

## Prioritizing multiple therapeutic targets in parallel using automated DNA-encoded library screening

Machutta, Carl A; Kollmann, Christopher S; Lind, Kenneth E; Bai, Xiaopeng; Chan, Pan F; Huang, Jianzhong; Ballell, Lluís; Belyanskaya, Svetlana; Besra, Gurdyal S; Barros-Aguirre, David; Bates, Robert H; Centrella, Paolo A; Chang, Sandy S; Chai, Jing; Choudhry, Anthony E; Coffin, Aaron; Davie, Christopher P; Deng, Hongfeng; Deng, Jianghe; Ding, Yun

DOI:

[10.1038/ncomms16081](https://doi.org/10.1038/ncomms16081)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Machutta, CA, Kollmann, CS, Lind, KE, Bai, X, Chan, PF, Huang, J, Ballell, L, Belyanskaya, S, Besra, GS, Barros-Aguirre, D, Bates, RH, Centrella, PA, Chang, SS, Chai, J, Choudhry, AE, Coffin, A, Davie, CP, Deng, H, Deng, J, Ding, Y, Dodson, JW, Fosbenner, DT, Gao, EN, Graham, TL, Graybill, TL, Ingraham, K, Johnson, WP, King, BW, Kwiatkowski, CR, Lelièvre, J, Li, Y, Liu, X, Lu, Q, Lehr, R, Mendoza-Losana, A, Martin, J, McCloskey, L, McCormick, P, O'Keefe, HP, O'Keefe, T, Pao, C, Phelps, CB, Qi, H, Rafferty, K, Scavello, GS, Steinginga, MS, Sundersingh, FS, Sweitzer, SM, Szewczuk, LM, Taylor, A, Toh, MF, Wang, J, Wang, M, Wilkins, DJ, Xia, B, Yao, G, Zhang, J, Zhou, J, Donahue, CP, Messer, JA, Holmes, D, Arico-Muendel, CC, Pope, AJ, Gross, JW & Evindar, G 2017, 'Prioritizing multiple therapeutic targets in parallel using automated DNA-encoded library screening', *Nature Communications*, vol. 8, 16081. <https://doi.org/10.1038/ncomms16081>

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

Download date: 12. May. 2021

ARTICLE

Received 3 Sep 2016 | Accepted 24 May 2017 | Published 17 Jul 2017

DOI: 10.1038/ncomms16081

OPEN

# Prioritizing multiple therapeutic targets in parallel using automated DNA-encoded library screening

Carl A. Machutta<sup>1,\*</sup>, Christopher S. Kollmann<sup>2,\*</sup>, Kenneth E. Lind<sup>2</sup>, Xiaopeng Bai<sup>2</sup>, Pan F. Chan<sup>1</sup>, Jianzhong Huang<sup>1</sup>, Lluís Ballell<sup>3</sup>, Svetlana Belyanskaya<sup>2</sup>, Gurdyal S. Besra<sup>4</sup>, David Barros-Aguirre<sup>3</sup>, Robert H. Bates<sup>3</sup>, Paolo A. Centrella<sup>2,†</sup>, Sandy S. Chang<sup>1</sup>, Jing Chai<sup>2</sup>, Anthony E. Choudhry<sup>1</sup>, Aaron Coffin<sup>2,†</sup>, Christopher P. Davie<sup>2</sup>, Hongfeng Deng<sup>2</sup>, Jianghe Deng<sup>1</sup>, Yun Ding<sup>2</sup>, Jason W. Dodson<sup>1</sup>, David T. Fosbenner<sup>1</sup>, Enoch N. Gao<sup>1</sup>, Taylor L. Graham<sup>1</sup>, Todd L. Graybill<sup>1</sup>, Karen Ingraham<sup>1</sup>, Walter P. Johnson<sup>1</sup>, Bryan W. King<sup>1</sup>, Christopher R. Kwiatkowski<sup>1</sup>, Joël Lelièvre<sup>3</sup>, Yue Li<sup>1</sup>, Xiaorong Liu<sup>2</sup>, Quinn Lu<sup>1</sup>, Ruth Lehr<sup>1</sup>, Alfonso Mendoza-Losana<sup>3</sup>, John Martin<sup>1</sup>, Lynn McCloskey<sup>1</sup>, Patti McCormick<sup>1</sup>, Heather P. O'Keefe<sup>2</sup>, Thomas O'Keefe<sup>2</sup>, Christina Pao<sup>1,†</sup>, Christopher B. Phelps<sup>2</sup>, Hongwei Qi<sup>1</sup>, Keith Rafferty<sup>1</sup>, Genaro S. Scavello<sup>1</sup>, Matt S. Steingina<sup>1</sup>, Flora S. Sundersingh<sup>2</sup>, Sharon M. Sweitzer<sup>1</sup>, Lawrence M. Szewczuk<sup>1,†</sup>, Amy Taylor<sup>1</sup>, May Fern Toh<sup>1,†</sup>, Juan Wang<sup>1</sup>, Minghui Wang<sup>1</sup>, Devan J. Wilkins<sup>2,†</sup>, Bing Xia<sup>2</sup>, Gang Yao<sup>2</sup>, Jean Zhang<sup>2</sup>, Jingye Zhou<sup>2,†</sup>, Christine P. Donahue<sup>2</sup>, Jeffrey A. Messer<sup>2</sup>, David Holmes<sup>1</sup>, Christopher C. Arico-Muendel<sup>2</sup>, Andrew J. Pope<sup>1</sup>, Jeffrey W. Gross<sup>1</sup> & Ghotas Evindar<sup>2</sup>

The identification and prioritization of chemically tractable therapeutic targets is a significant challenge in the discovery of new medicines. We have developed a novel method that rapidly screens multiple proteins in parallel using DNA-encoded library technology (ELT). Initial efforts were focused on the efficient discovery of antibacterial leads against 119 targets from *Acinetobacter baumannii* and *Staphylococcus aureus*. The success of this effort led to the hypothesis that the relative number of ELT binders alone could be used to assess the ligandability of large sets of proteins. This concept was further explored by screening 42 targets from *Mycobacterium tuberculosis*. Active chemical series for six targets from our initial effort as well as three chemotypes for DHFR from *M. tuberculosis* are reported. The findings demonstrate that parallel ELT selections can be used to assess ligandability and highlight opportunities for successful lead and tool discovery.

<sup>1</sup>GlaxoSmithKline, 1250 South Collegeville Road, Collegeville, Pennsylvania 19426, USA. <sup>2</sup>GlaxoSmithKline, 830 Winter Street, Waltham, Massachusetts 02451, USA. <sup>3</sup>GlaxoSmithKline, Severo Ochoa 2, Tres Cantos, Madrid 28760, Spain. <sup>4</sup>University of Birmingham, School of Biosciences, Edgbaston, Birmingham B15 2TT, UK. \* These authors contributed equally to this work. † Present addresses: Forma Therapeutics Inc., 500 Arsenal Street, Watertown, Massachusetts 02472, USA (C.S.K.); X-Chem Pharmaceuticals, 100 Beaver Street, Waltham, Massachusetts 02453, USA (P.A.C.); UT Southwestern Medical Center, 5323 Harry Hines BLVD., Dallas, Texas 75390, USA (A.C.); Otsuka Pharmaceuticals, 508 Carnegie Center, Princeton, New Jersey 08540, USA (C.P.); Janssen Pharmaceuticals, Molecular and Cellular Pharmacology, 1400 McKean RD, Spring House, Pennsylvania 19477, USA (L.M.S.); Alkermes Pharmaceuticals, 852 Winter Street, Waltham, Massachusetts 02451, USA (M.F.T.); Northeastern Reproductive Medicine, 105 W View Road, Colchester, Vermont 05446, USA (D.J.W.); Eli Lilly and Company, 222 Hubin Rd, Lu Wan Qu, Shanghai Shi 200021, China (J.Z.). Correspondence and requests for materials should be addressed to G.E. (email: ghotas.evindar@gsk.com).

