1   **TargetSeeker-MS: A Computational Method for Drug Target Discovery using Protein**

2   **Separation Coupled to Mass Spectrometry**

3   Mathieu Lavallée-Adam[1,2‡], Alexander Pelletier[1,2‡], Jolene K. Diedrich[1,3‡], William Low[3],

4   Antonio F. M. Pinto[1,3], Salvador Martínez-Bartolomé[1], Michael Petrascheck[1], James J.

5   Moresco[1,3], John R. Yates III[1,3*]

6

7   ‡ These authors contributed equally to this work.

8   [1] Department of Molecular Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Rd.,

9   La Jolla, California, 92037, USA.

10   [2] (Current affiliation) Department of Biochemistry, Microbiology and Immunology and Ottawa

11   Institute of Systems Biology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario K1H

12   8M5, Canada

13   [3] The Mass Spectrometry Core for Proteomics and Metabolomics, The Salk Institute for

14   Biological Studies, 10010 N Torrey Pines Rd., La Jolla, California, 92037, USA.

15

16   *Corresponding Author:

17   **Dr. John R. Yates III**

18   Department of Molecular Medicine

19   The Scripps Research Institute

20   10550 North Torrey Pines Road, SR302

21   La Jolla, California, 92037, USA

22   Phone: 1-858-784-8862

23   Fax: 1-858-784-8883

24   E-mail: jyates@scripps.edu

25

26

27

28

29

30

ABSTRACT

When coupled to mass spectrometry (MS), energetics-based protein separation (EBPS) techniques, such as thermal shift assay, have shown great potential to identify the targets of a drug on a proteome scale. Nevertheless, the computational analyses assessing the confidence of drug target predictions made by these methods have remained rudimentary and significantly differ depending on the protocol used to produce the data. To identify drug targets in datasets produced using different EBPS-MS techniques, we have developed a novel flexible computational approach named TargetSeeker-MS. We showed that TargetSeeker-MS reproducibly identifies known and novel drug targets in *C. elegans* and HEK293 samples that were treated with the fungicide benomyl and processed using two different EBPS techniques. We also validated a novel benomyl target in vitro. TargetSeeker-MS, which is available online, allows for the confident identification of targets of a drug on a proteome scale, thereby facilitating the evaluation of its clinical viability.

Keywords:

Bioinformatics / Drug target discovery / Mass spectrometry / Proteomics / Thermal shift assay.

INTRODUCTION

Methods that can identify the putative protein targets of a drug on a proteome-scale are critical to decipher the mechanism of action of a compound. Such methods include large-scale phenotypic screenings, which are often performed to evaluate the ability of a library of compounds to modulate a given target pathway and therefore provide a putative treatment for the associated disease process. High-throughput screening (HTS) methods, which use automated high-end instrumentation, can test the properties of up to 100,000 compounds per day. However, high-throughput screening is typically performed on isolated systems and offers minimal insights regarding secondary protein targets[1].

Mass spectrometry (MS)-based proteomics allows the large-scale identification and quantification of proteins in complex samples (e.g. cells, tissues, plasma). MS coupled to affinity-based enrichment strategies has routinely been used to identify protein interactions of compounds in the cell[2–4]. In the recent years, energetics-based protein separation (EBPS) techniques coupled to MS have emerged as a large-scale approach to rapidly and unbiasedly identify the protein targets of a given compound or drug. Such approaches rely on the hypothesis that the target of a drug will see its stability or thermodynamic properties changed upon binding with the compound. We previously demonstrated that Stability of Proteins from Rates of Oxidation (SPROX) combined with quantitative MS can be used to identify the protein targets of a drug[5]. This strategy uses the chemical denaturant-dependent oxidation rates of methionine residues to measure the thermodynamics of the unfolding or refolding reaction of proteins in drug-treated and untreated samples. A significant change in a given protein's thermodynamics would indicate its binding to the drug. In this workflow, proteins were quantified using 6-plex tandem mass tags (TMT)[6] stable isotope labeling coupled to MS. An increasingly popular EBPS

3

1   coupled to MS (EBPS-MS) approach assesses ligand binding by evaluating changes in thermal

2   stability[7,8]. This approach is based on the hypothesis that a protein bound by a given drug would

3   see its thermal stability changed by the binding. A thermal shift assay technique was also

4   employed in cells (CETSA)[9]. Savitski et al. demonstrated that combining a thermal shift assay

5   approach with quantitative mass spectrometry allows the large-scale unbiased identification of

6   drug targets[10]. These studies showed that EBPS-MS is complementary to HTS. Indeed, while

7   HTS can process tens of thousands of compounds simultaneously and EBPS-MS only

8   investigates one compound at a time, EBPS-MS allows the testing of this compound against the

9   proteome of a sample, a much more complex system than what is tested in traditional HTS.

10   　　While EBPS-MS-based techniques for drug target discoveries are emerging rapidly, no

11   general computational frameworks for the unbiased identification of drug-protein interactions

12   have been produced. One of the reasons why the adoption of EBPS-MS technologies has been

13   slow is because there are currently no implementations available to assess the confidence or

14   statistical significance that a protein is bound by a given drug based on its change in stability

15   evaluated by an EBPS-MS approach. The current practices involve the use of in-house

16   computational scripts with numerous custom thresholds, making the comparisons and

17   benchmarking of results across different laboratories extremely difficult[5,10]. Savitski et al.

18   presented a statistical approach using curve fitting and statistics, which rely on largely

19   unsupported assumptions of the normality distribution of the data[11]. In addition, this statistical

20   approach is tied to the thermal shift assay experimental protocol using a 10-plex tandem mass tag

21   as protein quantification technique[10] and is not readily applicable to the other protein EBPS-MS

22   quantification techniques that may be used to identify drug targets. For instance, we recently

23   developed DiffPOP[12], a novel EBPS technique that allows efficient separation of complex

1 protein samples using an increasing concentration of a solution of acetic acid and methanol.

2 DiffPOP was recently used to identify the targets of JIB-04, a compound that blocks the

3 expression and transactivation of HIV-1 Tat[12]. DiffPOP differentiates itself from the Savitski et

4 al. thermal shift assay approach by using MS to quantify proteins precipitated in each fraction

5 instead of proteins remaining in the supernatant. The statistical method proposed by Savitski et

6 al., which relies on melting curves based on supernatant analysis, is therefore not applicable for

7 such a technique nor the SPROX quantitative MS approach. Furthermore, the use of TMT 10-

8 plex reagents in Savitski et al.'s quantitative proteomics analysis suffers from a number of

9 drawbacks. As in any TMT labeling analysis, low abundance proteins are less likely to be

10 quantified, making it difficult to detect low abundance drug targets[13]. Furthermore, the price of

11 TMT 10-plex reagents[14] and the high resolution instruments necessary for TMT 10-plex

12 analysis[15] limits the democratization of Savitski et al.'s drug target discovery approach.
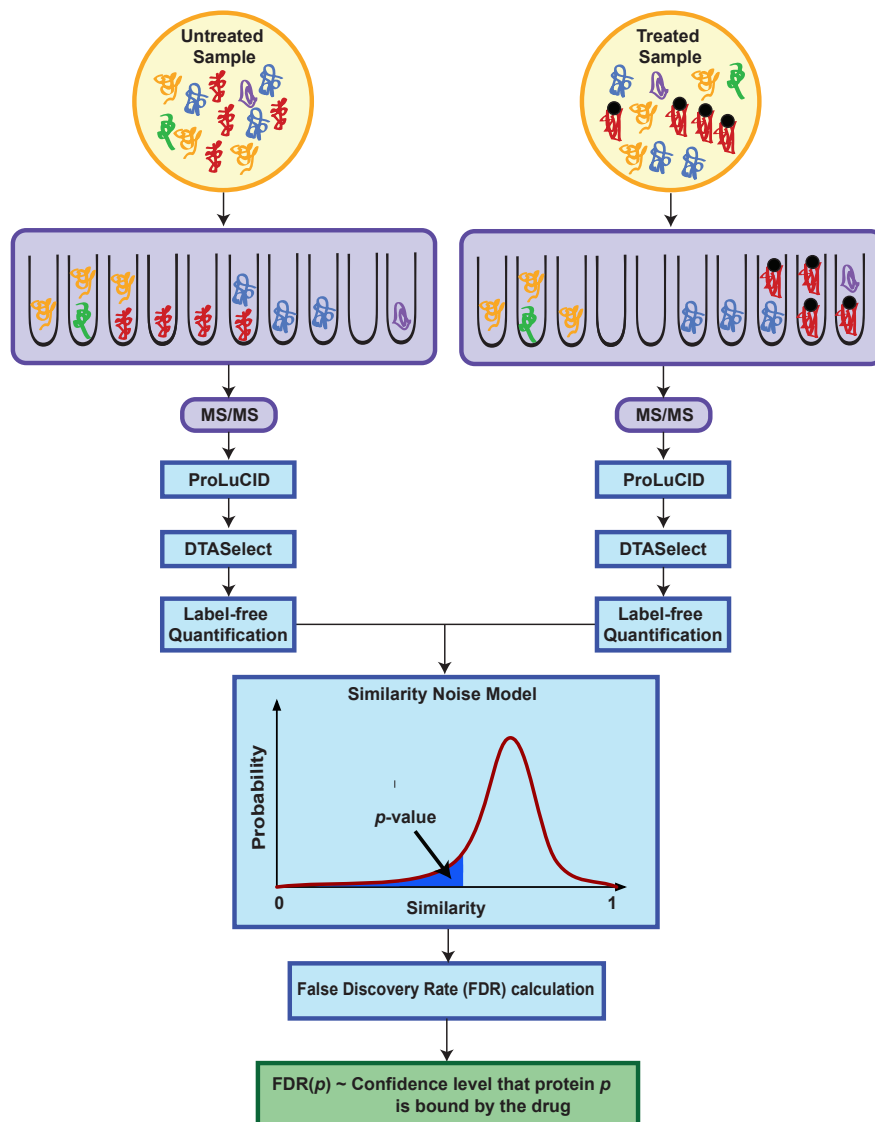
13 Herein, we propose a general computational framework, TargetSeeker-MS, for the

14 identification of drug targets using EBPS coupled to quantitative MS. TargetSeeker-MS

15 implements a Bayesian inference machine learning approach to assess the confidence that a

16 protein is bound by a given compound. We demonstrate that TargetSeeker-MS, which is open-

17 source and available as a user-friendly web-server, is hypothesis-free and flexible enough to

18 analyze datasets originating from any EBPS-MS techniques. TargetSeeker-MS identified

19 putative targets of benomyl, a fungicide putatively linked to Parkinson's disease[16,17], in two *C.*

20 *elegans* datasets analyzed using DiffPOP and thermal shift assay coupled to MS. We showed that

21 although both fractionation methods vary in nature, TargetSeeker-MS predictions in both

22 datasets share a significant overlap of confident targets. In addition, we demonstrated that the

23 TargetSeeker-MS algorithm predicts drug targets with a greater sensitivity than previously

1   proposed approaches. Benomyl is known to inhibit aldehyde dehydrogenase (ALDH), a

2   mechanism putatively leading to Parkinson's disease development[16]. TargetSeeker-MS identified

3   aldehyde dehydrogenase as a benomyl target along with other known and novel targets. Finally,

4   we highlight that TargetSeeker-MS identified human benomyl target orthologs when processing

5   a HEK293 cells dataset treated with the drug and validated the impact of benomyl on the

6   enzymatic activity of one of its novel predicted targets, GAPDH.

