

Perspective, Hit Discovery Methodology, SLAS Discovery (<https://doi.org/10.1177/2472555220942765>)

**Hit discovery for public target programmes in the European Lead Factory: Experiences and output from assay development and ultra-High Throughput Screening**

Saman Honarnejad<sup>1,\*</sup>, Stan van Boeckel<sup>1</sup>, Helma van den Hurk<sup>1</sup>, Steven van Helden<sup>1</sup>

Affiliations

<sup>1</sup> *Pivot Park Screening Centre B.V., Kloosterstraat 9, 5349 AB, Oss, The Netherlands*

\*Correspondence should be addressed to SH (email: [saman.honarnejad@ppscreeningcentre.com](mailto:saman.honarnejad@ppscreeningcentre.com))

## Abstract

The European Lead Factory (ELF) consortium provides European academics and small & medium enterprises access to ~0.5 million unique compounds, a state-of-the-art ultra-High Throughput Screening (u-HTS) platform and industrial early drug discovery (DD) expertise with the aim to deliver innovative DD starting points. From 2013-2018, 154 proposals for 8 target classes in 7 therapeutic areas were submitted to ELF consortium, 88 of which were accepted by the selection committee. During this period, 76 primary assays based on 7 different readout technologies were optimized and mainly miniaturized to 1536-well plates. In total, 72 u-HTS campaigns were carried out and follow-up work including hit triage through orthogonal, deselection, selectivity and biophysical assays were finalized. This ambitious project showed that besides the quality of the compound library and the primary assay, the success of centralized u-HTS of large compound libraries across many target classes, various assay types and different readout technologies is also largely dependent on the capacity and flexibility of the automation on the one hand and hit triaging phase on the other, particularly due to undesired compound-assay interference. Thus far, the delivered 'hit lists' from ELF consortium resulted in spin-offs, patents, *in-vivo* proof of concepts, pre-clinical development programmes, peer-reviewed publications<sup>1-3</sup>, PhD theses and much more, demonstrating early success indications.

## Keywords

Drug Discovery, European Lead Factory, Assay Development and Miniaturization, ultra-High Throughput Screening, Public-Private Partnerships

## Introduction

While success rates in discovery and development of new medicines remain low and unmet medical needs are still a huge burden on global health care<sup>4, 5</sup>, many academic discoveries with great potential for pharmaceutical exploitation have remained underutilized<sup>6, 7</sup>. The underlying reason for this is that on the one hand, academic labs typically lack the resources, infrastructure and expertise to progress early-stage biomedical discoveries towards development and on the other, pharmaceutical companies and venture capitals are mainly attracted to investing in the later-stage de-risked projects<sup>8</sup>. Over the last decade, the awareness about this so called ‘innovation gap’ has resulted in notable changes to the traditional pharmaceutical R&D model. A shift from ‘closed innovation’ to ‘open innovation’<sup>9</sup> based on complementarity of assets, collaborative efforts and shared-risk within academia, biotech, Contract Research Organizations (CROs) and big pharmaceutical companies<sup>10</sup>. Such public-private partnerships not only hold the promise of bridging the ‘innovation gap’ between academia and industry, but also integrating the fragmented efforts in the DD landscape through ‘open innovation’.

ELF consortium is a collaborative public-private partnership with over 30 partners (2013-2018), funded by the EU commission through Innovative Medicines Initiative (IMI) and the European Federation of Pharmaceutical Industries and Associations (EFPIA). In its first phase (IMI-1 funding period 2013-2018), this consortium provided researchers in Europe with a unique early DD platform to translate innovative biology and chemistry concepts into high-quality starting points for DD<sup>6</sup>. ELF consortium accomplished this by capitalizing on two pillars consisting of the first ‘Joint European Compound Library’ (JECL) and the first ‘European Screening Centre’ (ESC) (Figure 1A).

The JECL is a diverse pool of ~0.5 million drug-like compounds, partly contributed by the EFPIA companies and partly generated by chemistry CROs and academic labs (the latter called public compound collection). Two studies on the JECL showed that the library is high-quality with respect to drug-likeness and physiochemical profile, chemically diverse with each subset driven from a distinct region of chemical space and predicted to have activities on a wide spectrum of biological targets, based on Bayesian models, built using data in ChEMBL database<sup>11, 12</sup>. Another study on the public compound collection described the approaches and challenges for successful generation of a large screening collection by academic labs and chemistry CROs, while populating a biologically relevant chemical space that is distinct from the EFPIA subsets<sup>13</sup>. These overall attributes formed the foundation of a unique and attractive screening collection for small-molecule lead discovery. The ESC, consisting of five different partners, collectively assured the identification of most promising chemical starting points for further development through their unique and complimentary assets. Participating partners in ESC were Pivot Park Screening Centre (PPSC) (for assay development, u-HTS, data analysis and part of hit triage), University of Oxford (for protein production and structural biology), University of Dundee, Newhouse (for biophysical hit characterization and medicinal chemistry), BioAscent (for compound logistics) and Lygature (for programme management). Even though the ultimate goal of the ELF consortium is to facilitate and accelerate the discovery of new medicines, it also resulted in unique opportunities such as access to external innovation, knowledge and resource exchange, extended network and independent partnerships for participating pharmaceutical companies and CROs (<https://www.europeanleadfactory.eu/elf-2013-2018/results-2013-2018>).

In this perspective, we mainly focus on PPSC as an independent CRO with an emphasis on activities and services, the impact of ELF consortium on PPSC and contributions of PPSC to ELF consortium. Furthermore, we summarize the overall hit finding efforts in ELF consortium public target screening

programmes; crowd-sourced from biomedical researchers at European academic organizations and small biotech companies from 2013 to 2018 and describe a chemical biology approach to systematically evaluate assays prior to screening phase and how such an approach could help with tailoring screening assay conditions and guide post-screen follow-up activities. Moreover, we illustrate the application of cheminformatics tools to enrich blinded primary screen results and reflect on the lessons learnt during assay development and u-HTS operations.

### **Pivot Park Screening Centre (partner in ESC)**

PPSC is a quickly growing CRO with a main focus on u-HTS. The company is specialized in early DD services including cellular and biochemical assay development, u-HTS, hit-to-lead biological profiling and laboratory automation support with customers ranging from big pharma to small & medium biotech companies, non-profit/non-governmental organizations and universities in Europe, the US and Asia. Upon the takeover of Organon/Schering-Plough by Merck Sharp & Dohme and the decision to close the lead discovery research site in Oss, the Netherlands, PPSC was established as a spin-off in 2012. PPSC currently operates with ~30 employees including members of the management team, assay development and robot operator groups, as well as additional supporting administrative, finance and human resource staff.

PPSC is a project-driven organization and as such organized in a matrix structure, in which project teams are formed with internal (a project leader, 1-2 assay developers, 1-2 robot operators and an account manager) and external outsourcing partner members. After initiation of projects based on an agreed work package (with clear milestones, deliverables, acceptance criteria and timelines), outsourced projects typically start with a kick-off meeting and scheduling bi-weekly or monthly progress report teleconferences, followed by transfer of an existing assay cascade from customers or development of novel screening assay cascades from scratch and subsequent optimization and miniaturization to u-HTS-ready format. PPSC receives and screens customer libraries ranging from thousands to millions of compounds mainly in 384- or 1536-well formats of both echo-compatible and non-compatible plates or alternatively screens its own ~0.3 million diverse compound library, followed by fully automated data analysis. Identified active compounds are typically re-tested and then re-delivered in REMP tubes and profiled in primary and follow-up assays for hit triaging and subsequent Dose-Response Curve (DRC) testing for confirmed hit characterization. An assay development project typically takes 1-6 months and the duration is dependent on the complexity of the assays, availability of commercial reagents and the need for in-house reagent development. u-HTS projects including assay optimization, miniaturization and follow-up triaging usually take 2-4 months. Hit-to-lead biological profiling services often include delivery of 10s-100s of compounds to PPSC on weekly to monthly basis, followed by reporting of Structure-Activity Relationship (SAR) biological data to guide external chemical synthesis cycles. The current automation infrastructure at PPSC includes a state-of-the-art u-HTS lab for large-scale compound screening and an open-access lab with multiple automation islands and liquid handling stations for medium-scale HTS, hit-to-lead biological profiling support, plate preparations, biophysical testing and high-content screening. Other facilities at PPSC include assay development labs, tissue culture labs and molecular biology lab for wet lab work up to safety level 2, as well as storage rooms for compound libraries, chemicals, reagents and consumables.