Drug discovery and development is an extremely challenging, lengthy and expensive endeavour with an unsustainable success rate<sup>1,2</sup>. Understanding the ligandability of a target protein remains a central challenge for early small molecule drug discovery programmes<sup>3</sup>. An integrated lead discovery approach often employs biochemical target-based screening, cellular screening, phenotypic screening or a combination of methods. In most cases, a biochemical target-based approach is chosen to search for early lead molecules<sup>4</sup>. This approach requires extensive reagent generation, assay development, lead identification and optimization efforts, and can amount to years of invested time and millions of dollars in expense. The approach only interrogates a tiny fraction of the essential and druggable proteome and often results in failure after considerable investment. The alternative phenotypic and cellular screening methods probe multiple targets but often require target deconvolution efforts to support lead optimization as well as significant resource and technology investments to execute<sup>5,6</sup>. The antibacterial therapeutic area is particularly challenging, with the number of approved drugs steadily declining since 1980. Multiple factors have contributed to the lack of success, including the emergence of resistance, challenges in designing cell penetration properties into an antibacterial agent, a focus on genes essential for growth in rich media (only 7% in *Escherichia coli*) and a general lack of tools to understand the complex biology of host–pathogen interactions<sup>7–9</sup>. In addition to the many uncharacterized bacterial targets, 75% of protein research into human targets still focuses on only 10% of the proteins known before the human genome mapping<sup>10</sup>. A platform to prioritize this wider scope of targets based on ligandability and the rapid identification of lead and tool molecules for these proteins will aid in a deeper understanding of biology and the discovery of the next generation of therapeutics.

A variety of academic and pharmaceutical research has been directed towards addressing the need for understanding chemical tractability over the past decade and both experimental and computational strategies have been reported<sup>11–13</sup>. As described elsewhere, chemical tractability or ligandability is the ability of a protein to bind chemical matter, whereas druggability describes the possibility of discovering a molecule for the target that will modulate the disease state<sup>3</sup>. A technique to characterize small molecule ‘hot spots’ on proteins was reported by Hajduk *et al.*<sup>14</sup>, where the relationship between ligandability and fragment-binding measurements with NMR was described. Edfeldt’s more recent feature highlights the use of fragment technology to probe the ligandability of 36 targets in the AstraZeneca portfolio over an 8-year period<sup>15</sup>. Computational methods and docking approaches have been reported and are highly valuable<sup>16,17</sup>. These techniques have helped guide target selection for the past decade but are often limited by target throughput, experimental efficiency and the prerequisite that structural information is available.

A relatively recent addition to drug discovery screening is the use of DNA-encoded chemical libraries (DEL) or encoded library technology (ELT) to discover small molecule binders to targets of interest<sup>18</sup>. First conceptualized in 1992 (ref. 19), various methods for creating large DNA-encoded combinatorial libraries have been developed, such as split and pool encoding<sup>20</sup>, DNA-templated synthesis<sup>21,22</sup> and self-assembling libraries<sup>23,24</sup>. Regardless of the library synthesis technique, binders to individual proteins are selected by affinity from complex mixtures. Binding events are detected by harnessing the power of high-throughput sequencing to identify DNA tags encoding the bound chemotypes. Chemotypes of interest can then be re-synthesized off-DNA for assay testing and further development. The technology has become widely used throughout the

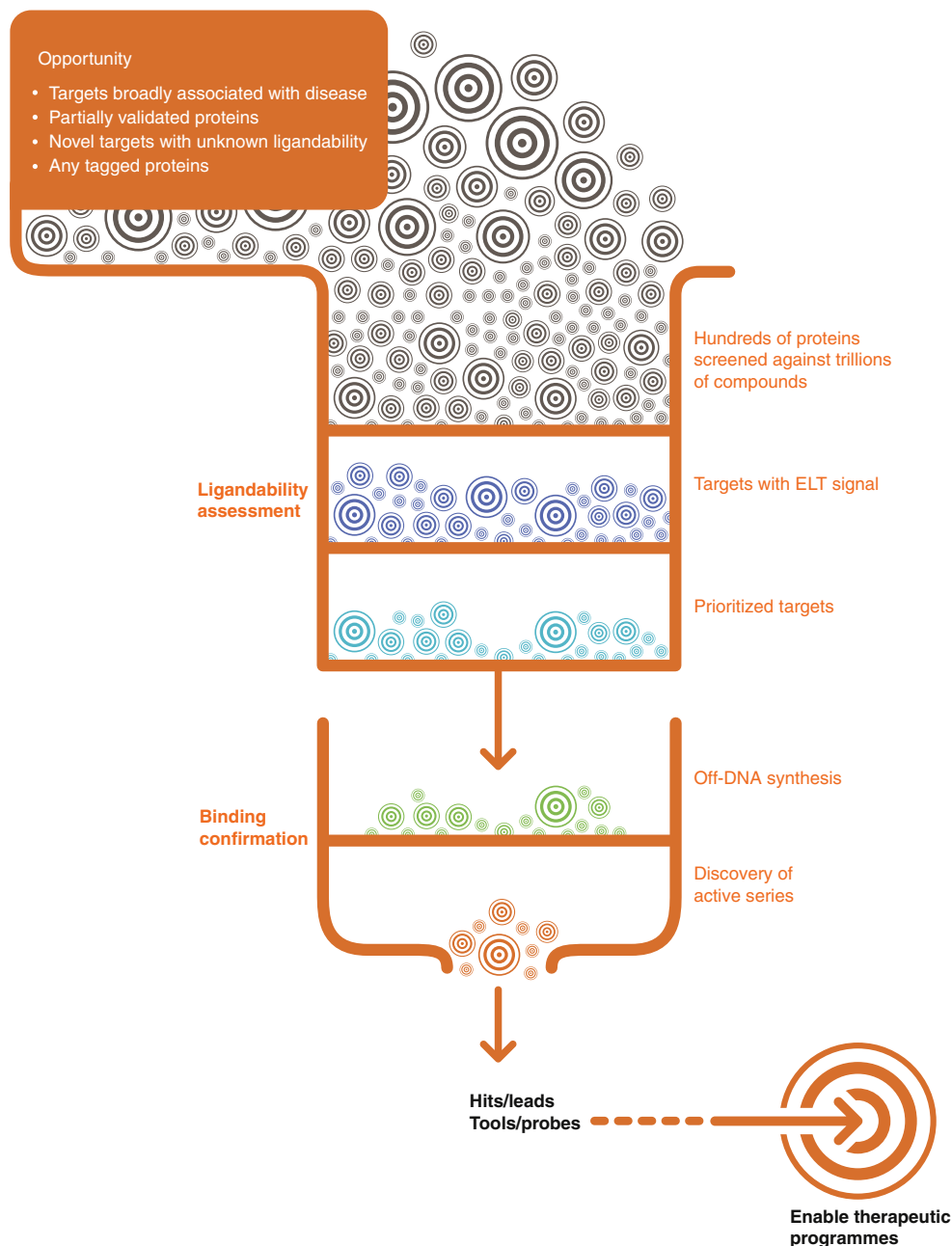
pharmaceutical, biotech and academic environments to discover ligands and is continually evolving<sup>25–27</sup>. ELT has been employed successfully to identify numerous molecules that bind to proteins from diverse target classes and sample deep chemical space as well as understand structure activity relationships directly from the screening output<sup>28–38</sup>.

At GlaxoSmithKline (GSK), we have created a DNA-encoded compound collection containing over 100 different libraries comprised of billions of unique molecules generated from over 40 reaction types using a split and pool encoding strategy<sup>39,40</sup>. These libraries are ultimately assembled and screened in a single pool, allowing for the rapid identification of chemical ligands. The technology has continued to mature and ELT is now considered for every small molecule screening effort in the GSK portfolio to discover new chemical matter. As high-throughput sequencing has continued to advance and we have refined our analysis methods, the target capacity of the technology has increased. Pairing these improvements with the automation of protein immobilization and affinity selections has allowed for the screening of multiple targets in a single experiment, as opposed to probing one target at a time, enabling a panel to be executed in a 5–7-month period<sup>41</sup>. Thus, the interrogation of a larger fraction of the proteome is possible while maintaining the simplicity and efficiency of a reductionist target-based approach.

