

The Druggable Genome as Seen from the Protein Data Bank

Jiayan Wang¹, Setayesh Yazdani¹, Ana Han¹, Matthieu Schapira^{1,2,*}

¹Structural Genomics Consortium, University of Toronto, MaRS South Tower, Suite 700, 101 College Street, Toronto, Ontario M5G 1L7, Canada.

²Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, M5S 1A8, Canada.

*matthieu.schapira@utoronto.ca

Abstract

Almost twenty years after the human genome was sequenced, the wealth of data produced by the international human genome project has not translated into a significantly improved drug discovery enterprise. This is in part because small molecule modulators that could be used to explore the cellular function of their target proteins and to discover new therapeutic opportunities are only available for a limited portion of the human proteome. International efforts are underway to develop such chemical tools for a few, specific protein families, and a “Target 2035” call to enable, expand and federate these efforts towards a comprehensive chemical coverage of the druggable genome was recently announced. But what is the druggable genome? Here, we systematically review structures of human proteins bound to drug-like ligands available from the protein databank (PDB) and use ligand desolvation upon binding as a druggability metric to draw a landscape of the human druggable genome. We show that the vast majority of druggable protein families, including some highly populated and deeply associated with cancer according to genomic screens, are almost orphan of small molecule ligands, and propose a list of 46 druggable domains representing 3440 human proteins that could be the focus of large chemical probe discovery efforts.

Introduction

Continuous advances in library screening technologies such as DNA-encoded libraries (Neri and Lerner, 2018), fragment screening (Erlanson et al., 2016) or free energy perturbation (Wang et al., 2015a) enable the discovery of drug-like small molecule ligands for an increasing number of protein targets. In parallel, the boundary of target druggability is constantly evolving, sometimes via a precise understanding of the structural dynamics of a ligand binding site (Kessler et al., 2019), or via the emergence of novel therapeutic modalities (Lai and Crews, 2017). These progresses and future technological advances are expected to enable a significantly expanded chemical coverage of the human proteome.

Born from both this favorable context and also the realization that chemical probes can play a decisive role in the discovery of novel therapeutic targets (Arrowsmith et al., 2015; Oprea et al., 2018), large-scale efforts are ongoing to increase the chemical coverage of specific protein families, such as protein kinases, solute carriers, GPCRs and epigenetic target classes (César-Razquin et al., 2015; Knapp et al., 2012; Roth et al., 2017; Scheer et al., 2019; Wu et al., 2019). Based on the progress with these efforts, a vision is emerging to create chemical ligands for the entire druggable genome by year 2035 (Carter et al., 2019). Delineating what is the druggable genome will help guide this ambitious enterprise (Finan et al., 2017; Hopkins and Groom, 2002; Nguyen et al., 2017; Oprea et al., 2018; Russ and Lampel, 2005; Santos et al., 2017).

Here, we systematically analyze protein domains that are in complex with drug-like ligands in the Protein Data Bank (PDB) to evaluate the druggability and structural coverage of target classes in the human genome. While restricted to protein families with bound ligands in the PDB, this approach, firmly anchored in experimental data, largely recapitulates previous classifications, but also reveals emerging target classes, some highly populated, that could be the focus of future chemical biology and drug discovery efforts.

