Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping

John S Reece-Hoyes[1–3], Alos Diallo[1–3], Bryan Lajoie[1,2,4], Amanda Kent[1–3], Shaleen Shrestha[1–3], Sreenath Kadreppa[1–3], Colin Pesyna[5], Job Dekker[1,2,4], Chad L Myers[5] & Albertha J M Walhout[1–3]

A major challenge in systems biology is to understand the gene regulatory networks that drive development, physiology and pathology. Interactions between transcription factors and regulatory genomic regions provide the first level of gene control. Gateway-compatible yeast one-hybrid (Y1H) assays present a convenient method to identify and characterize the repertoire of transcription factors that can bind a DNA sequence of interest. To delineate genome-scale regulatory networks, however, large sets of DNA fragments need to be processed at high throughput and high coverage. Here we present enhanced Y1H (eY1H) assays that use a robotic mating platform with a set of improved Y1H reagents and automated readout quantification. We demonstrate that eY1H assays provide excellent coverage and identify interacting transcription factors for multiple DNA fragments in a short time. eY1H assays will be an important tool for mapping gene regulatory networks in *Caenorhabditis elegans* and other model organisms as well as in humans.

Gene expression is governed by sequence-specific transcription factors that bind to regulatory genomic regions. To understand the mechanisms of gene regulation at a systems level, one needs to identify which factors contribute to the regulation of each gene, and under which developmental, physiological or pathological conditions. For this, it is crucial to know which transcription factors interact with which regulatory genomic regions. These interactions can be represented as gene regulatory networks that can provide insight into the design principles of gene control and, thereby, into the mechanisms of organismal development, growth, homeostasis and environmental responses[1].

Several approaches can detect interactions between transcription factors and DNA[1,2]. Transcription factor–centered (protein-to-DNA) methods such as chromatin immunoprecipitation (ChIP) target a transcription factor and determine the genomic regions with which it interacts. Gene-centered (DNA-to-protein) methods such as Y1H assays, in contrast, determine the repertoire of transcription factors that interact with a genomic region of interest. Y1H assays capture interactions in the yeast nucleus, which means that, in contrast to ChIP, interactions that occur in a few cells or under highly specific conditions *in vivo* can be detected more readily. Both approaches have limitations, but arguably their chief problem is achieving the throughput required to generate genome-scale and proteome-scale datasets.

Y1H assays involve two components: 'DNA baits' and 'protein preys' (**Fig. 1a**). Briefly, a DNA bait is placed upstream of two Y1H reporter genes: *LacZ* and *HIS3* (refs. 3,4). Each DNA bait–reporter construct is integrated into a fixed location within the yeast genome to generate 'DNA bait strains', ensuring that the DNA bait is incorporated into yeast chromatin. A plasmid that expresses a protein prey fused to the activation domain of the yeast Gal4 transcription factor (Gal4-AD) is then introduced into the DNA bait strain, and when the protein binds the promoter, the Gal4-AD moiety activates reporter gene expression. Activation of the *HIS3* reporter is visualized by growth on media lacking histidine and containing the competitive inhibitor 3-amino-1,2,4-triazole (3AT), and *LacZ* activation is visualized using a colorimetric assay that converts colorless 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) into a blue compound (Online Methods).

We have previously combined Y1H assays with Gateway cloning and demonstrated that large DNA fragments such as gene promoters can be used as DNA baits[3]. Gateway cloning is based on site-specific recombination and can use existing open reading frame (ORF) and promoter clone resources[5–7]. We have used Gateway-compatible Y1H assays to delineate several medium-scale *C. elegans* regulatory networks[8–11]. Each of these contains 50–100 gene promoters and ~100 transcription factors, and took several years to complete. Although Gateway-compatible Y1H assays increased the throughput of DNA bait generation, they still relied on time-consuming screening of complex cDNA and transcription factor libraries to identify interacting protein preys. This involved extensive colony picking, retesting and sequencing, and provided relatively modest coverage, largely owing to the low abundance of some transcription factors in cDNA libraries and the difficulty of reaching saturation in library-based Y1H screens[12].

[1]Program in Systems Biology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [2]Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [3]Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [4]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [5]Department of Computer Science and Engineering, University of Minnesota–Twin Cities, Minneapolis, Minnesota, USA. Correspondence should be addressed to A.J.M.W. (marian.walhout@umassmed.edu) or J.S.R.-H. (john.reece-hoyes@umassmed.edu).

