Non-parametric analysis of thermal proteome profiles reveals novel drug-binding proteins

Dorothee Childs[1,4†], Karsten Bach[2,3†], Holger Franken[4†], Simon Anders[5], Nils Kurzawa[1], Marcus Bantscheff[4], Mikhail Savitski[1] and Wolfgang Huber[1*]

[*]Correspondence:
wolfgang.huber@embl.de
[1]European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany
Full list of author information is available at the end of the article
[†]Equal contributor

Abbreviations

- CETSA: Cellular thermal shift assay
- FDR: False discovery rate
- H_0: Null hypothesis
- H_1: Alternative hypothesis
- NPARC: Non-parametric analysis of response curves
- ROC: Receiver operating characteristic
- RSS: Residual sum of squares
- T_m: Melting point
- TPP: Thermal proteome profiling

Abstract

Detecting the targets of drugs and other molecules in intact cellular contexts is a major objective in drug discovery and in biology more broadly. Thermal proteome profiling (TPP) pursues this aim at proteome-wide scale by inferring target engagement from its effects on temperature-dependent protein denaturation. However, a key challenge of TPP is the statistical analysis of the measured melting curves with controlled false discovery rates at optimal proteome coverage and detection power. We present non-parametric analysis of response curves (NPARC), a statistical method for TPP based on functional data analysis and nonlinear regression. We evaluate NPARC on five independent TPP datasets and observe that it is able to detect subtle changes in any region of the melting curves, reliably detects the known targets, and outperforms a melting point-centric, single-parameter fitting approach in terms of specificity and sensitivity. NPARC can be combined with established analysis of variance (ANOVA) statistics and enables flexible, factorial experimental designs and replication levels. To facilitate access to a wide range of users, a freely available software implementation of NPARC is provided.

Keywords: Drug Discovery; Proteomics; Functional Data Analysis

Introduction

Determining the cellular interaction partners of drugs and other small molecules remains a key challenge [1, 2, 3, 4]. In drug research, better assays to detect targets (and off-targets) would provide valuable information on drugs' mechanisms of action, reveal potential reasons for side effects, and elucidate avenues for drug repurposing. More broadly, in cell biology basic research, the dynamical landscape of binding partners of metabolites, messengers or chemical probes contains much

uncharted territory. Thermal proteome profiling (TPP) addresses these needs by screening for protein targets of drugs or small molecules in living cells on a proteome-wide scale [5, 6]. TPP combines multiplexed quantitative mass spectrometry with the cellular thermal shift assay (CETSA) [7], which identifies binding events from shifts in protein thermostability (see Figure S1 for a detailed explanation). A typical TPP experiment generates temperature dependent abundance measurements for a large part of the cellular proteome. Drug binding proteins can then be inferred by comparing the melting curves of proteins between samples treated with drug and vehicle (negative control without drug).

Applications of TPP successfully identified previously unknown protein-ligand interactions [5], protein complexes [8] and downstream effects of drugs in signaling networks [6, 9, 10, 11] in human cells. Recently it has also been extended to study drug resistance in bacteria [12] and targets of antimalarial drugs in plasmodium [13]. There is urgent interest in further advancing its component technologies, including experimental and computational aspects, in order to maximize its biological discovery potential [14, 15, 16, 17, 18, 19].

The central computational task in TPP data analysis is the comparison of the temperature dependent abundance measurements—which can be visualized as melting curves—for each protein with and without (or with various concentrations of) drug. The aim is to detect changes in thermostability from statistically significant changes in the melting curves.

A naïve approach is to summarize each curve into a single parameter, such as the melting point (T_m), which is defined as the temperature of half-maximum relative abundance (horizontal line in Figure 1A). Its value is estimated by fitting a parametric model separately for the control and treatment conditions and comparing the estimates. Statistical significance is assessed using replicates and hypothesis testing, such as a t- or z-test. While the approach has delivered valid and important results [5, 6, 20, 21], we will see in the following that it tends to lead to needlessly high rates of false negatives. There are three main reasons for that: first, drug-induced effects on thermostability do not always imply significant shifts in T_m (Figure 1B-C). Second, the true T_m of a protein can lie outside of the measured temperature range, which impairs its estimation (Figure 1D). Both scenarios can result in important targets being missed in the analysis (Figure 1E). The third reason is a more subtle statistical one: hypothesis tests using only the point estimates of T_m do not take into account goodness-of-fit of the parametric model or the confidence range of the estimates. Thus, important information is ignored, which statistically leads to loss of power.

Here, we propose an alternative approach that compares whole curves instead of summary parameters and does not rely on T_m estimation. The method, non-parametric analysis of response curves (NPARC), is based on a branch of statistical data analysis that works on continuous functions rather than individual numbers, termed *functional data analysis* [22]. It considers the measured melting curves as samples from an underlying stochastic process with a smooth mean function—which can be modelled parametrically or non-parametrically [23]—and constructs its hypothesis tests directly on these samples. Thus, all available information is used. Moreover, the method is robust against deviations of the data from the function

model and thus is applicable to proteins in which T_{m}-based implementations fail. This increases proteome coverage, which can make the difference between missing or detecting an important drug target.

We demonstrate NPARC on the five published datasets introduced in Table 1. We also compare its results to those of the T_{m}-based method used by [6]. Three of the experiments used the cancer drugs panobinostat or dasatinib in different concentrations, one investigated the effects of the high-affinity, ATP-competitive pan-kinase inhibitor staurosporine, one the cellular metabolite ATP. While the cancer drugs interact with limited sets of proteins, the two other compounds are promiscuous binders and affect the thermostability of a large fraction of the cellular proteome.

Experimental Procedures

Data sets and preprocessing

Five TPP data sets (Table 1) were obtained from the supplements of the respective publications. Each data set contained relative abundance measurements per protein and temperature which had been scaled to the value measured at 37°C (the lowest of the ten temperatures assayed) and subjected to the global normalization procedure described by Savitski et al. [5]. Only proteins quantified with at least one unique peptide in each of two replicates of the vehicle and compound treated conditions were included in the analysis; the resulting proteome coverages are listed in Table 1.

Curation of lists of expected targets

Lists of expected protein targets for the pan-kinase inhibitor staurosporine and ATP were obtained from Gene Ontology Consortium annotations via the Bioconductor annotation packages *AnnotationDbi* (version 1.36.2), *org.Hs.eg.db* (version 3.4.0) and *GO.db* (version 3.4.0). Terms and numbers of annotated proteins are shown in Table 2.