The chemical coverage of the human proteome is poor

We first extracted all drug-like ligands from the PDB and identified the Interpro domain of the protein to which each ligand was bound. For each domain, we next counted the number of different proteins in which the domain was bound to a drug-like ligand in the PDB (Supplementary Tables 1 and 2). Unsurprisingly, we found that kinases are the most ligand-bound protein family (195 proteins bound to 2091 ligands in 2536 structures). Far behind were GPCRs (37 proteins bound to 67 ligands in 98 structures), followed by nuclear hormone receptors (25 proteins bound to 515 ligands in 686 structures) and bromodomains (22 proteins bound to 297 ligands in 365 structures). The chemical coverage (as seen from the PDB) of bromodomains is however greater than that of kinases: 54% of bromodomains, 52% of nuclear hormone receptors and 40% of kinases are bound to a drug-like ligand (Fig. 1). This number drops to 5% for GPCRs, but this reflects the difficulty to crystallize these proteins and their limited presence in the PDB rather than the extent of their chemical coverage. Similarly, ion channels structures typically are solved at resolutions above 3.0 Å, which is the cut-off for this analysis. As a result, this important class of drug targets (Hopkins and Groom, 2002; Oprea et al., 2018; Santos et al., 2017) is under-represented in our survey. Together, these data indicate that vast areas of target classes known to be druggable, such as kinases or nuclear hormone receptors remain underexplored. The fact that over 75% of protein families composed of 30 or more proteins have less than 10% of their members in complex with drug-like ligands in the PDB clearly illustrates the paucity of the chemical coverage of the human genome (Fig. 1, Supplementary Table 1).

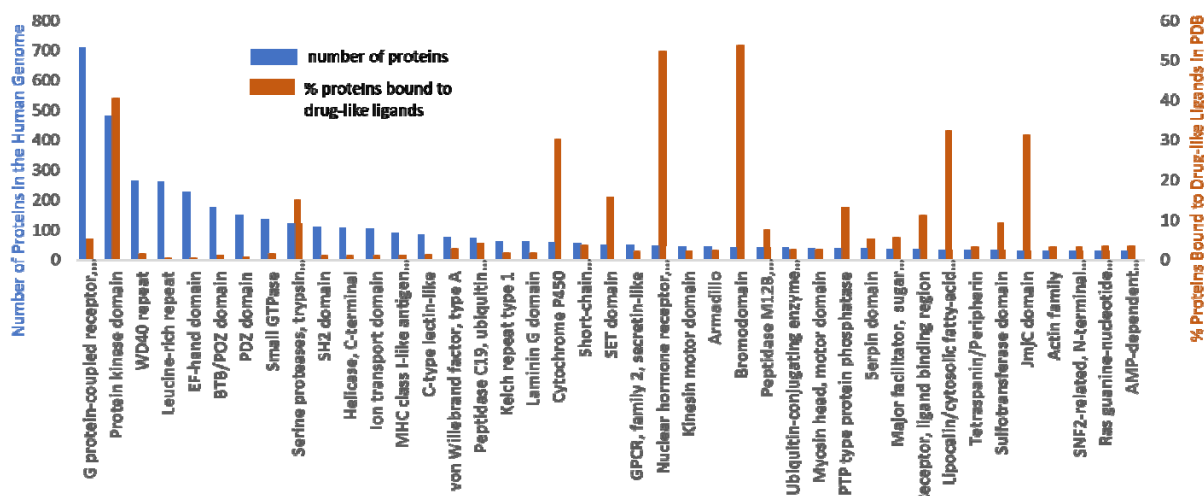


Figure 1: Chemical coverage in the PDB of the most frequent protein domains found in the human genome. Only protein domains with at least one structure in the PDB bound to a drug-like ligand with a resolution of 3Å or better are included in the analysis. Proteins domains found in less than 30 human proteins are not shown.

Some large and druggable protein families in the Protein Data Bank are almost entirely devoid of drug-like ligands

We note that some of the largest protein families have very few ligands bound. The WDR family is predicted to contain up to 610 proteins (Wang et al., 2015b), but only 4 WDR domains are bound to drug-like ligands in the PDB. Only 1 leucine-rich repeat (LRR) and 1 EF-hand containing protein are bound to a ligand, but each family includes over 200 genes. To evaluate whether lack of druggability is underlying this deficiency in chemical coverage, we estimated the druggability of protein families with two orthogonal approaches. In the first method, we calculated the percent surface of the ligand that is desolvated when bound to the protein. This value reflects the buriedness of the ligand in the pocket, and is expected to correlate with the druggability of the protein. The second method relies on the PDBbind database, which compiles the affinity of some of the ligands found in the PDB for their bound proteins (Wang et al., 2005). When multiple ligands were available for a protein, we used the strongest affinity and highest desolvation state, and the median value across all proteins was used for each domain. We purposely omitted the use of geometrical or electrostatic features of pockets to predict druggability, as is often done (Borrel et al., 2015; Halgren, 2009), including by us (Song et al.,