7

RESULTS

9   TargetSeeker-MS is a Bayesian inference-based approach that computes the probability that a

10   protein is bound by a given drug through the analysis of EBPS-MS datasets. Briefly,

11   TargetSeeker-MS takes as input a set of untreated (control) samples that were processed using

12   EBPS and quantified using MS and builds for each protein a noise model of the similarity of the

13   protein fractionation profiles in different biological replicates. It then evaluates the similarity of

14   these control protein fractionation profiles with that of a drug-treated sample that was also

15   separated using the same EBPS-MS approach. TargetSeeker-MS then assesses the confidence

16   that each protein is bound by the drug. Figure 1 provides a graphical representation of

17   TargetSeeker-MS' pipeline. In this study, we used TargetSeeker-MS to identify the proteins

18   bound by benomyl in three different datasets. The first two datasets analyzed *C. elegans* samples

19   and were produced using DiffPOP separation (see Methods) coupled to MS (Dataset 1 –

20   DiffPOP/*C. elegans*; Supplementary Figure S1) and Thermal Shift Assay (TSA) separation

21   coupled to MS (Dataset 2 – TSA/*C. elegans*; Supplementary Figure S1). The third dataset

1   represents the processing of Human Embryonic Kidney 293 cells (HEK 293) using DiffPOP-MS

2   (Dataset 3 – DiffPOP/HEK293; Supplementary Figure S1).

3

4   **Figure 1. Experimental and computational pipeline.** A graphical representation of the drug

5   target identification pipeline illustrating the protein separation, the mass spectrometry analysis,

6   and the TargetSeeker-MS algorithm.

7

1 **Protein fractionation profiles display a high level of reproducibility.**

2 To begin, we investigated the level of reproducibility of protein (p) fractionation profiles $\hat{F}_p^u$

3 across different biological replicates of untreated biological samples (u) in order to assess the

4 possibility of building a model of the noise of fractionation profile similarity values with a small

5 number of replicates. With this objective in mind, we computed the similarity $S_p^{u_1, u_2}$ for all

6 proteins p and all untreated samples u_1 and $u_2 \in U$, where U is the set of untreated samples in a

7 dataset. The vast majority of the protein fractionation profiles shared a high degree of similarity

8 (Figure 2A and Supplementary Figures S2 and S3). For instance, 79% of all pairs of

9 fractionation profiles in untreated samples have a similarity $S_p^{u_1, u_2} > 0.75$ in the DiffPOP-*C.*

10 *elegans* dataset (Dataset 1). These similarity values demonstrate that both EBPS-MS approaches

11 yield a high level of fractionation reproducibility, which is likely to be sufficient to compute the

12 probability matrix $\Pr\left[S_p^{u_1, u_2} \middle| \overline{S_p^u}\right]$ and to build an accurate noise model of the similarity values of

13 fractionation profiles under the null hypothesis (untreated samples) with a small number of

14 replicates (see Methods). Supplementary Figure S4 displays the probability matrix $\Pr\left[S_p^{u_1, u_2} \middle| \overline{S_p^u}\right]$

15 for Dataset 1. Protein fractionation profiles also shared a high similarity in benomyl treated

16 samples (Supplementary Figure S5, S6, and S7). To illustrate this, in Dataset 1 83% of all pairs

17 of fractionation profiles obtained a similarity $S_p^{t_1, t_2} > 0.75$, where t_1 and $t_2 \in T$ the set of all

18 treated samples. The similarity values between the untreated and benomyl treated samples are

19 similarly distributed in all three datasets with only the DiffPOP-HEK293 dataset (Dataset 3)

20 showing a slight difference between the two distributions (Figure 2B, Supplementary Figure S8,

21 and S9). Overall, the distributions of similarity values highlight the feasibility of
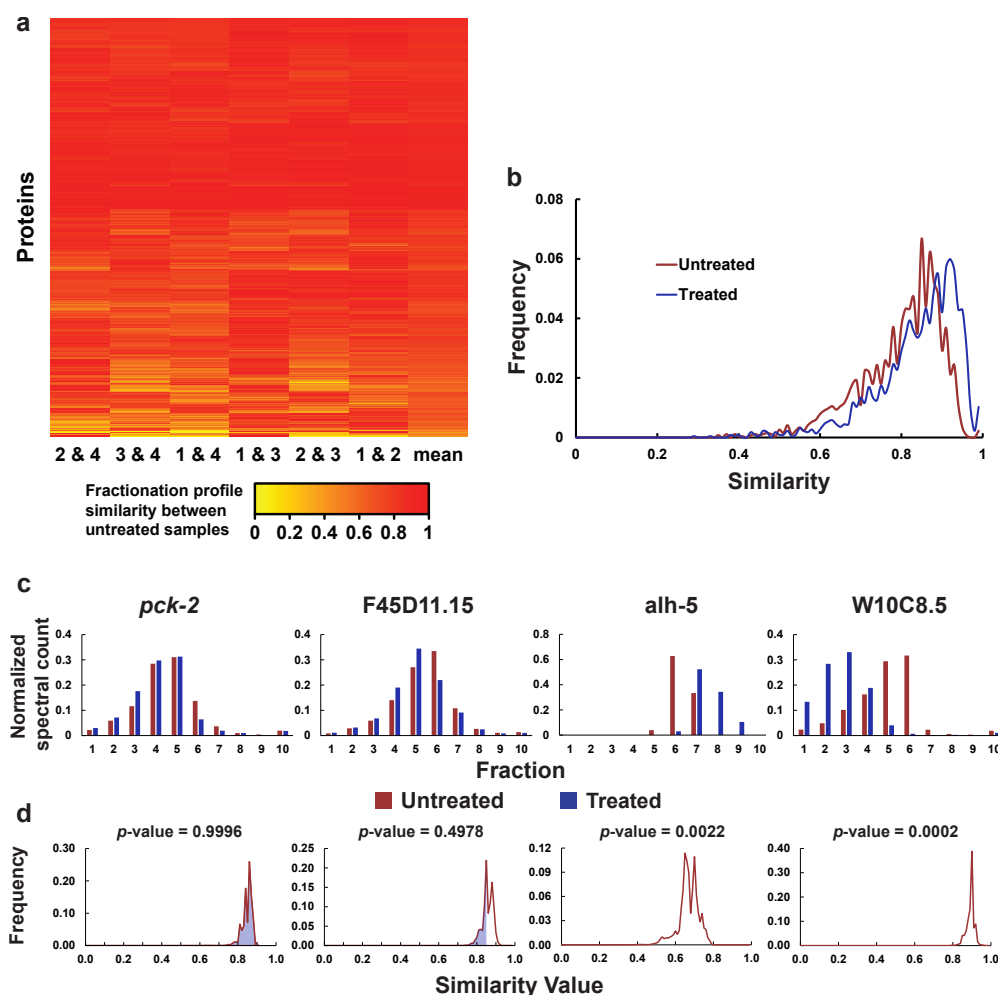
8

Figure 2. Fractionation profile similarity. (A) Heatmap representation of the similarity between protein fractionation profiles for all pairs of untreated samples of the DiffPOP/*C. elegans* dataset. All proteins with a sufficient abundance to compute a fractionation profile in all untreated samples are displayed. (B) Distributions of the fractionation profile similarity values in both untreated and benomyl treated samples of the DiffPOP/*C. elegans* dataset. (C) Fractionation profiles represented as normalized spectral counts in each fraction and (D) posterior distributions of $\overline{S_p^u}$ of four different C. *elegans* proteins identified using DiffPOP-MS. The shaded portion of the distributions represents the p-value$\left(\overline{S_p^{u,t}}\right)$ for each protein. The shaded portion of W10C8.5 and alh-5 is not visible since it spans from (0 to 0.62) and (0 to 0.38), respectively.

9

1   creating a noise model for the similarity between the fractionation profiles of a given protein

2   using a small number of biological replicates.

3

4   **TargetSeeker-MS assesses the statistical significance of protein fractionation profile changes**

5   **upon benomyl treatment.**

6   The ability of the algorithm to assess the significance of the change in the fractionation profile of

7   a protein upon drug treatment was illustrated with four protein examples from Dataset 1 (Figure

8   2C and 2D). *Pck-2* represents an example of a protein without a change in its fractionation

9   profile upon benomyl treatment (Figure 2C). Indeed, both distributions of normalized spectral

10   counts are almost identical (Similarity $\overline{S_p^{u,t}}$ = 0.89). The associated *p*-value (>0.99; FDR=1.0)

11   computed by TargetSeeker-MS is therefore very high (see Methods for *p*-value calculation). On

12   the other hand, F45D11.15 appears to display a small shift to the left in its fractionation profile

13   upon benomyl treatment (Similarity $\overline{S_p^{u,t}}$ = 0.85). However, this change in fractionation profile is

14   too minor to be deemed significant by TargetSeeker-MS (*p*-value = 0.5; FDR = 1.0), since it may

15   simply be due to noise as indicated by its high FDR. *Alh-5*, a protein known to be bound by

16   benomyl[16,18] clearly displays a significant change in its fractionation profile, showing an

17   increased precipitation resistance upon benomyl treatment (Similarity $\overline{S_p^{u,t}}$ = 0.38). Due to the

18   amplitude of this shift, TargetSeeker-MS assigned a *p*-value of 0.0022 (FDR=0.005) to *alh-5*.

19   Conversely, W10C8.5, an ortholog of a human creatine kinase, also sees its fractionation profile

20   drastically modified by benomyl (Similarity $\overline{S_p^{u,t}}$ = 0.62), but in this case a decrease in

21   precipitation resistance is observed (*p*-value = 0.0002; FDR< 0.005). This result is consistent

1   with previous reports, which showed that creatine kinase enzymatic activity is altered by

2   benomyl[19,20].

3

4   **TargetSeeker-MS identifies high-confidence benomyl targets in *C. elegans* samples that were**

5   **analyzed using DiffPOP-MS.**

6   We tested the ability of TargetSeeker-MS to identify benomyl targets in *C. elegans* through the

7   analysis of a dataset produced with DiffPOP-MS (Dataset 1). TargetSeeker-MS built the

8   similarity noise model with a set of four untreated biological replicates. With this noise model

9   established, we used TargetSeeker-MS to evaluate the confidence that the precipitation resistance

10   of proteins quantified in three biological replicates of benomyl treated *C. elegans* samples was

11   altered. We first analyzed all drug treated samples as a group in a single TargetSeeker-MS

12   analysis, computing the average of the similarity values between $\hat{F}_p^t$ and $\hat{F}_p^u$ for all treated

13   samples t and all untreated samples u. TargetSeeker-MS identified 59 proteins with a FDR <

14   0.01 and 101 with a FDR < 0.05 (Supplementary Table S1).

