ELSEVIER

KEYNOTE (GREEN)

The European Lead Factory: Results from a decade of collaborative, public–private, drug discovery programs

Herman van Vlijmen [1,*], Andrew D. Pannifer [2,†], Phil Cochrane [3,‡], Daniel Basting [4], Volkhart M Li [4], Ola Engkvist [5], Jean-Yves Ortholand [6], Markus Wagener [7], James Duffy [8], Dirk Finsinger [9], Jeremy Davis [10], Steven P. van Helden [11,§], Jon S.B. de Vlieger [12] European Lead Factory Consortium[*]

[1] Janssen Research & Development, Turnhoutseweg 30, 2340 Beerse, Belgium
[2] Biocity Scotland, Bo'ness Road, Newhouse ML1 5SH, UK
[3] Unit 8, Dock Offices Surrey Quays Rd, SE16 2XU, UK
[4] Bayer AG, Aprather Weg 18a, 42113 Wuppertal, Germany
[5] AstraZeneca Discovery Sciences, Pepparedsleden 1, SE-431 83 Mölndal, Sweden
[6] Edelris, 60, Avenue Rockefeller, F-69008 Lyon, France
[7] Grünenthal, 52099 Aachen, Germany
[8] Medicines for Malaria Venture, ICC 20, Rte de Pré-Bois, 1215 Geneva, Switzerland
[9] Merck KGaA, Frankfurter Str. 250, 64293 Darmstadt, Germany
[10] UCB Biopharma UK, 216 Bath Road, Slough, SL1 3WE, UK
[11] Pivot Park Screening Centre, Kloosterstraat 9, 5349 AB Oss, the Netherlands
[12] Lygature, Jaarbeursplein 6, 3521 AL Utrecht, the Netherlands

Herman van Vlijmen graduated with a Master's degree in biopharmaceutical sciences from Leiden University and received his PhD in physical chemistry from Harvard University. He worked for 9 years at the biotech company Biogen in Boston in the computational design of small molecule and protein therapeutics. In 2005, he joined Janssen in Belgium, where he is now global head of computer-aided drug design. Since 2008, he has also been adjunct professor of computational drug discovery at Leiden University.

The European Lead Factory (ELF) is a consortium of universities and small and medium-sized enterprises (SMEs) dedicated to drug discovery, and the pharmaceutical industry. This unprecedented consortium provides high-throughput screening, triage, and hit validation, including to non-consortium members. The ELF library was created through a novel compound-sharing model between nine pharmaceutical companies and expanded through library synthesis by chemistry-specialized SMEs. The library has been screened against ~270 different targets and 15 phenotypic assays, and hits have been developed to form the basis of patents and spin-off companies. Here, we review the outcome of screening campaigns of the ELF, including the performance and physicochemical properties of the library, identification of possible frequent hitter compounds, and the effectiveness of the compound-sharing model.

Keywords: Public-private partnership; High-throughput screening; Drug discovery; Compuond sharing

* *Corresponding author. van Vlijmen, H. (hvvlijme@its.jnj.com)*
† Current address: Exscientia Ltd, Oxford Science Park, Oxford, OX4 4GE, UK.
‡ Current address: Dassault Systemes UK Ltd, Cambridge Science Park, Milton Road, Cambridge, CB4 0FJ, UK.
§ Current address: Wega Informatik, St. Alban-Graben, 34051 Basel, Switzerland.
* A list of the members and their role is presented at the end of the paper.

Introduction

Translating publicly funded biomedical research into therapies is one of the key challenges for health policymakers and researchers. In 2012, public spending on biomedical research in Europe amounted to >US$28 billion[(p1)] and represented a significant investment of public resources. Identifying novel biological targets for drug discovery programs is an important part of this research, but the translational barriers to moving an interesting new target discovered in an academic setting to a full drug discovery program are high. The confidence of commercial organizations in early target validation is often insufficient to justify a high level of investment, while the integrated drug discovery infrastructure and expertise necessary to identify authentic, specific inhibitors (thus providing this confidence) are not widespread in academic environments because of their cost.[(p2),(p3)]

The ELF is a 30-member public–private partnership between the Innovative Medicines Initiative (IMI) of the European Union and the European Federation of Pharmaceutical Industries and Associations (EFPIA) established with the goal of lowering these barriers.[(p4),(p5)] ELF started its operations in 2011 and continued in a follow-up IMI project called ESCulab, allowing continued access to ELF infrastructure and capabilities for the wider scientific community.[(p5),(p6),(p7)] The ELF selects novel therapeutic targets through a competitive proposal system from academic groups and SMEs across Europe, carries out a high-throughput screen to identify hits, then validates and characterizes the hits to build confidence that they are developable and acting through a specific mechanism. Robust hits are taken forward to analog synthesis to explore the structure–activity relationships (SARs) and generate a data package that will attract the interest of other funding bodies or commercial organizations. The ELF has two parts: (i) a high-throughput, high-content screening and hit characterization lab (BioAscent, Pivot Park Screening Centre, University of Dundee, University of Oxford), compound storage and logistics (BioAscent), as well as expert capabilities in medicinal chemistry, structural biology, and computational chemistry; and (ii) a network of academic groups and chemistry SMEs designing and synthesizing novel compounds to expand the screening library. The ELF Compound Library[(p6)] was established with a Pharma set of 327 000 compounds contributed by the nine pharmaceutical company members of the consortium (Bayer AG, AstraZeneca AB, Janssen Pharmaceutica NV, H. Lundbeck A/S, Merck KGaA, Sanofi S.A, UCB, Servier, and Grünenthal GmbH), with each company contributing between 25 000 and 55 000 noncommercially available compounds. These compounds were selected within each company and brought together into a single collection (following deduplication) and stored in a central facility (BioAscent). The SMEs (Edelris, Symeres, Sygnature Discovery, and Taros) had the, ultimately successful, objective of synthesizing and adding a further 190 000 'public' compounds to the ELF Library (Public Compound Collection: PCC).[(p8)] Ideas for these libraries were crowdsourced from chemists across Europe proposing ideas through a web interface, and proposals were reviewed by an expert group of medicinal chemists from industry and academia. Successful proposals were synthesized by the network of chemistry SMEs and added to the library. These compounds have been designed to explore complementary chemical space to the Pharma set with emphasis on novel, sp^3-rich scaffolds and physicochemical (physchem) properties in line with the state of the art for compound-screening libraries.

Since the start of the ELF, 108 targets from academia and SMEs have been screened. In addition to these, the pharmaceutical companies contributing compounds to the ELF Library are also entitled to screen their targets against the ELF Library at their own sites, and 165 screens have been performed to date by these industry partners. An EFPIA-associated partner, the Medicines for Malaria Venture (MMV), was also able to screen several of their targets. Single-point, dose–response, and any other biological data from the screens are deposited in the Honest Data Broker (HDB),[(p9)] a central secure database and set of hit triaging tools hosted on the Biovia ScienceCloud platform, where compound information is also stored. The screening output is then triaged (chemical structures remain blinded during part of the process) in a controlled series of stages in which only compounds of highest interest to the program and increasing confidence in authentic bioactivity are carried forward, until a final output of up to 50 validated compounds is generated, the chemical structures of which are then unblinded to the target owner. To date, 100 million single-point primary assay values have been generated, together with results from downstream deselection data, orthogonal data, dose–response data and liquid chromatography-mass spectrometry (LC-MS) results. These results have created a data set comprising diverse biological targets, assay methodologies, and readout technologies.

Exploiting and analyzing these data needs to be carried out within the intellectual property (IP) agreement of the ELF. The fundamental aim of the ELF is to move academic research into drug discovery and safeguarding the IP of the participants is necessary to make the output of the ELF attractive to future funding partners. As a result, chemical structure information and knowledge of precise biological targets that have been screened are restricted.

However, assessment of the ELF Library performance and the progression of programs are key outcomes of the ELF. Thus, balancing these competing aims was an important consideration at the outset of the project and a system was put in place to enable this. Biological targets are registered within aggregated target classes (e.g., metalloproteases, serine proteases, or Ser/Thr kinases) to allow granularity in understanding compound bioactivity. Meanwhile, compounds registered in the ELF database are processed to calculate a range of physchem properties and other molecular descriptors, such as circular fingerprints and molecular scaffolds. Thus, the database provides a rich source of target class bioactivity, which can be viewed through a range of physchem properties and structural descriptors to enable decision-making and retrospective analysis. Here, we use this information to analyze the results of screening campaigns conducted during ELF 2013–2022, comparing and contrasting the results from the Pharma and PCC subsets. We describe biological target classes addressed by the collection, identify frequently hitting compounds, and track physchem properties of compounds as programs progress along the triage Box 1.

Box 1 Key points.

More than 250 high-throughput screens on a wide variety of targets have been completed by academic groups, small biotech, nonprofits, and large pharma on a shared collection of half a million compounds.

An effective mechanism was set up for sharing proprietary screening compounds with full enablement of IP creation.

Screening hits from the ELF have led to multiple active compound optimization programs at participating academic and industry organizations, and two companies have been created on the basis of hits from the ELF.

Analysis of the screening results indicate that compound promiscuity is correlated with lipophilicity and anti-correlated with the three-dimensionality of compounds.

There is no indication of the existence of dark chemical matter in this collection, as we continuously observed previously inactive compounds to be active in new screens.

ELF Library profile

The ELF Library is built from two complementary sets of compounds: (i) the Pharma 327 000 set contributed by the pharmaceutical company partners; and (ii), a set of 190 000 novel compounds synthesized within the ELF consortium (the PCC) through large-scale library synthesis. The Pharma set has been described elsewhere[(p6)] and provides a foundation for the ELF Library, balanced between diversity and having near neighbors to drive clustering and machine learning. The overarching aim of the PCC set of compounds is to explore chemical space complementary to that occupied by the Pharma set, while remaining within drug-like physchem properties and using tractable chemistry that can be followed up in a drug discovery project. The physchem properties of the two sets are compared in Figure 1.