RECEIVED 20 JUNE; ACCEPTED 19 AUGUST; PUBLISHED ONLINE 30 OCTOBER 2011; DOI:10.1038/NMETH.1748

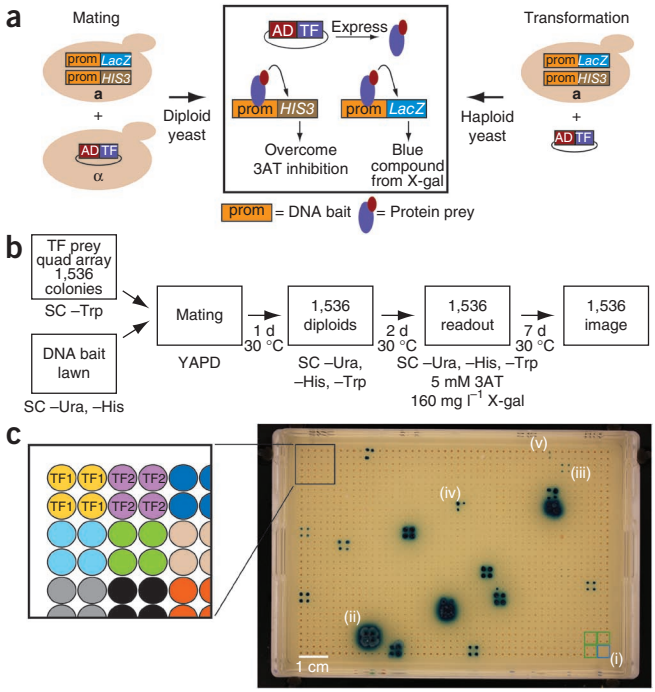

Figure 1 | eY1H assays. (**a**) Schematic of Y1H assays by mating and by transformation. prom, promoter (or DNA bait); AD, Gal4 transcription activation domain; TF, transcription factor. (**b**) Schematic of steps in the eY1H pipeline. SC, synthetic complete medium; –Trp, without tryptophan; –Ura, without uracil; –His, without histidine; YAPD, rich yeast medium with adenine. (**c**) Example of eY1H readout plate ('1,536 image' in **b**). Each transcription factor was tested in quadruplicate. (i) Control quads that lack yeast indicate plate identity and orientation (blue square), whereas yeast that contain an empty AD plasmid serve as a background above which interacting factors are detected (green squares); (ii) strong eY1H positive in which all four spots of a TF quad score positively and had bleed-over; (iii) weak eY1H positive in which all four spots of a TF quad score positively; (iv) medium eY1H positive in which three spots of a TF quad score positively; (v) very weak eY1H positive in which two spots of a TF quad score positively. We considered only quads in which at least two of the four colonies are positive because such an interaction is by definition retested.

An alternative to screening complex libraries is to screen individual transcription factors directly for binding to DNA baits. This can be done either by transforming plasmids encoding the prey proteins into a DNA bait strain or by mating a DNA bait strain with another yeast strain that expresses the prey protein (**Fig. 1a**). Both approaches detect more interacting transcription factors, take less time and reduce cost and effort as compared to screening complex libraries. However, current configurations of mating assays detect only about half as many interactions compared to haploid transformation assays[12].

Here we describe an eY1H pipeline with several modifications to improve performance over previously reported methods (**Table 1** and Online Methods). We have increased coverage by adjusting assay conditions and by mating with a new prey strain carrying a high-copy prey expression vector. We have increased throughput by using a 1,536-colony robotic platform, by standardizing readout analysis and interaction scoring, and by generating a new yeast strain into which DNA bait integration is more efficient. We demonstrate that eY1H assays provide a ~50-fold increase in throughput over traditional library screens and detect more interactions than transformation-based methods. Finally, we introduce 'SpotOn', an automated colony-quantification program that greatly increases the throughput of assessing positive colonies, and hence that of identifying interacting transcription factors. eY1H assays will provide an important tool for the genome-scale delineation of gene regulatory networks in *C. elegans*. As shown in the accompanying papers[13,14], the pipeline can also be applied to the analysis of gene regulation in other model organisms and in humans.