With the existing state-of-the-art screening infrastructure, over two decades of experience in screening and industrial expertise in lead discovery, PPSC held a competitive position to become a partner for ELF consortium. To this end, PPSC carried out all the assay development and miniaturization, u-HTS

campaigns and part of hit triaging for the public target screening programmes. The overall public target screening programmes and the interactions within ESC and the whole ELF consortium contributed to experience and knowledge advancement in novel biology, exposure to a wide range of target classes, therapeutic areas, assay readout technologies, hit triaging cascades design and the expansion of PPSC's network within academia, chemistry CRO space and pharmaceutical companies. PPSC aims to sustain its growth trajectory by maintaining an up-to-date u-HTS platform through investing in the integration of state-of-the-art screening technologies and novel complex biological assays to accelerate the early lead discovery efforts of its global clients.

### **Scientific review and proposal selection**

All the eligible public target proposals underwent a rigorous scientific review, while the proposals were kept confidential. This was carried out by the screening selection committee consisting of members within EFPIA, ESC and external members who evaluated the proposals based on scientific quality, target validation, novelty of target and/or approach, or clear need for hits, availability of high-quality assays and supporting data, compatibility of assays with u-HTS and available readouts, chemical tractability, level of innovation, differentiation compared to existing treatments and exploitation potential. Together with assessment of elements like patient benefit and risk diversity of the portfolio, proposals were selected that contributed to a scientifically sound, diverse, risk-balanced, and economically justifiable portfolio of screening programmes. Interestingly, the overall distributions of disease areas and target classes across 154 submitted proposals were largely maintained in the 88 selected proposals. Upon approval of the target programme proposals and signing of the legal documents, a programme team was formed including the Programme Owner (PO) and experts from the ESC, who wrote a comprehensive programme plan defining the strategy to retrieve the best <50 hits i.e., the Qualified Hit List (QHL)<sup>4</sup>.

### **Assay development and miniaturization**

There are many approaches to small-molecule lead discovery. One of the most prominent ones is random High Throughput Screening (HTS). Over the last decades, unbiased HTS has remained one of the core approaches of many pharmaceutical and biotechnology companies to identify novel starting points for drug development. In fact, a recent analysis of clinical candidates suggests that ~30% of novel leads are generated through random HTS<sup>14</sup>. The quality of the outcomes from HTS campaigns are dependent on several factors such as robustness and quality of the primary assay, buffer composition, readout technology, kinetics including incubation times and temperatures, composition of the compound library, quality of automation as well as the post-screen hit triaging strategy. A major challenge in evaluating the hits from HTS campaigns is to select hit compounds that are worth pursuing as a chemical lead in a drug discovery programme. Therefore, having appropriate follow-up assay cascades in place is crucial for any successful hit finding effort<sup>15</sup>.

Upon ELF consortium onset, u-HTS assay development procedures required rapid evolution into a pharma grade early drug discovery operation with smooth assay feasibility testing, validation, optimization and miniaturization pipeline. In order to standardize assay transfer from POs, several documents were created to support every programme. Among others, Assay Input Requirement (AIR) document (describing detailed assay reagents and protocols), programme plan (describing the outline of the screening cascade), and general guideline regarding ELF consortium (explaining the workflow and terminology) were used for all programmes. The screening cascade outline included in the final programme plan described the general public target programme screening cascade (Figure 1B). Upon

signing the legal documents, AIR documents were transferred and programme plan was agreed upon. PPSC tested the feasibility of original conditions and reagents in the assays as described in the AIR document and proposed alternative conditions, setups or readouts if e.g. the PO results were not reproducible, assay parameters were suboptimal when miniaturized, original assay setup was biased towards high hit rate or picking up assay interfering compounds despite optimization efforts, etc. Additionally, based on the identified liabilities in the screening cascades, alternative or tailored assay conditions were included for optimal hit triaging. u-HTS campaigns on the JECL were carried out at a single point (10 $\mu$ M) and depending on the primary active rate, active compounds were re-picked from the same source plates and re-tested in primary assay. Confirmed actives (plus a selection of non-active compounds chosen based on Near neighbor analysis of actives and Bayesian model activity predictions) were re-delivered from BioAscent and depending on the project, orthogonal assays (e.g. same assay setup with different readout, alternative enzymatic product with same or different readout, etc.) and deselection assays (counter-screen assays such as same assay readout without the target and/or liability-based assays such as resonance energy transfer quenching, sensitivity to non-drug-like mode of actions such as metal ion chelation, aggregation, etc. as well as viability assays) were employed to confirm activity and deselect false-positives due to e.g. technology interference and uninteresting compounds e.g. due to toxicities, chelation, aggregation, etc. Remaining compounds were re-delivered in DRCs and tested in the primary and appropriate follow-up assays in duplicates (n=2) at 7 different concentrations (7-point) with maximal dose of 20 $\mu$ M and 1/2 log ( $\sqrt{10}$ ) dilution series. Duplicate data points seemed to be sufficient due to having 7 dose points per compound and infrequent outliers. Combination of these activities at PPSC, biophysical characterization of hits at the University of Dundee, Newhouse and cheminformatics filtering formed the basis for selecting the Preliminary Hit List (PHL), followed by evaluation of compound structures by medicinal chemists at the ESC, selection of up to 100 most promising compounds for LC-MS analysis and delivery of QHL. On average the transfer of assays from POs to ESC took about 1 month. On few occasions, POs visited PPSC to accommodate assay transfers and/or learn about screening operations. After full assay transfer and arrival of reagents, ~2-3 month was spent on assay panel feasibility testing, optimization and miniaturization, ~1-2 weeks for u-HTS and active confirmation, ~2-3 month for hit validation, DRC testing, biophysical characterization and PHL registration, ~1-2 month for LC-MS analysis and ~1-2 month for QHL registration.

The general public target programme screening cascade as outlined in the programme plan may seem routine to pharmaceutical R&D community and academic screening groups. However, a great deal of discussion between ESC members and POs was vital to justify the basic criteria for u-HTS and acceptable parameters for u-HTS-ready assays; especially terms like assay robustness, reproducibility, window and quality which were interpreted in many different ways. Other challenges ranging from technical to biological aspects were encountered during the course of the ELF consortium operations. As an example, the interaction with over 40 potential POs prior to the submission of the application was essential to raise the ELF consortium awareness, guide assay development experiments and inculcate the concepts relevant to early DD process. Furthermore, 25 proposals passed the selection process, despite not meeting all requirements for ELF consortium and were as such 'conditionally' accepted. For instance, those programmes were based on assays with kinetic readout, large volumes, suboptimal Z-prime, inclusion of washing steps, no compatible readout (e.g. radioactivity) or lack of stable cell lines. All 25 conditionally approved programmes except one were further developed at PPSC to meet u-HTS standards. Although the total number of assay development activities needed for these programmes to reach u-HTS-ready stage was higher compared to the other programmes, accepting these programmes had an added value to ELF consortium 'public target portfolio' in form of innovativeness, novelty and attractive biology and knowledge expansion of POs. In general, the number of experiments per

programme prior to the transfer to screening mainly depended on familiarity of the POs with HTS. The overall ELF consortium public target portfolio in the first phase (IMI-1 funding period 2013-2018) belonged to eight different target classes including Enzymes (60%), Protein-Protein Interactions (PPIs) (15%), Receptors (~9%), Ion Channels (5%), G Protein-Coupled Receptors (GPCRs) (~4%), Protein-DNA interactions (~4%), Protein-RNA interactions (2.5%) and Chaperone (~1%) (Figure 1C). These targets related to seven different disease areas including oncology (40%), CNS and neurology (16%), infectious diseases (16%), metabolic diseases (11%), neglected tropical diseases (9%), cardiovascular diseases and hematology (6%) and inflammation and immunology (2%) (Figure 1D).