2017), because we set out to avoid any computational prediction in this analysis. We saw a correlation between the two metrics: protein families with high median ligand desolvation generally have low median ligand affinity values (i.e. measured binding is stronger). Small GTPases are an exception, as structures of KRAS bound to potent inhibitors were recently deposited in the PDB, leading to high desolvation values, but these ligands are not yet entered in the PDBbind database, where only weaker KRAS ligands are recorded (Supplementary Table 2).

We find that nuclear hormone receptors, cytochrome P450s and GPCRs are among the most druggable target classes, all having median ligand desolvation values of 95% or more, and affinity values (when available) below 100 nM (Fig. 2). LRR seem to position at the other end of the druggability spectrum, with a desolvation value of 57% (Supplementary Table 2). Though largely overlooked by the drug discovery and chemical biology community, the WDR family - one of the most populated in the human genome - seem eminently druggable, with a high median ligand desolvation of 77% and a low median affinity of 48 nM (Fig. 2, Supplementary Table 2). These data are derived though from only 4 targets (EED, WDR5, CDC20 and BTRC), and extrapolation to the entire protein family is uncertain, but it is nevertheless an encouraging indicator. WDR are β -propeller domains with a canonical toroidal shape (Schapira et al., 2017), and reported ligands exploit the central cavity of EED (He et al., 2017; Qi et al., 2017), WDR5 (Grebien et al., 2015) and BTRC (Simonetta et al., 2019) with low nanomolar potency, and the side wall of CDC20 (Sackton et al., 2014) with micromolar affinity.