Here we present the successful use of this strategy to identify tractable targets from *Staphylococcus aureus* and *Acinetobacter baumannii* and to discover individual lead/tool molecules for six different target proteins. We describe the chemical series and their activity against each protein. In addition, we propose an evolution of the panel screening, in which the output of the screen moves beyond identifying active pharmacophores and is used to rapidly assess targets based on their ligandability as determined by ELT. The refined approach was used to prioritize proteins from *Mycobacterium tuberculosis* and we present those results along with active pharmacophores against one of the highest-ranked targets as a proof of concept. Our results demonstrate that ELT can inform the allocation of resources within the drug discovery process towards the most chemically tractable targets. We envision the methodology will provide a tool to assess targets associated with virtually any therapeutic area. Parallel screening could include targets clustered in pathways, related targets, or a single target examined under multiple conditions perhaps using multiple constructs<sup>42,43</sup>. While this assessment is independent from target selection in the broader biological context, the technique could enable a deeper understanding of disease biology by rapidly providing much-needed tool molecules<sup>18,44,45</sup>. Here we report data supporting this assertion as well as several novel chemotypes as tools for the antibacterial field.

## Results

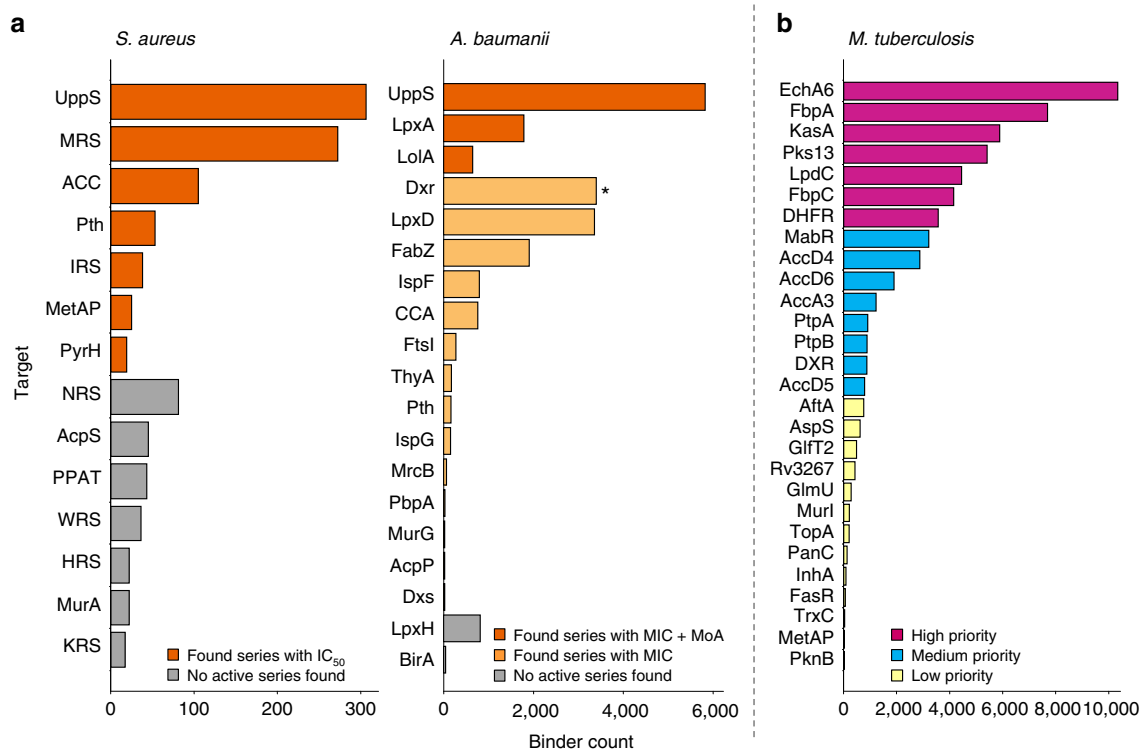
**ELT selection outcome.** A schematic showing the streamlined selection of targets for R&D efforts through the ELT tractability approach is illustrated in Fig. 1. The ELT selections were conducted by immobilizing affinity-tagged protein onto an affinity matrix, then exposing the protein to pooled compound libraries before washing away non-binders and recovering bound compounds by heat elution (detail below). This process was repeated to enrich bound species and reduce the population that does not bind to the protein of interest (described previously)<sup>20,31–33,37</sup>. The individual selection process was adapted to an automation platform such that hundreds of proteins could be evaluated in parallel. For each selection, final yields of 10<sup>7</sup>–10<sup>9</sup> sequences of DNA were obtained, quantified using qPCR and amplified for sequencing as described in the ‘Methods’ section and



**Figure 1 | Schematic representation of the ELT screening and ligandability assessment strategy.** Hundreds of potential targets of interest are immobilized and screened against GSK's DNA encoded libraries. Targets with signal are ranked by the counts of ELT binders which correlate to the protein's chemical ligandability. Data is used to plan off-DNA synthesis and confirm hits in assays. In addition to novel chemical starting points for lead optimization, the output is a target/chemotype pair that enables therapeutic programmes by providing tools for target validation as well as a ranking of targets to prioritize further studies.

previously<sup>20</sup>. The collected data were translated from the DNA barcode to the associated encoded molecule. On the basis of the library size and the number of sequences obtained, the noise level was calculated for each selection. Signal strength is reported as a value relative to that level (that is, signal value of 10 represents 10-fold greater measurement than noise). All data points with signal greater than two were included in subsequent data analysis steps. This output was then filtered to remove chemotypes that had been identified as binders to affinity matrix or multiple proteins in past selections (non-specific or frequent nuisance binders). This set of specific binders was clustered by chemical similarity (Tanimoto score > 0.85) or shared building blocks. The compounds can be used as tools to assess the validity of the target

or potentially as a lead molecule. Over the course of these panels, the number of libraries available for ELT screening expanded from 36 in the initial panel to 84 in the final screening panel and is currently over 100. This increase in library size was accompanied by a parallel increase in the reaction types used to create those libraries contributing to a modest increase in diversity and an ability to probe a greater breadth of chemical space. Generally, we assessed the target's tractability by examining the number of enriched chemical series as illustrated in Fig. 2 by a plot of the number of binders versus target protein. Committing chemistry resource to the most tractable targets first, clusters with dense representation, high signal strength and favourable chemical properties were prioritized for synthesis of



**Figure 2 | Count of binders that are specific to each target. (a)** Results of screening against *S. aureus* and *A. baumannii* panels. Targets are binned and coloured by the activity of molecules found for the target and sorted by the number of binders. **(b)** Proposed priority of targets for the *M. tuberculosis* panel. Targets are sorted by binder count and coloured by the proposed priority based on the number of specific binders found. \*Compounds for Dxr confirmed by biochemical assay.

representative chemical series<sup>46</sup>. In addition, a structure similarity search (Tanimoto >0.85) was conducted against the corporate collection for every target, including those with fewer binders, enabling rapid hit confirmation as well as interrogation of all targets with signal. Specific outcomes for each of the three panels are described below.