15   We benchmarked the TargetSeeker-MS analysis against two alternate approaches: the *Z*-

16   score method and the Savitski et al. statistical approach (see Methods) (Figure 3A). Using high-

17   confidence FDRs, TargetSeeker-MS reported more drug target predictions than either of the

18   other two methods. To maximize the stringency of TargetSeeker-MS drug target predictions, in

19   addition to the FDR threshold, a Fold-change of Similarity Difference (FSD) threshold of 0.20

20   was applied to each protein (see Methods). Using both thresholds, TargetSeeker-MS identified

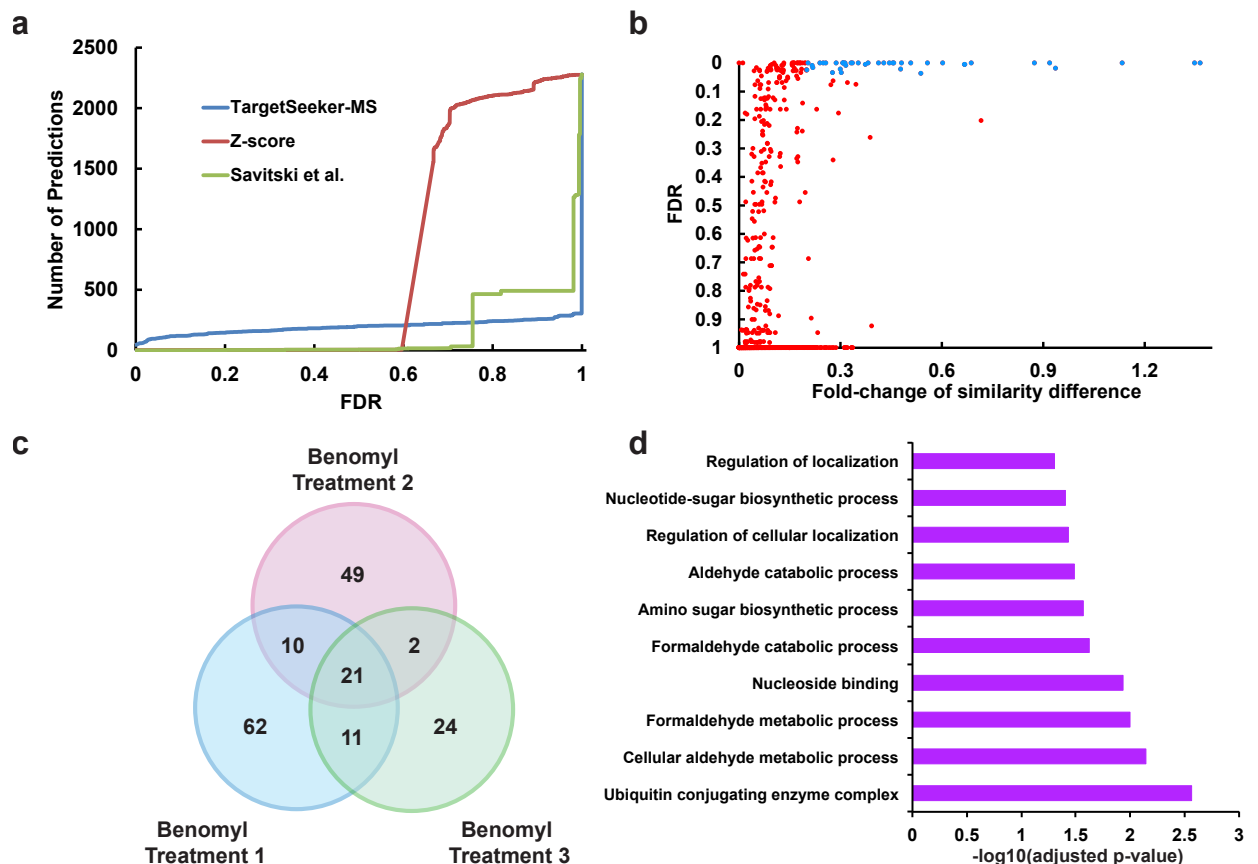21   41 proteins with a FDR < 0.05 and a FSD > 0.2 (Figure 3B).

22

Figure 3. Analysis of the DiffPOP/*C. elegans* dataset (A) Cumulative distributions of the number of target predictions at a given FDR obtained by TargetSeeker-MS, the Z-score method, and the Savitski et al. approach. (B) FDR and fold-change of similarity difference of all proteins to which a fractionation profile was assigned. High-confidence drug targets (FDR < 0.1, FSD > 0.2) are represented in blue, while low-confidence predictions are shown in red. (C) Venn diagram representation of the high-confidence predictions of TargetSeeker-MS when processing each drug treated biological replicate independently. (D) Gene Ontology terms that are enriched among the drug targets identified by TargetSeeker-MS in all three biological replicates.

12

1   While TargetSeeker-MS can assess the significance of the average of the similarity

2   values between $\hat{F}_p^t$ and $\hat{F}_p^u$ for all treated samples t and all untreated samples u, it can also assess

3   the significance of the similarity between $\hat{F}_p^t$ and $\hat{F}_p^u$ for all untreated samples and a single given

4   treated sample t. This allows TargetSeeker-MS to independently identify drug targets in each

5   treated sample. Proteins that are reproducibly predicted as drug targets in all treated samples

6   (biological replicates) therefore represent very high confidence predictions. To once again

7   maximize the stringency of our analysis, we only considered proteins as drug targets if they were

8   identified as high-confidence targets by TargetSeeker-MS (FDR <0.1, FSD > 0.2) in all treated

9   samples (Figure 3C; Supplementary Table S2-S5). The resulting 21 proteins are reported in

10   Table 1. This list includes alh-5, an aldehyde dehydrogenase. It was previously reported that

11   benomyl inhibits the low-Km hepatic mitochondrial aldehyde dehydrogenase of mice[18] and that

12   the inhibition of aldehyde dehydrogenase may lead to preferential development of Parkinson's

13   disease[16]. TargetSeeker-MS also identified a number of creatine kinases (F32B5.1, F44G3.2,

14   W10C8.5, ZC434.8) as high-confidence benomyl targets. These results are consistent with

15   studies demonstrating that the enzymatic activity of a creatine kinase was altered by benomyl in

16   rats[19] and in human serum[20]. Unc-25, a *C. elegans* ortholog of the GABA neurotransmitter

17   biosynthetic enzyme that was not previously reported as a benomyl target, was also identified by

18   TargetSeeker-MS to be affected by the compound. The 21 targets identified by TargetSeeker-MS

19   are significantly enriched for Gene Ontology (GO) terms (functions and biological processes)

20   that have been reported to be affected by benomyl, such as "Aldehyde catabolic process" and

21   "Cellular aldehyde metabolic process" (Figure 3D; Supplementary Table S6 for complete

22   enrichment analysis results)[16,18].

23

Table 1: Benomyl targets as predicted by TargetSeeker-MS in the *C. elegans*/DiffPOP dataset.[1]

Wormbase ID	Gene Name[2]	Untreated Similarity	Drug vs Untreated Similarity	*p*-value	FDR	Fold-change of similarity difference
F44G3.2	CKB*, CKM*, CKMT2*, CKMT1B*	0.90	0.70	3.2×10^{-5}	<0.005	0.29
Y54G2A.31	ubc-13	0.90	0.65	3.4×10^{-5}	<0.005	0.38
F20G2.2	RDH8, DHRS1 *	0.85	0.56	8.0×10^{-5}	<0.005	0.51
T07C4.9b	nex-2	0.87	0.60	2.0×10^{-4}	<0.005	0.44
T22D1.3a	IMPDH2*, IMPDH1*	0.81	0.55	2.0×10^{-4}	<0.005	0.48
W10C8.5	CKB*, CKMT2*, CKMT1B*	0.90	0.62	2.0×10^{-4}	<0.005	0.46
Y37D8A.23a	unc-25	0.83	0.52	2.0×10^{-4}	<0.005	0.60
Y37D8A.23b	unc-25	0.82	0.52	2.0×10^{-4}	<0.005	0.56
ZC434.8	CKB*, CKMT2*, CKMT1B*	0.89	0.71	2.0×10^{-4}	<0.005	0.25
F32B5.1	CKB*, CKM*, CKMT2*, CKMT1B*	0.89	0.61	4.0×10^{-4}	<0.005	0.46
T22D1.3b	IMPDH2*, IMPDH1*	0.77	0.45	5.0×10^{-4}	<0.005	0.69
Y42G9A.4c	mvk-1	0.82	0.63	1.1×10^{-3}	<0.005	0.31
C36A4.4	UAP1*, UAP1L1*	0.81	0.57	1.2×10^{-3}	<0.005	0.43
Y37D8A.23c	unc-25	0.67	0.36	1.2×10^{-3}	<0.005	0.87
K06B9.2	UAP1*, UAP1L1*	0.80	0.57	1.5×10^{-3}	<0.005	0.41
Y48G10A.1	ESD*	0.73	0.44	1.9×10^{-3}	0.005	0.67
T08B1.3	alh-5	0.64	0.38	2.2×10^{-3}	0.005	0.67
F21C3.3	hint-1	0.54	0.28	5.4×10^{-3}	0.019	0.94
D2063.3a	glrx-3	0.78	0.60	6.0×10^{-3}	0.021	0.30
R07E3.1a	CTSF	0.72	0.49	6.2×10^{-3}	0.021	0.48
D2063.3b	glrx-3	0.77	0.59	9.9×10^{-3}	0.034	0.30

[1] A yellow row represents a protein that was not quantified with enough confidence in the TSA/*C. elegans* dataset to allow TargetSeeker-MS to assess the significance of benomyl binding. A green row represents a protein that was also predicted as a benomyl target in the TSA/*C. elegans*.
[2] A * indicates the human ortholog gene name of a *C. elegans* protein without a name.

1   **TargetSeeker-MS identified high confidence benomyl targets in the TSA/*C. elegans* dataset.**

2   In order to test the ability of the TargetSeeker-MS algorithm to accurately identify drug targets in

3   samples that were separated using a different EBPS-MS method, we applied our algorithm to the

4   TSA-*C. elegans* dataset (Dataset 2). Combining two replicate TSA analyses of benomyl-treated

5   samples into a single TargetSeeker-MS analysis allowed the algorithm to identify 278 benomyl

6   targets with a FDR < 0.01 and 331 benomyl targets with a FDR < 0.05 (Supplementary Table

7   S7).

8   Even though Dataset 2 was processed using a similar approach to that described in

9   Savitski et al. (i.e. using the same mechanism: heat destabilization), TargetSeeker-MS again

10 identified more drug targets at high-confidence FDRs than the Z-score approach or the Savitski

11 et al. method (Figure 4A). Interestingly, 285 proteins in Dataset 2 that were assigned a FDR <

12 0.05, were associated with a relatively low FSD (< 0.2) (Figure 4B). This result is likely

13 explained by the smaller variance in similarity values of fractionation profiles in replicate

14 samples obtained with TSA-MS compared to DiffPOP-MS (Figure 2B and Supplementary

15 Figure S8). This decreased variation in fractionation profiles therefore allows TargetSeeker-MS

16 to assign high-confidence FDRs to even small changes in protein fractionation profiles upon

17 drug treatment. To maximize sensitivity we therefore opted to not apply a FSD threshold for all

18 data produced using TSA-MS. Nevertheless, in order to maintain a high level of confidence in

19 TargetSeeker-MS predictions, we only retained as putative drug targets proteins that were

20 assigned by the algorithm a FDR < 0.05 in both replicate TSA analyses of benomyl-treated

21 samples when analyzed independently (Supplementary Table S8, S9, and S10). This process

22 yielded a list of 154 benomyl-binding proteins in the TSA/*C. elegans* dataset. These proteins are