Upon transfer of all materials to PPSC, for every single project, assay feasibilities were tested using the detailed protocol and reagents provided by the PO. If meeting all requirements, the programme progressed to assay optimization and miniaturization. During assay development and optimization, several experiments were performed for each project to check for appropriate assay component concentrations, stability of reagents and their minimal required concentrations or volumes, kinetics (such as incubation time, temperature, Km/Kd determination), assay linearity, EC<sub>50</sub>/IC<sub>50</sub> of reference compounds, inter-/intra- plate variation, DMSO tolerance, optimal final reagents concentrations and volume per well required, plate effects as well as robustness of the assay using reagents, chemicals and micro-titer plates from different vendors. Furthermore, the awareness for the need of orthogonal, deselection, selectivity and other assays within academia was rather low. PPSC together with the University of Dundee, Newhouse team (currently largely employed at BioAscent), developed most of these assays from scratch, reaching >200 overall assays (Figure 2A). For this, several factors beyond and above assay robustness and window, such as the choice of appropriate protein (state) and biochemically sound assay composition including concentrations of assay components, types of salts, pH, reducing agents, detergents, primary assay pitfalls, etc. were taken into consideration.

The assays employed by PPSC covered a wide range of readout technologies including Fluorescence Polarization (FP), Fluorescence Intensity (FI), Fluorescence Resonance Energy Transfer (FRET), Homogeneous Time Resolved Fluorescence (HTRF), Time Resolved Fluorescence Resonance Energy Transfer (TR-FRET), Fluorescence Imaging Plate Reader (FLIPR), AlphaScreen (AS), Luminescence (Lum) and Absorbance (Abs) (Figure 2B). These readouts were based on biochemical (~85%), cellular (~13%) and bacterial (~1%) assays (Figure 2C). To reduce reagent consumption and processing time on the u-HTS equipment to cover as many programmes as possible, all assays, except three, were successfully miniaturized to 1536-well format with final volume per well ranging from 2-8 µl. From 82 transferred protocols, 2 were not feasible due to lack of active protein within reasonable price range and 2 others were stopped due to lack of time for further optimization within the first ELF consortium funding period. In summary, 80 primary assays were shown to be feasible, of which 8 could not reach the screening phase, partly due to end of project running period. Out of these 8, 4 POs received a u-HTS-ready protocol (meaning fully automated and miniaturized assay protocol) ready to be applied to screening in 1536-well format. The remaining 72 programmes progressed to u-HTS phase. The data from primary screens of these programmes showed that the average primary active rate (based on Z-score >4 or <-4) is neither dependent on the target class (Figure 2D), nor readout technology (Figure 2E) and large variation in primary active rates within each target class (CV=88-169%) and readout technologies (CV=58-184%) suggested that in HTS campaigns, biological targets from the same class do not necessarily behave the same. This observation was not entirely unexpected, as each target class contained variety of subclasses with different target protein structural domains, catalytic activities and assay components. Additionally, the data provided counterevidence to general assumptions with respect to superiority of certain readout technologies and confirms that the primary active rate is a poor

predictor of 'real' hit rate, specially for challenging targets such as protein interactions with typically large interaction surfaces. Nevertheless, the collected data from these campaigns showed that the median of primary active rates for ion channels, GPCRs and Protein-DNA interactions as well as Abs, FLIPR, FP and luminescence readouts are below 1% and luminescence primary assays had the least variation (CV=58%) compared to other readouts. It is of importance to note that Z-score >4 or <-4 was used here to make an unbiased comparison across readouts/targets and even though the same criteria was used to identify primary actives in many programmes, depending on the data spread; primary active selection for some programmes were based on %Effect alone or in combination with Z-score ranging from >10 to 4 or <-10 to -4.

### **Robustness set compound collection**

It is widely accepted that besides the quality of the library and the primary assay, the success of any u-HTS campaign is largely dependent on the post-screen hit triaging phase; particularly due to undesired compound-assay interference leading to false-positive data. The undesired false-positive effects of these compounds disturb the hit selection process, thus requiring orthogonal assay readouts and adjustments in primary assay conditions. For organizations running many u-HTS assays with the same compound library, such interfering molecules in the long run become recognized as 'frequent hitters'. The main underlying causes for such interferences are aggregation, metal ion chelation, redox activity, auto-fluorescence, absorbance, luciferase inhibition and chemical reactivity<sup>16-18</sup>. To dissect such liabilities early in assay development phase, together with University of Dundee, Newhouse, the 'Robustness Set' compound collection was established. Robustness set comprises of selected molecules from the literature with well-defined assay interfering properties and so called 'clean' compounds for which we did not expect assay interference. These were supplemented with many DMSO controls, all in a ready-to-employ single 1536-well plate (Figure 2F).

The application of the robustness set within ELF consortium target programmes not only provided a framework to identify assay liabilities and tailor u-HTS assay conditions such that minimal interferences occur, but also helped to design suitable follow-up assay cascades for successful hit triaging. Based on the lessons learnt from early ELF consortium target screening programmes, the established framework prior to calling an assay u-HTS-ready was to perturb the assay with different classes of molecules with known assay interfering properties and a small set of chemically diverse lead-like compounds without obvious assay interfering properties. Active compounds within the robustness set were then identified, by quantifying responses to these perturbations (%Effect relative to DMSO controls). Depending of the assay type and responses to reference compounds, typically a compound in the robustness set was called 'active' when it resulted in %Effect >25-50%. Once active compounds were identified, a post-analysis was done to calculate the percentage of active compounds per each class that showed a %Effect >25-50%. While percentages of actives in 'clean' compound class gave an indication of the expected hit rate, percentages of actives in other compound classes informed on assay liabilities and biases towards picking up undesired compounds. Depending on the nature of identified liabilities, appropriate adjustments were made to assay conditions.

As an example, we profiled the performance of a FP assay from a PO using the robustness set. The original assay protocol indicated dithiothreitol (DTT) as a reducing agent in assay buffer. Testing robustness set compounds resulted in identification of assay liabilities, mainly towards redox-cycling compounds (~45% of compounds in this class were active), auto-fluorescent (~25% of compounds in this class were active) and colored compounds (~40% of compounds in this class were active). DTT and oxygen present in HTS samples may generate H<sub>2</sub>O<sub>2</sub>, which can oxidize essential functional groups in

proteins (e.g. cysteines). Such processes are catalyzed by compounds capable of redox cycling; thus, resulting in false-positives<sup>19</sup>. To tailor the assay buffer such that it becomes less biased towards redox compounds, a weaker reducing agent (L-Cysteine) was used, resulting in picking up fewer redox compounds (~17% of compounds in this class were active) (Figure 2G). Even though using a weaker reducing agent is not expected to have an impact on liabilities towards auto-fluorescent and colored compounds, the results from robustness set test guided the decision on the type of orthogonal and deselection assays to be used in hit triaging phase.

In another example, a luminescence-based primary assay for the activity of an epigenetic enzyme was miniaturized to 1536-well plate format based on adaption and optimization of the PO protocol. Results from testing the robustness showed that the assay is responsive to luciferase inhibitors as expected (90%) and predicted a high hit rate based on percent of actives in clean class (~38%). Furthermore, the assay setup was prone to picking up different classes of assay interfering molecules such as auto-fluorescent, colored, redox and aggregators. Even though additional of 2% tergitol NP-9 to assay buffer reduced these liabilities, switching to HTRF readout technology significantly improved the primary assay quality (Figure 2H) and minimized efforts needed in hit triaging phase.