***S. aureus* panel.** The *S. aureus* ELT target panel was a collection of 39 essential enzyme targets. Table 1 shows the high-level progression of the targets in each panel; for detailed target information, see Supplementary Table 1. These targets were selected based on their potential utility as antibacterial drug targets and the availability of a biochemical assay system either inside GSK or rapidly adaptable from reported literature. N- and C-terminal dual-tagged proteins were subjected to ELT selections; one tag was used for rapid protein purification and the second for ELT selection immobilization. Miniaturization, automation and coupling of the purification and ELT selections in a parallel method enabled maximum efficiency (described in detail in Supplementary Methods). Data from the targets were examined individually and chemical series for 14 targets were chosen for follow-up by off-DNA synthesis (Fig. 2a). Approximately five compounds per chemical series and three chemical series for each target were synthesized for activity testing in biochemical assays. Following this confirmation, the activity of the compounds was assessed by cellular toxicity assays as described in the ‘Methods’ section. After compound synthesis and *in vitro* testing, seven targets gave at least one chemical series with confirmed biochemical activity measured with a minimum of two replicates, see Supplementary Table 1. This represents an attrition rate of roughly 50%, which is consistent with our previous ELT screening experience at GSK. Antibacterial activity was demonstrated against *S. aureus* for several chemical series identified from five

targets, methionyl-tRNA synthetase (MRS), isoleucyl-tRNA synthetase (IRS), methionine aminopeptidase (MetAP), undecaprenyl pyrophosphate synthase (UppS) and acetyl-CoA carboxylase (ACC). The *S. aureus* UppS ELT hit series, its medicinal chemistry efforts, X-ray co-crystal structure and genetic and biochemical confirmation of antibacterial MoA were recently reported in a separate publication<sup>47</sup>. Similarly, the acetyl-CoA carboxylase hit series and characterization studies will be published at a later time. See Supplementary Methods for more details on compound characterization.

In Table 2, we disclose a representative novel MRS inhibitor belonging to a phenylbenzimidazole chemical series. It has potent enzyme activity in an MRS biochemical assay with an  $IC_{50}$  of 830 pM and possesses moderate antibacterial activity with a minimum inhibitory concentration (MIC) value of  $0.5 \mu\text{g ml}^{-1}$  against *S. aureus*. Importantly, we conducted studies that suggest the cellular activity of the compound was linked to the inhibition of MRS. Increasing the levels of MRS in an overexpressing strain resulted in a higher MIC. We observed that benzimidazole 1 gave a significant MIC increase of 8-fold (uninduced) and >128-fold (induced with  $0.1 \mu\text{g ml}^{-1}$  anhydrotetracycline) as compared to a vector control strain. These studies were conducted with a minimum of two replicates. In contrast, we were unable to observe the on-target antibacterial activity of the IRS inhibitor represented by the *t*-butylphenyl-piperazine amide 2 shown in Table 2. We include this series as an example of a hit that binds and inhibits its selected target but whose antibacterial activity could not be established as on-target. An exemplar from a MetAP benzimidazole series is shown (benzimidazole-triazole 3) with an  $IC_{50}$  of  $0.35 \mu\text{M}$  in the biochemical assay (described in Supplementary Methods) but lacks antibacterial activity, hence its antibacterial MoA could not be confirmed in overexpression studies.

**Table 1 | ELT screening progression.**

Phase	<i>S. aureus</i>	<i>A. baumannii</i>	<i>M. tuberculosis</i>
<i>Ligandability assessment</i>			
Initial number of targets considered	39	80	42
Targets amenable to ELT selection	32	70	41
Targets with ELT signal	14	52	27
<i>Binding confirmation</i>			
Targets with off-DNA synthesis	14	18	13
Targets with confirmed active chemical series (IC <sub>50</sub> and/or MIC)	7	17	4*
Targets with confirmed MoA	2	3	ND

ND, not determined.

The table shows a summary of the progression of protein targets through each tractability panel. For additional detail and a full list of targets prosecuted, see Supplementary Table 1.

\*Off-DNA activity assessment ongoing for remaining nine targets.

**A. baumannii panel.** On the basis of our experience with the *S. aureus* campaign, we decided to expand the number of targets examined from the *A. baumannii* genome as a representative organism for the discovery of novel Gram-negative agents. To further simplify the process, we tested the antibacterial activity of all synthesized compounds against a panel of bacterial strains as shown in Supplementary Table 2. This eliminated the need to develop biochemical assays for each target and focused the effort on agents with cell penetration. Eighty proteins were selected for screening from the *A. baumannii* genome and, similar to the *S. aureus* panel described above, constructs were generated incorporating dual affinity tags (streptavidin-binding peptide and FLAG tag). *E. coli* was again used as a heterologous expression host and cell pellets, confirmed to be expressing protein, were thawed on the day of selection and lysed (described in Supplementary Methods). Each target protein was purified using streptavidin agarose resin, eluted with biotin and quantified using an Agilent Bioanalyzer all on the same day as the selection experiment. This removed the need for large-scale protein preparations to be conducted and removed a freeze/thaw cycle that can often compromise the quality of the reagent. Of the 80 targets, 70 were successfully purified using this rapid process. Our panel screen gave positive ELT selection signal for 52 targets and 18 were prioritized for off-DNA synthesis. After off-DNA synthesis of 3–5 chemical series per target and/or similarity searching of the GSK compound collection, 17 of the targets had at least one chemical series with activity in the antibacterial MIC assay panel and 3 targets had compounds that show positive MoA data. We disclose three chemical series identified from these targets with confirmed on-target activity shown in Table 2. LpxA, involved in lipid A biosynthesis, yielded two chemotypes with multiple compounds showing on-target activity. We show a general structure of a chemical series for one of the LpxA chemotypes and a particular aryl-urea series 4 with confirmed on-target MoA. UppS from *A. baumannii* was again found to produce a high number of binders and three unique series with positive MoA were identified, one of these (compound 5) is shown in Table 2. LolA, a lipoprotein chaperone, afforded a chemical series with positive MoA with a disclosed tetrahydropyrido-pyrimidine exemplar 6 shown in Table 2. See the Supplementary Methods for additional details on compound characterization, assay methods and activity results across a panel of bacteria.

**M. tuberculosis panel.** Having seen the ability of these panels to rapidly find ligands for a diverse set of protein targets, we proposed using the number of ELT binders to rank the tractability of proteins from *M. tuberculosis*. An additional goal of this screen was to obtain tool molecules that could help validate a target as

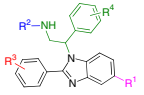
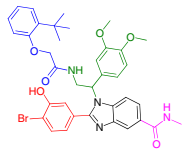
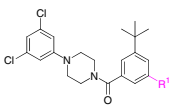
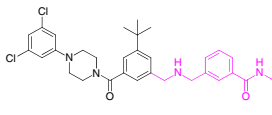
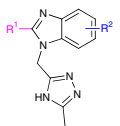
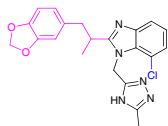
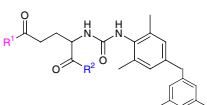
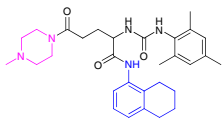
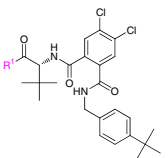
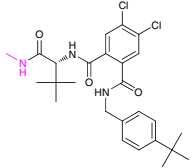
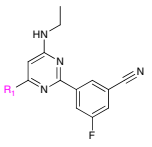
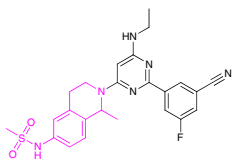
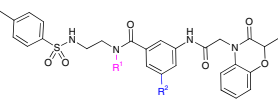
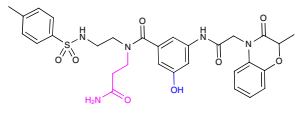
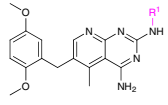
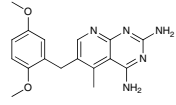
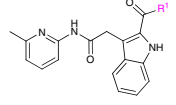
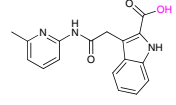
pharmacologically relevant. We did not place any requirements for target inclusion other than that the protein had to be purified and contain either a 6-histidine or biotin affinity tag. We obtained a diverse group of targets from academic collaborators for ELT screening. Eleven academic partners provided 42 protein preparations for screening (Supplementary Table 1). Although little to no previous information existed on the chemical tractability for most of these 42 proteins, a small number of known, chemically tractable targets were included as ‘controls’ (for example, InhA and DHFR). The ranking of target tractability is shown in Fig. 2b. Twenty-seven out of 42 targets afforded enriched chemotypes, and 13 targets with a greater number of binders were selected for further data analysis. Each of these targets produced multiple chemical series, and we chose three–five chemotypes for synthesis and off-DNA evaluation. The synthesized compounds 7–9 (Table 2) were easily assayed in a biochemical assay from the reported literature and we disclose three distinct chemical series with potency varying from 492 nM to 4.9 μM against dfrA (also known as DHFR). Compounds 7–9 were active against *M. tuberculosis* and the MIC values, measured in duplicate, are reported in Table 2. Although inhibitors of DHFR are already widely known, the mixture of familiar (compound 8 has a similar substructure to the known DHFR inhibitor methotrexate) and novel inhibitor chemotypes (7 and 9) identified by the ELT selection serves to validate the platform and demonstrate its powerful potential. InhA, the enoyl reductase targeted by the frontline therapeutic isoniazid, was screened but did not result in hits (see ‘Discussion’ section). Assessment of the compounds synthesized for other targets remains ongoing, with chemical series against four targets, KasA, LpdC, DHFR and DXR, confirmed in biochemical assays to date.

