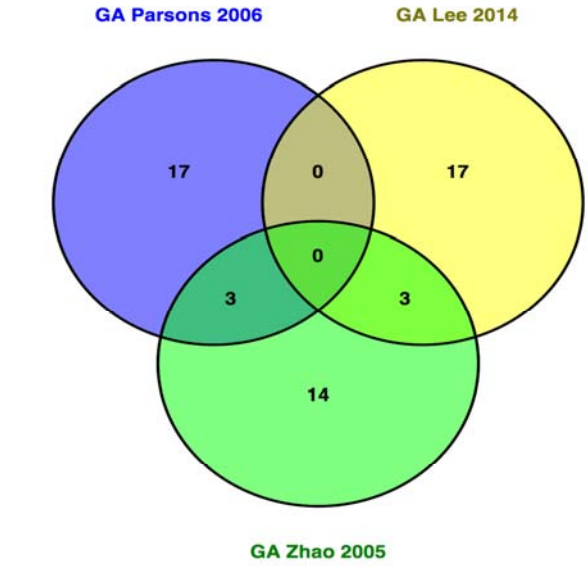
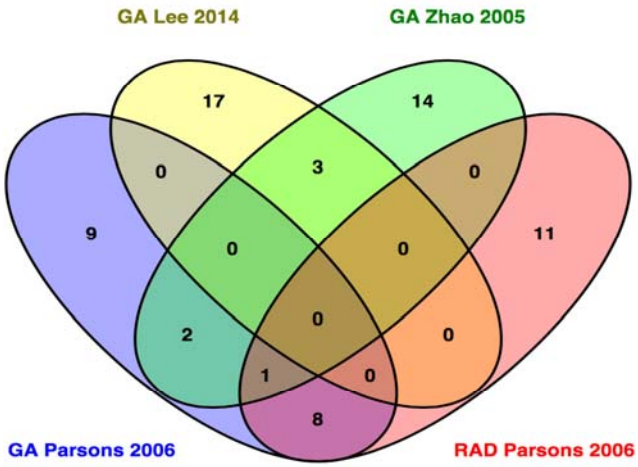
Supporting Information

Title: A Chemogenomic Screening Platform Used to identify Chemotypes Perturbing HSP90 Pathways

Authors: Fiona M Thomas[1], Kourtney M Goode[1], Bartek Rajwa[2], Andrew B. Bieberich[1], Larisa V. Avramova[2], Tony R. Hazbun[1], V. Jo Davisson[1*]

Supporting information contents





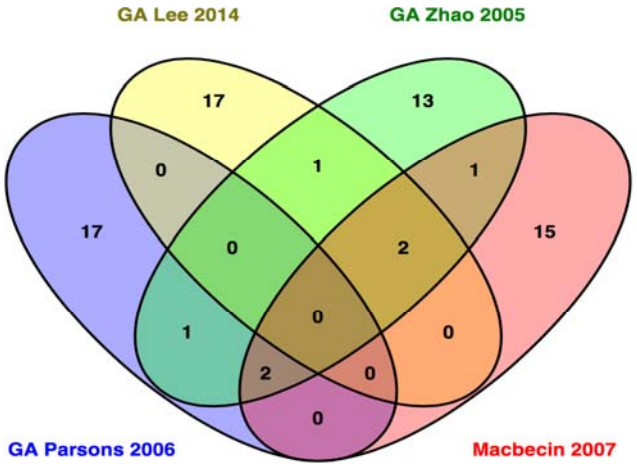


Figure S1: Comparison of the top 20 sensitive strains to several Hsp90 inhibitors from four publications. A. Comparison of geldanamycin (GA) from three separate studies shows that no deletion strains were observed in common and only 3 strains were shared between two studies. **B**. Additional comparison of the GA studies with radicicol (RAD). An overlap of 8 strains was observed by the Parsons 2006 study for GA and RAD. However, overlap with other studies was non-existent. **C**. Comparison of GA studies with macbecin sensitivity (McClellan et al. 2007). Only five strains exhibiting macbecin sensitivity overlapped with the other three GA studies.

References as cited in article:

6. Parsons AB, et al.: Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell* 2006; 126:611-625.

16. McClellan AJ, et al.: Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. *Cell* 2007; 131:121-135.

21. Lee AY, et al.: Mapping the cellular response to small molecules using chemogenomic fitness signatures. *Science* 2014; 344:208-211.