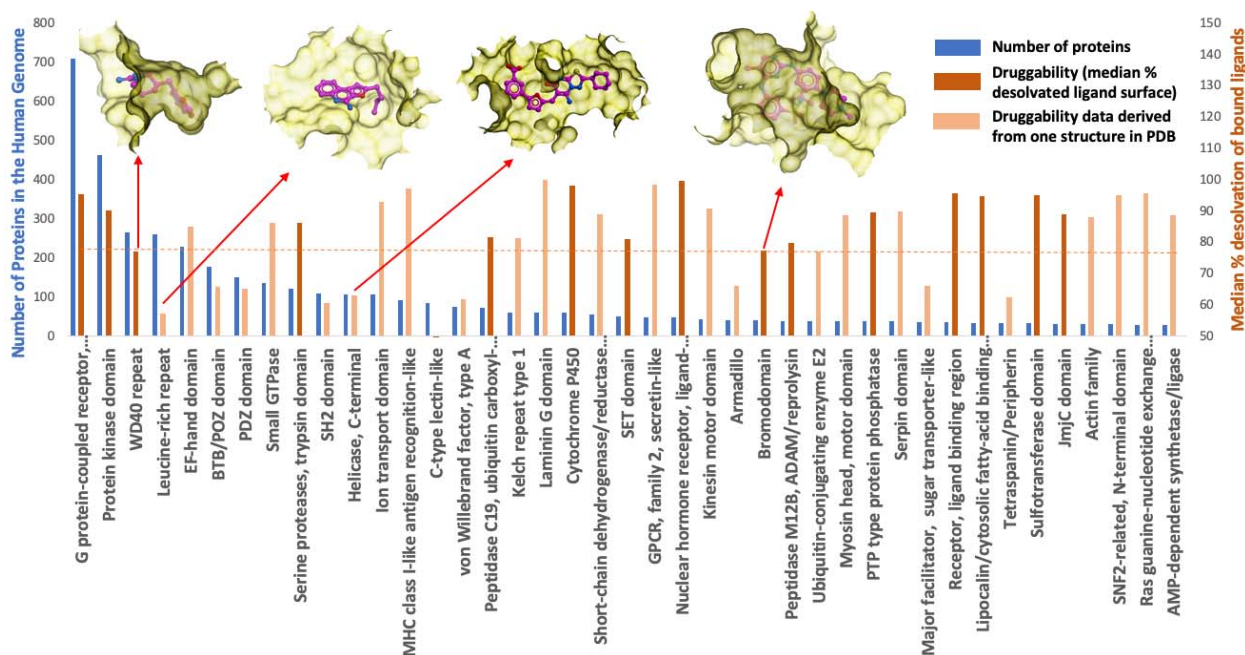


Figure 2: Druggability of target classes. For each protein family, the median desolvation value of bound ligands across all proteins in the PDB is used as an indicator of druggability (the most desolvated ligand is used for each protein). Dashed line: median desolvation of bromodomains. Faded orange color indicates families where the druggability data is based on less than three proteins (mean values are used instead of median). Crystal structures of ligand-bound pockets are shown for four representative domains: two with high ligand desolvation (WDR [PDB: 5u5t], Bromodomain [PDB: 4o76]) and two with low ligand desolvation (Leucine-rich repeat [PDB: 3wn4], helicase [PDB: 5urm]).

WDR proteins are a large, druggable and underexplored target class with strong disease association

To evaluate the disease association of severely underexplored but putatively druggable target classes, such as WDRs, and compare it with that of protein families more popular in drug discovery, we analyzed the essentiality of proteins across multiple cancer types based on CRISPR-knockout data available from the Broad and the Wellcome Sanger Institutes (Behan et al., 2019; Meyers et al., 2017). For each cancer type represented by at least 3 cell lines, the median essentiality score across all cell lines was determined for each gene, and the number of genes essential (essentiality score < -1.) in at least 3 cancer types was determined for each protein family (Fig. 3). Strikingly, we found that the WDR family is populated with the highest number of essential genes (Broad: 54 genes; Sanger: 64 genes), followed by helicases (Broad:

22 genes; Sanger: 32 genes) and protein kinases (Broad: 14 genes; Sanger: 20 genes). While helicases are low on our druggability ladder (the only ligand in the PDB is 63% desolvated when bound to its target, SNRNP200 [PDB code 5urm]), available data indicates that WDRs are druggable (see above), and the exceptional disease association of this target classes is a call for increased efforts to develop small-molecule WDR antagonists.

