

Topic: Open access chemogenomics library and chemical probes for the druggable genome

All information regarding future IMI Call topics is indicative and subject to change. Final information about future IMI Calls will be communicated after approval by the IMI Governing Board.

Topic details

Action type	Research and innovation action (RIA)
Submission and evaluation process	2 stages

Background and specific challenges to be addressed

In biomedicine, discoveries arising from novel enabling technologies and reagents have garnered a quarter of the Nobel Prizes for chemistry and medicine in this century. Among the myriad of these transformative techniques and reagents, bibliometric evidence shows that pharmacological modulators (chemical and biological probes) have both the greatest scientific citation impact, the greatest sway on exploratory biomedical research, and provide the best mechanism to understand the relevance of a protein as a potential drug target [1][1]. Indeed, the field of drug discovery and the development of new molecular entities are predicated on the availability of sound mechanistic principles. Unfortunately, our understanding of human disease remains inadequate, and as a result clinical success rates for novel mechanisms remain low. Currently only one out of ten clinical drug candidates reaches the Open Access Chemogenomics Library and Chemical Probes for the Druggable Genome market after an average of 10 years and at a cost of at least EUR 2 billion in R&D expenses per drug.

Ultimately, the most effective method of dramatically improving the efficiency of R&D is to initiate studies on the 'right' target, and this is possible only if we dramatically increase our understanding of disease mechanisms. Experts agree that genetics and big data are promising approaches to select the right target, the appropriate biomarkers and the patients that are most likely to respond to any given treatment. However, this promise is a long way from reality in most cases; experience has shown there still remains a difficult path from prioritising a candidate gene with human genetics through to a successful R&D project [1][2].

We urgently need to close the gap between establishing a genetic link and the underlying disease mechanism for potential drug targets, and, to this end, we believe that there is an immediate need to design a set of open access (*i.e.* unencumbered and free of intellectual property restrictions) chemical compounds for the entire druggable genome. The set, which would comprise an openly accessible chemogenomics library and selected high-quality chemical probes will provide scientists across the world with the tools to interrogate and validate independently new candidate genes identified by modern genetic studies and bioinformatics in a variety of informative biological systems [1][3][1][4][1][5][1][6][1][7] among which advanced, patient-derived assays will be the most relevant.

Need and opportunity for public-private collaborative research

The creation of an open access set of tools with which to interrogate the entire druggable genome is a challenge of scale and breadth that cannot be solved by a single institution. Moreover, a partnership between EFPIA members, mid-sized companies, Associated Partner organisations, academia and SMEs will be essential to achieve this goal, as these institutions have complementary resources and expertise necessary for success. For example, industry has extensive medicinal chemistry and screening facilities, but these

capabilities must be used in the context of biological validation, most commonly performed in academia. Access to patient samples that are genotyped and accompanied by their clinical histories must be accessed by involving physician scientists in academic institutions. New technological and experimental approaches will also be needed to speed up the process of creating new tool compounds. We also believe that impact is dependent on open science because a pre-competitive, shared risk investment model will allow the partnership to provide the wider community access to the generated reagents quickly and with no strings attached, thus amplifying its impact. Moreover, open science also better assures adherence to high standards of quality and the reproducibility of results (a major issue in biological research), with concomitant increases in productivity and innovation. Finally, through a public-private partnership, both funding and expertise will be highly leveraged.

Scope

Currently, the druggable human genome is estimated to consist of at least 3 000 genes. This topic aims to generate potent, well-characterised, functional, small-molecule modulators for a significant number of these and, at the same time, lay the foundation for identifying a set of openly accessible (i.e. unencumbered from restrictions on use) tool compounds for the entire druggable human genome. With this set of chemical tools available, scientists will be poised to interrogate the latest findings emerging from big data approaches and human genetic studies, thus compressing time from gene discovery to target prioritisation, and ultimately to patient benefits. Importantly, although we imagine the consortium can make great progress by assembling and characterising pre-existing compounds into an initial chemogenomics set, this is not enough. It is imperative to fill the significant gaps by discovering and developing novel chemical tools/probes against under-studied proteins (or protein families) that may be involved in the initiation and progression of disease.

The overall aim of this topic will be as follows:

- establish a framework to assemble an open-access chemogenomics library for the druggable genome – namely a physical library supported by compound meta-data;
- further enrich the open access library by inventing new, deeply characterised chemical probes to selected specific protein families;
- develop open-access assays from well characterised human disease tissue with a special emphasis on immunology, oncology (including immune-oncology) and neuroscience to profile the chemical tools and chemical probes;
- establish sustainable infrastructure, with high priority on accessible platforms and appropriate governance, for prolonged discovery and dissemination of tool compounds, assays, and associated data, beyond the lifetime of this project;
- develop a communication plan to facilitate the dissemination of the compound sets and to ensure their appropriate use.