Overall, the compounds in the PCC are larger than those in the Pharma set, but have a similar AlogP distribution, which is due to considerable effort during the library design to control lipophilicity. The most striking divergence between the two sets is the fraction of sp^3 hybridized carbon atoms (Fsp^3), which is significantly higher in the PCC, reflecting the selection process for the crowdsourced proposals.[(p10),(p11)] The higher Fsp^3 of the PCC translates to a higher three-dimensionality in principal moment of inertia (PMI) plots.[(p12)] Figure 1g–h shows the PMI plots of the Pharma and PCC sets in a heatmap representation. The areas of the PMI plot most highly populated (in red) are displaced away from the left-hand axis and toward molecules with a more 3D or spherical shape for the PCC. The Pharma set is richer in molecules along the linear/disc-shaped axis on the left side of the plot. This is also shown in Figure 1f, which takes slices parallel to the left-hand axis where the linear/planar molecules are located. The Pharma set is strongly biased to this 2D region of the map close to the left axis, whereas the PCC has more spherical character and the collection extends toward the center of the plot. The reasons for largely planar, sp2-rich compounds in pharmaceutical collections have been discussed extensively elsewhere, together with the advantages and disadvantages of using sp^3-rich compounds.[(p13),(p14),(p15)]

The major part of the Pharma set has been screened against all targets submitted to the ELF, whereas new PCC compounds are distributed to the screening centers at the start of each year and, hence, have been added on a rolling basis to the current ELF Library. They are then screened against any targets subsequently accepted through the competitive application process and against targets chosen by the pharmaceutical company partners for screening at their own sites. Thus, the ELF generates a substantial set of screening data from multiple screening sites with a variety of screening technologies and against a large range of target types.

ELF Library screening data analysis

The ELF triage process operates through a well-defined series of stages (Figure 2), designed to enable effective decision making in the triage but to minimize the amount of structural information disclosed about the library compounds to protect IP. We used these stages to track the progress of compounds through the triage and the stages are briefly described here.

The Screening Result List (SRL) represents the normalized single-point data from the high-throughput screen. Compound activity on the SRL is reported as the Z-Score (i.e., the number of standard deviations from 0% activity control). Compounds showing modulation of activity more than four standard deviations from the 0% activity controls are classified as active. DMSO was used as a negative control and positive controls were used when available. Triaging for academic and SME-submitted screens are carried out by a special team in the ESCulab project, whereas pharma-submitted screens are triaged by the companies themselves through external consultants. Triage scientists can analyze the SRL using HDB cheminformatics tools, incorporate further biological data from dose–response, deselection, orthogonal, and confirmatory assays, and assess the selectivity of the compounds by investigating their activity against other biological target classes (the classes, not the actual targets). The number of compounds that can be tested in these additional experiments is limited to a maximum of 1% of the number of compounds screened (~5000 compounds). At this stage, no compound structures are visible. The primary aim of this limitation is to prevent large-scale analysis of other partners' subsets and was necessary to allow the companies to participate in the ELF and share significant proportions of their libraries. It also has the effect of reducing selection bias toward particular chemotypes early in the triage process.

Effective triage is enabled by a comprehensive suite of cheminformatics tools in the HDB, including clustering, multiple similarity metrics, multiple calculated physchem properties, desirability scoring (e.g., the QED score[(p16)]), and the ability to see the activity of compounds across target classes, readout modalities, and assay technologies. Compounds showing interesting properties on the SRL are selected for the next stage in the triage: the Provisional Hit List (PHL).

The PHL is an intermediate list of typically several hundred to a thousand compounds reserved temporarily for the program, and the compound structures on this list are now made visible to the triage scientists. This enables a 'chemist's eye' inspection of the compounds and inclusion of LC-MS data to verify the

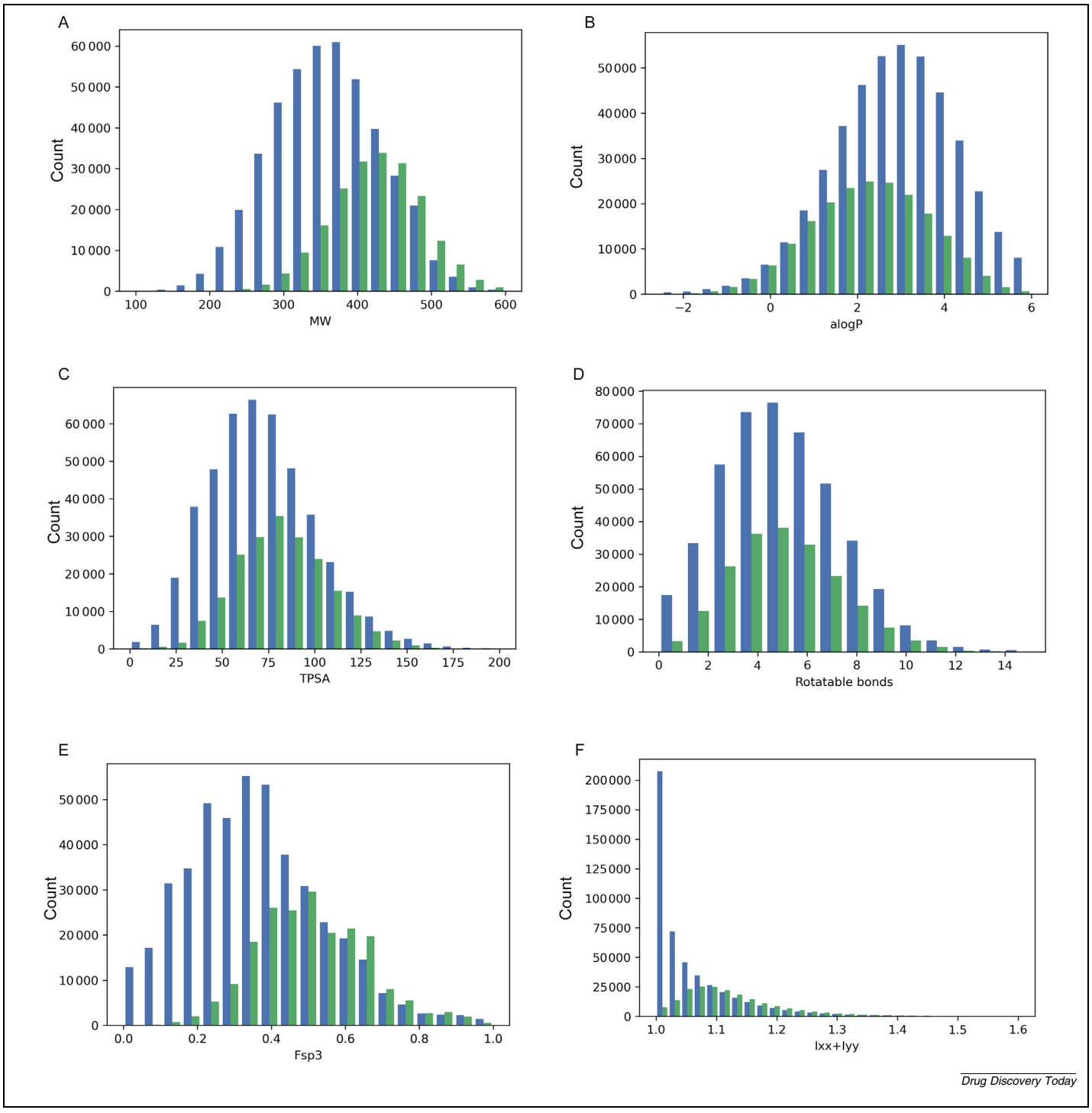

FIGURE 1

Physicochemical properties of the Joint European Compound Library (JECL). In (a–f), Pharma compounds are shown in blue and those from the Public Compound Collection (PCC) are shown in green. Molecular weight (MW) is given in Da and topological polar surface area (TPSA) in Å^2. Fsp^3 is the fraction of sp^3-hybridized carbon atoms in a molecule. In (f), Ixx + Iyy is the sum of the moments of inertia from a principal moment of inertia (PMI) plot, with each bar corresponding to a slice of the plot parallel to the left-hand axis of the PMI plot triangle. Linear and planar molecules are near the origin, whereas more 3D molecules are distal (g) PMI plot of the Pharma set. (h) PMI plot of the PCC set. Red shows highly populated regions, with blue showing less populated regions. Most of the Pharma set is concentrated along the left diagonal.

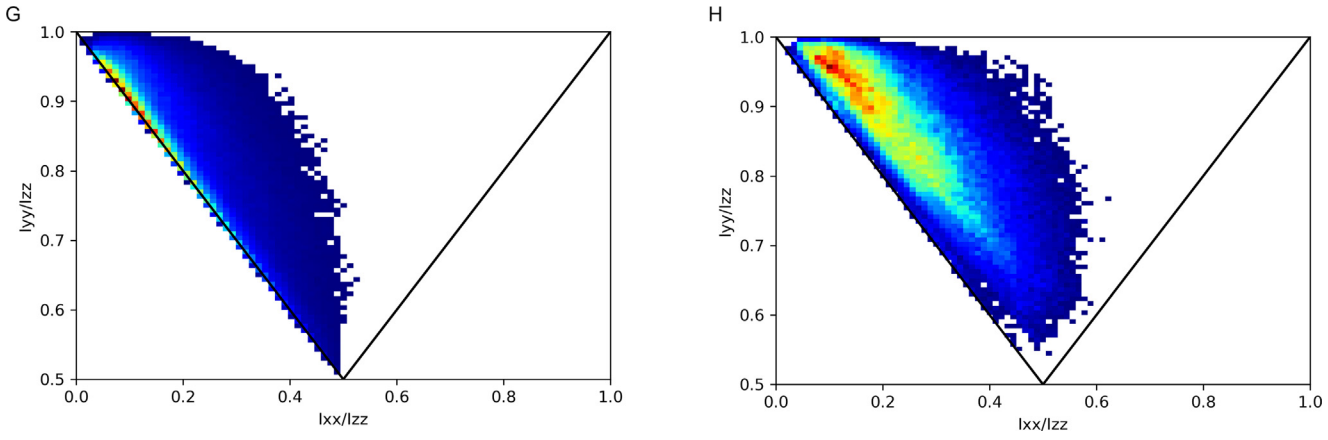

Fig. 1 (*continued*)