### **ultra-High Throughput Screening**

In addition to the assay development activities, all u-HTS campaigns for ELF consortium public programmes were performed at PPSC. This facilitated the use of screening equipment in assay development. PPSC started operations in 2013 with a newly installed, state-of-the-art u-HTS platform developed together with HighRes Biosolutions. The system was designed to run biochemical and cellular assays in a fully automated manner using 1536-well plates, running >20 full-deck screens per year on a 2.5 million compound library. It had built-in flexibility to support different types of assays and projects, and featured redundant instruments to allow for continuing operations in case of instrument failures. The u-HTS system consisted of 3 pods with Staubli robotic arms surrounded by >50 instruments and connected by a conveyer belt system. The 3 pods operated as a single system for large and complex projects or independently to perform different (smaller) tasks. Most instruments were positioned on carts, so that the configuration of the system could be easily adapted to the needs of individual programmes and future technology upgrades. The system was accompanied by a unique and custom-designed safety system consisting of enclosures and hurricane doors. The enclosures shielded the users from the robot arms and the hurricane doors were closed when needed, to separate individual instruments from the robot. This allowed for trouble shooting, reagent top-ups and even assay development activities for new programmes in parallel, without interrupting the robot activities. For example, one of the ELF consortium screening programmes required supplementing freshly prepared enzyme solution every hour due to enzyme instability, which was smoothly accommodated by the platform. The compound preparation pod featured a 12-sided pod including 4 incubators for storage of compounds in controlled conditions.

The JECL compounds, prepared in 1536-well format by BioAscent in Scotland, were stored online for short-term use in sealed plates within the environmentally controlled incubators (~10% humidity with constant N<sub>2</sub> flushes at room temperature), directly accessible by the robot for screening. To limit and control compound oxidation, the compound source plates were refreshed every 3 months or used in <10 operations to ensure that compound integrity is maintained. u-HTS-ready assays were tested for reproducibility, stability and robustness during pre-screen runs (methodology-phase). Protocols were optimized to eliminate and if not possible, minimize plate incubation and dispenser effects. Quality during screening was monitored by including DMSO dummy plates and quality control plates containing

serial dilutions of available reference compounds. During primary screens, samples were transferred into assay plates (mostly 1536-well format, but the system also supported 384-well format) using ECHO 555 acoustic dispensers. Source plates were re-sealed and returned to the on-line store while the assay plates were transported to the other pods using the conveyer belt system. The two assay pods on the system were both 9 sided and held all instruments that were required to perform various biological assays. Different types of dispensers were available to enable users to select optimal conditions for a wide range of reagents types and volumes and seven available Cytomat incubators allowed different incubation steps at temperatures ranging from 4 to 37 °C. Assay readouts were performed in pod 3 that covered a wide range of readout modalities. To keep the u-HTS platform up-do-date and accommodate a broader range of biological assays and readouts, several novel instruments and technologies were implemented into the u-HTS system. Among these were reliable, flexible and fast dispensers to increase accuracy, reduce dead volumes and reduce clogging (ECHO 555 and Certus), platewashers suitable for robust washing steps in 1536-well format and medium replacement (BlueCatBio BlueWasher), FLIPR-Tetra suitable for ion channels and fast kinetics readouts such as calcium fluxes and membrane potential measurement, next-generation of multimode readers (PHERAstar) and recently Rapiflex (Bruker) for label-free MALDI-TOF Mass Spectrometry readout capable of u-HTS in order to minimize use of costly detection reagents, limit technology interference and shorten hit triage.

Data processing was done using a combination of in-house and commercial software. The custom-built SAMTRACK software continuously tracked samples and monitored operations on the robotic system. Logfiles produced by dispensers like the ECHO were processed for every sample across the screening campaign. These plate maps were then automatically fed into ActivityBase software platform (from IDBS) together with the data files from the readers without manual intervention and users easily evaluated the data remotely in ActivityBase and performed data quality control. When necessary and if applicable, empirical orthogonal function in IDBS software was used to correct for reproducible plate effects. Z-factor ( $Z'$ )  $>0.6$  and Signal-to-Background (S/B)  $>3$  ( $>2$  in HTRF) were used as quality parameters for approving data from plates. For a few cellular screening programmes that were not robust,  $Z' >0.5$  was also accepted for data approval. After data approval, the final data were loaded into ScienceCloud in a batch process that ran automatically every night. The batch process was included later to increase efficiency. The ScienceCloud package was used for secure storage and analysis of all screening and triage data. In order to maintain confidential information (e.g. about the JECL structures), Honest Data Broker (HDB)<sup>20</sup> was specially integrated into ScienceCloud and used to analyze blinded u-HTS campaign results. To use the HDB as a tool throughout all ELF consortium processes, the consortium agreed on standardized metadata (e.g. for target classes, technology readout and mode of action) as well as abbreviations for different triage phases (for instance P for primary screen, C for confirmation, O for orthogonal, etc.). This was the basis to create a dedicated database that can be used for cross-programme analysis to determine e.g. frequent hitters, technology interfering compounds, unique compounds, target class related compounds, etc. For identification of primary actives, depending on respective mode of action Z-score of  $>4$  or  $<-4$  was used as minimum criteria for initial active selection.

If number of primary actives was above selection criteria, 'Active confirmation' was fulfilled via re-picking compounds from the full library. In general, active rate was typically lower in active confirmation than primary screen. This could be partly explained by statistical false-positives during primary screen resulting in lack of activity within re-picked compounds. In addition, the process of re-picking primary active compounds that are randomly distributed across the compound source plates is time consuming which results in 1536-well assay destination plate long unsealed time and risk of DMSO evaporation and compound precipitation. To avoid DMSO evaporation in assay destination plates during compound re-

picking, compounds were transferred to  $\leq 10$  columns of 1536-well destination assay plates. This reduced the destination plate unsealed time, resulting in better correlations with compound effects observed in primary screen. Furthermore, depending on the active rate, sometimes the selection criteria needed to become more stringent to stick within the compound limit of 1% to be re-ordered for testing as serial dilutions to generate DRC data. This cut-off made the follow-up work more economical.

To enrich the primary active compound list while carrying out blinded u-HTS campaigns, the power of cheminformatic tools available on ScienceCloud were used to reduce the risk of early abandonment of false-negatives and potentially generate early SAR data using Bayesian models and Near neighbor searches. Bayesian model used molecular fingerprints of the compounds together with experimentally determined biological activities to derive features of active molecules that were associated with biological activity. Then, the model cross-referenced molecular fingerprints of every molecule in the full library against features associated with biological activity to find potential false-negative compounds. Typically, the top 250-500 molecules predicted to be active were included for follow-up re-testing. Near neighbors were compounds showing similar structural scaffold as active compounds, yet they were not identified as actives in the primary screen. An acceptable chemical similarity score cut-off was set to find these molecules, which were later included in the primary active list for re-testing. 11 programmes had QHLs containing at least one Bayesian model- or Near neighbor-derived compound.

Even though during the project ramp-up phase fewer programmes were recruited than planned, PPSC managed to complete >20 full-deck screens for ELF consortium programmes in 2017 (Figure 3A), reaching u-HTS production of >12 million data points (in combination with non-ELF screening campaigns); an output typically expected from an established large pharmaceutical company. The throughput for every screen was optimized per assay based on required incubation times and stability of reagents. For screens carried out in 384-well formats, the maximum number of data points per day was up to 54,000 (Figure 3B). Every u-HTS campaign started with a pre-screen of 50 plates and continued if assay performed within acceptance criteria and by processing around 100-200 plates per day for screens based on 1536-well formats; the system routinely generated up to 200,000 data points per day and even ran an enzymatic assay on 350,000 compounds in less than 24 hours (Figure 3B-C). While most screens were carried out in 3-4 days, some expanded to 5 days for re-testing failed plates and a few campaigns based on 384-well format took up to 7-16 days, mainly due to assay instability. These were the same assays in which efforts to miniaturize them to 1536-well format failed.