N.B. A chemogenomics library describes the use of target family-directed chemical libraries in target or cell-based assays as a means of accessing new areas of biology and accelerating drug discovery research based on the assumption that similar receptors bind similar ligands. Such sets, although containing compounds that individually do not fulfill the stringent criteria of a chemical probe, can still be used to interrogate multiple members of protein families to help prioritise the most therapeutically relevant ones that could then form the basis of a chemical probe project. In contrast, a chemical probe is a small molecule that modulates the function of a protein in a specific and selective way. The compound must exhibit a defined *in vitro* potency for a single target and possess a minimum 30-fold selectivity relative to other sequence-related proteins of the same family. Furthermore, the probe must be profiled against a standard selection of other unrelated, pharmacologically relevant targets and large protein families of relevance to drug discovery (specificity), and, finally, have demonstrated on-target effects in cells (cellular activity).

Expected key deliverables

The consortium will generate an open access chemogenomics library consisting of about 5 000 compounds that cover roughly 1 000 protein targets (i.e. one third of the current druggable genome). Here, the term open access includes not only the right to publish findings using these tools, but also includes the unencumbered and

pre-publication dissemination of the results, the tools themselves, the assay protocols, and all the associated data packages. This open access chemogenomics set will serve as a substantial head start on generating a library covering the entire druggable genome. In addition, the consortium will develop chemical probes for two to three jointly agreed target families with an initial focus on E3 ligases and solute carriers (SLCs), which may be carried out in partnership with existing IMI consortia, such as ReSolute. For this component, up to 100 novel, well characterised, high-quality chemical probes, as defined by leaders in chemical biology [1][5][1][6][1][7][1][8] are intended to be generated. To achieve this goal, the consortium will generate recombinant proteins, solve crystal structures, and establish all biochemical and cellular assays needed to ensure that the probes meet the established stringent quality criteria, including target engagement in cells [1][9]. Finally, the consortium will develop scientific and sociological mechanisms to extract biological and disease information from the chemogenomics libraries and chemical probes – and their targets. Given the technical issues that plague interpretation of data from established cell lines, we strongly believe that this will depend on accessing more relevant assays through which to profile the compounds. These assays must be shown to be reproducible, to be derived from genotyped and deeply phenotyped patient-derived tissue and the results to be made available broadly, so that biological data from all the assays can be combined and mined [1][10]. The partnership is expected to develop around 20 novel human tissue-derived assays in three major therapeutic areas of immunology, oncology (including immune-oncology) and neuroscience and test tool molecules and chemical probes in these assays.

Project success will require the partnership to establish enabling infrastructure to generate the probes. This includes cell and biochemical assay panels for characterisation of the compounds, including off-target analysis, a complementary database and a modern, scalable compound store and compound logistics. In addition, the partnership must explore new technologies in the field of fragment-based screening and high-throughput proteomics to accelerate the process of tool compound generation and characterisation. Finally, the consortium will set up platforms that permit broader access to these technologies both from within and outside the consortium – so the community can participate and help achieve, or exceed, the projects goals.

With this proposal, it is planned to lay the foundation on which to build and organise a worldwide network of laboratories to generate chemical tools for the entire druggable genome. Outstanding scientific leadership will be required in order to achieve this vision. It is also envisaged that this endeavor will help identify new scientific opportunities, identify and build strategic partnerships with other projects (e.g. patient groups, international consortia, other IMI projects), and promote truly open accessible science.

The chemogenomics library, the chemical probes, and the accompanying metadata (potency, selectivity, activity in cell-based assays) are intended to be open access, i.e. use of the compounds will be made available unencumbered, in a pre-publication state and free from restrictions on use. After quality control, assays and data generated by the consortium will also be made publicly available without restrictions. In addition, technologies developed throughout the project must be likewise non-exclusive, platform-oriented, ready for application, openly and freely available for use, dissemination, and commercialisation during and following the completion of the project. Finally, as part of a sustainability concept, the partnership will provide non-exclusive access to the synthetic routes of these compounds to large and/or mid-size vendors that are willing to distribute the chemogenomics library and/or chemical probes and their controls worldwide.

Expected impact

This project will provide the wider academic community with unencumbered access to the highest quality tool compounds for a large number of novel targets, and the expected impact should therefore be transformative. Presently, many companies and organisations are already in the process of setting up their own chemogenomics libraries. Although these have the potential to be phenomenal resources for the companies, their utility is also limited: they are not widely available to academia, they are likely to overlap, and each compound set is not as deeply characterised as could be managed within a larger, more focused, more resourced and more transparent project. By making a high-quality, broader compound set available, the consortium will seed a massive community target prioritisation and target deconvolution effort [1][2]. Moreover, in providing chemical tools without restriction, the consortium will also make available tools to invent new assays and unencumbered starting points for probe development or drug discovery. The consortium's centralised, cell-based and biochemical assay panels will serve as a resource for the entire chemical biology community. The ability to access these capabilities will provide significant incentives for external scientists to contribute innovative compounds to the network, thus expanding the impact with donated

resources. The cell and tissue platform with the high-quality, patient-derived cell assays will provide the opportunity for clinical scientists to undertake translational medical research and biomarker discovery, and will provide the roadmap for other clinical centers to access the libraries and make important translatable discoveries.