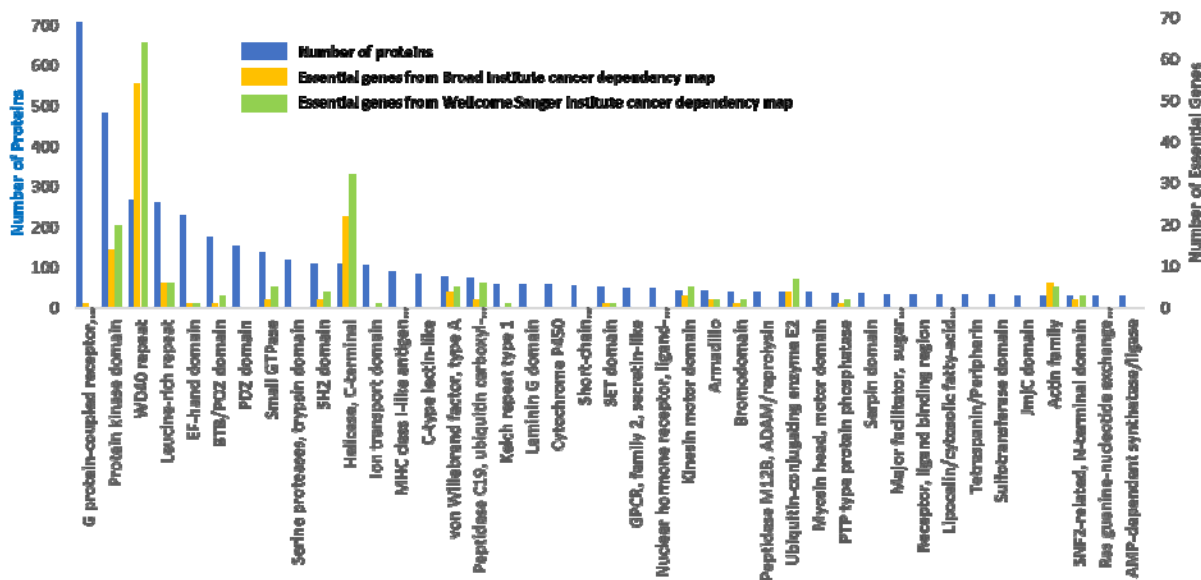


Figure 3: Essentiality of protein families in cancer. The number of proteins essential in at least 3 cancer types based on cancer dependency maps from the Broad and the Wellcome Sanger Institutes are indicated (details provided in the supplementary information).

Structural data currently in the PDB suggests that a large number of protein families are more druggable than bromodomains in the human genome

Overall, we find that 163 protein families have a median ligand desolvation level greater than that of bromodomains (desolvation value: 77%), a druggable target class (Supplementary Table 2). Of these, 46 are populated with 20 or more proteins, and could be the object of a focused medicinal chemistry effort where knowledge acquired and ligands discovered for one protein may guide and accelerate the design of ligands for another, as was done for kinases (Elkins et al., 2016) or bromodomains (Wu et al., 2019). We believe that these 46 protein families (Table 2) representing 3440 proteins (Supplementary Table 3) could be a good place to start a global

effort towards the discovery of chemical ligands for the druggable proteome (Carter et al., 2019).