23 statistically significantly enriched for GO terms such as "Microtubule" (adjusted *p*-value =
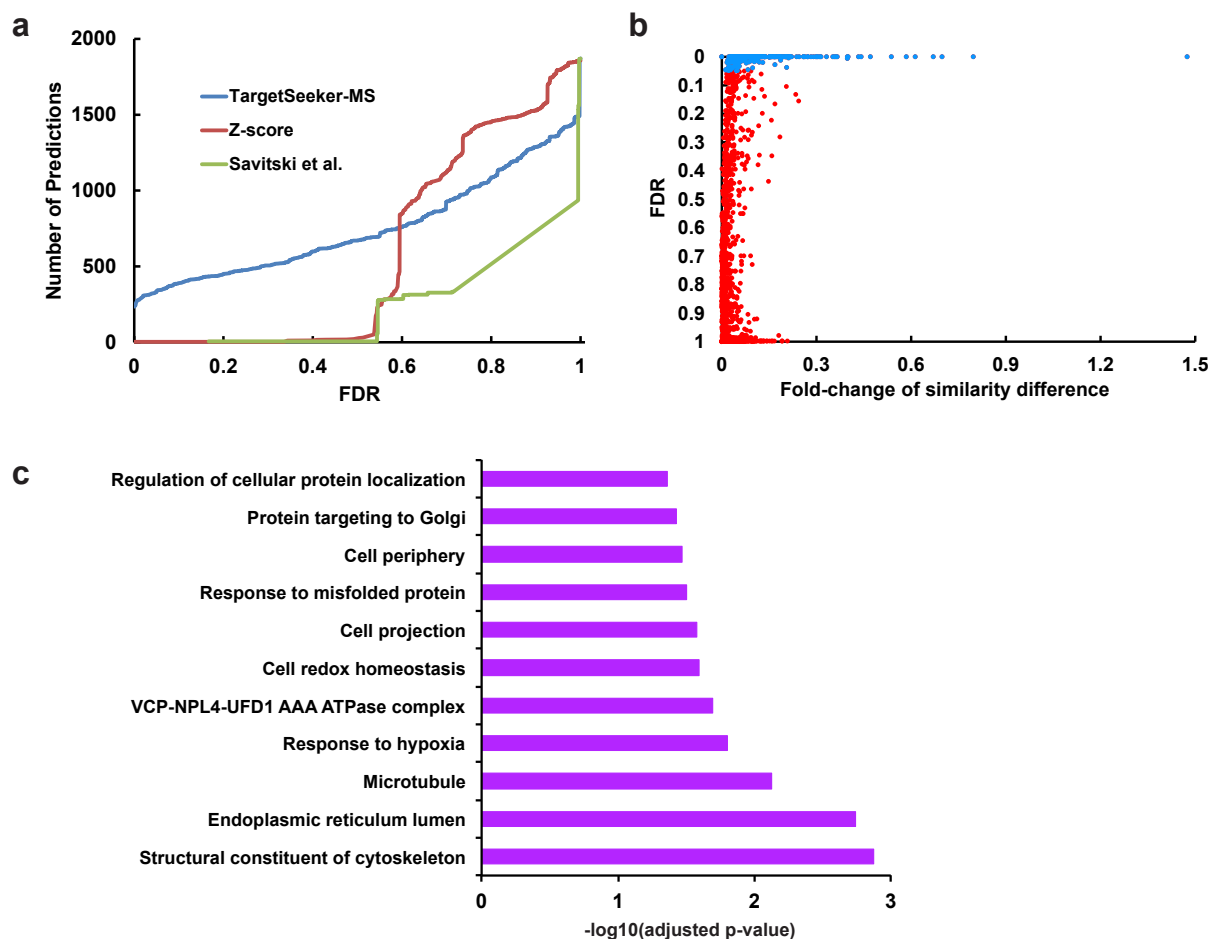
15

Figure 4. Analysis of the TSA/*C. elegans* dataset (A) Cumulative distributions of the number of target predictions at a given FDR obtained by TargetSeeker-MS, the Z-score method, and the Savitski et al. approach. (B) FDR and fold-change of similarity difference of all proteins to which a fractionation profile was assigned. High-confidence drug targets (FDR < 0.05) are represented in blue, while low-confidence predictions shown in red. (C) Gene Ontology terms that are enriched among the drug targets identified by TargetSeeker-MS in both biological replicates.

1   0.0075), "Regulation of cellular protein localization" (adjusted p-value = 0.043), "Structural

2   constituent of cytoskeleton" (adjusted p-value = 0.0013), and "Cell periphery" (adjusted p-value

3   = 0.033) (Figure 4C and Supplementary Table S11). These GO term enrichments are consistent

4   with the role of benomyl, which was previously reported as a compound that depolymerizes

5   microtubules near the cell periphery[21–23].

6

7   **Drug targets identified in the TSA/*C. elegans* dataset significantly overlap with those**

8   **identified in the DiffPOP/*C. elegans* dataset.**

9   Among the *C. elegans* benomyl targets identified in the TSA dataset, we count a number of

10   creatine kinases (F44G3.2, W10C8.5, and F32B5.1) that were also identified in the DiffPOP-

11   MS/*C. elegans* dataset. However, alh-5 a known target of benomyl was not quantified with

12   enough spectral counts to be associated with a fractionation profile and could therefore not be

13   assessed by TargetSeeker-MS (see Methods). Even though TSA and DiffPOP use different

14   mechanisms to separate proteins, the intersection of the sets of drug targets identified by

15   TargetSeeker-MS in both datasets is significant (hypergeometric test p-value = 0.02). Indeed, 8

16   out of the 11 drug targets from the DiffPOP/*C. elegans* that were quantified accurately enough in

17   the TSA/*C. elegans* dataset to be analyzed by TargetSeeker-MS were also identified as drug

18   targets in the latter dataset (Table 1). These results highlight the ability of TargetSeeker-MS to

19   reproducibly detect protein targets in datasets originating from different EBPS-MS techniques.

20

21

1 **TargetSeeker-MS identifies high-confidence benomyl targets in HEK 293 that are *C. elegans***

2 **orthologs.**

3 We executed the TargetSeeker-MS algorithm on the DiffPOP/HEK293 datasets (Dataset 3) to

4 obtain some insights regarding human proteins that may be bound by benomyl. TargetSeeker-

5 MS identified 349 human drug targets with a FDR < 0.01 and 566 proteins with an FDR < 0.05

6 (Supplementary Table S12). Once again, TargetSeeker-MS outperformed the *Z*-score method

7 and the Savitski et al. statistical approach by detecting more benomyl targets at high confidence

8 FDRs (Figure 5A). To maximize the specificity of the analysis, a FSD threshold of 0.2 was

9 applied yielding 77 high-confidence drug predictions with a FDR < 0.01 and 94 with a FDR <

10 0.05 (Figure 5B; Supplementary Table S12). To maximize the specificity of the drug predictions

11 as for the other datasets, we only retain as putative drug targets proteins that were assigned by

12 the algorithm a FDR < 0.05 in both benomyl-treated replicates when analyzed independently

13 (Supplementary Table S13, S14, and S15). This resulted in a list of 22 high-confidence benomyl

14 targets (Supplementary Table S15). Among these targets we find three proteins, CKMT1A (FDR

15 < 0.001), CKMT2 (FDR < 0.001), and UBE2O (FDR = 0.005) that correspond to orthologs of *C.*

16 *elegans* proteins (E-value < 10^{-6} and sequence identity > 50%; see Methods) that were also

17 predicted by TargetSeeker-MS to be high-confidence benomyl targets in the DiffPOP/*C. elegans*

18 dataset (Dataset 1). In addition, CKMT1A and CKMT2 orthologs were also identified as

19 benomyl targets in the TSA/*C. elegans* dataset (Dataset 2). Of note, creatine kinases were

20 identified as benomyl targets in all three datasets that involve two different EBPS-MS techniques

21 analyzing two different organisms. Among the notable GO term enrichments present in the list of

22 22 benomyl targets, we find "Creatine metabolic process" and "Creatine kinase activity", which

1 are biological processes that have been reported to be linked to benomyl[19,20] (Figure 5C and
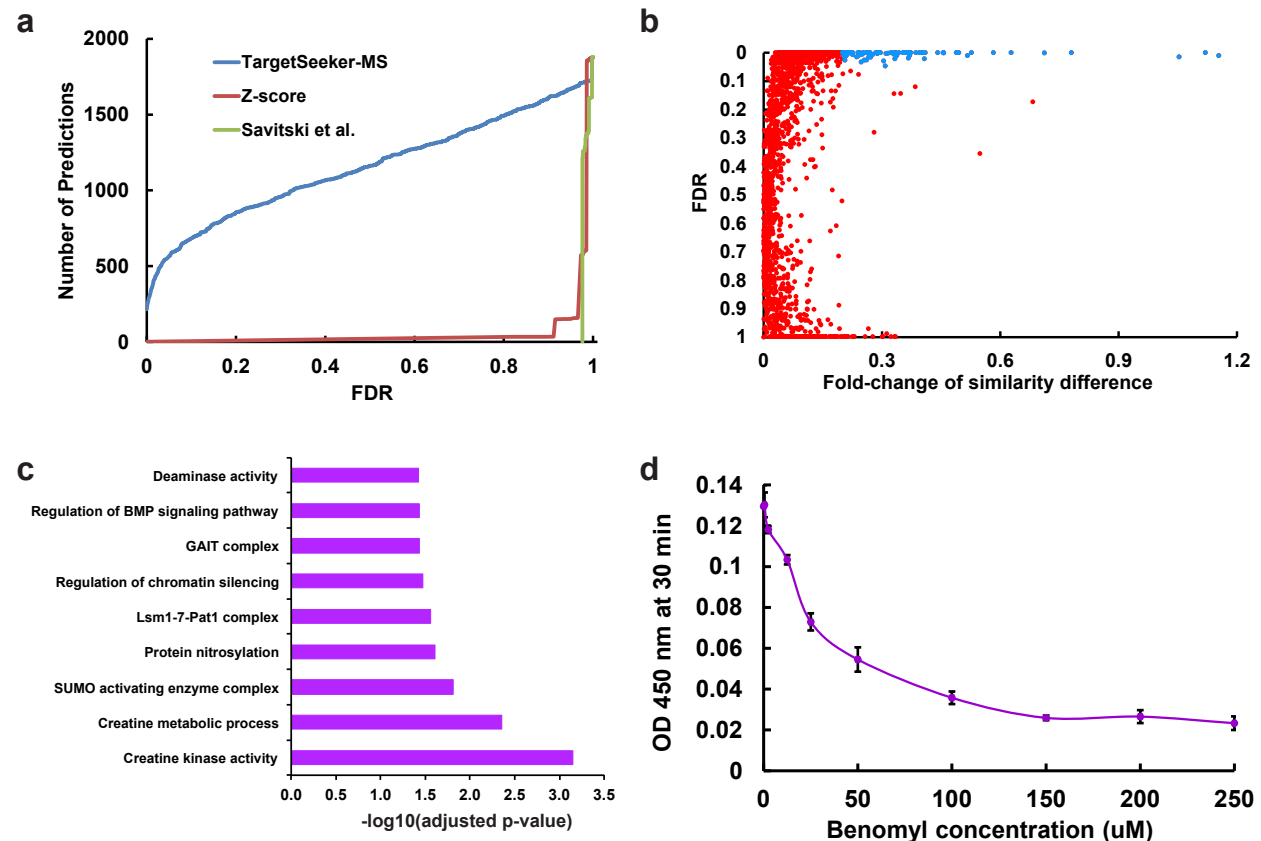
2 Supplementary Table S16).

4 **Figure 5. Analysis of the DiffPOP/HEK 293 dataset** (A) Cumulative distributions of the

5 number of target predictions at a given FDR obtained by TargetSeeker-MS, the Z-score method,

6 and the Savitski et al. approach. (B) FDR and fold-change of similarity difference of all proteins

7 to which a fractionation profile was assigned. High-confidence drug targets (FDR < 0.05, FSD >

8 0.2) are represented in blue, while low-confidence predictions shown in red. (C) Gene Ontology

9 terms that are enriched among the drug targets identified by TargetSeeker-MS in both biological

10 replicates. (D) Efficiency of benomyl inhibition of GAPDH enzymatic activity upon mixing with

11 different benomyl concentrations. Points represent the mean Optical Density (OD) and error bars

12 represent the standard deviation across three replicates.