The availability of chemical probes to unprecedented targets will also open up exciting new research avenues. As an example, open access, novel E3 ligase binders will provide much needed starting points for the development of new protein-targeting chimeras (PROTAC). In addition, the research strategies undertaken in this topic may serve as a template for the technology development to expand the project to include tool generation for areas of the genome currently not considered as druggable.

The management and data infrastructure and the assembled global collaborative network will lay the foundation for unparalleled progress in providing high-quality, open-access tool compounds as a basis for reproducible research. Indeed, this topic has the potential to cause a fundamental shift towards a more open and pre-competitive approach to the costly field of target prioritisation and discovery without compromising the proprietary research models required in industry. Finally, the significant amount of freely accessible, high-quality data generated within this consortium will be a rich source for future analyses by data scientists. Artificial intelligence and machine learning applications using high-quality data on highly diverse compounds across many pharmacological mechanisms will spur research in new fields of biology and generate a source of targets for proprietary projects in various therapeutic areas.

Open access – additional dissemination obligation

Considering the specific nature of this topic, it foresees application of an additional dissemination obligation (IMI2 JU MGA art. 29.1). All results of this project will be made available to the scientific community by open access (i.e. unencumbered, pre-publication, and free from restriction on use). Open access parameters include not only the right to publish findings using these tools, but also the right to disseminate the tools, results, assay protocols, and all the associated data packages, including cell-based assays.

Potential synergies with existing Consortia

Applicants should take into consideration, while preparing their short proposal, relevant national, European (both research projects as well as research infrastructure initiatives), and non-European initiatives. Synergies and complementarities should be considered in order to incorporate past achievements, available data and lessons learnt where possible, thus avoiding unnecessary overlap and duplication of efforts.

Therefore, the synergies with the following past and ongoing IMI1 & IMI2 projects could be considered by the applicants:

- Unrestricted Leveraging of Targets for Research Advancement and Drug Discovery (ULTRA-DD): <https://www.imi.europa.eu/projects-results/project-factsheets/ultra-dd>
- Research Empowerment on Solute carriers (ReSOLUTE): <https://re-solute.eu/>
- Open PHACTS (terminated): <https://www.imi.europa.eu/projects-results/project-factsheets/open-phacts>

Please note that during the project implementation phase the applicants could also consider other potential knowledge generated by the forthcoming projects under IMI2 JU:

- European Screening Centre: unique library for attractive biology (ESCuLab): https://www.imi.europa.eu/sites/default/files/uploads/documents/apply-for-funding/open-calls/IMI2_Call12_CallText.pdf
- Development of a platform for federated and privacy-preserving machine learning in support of drug discovery: https://www.imi.europa.eu/sites/default/files/uploads/documents/apply-for-funding/call-documents/imi2/IMI2%20CALL%2014%20TOPICS%20TEXT_EN.PDF

Synergies with the following European and international initiatives could also be highly relevant:

- Structural Genomics Consortium (SGC, <https://www.thesgc.org/>) that has in depth expertise concerning all aspects of this proposal;
- US National Institutes of Health initiative, Illuminating the Druggable Genome (<https://ncats.nih.gov/idg>), which will provide the bioinformatics tools to help improve the understanding of the properties and functions of proteins that are currently not well studied within commonly drug-targeted protein families;
- European Bioinformatics Institute (EMBL-EBI, <https://www.ebi.ac.uk/>) for data handling and analyses;
- Open Targets (<https://www.opentargets.org/>) for target identification;
- H2020 iNEXT consortium (https://cordis.europa.eu/project/rcn/194892_fr.html) for fragment screening;
- H2020 OpenRiskNet (https://cordis.europa.eu/project/rcn/206759_fr.html) for compound design;
- ERIC EU-OPENSREEN (www.eu-openscreen.eu) for screening;
- ESFRI-consortium ELIXIR (www.elixir-europe.org) for sustainable infrastructure for biological information;
- ERIC INSTRUCT (www.instruct-eric.eu) for structural biology infrastructure.

Industry consortium

The industry consortium is composed of the following EFPIA companies:

- Bayer (project co-lead)
- Boehringer Ingelheim (project co-lead)
- Pfizer
- Servier
- Takeda

As part of this endeavour, each pharmaceutical industry partner is willing to contribute at least 10 high-quality chemical probe compounds from their current or previously terminated R&D projects; at least 50 chemogenomics tool compounds from their own compound collections; and support the development of a minimum of 5 chemical probes by in-kind (especially chemistry).