22. Zhao R, et al: Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell* 2005; 120:715-727.

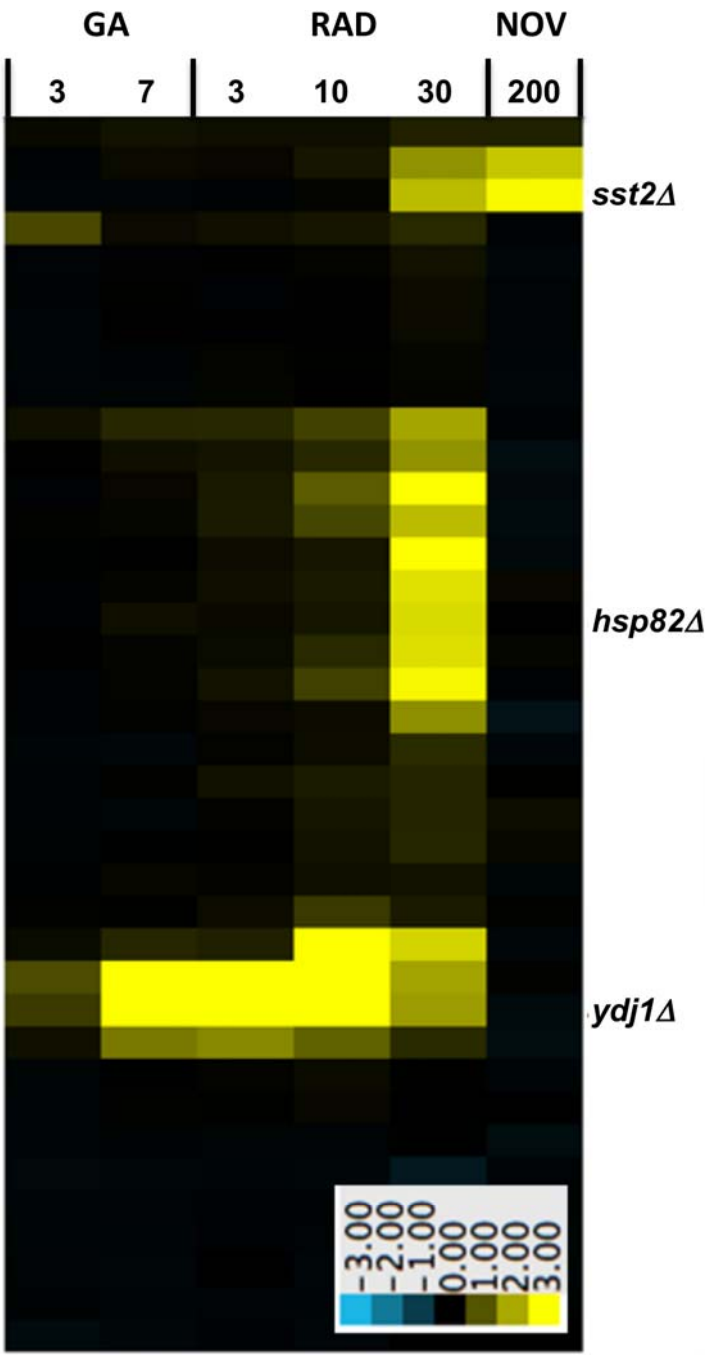


Figure S2: Sensitivity of yeast deletion strains to geldanamycin (GA), radicicol (RAD) and novobiocin (Nov). Growth rates of strains at several concentrations (μM) displayed on the x-axis were measured and a growth ratio relative to wild-type strain was determined followed by clustering using the centroid linkage algorithm. Yellow indicates increased sensitivity and blue indicates resistance relative to the wild-type strain at each condition – nearly all strains behaved similarly to wild-type or were more sensitive. The three deletion strains selected for further screening are labeled. Identities of all strains, time to reach OD600=0.8 and raw data curves are present in an excel file named Table S1.

Term	Definition
% change	$\left[\frac{Distance\ between\ mean\ drug\ curve\ and\ mean\ control\ curve)}{Distance\ between\ mean\ control\ curve\ and\ a\ no\ growth\ curve}\right] * 100$
Sum Index	Sum of % change of all four strains under a particular drug treatment Sum Index = %Hsp40 + %Hsp82 + %Sst2 + %WT
Diversity Index	Largest pairwise difference calculated between % change of all four strains under a particular treatment i.e. the largest difference between |%Hsp40-%Hsp82|; |%Hsp40-%Sst2|; |%Hsp40-%WT|; |%Hsp82-%Sst2| etc
v-value	$[Distance\ between\ mean\ drug\ curve\ and\ mean\ control\ curve] \div [distance\ measured\ at\ 95th\ percentile\ (critical\ value)]$