## Discussion

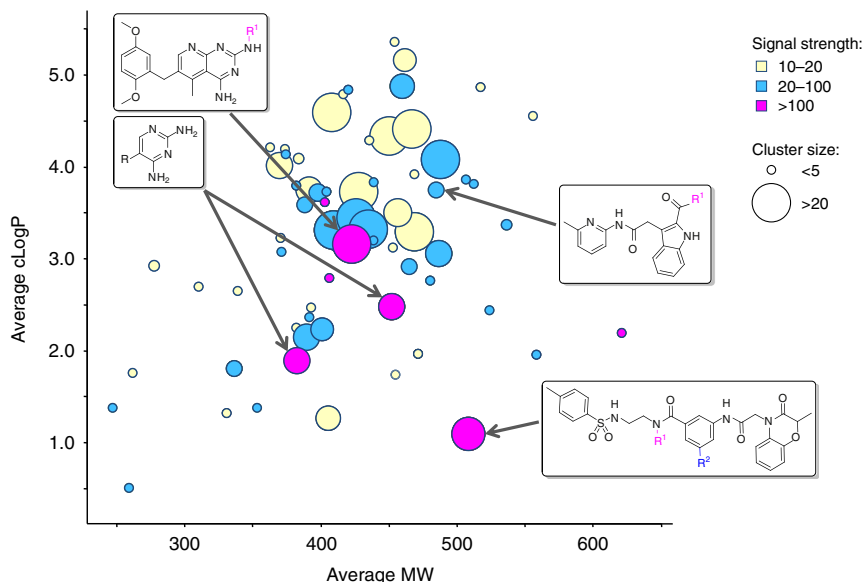
Tools to prioritize the abundance of uncharacterized proteins that will serve as the next generation of drug targets are crucial to the successful development of new medicines. The ELT target panel platform described here provides an opportunity to select protein targets based on an experimental and data-driven ligandability assessment. Starting with a diverse set of proteins that are of interest to a particular therapy area, the strategy rapidly identifies tractable targets and simultaneously defines novel starting points for therapeutic programme teams to explore. To our knowledge, no other report provides an experimental approach to tractability assessment that can evaluate a large set of proteins *de novo* in such a short period of time<sup>15,17</sup>.

Initially, the concept of screening many targets in parallel was proposed as a way to simply identify novel chemical matter for antibacterial drug discovery with less investment of both time and effort. The dual tagging, automation and miniaturization

**Table 2 | Representative chemical series discovered in three tractability campaigns.**

Campaign	Target	Compound	Scaffold	Exemplar	IC <sub>50</sub> (μM)	MIC* (μg ml <sup>-1</sup> )	MoA
<b><i>S. aureus</i></b>							
	MRS	1			0.00083 ± 0.00005	0.5	Y
	IRS	2			0.75 ± 0.21	4	N
	MetAP	3			0.35 ± 0.01	> 128	ND
<b><i>A. baumannii</i></b>							
	LpxA	4			ND	>64	Y
	UppS	5			0.043 ± 0.025	>128	Y
	LolA	6			ND	>128	Y
<b><i>M. tuberculosis</i></b>							
		7			0.492 ± 0.004	80 <sup>†</sup>	ND
	DHFR	8			0.61 ± 0.42	2.5 <sup>†</sup>	ND
		9			4.9 ± 2.2	80 <sup>†</sup>	ND

ND, not determined; MoA, mode of action.  
The scaffold column represents the selected pharmacophore with areas of substitution indicated by R. All data are reported for the single exemplar shown. IC<sub>50</sub> values are reported as the average of two replicate experiments with s.d. values calculated using the n - 1 method.  
\*Minimum inhibitory concentration measured against *S. aureus* WCUH29 wild type or *A. baumannii* BM652 efflux strains for *S. aureus* or *A. baumannii* campaigns with a minimum of two independent experiments.  
<sup>†</sup>MIC unit is in μM and is determined against *M. tuberculosis* H37Rv.



**Figure 3 | Representative data set of enriched binders for the *M. tuberculosis* target DHFR.** The average molecular weight of clustered chemotypes is plotted versus average cLogP. Clusters are sized by the number of members and coloured by maximum signal strength. Chemical series with active molecules are indicated by their scaffolds.

strategies described in the ‘Methods’ section (and previously) enabled the purification and screening of multiple targets in a single experiment<sup>41</sup>. We screened 39 proteins from *S. aureus* in the first panel, utilizing 36 libraries combined in a single pool, highlighting 14 tractable targets and, with limited chemistry investment, we discovered active chemical series for seven of the targets. Building on the success of the *S. aureus* panel, we initiated a refined strategy against a panel of protein targets from the Gram-negative pathogen *A. baumannii*. While most of these targets are essential for growth to the bacteria, some targets were added in this iteration, like LpxA, that are non-essential for *A. baumannii*. These are of interest to Gram-negative discovery projects in general as they are essential in most other Gram-negative pathogens and show severe growth impairment in *A. baumannii* when knocked out<sup>48</sup>. To rapidly identify potential lead compounds as well as explore a wider set of targets, we directly measured MIC, rather than developing individual biochemical assays for each target. Thus, the ability of hits to overcome bacterial permeability was added as a filter. This refinement increased the number of targets available for screening since no biochemical assay development was needed. However, the lack of biochemical data reduced our ability to confirm the activity of every chemotype as a tool rather than a cell penetrant lead. We observed active compounds by MIC testing, with hits from 17 of 18 targets showing activity against *A. baumannii*. Due to resource constraints and the focus of the therapeutic programme, only 3 of the 17 hits were fully characterized in mechanistic studies and are reported here.

By applying ELT screening to the antibacterial panels, we were able to rapidly interrogate 102 targets and discover active chemical matter against 24, a success rate of 24%. A similar rate of success for finding hits from high-throughput screening (HTS) against antibacterial targets has been reported earlier by GSK<sup>49</sup>. Indeed, the relatively low success in HTS along with the time required for lead molecule development as well as regulatory and commercial concerns are cited as reasons why many large pharmaceutical companies have abandoned this therapeutic area<sup>50</sup>. A comparison of 29 *S. aureus* targets common to both HTS and ELT approaches showed that in 21 of the 29 HTS campaigns, our ELT strategy would have predicted the outcomes (summarized in Table 3; for

details, see Supplementary Table 3). In addition, for the remaining eight targets, the ELT screen successfully found hits for five targets where HTS did not. Another benefit of the ELT parallel screening method is that significantly less resources and time are needed to screen the collection of targets when compared to HTS. The time required to proceed from the set of purified proteins to an analysis of the chemical output averaged 3–4 months with an additional 2–3 months of time committed to chemical synthesis and activity testing. HTS assay development through hit analysis and hit qualification happens on a similar timescale (9–12 months). However, while only a single target is probed in HTS, here we describe a parallel ELT strategy that can interrogate many proteins in the same timeframe using only microgram amounts of reagent and far less human resource.