In addition, the industry consortium includes the following IMI2 Associated Partners:

- Diamond Light Source (UK)
- Ontario Institute for Cancer Research (Canada)
- The Montreal Neurological Institute at McGill University (Canada)
- The Royal Institute of Technology (Sweden)

Indicative duration of the action

The indicative duration of the action is 60 months.

Indicative budget

The indicative in-kind and financial contribution from EFPIA partners and IMI2 JU Associated Partners is EUR 30 257 000.

This contribution comprises an indicative EFPIA in-kind contribution of EUR 23 800 000, of which EUR 9 930 000 financial contribution to the beneficiaries receiving JU funding in the selected action and an indicative IMI2 JU Associated Partners in-kind contribution of EUR 6 457 000.

Due to the global nature of the participating industry partners, it is anticipated that some elements of the contributions will be non-EU/H2020 Associated Countries in-kind contributions.

The financial contribution from IMI2 JU is a maximum of EUR 27 935 000.

Applicant consortium

The applicant consortium (academic groups and SMEs) is expected to demonstrate expertise, leadership and a proven track record in all scientific areas addressed in the topic, including:

- adherence to open-access principles, and expertise in developing and managing open-access projects, which are essential to provide unencumbered and pre-publication access to the chemogenomics tools, chemical probes, patient-derived assays, and associated data packages to the scientific community free of any restrictions on use;
- expression, characterisation and structure determination of soluble proteins, integral membrane proteins, and protein complexes in an integrated project at large scale;
- assay development across a large number of different proteins and protein classes, including cell-based target engagement assays;
- screening compound libraries at scale, using a variety of approaches including high-throughput, focused, computational, fragment and DNA-encoded libraries;
- 'hit-to-probe' capabilities, including using structure-guided methods to improve efficiency and systematic characterisation in relevant biophysical, biochemical, and especially cellular/phenotypic assays;
- strategies to systematically map the knowledge space of protein families, including developing computational approaches and physical reagents to facilitate cross-screening;
- generation, characterisation and dissemination of chemogenomics libraries, including systematic compound characterisation *in vitro* and in cells;
- establish quality-control metrics and both demonstrate and record their use in practice, including mechanisms to solicit independent input into quality;
- development of innovative technologies to speed up the generation of tool compounds, and innovative approaches to accelerate their wider adoption in academia and industry;
- development of strategies to ensure that chemical probes are appropriately used by the community in biological assays;
- track record of scientific success in partnerships with clinical centres, and success in managing contracts and ethical issues;
- track-record of obtaining project-specific ethical approvals for clinical research collaborations;
- using patient-derived samples to advance drug discovery in close cooperation with industrial partners, including the development of novel assays.

SMEs can be of great benefit to IMI2 JU actions and can strengthen the competitiveness and industrial leadership of the European Union. Their involvement in the action might offer a complementary perspective to industry and the academia, and help deliver the long-term impact of the funded action. For these reasons, applicants should consider engaging SMEs throughout the proposal. For example, under this topic, the contribution of SMEs would be considered beneficial for broad profiling of chemogenomics compounds and chemical probes.

Members of the applicant consortium are also expected to demonstrate excellence and a proven record of accomplishment (evidenced by collaborative publications) in establishing networks of recognised thought leaders in all relevant sectors indicated in the topic, including:

- a global network that spans medicinal chemistry, biological assays, human biology, experimental medicine and clinical research;
- proven track record of achieving high-value/high-impact outcomes catalysing research in pioneer target areas of drug discovery;
- ability and history of leveraging additional funds with diverse and international organisations, including patient groups, foundations, philanthropy and SMEs;
- history of making research output widely available and evidenced commitment to open science principles;
- mechanism in place to efficiently and effectively disseminate chemical and biological research materials (e.g. chemical probes, protein constructs, antibodies).

Members of the applicant consortium are expected to have successfully collaborated with a network of scientific researchers especially with industry and should demonstrate:

- previous impact on launching or adding value to internal drug discovery projects in the pharmaceutical industry;
- previous impact on providing the foundation for experimental medicine studies in the public sector;
- previous success in collaborations among networks of academics and SMEs – as evidenced through shared projects and co-authored publications;
- previous success in governing and managing large projects, including e.g. finance, intellectual property and inter-institutional contracts;
- a track record of consistently achieving (or even exceeding) milestones and deliverables on time and within budget;
- a track record of making new technologies widely available, for example as generally accessible platforms or commercial products;
- experience in managing varying interests of multiple stakeholders.

Suggested architecture of the full proposal

The applicant consortium should submit a short proposal that includes their suggestions for creating the full proposal architecture, taking into consideration the industry contributions and expertise provided below.