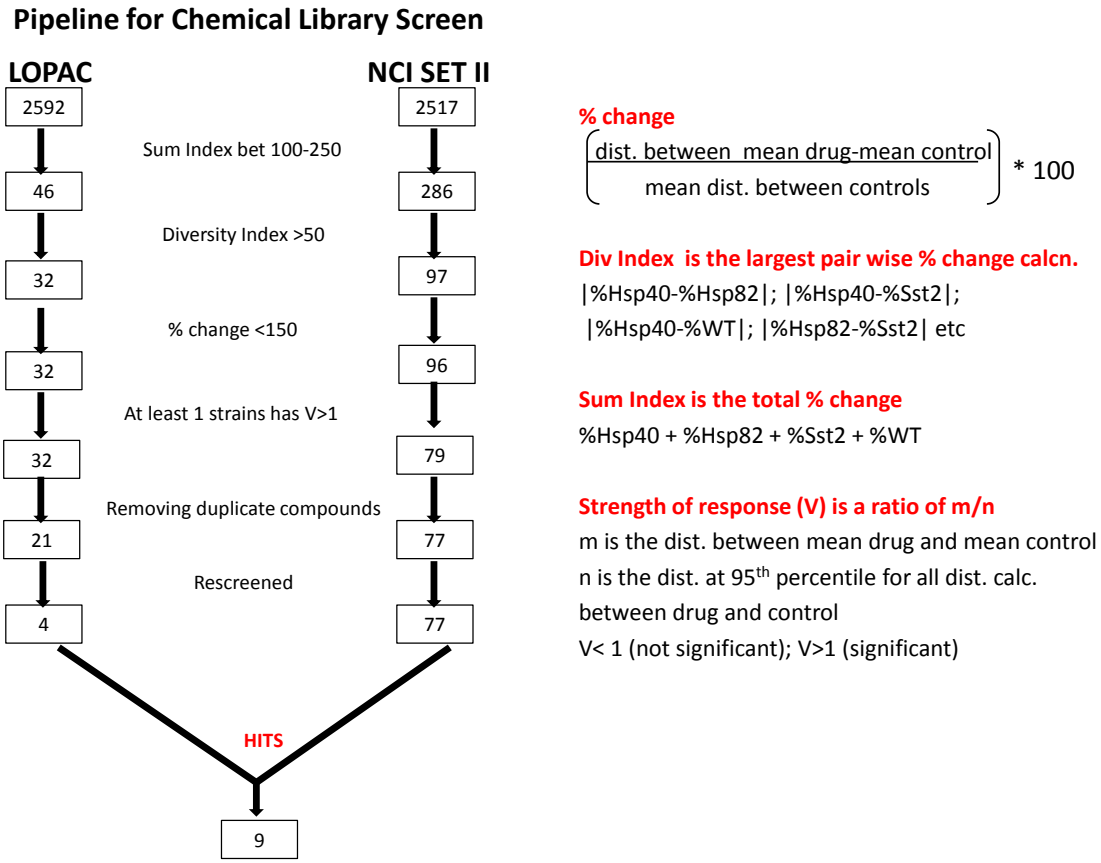
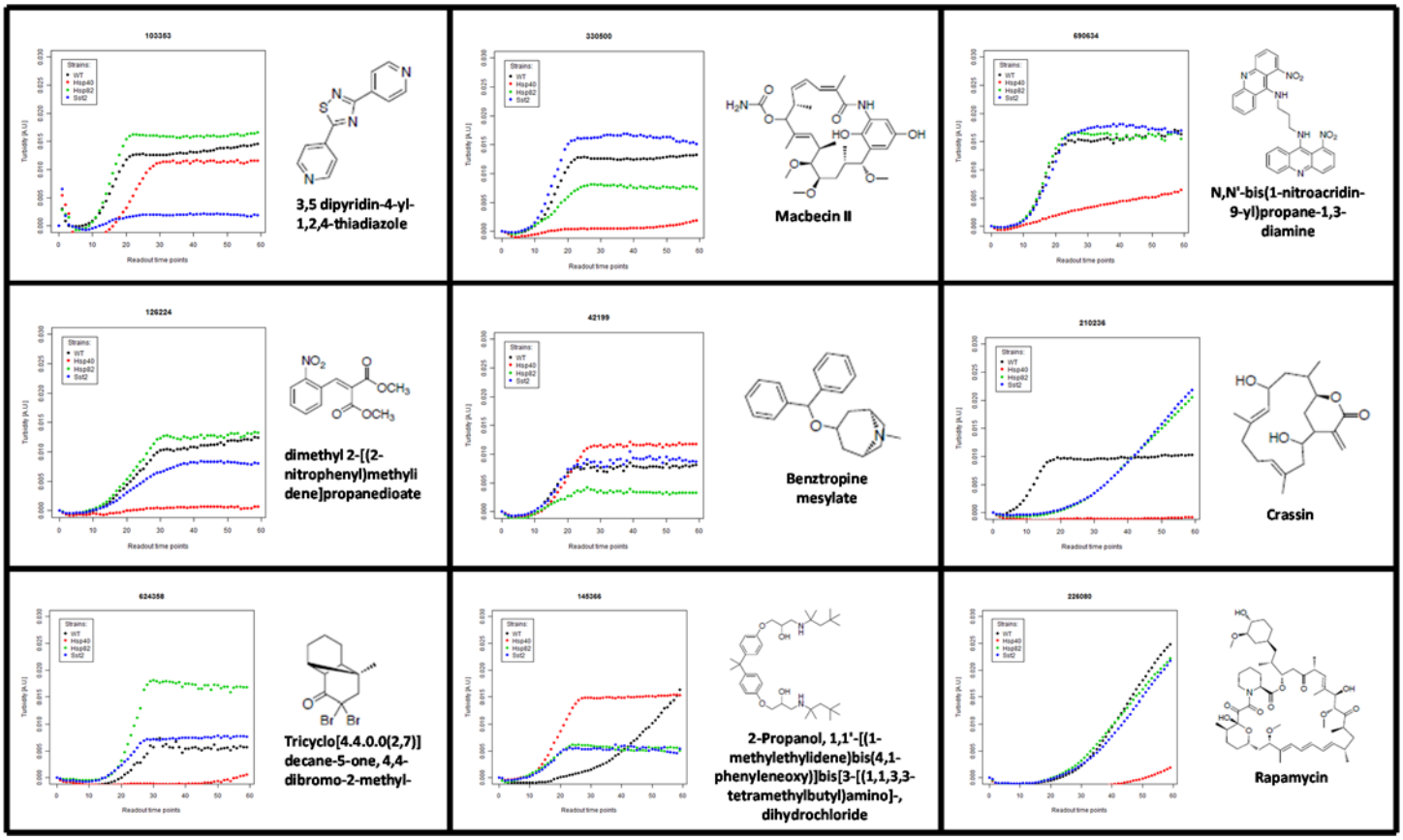


Figure S3. A) Summary of calculations performed to determine % change, Sum Index, Diversity Index and V-values; **B)** A schematic of the filtering process as applied to primary screening data for two chemical libraries that lead to the defining of the Class I hits.