Index	Interpro ID	Domain name	Proteins in the Human Genome	Proteins bound to approved drugs in the PDB	Targets in clinical trials	Essential genes from "[Broad] - [SANGER]" institutes cancer dependency maps	Median ligand desolvation	Median affinity [nM]	Druglike ligands in the PDB	Structures with bound ligands in the PDB	Proteins with bound ligands in the PDB	Proteins in the PDB
1	IPR000276	G protein-coupled receptor, rhodopsin-like	710	22	57	1-0	95	2.	67	98	37	39
2	IPR000719	Protein kinase domain	482	52	95	14-20	90	4	2091	2536	195	268
3	IPR001680	WD40 repeat	266	0	1	54-64	77	48	31	33	4	53
4	IPR002048	EF-hand domain	228	1	6	1-1	85	18.	3	7	1	54
5	IPR001806	Small GTPase	137	0	3	2-5	86	340000.	33	35	2	73
6	IPR001254	Serine proteases, trypsin domain	121	1	14	0-0	86	3	427	449	18	53
7	IPR005821	Ion transport domain	106	1	14	0-1	93	no data	3	3	1	2
8	IPR011161	MHC class I-like antigen recognition-like	92	0	0	0-0	97	no data	1	3	1	17
9	IPR001394	Peptidase C19, ubiquitin carboxyl-terminal hydrolase	74	1	0	2-6	82	65.	15	15	3	17
10	IPR001791	Laminin G domain	61	1	2	0-0	100	0.2	4	7	1	7
11	IPR006652	Kelch repeat type 1	61	0	0	0-1	81	1.3	13	14	1	9
12	IPR001128	Cytochrome P450	60	16	8	0-0	98	no data	65	80	18	24
13	IPR002347	Short-chain dehydrogenase/reductase SDR	55	0	4	0-0	89	no data	8	8	2	22
14	IPR001214	SET domain	51	0	2	1-1	81	27	21	27	8	29
15	IPR000832	GPCR, family 2, secretin-like	49	0	8	0-0	98	7	1	1	1	2
16	IPR000536	Nuclear hormone receptor, ligand-binding domain	48	19	6	0-0	100	1.	515	686	25	36
17	IPR001752	Kinesin motor domain	45	0	1	3-5	91	1.6	25	29	1	15
18	IPR001487	Bromodomain	41	2	5	1-2	77	43	297	365	22	34
19	IPR001590	Peptidase M12B, ADAM/reprolysin	40	0	3	0-0	80	5.	8	8	3	9
20	IPR000608	Ubiquitin-conjugating enzyme E2	40	0	1	4-7	77	1720.	1	1	1	32
21	IPR001609	Myosin head, motor domain	39	0	1	0-0	89	290.	1	2	1	5
22	IPR023796	Serpin domain	38	1	4	0-0	90	290.	2	3	2	21
23	IPR000242	PTP type protein phosphatase	38	0	4	1-2	90	166.	66	81	5	30
24	IPR001828	Receptor, ligand binding region	36	0	5	0-0	96	81	5	6	4	13
25	IPR000566	Lipocalin/cytosolic fatty-acid binding domain	34	6	0	0-0	95	29.	35	109	11	24
26	IPR000863	Sulfotransferase domain	33	3	0	0-0	95	83200.	4	5	3	15
27	IPR000330	SNF2-related, N-terminal domain	32	0	0	2-3	95	no data	2	2	1	1
28	IPR003347	JmjC domain	32	0	0	0-0	89	60.	39	48	10	24
29	IPR004000	Actin family	32	0	0	6-5	88	no data	2	29	1	4
30	IPR001895	Ras guanine-nucleotide exchange factors catalytic domain	30	0	0	0-0	96	322000.	31	31	1	7
31	IPR000873	AMP-dependent synthetase/ligase	30	1	0	0-0	89	no data	1	1	1	1
32	IPR000073	Alpha/beta hydrolase fold-1	28	0	3	0-0	97	0.7	46	47	1	6
33	IPR003594	Histidine kinase/HSP90-like ATPase	28	0	3	1-2	84	229	180	205	6	9
34	IPR001251	CRAL-TRIO lipid binding domain	27	1	0	0-0	100	10.	3	9	3	7
35	IPR000086	NUDX hydrolase domain	27	1	0	1-2	95	0.4	29	30	2	19
36	IPR004045	Glutathione S-transferase, N-terminal	27	3	1	1-0	89	300.	27	33	7	19
37	IPR005135	Endonuclease/exonuclease/phosphatase	25	0	2	0-0	81	17	1	1	1	11
38	IPR003008	Tubulin/FtsZ, GTPase domain	24	0	1	2-5	98	no data	1	1	1	4
39	IPR004046	Glutathione S-transferase, C-terminal	24	1	2	4-3	89	1950.	18	19	5	18
40	IPR002130	Cyclophilin-type peptidyl-prolyl cis-trans isomerase domain	24	1	0	2-6	79	1600.	18	19	2	14
41	IPR000834	Peptidase M14, carboxypeptidase A	23	0	1	0-0	98	8.	1	1	1	8
42	IPR001818	Peptidase M10, metallopeptidase	23	0	5	0-0	90	1.7	65	73	8	11
43	IPR000313	PWWP domain	23	0	1	0-0	78	no data	3	3	1	13
44	IPR017978	GPCR family 3, C-terminal	22	1	3	0-0	97	no data	2	1	1	1
45	IPR002073	3'S-cyclic nucleotide phosphodiesterase, catalytic domain	21	5	4	0-0	91	0.5	170	202	11	12
46	IPR005123	Oxoglutarate/iron-dependent dioxygenase	21	0	3	0-0	90	16	9	22	1	7

Table 2: Human druggable genome, represented by protein families of 20 or more proteins, with at least one protein bound to a drug-like inhibitor in the PDB (with a resolution better than 3 Å), and median ligand desolvation upon binding equal or better than that of bromodomains.

Conclusion

Motivated by the principle that experimental structures of drug-like ligands in complex with their receptor proteins are among the most reliable data linking ligands and protein domains, we provide here an analysis of the druggable genome as seen from the PDB. We find that a significant portion of well-known druggable target classes remain underexplored, and that

other druggable protein families, some large and with strong disease association, are almost entirely orphan of inhibitors. The list of 3440 proteins from 46 druggable target classes compiled here may help strategize emerging efforts to systematically develop chemical reagents for the human proteome (Carter et al., 2019).