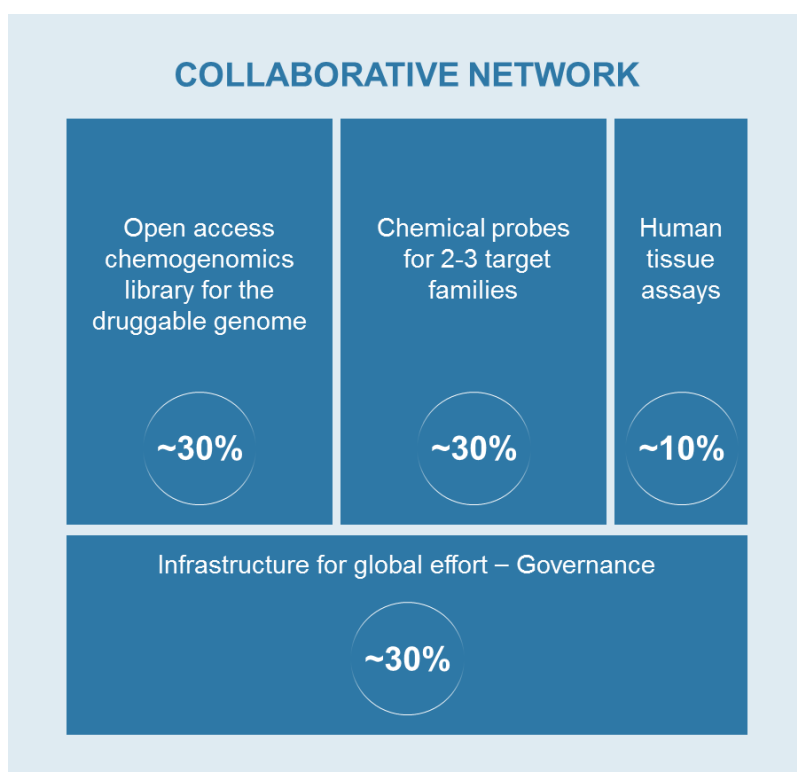
In the spirit of the partnership, and to reflect how IMI2 JU call topics are built on identified scientific priorities agreed together with EFPIA beneficiaries/large industrial beneficiaries, these beneficiaries intend to significantly contribute to the programme and project leadership as well as project financial management.

The final architecture of the full proposal will be defined by the participants in compliance with the IMI2 JU rules and with a view to the achievement of the project objectives. The allocation of a leading role within the consortium will be discussed in the course of the drafting of the full proposal to be submitted at stage 2. To facilitate the formation of the final consortium, until the roles are formally appointed through the consortium agreement, the proposed project leader from among EFPIA beneficiaries/large industrial beneficiaries shall facilitate an efficient negotiation of project content and required agreements. All beneficiaries are encouraged to discuss the project architecture and governance and the weighting of responsibilities and priorities therein.

The full consortium will define project aspects such as governance, guiding principles and project plan. The architecture below for the full proposal is a suggestion.

A plan for aspects related to sustainability, facilitating continuation beyond the duration of the project should also be proposed.

The planned endeavour consists of four parallel pillars that include an underlying sustainable network infrastructure. The expected resource distribution to the four pillars is indicated below (% of overall resources):



Pillar 1 – Open-access chemogenomics library for the druggable genome

In the first pillar, it is planned to establish a chemogenomics library consisting of compounds meeting predefined target-specific criteria (biochemical activity, selectivity, physico-chemical profile sufficient for cell-based assays, evidence of cellular target engagement and no general cytotoxicity). To ensure transparency and quality, acceptance of compounds into this library will be governed by an arm's length committee of independent experts from academia. The acceptance criteria for Pillar 1 compounds (e.g. selectivity, potency) will not be as stringent as for chemical probes (Pillar 2 compounds), but because of the extensive annotation, Pillar 1 compounds will be very valuable for target prioritisation, target deconvolution, and as starting point for chemical probes. They will also enable a fast-track approach to develop drug leads for exciting new targets.

The following work packages are planned to achieve Pillar 1 goals:

Work package 1 – Collection of available compounds from academia and industry

Deliverables: 30-100 mg of pure material for ~2 000 compounds identified and collected from the following sources meeting predefined criteria:

- compounds (1 000-2 000 compounds covering a variety of targets) identified and collected from known literature compounds;
- inclusion of openly available chemogenomics compound sets that fulfil stringent quality criteria;
- acquisition of compounds provided by participating pharmaceutical companies (at least 50 per participating EFPIA partner);
- compounds for selected target families generated within this IMI2 project (see Pillar 2);

- establish an independent review mechanism to assess the quality of the compound to be included in the set.

Work package 2 – Annotation of library compounds

Deliverables:

- data packages necessary to enable use of the compounds in pre-clinical studies by scientists globally;
- making data available to the scientific community via a publicly accessible database (see Pillar 4, work package 10), either generated within the consortium, or in collaboration with an existing public partnership.