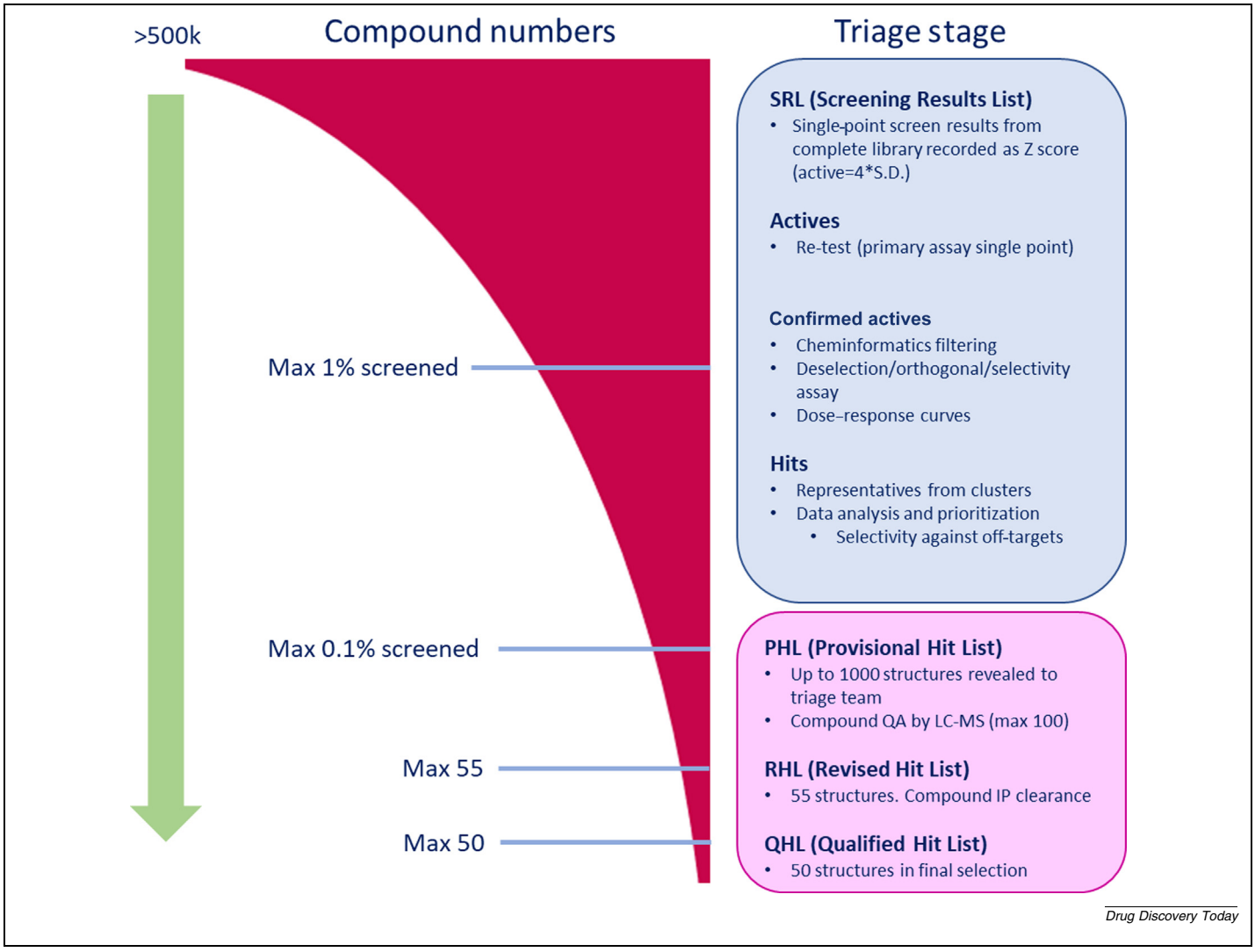

FIGURE 2

The European Lead Factory (ELF) triage process. The triage input is single-point data from the primary high-throughput screen (HTS) and the output is a list of up to 50 compounds. The first part of the triage, in blue, is performed without access to compound structures and relies on biological data from primary, confirmatory, deselection, and orthogonal assays, together with a large suite of cheminformatics tools. The second phase of the triage (pink) is performed with access to compound structures plus all other data and enables a final selection from a prioritized subset of the hits. Abbreviations: IP, intellectual property; LC, liquid chromatography; MS, mass spectrometry.

identity and purity of the compound before progressing compounds further. A maximum of 100 compounds can be selected for LC-MS analysis. Up to 55 compounds are selected from the PHL and registered on the Revised Hit List (RHL). Throughout the triaging process, the 'assay owners' (i.e., organizations that submitted the target and assay) indicate what compound properties they would like to see in the RHL set. These are the compounds that the program wishes to take as the output of the screen. The Qualified Hit List (QHL) is the final output list after IP checks have been made on the RHL by the compound owners. Compounds on the PHL and RHL that were not selected for the QHL are released and returned to the screening pool. Therefore, each stage represents increasing confidence in the quality of the compounds and provides a framework along which we can track compound progress and a measure of how attractive the compound was to the program team.

Target Class activity profile

The Pharma and PCC subsets were investigated to understand their activity profile against different target classes. Compounds in the Pharma set are contributed by companies each with their own history in therapeutic areas, whereas the PCC compounds are designed with an emphasis on novelty, sp^3 hybridized-rich scaffolds, and dissimilarity to patent and commercial chemical space. Therefore, the activity profiles of the two subsets in the screens might contrast. However, identifying active compounds in high-throughput screening campaigns is fraught with difficulty; nonspecific activity, impure samples, and assay-interfering compounds can overwhelm the small number of true actives and give misleading results.[(p17),(p18),(p19)] To address this problem, we exploited the ELF triage framework and used representation on the PHL as the criterion for a compound showing interesting activity to the program. Although this introduces some subjectivity, it incorporates the analysis expertise of the scientists performing the triage. It also represents how compounds are selected in 'real life' by the independent pharma industry, nonprofit, SME, and academic screening organizations.

The target classes that have been screened in the consortium to date are summarized in Figure 3. Kinases, G-protein-coupled receptors (GPCRs), ion channels, and proteases are well-known target classes and form a significant fraction of the target set. Target classes of more recent interest, such as protein–protein interactions (PPIs), transcription factors/regulators, and methyltransferases, are also well represented. Figure S1 in the supplemental information online gives a breakdown of the

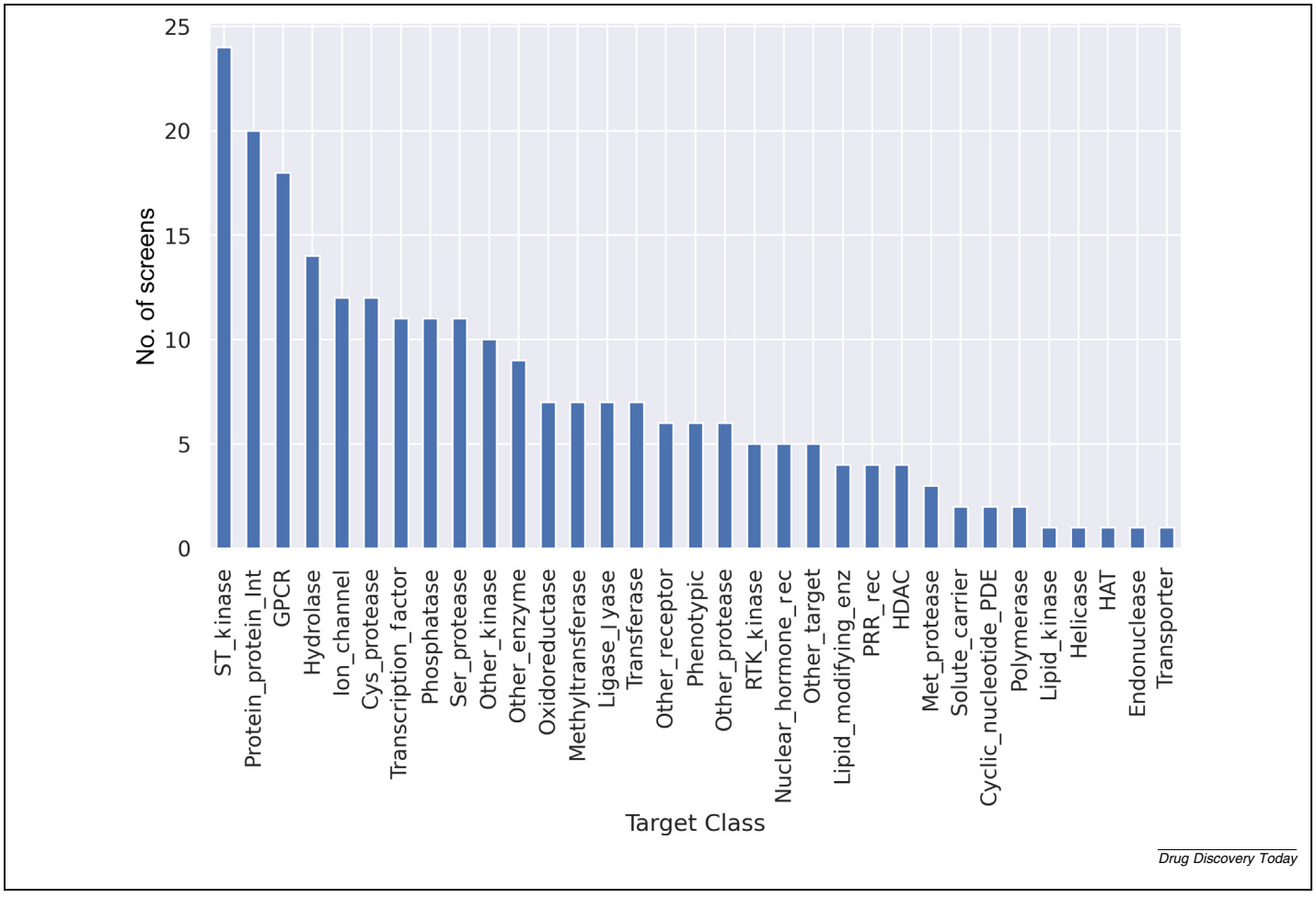

FIGURE 3

Overview of targets in screens. An overview of the number of screens run so far in the European Lead Factory (ELF) grouped by the target classes defined in the project.

screening assay types and categories used in the project. Biochemical assays were used most often (75%) compared with cellular assays (25%). Looking at the assay categories that were defined, it can also be seen that, in most cases (72%) functional assays were used, as were various binding assay formats (28%).

Screening data were grouped by target class and compound owner, belonging to either the Pharma set or to the PCC set. The frequency with which Pharma set or PCC compounds were observed on PHLs was then normalized by the number of times the compounds in each group were screened. The profile of PHL frequency of Pharma and PCC compounds is shown in Figure 4. In general, Pharma set compounds showed a higher normalized PHL hit rate, indicating a generally higher rate of selection by the program teams on the basis of bioactivity and physchem properties. This effect is particularly striking for kinases, whereas, by contrast, PPIs, proteases, GPCRs, and ligand-gated ion channels showed a less pronounced difference. The relative activities of the two subsets toward kinases is not surprising because this is a target class intensely worked by pharmaceutical companies and their libraries will contain significant numbers of compounds synthesized for this class. Most kinase inhibitors bind at the adenine site and analogs of the adenine moiety are often also sp2 rich; in an analysis of kinase scaffolds,[p20] the average Fsp^3 of the set of kinase scaffolds identified was 0.22. This was in a highly populated region of the Pharma collection but a very sparsely populated part of the PCC collection and would tend to disfavor sp^3-rich public compounds (Figure 1e). Differences in molecular weight (MW) and topological polar surface area (TPSA) between the Pharma and PCC collections could also have a role. PPIs, proteases, ion channels, and GPCRs are also well-established target classes and the PCC was notably active for these. This indicates that the ligand-binding sites of these target classes are better modulated by the sp^3-rich public collection compared with kinases. An overview of PHL hit rates, Z-Score (primary) hit rates, and a comparison of Pharma versus PCC Z-Score hit rates across target classes are shown in Figures S2–S4 in the supplemental information online.