To identify generalizable trends in the data collected during the screening cascades for all programmes, Principal Component Analysis (PCA) was carried out on different variables such as excitation and emission wavelength (except luminescence and for absorption), plate format (number of wells in 384-well and 1536-well formats), volume per well, library size (the size of the library increased over the course of 5 years), number of screening days, assay robustness ( $Z'$ ) and window ( $S/B$ ), primary active rate, confirmed active rate, number of re-ordered compounds, PHL, number of LC-MS qualified compounds, QHL, number of Bayesian model- and Near neighbor-derived compounds in QHL (Figure 3D). In a two-dimensional PC space, PC1 mainly captured the variation in number of compounds registered in QHL with higher numbers associated with some of the cellular assays and PC2 was a function of screening days and well formats. This orientation in PC2 space was partly trivial, as the u-HTS campaigns based on some of the cellular assays that were carried out in 384-well format had larger volumes per well and took more days to finalize than others. Interestingly, these programmes mainly focused on identification of agonists and antagonists, but not inhibitors and had the highest possible compounds (50) registered in QHL. In contrast, programmes focused on inhibitor finding mainly against

Protein-DNA, Protein-RNA, Receptors and some PPIs and enzymes had up to 10 compounds in QHL, in line with known challenges to identify inhibitors of protein interactions. Low number of compounds in QHL list did not seem to be readout dependent. We carried out the same analysis on biochemical programmes (Figure 3E) and cellular programmes (including a bacterial programme) (Figure 3F) separately to avoid biases towards assay type specific trends. While in PCA of biochemical screening campaigns beside relative short distances between Protein-RNA, Protein-DNA and some Receptors, no distinct clusters were identifiable in PC space; PCA of cellular screening programmes mainly captured differences in screening days and plates formats across similar targets and readout technologies, suggesting the potential for miniaturization is not necessarily target class or readout technology dependent.

## Conclusion

Discovery and development of new medicines is a risky and time-consuming business. The increasing costs and declining R&D productivity remains a major challenge for large pharmaceutical companies and global health care. In response to this sustainability crisis, collaborative efforts and open innovation models have gained much attraction. ELF consortium is an example of such collaborative efforts to accelerate the discovery of new medicines. This pan-European platform not only facilitates the translation of novel biology and chemistry into high-quality starting points for DD, but also stimulates the cooperation of public institutes and the private sector.

Over the course of 5.5 years within the first phase of ELF consortium, 154 target programmes were submitted to the consortium. After review by the selection committee, 88 proposals were accepted, of which 6 were not progressed for legal reasons. The distribution of target classes and disease areas of accepted proposals were roughly similar to the distributions across all submitted proposals. For the remaining 82 programmes, feasibility of 78 primary assays were validated and 76 assays were successfully optimized and miniaturized to 1536-/384-well format. u-HTS campaign for 72 programmes was completed, which followed the finalization of follow-up triaging and delivery of QHLs (Figure 3D).

The results from these campaigns made it evident that usually HTS assays that are not adaptable to 1536-well format, also perform sub-optimally in lower plate formats and re-confirmed the challenges in identifying small-molecules against protein interactions with other proteins and nucleic acids. In addition, these results demonstrated that even within the same target class, outputs from different readout technologies are context dependent and in-depth understanding of the molecular interactions between targets, compounds, assay components, additives and buffer conditions is critical to design optimal u-HTS assays. Moreover, we showed that systematic perturbation of screening assays with well-characterized molecules is a reliable method to determine assay liabilities and design tailored follow-up cascades accordingly. This approach fits well with the existing infrastructure in screening facilities and provides a framework to execute successful screening programmes rapidly, particularly important for project-based organizations with high project turnovers. On the technical side, we exemplified the importance of miniaturization for maximized throughput and showed how flexibility of the automation could in the one hand accommodate biologically and technically diverse screening programmes and in the other, facilitate the integration of various new instruments to keep a u-HTS platform up-to-date.

Together with its partners at ESC, PPSC actively contributes to ELF consortium by disseminating knowledge and knowhow to public target portfolio, streamlining industry-standard operations from

assay concepts all the way to confirmed hits, already delivering >200 high-quality assays and successful u-HTS campaigns and follow-up triaging results. In the meantime, PPSC has positioned itself as a well-known service provider in international early DD market and academic space, thanks to the knowledge and experiences gained during the project as well as the network that accompanied the ELF consortium. On the EFPIA side of the ELF consortium, independent HTS campaigns on the JECL were carried out by partner companies (at their own in-house screening facilities); resulting in combined 'industry target portfolio', adding value to the internal drug discovery programmes of each partner through the expanded chemical space facilitated by the ELF consortium.

Access to ever-expanding biologically relevant chemical space, enabled through shared high-quality compound libraries, together with the widespread use of novel label-free u-HTS technologies such as MALDI-TOF Mass Spectrometry not only hold the promise of accelerating the discovery of de-risked new chemical entities for further development, but also reducing the cost of large-scale screening and the duration of hit triage as well as hit-to-lead phases. We envision that in parallel to other lead generation methods such as structure-based drug design, screening DNA-encoded libraries, fragment-based screening, etc., method development efforts for label-free multi-parametric/phenotypic readouts and maturation of physiologically relevant disease modeling in miniaturized assays based on organoid, iPSC and Crispr/Cas technologies have the potential to pave the way for the discovery of the next generation of high-quality starting points for drug development and ultimately improve pharmaceutical R&D productivity by reducing late stage attrition rates.

It is commonly known that on average it takes 10-20 years to bring a single new medicine to the market. Therefore, it is still too early to truly evaluate the direct impact of initiatives such as ELF consortium on health care. Nevertheless, early success stories and achievements from ELF consortium now form the basis for continued operations of the ELF consortium concept through European Screening Centre; unique library for attractive biology (ESCulab; <https://www.europeanleadfactory.eu/>), an IMI-2 Project.

### **Acknowledgements**

We thank all our colleagues in assay development and operator groups at PPSC, members of the ESC, especially our collaborators at BioAscent (formerly at University of Dundee, Newhouse), public target programme owners and all members of ELF consortium for their feedback and helpful discussions. We would also like to show our gratitude to Jon de Vlieger (Lygature) for critical comments to the manuscript.

### **Author contributions**

All authors wrote the manuscript.

### **Declaration of Conflicting Interests**

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115489, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in-kind contribution.

### Figure legends

**Figure 1)** The European Lead Factory (ELF) consortium and the associated screening cascade for public target programmes in different target classes and disease areas. A) ELF consortium pillars: Joint European Compound Library (JECL) consisting of ~0.5 million compounds partly contributed by EFPIA companies (~0.3 million) and partly synthesized by chemistry CROs and academic labs (~0.2 million) and European Screening Centre (ESC) pillar with expertise, infrastructure and logistics capabilities to support early drug discovery programmes; B) The public target programme screening cascade from proposal accession to the generation of QHL; C) ELF consortium public target portfolio consisting of eight different target classes including Enzymes (60%), Protein-Protein Interactions (PPIs) (15%), Receptors (~9%), Ion Channels (5%), G Protein-Coupled Receptors (GPCRs) (~4%), Protein-DNA interactions (~4%), Protein-RNA interactions (2.5%), and Chaperone (~1%); D) Public target portfolio disease areas consists of oncology (40%), CNS and neurology (16%), infectious diseases (16%), metabolic diseases (11%), neglected tropical diseases (9%), cardiovascular diseases and hematology (6%), and inflammation and immunology (2%).