Acknowledgments:

The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada through Ontario Genomics Institute [OGI-055], Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no. 115766], Janssen, Merck KGaA, Darmstadt, Germany, MSD, Novartis Pharma AG, Innovation and Science (MRIS), Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and Wellcome. M.S. gratefully acknowledges support from NSERC [Grant RGPIN-2019-04416]. We thank Aled Edwards for comments on the manuscript.

References:

- Arrowsmith, C.H., Audia, J.E., Austin, C., Baell, J., Bennett, J., Blagg, J., Bountra, C., Brennan, P.E., Brown, P.J., Bunnage, M.E., et al. (2015). The promise and peril of chemical probes. *Nat. Chem. Biol.* *11*, 536–541.
- Behan, F.M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C.M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., et al. (2019). Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* *568*, 511–516.
- Borrel, A., Regad, L., Xhaard, H., Petitjean, M., and Camproux, A.-C. (2015). PockDrug: A Model for Predicting Pocket Druggability That Overcomes Pocket Estimation Uncertainties. *J Chem Inf Model* *55*, 882–895.
- Carter, A.J., Kraemer, O., Zwick, M., Mueller-Fahrnow, A., Arrowsmith, C.H., and Edwards, A.M. (2019). Target 2035: probing the human proteome. *Drug Discov. Today*.
- César-Razquin, A., Snijder, B., Frappier-Brinton, T., Isserlin, R., Gyimesi, G., Bai, X., Reithmeier, R.A., Hepworth, D., Hediger, M.A., Edwards, A.M., et al. (2015). A Call for Systematic Research on Solute Carriers. *Cell* *162*, 478–487.

Elkins, J.M., Fedele, V., Szklarz, M., Abdul Azeez, K.R., Salah, E., Mikolajczyk, J., Romanov, S., Sepetov, N., Huang, X.-P., Roth, B.L., et al. (2016). Comprehensive characterization of the Published Kinase Inhibitor Set. *Nat. Biotechnol.* *34*, 95–103.

Erlanson, D.A., Fesik, S.W., Hubbard, R.E., Jahnke, W., and Jhoti, H. (2016). Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov* *15*, 605–619.

Finan, C., Gaulton, A., Kruger, F.A., Lumbers, R.T., Shah, T., Engmann, J., Galver, L., Kelley, R., Karlsson, A., Santos, R., et al. (2017). The druggable genome and support for target identification and validation in drug development. *Sci Transl Med* *9*.

Grebien, F., Vedadi, M., Getlik, M., Giambruno, R., Grover, A., Avellino, R., Skucha, A., Vittori, S., Kuznetsova, E., Smil, D., et al. (2015). Pharmacological targeting of the Wdr5-MLL interaction in C/EBP α N-terminal leukemia. *Nat. Chem. Biol.* *11*, 571–578.

Halgren, T.A. (2009). Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* *49*, 377–389.

He, Y., Selvaraju, S., Curtin, M.L., Jakob, C.G., Zhu, H., Comess, K.M., Shaw, B., The, J., Lima-Fernandes, E., Szewczyk, M.M., et al. (2017). The EED protein–protein interaction inhibitor A-395 inactivates the PRC2 complex. *Nature Chemical Biology* *13*, 389–395.

Hopkins, A.L., and Groom, C.R. (2002). The druggable genome. *Nat Rev Drug Discov* *1*, 727–730.

Kessler, D., Gmachl, M., Mantoulidis, A., Martin, L.J., Zoepfel, A., Mayer, M., Gollner, A., Covini, D., Fischer, S., Gerstberger, T., et al. (2019). Drugging an undruggable pocket on KRAS. *Proc. Natl. Acad. Sci. U.S.A.* *116*, 15823–15829.

Knapp, S., Arruda, P., Blagg, J., Burley, S., Drewry, D.H., Edwards, A., Fabbro, D., Gillespie, P., Gray, N.S., Kuster, B., et al. (2012). A public-private partnership to unlock the untargeted kinome. *Nature Chemical Biology* *9*, 3–6.