Selection of compounds for the PHL from the screen is done without knowing the compound structures; once the compounds reach the PHL, a small group of triage chemists can see the compound structures for further selection onto the RHL. Therefore, it was of interest to understand whether Pharma or PCC compounds were more likely to be chosen by chemists once they were on a PHL and their structures were known. We calculated conditional probabilities that a PCC or Pharma compound was present on an RHL given that it was present on a PHL, and grouped the results by target class and Pharma or PCC subset (Figure 5). Here, the relative frequency of Pharma and PCC compounds on the RHL was closer, with Pharma compounds being preferred in some classes and PCC ones in others, indicating that the scientists choosing the compounds had no strong preference for PCC or Pharma compounds once they were visible. Interestingly, the conditional probabilities for PCC compounds were higher for Ser/Thr kinases, GPCRs, and ion channels. These are targets for many known drugs and this preference for PCC compounds suggests that the novel chemotypes offered by the

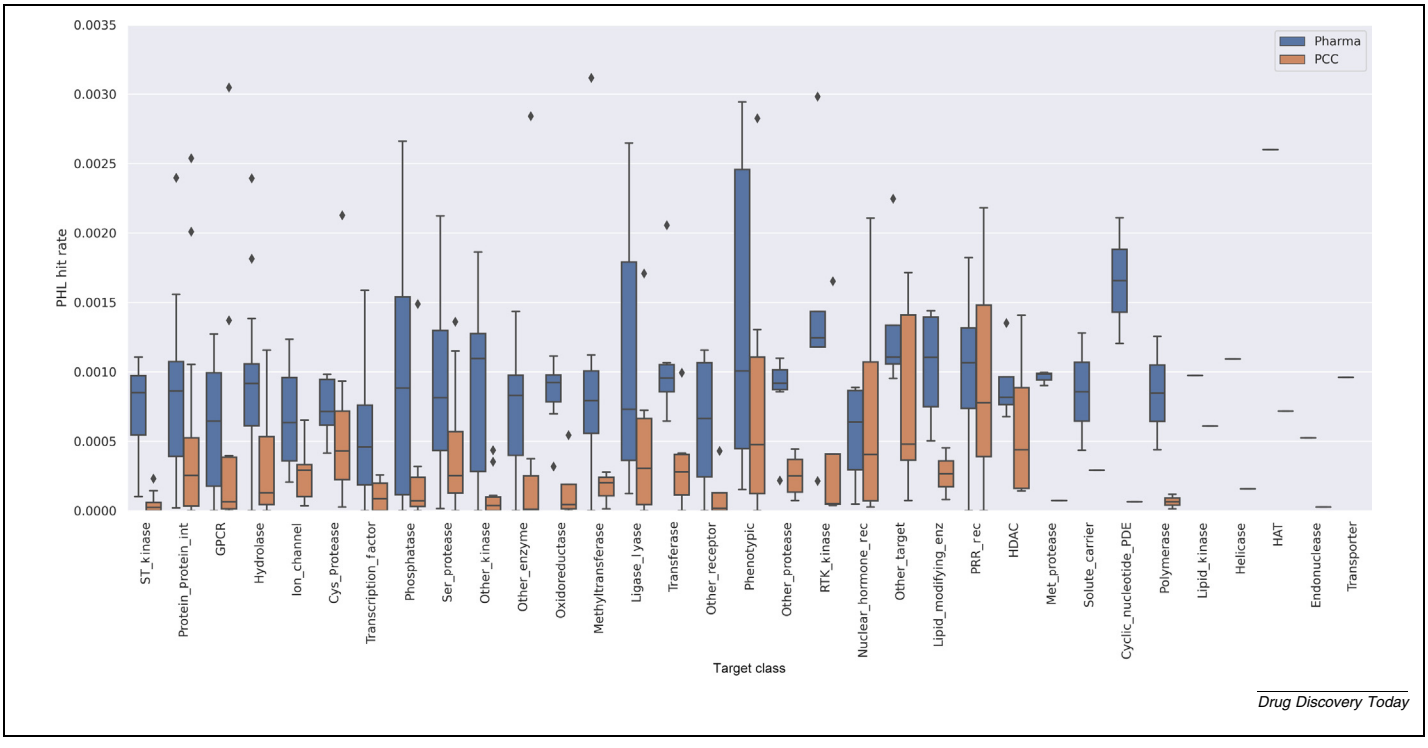

FIGURE 4

Hit probability of compounds. Boxplot of the probability of a compound appearing on a Provisional Hit List (PHL) normalized by the number of times screened, grouped by target class, over all screens shown in Figure 3 in the main text. Pharma compounds are shown in blue and Public Compound Collection (PCC) compounds in orange. Outlier screens are shows as diamonds.

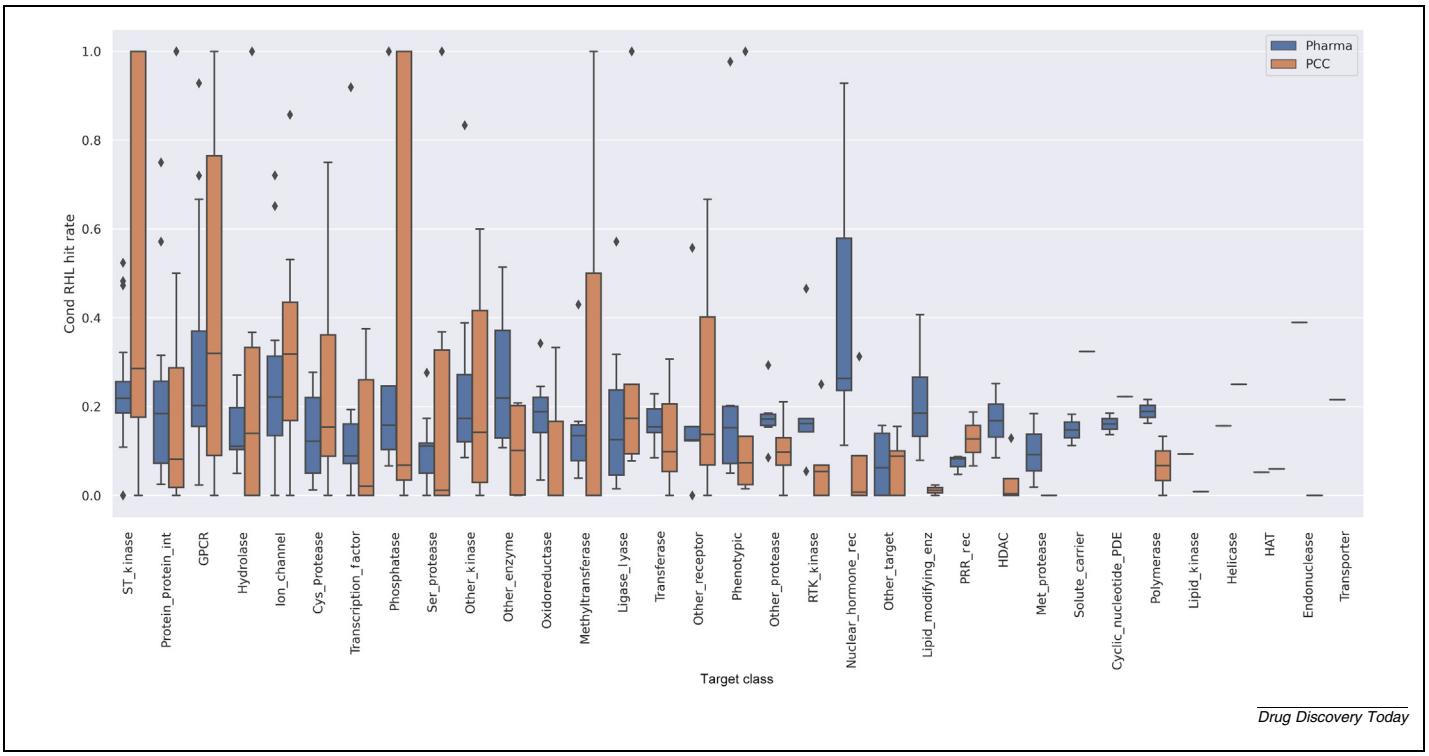

FIGURE 5

Conditional compound selection probability. Boxplot of the probability of a compound appearing on a Revised Hit List (RHL) after it was selected for the Provisional Hit List (PHL), grouped by target class. Pharma compounds are shown in blue and Public Compound Collection (PCC) compounds in orange.

PCC collection were of particularly high interest to the triage teams. Conditional probabilities for selecting compounds for the PHL given that they are a Z-score hit are shown in Figure S5 in the supplemental information online. Here, we see a slight preference for Pharma compounds for most target classes.

Properties of progressed compounds

The ELF triage framework can also be exploited to track the physchem properties of compounds as they progress from screen to final output. The mean AlogP, Fsp^3, and MW of SRLs, Z-score hits, PHLs, and RHLs grouped by target class were calculated (Figure 6; Figure S6 and Table S1 in the supplemental information online). This allows monitoring of how these physchem properties change when activity in assays is measured and in response to compound selection by the triage team at each stage. The average property value of the SRL (all compounds screened) represents a reference from which the relative changes in value can be followed. Figure 6 shows that there was a trend toward compounds with higher AlogP being active and being selected on PHLs compared with SRLs. AlogP information is available to the triage scientists when selecting compounds, and controlling the upper limit of AlogP is usually an aim of a triage. Therefore, it is likely that it is the biological targets that are driving this higher AlogP rather than selection by the project teams. The largest AlogP average values in the final selection lists (the RHLs) were for ion channels, transporters, helicases, and solute carriers, although there were very few screens for the latter three classes. The smallest AlogP average values in the RHL were for histone acetyltransferases (HATs; one screen only) and pattern recognition receptors (four screens). Another distinctive trend across all screens is the lower average Fsp^3 on Z-score active hits, PHLs, and RHLs. This is partially due to complex, sp3-rich molecules having a lower hit rate in the screen, as predicted by the Hann complexity hypothesis and demonstrated by Hansson *et al.*[p21] It also suggests that triage scientists on average do not favor sp^3-rich molecules over flatter compounds, even though their advantages have been published.[p13],[p14] MW shows a less obvious trend from SRL to Z-score hits to PHL with a relatively small upwards trend observed for average value here, but with some PHLs showing a lower average MW compared with the library as a whole (Figure S6 in the supplemental information online). The highest MWs on RHLs are for helicases, solute carriers, phenotypic screens, and GPCRs (all MW > 400).