Mathematical model

NPARC is based on fitting two competing models to the data, a null model and an alternative model. The null model states that the relative protein abundance at temperature t (given in °C) is explained by a single smooth function $\mu_0(t)$ irrespective of the treatment condition (Figure 2A). The alternative model posits two condition-specific functions: $\mu_T(t)$ for the treatment condition and $\mu_V(t)$ for the vehicle condition (Figure 2B). Deviations between observed data and fitted model are referred to as residuals, and the sum of squared residuals (RSS) serves as an indicator of each model's goodness-of-fit. We then compute

$$RSS_0 = \sum_{t,i,c} \left(x_{t,i,c} - \mu_0(t)\right)^2 , \tag{1}$$

$$RSS_1 = \sum_{t,i,c} \left(x_{t,i,c} - \mu_c(t)\right)^2 , \tag{2}$$

where $x_{t,i,c}$ is the measured value at temperature t for experimental replicate i and condition $c \in \{V, T\}$, and the summations extend over all temperatures, replicates, and conditions.

Choice of the mean function

The mean functions $\mu_0(t)$, $\mu_T(t)$ and $\mu_V(t)$ are each chosen from the same space of smooth functions $f : \mathbb{R}_+ \to [0,1]$ spanned by the three parameters $a, b \in \mathbb{R}$, $f_\infty \in [0,1]$ and the prescription

$$f(t) = \frac{1 - f_\infty}{1 + e^{-(\frac{a}{t} - b)}} + f_\infty. \tag{3}$$

The shape of these functions is sigmoid, and the functional form (3) can be motivated by simplifying protein thermodynamics considerations [5]. The mean functions and the RSS values are computed separately from the data for each protein. In order not to overburden the notation, we omit the protein indices.

Hypothesis test statistic

To discriminate between null and alternative models, we compute the F-statistic

$$F = \frac{d_2}{d_1} \frac{RSS_0 - RSS_1}{RSS_1}, \tag{4}$$

with $d_2/_1 > 0$ defined as below. F quantifies the relative reduction in residuals from null to alternative model. While F is by definition always positive, it will be small for proteins not affected by the treatment, while a high value of F indicates a reproducible change in thermostability.

Null distribution

To compute a p-value from a value of the F-statistic (4), we need its null distribution, i.e., its statistical distribution if the data generating process is described by a common mean function $\mu_0(t)$. If the residuals were independent and identically normal distributed, this distribution would be given by an analytical formula, namely that of the $F(d_1, d_2)$-distribution with parameters $d_1, d_2 > 0$, and these parameters—sometimes called *degrees of freedom*—would be explicitly given from the number of measurements and number of model parameters that go into the computation of RSS_0 and RSS_1. In practice, this is not the case, since the residuals are heteroscedastic (i.e. have different variances at different temperatures) and correlated. However, the family of $F(d_1, d_2)$-distributions is quite flexible, and we can approximate the distribution of the F-statistic (4) on data occurring in practice with an $F(d_1, d_2)$-distribution with different "effective degrees of freedom" d_1, d_2. To this end, we separately approximate the numerator and denominator of F as

$$\frac{RSS_0 - RSS_1}{d_1} \overset{H_0}{\sim} \frac{\sigma_0^2}{d_1} \cdot \chi^2(d_1), \tag{5}$$

$$\frac{RSS_1}{d_2} \overset{H_0}{\sim} \frac{\sigma_0^2}{d_2} \cdot \chi^2(d_2), \tag{6}$$

and use the fact that the ratio of two χ^2-distributed random variables $\chi^2(d_1)$, $\chi^2(d_2)$ has an $F(d_1, d_2)$-distribution [24]. The scale parameter σ_0^2 and the effective degrees of freedom d_1 and d_2 are estimated from the the empirical distributions—across

proteins—of RSS_0 and RSS_1. Thus, we assume that σ_0^2, d_1, d_2 are the same for all proteins. In particular, we estimate σ_0^2 from the moments of $RSS_1 - RSS_0$ as

$$\sigma_0^2 = \frac{1}{2}\frac{\mathrm{var}(RSS_0 - RSS_1)}{\mathrm{mean}(RSS_0 - RSS_1)}, \tag{7}$$

where mean and variance are computed across proteins on the observed values of $RSS_0 - RSS_1$ (see Supplementary Methods for details). Then, d_1 and d_2 are obtained by numerical optimization of the likelihoods for models (5) and (6) using the `fitdistr` function of the `R` package `MASS` [25].

p-values

For each protein, a p-value is computed from its F-statistic and the cumulative F-distribution with parameters d_1, d_2 as described above. The multiset of p-values across all proteins is corrected for multiple testing with the method of Benjamini and Hochberg [26]. The outcome of such an analysis is exemplarily shown in Figure 2C.

Results

Application to panobinostat

We assessed the ability of NPARC to detect drug targets on a dataset on panobinostat (Table 1). Panobinostat is a broad-spectrum histone deacetylase (HDAC) inhibitor known to interact with HDAC1, HDAC2, HDAC6, HDAC8, HDAC10, and tetratricopeptide repeat protein 38 (TTC38) [6].

Out of 3649 proteins reproducibly quantified across both biological replicates in both treatment conditions, NPARC yielded 16 proteins with Benjamini-Hochberg adjusted p-values ≤ 0.01. They contained the expected HDAC targets (Figure 3A-E) as well as TTC38, the histone proteins H2AFV or H2AFZ (the two variants could not be distinguished by mass spectrometry), and zinc finger FYVE domain-containing protein 28 (ZFYVE28) (Figure 3F-H). These proteins were previously identified as direct or indirect targets of panobinostat [6, 11]. In addition, eight more proteins were detected for which no direct or indirect interactions with panobinostat have been described (Figure S3). They reached statistical significance because they either showed effect sizes comparable to known panobinostat targets, or more subtle but highly reproducible changes in a similar strength to those already described for Dasatinib target BTK (Figure 1B).

Beyond two-group comparisons

Since NPARC is based on analysis of variance (ANOVA), it admits experimental designs in which the covariate has multiple levels. An examples is the data set for the BCR-ABL inhibitor dasatinib, which comprises measurements on cells treated at two different concentrations as well as untreated cells. NPARC successfully identified known targets of dasatinib (Figure S4).

Replicate agreement and model fit diagnostics

Application of the T_m-based approach by [6] to the panobinostat data failed to detect HDAC1 and HDAC2. This was because the data for these proteins had relatively high variance in the drug-treated condition, as is visible in Figure 3A-B.