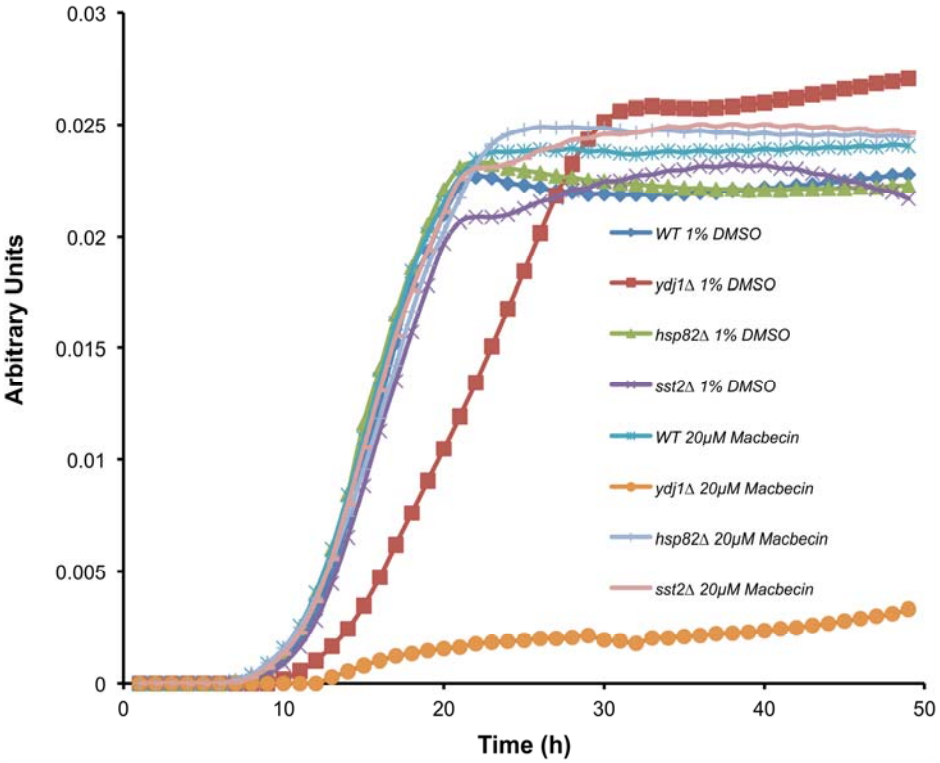


Figure S4. A) Examples of averaged raw data curves for Class I hit compounds screened at 20 μM; **B)** Normalized turbidity curves of deletion mutants demonstrating sensitivity of the *ydj1Δ* strain to 20 μM Macbecin.

Strain List	Inclusion Criteria
Zhao and Gong - 75	Strains corresponding to chaperone genes in Gong et al and the 12 strains in the top 20 GA-sensitive strains in Zhao et al
Parsons - 150	Strains in the top 100 sensitive to Radicicol or Geldanamycin
Hsp90 literature - 90	Strains corresponding with connection to Hsp90 or sensitivity to inhibitors in the literature
McClellan - 99	Strains most sensitive to Macbecin at 30C or 37C
Insensitive Strains - 12	Randomly selected strains
HAS - 16	Final Heat shock associated sensitive strains

* Essential chaperone genes so heterozygous diploid strains were used

6. Parsons, A. B.; Lopez, A.; Givoni, I. E.; et al. Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. Cell 2006, 126 (3), 611-25.

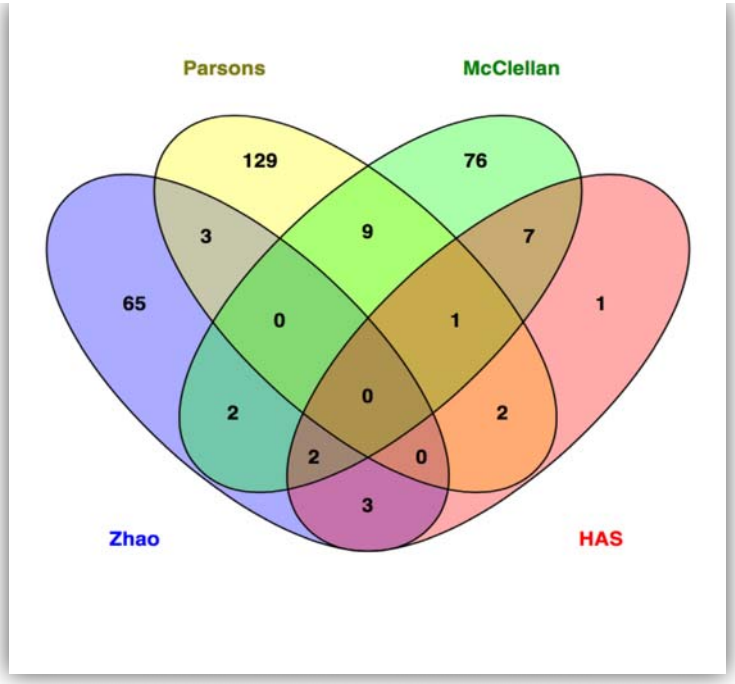
14. Zhao, R.; Davey, M.; Hsu, Y. C.; et al. Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. Cell 2005, 120 (5), 715-27.

16. McClellan, A. J.; Xia, Y.; Deutschbauer, A. M.; et al. Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. Cell 2007, 131 (1), 121-35.

33. Gong, Y.; Kakihara, Y.; Krogan, N.; et al. An atlas of chaperone-protein interactions in Saccharomyces cerevisiae: implications to protein folding pathways in the cell. Mol Syst Biol 2009, 5, 275.