**Figure 2)** Assay development and miniaturization at European Screening Centre (ESC). A) >200 assays including primary, deselection, orthogonal, selectivity and other assays were optimized or developed from scratch at PPSC (together with university of Dundee, Newhouse team); B) u-HTS readout technologies including Fluorescence Intensity (FI; ~27%), Fluorescence Polarization (FP; ~17%), Homogeneous Time Resolved Fluorescence and (Time Resolved) Fluorescence Resonance Energy Transfer technology (HTRF and (TR-)FRET; ~16%), AlphaScreen (AS; ~14%), Luminescence (Lum; ~12%) Fluorescent Imaging Plate Reader (FLIPR; ~8%), and Absorbance (Abs; ~6%) were employed at PPSC; C) Biochemical (~85%), cellular (~13%) and bacterial (~1%) assay types in public target screening programmes; D) Primary active rate based on Z-score >4 or <-4 across different target classes including %CV per class; E) Primary active rate based on Z-score >4 or <-4 across different readout technologies including %CV per readout type; F) Robustness set compound collection comprising of clean compounds (~40%), DMSO (~24%), Aggregators (~11%), Redox (~7%), Luciferase inhibitors (~5%), Auto-fluorescent (~5%), Reactive (~3%) and Colored (~3%) compound classes; G) Performance of an FP assay in robustness set compounds using dithiothreitol (DTT) and L-Cysteine as a reducing agent; H) Performance of a luminescence assay with and without 2% tertigitol NP-9 compared to HTRF assay in response to robustness set compounds.

**Figure 3)** ultra-High Throughput Screening (u-HTS) output within ELF consortium. A) The predicted and actual u-HTS production between 2013 and 2017 at PPSC; B) Distribution of data points per day across all programmes in 384-well and 1536-well plates; C) Number of plates screened per day for ELF consortium public target programmes and overall screening days; D) PCA analysis of 72 u-HTS campaigns. PC1 mainly captured the variation in number of compounds registered in QHL with higher numbers associated with cellular assays and PC2 was a function of screening days and well formats, with biochemical assays having less screening days and higher number of wells. Increasing size corresponds to increasing (log) active rate; E) PCA analysis of biochemical u-HTS campaigns. PC1 mainly captures number of compounds in QHL with Protein-RNA, most Protein-DNA and some Receptors having lower numbers and PC2 variations in volume and Z'. Increasing size corresponds to increasing (log) active rate;

F) PCA analysis of cellular u-HTS campaigns. PC1 is a function of compound numbers in QHL and PC2 mainly screening days. Increasing size corresponds to increasing (log) active rate; G) Number of submitted proposals (154), accepted proposals (88), u-HTS-ready assays developed (76), u-HTS campaigns (72) and follow-up work completed (72) and QHL list delivered (72).

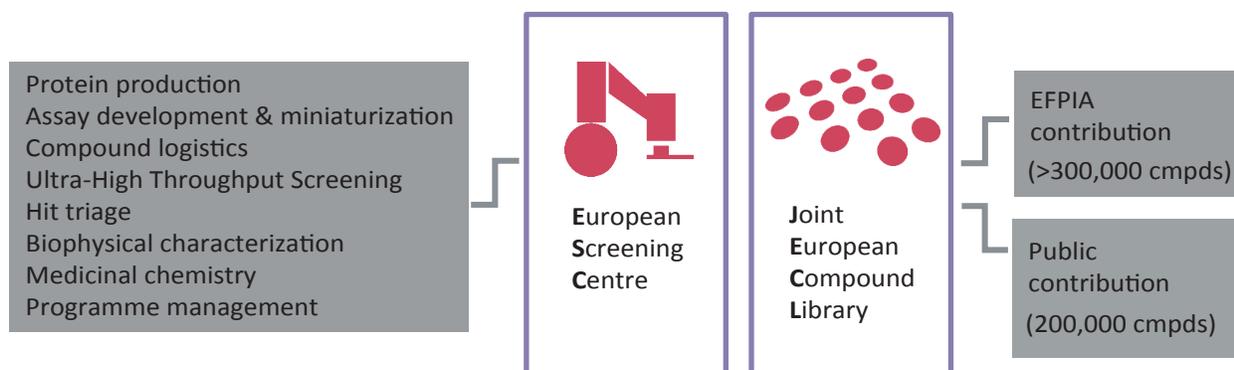
## References

1. Beck, E. M.; Parnell, E.; Cowley, A.; et al. Identification of A Novel Class of Benzofuran Oxoacetic Acid-Derived Ligands that Selectively Activate Cellular EPAC1. *Cells* **2019**, *8*, 1425.
2. De Vita, E.; Smits, N.; van den Hurk, H.; et al. Synthesis and Structure-Activity Relationships of N-(4-Benzamidino)-Oxazolidinones: Potent and Selective Inhibitors of Kallikrein-Related Peptidase 6. *ChemMedChem* **2020**, *15*, 79-95.
3. Mock, E. D.; Mustafa, M.; Gunduz-Cinar, O.; et al. Discovery of a NAPE-PLD inhibitor that modulates emotional behavior in mice. *Nat Chem Biol* **2020**, *16*, 667-675.
4. Laverty, H.; Orrling, K. M.; Giordanetto, F.; et al. The European lead factory—an experiment in collaborative drug discovery. *Journal of Medicines Development Sciences* **2015**, *1*, 20-33.
5. Giordanetto, F.; Jones, P.; Nelson, A.; et al. European Lead Factory. In *Comprehensive Medicinal Chemistry III*; 2017, pp. 505-519.
6. Karawajczyk, A.; Orrling, K. M.; de Vlieger, J. S.; et al. The European Lead Factory: A Blueprint for Public-Private Partnerships in Early Drug Discovery. *Front Med (Lausanne)* **2016**, *3*, 75.
7. Takebe, T.; Imai, R.; Ono, S. The Current Status of Drug Discovery and Development as Originated in United States Academia: The Influence of Industrial and Academic Collaboration on Drug Discovery and Development. *Clin Transl Sci* **2018**, *11*, 597-606.
8. Parrish, M. C.; Tan, Y. J.; Grimes, K. V.; et al. Surviving in the Valley of Death: Opportunities and Challenges in Translating Academic Drug Discoveries. *Annu Rev Pharmacol Toxicol* **2019**, *59*, 405-421.
9. Bentzien, J.; Bharadwaj, R.; Thompson, D. C. Crowdsourcing in pharma: a strategic framework. *Drug Discov Today* **2015**, *20*, 874-83.
10. McElroy, S. P.; Jones, P. S.; Barrault, D. V. The SULSA Assay Development Fund: accelerating translation of new biology from academia to pharma. *Drug Discov Today* **2017**, *22*, 199-203.
11. Besnard, J.; Jones, P. S.; Hopkins, A. L.; et al. The Joint European Compound Library: boosting precompetitive research. *Drug Discov Today* **2015**, *20*, 181-6.
12. Karawajczyk, A.; Giordanetto, F.; Benningshof, J.; et al. Expansion of chemical space for collaborative lead generation and drug discovery: the European Lead Factory Perspective. *Drug Discov Today* **2015**, *20*, 1310-6.
13. Morgentin, R.; Dow, M.; Aimon, A.; et al. Translation of innovative chemistry into screening libraries: an exemplar partnership from the European Lead Factory. *Drug Discovery Today* **2018**, *23*, 1578-1583.
14. Brown, D. G.; Bostrom, J. Where Do Recent Small Molecule Clinical Development Candidates Come From? *J Med Chem* **2018**, *61*, 9442-9468.
15. Jones, P.; McElroy, S.; Morrison, A.; et al. The importance of triaging in determining the quality of output from high-throughput screening. *Future Med Chem* **2015**, *7*, 1847-52.
16. Aldrich, C.; Bertozzi, C.; Georg, G. I.; et al. The Ecstasy and Agony of Assay Interference Compounds. *ACS Cent Sci* **2017**, *3*, 143-147.
17. Baell, J.; Walters, M. A. Chemistry: Chemical con artists foil drug discovery. *Nature* **2014**, *513*, 481-3.