19

TargetSeeker-MS novel benomyl target discoveries can be confirmed in vitro.

GAPDH is among the *H. sapiens* proteins for which a *C. elegans* ortholog was not a predicted as a benomyl target. To our knowledge, GAPDH activity was never previously reported to be affected by benomyl nor was the protein shown to be bound by the compound. We therefore validated the TargetSeeker-MS' prediction from the Dataset 3 using an orthogonal in vitro approach (see Methods). Figure 5D shows that the enzymatic activity of GAPDH is reduced with an increasing concentration of benomyl, with activity reduced approximately by 50% at 25 uM. This effect confirms the TargetSeeker-MS prediction that GAPDH is bound by benomyl and that this binding affects the in vitro functionality of GAPDH.

DISCUSSION

TargetSeeker-MS is compatible with most MS quantification techniques.

In this study, spectral counting was used to build protein fractionation profiles to illustrate that TargetSeeker-MS can confidently identify drug targets with simple and cost-effective MS quantitative approaches (label-free quantification). Nevertheless, most types of intensity-based protein quantification strategies (extracted ion chromatogram[24,25], SILAC[26], TMT[27], iTRAQ[28], etc.) can also be processed by TargetSeeker-MS. In fact, TargetSeeker-MS performance would likely benefit from quantification techniques with improved accuracy. Indeed, spectral counting may not possess the quantification accuracy to enable TargetSeeker-MS to detect slight changes in protein fractionation profiles that are consistent across all replicates. This may particularly be the case for low abundance proteins for which slight changes may be easier to detect with an increase quantification resolution provided by intensity-based quantification techniques.

1 **TargetSeeker-MS' applications for therapeutic discovery**

2 EBPS-MS approaches coupled to TargetSeeker-MS can play an important role in the drug

3 discovery process. For instance, high-throughput screenings can allow the identification of a

4 small set of compounds modulating a certain target pathway. These compounds can then be

5 tested against a complex proteome such as the one of a human cell to identify drug bindings and

6 any potential deleterious protein interactions. In the future, EBPS-MS approaches could be

7 applied using TargetSeeker-MS to identify the targets of drugs in induced pluripotent stem cells

8 (iPSCs) and potentially allow the design of better therapeutics.

9

10 **Scaling TargetSeeker-MS analysis to multiple compounds.**

11 While this study focused on analyzing a single compound in different samples using different

12 EBPS-MS approaches, TargetSeeker-MS' analysis can be scaled for the analysis of multiple

13 compounds. A given set of untreated samples (four control biological replicates) under the same

14 experimental conditions can be used to identify the drug targets of multiple compounds. Using

15 four controls, TargetSeeker-MS can confidently build its noise model and assess the targets of

16 the tested drugs with two treated biological replicates for a total of six samples, which is the

17 standard number of samples processed in MS-based proteomics experiments comparing two

18 experimental conditions (three replicates from one condition and three replicates from a second

19 condition). Additionally, the noise model built by TargetSeeker-MS can be reused to evaluate

20 supplementary compounds without the analysis of more untreated samples. Since the algorithm

21 only requires the analysis of duplicate drug treated samples, the total number of samples needed

22 to identify the targets of two compounds under the same experimental conditions would be eight,

21

1   and three compounds would require ten samples (i.e. required samples = 4 + number of

2   compounds × 2). TargetSeeker-MS therefore eases the scalability of the analysis of multiple

3   compounds in contrast to an approach that requires a paired control with each treated sample.

4

5   **Drug repositioning applications**

6   The rapidly growing field of drug repositioning depends on the use of techniques like EBPS-MS

7   for the systematic identification of the targets of a drug. The TargetSeeker-MS software package

8   will facilitate the implementation of these techniques and will have a significant impact on the

9   field. Over time the drug discovery pipeline has become increasingly long and costly. Indeed, it

10   takes on average 13.5 years for a drug to reach the market from the start of the investigation[29].

11   The drug discovery process is also highly failure-prone with a success rate of less than 10%[29].

12   Drug repositioning is an effort to test whether compounds for which the safety is known and that

13   were approved by organizations such as the FDA could have applications for treatment of

14   disease processes other than the one they were originally designed for. This accelerates the drug

15   discovery process and minimizes the expense and failure risks. Success stories such as

16   duloxetine, which was originally developed to treat depression and is now also used to improve

17   the condition of stress urinary incontinence victims[30], and crizotinib, which was created to treat

18   anaplastic large-cell lymphoma and was repurposed for the treatment of non-small-cell lung

19   cancer[31], will hopefully become more common with the increased popularity of EBPS-MS and

20   the use of TargetSeeker-MS.

21

22

1   While EBPS-MS approaches are emerging as fast efficient methods to identify drug

2   targets in complex proteome, no user-friendly approaches have been proposed to assess the

3   confidence of drug binding predictions. We have shown that TargetSeeker-MS can identify high-

4   confidence drug targets with a limited number of control samples. We also demonstrate that

5   TargetSeeker-MS recapitulates predictions of benomyl targets using different EBPS-MS

6   techniques and processing samples from different organisms. In conclusion, our algorithm will

7   favor the growth of the applications of EBPS-MS techniques and improve their impact the field

8   of drug discovery.

9   **METHODS**

10   **Method overview**

11   We propose a computational approach, TargetSeeker-MS, for the identification of drug targets in

12   energetics-based protein separation (EBPS) coupled to mass spectrometry (MS) datasets. The

13   Bayesian inference approach assesses the confidence that a protein is bound by a given

14   compound. We first describe the two EBPS procedures used in this manuscript (DiffPOP and

15   Thermal shift assay) and the mass spectrometry analysis. We then present TargetSeeker-MS

16   machine learning algorithm. Figure 1 graphically depicts our approach.

17

18   **Sample preparation**

19   Mixed stage *C. elegans* were grown and homogenized using a chilled Precellys24 homogenizer

20   (Bertin Instruments) in 80% Phosphoprotein Kit Buffer A (ClonTech, catalog number 635626).

21   Human embryonic kidney cell lysate (HEK293) was prepared from HEK293 cells grown in

22   Dulbecco's modified Eagle's medium (D-MEM) with 10% fetal bovine serum (FBS)

23

1 supplemented with penicillin and streptomycin. Cells were grown (37°C/5% CO_2) to

2 approximately 80% confluence in tissue culture flasks. Cells were washed twice with DPBS,

3 scrapped from flasks, supplemented with protease inhibitor cocktail (Roche) and lysed by

4 sonication. Protein concentration was determined by BCA assay, lysate was kept at -80°C until

5 use.

6   *C. elegans* and HEK293 lysate samples were aliquoted and volume adjusted so that they

7 contained five hundred micrograms of proteins in a final volume of 250 µL containing 40 %

8 Phosphoprotein Buffer A (ClonTech). Nineteen aliquots of *C. elegans* lysate were utilized, seven

9 of them were treated with benomyl, while twelve were left untreated (controls). In addition, six

10 aliquots of HEK293 human embryonic kidney cell lysate were utilized, four were left untreated

11 (controls), while two were treated with benomyl (see below).

12

13 **Benomyl treatment and DiffPOP protein separation**

14 Lysates of HEK293 and *C. elegans* were incubated with benomyl prior to fractionation. Benomyl

15 (5ul of 10mM stock dissolved in DMSO) was added for the drug treated samples, 5ul of DMSO

16 was added to the control samples. The DiffPOP method was carried out with sequential additions

17 of denaturing solution of 90 % methanol/1% acetic acid (3.75, 8.25, 12.5, 16.25, 20, 42.5, 65,

18 212.5 and 2000 µL). Each addition was followed by vigorous vortexing and centrifugation

19 (18000 x g for 10 min at 4ºC). The supernatant was transferred to a new Eppendorf tube,

20 denaturing solution was added, sample was vortexed and centrifuged. The process was repeated

21 to produce ten pelleted fractions. All resulting pellets were washed with 400 µL ice-cold acetone

22 and centrifuged (18000 x g for 10 min at 4ºC). Pellets were air-dried and digested with trypsin

23 (see below).

24

Thermal shift assay protein separation

Four *C. elegans* untreated samples and two benomyl treated samples were each separated into 10 fractions using a thermal shift assay procedure slightly modified from the version presented by Savitski et al.[10] Lysates of *C. elegans* were prepared and incubated with benomyl, as described above, prior to the thermal shift assay. Lysates were heated in an Eppendorf heated shaker block at 20°C for 10 min. Samples were centrifuged at 4°C for 5min. The supernatant was transferred to a new Eppendorf tube and heated to 25°C for 5min. The process was repeated for 30, 35, 40, 50, 60, 75, 90°C to produce 9 pelleted fractions and one supernatant sample. The protein in the final soluble sample was pelleted by addition of methanol/chloroform. All resulting pellets were washed with 400 µL ice-cold acetone and centrifuged (18000 x g for 10 min at 4°C). Pellets were air-dried and digested with trypsin (see below).

Liquid chromatography coupled to MS/MS analysis

Dried pellets from the DiffPOP and thermal shift assay were dissolved in 8 M urea/100 mM TEAB, pH 8.5. Proteins were reduced with 5 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Sigma-Aldrich) and alkylated with 10 mM chloroacetamide (Sigma-Aldrich). Proteins were digested overnight at 37 ^{o}C in 2 M urea/100 mM TEAB, pH 8.5, with trypsin (Promega) at a ratio of 1:100 (enzyme:protein). Digestion was stopped with formic acid (5 % final concentration). Debris was removed by centrifugation. Final volume of each digest was 100ul, and 10ul of each digested fraction was used for analysis by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS).

The digested samples were analyzed on a Q Exactive mass spectrometer (Thermo Fisher Scientific). The digests were injected directly onto a 2 cm desalting column attached to a 20cm,

1   100um ID analytical column with pulled tip. Both were packed with 5 µm ODS-AQ C18 resin,

2   (YMC). Samples were separated at a flow rate of 400nl/min on an Easy nLCII (Thermo Fisher

3   Scientific). Buffer A was 5 % acetonitrile and 0.1 % formic acid, buffer B was 80 % acetonitrile

4   and 0.1 % formic acid. The following gradient was utilized: 1-10% B over 5 min, an increase to

5   45% B over 90 min, an increase to 80% B over another 15 min and held at 80% B for 5 min of

6   washing before returning to 1% B during the final 5 min for a 120 min total run time. Column

7   was re-equilibrated with 10ul of buffer A prior to the injection of sample. Peptides were eluted

8   directly from the tip of the column and nanosprayed directly into the mass spectrometer by

9   application of 2.5kV voltage at the back of the column. The Q Exactive was operated in a data

10 dependent mode. Full MS[1] scans were collected in the Orbitrap at 70K resolution with a mass

11 range of 400 to 1800 m/z and an AGC target of $1e^6$. A top 10 acquisition method was utilized

12 with HCD fragmentation at 25NCE, resolution of 17.5k, AGC target of $1e^5$ and an underfill ratio

13 of 0.1%. Maximum fill times were set to 60ms and 120ms for MS and MS/MS scans

14 respectively. Quadrupole isolation at 2 m/z was used, singly charged and unassigned charge

15 states were excluded, and dynamic exclusion was used with exclusion duration of 15 sec.