RESULTS

eY1H assays

eY1H assays start with a lawn of a DNA bait strain that is mated with colonies of prey strains expressing transcription factors (**Fig. 1b**). For the prey resource, we generated a collection of 865

Table 1 | Summary of changes in eY1H assays

Change	Improvement	Comments
Modified host yeast strain for DNA bait generation: Y1H-aS2 (*his3-Δ1*)	Increased throughput of DNA bait generation	More efficient double integrations by simultaneously transforming DNA bait–*HIS3* and DNA bait–*LacZ* reporter constructs
New host yeast strain for transcription factor preys: Yα1867	Increased coverage	Out of a panel of different yeast backgrounds, this strain performed best
2μ Gal4-AD Destination prey vector	Increased coverage	The 2μ vector provides higher expression levels of the Gal4-AD transcription factor preys
TF quad arrays	Inherent retest, reduction in false positives and false negatives	We consider only TF quads for which two or more colonies score positively
High-density 1,536-colony format	Increased assay throughput	Only three plates are needed to screen quads for all 865 available *C. elegans* protein preys
Robotic yeast transfers	Increased assay throughput	RoToR HDA robot (Singer Instruments)
Single readout plate to detect *HIS3* and *LacZ* activation	Increased assay throughput, reduction in false positives	Fewer plates are needed, less error-prone than comparing the two readouts
Reduction of 3AT concentration (20 mM to 5 mM)	Increased coverage	Using less 3AT, we detect more interactions in diploids, without an increase in background
Single readout image after incubation for 7 d	Increased assay throughput	This fixed time point captures all detectable interactions for the majority of DNA baits
SpotOn: software for fully automated analysis of assay readout	Increased assay throughput, generation of quantitative interaction data	A list of interacting protein preys for each DNA bait is generated in minutes

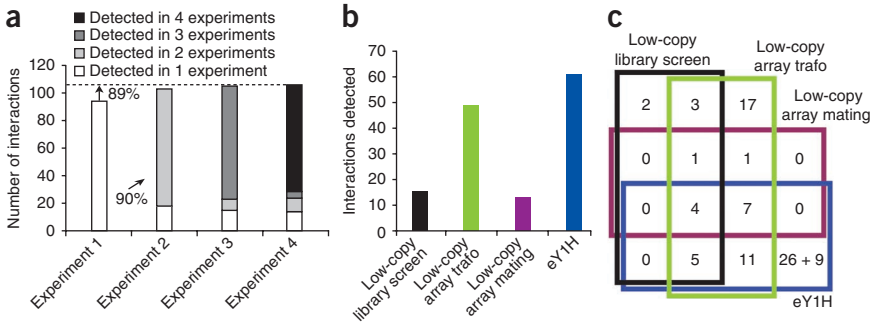

Figure 2 | Sampling sensitivity and reproducibility of eY1H assays. (**a**) Two promoters, *Pvha-15* and *Pcog-1*, were screened four times against the worm transcription factor array; the cumulative number of times an interaction was detected is indicated. In a single experiment we detected 89% of all interactions collectively detected in four experiments. In a single experiment we detected 90% of the interactions detected in a second experiment. (**b**) Bar graph indicating the number of interactions detected using *Pvha-15* as a DNA bait. Trafo, transformation. (**c**) Venn diagram of the interactions graphed in **a**. The label 26 + 9 indicates 26 transcription factors found exclusively by eY1H assays and 9 factors newly detected with clones that were heretofore not available, but that we cloned based on improved gene models.

C. elegans proteins (**Supplementary Table 1**), including 834 transcription factors (89% of the 937 present in transcription factor compendium wTF2.2; **Supplementary Table 2**). It also includes 31 unconventional DNA binding proteins (uDBPs) that can bind DNA but that lack a recognizable DNA binding domain[8] (**Supplementary Table 3**). This prey resource contains 85 additional factors compared to our previous collection[12].

