protein physically interacting with T (black dots).



Supplementary Figure 12 | Supporting genetic evidence for 50 drugs approved by the FDA in 2021. Evidence for each drug target (T) – disease (D) indication is divided by oncology drugs (top panel), drugs for other indications with human targets (middle panel) and drugs without a human target (bottom panel). The x-axis is grouped by the predominant genetic data in the respective resources. Support by data source is displayed as: a) evidence for T and D or any other descendant in the EFO ontology aggregated using the Open Targets Platform association score (blue heatmap colours); b) evidence for T and D or closely related phenotypes to D added by manual curation (open circles); and/or c) evidence for D and T or any protein physically interacting with T (black dots).



Supplementary Figure 13 | Supporting genetic evidence for 37 drugs approved by the FDA in 2022. Evidence for each drug target (T) – disease (D) indication is divided by oncology drugs (top panel), drugs for other indications with human targets (middle panel) and drugs without a human target (bottom panel). The x-axis is grouped by the predominant genetic data in the respective resources. Support by data source is displayed as: a) evidence for T and D or any other descendant in the EFO ontology aggregated using the Open Targets Platform association score (blue heatmap colours); b) evidence for T and D or closely related phenotypes to D added by manual curation (open circles); and/or c) evidence for D and T or any protein physically interacting with T (black dots).

References

1. Mullard, A. 2013 FDA drug approvals. *Nat. Rev. Drug Discov.* **13**, 85–89 (2014).
2. Mullard, A. 2014 FDA drug approvals. *Nat. Rev. Drug Discov.* **14**, 77–81 (2015).
3. Mullard, A. 2015 FDA drug approvals. *Nat. Rev. Drug Discov.* **15**, 73–76 (2016).
4. Mullard, A. 2016 FDA drug approvals. *Nat. Rev. Drug Discov.* **16**, 73–76 (2017).
5. Mullard, A. 2017 FDA drug approvals. *Nat. Rev. Drug Discov.* **17**, 81–85 (2018).
6. Mullard, A. 2018 FDA drug approvals. *Nat. Rev. Drug Discov.* **18**, 85–89 (2019).
7. Mullard, A. 2019 FDA drug approvals. *Nat. Rev. Drug Discov.* **19**, 79–84 (2020).
8. Mullard, A. 2020 FDA drug approvals. *Nat. Rev. Drug Discov.* **20**, 85–90 (2021).
9. Mullard, A. 2022 FDA approvals. *Nat. Rev. Drug Discov.* **22**, 83–88 (2023).
10. Mullard, A. 2021 FDA approvals. *Nat. Rev. Drug Discov.* **21**, 83–88 (2022).
11. Ochoa, D. *et al.* Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs. *Nat. Rev. Drug Discov.* **21**, 551–551 (2022).
12. Gaulton, A. *et al.* The ChEMBL database in 2017. *Nucleic Acids Res.* **45**, D945–D954 (2017).
13. MacNamara, A. *et al.* Network and pathway expansion of genetic disease associations identifies successful drug targets. *Sci. Rep.* **10**, 1–11 (2020).
14. del Toro, N. *et al.* The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.* **50**, D648–D653 (2021).
15. Malone, J. *et al.* Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics* **26**, 1112–1118 (2010).
16. Vasilevsky, N. A. *et al.* Mondo: Unifying diseases for the world, by the world. *bioRxiv* (2022) doi:10.1101/2022.04.13.22273750.
17. Côté, R. G., Jones, P., Apweiler, R. & Hermjakob, H. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics* **7**, 97 (2006).
18. Ochoa, D. *et al.* The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res.* **51**, D1353–D1359 (2022).
19. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2018).
20. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* **10**, 1081–1082 (2013).
21. Tamborero, D. *et al.* Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.* **10**, (2018).
22. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062 (2018).
23. Dwane, L. *et al.* Project Score database: a resource for investigating cancer cell dependencies and prioritizing therapeutic targets. *Nucleic Acids Res.* **49**, D1365–D1372 (2020).
24. Iorio, F. *et al.* Pathway-based dissection of the genomic heterogeneity of cancer hallmarks' acquisition with SLAPenrich. *Sci. Rep.* **8**, 6713–6713 (2018).
25. Schubert, M. *et al.* Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **9**, 1–11 (2018).
26. Ghoussaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2020).
27. Rehm, H. L. *et al.* ClinGen — The Clinical Genome Resource. (2015) doi:10.1056/NEJMsr1406261.
28. Scaling national and international improvement in virtual gene panel curation via a collaborative approach to discordance resolution. *Am. J. Hum. Genet.* **108**, 1551–1557 (2021).
29. Thormann, A. *et al.* Flexible and scalable diagnostic filtering of genomic variants using G2P with Ensembl VEP. *Nat. Commun.* **10**, 1–10 (2019).
30. The UniProt Consortium *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2016).
31. Groza, T. *et al.* The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res.* **51**, D1038–D1045 (2022).