We investigated whether the difference between the mean Fsp^3, AlogP, and MW on SRLs, Z-score hits, PHLs, and RHLs were significantly different. Significance was tested using paired sample *t*-tests between columns in Table S1 in the supplemental information online. For MW, a significant increase was found from SRL to Z-score hits and from Z-score hits to PHL (both $P < 0.01$). The magnitudes of the differences are perhaps of more interest. The mean MW difference was, for practical purposes, very small: equivalent to around a carbon atom in going from SRL to PHL for example. The differences in mean AlogP in moving from SRL to PHL was ~0.5 log units, and these ALogP differences were highly significant from SRL to Z-score actives ($P < 10^{-12}$) and from Z-score active to PHL ($P < 0.03$). Fsp^3 differences showed a distinct reduction in the proportion of sp^3-hybridized

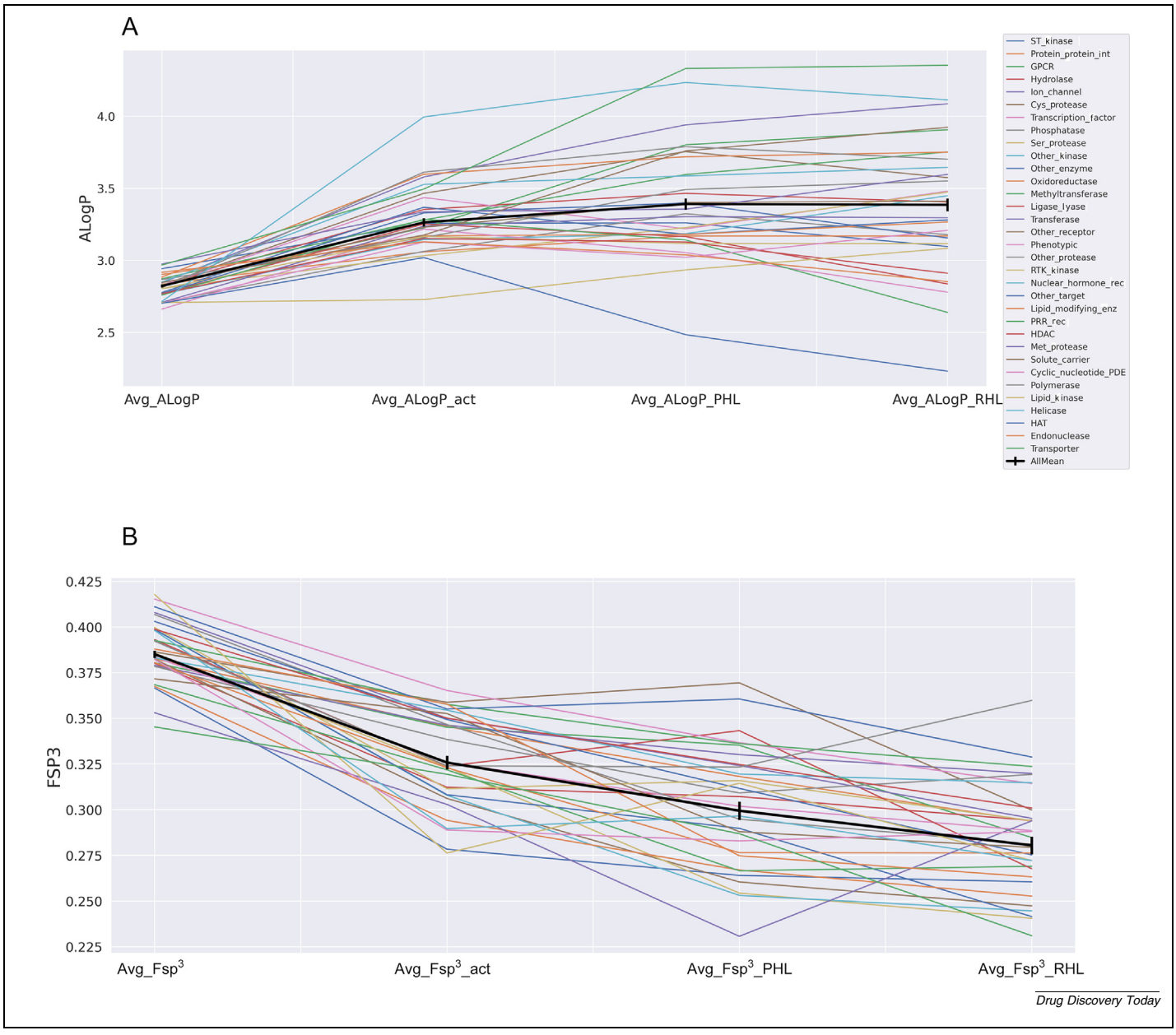

FIGURE 6

Compound properties along the triaging process. Mean values of physicochemical properties [(a) AlogP, (b) Fsp^3) of compounds along the triaging process from full screening set to Z-score actives, Provisional Hit List (PHL) and Revised Hit List (RHL) selections. Each line represents a target class. Differences in the values for the full screening set stage result from library changes during the project. The overall mean values are shown by the black line, with standard error of the mean indicated.

carbon atoms in molecules on the PHL compared with the SRL, with mean values of 0.39 (SRL) and 0.30 (PHL). The reduction in Fsp^3 between SRL, Z-score active, PHL, and RHL was highly significant in all cases ($P < 10^{-13}$, 10^{-5}, and 10^{-3}, respectively).

High hit rate compounds

An analysis of compounds active in multiple primary screens was done to identify 'frequent hitter' (FH) compounds. FHs are undesirable compounds that are active in assays against multiple biological targets and could act via a nonspecific mechanism that makes them unsuitable for further development.[(p18),(p22),(p23),(p24),(p25)] This activity might result, for example, from strong fluorescence, aggregation, nonspecific covalent modification, or redox properties. The ELF Library data set is attractive for this analysis because it represents active compounds identified in multiple screening locations and with multiple assay technologies, reducing systematic bias. However, elevated hit rate alone is not a sufficient criterion for identification as a FH. Privileged structures, such as benzodiazepines, also give rise to molecules that hit more frequently than average but in a useful manner that can be followed up.[(p26)]

Candidates for FHs were identified by collecting compounds showing activity across multiple target classes (Figure 2) and multiple assay technologies (Table S2 in the supplemental information online). A threshold of activity in five or more target classes was set to distinguish potential candidates for FHs from well-behaved but unselective compounds. The latter class of compounds could include, for instance, molecules showing activity across closely related target classes, such as Ser/Thr kinases and Tyrosine kinases. FH candidates were also required to show activity in at least five assay technologies. All technologies used by the ELF partners are light based and span colorimetric, absorbance, and multiple fluorescence formats. Table S2 in the supplemental information online lists the ten assay technology classes that were used, and their frequency of use, in the programs.

Before carrying out these identifications of candidates for FHs, assays with improbably high Z-score hit rates were excluded from the analysis. The hit rates in these assays were likely the result of issues in the performance of the assay rather than problematic compounds and could lead to misclassification of well-behaved compounds as FHs. The hit rate of each assay was calculated (compounds with Z-Score > 4 in the primary high-throughput screen) and the median hit rate + (4 × median absolute deviation) was used as the upper limit of assays used in the FH identification. This measure was used rather than a variance-based method because the distribution deviates widely from normal and the upper limit corresponds to a maximum allowable hit rate of just under 5%. In the analysis, we ignored all primary screens with a hit rate of 5% or higher.

In total, 1168 compounds were identified as FHs, showing activity in five or more target classes and five or more assay technologies. We did not see a relatively higher occurrence of activity in cell-based assays versus biochemical assays for FHs. A 2D histogram of these compounds is shown in Figure S7 in the supplemental information online. It is possible that some target classes are more likely to be hit by FHs, such as Cys proteases, which have a free Cys residue and could lead to nonspecific binding of compounds with an electrophilic group. We did not observe a strong bias in the target classes for FHs, normalized by the number of screens done for that target class (Figure S8 in the supplemental information online). Target classes that were represented above average were transcription factors, methyltransferases, and receptor Tyr kinases. Targets represented below average were GPCRs, ion channels, oxidoreductases, and ligases/lyases. Endonucleases and helicases also had high FH presence but only had one screen each (Figure 3).

The physchem properties of FHs are of interest because these are easily tunable in the design or purchase of screening compounds. We analyzed MW, AlogP, Fsp^3, and number of rotatable bonds. For AlogP, we observed a clear difference in the distribution of values between the overall collection and the FHs (Figure S9a in the supplemental information online). The mean AlogP of the full collection was ~2.5, whereas the mean AlogP for the FHs was ~3.5. This illustrates the benefit of designing compounds with lower AlogP values, as illustrated in Figure 1b for the PCC collection. The other property that showed a different distribution was Fsp^3 (Figure S9b in the supplemental information online). FHs tended to have lower Fsp^3 compared with the full compound set, confirming results published previously.[p14] No differences were seen for MW and number of rotatable bonds (Figure S9c,d in the supplemental information online).

Given that compound structures are blinded in the project, we analyzed the 1168 FHs by their Bemis–Murcko (BM) scaffold.[p27] These scaffolds result from preserving all ring systems and linkers between ring systems in a molecule and removing all side chains from the rings and linkers. Double bonds directly attached to rings and linkers are kept, as well as the atom types in the scaffold. The 1168 FHs belonged to 851 different BM scaffolds; thus, most were singletons. When we focused on the BM scaffolds that appeared four or more times in the list, seven occurred ten or more times (Figures S10 and S11 in the supplemental information online). Normalizing the numbers by the frequency at which the BM scaffold appeared in the overall collection reduced the (normalized) frequency of common scaffolds, such as quinoline, phenyl, 2-phenylquinoline, and stilbene. The FH scaffold that occurred the most was acridine (Figure S11a in the supplemental information online), which has been used historically in dye compounds. The selection criteria for compounds in the collection contained guidelines (without hard cut-offs) to adhere to Lipinski's Rule of Five and a list of unwanted chemical substructures, but this did not preclude the acridine scaffold being selected. The FH thienoimidazole scaffold in Figure S11d in the supplemental information online and two related scaffolds with high normalized frequencies (Figure S10k,l in the supplemental information online) were found in ChEMBL[p28] to have ATPase and hydrolase inhibitory activity. All molecules in ChEMBL with this scaffold had the sulfur in the linker present as a sulfoxide. Interestingly, the third most commonly occurring FH scaffold (Figure S11c in the supplemental information online) had no direct analogs in the ChEMBL database.