In eY1H assays, each DNA bait-prey combination is tested four times using transcription factor 'quad arrays' (TF quad arrays) that contain each prey in quadruplicate (**Fig. 1b**,**c** and **Supplementary Fig. 1**). This approach provides independent technical replicates, thus reducing both false positives and false negatives. By using a high-density 1,536-colony format[15,16], we covered all available *C. elegans* factors on only three plates. After mating of a DNA bait yeast strain with the TF quad array and selecting for diploids, yeast were transferred to a single readout plate per quad array plate. After 7 d of incubation, a single image was taken per plate, which was manually or computationally evaluated for blue quads. Using a single readout plate halves the number of plates per assay and, importantly, removes the error-prone step of comparing the readouts for *HIS3* and *LacZ* reporter activation that were previously generated independently. Similarly, capturing a single image at one time point for each readout plate increases throughput. Typically, images captured after 7 d of incubation displayed all interactions. For the minority of DNA bait strains that show extremely high or low background reporter expression (**Supplementary Fig. 2**), these fixed conditions may lead to missing interactions, and examining the readout plates at an earlier or later time point, or using a more or less stringent readout plate (for example, 3AT concentration) may be optimal.

eY1H assay sampling sensitivity and reproducibility

The sampling sensitivity of an assay is a measure of how many of the total detectable events one screen will identify, whereas reproducibility is defined as how many events detected by one screen are reproduced in a second[17]. Both parameters are directly linked to the rate of technical false positives and false negatives inherent to the technique. To evaluate these parameters for eY1H assays, we screened two *C. elegans* gene promoters (*Pcog-1* and *Pvha-15*) four times. We selected these baits because they show low background and can be bound by numerous transcription factors[8,9]. eY1H assays were essentially saturated by the third screen (**Fig. 2a**, **Supplementary Figs. 3** and **4**, and **Supplementary Table 4**). Notably, 89% of all interactions were found in a single screen (sampling sensitivity), and 90% of the interactions found in a single screen were reproduced in a second (reproducibility).

Coverage of eY1H assays

To compare the coverage of eY1H assays to previous Y1H methodologies, we focused on *Pvha-15* because we had previously tested this promoter with individual low-copy prey clones by both haploid transformation and mating, as well as in screens of low-copy vector libraries[12]. These previous methods assayed interactions at various concentrations of 3AT for up to 2 weeks, in contrast to the standardized readout conditions in eY1H. We collated the results of the four eY1H screens described above to compare to the previously reported interactions (**Supplementary Table 5**). The eY1H pipeline detected 62 interactions (**Fig. 2b**). Nine of these involved factors that were newly added to the TF quad array and so were tested only with eY1H and cDNA library screens. Of 77 interactions tested by all four methods, 27 were found by eY1H and at least one of the other approaches, 26 interactions were detected exclusively by eY1H assays and 24 were not detected by eY1H assays (**Fig. 2c**).

To determine whether interactions detected exclusively by eY1H assays can occur *in vivo*, we took advantage of ChIP data recently made publicly available for 14 transcription factors (http://modencode.org/) that were detected at least once in the Y1H assays we analyzed (that is, we compared only transcription factors that function in both assay types). Querying the ChIP data revealed nine binding events with the *Pvha-15* DNA fragment (**Supplementary Table 5**), five of which were detected by eY1H assays. This indicates that eY1H can detect *in vivo* binding events. For two additional transcription factors, an interaction with *Pvha-15* was detected by eY1H screens but not by ChIP, which may reflect binding that occurs under *in vivo* conditions not assayed in the ChIP experiments. Four other interactions were detected by ChIP but were not observed in eY1H assays. Aside from being putative ChIP false positives, there are several potential explanations for this that relate to inherent limitations of Y1H. Together, these results indicate that eY1H assays find more interactions for *Pvha-15* than the previous methods combined, and they detect interactions that have *in vivo* support.

Throughput of eY1H assays

We used 50 previously analyzed *C. elegans* DNA bait strains to evaluate the throughput of the eY1H pipeline and to further characterize coverage. We had previously assayed these baits by transforming haploid DNA bait strains with libraries of low-copy