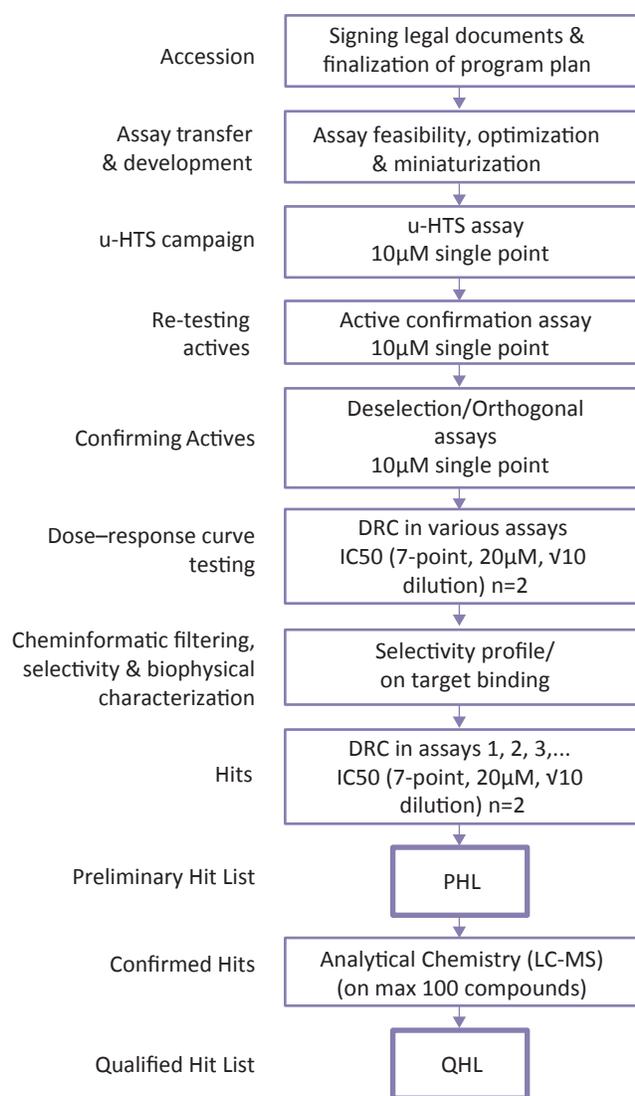
18. Rainard, J. M.; Pandarakalam, G. C.; McElroy, S. P. Using Microscale Thermophoresis to Characterize Hits from High-Throughput Screening: A European Lead Factory Perspective. *SLAS DISCOVERY: Advancing the Science of Drug Discovery* **2018**, *23*, 225-241.
19. Johnston, P. A. Redox cycling compounds generate H<sub>2</sub>O<sub>2</sub> in HTS buffers containing strong reducing reagents--real hits or promiscuous artifacts? *Curr Opin Chem Biol* **2011**, *15*, 174-82.
20. Paillard, G.; Cochrane, P.; Jones, P. S.; et al. The ELF Honest Data Broker: informatics enabling public-private collaboration in a precompetitive arena. *Drug Discovery Today* **2016**, *21*, 97-102.