These results show that, although there are some FH scaffolds, such as acridine, that could be expected to cause fluorescence interference, there are also apparently benign compounds that are active in multiple target classes assayed in multiple assay technologies. Their activity could be caused by breakdown products accumulated over time, reactive components from the synthesis (e.g., metal ions), or aggregating activity. A key result from this work indicates that even attractive compounds should be carefully characterized to ensure that the activity shown in the assay is related to the structure in the database. Leveraging the accumulated experimental knowledge of compound activities across the portfolio is a key step in filtering such compounds out, together with the application of deselection assays on prioritized compounds. Machine learning methods are also showing some promise in identifying these compounds despite their diffuse structural similarity.[p17]

We found that 794 compounds appeared on more than four PHLs (Figure S12 in the supplemental information online). These PHLs can contain up to 2% of the full compound set (~1000 compounds) and are the first time during the screening triage when actual chemical structures can be seen. Compounds appearing on multiple PHLs are of interest because they have been attractive enough to be selected in multiple programs based on bioactivity and physchem properties, but repeatedly rejected

once the structures were seen by the triage chemists and then returned to the screening pool. Two compounds appeared on nine PHLs, indicating a high hit rate in the primary screens and sufficiently attractive to pass typical filters, but were subsequently assessed as unattractive for inclusion on an RHL. The compounds are not particularly unattractive chemically and it is likely that aggregation, sample impurities, or compound fluorescence are responsible for the apparent activity. A striking result is that there were few PCC compounds (five out of 794) selected on multiple PHLs. This is likely the result of multiple factors. They have in general been screened fewer times, although this does not explain such a large difference. In addition, they were all recently synthesized and purified by modern techniques; thus, reactive impurities or breakdown products will be less likely. The compounds are also larger and more complex than those in the Pharma subset and, again, the complexity hypothesis predicts that this would result in few promiscuous compounds. Within these 794 compounds, several BM scaffolds appeared relatively often. Perhaps not surprisingly, several of these scaffolds are also in the FH list in Figure S11 in the supplemental information online. These include the acridine, quinoline, benzoxazole, phenyl, 1,3-diphenylpropene, and scaffolds (Figure S11a,b,c,e,j, respectively). The BM scaffold that occurred the most in four or more PHLs contained the anilinepyrimidine moiety, which is found in many kinase inhibitors, including imatinib and nilotinib (Figure S13 in the supplemental information online). Lack of selectivity or novelty are possible reasons for not retaining these compounds for the final selected RHL list.

Dark chemical matter

Another perspective on the bioactivity of a molecule is whether it can be described as 'dark chemical matter' (DCM).[p29] DCM compounds are those that show no biological activity in a large number of screens (>100 was used in the original publication). Counted over all assays with a hit rate < 5%, the fraction of thus far inactive compounds in the Pharma collection was 31%, whereas that in the PCC collection was 44%. If we restrict the compounds that have been tested between 75 and 200 times, to allow a more balanced comparison between the Pharma and PCC sets, the fractions were 26% and 43%, respectively. It is important to bear in mind that 'darkness' is not by itself a desirable or undesirable property and can be seen as a selective molecule waiting for the right target. This is indeed the case, as can be seen in Figure 7, which shows a histogram of the number of compounds versus the number of screens in which they were a hit. This includes the numbers for 25, 50, 75, and 100% of the screens that were run in chronological order. As more screens were done, we see that the first bar (hit in 0 screens) decreases significantly, representing previously nonhit compounds that show up as active in later screens. Figure S14 in the supplemental information online shows how many screens were done before compounds were first active, split by Pharma and PCC sets. For both collections, we observe a long tail that clearly indicates that compounds can remain inactive for many screens before showing up as active in a screen (the large peak at 0 represents compounds that have not yet been active). One difficulty with identifying truly biologically inert compounds is the very large amount of

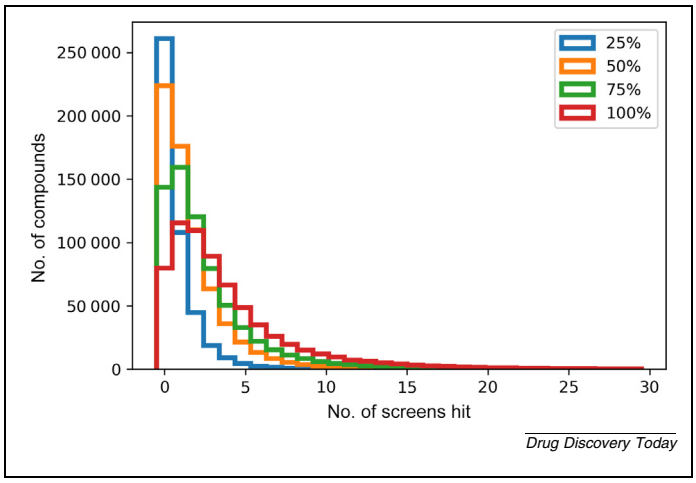

FIGURE 7

Compound hits over time. Number of compounds versus the number of screens in which they were a hit. Different colors indicate the statistics for the first 25, 50, 75, and 100% of screens in chronological order.

data necessary to show this. A typical library compound can be assumed be active in a high-throughput screen with a probability of ~1%. Using this probability, even screening such a compound 300 times, there is a 0.99^{300} (~5%) chance that a compound with a perfectly normal bioactivity profile will not show activity in any screen. Assuming a slightly lower but still reasonable activity probability of 0.5%, this overall chance increases to 22%.

The ELF compound sharing model

Nine large and medium-sized pharmaceutical companies contributed proprietary compounds from their collections to create the Pharma set of the ELF Library. This level of cooperation is unprecedented and brings together decades of high-throughput screening knowledge and diverse therapeutic area experience. The potential benefits of this sharing approach were investigated. Previous work showed that the compounds in each pharmaceutical company subset show relatively little similarity to those in other subsets with a mode Tanimoto similarity of ~0.2–0.3.[p6] Does this chemical dissimilarity between subsets follow through into differing biological activity profiles and appearance on the PHLs? The raw number of compound appearances on PHLs from each compound owner, including the PCC, against each target class are shown in Figure 8. The large peaks seen, for example, in the kinase subclasses are due to a large number of kinase screens being carried out. It is apparent within each target class that there is wide variation in the number of compounds represented on PHLs from different compound owners and that the relative contributions from each compound owner in target classes fluctuates widely. This suggests that combining subsets from different companies is a relatively simple way of maximizing the pharmacological space addressed by the full collection. This profile of activity shown by the subsets from the different companies can also be used to identify complementary collections that would be most beneficial to combine.

To gain an understanding of the practical effect of compound sharing on the screening outcome, we compared the profiles of the individual company subsets with that of the combined

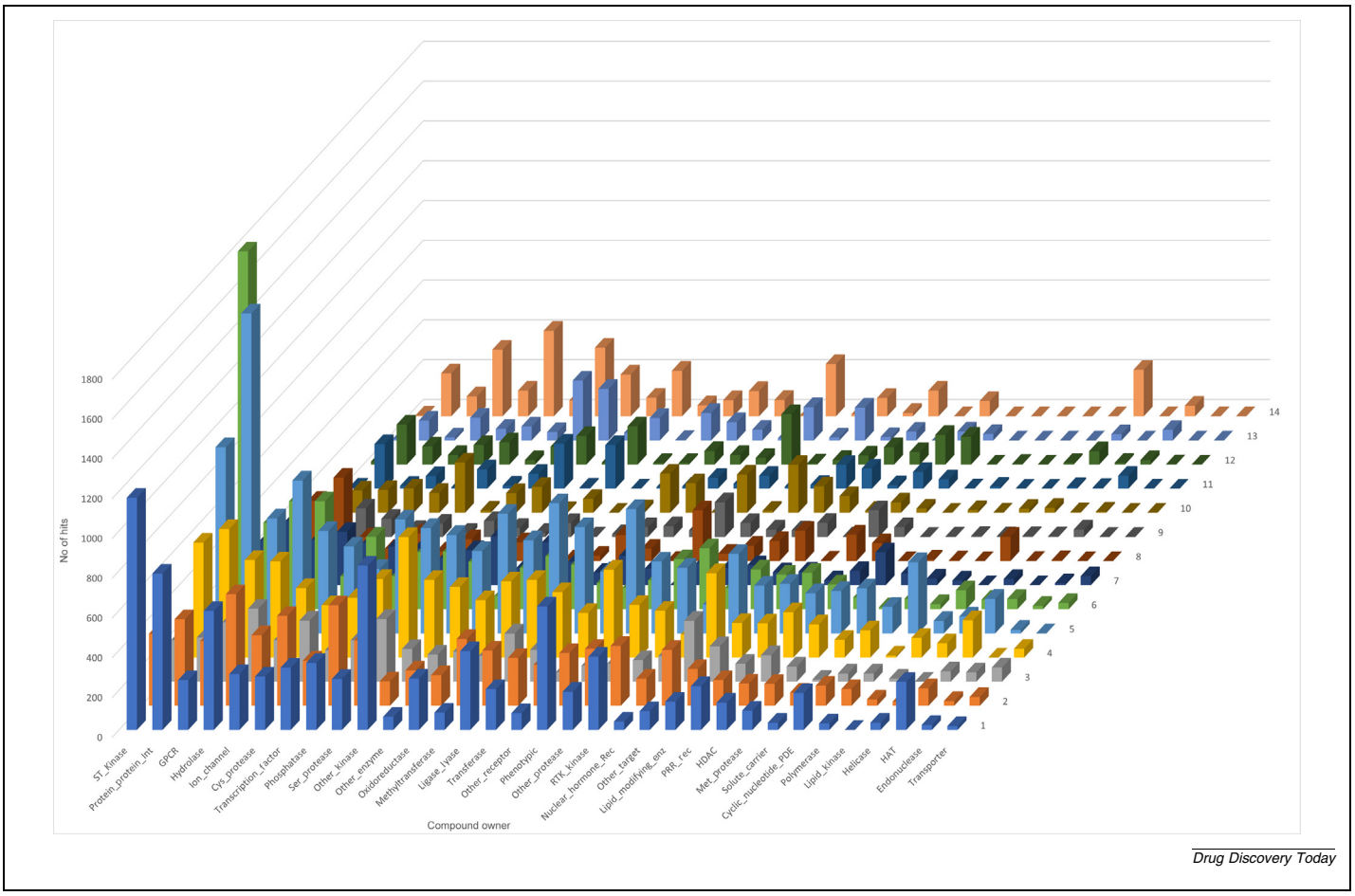

FIGURE 8

Target class hits per compound owner. Number of Provisional Hit List (PHL) appearances of compounds categorized by target class and compound owner (1–14). The Public Compound Collection (PCC) compounds are represented by compound owners 8–12.