We hypothesized that applying chemistry resource to only the targets with the greatest number of binders would further improve the screening efficiency. The relationship between the ELT output and success in discovering active chemical series is illustrated by Fig. 2a. Here we have binned targets by outcome and ordered by the number of binders, finding that the successful targets were those with a greater number of binders. In the *S. aureus* panel, 7 of 14 highly ranked targets were confirmed as active and in the *A. baumannii* panel, three prioritized targets yielded compounds with confirmed MoA. The finding that ELT could rapidly winnow down a diverse set of targets and highlight those that were chemically ligandable provided an opportunity to improve our decision making and prioritize resources onto the most tractable targets. Utilizing the output of the ELT selections to predict downstream target success, both time and money can be saved while simultaneously identifying novel chemical starting points. We therefore hypothesize that ELT panels can additionally be used for pure tractability assessments for any group of interesting targets.

This was explored in a third panel by screening against *M. tuberculosis* proteins. As focus shifted from the identification of new chemical matter to tractability assessment, we chose to screen only against high-quality purified protein. The benefit of this can be seen in the reduced attrition of targets entering selection, with 41 of 42 proteins being amenable to ELT at the selection stage (Table 1 and Supplementary Table 1). Targets were



**Table 3 | Consistency of HTS and ELT outcomes for 29 *S. aureus* targets.**

	ELT active	ELT inactive
HTS active	4	3
HTS inactive	5	17

Comparison of target outcomes from ELT and HTS where the same target was screened by both methods. ELT and HTS outcomes agreed in 21 of 29 campaigns. For the remaining eight targets, ELT found hits for five where HTS did not.

ranked by decreasing number of binders (Fig. 2b). Active chemical matter for three high-priority targets (DHFR, KasA and LpdC) and one medium-priority target (DXR) has been confirmed, and chemotypes for the nine remaining targets are currently being evaluated. The effects of refinements and advances in the ELT process and data analysis are evident when comparing the panels shown in Fig. 2, which illustrates increases in both the number of binders found and number of targets with signal. This increased information density enabled the ranking of targets by tractability, allowed for improved discrimination between targets and facilitated better decision making.

The challenges in interpreting data and planning chemical synthesis have been met by advances in analysis tools and streamlining the planning of chemical synthesis. Historically, we examined an individual library through cube analysis to provide guidance for off-DNA hit to lead efforts<sup>30–33,35,37,51</sup>. Figure 3 shows a representative data set for a single target, DHFR, where chemical clusters are plotted with their average molecular weight versus the average cLogP. This is a common representation of the data that allows for rapid comparison of chemical series by signal intensity and also by physicochemical properties. Chemical series with higher signal intensity and larger cluster size that also fell into more attractive physicochemical space were selected for follow-up<sup>46</sup>. All targets were evaluated in parallel and common chemical reactions were executed in a concerted manner to increase efficiency. Representatives from each of the highlighted chemical series were then synthesized off-DNA and tested in biochemical or cellular assays. Additional visualizations were generated to examine potential structure activity relationships<sup>37,51</sup>. The structure activity relationships analysed within each chemical cluster allows for rapid pharmacophore identification and expansion of a single hit into a chemical series.

A key component of our strategy is that attrition is expected and that a negative result is not predictive, it simply helps focus on success at each step. It should be emphasized that we aim to expand the possibility of discovery to a wider range of targets and then redirect focus onto ligandable targets with speed and efficiency and do not advocate ignoring difficult targets. It is also important to note that targets with fewer ELT binders under our screening conditions may in fact be highly tractable under different conditions. InhA, the target of tuberculosis drug isoniazid, was found to have low tractability despite having known ligands. This enzyme is known to require cofactor to bind inhibitors and has an active site loop that plays a critical role in inhibition<sup>52–54</sup>. Recent publications report ELT selections for InhA consistent with the notion that the presence of cofactor improves this target's tractability<sup>33,42</sup>. In fact, the method described here could be used to test the same target under many conditions or multiple constructs of the same target for optimization. These conditions may influence the stability of the protein or the conformation of the active site, which will impact the outcome of the ELT selection. Previous reports have focused on the discovery of ligands specific for protein conformations or complexes, either by screening in the presence of cofactors or against homologous proteins and our technique can be easily adapted for similar screens against any group of proteins to produce a relevant ranking<sup>33,42,55</sup>.

The ELT screening campaigns presented here have successfully identified multiple antibacterial compounds with confirmed MoA and have led to multiple lead optimization programmes. We have also applied this principle to other novel target areas, such as deubiquitinases, methyl readers, bromodomains and histone methyltransferases, and current studies are underway to include additional target areas (for example, the kinome), expanding the potential impact of this strategy. In cases where modulation of many targets converges on a single disease or phenotype, we are designing ELT panels where all of the off-DNA synthesis converges on a single phenotypic assay. The result will be a tractability ranking for all of the targets connected to a particular disease state along with potential tool and lead molecules to further explore the biology. Given the highly competitive, challenging and expensive endeavour of drug discovery, it is crucial to focus resource where it is most likely to have a positive impact. Our strategy enables teams to quickly identify targets to focus on and we see tremendous potential for this approach to impact other areas throughout the R&D community.

## Methods

**Just-in-time protein purification.** *E. coli* cell culture expressing target protein was stored as a frozen pellet until the day of selections. Approximately 200 mg of cell pellet was dissolved in B-PER cell lysis solution (Thermo 90084) and incubated at room temperature for 20 min with occasional stirring. Solutions were divided evenly into eppendorf tubes and centrifuged at 18,000 g, 4 °C for 20 min. A 10 ml filter spin column (Pierce 89898) was prepared for each sample. The column was loaded with 400 µl volume of 50% slurry of Streptavidin Ultra-link Resin (Pierce 53114). Solution was allowed to flow through and 5 ml of ELT buffer (50 mM Tris-HCl pH 7.5, 0.15 M NaCl, 0.1% Tween 20) was added to equilibrate resin bed. The resin was centrifuged for 3 min at 500 r.p.m., the column capped and 100 µl of buffer added to keep the resin moist. A measure of 3 ml of clarified cell lysate was loaded onto the prepared resin bed and incubated on ice for 15–20 min. The solution was allowed to flow through the column and centrifuged for 3 min at 500 r.p.m.. A measure of 10 ml of ELT buffer was loaded to wash the resin and centrifuged for 3 min at 500 r.p.m.. A measure of 500 µl of 2.5 mM D-biotin was added (Sigma B-4501). To make this stock solution solid, D-biotin was dissolved to 0.5 M in 1 M NaOH, diluted to 2.5 mM with ELT buffer and incubated on ice for 45 min with gentle and occasional stirring. Column was transferred to a clean tube (Falcon) and eluted by centrifugation at 500 r.p.m. for 3 min. Purified proteins were quantified using a Bioanalyzer (Agilent 230 Protein Chip Kit, Fisher NC9415734) where purity and approximate concentrations were accessed. A yield of 30–200 µg of target protein was observed with purity varying from ~60 to >95%.

**Encoded library preparation for selection.** ELT libraries were combined into a pooled library for screening purposes. For each screening effort, the then available libraries were ligated to contain a unique library identifier and then combined into an equimolar mixture at a concentration of 0.5 mM. After final library closing to add a pool identification sequence, these libraries could then be identified as associated with each individual condition after sequencing (see Supplementary Methods).

**PhyNexus tip preparation and automated selections.** Each condition was screened using a Beckman Coulter, BioMek FX robot. Four 200 µl PhyNexus tips packed with 5 µl beds of either M2 anti-FLAG, IMAC or streptavidin resin were employed per target and equilibrated before selections using ELT selection buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 1 mM CHAPS, 10 mM imidazole, 0.1 mg ml<sup>-1</sup> sheared salmon sperm DNA, 1 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 1 mM DTT). Tips were used within 2 h of equilibration. Ten nmoles of pooled library were suspended in 100 µl of selection buffer and used as input for each selection. Ninety-six-well plates were prepared containing 10 µg of target protein suspended in 110 µl of selection buffer per well in row A for each of four rounds of selection. A measure of 110 µl selection buffer was pipetted per row into rows B–G to serve as washes. In the case of streptavidin tips only, 1 mM biotin was included in the wash

buffer in row B to reduce enrichment of streptavidin binders. In addition, 120  $\mu$ l of selection buffer in row H was used for heat elution. A BioMek FX protocol was developed in which purified protein from row A was captured for 15 min on Phynexus tips, washed once with selection buffer from row B, exposed to library pool for 30 min, washed five times with selection buffer (rows C–G) and heat eluted in selection buffer from row H at 90 °C to recover binders. Four rounds of selection were performed in this manner with the output of each round carried forward as the input to the successive round with fresh protein and new tips.