16

17 **Peptide and protein identification and quantification**

18 Protein and peptide identification were done with the Integrated Proteomics Pipeline – IP2

19 (Integrated Proteomics Applications). Tandem mass spectra were extracted from raw files using

20 RawConverter[32] and searched with ProLuCID[33] against human SwissProt UniProt[34] (downloaded

21 on March 25, 2014) and Wormbase[35] (version WS234) protein sequence databases. The search

22 space included all fully-tryptic and half-tryptic peptide candidates. Carbamidomethylation on

23 cysteine was considered as a static modification. Data was searched with 50 ppm precursor ion

1 tolerance and 600 ppm fragment ion tolerance. Data was filtered to 10 ppm precursor ion

2 tolerance post database search. Identified proteins were filtered using DTASelect[35] and utilizing

3 a target-decoy database search strategy to control the false discovery rate to 1% at the protein

4 level. To maximize identification specificity only proteins that were identified by at least two

5 different peptides in a given fraction were considered. A single protein identification was

6 retained for the analysis when multiple proteins were inferred from the same peptides across all

7 fractions. Redundant protein identifications were discarded. Proteins were quantified in each

8 fraction by the sum of the spectral counts of their corresponding peptides identified in the

9 fraction.

10

11 **Validation of enzymatic activity alteration**

12 A GAPDH activity assay kit (BioVision Inc. Milpitas, CA, Catalog # K680-100) was used to

13 measure GAPDH activity in the presence of benomyl. Benomyl was dissolved in dimethyl

14 sulfoxide (DMSO) and diluted in DMSO to concentrations ranging from 10 mM to 250 uM. The

15 enzyme mixture consisted of 0.5 ul of GAPDH positive control (prepared as instructed by the

16 kit) and 44.5 ul Assay Buffer. The reaction mixture was 2 ul GAPDH Substrate, 2 ul GAPDH

17 Developer and 46 ul Assay Buffer. For each test concentration, 5 ul of drug in DMSO was added

18 followed by 45 ul of enzyme mixture. After a 5 minute incubation at RT, 50 ul of reaction

19 mixture was added. The optical density (OD) at 450 nm was measured at 30 minutes.

20

21 **Protein fractionation profiles**

22 Protein fractionation profiles are defined as follows. For a drug treated sample t, let a protein p

23 be associated with a fractionation profile $F_p^t = \left[F_{p_1}^t \ F_{p_2}^t \ ... \ F_{p_f}^t \right]$, where f is the number of

1   fractions the sample is divided into and $F_{p_i}^t$, where $i = 1, \ldots, f$, is the spectral count of p in

2   fraction i. F_p^u is defined similarly for untreated samples u. Fractionation profiles F_p^u and F_p^t are

3   computed for all $p \in P$, where P is the set of all proteins that were identified with more than 5

4   spectral counts across all fractions of a sample. Finally, let the fractionation profiles be

5   normalized such that final protein fractionation profiles are defined as follows: $\hat{F}_p^t =$

6   $\left[\hat{F}_{p_1}^t \ \hat{F}_{p_2}^t \ \ldots \ \hat{F}_{p_f}^t\right]$, where $\hat{F}_{p_i}^t = F_{p_i}^t / \sum_{j=1}^f F_{p_j}^t$. $\hat{F}_p^u$ is obtained similarly.

7

8   We hypothesize that upon the binding of a drug to a protein, the stability of this protein

9   will be changed. This will affect the fractionation profile of that protein. To measure the change

10 in the fractionation profiles of a protein in the treated and untreated samples we compute the

11 distance between the fractionation profiles of a protein p between the drug treated sample $\hat{F}_p^t$ and

12 the untreated samples $\hat{F}_p^u$ as the Euclidean distance $d\left(\hat{F}_p^u, \hat{F}_p^t\right) = \sqrt{\sum_{j=1}^f (\hat{F}_{p_j}^u - \hat{F}_{p_j}^t)^2}$. We then

13 compute $d'\left(\hat{F}_p^u, \hat{F}_p^t\right) = \frac{d\left(\hat{F}_p^u, \hat{F}_p^t\right)}{\sqrt{2}}$, to ensure that the measure falls between 0 and 1 inclusively. We

14 then define the similarity $S_p^{u,t}$ for p between two samples' fractionation profiles $\hat{F}_p^u$ and $\hat{F}_p^t$ as 1-

15 $d'\left(\hat{F}_p^u, \hat{F}_p^t\right)$.

16

17 **Significance assessment of fractionation profile changes**

18 Biological and technical variation may cause fractionation profiles to differ. Such a variation can

19 be captured by computing the similarity $S_p^{u_1,u_2}$ of two fractionation profiles $\hat{F}_p^{u_1}$ and $\hat{F}_p^{u_2}$ of a

20 protein obtained from two untreated samples u_1 and u_2 under the same experimental conditions.

21 A model specifying for each protein p the null distribution of the similarity between untreated

22 protein fractionation profiles could therefore be built. Nevertheless, a large number of untreated

1   samples would be required to estimate this distribution accurately. Processing such large

2   numbers of samples with an EBPS-MS would be lengthy and consume vast amounts of

3   resources. However, we showed that using a small number of untreated samples (e.g. four) and

4   assuming that the proteins with similar similarity average values between the different samples

5   will share a similar null distribution, we can build a noise model of the fractionation profiles by

6   pooling proteins within a given similarity range. Such a model can then allow us to evaluate the

7   significance of the difference of the similarity between a treated sample and the untreated ones.

8   We used an approach inspired by a previously described method used to assess the confidence of

9   protein-protein interactions[36] to build this model and to assess the significance of the similarity

10   difference between treated and untreated samples. Our novel computational approach is

11   described below.

12

13   *Step 1: Building the similarity noise model from untreated samples*

14   For all untreated sampled that were processed using a EBPS-MS approach TargetSeeker-

15   MS first computes the protein fractionation profiles $\hat{F}_p^u$ of all identified protein p. TargetSeeker-

16   MS then estimates $\Pr[S_p^{u_1,u_2}|\overline{S_p^u}]$, which represents the probability of observing a similarity

17    $S_p^{u_1,u_2}$ given the true mean similarity of fractionation profiles of p in untreated samples $\overline{S_p^u}$. Our

18   algorithm estimates $\Pr[S_p^{u_1,u_2}|\overline{S_p^u}]$ using a leave-one-out scheme on pairs of untreated samples.

19   More specifically, for all untreated sample pairs $(u_1, u_2) \in U \times U$, where U is the set of all

20   untreated samples, we compare $S_p^{u_1,u_2}$ to

21
$$\mu_p^{\neq u_1,u_2} = \frac{\sum_{\dot{u},\ddot{u} \in (U \times U) \setminus \{(u_1,u_2),(u_2,u_1)\}} S_p^{\dot{u},\ddot{u}}}{2\left(\binom{|U|}{2}-1\right)},$$

1   the average of the similarity between all fractionation profiles except those of the sample pair

2   (u_1, u_2). We therefore assume that $\mu_p^{\neq u_1, u_2}$ represents a fair approximation of $\overline{S_p^u}$. Now let M be a

3   100×100 matrix and $M(a, b)$ represent the frequency where $\left\lceil S_p^{u_1, u_2} \cdot 100 \right\rceil = a$ and

4     $\left\lceil \mu_p^{\neq u_1, u_2} \cdot 100 \right\rceil = b$. Our goal is to use the frequency matrix M to estimate $\Pr\left[S_p^{u_1, u_2} \middle| \overline{S_p^u} \right]$,

5   however due to its dimensions, an important number of entries in M have a zero value. This

6   results in an estimator for $\Pr\left[S_p^{u_1, u_2} \middle| \overline{S_p^u} \right]$, which would often yield probabilities that are equal to

7   zero. In order to solve this problem, we used a k -nearest neighbor smoothing algorithm we have

8   previously described[36]. to build the smoothed frequency matrix M'. Briefly, let $N_\delta(a, b) =$

9     $\{(a', b'): |a - a'| \le \delta, |b - b'| \le \delta\}$ be the set of cells surrounding $M(a, b)$ with a distance δ.

10   For each entry in M, δ is computed such that $W_\delta(a, b) = \sum_{(a', b') \in M_\delta(a, b)} M(a', b') \le k$ and

11     $W_{\delta+1}(a, b) = \sum_{(a', b') \in M_{\delta+1}(a, b)} M(a', b') > k$. The smoothed matrix M' is therefore computed

12   as following:

13
$$M'(a, b) = \frac{\sum_{(c,d) \in N_\delta(a,b)} w_{c,d} \cdot M(c, d)}{\sum_{(c,d) \in N_\delta(a,b)} w_{c,d}}$$

14   where

15
$$w_{c,d} = \begin{cases} 1 & \text{if } (c, d) \in N_\delta(a, b) \\ \dfrac{k - W_\delta(a, b)}{W_{\delta+1}(a, b) - W_\delta(a, b)} & \text{if } (c, d) \in N_{\delta+1}(a, b) \backslash N_\delta(a, b) \end{cases}$$

16   Upon the empirical analysis from a large range of k values, $k = 20$ was determined to yield a

17   good balance of smoothing, while conserving the original signal of the data and therefore

18   providing the best results. Therefore, $\Pr\left[S_p^{u_1, u_2} = x \middle| \overline{S_p^u} = y \right] = M'(x, y) / \sum_{x'=0}^{99} M'(x', y)$

19   provides us a good estimate of the noise of the similarity between two fractionation profiles.

20

1   *Step 2: Posterior distribution of the mean similarity between fractionation profiles from*

2   *untreated samples*

3   Using Bayes' rule and assuming the conditional independence of the similarity observations

4   given their true means, the posterior distribution of $\overline{S_p^u}$ is computed as follows:

5
$$\Pr\left[\overline{S_p^u}\Big|S_p^{u_1,u_2},\dots,S_p^{u_{|U|-1},u_{|U|}}\right] = \Pr\left[\overline{S_p^u}\right] \cdot \prod_{i=1,j>i}^{i<|U|,j\le|U|} \Pr\left[S_p^{u_i,u_j}\Big|\overline{S_p^u}\right]/\zeta$$

6   where ζ is a normalizing constant, and $\Pr\left[\overline{S_p^u}=\alpha\right]$ corresponds to the fraction of proteins with

7    $\left\lceil \overline{S_p^u}\right\rceil = \alpha$. This calculation provides us the distribution of the mean similarity of the fractionation

8   profile of a protein between untreated samples given the set of similarity observations

9    $S_p^{u_1,u_2},\dots,S_p^{u_{|U|-1},u_{|U|}}$

10

11   *Step 3: Significance assessment*

12   The goal of TargetSeeker-MS is to assess the significance of the change in the fractionation

13   profile of each protein in untreated samples and a drug treated sample. In order to do so, we first

14   compute the average of the similarity values between $\hat{F}_p^t$ and $\hat{F}_p^u$ $\forall\, u \in U$ as follows, $\overline{S_p^{u,t}} =$

15    $\frac{\Sigma_{u\in U} S_p^{u,t}}{|U|}$. We then assess the significance of the change in similarity for the drug treated

16   fractionation profile for each protein by computing a p-value using our noise model of the

17   similarity of the fractionation profiles of a protein in untreated samples (computed in *Step 2*):

18    $p\text{-value}\left(\overline{S_p^{u,t}}\right) = \Pr\left[\overline{S_p^u} \le \overline{S_p^{u,t}}\Big|S_p^{u_1,u_2},\dots,S_p^{u_{|U|-1},u_{|U|}}\right] = \sum_{i=0}^{\overline{S_p^{u,t}}} \Pr\left[\overline{S_p^u} = i\Big|S_p^{u_1,u_2},\dots,S_p^{u_{|U|-1},u_{|U|}}\right].$

19

20

21

1   *Step 4: False discovery rate estimation*

2   The p-values computed in *Step 3* may be underestimated as a result of the possible violation of

3   some of the assumptions made when building our noise model. To assess the specificity of its

4   predictions, TargetSeeker-MS uses a hypothesis-free approach to compute a False Discovery

5   Rate (FDR) for each protein. Given a p-value threshold p, TargetSeeker-MS computes $FDR(p)$,

6   which represents the fraction of false positive predicted drug binders that obtained a p-value $< p$.