Pharma set. The company subsets provide a benchmark for the performance of diverse compounds selected from individual collections on different target classes, while the combined Pharma set shows the effect of sharing compounds. First, screening results from the nine company subsets were selected and grouped by target class. The PHL frequency (number of compounds on a PHL)/number of Z-Scores) was calculated for each company subset and target class group. This gives the probability of success (p) for a compound being on a PHL in that target class. The expected number E of compounds necessary to be screened to obtain a PHL compound for a given target class is $1/p$. Figure 9 shows boxplots with each box representing a company subset. The range spanned by each box/whisker shows the expected number of compounds necessary to screen to obtain a PHL compound for each target class. Tractable target classes, where hits are easier to find, will have a lower value for expected number of compounds compared with more challenging target classes and will be in the lower quartiles of each box, while more difficult target classes will require more compounds to be screened and occupy the upper quartiles. The range within each boxplot shows that the number of compounds necessary to be screened to find a PHL compound for different target classes can vary considerably within each company subset, whereas differences between boxplots indicate differences between company subsets. The 'Comb' box shows the values for the combined Pharma set. The mean value for the combined Pharma set is equal to the mean of the nine subsets, but the scatter around the average and upper limits of the boxes is of greater interest. The values for the combined Pharma set are more tightly distributed about the mean, indicating that the number of compounds necessary to be screened to find a PHL compound for most target classes varies less than in individual company subsets. Furthermore, the height of the second quartile gives an upper bound to the expected number of compounds needed to be screened to find a PHL compound for different target classes.

Figure 9 also shows that the combined Pharma set has no outliers, suggesting that the 'blind spots' for specific target classes that any corporate collection might have can be relieved by the shared compound collection. Collections 5, 6, and 14 of Pharma have a lower average E value compared with the combined Pharma set, suggesting that these are effective collections for the target classes screened in the project, even though there are some outliers with higher E values. In summary, it is apparent that, for most target classes, the compound sharing model is an efficient approach for consistently finding hits registered on PHLs.

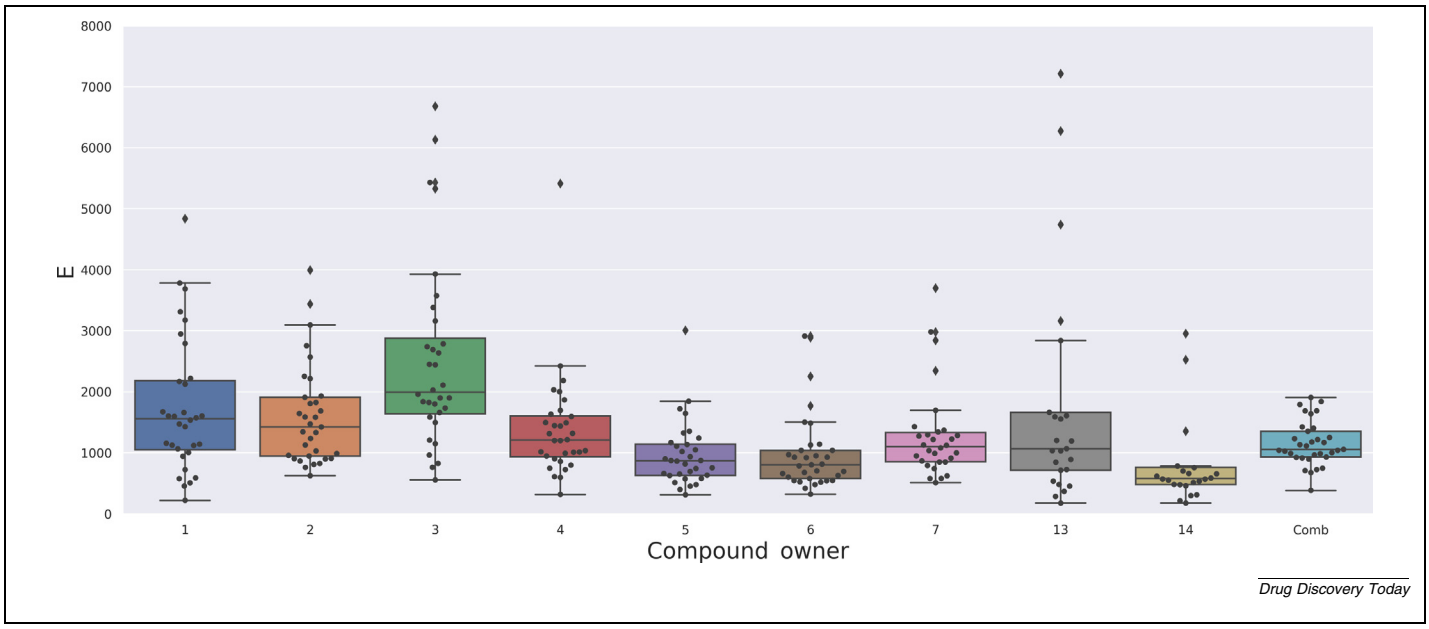

FIGURE 9

The value of compound sharing. Boxplot of expected number *E* of compounds necessary to be screened to obtain a Provisional Hit List (PHL) compound for all screens grouped by Pharma compound owner (1–14) and for the combined Pharma set (Comb). Each point represents all screening programs of a particular target class.

Discussion

The ELF was established to facilitate screening of therapeutically interesting biological targets proposed by academic groups or SMEs in Europe, followed by validation of the hits to build confidence in their developability. Three features of the ELF distinguish it from previous public screening initiatives: (i) the primary high-throughput screen is closely linked to a hit characterization and chemistry infrastructure; (ii) access to chemical structures and bioactivity information is tightly controlled; and (iii) access to a unique screening library comprising compounds contributed by multiple pharmaceutical companies along with newly designed and produced compounds using some of the latest organic chemistry methodologies available from academic universities and chemistry-focused SMEs.[p30] These features are aimed at generating downstream value by generating high-quality hits, building confidence in the hits, and maintaining the IP of the compounds to make them attractive for further public or private funding opportunities. This approach is complementary to other public screening initiatives where rapid disclosure of screening data and compound structures is fundamental. Although this places restrictions on the detailed data that can be disclosed, aggregation of programs into target classes coupled with a large array of molecular descriptors and cheminformatics tools enables effective triage by project teams and subsequent analysis.

The spectrum of biological targets for which hits (i.e., compounds on PHLs and RHLs) have been found demonstrates the effectiveness of the ELF Library against multiple target classes. However, it is critical not to rely on results from a single assay format to identify hits and multiple orthogonal validation and deselection is required to prevent wastage of resources in pursuing nonspecific modes of inhibition. Compounds emerging from these screens and showing appropriate behavior through validation have subsequently been developed into molecules with potent and selective activity in classes including PPIs, protein kinases, lipid-modifying enzymes, metalloproteases, serine proteases, type I and type II GPCRs, and ion channels.

These hits were derived from both the Pharma and PCC collections. In terms of the Pharma collection, compound sharing appears to be an effective way to efficiently explore pharmacological space and reduce bias toward or against particular target classes. Compounds from the PCC collection also found their place on PHLs across multiple target classes with slightly lower hit rates. It is clear from the data that the PCC compounds were attractive to triage teams; once the triage reached the PHL stage and compound structures became visible, both public and pharmaceutical company compounds were selected for RHLs.

An interim analysis of the performance of the EFPIA screening campaigns within each company revealed that screening of the PCC collection was of value for the companies (personal communications, stakeholder meeting of consortium November 2019). Companies were pleased that compounds on some of the qualified hit lists (QHLs) outperformed 'competing' in-house series or rescued programs where in-house screening campaigns failed to identify attractive chemical matter. Approximately 40% of QHLs received by pharma partners led to additional internal 'wet chemistry' work, which is a substantial percentage given the modest size of the QHLs compared with their internal screening compound collections. It was also noticeable that, although addition of PCC compounds was incremental, hits from these subsets appeared on QHLs as soon as they were added to the screening library. There are examples of hits from these subsets

on the QHLs of so-called poorly druggable targets (personal communication, stakeholder meeting of consortium November 2019).

Crowdsourcing of biological targets for screening has also been successful and the ELF has succeeded in attracting many interesting target programs from non-consortium members. Indeed, programs with a high societal impact but less often found in the research and development (R&D) pipeline of the pharmaceutical industry have benefited in particular from access to this consortium to obtain high-quality, new chemical matter. Screening of a Middle East respiratory syndrome target program has provided attractive tool compounds to researchers working in the field of a corona virus-mediated disease long before the devastating outbreak of Coronavirus 2019 (COVID-19) in 2020. In addition, a dedicated screening campaign has resulted in interesting new scaffolds of small-molecule inhibitors of viral entry into host cells as a potential prophylactic and therapeutic option for severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) infection (https://www.europeanleadfactory.eu/newsroom/cross-consortium-cooperation-strengthens-european-efforts-discover-and-develop-novel).

It is a notable outcome of the ELF screens that the ELF Library has generated some highly potent, validated chemical matter that has gone on to create value further along the drug discovery chain. The public and SME screens alone have generated over 240 compounds with measured pEC_{50}s > 7.3 (i.e., EC_{50} < 50 nM) in the primary assay dose response across multiple targets, with compounds from all the industry partners and from the PCC collection being represented in this potent set. An antimicrobial program generated a hit with a pEC_{50} of 7.5 (EC_{50} 30 nM) and, following optimization within the ELF, reached subnanomolar activity against the primary target and broad-spectrum activity against pathogenic bacteria.

Two companies (Scandicure and Keapstone Therapeutics) have been created on the basis of validated and optimized hits emerging from the ELF Library for two screening programs, respectively. One of these programs targets a kinase modulating lipid metabolism, whereas the other is a PPI involved in neurodegeneration. Both reached significant commercial milestones recently. A third SME (Metabomed) recently announced a successful closing of a financing round providing the company with over €12 million to bring their optimized lead compound on an oncology target closer to the clinic (Phase I studies have started). Other target classes, including GPCRs and ion channels, have also generated sub-100-nM hits that were subsequently validated with orthogonal technologies, such as surface plasmon resonance, patch clamp, or cell-based assays.

The ELF has shown that the public–private model can reach a high screening productivity, with 108 screens completed for academic and SME target owners and 168 at partner pharmaceutical companies in the period 2013–2022. Challenging targets have been taken on and multiple, investable, validated compound series have been generated. The confidentiality concerns that companies may have had in sharing significant proportions of their screening libraries were addressed by two factors. First, the selection of the Pharma compounds for the ELF Library was done without access to chemical structures. Chemical fingerprints and physchem properties were sufficient for integrating the subsets.[p6] Second, the staged triage process, which releases only a small proportion of the compound structures within a program, prevents large-scale analysis of compounds from other partners and enables the creation of IP on the compounds in the library. Confidentiality concerns of academics or companies in proposing their targets for screening were addressed by grouping targets into classes and not disclosing precise target information. Conversely, concerns over the ability to execute an effective triage without access to all compound structures were addressed with a full suite of filtering and cheminformatics tools in the HDB. These enable elimination of many compounds before a final manual selection from the prioritized list with access to structures and all data.