This led to their exclusion according to one of the data quality filter criteria of that method (Table 3), namely the criterion that asks for sufficiently high coefficients of determination (R^2). In contrast, a better and statistically sound trade-off between variability and effect size is an integral part of NPARC, and does not require an *ad hoc* filter criterion.

To further assess the price of the various filter criteria of the T_{m}-based approach by [6], we tabulated the numbers of proteins affected by them in each of the five data sets. These proteins would, in principle, not be detectable by that method, no matter how strong the effect. Their numbers amounted to 14-27% of the total numbers of proteins (Figure 1E and Table 4, and in more detail, Figure S2). In contrast, the F-test of NPARC could be applied to all proteins irrespective of these or similar criteria, a fact which contributed to the increased protein coverage and sensitivity of NPARC.

Effects beyond those on the melting point

Many of the proteins detected by NPARC displayed reproducible changes in curve shape, while their T_{m}-shifts were small, and not considered significant by the T_{m}-based approach (Figure 4). An example is the effect of staurosporine on protein kinase C beta (PRKCB), shown in Figure 1C. PRKCB is part of the PKC family, whose members were the first reported staurosporine targets [27, 4] and also exhibit similar characteristics (Figure S5).

Further examples include the effects of staurosporine on RanGTP binding tRNA export receptor exportin-T (XPOT) and two members of the p38 MAPK signaling pathway: Mitogen-activated protein kinase 14 (MAPK14) and MAP kinase-activated protein kinase 2 (MAPKAPK2) (Figure 4); and the effect of dasatinib on Bruton tyrosine kinase (BTK), an important drug target in B-cell leukemia (Figure 1B).

Missing melting point estimates

For highly thermostable proteins, the T_{m} in one or more of the treatment conditions can be outside of the tested temperature range of a TPP experiment (Figure 1E). One example is NAD(P)H quinone dehydrogenase 2 (NQO2), a cytosolic flavoprotein and a common off-target of kinase inhibitors [28, 29, 30]. In concordance with previous CETSA studies that found NQO2 to be highly stable [31], we observed denaturation only beginning at 67 °C (Figure 1D). Staurosporine treatment further stabilized NQO2 to an extent that it showed no sign of melting in the tested temperature range. The T_{m}-based approach by [6] will discard such proteins, in order to avoid potential problems from extrapolation of the fit beyond the measured temperature range. In contrast, the functional data analysis approach of NPARC is able to detect changes in any part of the melting curves, without reference to a single point such as T_{m}.

Sensitivity and specificity

So far, we have described increased sensitivity of NPARC, i.e., its ability to detect more true targets. However, this is only useful if at the same time specificity is maintained, i.e., if false positive detection remains under control. To compare

these performance characteristics between NPARC and the T_{m}-based approach, we computed pseudo receiver operator characteristic (ROC) curves for each of these methods on the staurosporine data and the ATP data, using as pseudo ground truth lists of expected targets from Gene Ontology annotation (Table 2). Here, the term *pseudo* refers to the fact that these target lists, and hence the ROC curves, are only approximations of the truth; however, the relative ranking of two methods in such a pseudo-ROC comparison is likely to be faithful even in the presence of such approximation error [32].

Figure 5 shows the results of NPARC on both datasets, as well as those of the z-test of the T_{m}-based approach by [6] applied to the individual replicates (displayed as a continuous lines parameterized by the z-cutoff), and those of the full procedure of [6] (shown by isolated points, due to its single, fixed cutoff). On the staurosporine data, the full procedure of [6] performs close to NPARC. For the individual z-tests, as well as overall on the ATP data, NPARC shows superior performance. Given that the decision rule set of [6] (listed in Table 5) and its cutoff parameters were developed and tuned partly on the staurosporine data, these results indicate that NPARC has fewer "fudge parameters" and is likely to be superior in applications to new data sets.

Discussion

Thermal proteome profiling offers the possibility to comprehensively characterize ligand-protein interactions on a proteome-wide scale in living cells. However, the method poses the analytical challenge of how to identify statistically significant shifts in thermostability amongst thousands of measurements.

To address this challenge, we introduced a functional data analysis approach to test for treatment effects by comparing competing models by their goodness-of-fit. This enables detection of treatment effects even if a (de-)stabilization of a protein is not captured by a single summary parameter like the T_{m}. The presented method is based on a sound statistical foundation and does not rely on hard-to-choose cutoff or tuning parameters. We showed that our method compares favourably to previous approaches with respect to sensitivity and specificity for several exemplary data sets, including ones with specific and ones with promiscuous binders.

The approach fits into the framework of analysis of variance (ANOVA) or linear models, and can thus be extended to experimental designs more complex than treatment-control comparisons, such as multiple levels (e.g. drug concentrations) per covariate, multiple covariates and interactions.

The suggested framework is flexible with regard to the mean function used to represent the melting behavior and can be adapted to the particular biological process of interest. To represent nonlinear relationships, approaches include locally linear regression [33], spline regression [34, 35] and nonlinear parametric regression. Here, we chose the latter as it incorporates *a priori* knowledge about the data and thus has favourable estimation efficiency. For example, sigmoid curves have horizontal asymptotes at both sides of the temperature range. In contrast, splines and local regression tend to overfit data near the boundaries of the observation range.

In a cellular environment we occasionally observe non-sigmoid melting curves for subsets of proteins. One possible reason is the presence of protein subpopulations

each with distinct melting curves [16]. For example, the formation of protein complexes, the binding to other molecules, or the localization in cellular compartments can lead to deviations from the idealized sigmoid melting curve expected from the same protein in purified form. Our model currently does not account for such systematic and reproducible shape deviations. This could be adapted in future work by adding a low-parametric systematic modification to the sigmoid mean function.

We have considered CETSA experimental designs, where the temperature is the major experimental variable and drug concentration is either zero or a chosen value. It appears relatively straightforward to extend NPARC to the isothermal dose response (ITDR) design [6, 7] where temperature is held constant and the drug concentration is varied across a range of values. A further extension of interest would be to 2D-TPP [11] where both factors are changed.

We employ the same "average" null distribution for all proteins, which we obtain by estimating its parameters (d_1, d_2, σ_0) from the distributions of residuals across all proteins. It is conceivable that determining null distributions in a protein-dependent manner, for instance by stratification, could increase the overall power of the method.

The here presented approach is likely to increase the accuracy of profiling protein-ligand interactions in living cells. To facilitate access to a wide range of users we have implemented NPARC in a freely available R workflow [36].