Zhao - 75	Parsons - 150		Hsp90 literature - 90		McClellan - 99		Insensitive Strains - 12	HAS - 16	
YDR320C	YAL005C	YHR139C	YLR024C	YDR283C	YAL002W	YNL197C	YAL007C	YBL008W	HIR1
YER048C	YAL019W	YHR167W	YML057W	YGL020C	YAL024C	YNL235C	YAL008W	YGL005C	COG7
YFL016C	YAL024C	YJR056C	YAL019W	YER103W	YBL051C	YNL294C	YAL015C	YGR285C	ZUO1
YFR041C	YBL058W	YJR058C	YML032C	YER110C	YBR061C	YNL298W	YAL018C	YIR019C	FLO11
YGL018C*	YBL101C	YKL057C	YMR186W	YAL024C	YBR164C	YNL322C	YAL020C	YJL179W	PFD1
YGL128C*	YBR101C	YKL092C	YAL005C	YMR097C	YCL005W	YNL325C	YAL027W	YJR073C	OPI3
YGR285C	YBR156C	YLR048W	YLL026W	YMR098C	YCL008C	YNR006W	YAL035W	YKL037W	AIM26
YIR004W	YDL134C	YLR073C	YLR113W	YER139C	YDL077C	YNR051C	YAL036C	YLR452C	SST2
YJL073W	YDR192C	YLR328W	YLL024C	YJR032W	YDL117W	YOL050C	YAL037W	YML071C	COG8
YJL162C	YDR200C	YLR329W	YOR007C	YER151C	YDR136C	YOL061W	YAL039C	YMR095C	SNO1
YJR097W	YDR435C	YLR344W	YOR027W	YER156C	YDR253C	YOL089C	YAL040C	YNL064C	YDJ1
YLR008C*	YEL025C	YML005W	YOR023C	YER145C	YDR260C	YOR027W	YAL046C	YNL098C	RAS2
YLR090W	YER056C-A	YML108W	YOR036W	YBR101C	YDR276C	YOR061W		YNL197C	WHI3
YMR161W	YER139C	YMR269W	YNL325C	YGR257C	YDR335W	YOR089C		YNL322C	KRE1
YMR214W	YFR035C	YNR009W	YPL253C	YNL307C	YDR443C	YOR275C		YPL106C	SSE1
YNL007C*	YFR043C	YOL004W	YNL281W		YER048C	YPL002C		YPL240C	HSP82
YNL064C	YFR053C	YOL076W	YPL240C		YER072W	YPL051W			
YNL077W	YGL257C	YOR087W	YOL018C		YER119C-A	YPL065W			
YNL227C	YGR235C	YOR228C	YDR073W		YER120W	YPL084W			
YNL328C	YGR243W	YOR239W	YPL193W		YER151C	YPL225W			
YOR254C*	YHL007C	YOR270C	YDR418W		YER177W	YPL240C			
YPR061C	YHL032C	YOR284W	YPL152W		YFL025C	YPR119W			
YLR259C*	YHR029C	YPL036W	YPL106C		YFR036W				
YAL005C	YHR124W	YPL060W	YHR034C		YGL005C				
YBL075C	YHR162W	YPL070W	YGR178C		YGL013C				
YBR169C	YHR184W	YFR043C	YCL037C		YGL045W				
YDL229W	YIL160C	YPL253C	YER072W		YGL046W				
YEL030W	YIL161W	YMR095C	YGR123C		YGL151W				
YER103W	YIR004W	YGR188C	YHL007C		YGL212W				
YHR064C	YIR019C	YML124C	YHR030C		YGR289C				
YJL034W*	YJL013C	YDR415C	YHR167W		YHL027W				
YJR045C*	YJR054W	YOL081W	YKL117W		YHR073W				
YKL073W	YJR109C	YOR027W	YLR216C		YHR079C				
YLL024C	YKL161C	YKL205W	YOR216C		YHR167W				
YLR369W	YLL024C	YER072W	YOR228C		YIL029C				
YNL209W	YLR004C	YOR216C	YLR362W		YIL035C				
YPL106C	YLR201C	YMR311C	YLR318W		YIL041W				
YMR186W	YML027W	YOL032W	YLR319C		YIL053W				
YPL240C	YML086C	YMR096W	YJL187C		YIL121W				
YEL003W	YML116W	YML123C	YGL253W		YIL154C				
YGR078C	YML117W	YBL068W	YDR214W		YIL170W				

Figure S5. Source and process for 360 strains screen to identify heat shock associated strains as indicators for follow up profiles.



Common Elements between Datasets

7 common elements in "McClellan" and "HAS":

YGL005C YNL098C

YJR073C YNL197C

YKL037W YNL322C

YML071C

1 common element in "Parsons", "McClellan" and "HAS":

YMR095C

2 common elements in "Parsons" and "HAS":

YIR019C

YBL008W

2 common elements in "Zhao", "McClellan" and "HAS":

YPL240C

YJL179W

3 common elements in "Zhao" and "HAS":

YGR285C

YNL064C

YPL106C

1 element included exclusively in "HAS":

YLR452C

Figure S6. Venn diagram demonstrating overlap between Heat shock associates strains (HAS) and previous screening efforts