Underlying the screening performance, the expertise of the computational scientists in the pharmaceutical companies who selected the nine subsets that formed the Pharma ELF Library has been important. Ongoing engagement in the project by these pharmaceutical companies throughout the course of the collaboration (extending beyond the initial 5 years into a second IMI-funded stage) was also beneficial, and close integration of their knowledge is a strength of the ELF model. Finally, the requirement for a flexible software platform to triage the hits should not be underestimated. The cloud-based HDB ensured consistent application of the ELF project agreement across a widely distributed consortium and could respond to changes in the numbers of users with complete flexibility. It also provided the tools to execute the unique early part of the ELF triage where structures are blinded to the user and multiple ways of grouping and prioritizing compounds are necessary.

With the follow-up grant called ESCulab, operating under the ELF brand, the consortium continued to make available the full library to academics, SMEs and participating industry partners for screening. Previously only open to European researchers eligible for IMI funding, the ELF launched a partnering option for charities and foundations around the world, hoping to build further on the invaluable experience obtained during the first 10 years of ELF and help charitable organizations finding new starting points for their drug discovery (https://www.europeanleadfactory.eu).

Concluding remarks

The results from the ELF show that it provides a model for future drug discovery efforts where diverse partners can operate within a project agreement that protects their interests while sharing enough data to execute high-quality science. Although the pharmaceutical companies have larger proprietary in-house libraries available for their screening campaigns, library sharing across companies has been of particular interest because of the mutual complementarity between subsets. In addition, the generation of unprecedented, *de novo* PCC compounds, which became available for screening, resulted in validated hits that initiated follow-up chemistry. The ELF has already resulted in > 100 publications, nine patents, three compounds being tested in clinical studies, several investigational new drugs in preparation, and several licensing deals. These numbers are expected to increase as a result of the progression of the individual programs screened within the ELF framework. Given the timelines of drug

discovery, it is expected that additional success stories from academia, the pharmaceutical industry, and biotech will be published in the coming years.

European Lead Factory Consortium

Daniel Basting, Jorg Benningshof, Stan van Boeckel, Iuliana Botez, Sylviane Boucharens, Nicola A. Burgess-Brown, Mads H. Clausen, Phil Cochrane, Jeremy Davis, Alexander Dömling, Thierry Dorval, James Duffy, Ola Engkvist, Amaury E. Fernandez Montalvan, Dirk Finsinger, Rutger H.A. Folmer, Hubert Haag, Daniel Hamza, Steven P. van Helden, Saman Honarnejad, Helma van den Hurk, Jörg Hüser, Freek J. Janssen, Stefan Jaroch, Geraint Jones, Philip S. Jones, Mirek Jurzak, Markus Kaiser, Christopher Kallus, Michael Koch, Patricia M. Kramer-Verhulst, Sven M. Kühnert, Wen H. Lee, Frank Leemhuis, Brigitte Lesur, Volkhart M. Li, Stuart P. McElroy, Mark Meininghaus, Angus Morrison, Hartmut Mors, Stefan Mundt, Adam S. Nelson, Vera J.M. Nies, Peter Nussbaumer, Kristina M. Orrling, Romano V.A. Orru, Jean-Yves Ortholand, Eckhard Ottow, Andrew D. Pannifer, Garry Pairaudeau, Christopher Pearce, Karien Peeters, Alexander Piechot, Fernando Ramón Olayo, Ton Rijnders, Didier Roche, Peter W.M. Roevens, Eelco Ruijter, Floris P.J.T. Rutjes, Klaus B. Simonsen, Peter B. Simpson, Mario van der Stelt, Robert A. Stockman, Jason R. Swedlow, Lena Tagmose, Dimitrios Tzalis, Natasja van Velthoven, Céline Vermeiren, Jon S.B. de Vlieger, Herman van Vlijmen, Ton Vries, Markus Wagener, Herbert Waldmann.

Declaration of interests

No competing interests declared by authors.

CRediT authorship contribution statement

Herman van Vlijmen: Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft. **Andrew D. Pannifer:** Conceptualization, Investigation, Writing – original draft. **Phil Cochrane:** Data curation, Investigation, Software, Writing – review & editing. **Daniel Basting:** Project administration, Supervision, Writing – review & editing. **Volkhart M Li:** Investigation, Project administration, Writing – review & editing. **Ola Engkvist:** Investigation, Project administration, Writing – review & editing. **Jean-Yves Ortholand:** Project administration, Writing – review & editing. **Markus Wagener:** Investigation, Writing – review & editing. **James Duffy:** Investigation, Project administration, Writing – review & editing. **Dirk Finsinger:** Investigation, Project administration, Writing – review & editing. **Jeremy Davis:** Investigation, Supervision, Writing – review & editing. **Steven P. van Helden:** Investigation, Methodology, Project administration, Supervision, Writing – review & editing. **Jon S.B. de Vlieger:** Investigation, Project administration, Supervision, Writing – review & editing.

Data availability

Relevant summary data is reflected in figures. Detailed data on compounds, targets, assays are confidential and cannot be shared.

Acknowledgments

The European Lead Factory is a public–private partnership with >30 partners, together engaging >200 scientists working on parts of the work presented in this paper. We thank everyone involved over the years for their important contribution to the success of the initiative. We also explicitly thank the academic and Small and Medium sized Enterprises (SME) scientists who have trusted us to work on their drug targets and provide them with starting points for their drug discovery projects, as well as the chemistry-contributing third parties. Without their high-quality chemistry and screening proposals, none of the work described in this paper would have been possible. The ELF has received support from the IMI Joint Undertaking under grant agreement no. 115489, resources of which comprise financial contribution from the EU's Seventh Framework Programme and EFPIA companies' in- kind contributions. This project also received funding from the IMI-2 Joint Undertaking (JU) under grant agreement no. 806948. The JU receives support from the EU's Horizon 2020 research and innovation programme, the EFPIA and MMV. This communication reflects the views of the authors and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.drudis.2024.103886.

References

1. Chakma J, Sun GH, Steinberg JD, Sammut SM, Jagsi R. Asia's ascent—global trends in biomedical R&D expenditures. *N Engl J Med*. 2014;370:3–6.
2. Frye S. US academic drug discovery. *Nat Rev Drug Discov*. 2011;10:409–410.
3. Shanks E, Ketteler R, Ebner D. Academic drug discovery within the United Kingdom: a reassessment. *Nat Rev Drug Discov*. 2015;14:510.
4. Mullard A. European Lead Factory opens for business. *Nat Rev Drug Discov*. 2013;12:173.
5. Karawajczyk A, Orrling KM, de Vlieger JSB, Rijnders T, Tzalis D. The European Lead Factory: a blueprint for public–private partnerships in early drug discovery. *Front Med*. 2016;3:75.
6. Besnard J, Jones PS, Hopkins AL, Pannifer AD. The Joint European Compound Library: boosting precompetitive research. *Drug Discov Today*. 2015;20:181–186.
7. Jones PS, Boucharens S, McElroy SP, Morrison A, Honarnejad S, van Boeckel S, et al.. IMI European Lead Factory—democratizing access to high-throughput screening. *Nat Rev Drug Discov*. 2022;21:245–246.
8. Karawajczyk A, Giordanetto F, Benningshof J, Hamza D, Kalliokoski T, Pouwer K, et al.. Expansion of chemical space for collaborative lead generation and drug discovery: the European Lead Factory Perspective. *Drug Discov Today*. 2015;20:1310–1316.
9. van Vlijmen H, Ortholand J-Y, Li V-M-J, de Vlieger JSB. The European Lead Factory: an updated HTS compound library for innovative drug discovery. *Drug Discov Today*. 2021;26:2406–2413.
10. ELF PCC collection analysis. https://www.europeanleadfactory.eu/newsroom/new-analysis-elf-public-compound-collection [Accessed 11 January 2024].
11. Paillard G, Cochrane P, Jones PS, van Hoorn WP, Caracoti A, van Vlijmen H, et al.. The ELF Honest Data Broker: informatics enabling public-private collaboration in a precompetitive arena. *Drug Discov Today*. 2016;21:97–102.
12. Sauer WHB, Schwarz MK. Molecular shape diversity of combinatorial libraries: a prerequisite for broad bioactivity. *J Chem Inf Comput Sci*. 2003;43:987–1003.

13. Lovering F, Bikker J, Humblet C. Escape from flatland: increasing saturation as an approach to improving clinical success. *J Med Chem*. 2009;52:6752–6756.
14. Lovering F. Escape from Flatland 2: complexity and promiscuity. *MedChemComm*. 2013;4:515–519.
15. Kenny PW, Montanari CA. Inflation of correlation in the pursuit of drug-likeness. *J Comput Aided Mol Des*. 2013;27:1–13.
16. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. *Nat Chem*. 2012;4:90–98.
17. Babaoglu K, Simeonov A, Irwin JJ, Nelson ME, Feng B, Thomas CJ, et al.. Comprehensive mechanistic analysis of hits from high-throughput and docking screens against β-lactamase. *J Med Chem*. 2008;51:2502–2511.
18. Feng BY, Simeonov A, Jadhav A, Babaoglu K, Inglese J, Shoichet BK, et al.. A high-throughput screen for aggregation-based inhibition in a large compound library. *J Med Chem*. 2007;50:2385–2390.
19. Shoichet BK. Screening in a spirit haunted world. *Drug Discov Today*. 2006;11:607–615.
20. Hu Y, Bajorath J. Exploring the scaffold universe of kinase inhibitors. *J Med Chem*. 2015;58:315–332.
21. Hansson M et al. On the relationship between molecular hit rates in high-throughput screening and molecular descriptors. *J Biomol Screen*. 2013;19:727–737.
22. Dahlin JL, Walters MA. The essential roles of chemistry in high-throughput screening triage. *Future Med Chem*. 2014;6:1265–1290.
23. Dahlin JL et al. PAINS in the assay: chemical mechanisms of assay interference and promiscuous enzymatic inhibition observed during a sulfhydryl-scavenging HTS. *J Med Chem*. 2015;58:2091–2113.
24. Irwin JJ et al. An aggregation advisor for ligand discovery. *J Med Chem*. 2015;58:7076–7087.
25. Jones P, McElroy S, Morrison A, Pannifer A. The importance of triaging in determining the quality of output from high-throughput screening. *Future Med Chem*. 2015;7:1847–1852.
26. Welsch ME, Snyder SA, Stockwell BR. Privileged scaffolds for library design and drug discovery. *Curr Opin Chem Biol*. 2010;14:347–361.
27. Bemis GW, Murcko MA. The properties of known drugs. 1. Molecular frameworks. *J Med Chem*. 1996;39:2887–2893.
28. Mendez D et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res*. 2019;47:D930–D940.
29. Wassermann AM et al. Dark chemical matter as a promising starting point for drug lead discovery. *Nat Chem Biol*. 2015;11:958–966.
30. Morgentin R et al. Translation of innovative chemistry into screening libraries: an exemplar partnership from the European Lead Factory. *Drug Discov Today*. 2018;23:1578–1583.