Table 1 Datasets and sample sizes.

Dataset	Treatment	Concentration	Buffer	Cell line	Intact cells or lysate	Proteins	Reference
ATP data	MgATP	2 µM	PBS	K562	Lysate	4177	[9]
Dasatinib 0.5 µM data	Dasatinib	0.5 µM	PBS	K562	Intact Cells	4625	[5]
Dasatinib 5 µM data	Dasatinib	5 µM	PBS	K562	Intact Cells	4154	[5]
Panobinostat data	Panobinostat	1 µM	PBS	K562	Intact Cells	3649	[6]
Staurosporine data	Staurosporine	20 µM	PBS	K562	Lysate	4505	[5]

Table 2 Expected targets per dataset.

Dataset	Gene Ontology term	Number of proteins in data set with this term
ATP	ATP-binding	558
Staurosporine	protein kinase activity	187

Table 3 A priori filters applied in the original TPP analysis workflow [6] to select proteins for hypothesis testing.

Rule number	Rule
1	Both fitted curves for the vehicle and compound treated condition have a coefficient of determination $R^2 > 0.8$, where $R^2 := 1 - \frac{\sum_t (x_t - \mu_c(t))^2}{\sum_t (x_t - \bar{x})^2}$, with $\mu_c(t)$ being the model prediction for condition c at temperature t, $\bar{x}$ being the mean of all measurements for the protein within a particular condition and replicate, and the summation extending over all temperatures.
2	The two curves fitted to the two replicates of the vehicle conditions have a plateau $f_\infty < 0.3$.
3	In each biological replicate, the steepest slope of the melting curve in the vehicle and treatment condition needs to be $< -0.06\ ^\circ\mathrm{C}^{-1}$.

Table 4 Coverage of proteins applicable for hypothesis testing by the original TPP analysis workflow [6].

Dataset	T_m outside measured range	Curves not passing *a priori* QC-filters	T_m available
ATP data	220	1004	2953
Dasatinib 0.5 µM data	689	768	3168
Dasatinib 5 µM data	667	507	2980
Panobinostat data	320	461	2868
Staurosporine data	621	631	3253

Table 5 Decision ruleset applied in the original TPP analysis workflow [6] to combine z-test p-values across replicates in an experimental design with two biological replicates.

Rule number	Rule
1	The Benjamini-Hochberg adjusted z-test p-values fulfill predefined thresholds in each replicate.
2	Both T_m differences are either positive or negative in the two biological replicates.
3	The smallest absolute difference between treatment and vehicle T_m is greater than the absolute T_m difference between the two vehicle experiments.

Additional Files

The following Figures and Tables can be found in the Supplementary Material.

Additional file 1 — Supplementary Methods

Detailed description of the fitting procedures for the scaling parameter, melting points, and mean functions of the model.

Additional file 2 — Table S1

Spreadsheet containing the results of the NPARC approach and of the T_m-based approach for all datasets listed in Table 1.

Additional file 3 — Supplementary Figures S1-S5

PDF containing Supplementary Figures S1-S5.

Additional file 4 — Supplementary Figure S6
All proteins detected by the NPARC approach with Benjamini-Hochberg adjusted F-test p-values ≤ 0.01 in the staurosporine data.

Additional file 5 — Supplementary Figure S7
All proteins detected by the NPARC approach with Benjamini-Hochberg adjusted F-test p-values ≤ 0.01 in the ATP data.

Acknowledgments
We would like to thank Sindhuja Sridharan (EMBL Heidelberg) for help with Figure preparation. We thank EMBL and GlaxoSmithKline for supporting the work. KB is supported by a Cambridge Cancer Centre studentship. SA is funded by the Deutsche Forschungsgemeinschaft, SFB 1036. WH acknowledges funding from the European Commission's H2020 Programme, Collaborative research project SOUND (Grant Agreement no 633974).

Software and data availability
The TPP-TR experiments based on staurosporine-, panobinostat-, ATP-, and dasatinib treatments are included in the supplementary materials of references [5, 9, 6]. All results generated from this data are provided in the supplementary material attached to this work. A vignette describing the workflow and relevant code to reproduce the results is available for download [36]. An open-source R/Bioconductor package will soon be released.

Competing interests
HF, MS, DC and MB are employees and/or shareholders of GlaxoSmithKline.

Authors' contributions
KB, DC and HF contributed equally. KB, DC, HF, SA, MS and WH developed the model. KB and DC implemented the software package and the analytical computations. KB, DC, HF, NK, MS, MB and WH interpreted the results. KB, DC, HF, NK, SA and WH wrote the manuscript. All authors read and approved the final manuscript.

Author details
[1]European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany. [2]Department of Pharmacology, University of Cambridge, CB2 1PD, Cambridge, UK. [3]Cancer Research UK Cambridge Cancer Centre, CB2 0RE, Cambridge, UK. [4]Cellzome GmbH, GlaxoSmithKline, Meyerhofstraße 1, 69117 Heidelberg, Germany. [5]Center for Molecular Biology of Heidelberg University (ZMBH), Im Neuenheimer Feld 282, 69120 Heidelberg, Germany.