7    $FDR(\text{p})$ is computed using a leave-one-out strategy, where each untreated biological replicate is

8   alternately left out from the rest of the untreated samples for the p-value calculation and

9   considered as if it was a treated sample. Specifically, $\forall\, u' \in U$, TargetSeeker-MS computes p-

10   value($\overline{S_p^{u,u'}}$) for all proteins p quantified in the untreated sample u' (left out) and p-value($\overline{S_p^{u,t}}$)

11   for all proteins p quantified in the treated sample t using the set of untreated samples excluding

12   u' to build the noise model. TargetSeeker-MS then computes for the FDR for the p-value p as

13   follows:

$$FDR(p) = \frac{\dfrac{\sum_{u' \in U} \sum_{p \in P_U} f\left(p - \text{value}\left(\overline{S_p^{u,u'}}\right)\right)}{|U| \cdot |P_U|}}{\dfrac{\sum_{u' \in U} \sum_{p \in P_t} f\left(p - \text{value}\left(\overline{S_p^{u,t}}\right)\right)}{|U| \cdot |P_t|}}$$

14

15   where

$$f(z) = \begin{cases} 1 & \text{if } z < p \\ 0 & \text{if } z \geq \text{p} \end{cases}$$

16

17

18   and P_U and P_t are the sets of proteins with fractionation profiles in the untreated and treated

19   samples respectively. TargetSeeker-MS' algorithm design, involving two independent leave-one-

32

1   out schemes, therefore requires a minimum of four untreated sample replicates to allow the

2   confidence assessment of putative drug targets.

3

4   *Step 5: Biological replicates of drug treated samples*

5   To maximize the stringency of its drug target predictions when given biological replicates of

6   drug treated samples, TargetSeeker-MS allows the user to analyze both replicates independently

7   and to report the proteins that are considered high-confidence drug targets in all replicates.

8   Alternatively, TargetSeeker-MS can also analyze all biological replicates of drug treated samples

9   simultaneously. In this context, TargetSeeker-MS computes the average of the similarity values

10   between $\hat{F}_p^t$ and $\hat{F}_p^u$ $\forall u \in U, t \in T$, such that $\overline{S_p^{u,t}} = \frac{\Sigma_{u \in U, t \in T} S_p^{u,t}}{|U| \cdot |T|}$, where T is the set of treated

11   samples. The two approaches tend to yield similar results with the former being slightly more

12   stringent (see Results).

13

14   **Computation of fold-change of similarity difference**

15   To maximize the specificity of TargetSeeker-MS' predictions, putative drug targets must be

16   associated with a FDR < 0.10, but a Fold-change of Similarity Difference (FSD) above a given

17   threshold can also be used. The FSD of a protein p is calculated as follows:

18
$$FSD(p) = \frac{\left| \overline{S_p^{u,t}} - \overline{S_p^u} \right|}{\overline{S_p^{u,t}}}$$

19   **Implementation and availability**

20   TargetSeeker-MS is implemented as a web-based fast Java program that is available at this

21   address: http://targetseeker.scripps.edu/. Example input and output files are provided online.

22   Given a set of untreated and drug treated samples that were fractionated using an EBPS

1   approach, TargetSeeker-MS computes a FDR for all proteins with a fractionation profile in both

2   conditions. Note that intensity-based protein quantification can also be provided as input to

3   TargetSeeker-MS. The methods are also implemented as stand-alone, open-source, platform

4   independent, command-line-based Java program, which is available at this address

5   http://targetseeker.scripps.edu/files/ and on GitHub: https://github.com/proteomicsyates. The

6   mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium

7   via the PRIDE partner repository with the dataset identifier PXD010799. Data are stored in Pride

8   Archive. To access the data files, please visit https://www.ebi.ac.uk/pride/archive/login and use

9   the following user name: reviewer02058@ebi.ac.uk and password: 4nyA0clB.

10

11   **Alternative approaches**

12   *Z-score filtering*

13   We developed an alternative method to benchmark the TargetSeeker-MS algorithm. The *Z*-score

14   method computes a Z-score for each protein p by comparing the similarity of its fractionation

15   profiles between the treated and untreated samples $\overline{S_p^{u,t}}$ to the mean $\overline{S_p^u}$ and the standard

16   deviation $\sigma(S_p^u)$ of the similarity of the fractionation profiles in untreated samples: $Z\text{-score}(\overline{S_p^{u,t}})$

17   $= (\overline{S_p^{u,t}} - \overline{S_p^u})/\sigma(S_p^u)$. This alternative approach may outperform the TargetSeeker-MS algorithm if

18   the variance of the fractionation profile similarity values in untreated samples of proteins with

19   close mean similarity values differs significantly from protein to protein. This is due to the

20   pooling of similarity values from different proteins in TargetSeeker-MS, which makes the

21   assumption that this variation is low. Nevertheless, since no data pooling is performed with the

22   *Z*-score method, the estimation of the variance is performed using a small number of values (i.e.

23   the total number of untreated sample pairs). Furthermore, the *Z*-score method assumes that the

34

1   noise of the fractionation profile similarity values in untreated samples is normally distributed,

2   which cannot be unequivocally verified. FDR values for all Z-scores were estimated using the

3   same leave-one out approach as described in *Step 4*.

4

5   *Savitski et al. statistical approach.*

6   We implemented the statistical approach adapted from the article from Cox et al.[11] We analyzed

7   all EBPS-MS datasets (Dataset 1, 2 and 3) with this implementation in the fashion it was applied

8   in the Savitski et al.'s article[10], with the only difference that melting curve slopes were replaced

9   with similarity values. It should be noted that this modification does not affect the validity of the

10 statistical approach nor does it change any of its assumptions about input values. Due to the

11 different nature of the algorithm, FDR values were estimated with a slightly modified procedure

12 than the one described in *Step 4*. The algorithm was fed the average similarity values of all

13 proteins $p \in P_u$ for which a fractionation profile was computed in untreated samples to which p-

14 values were associated. p-values were then calculated for all proteins $p \in P_t$ for which a

15 fractionation profile was computed based on the average similarity values of these profiles

16 between treated and untreated samples. FDRs were then associated to each p-value p as follows:

17
$$\mathrm{FDR}(p) = \frac{\dfrac{\sum_{p \in P_u} f(p - \mathrm{value}(\overline{S_p^u}))}{|P_u|}}{\dfrac{\sum_{p \in P_t} f(p - \mathrm{value}(\overline{S_p^{u,t}}))}{|P_t|}}$$

18 where

19
$$f(z) = \begin{cases} 1 & \text{if } z < p \\ 0 & \text{if } z \geq p \end{cases}$$

20

21

35

Gene Ontology enrichment analysis

To investigate the mechanism of action of benomyl, we evaluated the statistical enrichment of Gene Ontology terms[37] among the proteins predicted as its targets by TargetSeeker-MS in the three different datasets using Ontologizer[38]. We tested the enrichment of molecular functions, biological processes, and cellular components (with the complete set of proteins associated with a fractionation profile as background). Ontologizer uses a modified Fisher's exact test to assess the statistical significance of the enrichment of Gene Ontology terms and the Bonferroni correction to correct for multiple hypothesis testing[38].

Protein ortholog determination

Orthologous protein targets between *H. sapiens* and *C. elegans* were determined using the Blastp algorithm[39]. Proteins associated with a sequence identity between the two species > 50% and an E-value $< 10^{-10}$ were considered orthologs.

1

AUTHOR CONTRIBUTIONS

M.L.A. designed the computational approach, which was implemented by A.P. with help from M.L.A. and S.M.B. The DiffPOP method was developed by J.K.D., A.F.M.P. and J.J.M. The TSA approach was adapted by J.K.D. The DiffPOP, TSA, and LC-MS/MS procedures were performed by J.K.D., A.P. and W.L. The enzymatic assay was performed by J.J.M. C. elegans sample processing was performed by M.P. Finally, M.L.A., A.P., J.K.D., J.J.M. and J.R.Y. wrote the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1. Szymański, P., Markowicz, M. & Mikiciuk-Olasik, E. Adaptation of high-throughput screening in drug discovery-toxicological screening tests. *Int. J. Mol. Sci.* **13**, 427–52 (2012).

2. Bantscheff, M., Hopf, C., Savitski, M. M., Dittmann, A., Grandi, P., Michon, A.-M., Schlegl, J., Abraham, Y., Becher, I., Bergamini, G., Boesche, M., Delling, M., Dümpelfeld, B., Eberhard, D., Huthmacher, C., Mathieson, T., Poeckel, D., Reader, V., Strunk, K., Sweetman, G., Kruse, U., Neubauer, G., Ramsden, N. G. & Drewes, G. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat. Biotechnol.* **29**, 255–65 (2011).

3. Bantscheff, M., Eberhard, D., Abraham, Y., Bastuck, S., Boesche, M., Hobson, S., Mathieson, T., Perrin, J., Raida, M., Rau, C., Reader, V., Sweetman, G., Bauer, A., Bouwmeester, T., Hopf, C., Kruse, U., Neubauer, G., Ramsden, N., Rick, J., Kuster, B. & Drewes, G. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat. Biotechnol.* **25**, 1035–44 (2007).

4. Frei, A. P., Jeon, O.-Y., Kilcher, S., Moest, H., Henning, L. M., Jost, C., Plückthun, A., Mercer, J., Aebersold, R., Carreira, E. M. & Wollscheid, B. Direct identification of ligand-receptor interactions on living cells and tissues. *Nat. Biotechnol.* **30**, 997–1001 (2012).

5. West, G. M., Tucker, C. L., Xu, T., Park, S. K., Han, X., Yates, J. R. & Fitzgerald, M. C. Quantitative proteomics approach for identifying protein-drug interactions in complex mixtures using protein stability measurements. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9078–82 (2010).

6. Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D. F., Burkhard, P. R. & Sanchez, J.-C. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal. Chem.* **80**, 2921–31 (2008).

7. Vedadi, M., Niesen, F. H., Allali-Hassani, A., Fedorov, O. Y., Finerty, P. J., Wasney, G. A., Yeung, R., Arrowsmith, C., Ball, L. J., Berglund, H., Hui, R., Marsden, B. D., Nordlund, P., Sundstrom, M., Weigelt, J. & Edwards, A. M. Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15835–40 (2006).

8. Pace, C. N. & McGrath, T. Substrate stabilization of lysozyme to thermal and guanidine hydrochloride denaturation. *J. Biol. Chem.* **255**, 3862–5 (1980).

9. Martinez Molina, D., Jafari, R., Ignatushchenko, M., Seki, T., Larsson, E. A., Dan, C., Sreekumar, L., Cao, Y. & Nordlund, P. Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341**, 84–7 (2013).

10. Savitski, M. M., Reinhard, F. B. M., Franken, H., Werner, T., Savitski, M. F., Eberhard, D., Martinez Molina, D., Jafari, R., Dovega, R. B., Klaeger, S., Kuster, B., Nordlund, P., Bantscheff, M. & Drewes, G. Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **346**, 1255784 (2014).

11. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–72 (2008).