Lai, A.C., and Crews, C.M. (2017). Induced protein degradation: an emerging drug discovery paradigm. *Nature Reviews Drug Discovery* *16*, 101–114.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nature Genetics* *49*, 1779–1784.

Neri, D., and Lerner, R.A. (2018). DNA-Encoded Chemical Libraries: A Selection System Based on Endowing Organic Compounds with Amplifiable Information. *Annu. Rev. Biochem.* *87*, 479–502.

Nguyen, D.-T., Mathias, S., Bologna, C., Brunak, S., Fernandez, N., Gaulton, A., Hersey, A., Holmes, J., Jensen, L.J., Karlsson, A., et al. (2017). Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Res.* *45*, D995–D1002.

Oprea, T.I., Bologa, C.G., Brunak, S., Campbell, A., Gan, G.N., Gaulton, A., Gomez, S.M., Guha, R., Hersey, A., Holmes, J., et al. (2018). Unexplored therapeutic opportunities in the human genome. *Nature Reviews Drug Discovery* *17*, 317–332.

Qi, W., Zhao, K., Gu, J., Huang, Y., Wang, Y., Zhang, H., Zhang, M., Zhang, J., Yu, Z., Li, L., et al. (2017). An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED. *Nature Chemical Biology* *13*, 381–388.

Roth, B.L., Irwin, J.J., and Shoichet, B.K. (2017). Discovery of new GPCR ligands to illuminate new biology. *Nature Chemical Biology* *13*, 1143–1151.

Russ, A.P., and Lampel, S. (2005). The druggable genome: an update. *Drug Discov. Today* *10*, 1607–1610.

Sackton, K.L., Dimova, N., Zeng, X., Tian, W., Zhang, M., Sackton, T.B., Meaders, J., Pfaff, K.L., Sigoillot, F., Yu, H., et al. (2014). Synergistic blockade of mitotic exit by two chemical inhibitors of the APC/C. *Nature* *514*, 646–649.

Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I., et al. (2017). A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* *16*, 19–34.

Schapira, M., Tyers, M., Torrent, M., and Arrowsmith, C.H. (2017). WD40 repeat domain proteins: a novel target class? *Nature Reviews Drug Discovery* *16*, 773–786.

Scheer, S., Ackloo, S., Medina, T.S., Schapira, M., Li, F., Ward, J.A., Lewis, A.M., Northrop, J.P., Richardson, P.L., Kaniskan, H.Ü., et al. (2019). A chemical biology toolbox to study protein methyltransferases and epigenetic signaling. *Nature Communications* *10*, 19.

Simonetta, K.R., Taygerly, J., Boyle, K., Basham, S.E., Padovani, C., Lou, Y., Cummins, T.J., Yung, S.L., von Soly, S.K., Kayser, F., et al. (2019). Prospective discovery of small molecule enhancers of an E3 ligase-substrate interaction. *Nat Commun* *10*, 1402.

Song, R., Wang, Z.-D., and Schapira, M. (2017). Disease Association and Druggability of WD40 Repeat Proteins. *J. Proteome Res.* *16*, 3766–3773.

Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M.K., Greenwood, J., et al. (2015a). Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* *137*, 2695–2703.

Wang, R., Fang, X., Lu, Y., Yang, C.-Y., and Wang, S. (2005). The PDBbind database: methodologies and updates. *J. Med. Chem.* *48*, 4111–4119.

Wang, Y., Hu, X.-J., Zou, X.-D., Wu, X.-H., Ye, Z.-Q., and Wu, Y.-D. (2015b). WDSpdb: a database for WD40-repeat proteins. *Nucleic Acids Res.* *43*, D339–344.

Wu, Q., Heidenreich, D., Zhou, S., Ackloo, S., Krämer, A., Nakka, K., Lima-Fernandes, E., Deblois, G., Duan, S., Vellanki, R.N., et al. (2019). A chemical toolbox for the study of bromodomains and epigenetic signaling. *Nat Commun* *10*, 1915.