References
1. Comess, K.M., McLoughlin, S.M., Oyer, J.A., Richardson, P.L., Stöckmann, H., Vasudevan, A., Warder, S.E.: Emerging approaches for the identification of protein targets of small molecules - a practitioners' perspective. Journal of Medicinal Chemistry **61**(19), 8504–35 (2018)
2. Simon, G.M., Niphakis, M.J., Cravatt, B.F.: Determining target engagement in living systems. Nature Chemical Biology **9**, 200–5 (2013)
3. Bunnage, M.E., Chekler, E.L.P., Jones, L.H.: Target validation using chemical probes. Nature Chemical Biology **9**, 195–9 (2013)
4. Arrowsmith, C.H., Audia, J.E., Austin, C., Baell, J., Bennett, J., Blagg, J., Bountra, C., Brennan, P.E., Brown, P.J., Bunnage, M.E., Buser-Doepner, C., Campbell, R.M., Carter, A.J., Cohen, P., Copeland, R.A., Cravatt, B., Dahlin, J.L., Dhanak, D., Edwards, A.M., Frederiksen, M., Frye, S.V., Gray, N., Grimshaw, C.E., Hepworth, D., Howe, T., Huber, K.V.M., Jin, J., Knapp, S., Kotz, J.D., Kruger, R.G., Lowe, D., Mader, M.M., Marsden, B., Mueller-Fahrnow, A., Müller, S., O'Hagan, R.C., Overington, J.P., Owen, D.R., Rosenberg, S.H., Ross, R., Roth, B., Schapira, M., Schreiber, S.L., Shoichet, B., Sundström, M., Superti-Furga, G., Taunton, J., Toledo-Sherman, L., Walpole, C., Walters, M.A., Willson, T.M., Workman, P., Young, R.N., Zuercher, W.J.: The promise and peril of chemical probes. Nature Chemical Biology **11**, 536–41 (2015)
5. 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)
6. Franken, H., Mathieson, T., Childs, D., Sweetman, G.M., Werner, T., Tögel, I., Doce, C., Gade, S., Bantscheff, M., Drewes, G., Reinhard, F.B.M., Huber, W., Savitski, M.M.: Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. Nature Protocols **10**, 1567–93 (2015)
7. 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)
8. Tan, C.S.H., Go, K.D., Bisteau, X., Dai, L., Yong, C.H., Prabhu, N., Ozturk, M.B., Lim, Y.T., Sreekumar, L., Lengqvist, J., Tergaonkar, V., Kaldis, P., Sobota, R.M., Nordlund, P.: Thermal proximity coaggregation for system-wide profiling of protein complex dynamics in cells. Science **359**, 1170–7 (2018)
9. Reinhard, F.B.M., Eberhard, D., Werner, T., Franken, H., Childs, D., Doce, C., Savitski, M.F., Huber, W., Bantscheff, M., Savitski, M.M., Drewes, G.: Thermal proteome profiling monitors ligand interactions with cellular membrane proteins. Nature Methods **12**, 1129–31 (2015)
10. Huber, K.V.M., Olek, K.M., Müller, A.C., Tan, C.S.H., Bennett, K.L., Colinge, J., Superti-Furga, G.: Proteome-wide drug and metabolite interaction mapping by thermal-stability profiling. Nature Methods **12**, 1055–7 (2015)
11. Becher, I., Werner, T., Doce, C., Zaal, E.A., Tögel, I., Khan, C.A., Rüger, A., Mülbaier, M., Salzer, E., Berkers, C.R., Fitzpatrick, P.F., Bantscheff, M., Savitski, M.M.: Thermal profiling reveals phenylalanine hydroxylase as an off-target of panobinostat. Nature Chemical Biology **12**, 908–10 (2016)

12. Mateus, A., Bobonis, J., Kurzawa, N., Stein, F., Helm, D., Hevler, J., Typas, A., Savitski, M.M.: Thermal proteome profiling in bacteria: probing protein state in vivo. Molecular Systems Biology **14**, 8242 (2018)
13. Dziekan, J.M., Yu, H., Chen, D., Dai, L., Wirjanata, G., Larsson, A., Prabhu, N., Sobota, R.M., Bozdech, Z., Nordlund, P.: Identifying purine nucleoside phosphorylase as the target of quinine using cellular thermal shift assay. Science Translational Medicine **11**(473), 3174 (2019)
14. Türkowsky, D., Lohmann, P., Mühlenbrink, M., Schubert, T., Adrian, L., Goris, T., Jehmlich, N., von Bergen, M.: Thermal proteome profiling allows quantitative assessment of interactions between tetrachloroethene reductive dehalogenase and trichloroethene. Journal of Proteomics **192**, 10–7 (2019)
15. Dai, L., Zhao, T., Bisteau, X., Sun, W., Prabhu, N., Lim, Y.T., Sobota, R.M., Kaldis, P., Nordlund, P.: Modulation of protein-interaction states through the cell cycle. Cell **173**(6), 1481–94 (2018)
16. Becher, I., Andrés-Pons, A., Romanov, N., Stein, F., Schramm, M., Baudin, F., Helm, D., Kurzawa, N., Mateus, A., Mackmull, M.-T., Typas, A., Müller, C.W., Bork, P., Beck, M., Savitski, M.M.: Pervasive protein thermal stability variation during the cell cycle. Cell **173**(6), 1495–507 (2018)
17. Azimi, A., Caramuta, S., Seashore-Ludlow, B., Boström, J., Robinson, J.L., Edfors, F., Tuominen, R., Kemper, K., Krijgsman, O., Peeper, D.S., Nielsen, J., Hansson, J., Egyhazi Brage, S., Altun, M., Uhlen, M., Maddalo, G.: Targeting CDK2 overcomes melanoma resistance against BRAF and Hsp90 inhibitors. Molecular Systems Biology **14**, 7858 (2018)
18. Miettinen, T.P., Peltier, J., Härtlova, A., Gierliński, M., Jansen, V.M., Trost, M., Björklund, M.: Thermal proteome profiling of breast cancer cells reveals proteasomal activation by CDK4/6 inhibitor palbociclib. The EMBO Journal **37**, 98359 (2018)
19. Savitski, M.M., Zinn, N., Fälth-Savitski, M., Pöckel, D., Gade, S., Becher, I., Mülbaier, M., Wagner, A.J., Strohmer, K., Werner, T., Melchert, S., Petretich, M., Rutkowska, A., Vappiani, J., Franken, H., Steidel, M., Sweetman, G.M., Gilan, O., Lam, E.Y.N., Dawson, M.A., Prinjha, R.K., Grandi, P., Bergamini, G., Bantscheff, M.: Multiplexed proteome dynamics profiling reveals mechanisms controlling protein homeostasis. Cell **173**, 260–74 (2018)
20. Childs, D., Kurzawa, N., Franken, H., Doce, C., Savitski, M., Huber, W.: TPP: Analyze Thermal Proteome Profiling (TPP) Experiments. (2017). R package version 3.9.6
21. Volkening, J.: Mstherm: Analyze MS/MS Protein Melting Data. (2017). R package version 0.4.7. https://CRAN.R-project.org/package=mstherm
22. Ramsey, J.O., Silverman, B.W.: Functional Data Analysis, 2nd edn. Springer, New York (2005)
23. Wang, J.-L., Chiou, J.-M., Müller, H.-G.: Functional data analysis. Annual Review of Statistics and its Application **3**, 257–95 (2016)
24. DeGroot, M.H., Schervish, M.J.: Probability and Statistics, 4th edn. Pearson Education, Boston (2012)
25. Venables, W.N., Ripley, B.D.: Modern Applied Statistics with S, 4th edn. Springer, New York (2002)
26. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B **57**, 289–300 (1995)
27. Ward, N.E., O'Brian, C.A.: Kinetic analysis of protein kinase C inhibition by staurosporine: evidence that inhibition entails inhibitor binding at a conserved region of the catalytic domain but not competition with substrates. Molecular Pharmacology **41**, 387–92 (1992)
28. Leung, K.K.K., Shilton, B.H.: Quinone reductase 2 is an adventitious target of protein kinase CK2 inhibitors TBBz (TBI) and DMAT. Biochemistry **54**, 47–59 (2015)
29. Winger, J.A., Hantschel, O., Superti-Furga, G., Kuriyan, J.: The structure of the leukemia drug imatinib bound to human quinone reductase 2 (NQO2). BMC Structural Biology **9**, 7 (2009)
30. Bantscheff, M., Eberhard, D., Abraham, Y., Bastuck, S., Boesche, M., Hobson, S., Mathieson, T., Perrin, J., Raida, M., Rau, C., Reader, V., Sweetman, G.M.A., Bauer, A., Bouwmeester, T., Hopf, C., Kruse, U., Neubauer, G., Ramsden, N.G., Rick, J., Kuster, B., Drewes, G.: Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. Nature Biotechnology **25**, 1035–1044 (2007)
31. Miettinen, T.P., Björklund, M.: NQO2 is a reactive oxygen species generating off-target for acetaminophen. Molecular Pharmaceutics **11**, 4395–404 (2014)
32. Bourgon, R.W.: Chromatin immunoprecipitation and high-density tiling microarrays: a generative model, methods for analysis, and methodology assessment in the absence of a" gold standard". PhD thesis, Citeseer (2006)
33. Loader, C.: Local Regression and Likelihood. Springer, New York (1999)
34. Storey, J.D., Xiao, W., Leek, J.T., Tompkins, R.G., Davis, R.W.: Significance analysis of time course microarray experiments. Proceedings of the National Academy of Sciences **102**, 12837–42 (2005)
35. Berk, M., Ebbels, T., Montana, G.: A statistical framework for biomarker discovery in metabolomic time course data. Bioinformatics **27**, 1979–85 (2011)
36. Childs, D., Kurzawa, N.: Non-parametric analysis of thermal proteome profiles: Workflow (2019). https://git.embl.de/childs/TPP-data-analysis