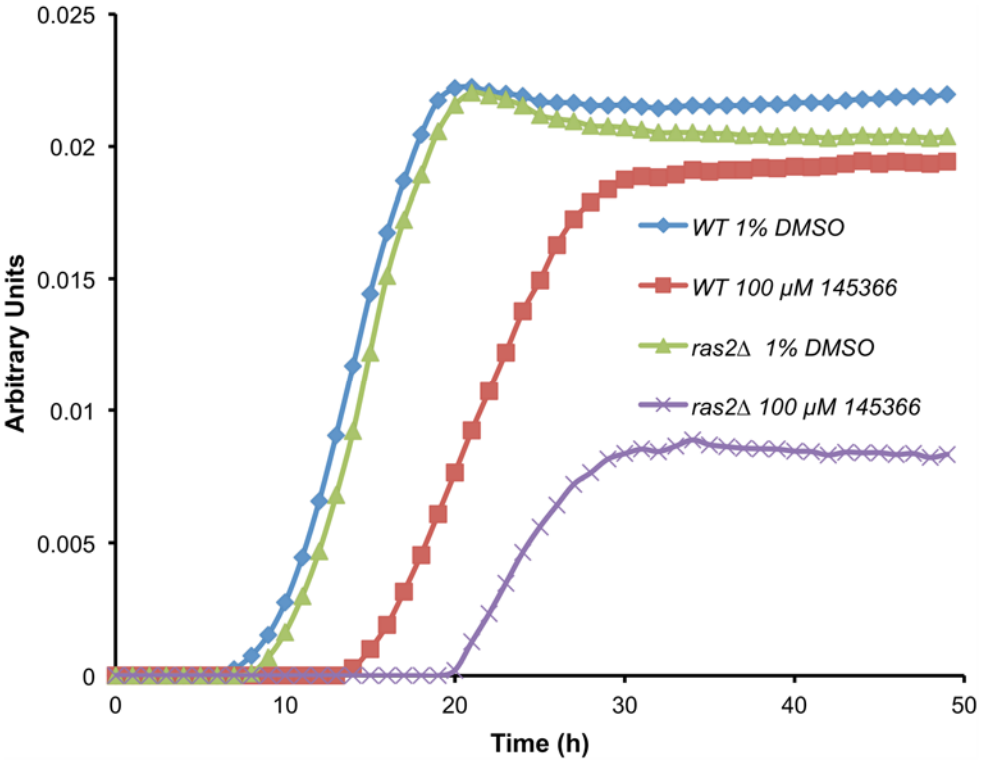


Figure S7. Normalized growth curves of *ras2⊿* strain treated with 1% DMSO and 100 µM NSC145366.