Figure 1

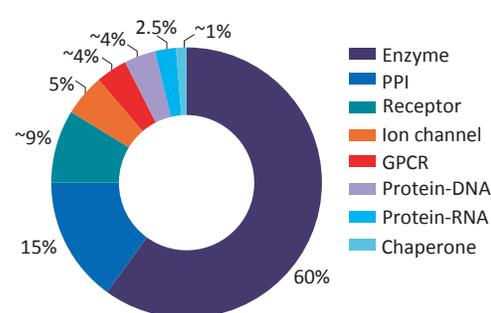
A) European Lead Factory consortium



B) Public target programme screening cascade



C) Target classes



D) Disease areas

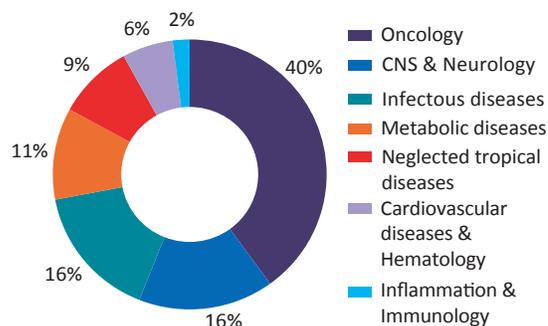


Figure 2

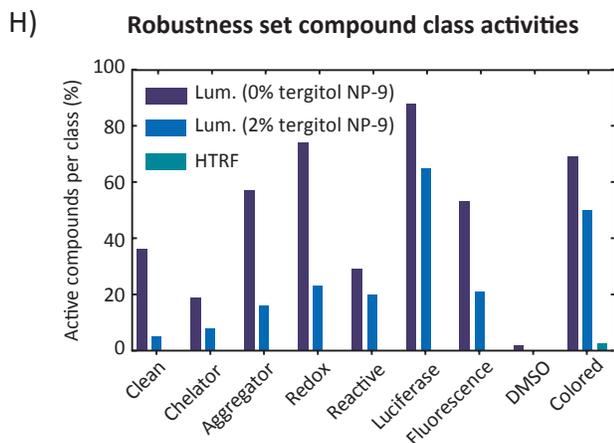
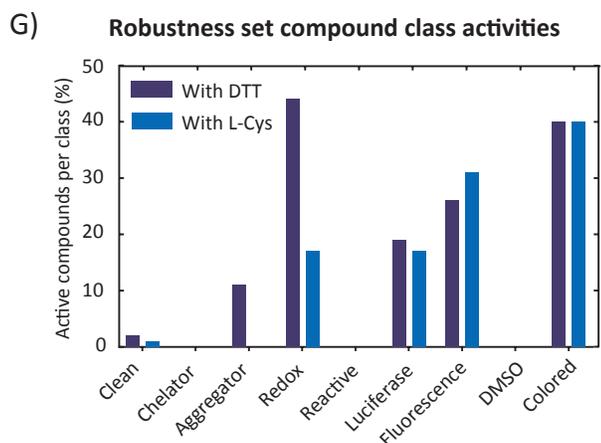
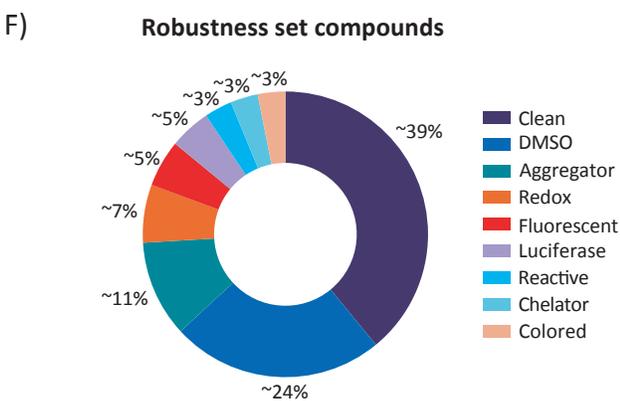
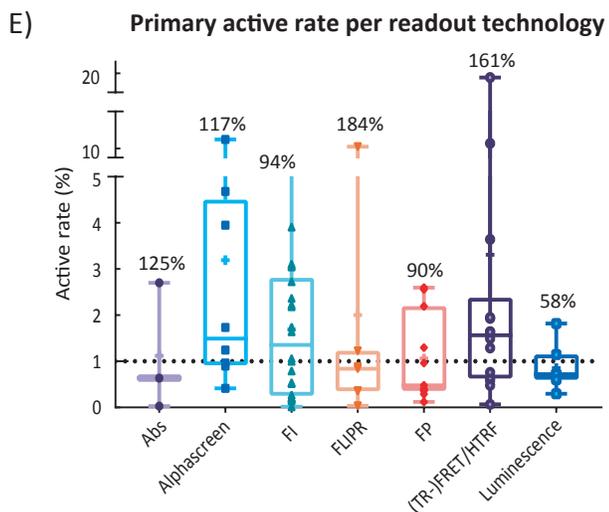
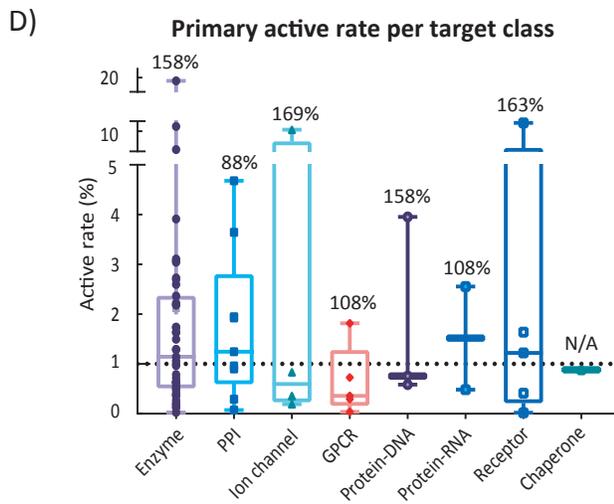
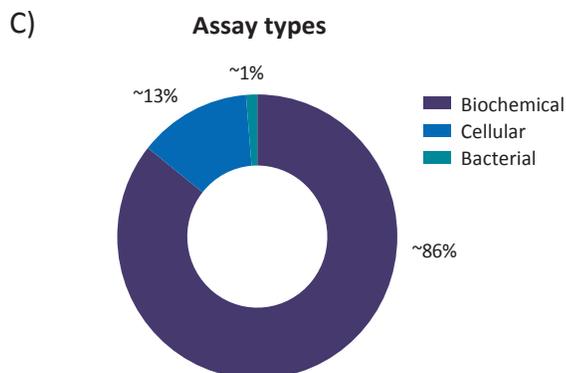
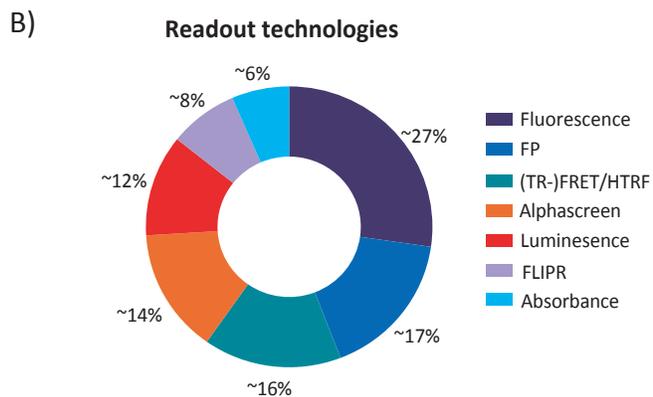
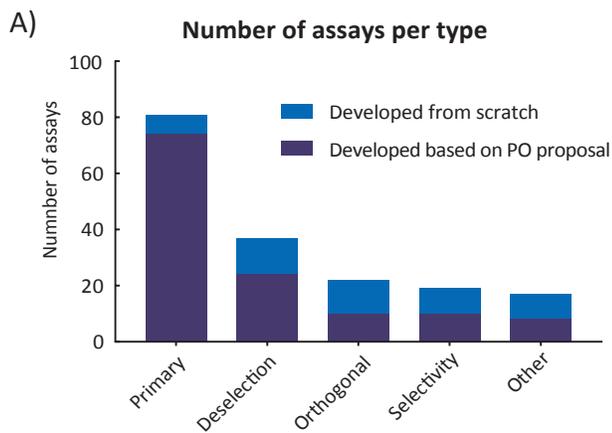
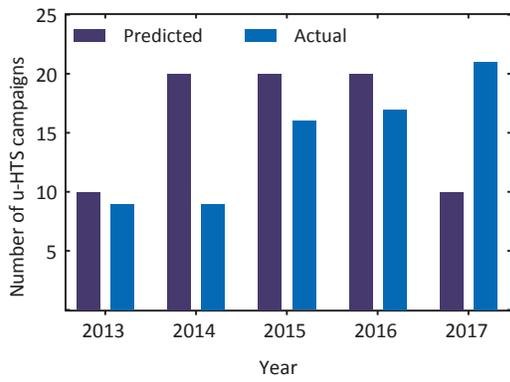
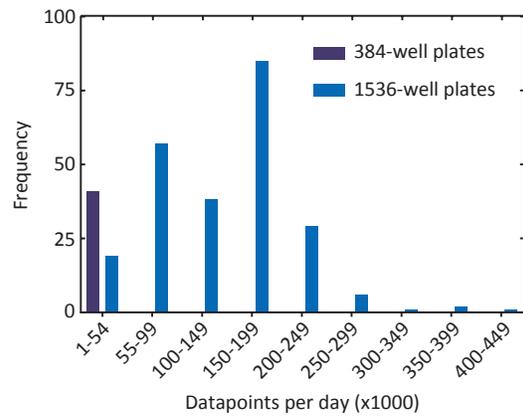


Figure 3

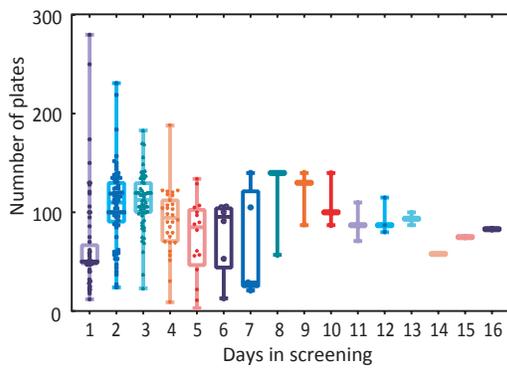
A) Annual u-HTS output



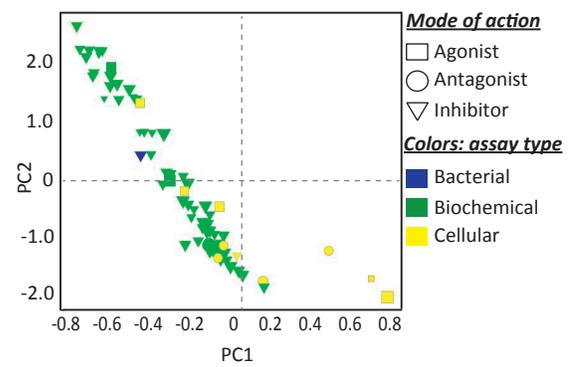
B) Distribution of datapoints per day



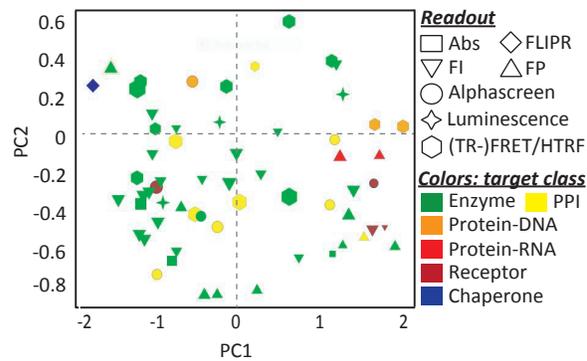
C) Number of plates per screening day



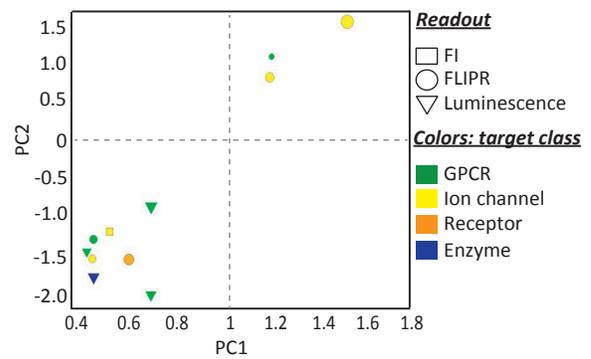
D) PCA for all programmes



E) PCA for biochemical programmes



F) PCA for cellular programmes



G) Public target programme screening results