Figures

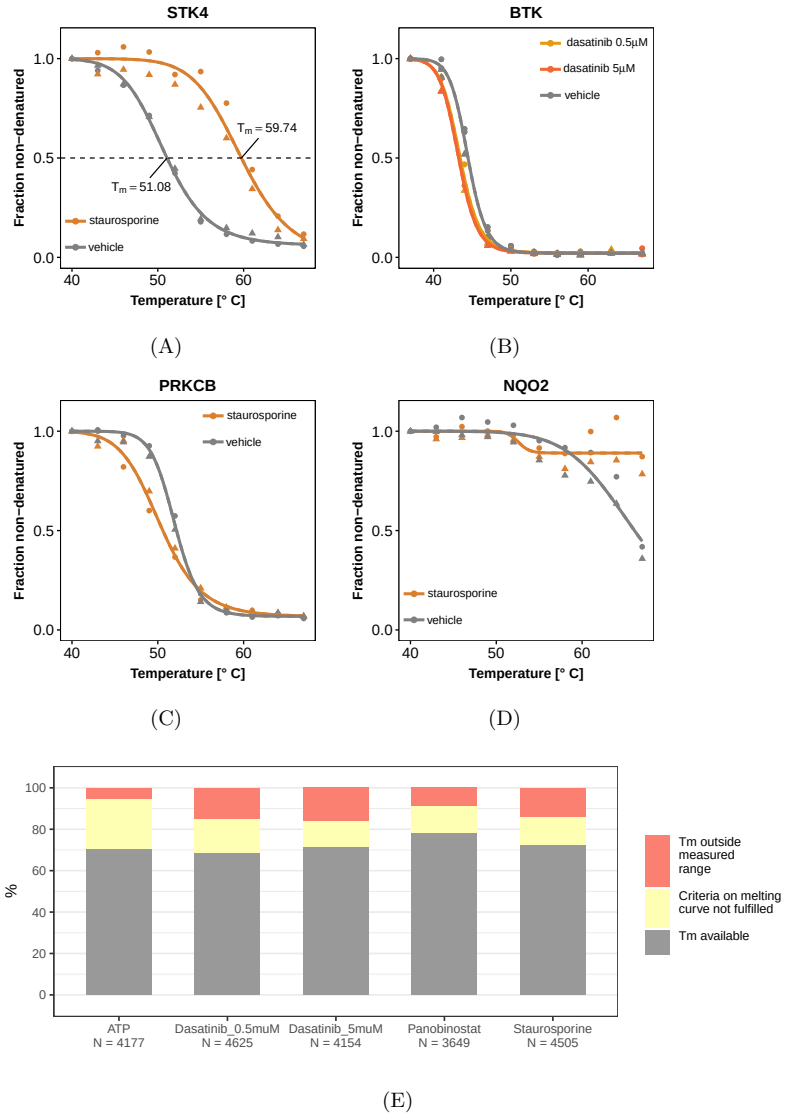

Figure 1 TPP data analysis challenges. (A)-(D) Examples for protein melting curves with and without drug (see color keys). In each case, ten temperatures were assayed, and two experimental replicates were made per condition, indicated by circle and triangle symbols. Fits of the sigmoid model (Eqn. (3)) to both replicates jointly are shown by smooth lines. **(A)** For serine/threonine protein kinase 4 (STK4), the binding of staurosporine is reflected by a marked shift between the curves. The fitted values for the melting points (T_m) are shown. **(B)** For Bruton's tyrosine kinase (BTK), there is a small but reproducible shift between the curves. **(C)** Protein kinase C beta (PRKCB) is destabilized by staurosporine; the effect occurs mainly at lower temperatures. **(D)** NAD(P)H quinone dehydrogenase 2 (NQO2) is strongly stabilized by staurosporine. While in each case, the effects of drug binding are clearly reproducible between replicates, the T_m-based approach of [6] only detects (A) and misses (B)-(D). In the case of (B) and (C), the fitted T_m are too similar, so that the statistical test does not assess the difference as significant. In the case of (D), no reasonable estimate for T_m in the staurosporine treated condition can be obtained, as it would lie outside the measured temperature range, and the protein is discarded from the analysis. In contrast, NPARC, the method proposed in this article, detects all four cases. **(E)** The fraction of proteins in each of the data sets of Table 1 that is missed by the T_m-based approach due to failure to estimate T_m or to meet the goodness-of-fit criterion (Table 3, Figure S2).