**Selection output processing.** Outputs of each selection condition were purified using a Qiagen Qiacube running a nucleic acid purification protocol and then quantified by qPCR run on a Roche Lightcycler. Total DNA output ranged from  $2 \times 10^7$  to  $1 \times 10^9$  copies, and was PCR amplified with primers adaptors to add sequences compatible with Illumina sequencing flowcells. PCR output was purified using Agencourt AMPure XP SPRI beads according to the manufacturer's instructions and then quantitated on an Agilent BioAnalyzer using a high sensitivity DNA kit. Final concentration of amplicon for each sample was between 2 and 40 nM. Portions of products were loaded to generate ~20 million clusters per selection condition on an Illumina GAI or HiSeq platform.

**Sequencing and data analysis.** Libraries were pooled in equal volume and each library had a unique DNA tag that enabled sequence deconvolution. Each warhead was uniquely tagged with a specific DNA tag combination (described previously<sup>20</sup>) and also included a degenerate region to account for sequencing or amplification artifacts. Compounds with identical degenerate regions were counted as a single occurrence.

For every library and for each target, the total number of unique warhead sequences was counted. Compounds were then grouped by the different possible combinations of building blocks (for example, single cycle 'mono' synthons, two cycle 'di' synthons and so on). The total count for every combination was then used to calculate signal as the fold value of number of copies divided by the expected noise level (as defined by the theoretical distribution of copies). Every chemotype with a signal value > 2 was reported. All chemotypes with signal were compared against a database of historic target. Any chemotypes that also had signal against a historic target were flagged as promiscuous and removed from further consideration. Any chemotype with signal against empty matrix (no-target control) was similarly removed. Of the remaining chemotypes, only those that had signal against a single target from the panel were considered for final evaluation, except in cases where the target had highly similar binding pockets, substrates or cofactors.

A data table was created containing all chemotypes and their associated signal values for each target/library combination. These data were graphed in Spotfire and used to rank the targets by total number of chemotypes. To prioritize off-DNA follow-up synthesis, chemotypes with signal were clustered together based on calculated Tanimoto similarity by two-dimensional fingerprints. Clusters with high representation and favourable physical chemical properties (lower average molecular weight, lower average cLogP) were chosen for further evaluation and off-DNA chemistry planning.

**Data availability.** All relevant data are available from the authors on request.

## References

- Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **11**, 191–200 (2012).
- Smietana, K., Ekstrom, L., Jeffery, B. & Moller, M. Improving R&D productivity. *Nat. Rev. Drug Discov.* **14**, 455–456 (2015).
- Hann, M. M. & Keseru, G. M. Finding the sweet spot: the role of nature and nurture in medicinal chemistry. *Nat. Rev. Drug Discov.* **11**, 355–365 (2012).
- Russ, A. P. & Lampel, S. The druggable genome: an update. *Drug Discov. Today* **10**, 1607–1610 (2005).
- Mullard, A. The phenotypic screening pendulum swings. *Nat. Rev. Drug Discov.* **14**, 807–809 (2015).
- Dominguez, E. et al. Integrated phenotypic and activity-based profiling links Ces3 to obesity and diabetes. *Nat. Chem. Biol.* **10**, 113–121 (2014).
- Forsyth, R. A. et al. A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol. Microbiol.* **43**, 1387–1400 (2002).
- Brown, E. D. & Wright, G. D. Antibacterial drug discovery in the resistance era. *Nature* **529**, 336–343 (2016).
- Hopkins, A. L. et al. Rapid analysis of pharmacology for infectious diseases. *Curr. Top. Med. Chem.* **11**, 1292–1300 (2011).
- Edwards, A. M. et al. Too many roads not taken. *Nature* **470**, 163–165 (2011).
- Hajduk, P. J., Huth, J. R. & Tse, C. Predicting protein druggability. *Drug Discov. Today* **10**, 1675–1682 (2005).
- Egner, U. & Hillig, R. C. A structural biology view of target druggability. *Expert Opin. Drug Discov.* **3**, 391–401 (2008).
- Kutilek, V. D. et al. Integration of affinity selection-mass spectrometry and functional cell-based assays to rapidly triage druggable target space within the NF- $\kappa$ B pathway. *J. Biomol. Screen.* **21**, 608–619 (2016).
- Hajduk, P. J., Huth, J. R. & Fesik, S. W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **48**, 2518–2525 (2005).
- Edfeldt, F. N., Folmer, R. H. & Breeze, A. L. Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov. Today* **16**, 284–287 (2011).
- Ward, R. A. Using protein-ligand docking to assess the chemical tractability of inhibiting a protein target. *J. Mol. Model.* **16**, 1833–1843 (2010).
- Perola, E., Herman, L. & Weiss, J. Development of a rule-based method for the assessment of protein druggability. *J. Chem. Inf. Model.* **52**, 1027–1038 (2012).
- Salamon, H., Klika Skopic, M., Jung, K., Bugain, O. & Brunschweiler, A. Chemical biology probes from advanced DNA-encoded libraries. *ACS Chem. Biol.* **11**, 296–307 (2016).
- Brenner, S. & Lerner, R. A. Encoded combinatorial chemistry. *Proc. Natl Acad. Sci. USA* **89**, 5381–5383 (1992).
- Clark, M. A. et al. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol.* **5**, 647–654 (2009).
- Kleiner, R. E., Dumelin, C. E. & Liu, D. R. Small-molecule discovery from DNA-encoded chemical libraries. *Chem. Soc. Rev.* **40**, 5707–5717 (2011).
- Gartner, Z. J. & Liu, D. R. The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *J. Am. Chem. Soc.* **123**, 6961–6963 (2001).
- Hansen, M. H. et al. A yoctoliter-scale DNA reactor for small-molecule evolution. *J. Am. Chem. Soc.* **131**, 1322–1327 (2009).
- Melkko, S., Scheuermann, J., Dumelin, C. E. & Neri, D. Encoded self-assembling chemical libraries. *Nat. Biotechnol.* **22**, 568–574 (2004).
- Mullard, A. DNA tags help the hunt for drugs. *Nature* **530**, 367–369 (2016).
- Yuen, L. H. & Franzini, R. Achievements, challenges, and opportunities in DNA-encoded library research: an academic point of view. *ChemBiochem* **18**, 1–9 (2016).
- Lerner, R. A. & Brenner, S. DNA-encoded compound libraries as open source: a powerful pathway to new drugs. *Angew. Chem. Int. Ed. Engl.* **56**, 1164–1165 (2017).
- Southan, C., Varkonyi, P., Boppana, K., Jagarlapudi, S. A. & Muresan, S. Tracking 20 years of compound-to-target output from literature and patents. *PLoS ONE* **8**, e77142 (2013).
- Gilmartin, A. G. et al. Allosteric Wip1 phosphatase inhibition through flap-subdomain interaction. *Nat. Chem. Biol.* **10**, 181–187 (2014).
- Yang, H. et al. Discovery of a potent class of PI3K $\alpha$  inhibitors with unique binding mode via encoded library technology (ELT). *ACS Med. Chem. Lett.* **6**, 531–536 (2015).
- Deng, H. et al. Discovery of highly potent and selective small molecule ADAMTS-5 inhibitors that inhibit human cartilage degradation via encoded library technology (ELT). *J. Med. Chem.* **55**, 7061–7079 (2012).
- Disch, J. S. et al. Discovery of thieno[3,2-d]pyrimidine-6-carboxamides as potent inhibitors of SIRT1, SIRT2, and SIRT3. *J. Med. Chem.* **56**, 3666–3679 (2013).
- Encinas, L. et al. Encoded library technology as a source of hits for the discovery and lead optimization of a potent and selective class of bactericidal direct inhibitors of *Mycobacterium tuberculosis* InhA. *J. Med. Chem.* **57**, 1276–1288 (2014).
- Gentile, G. et al. 5-Aryl-4-carboxamide-1,3-oxazoles: potent and selective GSK-3 inhibitors. *Bioorg. Med. Chem. Lett.* **22**, 1989–1994 (2012).
- Kollmann, C. S. et al. Application of encoded library technology (ELT) to a protein-protein interaction target: discovery of a potent class of integrin lymphocyte function-associated antigen 1 (LFA-1) antagonists. *Bioorg. Med. Chem.* **22**, 2353–2365 (2014).
- Thalji, R. K. et al. Discovery of 1-(1,3,5-triazin-2-yl)piperidine-4-carboxamides as inhibitors of soluble epoxide hydrolase. *Bioorg. Med. Chem. Lett.* **23**, 3584–3588 (2013).
- Franzini, R. M. et al. Identification of structure-activity relationships from screening a structurally compact DNA-encoded chemical library. *Angew. Chem. Int. Ed. Engl.* **54**, 3927–3931 (2015).
- Goodnow, Jr R. A., Dumelin, C. E. & Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat. Rev. Drug Discov.* **16**, 131–147 (2016).
- Ding, Y. & Clark, M. A. Robust Suzuki-Miyaura cross-coupling on DNA-linked substrates. *ACS Comb. Sci.* **17**, 1–4 (2015).
- Satz, A. L. et al. DNA compatible multistep synthesis and applications to DNA encoded libraries. *Bioconjug. Chem.* **26**, 1623–1632 (2015).
- Decurtins, W. et al. Automated screening for small organic ligands using DNA-encoded chemical libraries. *Nat. Protoc.* **11**, 764–780 (2016).
- Soutter, H. H. et al. Discovery of cofactor-specific, bactericidal *Mycobacterium tuberculosis* InhA inhibitors using DNA-encoded library technology. *Proc. Natl Acad. Sci. USA* **113**, E7880–E7889 (2016).
- Cuozzo, J. W. et al. Discovery of a potent BTK inhibitor with a novel binding mode using parallel selections with a DNA-encoded chemical library. *ChemBiochem* **18**, 1–9 (2017).