Work package 3 – New methods for chemogenomics compound generation and profiling

Deliverables:

- protocols for novel and broadly applicable assay principles for biochemical, biophysical and cell-based assays to speed up generation of chemogenomics compounds and their characterisation;
- broadly applicable, novel technological and experimental approaches with the potential to speed up the hit-to-probe process by more than 6 months;
- generation of the remaining 2 000 to 3 000 compounds needed to cover one third of the druggable genome (assumption: 5 000 compounds needed to cover 1 000 targets, i.e. 1/3 of the druggable genome);
- establishing the strongest possible chemogenomics open source network of collaborations, allowing efficient sourcing of externally generated high-quality chemogenomics compounds and sets.

Industry and Associated Partner contribution to Pillar 1:

- provide at least 50 chemogenomics compounds per EFPIA partner from proprietary compound collections as open access chemogenomics compounds; solid material for testing;
- compound profiling in established assay panels that are available within the companies;
- access (free of charge) to the Diamond Light Source (Associated Partner) platforms for external groups contributing to the project deliverables;
- membership in scientific and decision-making committees (e.g. definition of target family-specific criteria and assessment of candidate compounds).

Expected applicant consortium contribution to Pillar 1:

- develop tools to identify chemogenomics compounds from patents, scientific publications and other sources;
- synthesis to provide solid material of chemogenomics compounds for testing;
- provide compound profiling to confirm that they meet the agreed upon criteria;
- experience in high-throughput, fragment-based screening;
- experience in covalent-ligand chemoproteomics approaches;
- assemble the remaining 2 000 to 3 000 chemogenomics compounds to cover one third of the druggable genome via internal activities or through collaborations and/or crowdsourcing;

- characterise selected compounds by determining three-dimensional protein-small molecule complex structures to high resolution and accuracy;
- membership in scientific and decision-making committees (definition of target family specific criteria; assessment of candidate compounds);
- manage independent peer-review mechanism to assess suitability of compounds for inclusion in the set.

Pillar 2 – Chemical probes for 2 – 3 emerging target families

For the second pillar, the aim will be to generate chemical probes for proteins within a minimum of 2-3 priority target families of high therapeutic interest. The initial priority will be on E3 ligases and solute carriers (SLCs), although we will not limit the scope to these target families. E3 ligases and SLCs were selected as initial priority families due to their high therapeutic importance, coupled with their relative intractability. This combination makes them ideally suited for a consortium-based approach for developing high-quality chemical tools. The inclusion of other target families will be agreed jointly by the consortium.

The number of chemical probes generated per family will depend on a number of considerations, including the strength of the genetic links to disease, experimental feasibility, therapeutic potential, as well as the number of members of the target family class. While working on probe projects, the consortium will also generate compounds that may not meet the stringent probe criteria but will be valuable as chemogenomics compounds thereby enriching the collection described in Pillar 1.

The following work packages are planned to achieve Pillar 2 goals:

Work package 4 – Protein production

Deliverables:

Validated protein expression clones, protein purification protocols, recombinant proteins for assay development and for 3D-structure determination; recombinant antibodies to facilitate assay development.

Work package 5 – Assay development for target engagement

Deliverables:

- protocols for target-specific biophysical, biochemical and cell-based assays and use of those assays for probe generation and confirming target engagement;
- genetically-engineered cell lines (such as knockout cells) to inform on target selectivity.

Work package 6 – Structure determination and chemical starting matter

Deliverables:

Generation of protein structures necessary to support probe generation. Depending on the target families selected, protein structures for both soluble and membrane proteins will be required. All protein structures generated in this project will be deposited in the Protein Data Bank (<https://www ww pdb.org/>). In addition, fragment screens will be conducted to identify starting points for probe generation.

Work package 7 – Generation of chemical probes

Deliverable:

At least 100 well-characterised chemical probes, meeting stringent criteria for potency, selectivity and with demonstrated on-target effects in cells. This WP includes medicinal chemistry capabilities.

Work package 8 – Technology development

Deliverables:

Development of transferrable technologies for broadly applicable methods to speed up probe development and characterisation along the whole value chain from target selection to probe characterisation. This will include (but not exclusively) automation approaches, cloud-computing platforms, algorithms, parallelisation, reagents, devices, protocols and documentation.

Industry and Associated Partner contribution to Pillar 2:

- design and access to fragment or other bespoke libraries;
- access to larger compound screening collections;
- high-throughput screening (HTS) or focused screens to identify hits;
- crystal-based fragment screening at Diamond Light Source (Associated Partner);
- access (free of charge) to the Diamond platforms for external groups contributing to the project deliverables;
- expertise in triage and validation of screening hits;
- design and synthesis of research chemical probes;
- medicinal chemistry to optimise hits;
- protein expression and purification for selected priority targets;
- assays (e.g. selectivity screening panels) and structure determination to support probe development;
- establish quantitative chemical probe criteria, in conjunction with the applicant consortium.