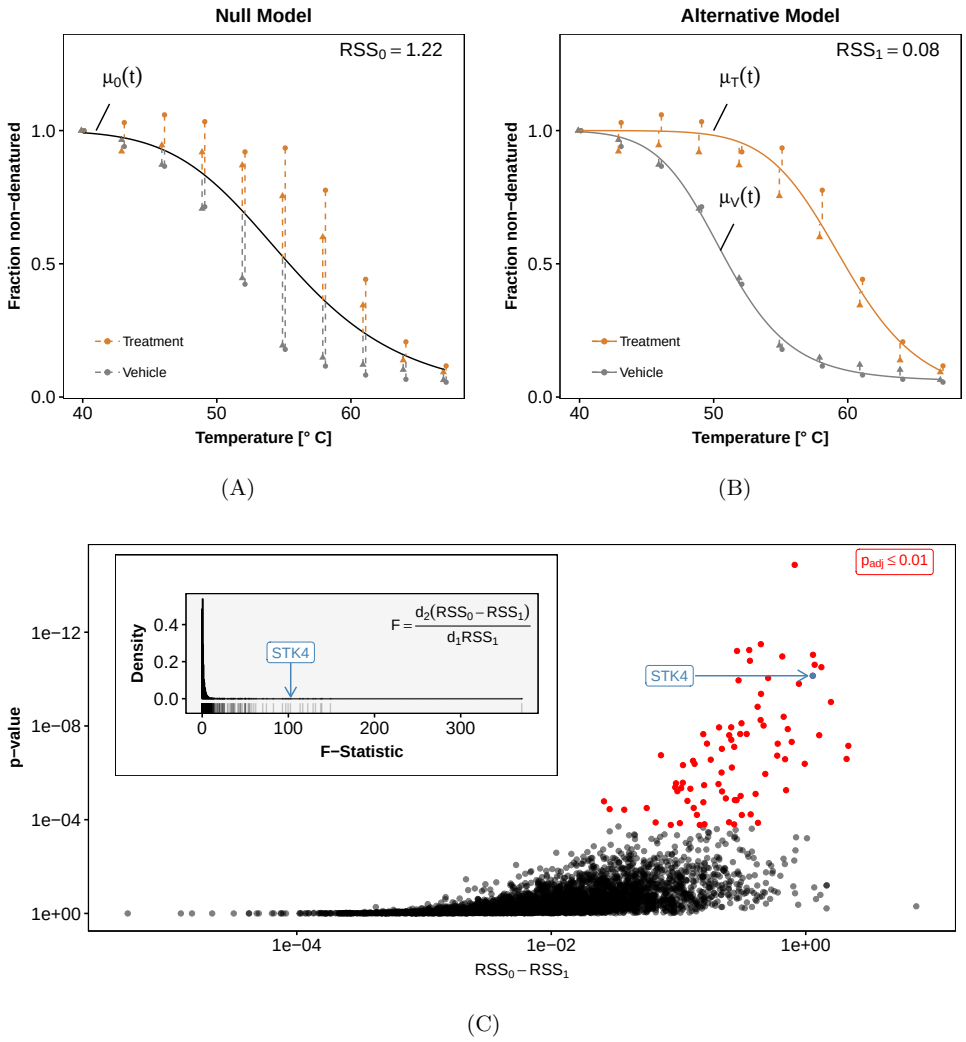

Figure 2 Principles of NPARC, illustrated for protein STK4 under staurosporine treatment. **(A)** Fit of the null model, i.e., no treatment effect (black line). The goodness-of-fit is quantified by RSS_0, the sum of squared residuals (dashed lines). As in Figure 1, the triangle and circle symbols indicate the experimental replicates. **(B)** Fit of the alternative model, with separate curves for the treated (orange) and the vehicle condition (grey). Due to the higher flexibility of the model, the sum of squared residuals RSS_1 is always less than or equal to RSS_0. **(C)** The question whether the improvement in the goodness-of-fit, i.e., the difference $RSS_0 - RSS_1$, is strong enough to reject the null hypothesis can be addressed with the variant of the F-test described in the main text. Each point in the plot corresponds to a different protein. The highlighted example STK4 has a large F-statistic and a small p-value.