12. Xu, M., Moresco, J. J., Chang, M., Mukim, A., Smith, D., Diedrich, J. K., Yates, J. R. & Jones, K. A. SHMT2 and the BRCC36/BRISC deubiquitinase regulate HIV-1 Tat K63-ubiquitylation and destruction by autophagy. *PLoS Pathog.* **14**, e1007071 (2018).

13. Megger, D. A., Pott, L. L., Ahrens, M., Padden, J., Bracht, T., Kuhlmann, K., Eisenacher, M., Meyer, H. E. & Sitek, B. Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines. *Biochim. Biophys. Acta* **1844**, 967–76 (2014).

14. Zhou, Y., Shan, Y., Zhang, L. & Zhang, Y. Recent advances in stable isotope labeling based techniques for proteome relative quantification. *J. Chromatogr. A* **1365**, 1–11 (2014).

15. Rauniyar, N. & Yates, J. R. Isobaric labeling-based relative quantification in shotgun proteomics. *J. Proteome Res.* **13**, 5293–309 (2014).

16. Fitzmaurice, A. G., Rhodes, S. L., Lulla, A., Murphy, N. P., Lam, H. A., O'Donnell, K. C., Barnhill, L., Casida, J. E., Cockburn, M., Sagasti, A., Stahl, M. C., Maidment, N. T., Ritz, B. & Bronstein, J. M. Aldehyde dehydrogenase inhibition as a pathogenic mechanism in Parkinson disease. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 636–41 (2013).

17. Fitzmaurice, A. G., Rhodes, S. L., Cockburn, M., Ritz, B. & Bronstein, J. M. Aldehyde dehydrogenase variation enhances effect of pesticides associated with Parkinson disease. *Neurology* **82**, 419–26 (2014).

18. Staub, R. E., Quistad, G. B. & Casida, J. E. Mechanism for benomyl action as a mitochondrial aldehyde dehydrogenase inhibitor in mice. *Chem. Res. Toxicol.* **11**, 535–43 (1998).

19. . S. B. & . T. A. Study on the Liver Functions in Rats Exposed to Benomyl. *J. Biol. Sci.* **5**, 666–669 (2005).

20. Kara, M. & Çelik, İ. The effects of some fungicides on activity of eight serum enzymes in vitro. *J. Environ. Sci. Heal. . Part A Environ. Sci. Eng. Toxicol.* **32**, 1377–1382 (1997).

21. Stearns, T., Hoyt, M. A. & Botstein, D. Yeast mutants sensitive to antimicrotubule drugs define three genes that affect microtubule function. *Genetics* **124**, 251–62 (1990).

22. Davidse, L. C. & Flach, W. Differential binding of methyl benzimidazol-2-yl carbamate to fungal tubulin as a mechanism of resistance to this antimitotic agent in mutant strains of Aspergillus nidulans. *J. Cell Biol.* **72**, 174–93 (1977).

23. Mahowald, J., Arcizet, D. & Heinrich, D. Impact of external stimuli and cell micro-architecture on intracellular transport states. *Chemphyschem* **10**, 1559–66 (2009).

24. Higgs Jr., R. E., Julian Jr., R. K. & Kaiser Jr., R. E. System and methods for qualitatively and quantitatively comparing complex admixtures using single ion chromatograms derived from spectroscopic analysis of such admixtures. (1999).

25. Higgs, R. E., Knierman, M. D., Gelfanova, V., Butler, J. P. & Hale, J. E. Comprehensive label-free method for the relative quantification of proteins from biological samples. *J. Proteome Res.* **4**, 1442–50

26. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. & Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–86 (2002).

27. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K. A. & Hamon, C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–904 (2003).

28. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. & Pappin, D. J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–69 (2004).

29. Li, Y. Y. & Jones, S. J. Drug repositioning for personalized medicine. *Genome Med.* **4**, 27 (2012).

30. Ashburn, T. T. & Thor, K. B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **3**, 673–83 (2004).

31. Shaw, A. T., Yasothan, U. & Kirkpatrick, P. Crizotinib. *Nat. Rev. Drug Discov.* **10**, 897–8 (2011).

32.   He, L., Diedrich, J., Chu, Y.-Y. & Yates, J. R. Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. *Anal. Chem.* **87**, 11361–7 (2015).

33.   Xu, T., Park, S. K., Venable, J. D., Wohlschlegel, J. A., Diedrich, J. K., Cociorva, D., Lu, B., Liao, L., Hewel, J., Han, X., Wong, C. C. L., Fonslow, B., Delahunty, C., Gao, Y., Shah, H. & Yates, J. R. ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *J. Proteomics* **129**, 16–24 (2015).

34.   The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).

35.   Harris, T. W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W. J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R., Fernandes, J., Han, M., Kishore, R., Lee, R., Müller, H.-M., Nakamura, C., Ozersky, P., Petcherski, A., Rangarajan, A., Rogers, A., Schindelman, G., Schwarz, E. M., Tuli, M. A., Van Auken, K., Wang, D., Wang, X., Williams, G., Yook, K., Durbin, R., Stein, L. D., Spieth, J. & Sternberg, P. W. WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res.* **38**, D463-7 (2010).

36.   Lavallée-Adam, M., Cloutier, P., Coulombe, B. & Blanchette, M. Modeling contaminants in AP-MS/MS experiments. *J. Proteome Res.* **10**, 886–95 (2011).

37.   Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–9 (2000).

38.   Bauer, S., Grossmann, S., Vingron, M. & Robinson, P. N. Ontologizer 2.0--a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* **24**, 1650–1 (2008).

39.   Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–402 (1997).

SUPPLEMENTARY FIGURES

Supplementary Figure S1. Graphical depiction of the datasets presented in this study.

Supplementary Figure S2. Heatmap representation of the similarity between protein

fractionation profiles for all pairs of untreated samples of the TSA/*C. elegans* dataset. All

1   proteins with a sufficient abundance to compute a fractionation profile in all untreated samples

2   are displayed.

3

4   **Supplementary Figure S3.** Heatmap representation of the similarity between protein

5   fractionation profiles for all pairs of untreated samples of the DiffPOP/HEK293 dataset. All

6   proteins with a sufficient abundance to compute a fractionation profile in all untreated samples

7   are displayed.

8

9   **Supplementary Figure S4.** Three-dimensional plot representation of the probability matrix of

10    $\mathbf{Pr}\left[S_p^{u_1,u_2} \middle| \overline{S_p^u}\right]$ based on the smoothed frequency matrix M' for four untreated samples from the

11   DiffPOP/*C. elegans* dataset.

12

13   **Supplementary Figure S5.** Heatmap representation of the similarity between protein

14   fractionation profiles for all pairs of benomyl treated samples of the DiffPOP/*C. elegans* dataset.

15   All proteins with a sufficient abundance to compute a fractionation profile in all benomyl treated

16   samples are displayed.

17

18   **Supplementary Figure S6.** Heatmap representation of the similarity between protein

19   fractionation profiles for the two benomyl treated samples of the TSA/*C. elegans* dataset. All

20   proteins with a sufficient abundance to compute a fractionation profile in both benomyl treated

21   samples are displayed.

22

1   **Supplementary Figure S7.** Heatmap representation of the similarity between protein

2   fractionation profiles for the two benomyl treated samples of the DiffPOP/HEK293 dataset. All

3   proteins with a sufficient abundance to compute a fractionation profile in both benomyl treated

4   samples are displayed.

5

6   **Supplementary Figure S8.** Distributions of the fractionation profile similarity values in both

7   untreated and benomyl treated samples of the TSA/*C. elegans* dataset.

8

9   **Supplementary Figure S9.** Distributions of the fractionation profile similarity values in both

10 untreated and benomyl treated samples of the DiffPOP/HEK293 dataset.

11

12 **SUPPLEMENTARY TABLES**

13

14 **Supplementary Table S1:** *C. elegans* proteins identified through DiffPOP-MS analysis along

15 with their associated statistics as calculated by TargetSeeker-MS when combining all treated

16 samples.

17

18 **Supplementary Table S2:** Benomyl treated sample 1 *C. elegans* proteins identified through

19 DiffPOP-MS analysis along with their associated statistics as calculated by TargetSeeker-MS.

20

21 **Supplementary Table S3:** Benomyl treated sample 2 *C. elegans* proteins identified through

22 DiffPOP-MS analysis along with their associated statistics as calculated by TargetSeeker-MS.

23

1   **Supplementary Table S4:** Benomyl treated sample 3 *C. elegans* proteins identified through

2   DiffPOP-MS analysis along with their associated statistics as calculated by TargetSeeker-MS.

3

4   **Supplementary Table S5:** *C. elegans* proteins identified as benomyl targets in each treated

5   sample processed with DiffPOP-MS.

6

7   **Supplementary Table S6:** Gene Ontology enrichment analysis complete results for the benomyl

8   targets identified in the DiffPOP-MS-*C. elegans* dataset.

9

10   **Supplementary Table S7:** *C. elegans* proteins identified through Thermal Shift Assay-MS

11   analysis along with their associated statistics as calculated by TargetSeeker-MS when combining

12   all treated samples.

13

14   **Supplementary Table S8:** Benomyl-treated sample 1 *C. elegans* proteins identified through

15   Thermal Shift Assay-MS analysis along with their associated statistics as calculated by

16   TargetSeeker-MS.

17

18   **Supplementary Table S9:** Benomyl-treated sample 2 *C. elegans* proteins identified through

19   Thermal Shift Assay-MS analysis along with their associated statistics as calculated by

20   TargetSeeker-MS.

21

22   **Supplementary Table S10:** *C. elegans* proteins identified as benomyl targets in both treated

23   samples processed with TSA-MS.

1

2 **Supplementary Table S11:** Gene Ontology enrichment analysis complete results for the

3 benomyl targets identified in the TSA/*C. elegans* dataset.

4

5 **Supplementary Table S12:** HEK 293 proteins identified through DiffPOP-MS analysis along

6 with their associated statistics as calculated by TargetSeeker-MS when combining both treated

7 samples.

8

9 **Supplementary Table S13:** Benomyl-treated sample 1 HEK 293 proteins identified through

10 DiffPOP-MS analysis along with their associated statistics as calculated by TargetSeeker-MS.

11

12 **Supplementary Table S14:** Benomyl-treated sample 2 HEK 293 proteins identified through

13 DiffPOP-MS analysis along with their associated statistics as calculated by TargetSeeker-MS.

14

15 **Supplementary Table S15:** HEK 293 proteins identified as benomyl targets in both treated

16 samples processed with DiffPOP-MS.

17

18 **Supplementary Table S16:** Gene Ontology enrichment analysis complete results for the

19 benomyl targets identified in the DiffPOP/HEK 293 dataset.

20

21 **SOURCE DATA**

22

23 **Figure 2A - Source Data:** Csv file containing the data presented in the heatmap of Figure 2A.

24

1

2   **Figure 2B - Source Data:** Excel file containing the data of the distributions presented in Figure

3   2B.

4

5   **Figure 2C-D - Source Data:** Excel file containing the data of the distributions and protein

6   fractionation profiles presented in Figures 2C and 2D.

7

8   **Figure 3A - Source Data:** Excel file containing the data of the benchmarking presented in

9   Figure 3A.

10

11   **Figure 4A - Source Data:** Excel file containing the data of the benchmarking presented in

12   Figure 4A.

13

14   **Figure 5A - Source Data:** Excel file containing the data of the benchmarking presented in

15   Figure 5A.

16

17   **Figure 5D - Source Data:** Excel file containing the data of the results of the GAPDH activity

18   assay presented in Figure 5D.

19