Expected applicant consortium contribution to Pillar 2:

- ability to access chemical libraries from leading academic chemists and chemical biologists;
- small- and medium-scale screening of EFPIA partner chemical libraries;
- secondary biochemical screens to validate and prioritise hits;
- off-target biochemical and cell-based screens;
- crystallographic fragment screening and protein-ligand structure determination to support probe development;
- design and synthesis of chemical probes;
- medicinal chemistry to optimise hits;
- assays (e.g. selectivity screening panels) and 3D-structure determination to support probe development;
- high-throughput cloning, expression, purification, and novel 3D structure solution (if necessary);
- established quantitative chemical probe criteria, in conjunction with industry;

- technology development to improve quality and speed up the development and dissemination of chemogenomics compounds and of chemical probes.

Pillar 3 – Human tissue assays

All chemical probes and selected chemical tools will be subjected to unbiased phenotypic screening in patient-cell-derived assays for target validation in the human disease context. Specific assays will be in new and emerging areas of immunology (e.g. inflammatory bowel disease, scleroderma, interstitial lung disease, lupus, arthritis, and fibrosis in different organs), oncology (including immuno-oncology) or neurological areas (e.g. neuroinflammation and neurodegeneration). Assays developed previously in the Ultra-DD and other IMI programs might be leveraged and expanded to include new and emerging areas of research.

The following work packages are planned to achieve Pillar 3 goals:

Work package 9 – Human tissue assays

Deliverables:

- develop at least 20 novel human tissue-derived assays in three major therapeutic areas of immunology, oncology (including immune-oncology) and neuroscience. Selected established high-quality translational assays will be optimised and miniaturised and others (such as more complex co-culture systems) will be developed within the project;
- validate these assays by using tool molecules and test chemical probes, including gold-standard positive and negative controls. The cell-based assays will be derived from human material, such as blood and tissue biopsies;
- when possible and if scientifically appropriate, the consortium will convert primary cells into a renewable resource, such as human stem cells and spheroids as well as organoids;
- both primary and stem-cell derived cells will be deeply characterised phenotypically, and to the extent possible within the funding frame, also characterised by deep -omics technologies.

Industry and Associated Partner contribution to Pillar 3:

- contribute high-quality compounds for screening in these biological assays;
- provide scientific expertise and advice to support setup and develop the human tissue assays (including details on protocols, throughput formats and patient-genetic stratification for sample collection as needed);
- access to patient-derived assays for neurodegeneration;
- profile the compounds emerging from Pillars 1 and 2 above into assays and generate target validation data packages collaboratively with the consortium partners.

Expected applicant consortium contribution to Pillar 3:

- network of target and disease experts to profile each probe in disease-relevant assays (e.g. immunology, cancer and neurology);
- access to patient-derived human material (fluids, blood, tissue, other);
- ethical and legal frameworks to engage in such collaborations;
- strategies to include genotyping and deep phenotyping of patient-derived cells and tissue;

- mechanism to characterise probes in other consortia with panels of cell-based assays, e.g. Sanger Institute (<https://www.sanger.ac.uk/>)
NCI panel (https://ntp.cancer.gov/discovery_development/nci-60/cell_list.htm)
BTCure (<https://www.imi.europa.eu/projects-results/project-factsheets/btcure>)
STEMBANCC (<https://www.imi.europa.eu/projects-results/project-factsheets/stembancc>);
- engage additional collaborators who are leading the field in functional cell assays and disease models for particular targets;
- mechanism to access additional, relevant phenotypic assay panels in priority areas.

Pillar 4 – Infrastructure and governance to lay the foundation for a global effort on the whole druggable genome

To establish an efficient and coordinated effort within this project, an additional goal of the fourth pillar is to align this project with similar projects or individual efforts globally, in order to reduce duplication of effort and to leverage the IMI investment. The consortium will work with global efforts to adopt such standards, and to this end, will establish or implement standardised, broad cell-based and biochemical assay panels to characterise chemogenomics compounds and chemical probes. The consortium will also establish a database for all data generated, as well as a central compound store and compound logistics (e.g. via a contract service organisation). The intention is for the chemogenomics library and the corresponding sustainable infrastructure to form the nucleus of a coordinated, worldwide, open-access effort to put together a reference compound library covering the entire druggable genome. Consequently, an important aspect of this project will be to provide leadership and a governance structure for the network, which will include investigators not only within the IMI consortium but also from complementary projects around the world. Although many members of the network will be working independently, and with independent funding, the aim is for all partners within the network to follow a jointly agreed masterplan to maximise synergies. In addition to the network, the consortium will find innovative ways to add compounds to the library, including e.g. setting up competitions for young scientists via crowdsourcing to add to the project deliverables. In summary, it is essential for the consortium to develop an international partnership comprising screening centres and chemical biologists around the world. We plan to encourage open-access publication of the results of the research in open-access scientific journals, help create platforms to share results, and work with commercial vendors to make the physical samples of tool compounds available for years to come to the biomedical community.