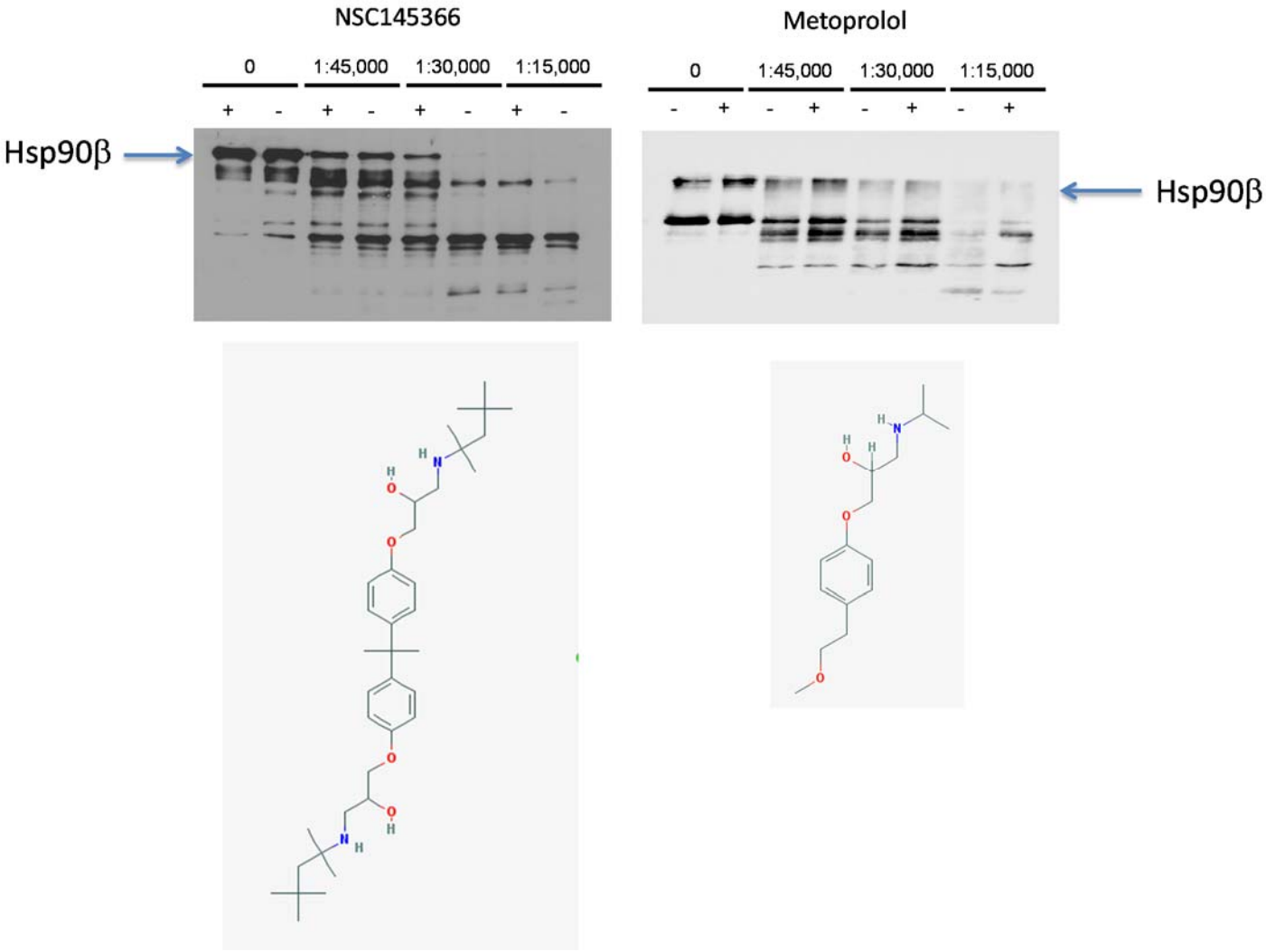


Figure S8: Negative control data for DARTs experiments with recombinant Hsp90 beta. The panel on the left is identical to Figure 4A and is used here for visual comparison with the data for the negative control compound metoprolol selected because of common chemical features to NSC145366. Both compounds are at 200 μM.

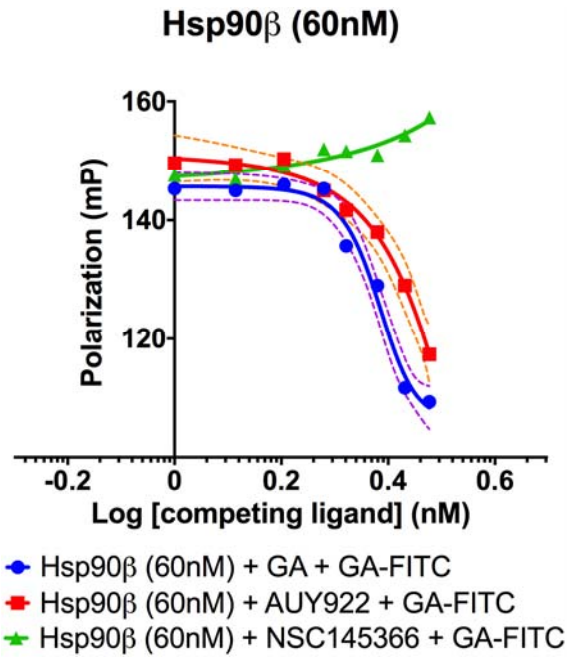


Figure S9. Competition assay using Hsp90 inhibitors Geldanamycin (GA), AUY922 and NSC145366 to displace FITC-GA binding to Hsp90β. Known N-terminal ATP site binding compounds GA and AUY922 displace FITC-GA but NSC145366 does not, indicating that NSC145366 does not bind to the Hsp90 βN-terminal ATP site.

Detailed method description of distance measurements:

Data Analysis for Identification of haploid deletion strains sensitive to Hsp90 inhibitors. Time-dependent changes in microplate well turbidities were measured by assessing optical density. These data vectors were first normalized using their integrals and the values of the initial turbidity measurements. Mean control- and drug-treated curves for each of the strains were computed. The curves were smoothed employing a robust locally weighted regression function (*lowess*)[1]. For analysis, pairwise distances between curves (yeast strains exposed to a treatment compared to the corresponding DMSO curves) were measured using quadratic form distance (QFD), Euclidean distance, KS distance metric, spectral angle distance (cosine distance), and dynamic time warping (DTW) distance. The final assessment of the curve characteristics was performed using DTW and QFD distances.

The distances were defined as follows:

Euclidean distance: $$d(p,q)=\left(\sum_{i=1}^{n}(q_i-p_i)^2\right)^{\frac{1}{2}}$$

KS distance: $$d(p,q)=\sup\left(\frac{1}{n}\left|\sum_{i=1}^{n-X}p_i-\sum_{i=1}^{n-X}q_i\right|\right),\quad X=0,1,\ldots,n$$

Cosine distance: $$d(p,q)=1-\frac{\sum_{i=1}^{n}p_iq_i}{\left(\sum_{i=1}^{n}p_i^2\sum_{i=1}^{n}q_i^2\right)^{\frac{1}{2}}}$$

Quadratic form distance: $$d(p,q)=\left(\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}(p_i-q_i)(p_j-q_j)\right)^{\frac{1}{2}},\quad a_{ij}=\left(1+(i-j)^2\right)^{\frac{1}{2}}$$

The DTW aligns two time series (or other data vectors) by allowing a given data point from one curve to be matched with one or several points on the other curve. The DTW algorithm finds the alignment between the curves that has the minimal cost (in the sense of a local distance measure). This cost of alignment is the DTW distance [2; 3].

The calculated distances were used to determine the curve-Z' factor [4]. In contrast to the univariate Z', the curve-Z' factor is defined employing curves (data vectors), rather than scalar measurements. The curve-Z' uses the 99-percentile of pairwise distances between curves obtained from the same group as a measure of dispersion within a group (instead of the standard deviations used in the univariate Z'), and 50-percentile pairwise distance between two groups of curves as a measure of the expected distance (instead of a difference between means). Note that this definition of Z' is different from the multivariate extension of Z' proposed by Kümmel et al., who performed a projection of multiple readouts to calculate an equivalent to a univariate Z′ factor [5].