44. Newman, R. H. & Zhang, J. Small molecules and chemical tools at the interface. *Nat. Chem. Biol.* **4**, 382–386 (2008).
45. Arrowsmith, C. H. *et al.* The promise and peril of chemical probes. *Nat. Chem. Biol.* **11**, 536–541 (2015).
46. Bayliss, M. K. *et al.* Quality guidelines for oral drug candidates: dose, solubility and lipophilicity. *Drug Discov. Today* **21**, 1719–1727 (2016).
47. Concha, N. *et al.* Discovery and characterization of a Class of pyrazole inhibitors of bacterial undecaprenyl pyrophosphate synthase. *J. Med. Chem.* **59**, 7299–7304 (2016).
48. Moffatt, J. H. *et al.* Colistin resistance in *Acinetobacter baumannii* is mediated by complete loss of lipopolysaccharide production. *Antimicrob. Agents Chemother.* **54**, 4971–4977 (2010).
49. Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat. Rev. Drug Discov.* **6**, 29–40 (2007).
50. Payne, D. J., Miller, L. F., Findlay, D., Anderson, J. & Marks, L. Time for a change: addressing R&D and commercialization challenges for antibacterials. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370**, 20140086 (2015).
51. Deng, H. *et al.* Discovery, SAR, and X-ray binding mode study of BCATm inhibitors from a novel DNA-encoded library. *ACS Med. Chem. Lett.* **6**, 919–924 (2015).
52. Li, H. J. *et al.* A structural and energetic model for the slow-onset inhibition of the Mycobacterium tuberculosis enoyl-ACP reductase InhA. *ACS Chem. Biol.* **9**, 986–993 (2014).
53. Pan, P. *et al.* Time-dependent diaryl ether inhibitors of InhA: structure-activity relationship studies of enzyme inhibition, antibacterial activity, and *in vivo* efficacy. *ChemMedChem* **9**, 776–791 (2014).
54. Tonge, P. J., Kisker, C. & Slayden, R. A. Development of modern InhA inhibitors to combat drug resistant strains of *Mycobacterium tuberculosis*. *Curr. Top. Med. Chem.* **7**, 489–498 (2007).
55. Franzini, R. M., Nauer, A., Scheuermann, J. & Neri, D. Interrogating target-specificity by parallel screening of a DNA-encoded chemical library against closely related proteins. *Chem. Commun.* **51**, 8014–8016 (2015).

## Acknowledgements

We gratefully acknowledge our collaborators for supplying the purified proteins used in the Tuberculosis screen: Prof. Yuk-Ching Tse-Dinh, Prof. John Blanchard, Prof. Ruslana Bryk, Prof. Carl Nathan, Prof. Courtney Aldrich, Prof. David Sherman, Prof. Yossef Av-Gay, Prof. Matthias Wilmanns, Dr Stacey Southall, Dr Argyrides Argyrou, Prof. Lluís Ribas and Prof. Hugo Gramajo. We thank the designers at The Moment for their contributions to Fig. 1, Stacey Walsh for sequencing work and Stephen Bach for contributions to the bamA MoA studies. We also thank Craig Volker and Jim Brown for bioinformatic analysis and Allison L. Haigney for help with editing the manuscript.

## Authors contributions

P.A.C., J.C., C.P.D., Y.D. and X.L. synthesized DNA-encoded libraries. C.A.M., P.F.C., J.H., L.B., G.S.B., D.B.-A., R.H.B., J.L., A.M.-L., S.M.S., L.M.S., C.P.D., D.H., A.J.P. and

J.W.G. selected targets for screening. C.A.M., C.S.K., K.E.L., X.B., D.B.-A., Q.L., A.M.-L., S.M.S., D.H., C.C.A.-M., A.J.P., J.W.G. and G.E. proposed and evolved the concept. K.L., T.L.G., Q.L., R.L., J.M., P.M., H.Q., A.T. and J.W. performed protein expression, cloning and purification. C.A.M., C.S.K., K.E.L., X.B., J.H., S.B., D.B.A., R.H.B., A.M.-L., C.P., C.B.P., C.P.D., C.C.A.-M., A.J.P., J.W.G. and G.E. defined screening strategy and supervised projects. C.A.M., C.S.K., E.N.G., Q.L., H.P.O.K., T.O.K., C.P., M.S.S., F.S.S., L.M.S. and J.Z. performed ELT selections. S.S.C., T.O.K., G.S.S., M.S.S., F.S.S. and J.Z. did amplification, DNA sequencing, data entry and analysis. C.A.M., C.S.K., X.B., K.E.L., R.H.B., D.T.F., T.L.G., B.W.K., B.X., G.Y., J.Z., M.S.S., J.A.M. and G.E. did selection data analysis and off-DNA chemistry planning. X.B., R.H.B., A.C., H.D., J.D., J.W.D., D.T.F., T.L.G., B.W.K., Y.L., D.J.W., B.X., G.Y., J.Z. and G.E. did chemistry planning and compound synthesis. P.F.C., J.H., A.E.C., L.M., K.R. and M.F.T. did assay development and compound testing. W.P.J., C.R.K. and M.W. did compound purification, HRMS and NMR of final compounds. C.A.M., C.S.K., K.E.L., X.B., P.F.C. and G.E. did article write-up.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing interests:** With the exception of G.S.B., all the other authors are either currently or in past have been employees and stockholders of GlaxoSmithKline.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Machutta, C. A. & Kollmann, C. S. *et al.* Prioritizing multiple therapeutic targets in parallel using automated DNA-encoded library screening. *Nat. Commun.* **8**, 16081 doi: 10.1038/ncomms16081 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017