The following work packages are planned to achieve Pillar 4 goals:

Work package 10 – Infrastructure and platforms

Deliverables:

- compound logistics to handle distribution of all chemogenomics compounds and probes as well as compound exchange between partners;
- more than 500 assays established/accessible to annotate chemogenomics compounds and probes generated within this project with a potential to test compounds from network;
- easily accessible database containing all data generated within this project with a potential to hold data from related endeavours; format suitable for chemists and biologists; these data will be generated and made accessible according to FAIR (findable, accessible, interoperable, reusable) principles;
- long-lived platforms and transferrable infrastructure (e.g. open source code, commercially available services, cloud-hosted servers) to make the new technologies available to a wide audience beyond the consortium;
- open access and dissemination framework established.

Work Package 11 – Global framework

Deliverables:

- the framework for a global network with partners around the world that work on related goals, established with a governance structure that supports efficient collaboration and sustainability;
- partnership agreements with major European and international efforts in screening assay development; patient-derived cell assays, chemical screening, chemical probe generation and compound profiling;
- a process for recruitment and rigorous triage of external activity and contributions.

Work package 12 – Project management

Deliverables:

A management and governance structure which ensures that the project completes all deliverables in a timely and efficient manner.

Industry and Associated Partner contribution to Pillar 4:

- director or senior scientist/manager to represent company on joint steering committee;
- experts in drug discovery to manage collaborations in specific scientific areas or on specific targets/target families;
- contributions to collaborative scientific meetings, management of internal versus external activities;
- advice, involvement or secondment on infrastructure development, e.g. compound management, database, platform technologies, partnering opportunities and governance framework.

Expected Applicant consortium contribution to Pillar 4:

- experienced managers to ensure that the key consortium deliverables are completed;
- senior scientists to manage project deliverables, to disseminate the project outputs and to engage in collaborations to maximise impact;
- database, loader and visualisation tools to enable open access use of all data generated in this project and within related initiatives; partnering with public databases (e.g. ChEMBL) if possible;
- development of compound logistics for this project and for related initiatives to enable easy access to the chemogenomics compounds and the chemical probes, e.g. in collaboration with established SMEs;
- management of finance, valuation of deliverables, communication etc;
- create international alliance of screening, probe generation and compound profiling initiatives and align toward consortium objectives;
- dissemination of results in the form of publications, meeting presentations, and via the consortium's website;
- screening assays for broad profiling, e.g. broad panels for kinases, G-protein-coupled receptors (GPCRs); assays for cell permeability and unspecific toxicity;
- plan for sustainability of infrastructure after the end of this project, e.g. via partnering with contract research organisations (CROs), national facilities and vendors.

Sustainability

Sustainability measures beyond the duration of the proposed action should be considered based on the expected results generated by the action. At stage 2 the full consortium would have to propose a sustainability plan to be implemented during the project duration and including relevant resources and budget. Sustainability is of utmost importance for this project.

- The chemogenomics library and the chemical probes, and the existing and new data from their use, should be easily accessible on a continuous basis. The applicant consortium should have a convincing plan how to achieve this, e.g. via non-exclusive access to the synthetic routes of these compounds to vendors be they large, mid-size or even start-ups, that are willing to distribute the chemogenomics library and/or chemical probes and their controls.
- A significant investment in hardware, software and expertise for compound logistics, database and assay panels will be needed to make this project a success. To make best use of the investment, the applicant consortium should already have an initial plan for sustainability.
- This project is planned as part of a global initiative for creating an open-access chemogenomics library for the entire druggable genome. This ambitious goal, which is beyond the scope of this particular call, will not be achieved within the timeframe of this IMI project, thus, sustainability of the infrastructure and platforms is of utmost importance for the overall mission.

The applicant consortium should already have an initial plan for sustainability, e.g. via CROs that are interested to continue operations as part of their business or via letters of intent from universities or other research organisations concerning the continued use of the research tools and the infrastructure. A detailed plan will be developed and implemented within the project.

References

- [1] Edwards *et al*, Nature 470, 163, 2011
- [2] K. Asadullah *et al*, Nat. Rev. Drug. Discov., 2011, 10, 712-713
- [3] M.E. Bunnage, E.L.P. Chekler, L.H. Jones, Nat. Chem. Biol. 2013, 9, 195-199
- [4] S.V. Frye, Nat. Chem. Biol. 2010, 6, 159-161
- [5] C.H. Arrowsmith *et al*, Nat. Chem. Biol. 2015, 11, 536-541
- [6] P. Workman & I. Collins, Chem. & Biol. 2010, 17, 561-577
- [7] S. Müller, S. Ackloo, ChL Arrowsmith, M. Bauser J.L. Baryza, J. Blagg, *et al.*, Donated Chemical Probes For Open Science, eLife 2018;7:e34311
- [8] J. Blagg & P. Workman, Cancer Cell 2017, 32, 9-25
- [9] J. Baell, M. A. Walters, Nature 2014, 513, 481-483
- [10] C.H. Arrowsmith *et al*. Nat. Rev. Drug. Disc. 2015, 11(8), 536-41