We computed the 99-percentile of pairwise distances between curves in the control and treated groups ($d_{1,1}$, and $d_{2,2}$), as well as the 50-percentile pairwise distance between curves belonging to two different groups ($d_{1,2}$). The curve-Z' is defined as follows:

$$\text{curve-Z'}=1-\frac{d_{1,1}^{[0.99]}+d_{2,2}^{[0.99]}}{d_{1,2}^{[0.5]}}$$

The curve-Z' shows the size of the effect as an overlap/separation between treated and untreated groups. As with the univariate Z' factor, this value indicates to what extent the treatment can cause a large, easily observable effect. In contrast to the traditional univariate screens designed to assist in defining a binary-screen result (hit or no hit), the expectations regarding the curve-Z' values are much more relaxed. In this setting, we do not expect the curve-Z' factor to extend well above 0. For curve-Z' =0 the average separation between the curves from treated and untreated groups would be twice as large as the 99-percentile of inter-curve distance within a group (assuming an identical variance within compared groups). This would indicate an almost complete separation between the groups (equivalent to a 6-standard deviation separation between the means in the univariate case).

The large number of repeated treatments per strain allowed calculation of p-values using an asymptotic Monte-Carlo permutation test comparing the growth vectors. The statistic used followed the idea presented by Hayden et al. and was defined as follows [6]:

$$s = d_{1,2}^{[0.5]} - \frac{d_{1.1}^{[0.5]} - d_{2,2}^{[0.5]}}{2}$$

where $d_{x,x}^{[0.5]}$ is the 50-percentile computed from the distribution of pairwise distances between curves in group x, and $d_{x,y}^{[0.5]}$ is the 50-percentile from the distribution of distances between curves belonging to groups x and y. The p-value for this one-sided test is the probability that the value of the resampled statistic s is equal to or greater than the observed value s_{obs}. Therefore

$$\text{p-value} = \frac{P\left(s \geq s_{obs}\right) + 1}{P + 1},$$

where $P\,(s \geq s_{obs})$ is the number of instances in which s is equal to or exceeds the s_{obs} value. P is the total number of permutations. All calculations were performed using R language for statistical computing (http://www.r-project.org/) [7].

Data Analysis for Chemical Library Screens. The raw time and optical-density data were normalized using the curve integrals. For the purpose of comparison and quantification, reference readouts/curves were established using WT, *sst2Δ*, *ydj1Δ,* and *hsp82Δ* strains exposed to 1% DMSO. The reference curves for every analyzed strain were compared pairwise to establish a distribution of DTW distances. The distance values at the 95th-percentile of the computed pairwise distance distributions were selected as the parameter above which the measured two curves were considered "significantly dissimilar" (i.e., likely originating from different growth phenotypes). Therefore, any distance between given curves that was below this 95-percentile point were considered not significant and likely occurring owing to natural biological variability.

For further analysis, mean control curves were computed for each of the strains using trimmed means of the repeated measurements at all the given time points. For every mean curve, a reference growth-retardation effect was defined as the DTW distance between a mean curve and a curve from wells containing medium only. Using the significance cut-off for DTW distance at the 95th percentile, the fold increase in dissimilarity was calculated for every measured curve / mean-curve pair. The computed parameter was called "response dissimilarity" or v-value. For example, a v-value of 10 means a dissimilarity value 10 ten times larger than the observed variance within the group of curves.

Additional scoring parameters were also computed. A relative difference between the DTW (control, treated sample), and DWT (control, no-growth sample) was scored as %-change values; the value of 1 (or 100%) indicated a DTW distance identical to the distance between the no-growth curve and the mean curve. It is important to point out that the practically observed relative difference could be higher than 100%, since some of the curves obtained from a treated sample may demonstrate a complex or chaotic growth pattern that results in large distance values. It is also important to recognize that the value of 100% reflects different absolute changes in growth (different level of growth retardation) for every one of the four strains, since their respective control mean curves were different.

Summary statistics were computed on the basis of the values defined above (Table S2). The summary included the sum (SUM Index) of all the dissimilarities (for all strains). This parameter was dominated by the greatest dissimilarity value (the strongest response). A high value of the SUM Index indicates that at least one of the strains demonstrated a very strong response to the presence of a drug, or that a number of strains demonstrated modestly strong responses. The diversity index (DIV), another computed summary statistic is defined as the maximum difference between responses exhibited by the tested strains. A small diversity index indicates that all the strains responded in a similar fashion, whereas a high value shows that some strains respond differently than the others did. All these calculations were performed using R language for statistical computing (http://www.r-project.org/). The compounds displaying SUM > 100 <250, DIV >50, % change <150, and V>1 were class I compounds. Class III compounds were defined as those that showed weak or no response in all four strains, i.e., having a DIV Index <40. Artifacts were defined as compounds having Sum Index >400 and DIV Index >100.

Additional references for data analysis methods.

1. Cleveland WS: LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. *The American Statistician* 1981; 35.

2. Gaudin R, Nicoloyannis N: An Adaptable Time Warping Distance for Time Series Learning. Paper presented at the Machine Learning and Applications, 2006 ICMLA '06 5th International Conference on, Dec. 2006 2006.
3. Müller M: Dynamic Time Warping. *Information Retrieval for Music and Motion*: Springer Berlin Heidelberg, 2007:69-84.
4. Zhang J-H, Chung TDY, Oldenburg KR: A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *Journal of Biomolecular Screening* 1999; 4:67-73.
5. Kümmel A, Gubler H, Gehin P, Beibel M, Gabriel D, Parker CN: Integration of Multiple Readouts into the Z′ Factor for Assay Quality Assessment. *Journal of Biomolecular Screening* 2010; 15:95-101.
6. Hayden D, Lazar P, Schoenfeld D, for The Inflammation and the Host Response to Injury I: Assessing Statistical Significance in Microarray Experiments Using the Distance Between Microarrays. *PLoS ONE* 2009; 4:e5838.
7. Ihaka R, Gentleman R: R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 1996; 5:299-314.