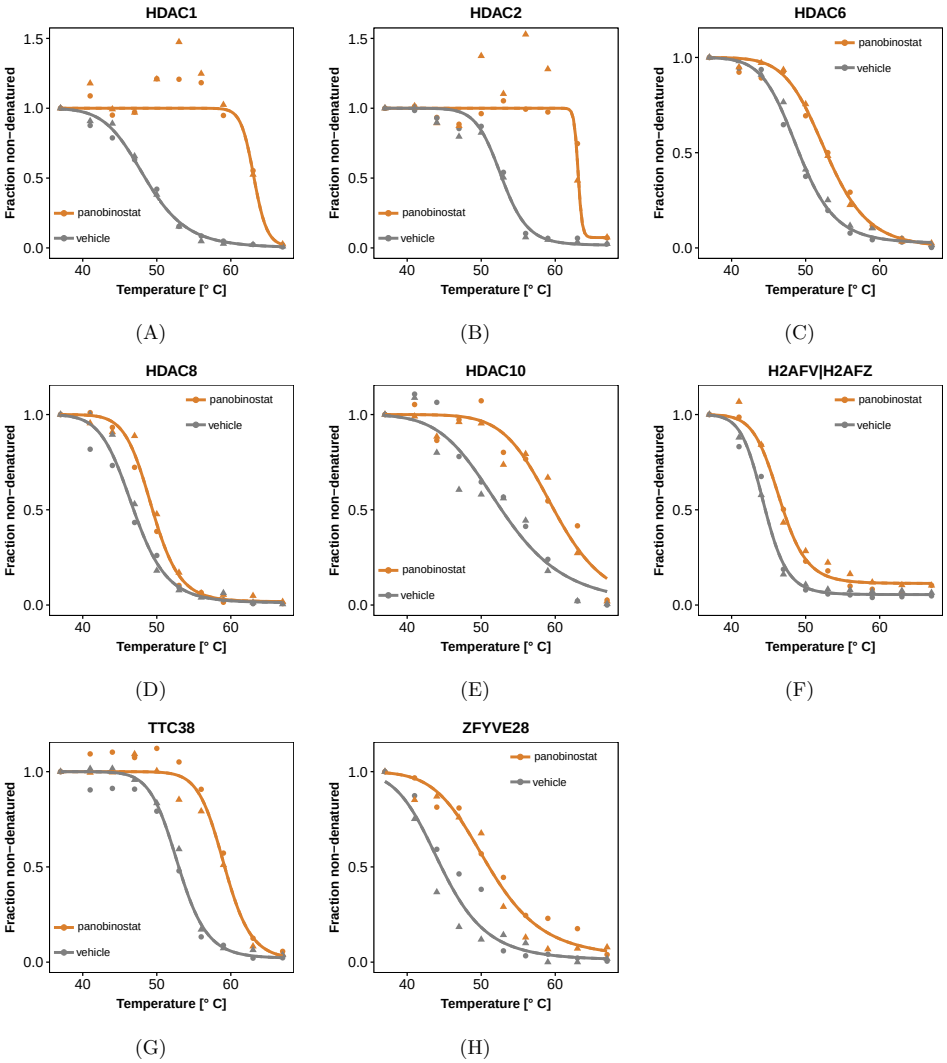

Figure 3 Direct and indirect targets of the HDAC inhibitor panobinostat detected by NPARC (FDR ≤ 0.01 according to the method of Benjamini and Hochberg [26]). **(A)-(E)** Data and curve fits for five HDACs that show significant shifts in their thermostability. HDAC1 and HDAC2 are not detected by the T_{m}-based approach of [6], since the higher variance between the replicates of the panobinostat-treated condition leads to them being eliminated by the filter heuristics of that method (Table 3). In contrast, NPARC naturally takes the variance into account in the computation of the F-statistic and does not require such filtering steps. **(F)-(H)** Data and curve fits for known non-HDAC targets.

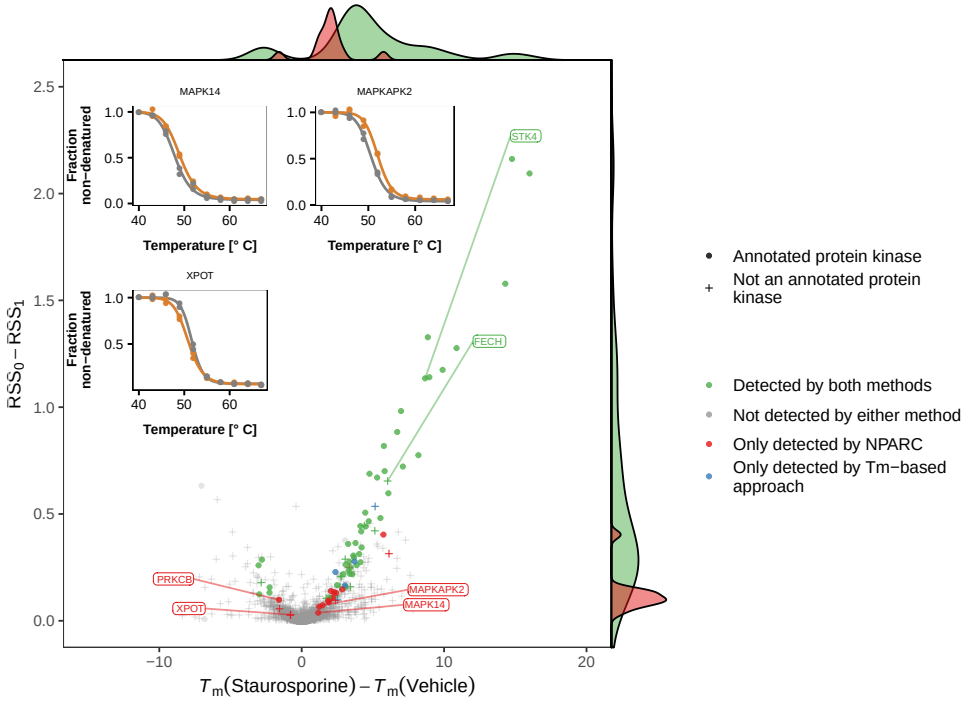

Figure 4 NPARC is sensitive to small but reproducible T_m-shifts. The plot compares the effect size measure used by NPARC, namely $RSS_0 - RSS_1$ (y-axis), to the T_m-difference (x-axis) for those proteins in the staurosporine data set for which T_m estimates could be obtained. Proteins with Benjamini-Hochberg adjusted p-values ≤ 0.01 are marked in red if they were exclusively found by NPARC, and in green if they were also detected by the T_m-based approach of [6]. NPARC detects targets with small T_m-differences if the measurements are reproducible between replicates.

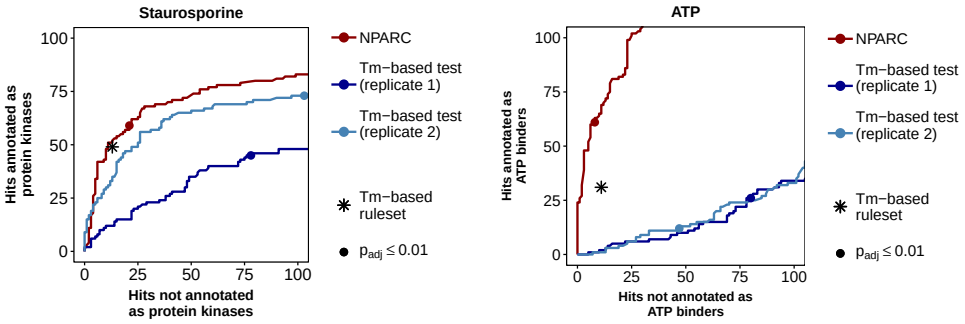

Figure 5 Sensitivity and specificity. Shown are pseudo-ROC curves, with expected hits (as a proxy for true positives) along the y-axis and unexpected hits (as a proxy for false positives) along the x-axis. The curves are obtained by varying the p-value cutoff of the F-test of NPARC (which is computed across replicates), and of the z-test of the T_{m}-based approach [6] (which is computed separately for each replicate). The asterisks indicate the result from the decision rules of [6] on the z-test results (Table 5). The dots indicate a threshold of 0.01 on the Benjamini-Hochberg adjusted p-values from NPARC. NPARC is modestly better than the T_{m}-based approach on the staurosporine data **(A)**, and substantially better on the ATP data **(B)**. The proteins found by NPARC at Benjamini-Hochberg adjusted p-value ≤ 0.01 are also shown in Supplementary Figures S6 and S7.