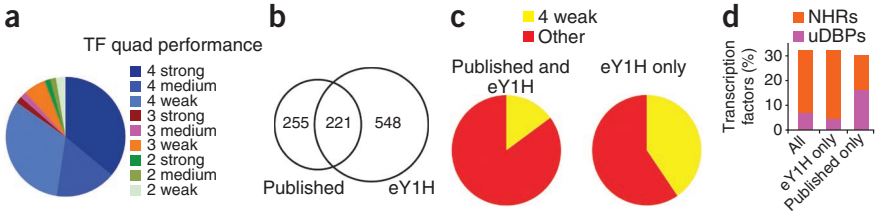

Figure 3 | eY1H assays with 50 previously published *C. elegans* gene promoters as DNA baits. (**a**) Pie chart of TF quad performance in eY1H assays. The number (2, 3 or 4) indicates the number of colonies in a TF quad that were scored positively. We did not consider transcription factors for which only a single colony was scored. (**b**) Venn diagram illustrating overlap between published and eY1H interactions. (**c**) Proportion of interactions detected in which all four colonies scoring positively were weak. (**d**) Percentage NHRs and uDBPs detected in different subsets of the data.

prey vectors, followed by low-throughput directed experiments using individual clones[8–12]. Using eY1H, all 50 DNA bait strains were screened once in a single eY1H batch and the interactions were scored manually, all in 13 d. We detected a total of 769 eY1H interactions (**Supplementary Table 6**) involving 48 DNA baits and 160 transcription factors. We also detected binding of seven uDBPs, further supporting their role as novel DNA-binding proteins. For the vast majority (~85%) of positive TF quads, all four colonies were blue (**Fig. 3a**). Previously, we had reported 476 interactions for these 50 DNA baits (**Supplementary Table 6**). Of these, 221 (46%) were detected by eY1H assays, whereas 548 interactions were exclusive to eY1H (**Fig. 3b** and **Supplementary Table 6**).

There are several reasons why more interactions are detected by eY1H assays (**Supplementary Table 7**). First, transcription factors encoded by long ORFs, or expressed at low levels or in a few cells, may not be represented in cDNA libraries. Second, eY1H assays test each available transcription factor directly and do not depend on library sampling, which results in higher coverage (**Fig. 2a**). Third, because transcription factors are compared directly to a negative control (empty vector, **Fig. 1c**), it is easier to detect weaker interaction phenotypes in eY1H (**Fig. 3c**). Fourth, some transcription factors were uniquely detected in eY1H assays, possibly owing to higher prey expression levels. Finally, some eY1H interactions could be technical false positives. To address this final point, we took advantage of the fact that the DNA baits used are mostly from our metabolic gene regulatory network[11], which is enriched for nuclear hormone receptors (NHRs). The proportion of NHRs detected exclusively by eY1H assays was similar to that in the combined data (**Fig. 3d**), suggesting that eY1H data are of high quality.

There are several possible reasons why some published Y1H interactions were missed in eY1H (**Supplementary Table 7**). First, the sampling sensitivity of a single eY1H screen is 89% (**Fig. 2a**), so we expect to miss 11% of interactions. Second, eY1H assays use fixed assay conditions, and some interactions may be detectable only under different Y1H conditions. Third, some transcription factors may not work well in eY1H assays because of the clone (for example, a full-length variant does not work as well as a DNA binding domain–only clone, or the Gateway tails interfere with binding) or because the assay is performed in diploids (for example, GATA-type factors seem to work better in haploid experiments). Finally, it is possible that some previously published interactions are technical false positives that cannot be repeated. In support of this, interactions with the highest scores in a system we have previously used to classify interactions[9] were more likely to be reproduced in eY1H assays (data not shown). Similarly, compared with interactions detected by eY1H, a higher proportion of 'published only' interactions involved uDBPs (**Fig. 3d**), and these may be more likely to be incorrect. Overall, processing of 50 baits in eY1H assays detected 60% more interactions (769 compared with 476) in one-fiftieth of the time.

SpotOn: a program for automated eY1H quantification

For the eY1H assays performed in this study, all readout plates were scored manually. However, the eY1H methodology is designed for screening hundreds or thousands of DNA baits in large-scale projects for which manual scoring will be too time-consuming and error-prone. We developed a custom Perl-based program called SpotOn that automatically quantifies eY1H assay results (Online Methods). SpotOn imports a JPEG image of an eY1H readout plate (**Fig. 4a**) and performs the following tasks. First, a grid is fitted to the 1,536 colonies to associate each colony with the transcription factor it expresses (**Supplementary Fig. 5**). Second, the intensity of β-galactosidase expression (that is, the blueness) of each colony is determined (**Fig. 4b**). Third, these intensity values are normalized for noise owing to growth differences resulting from uneven nutrient availability within

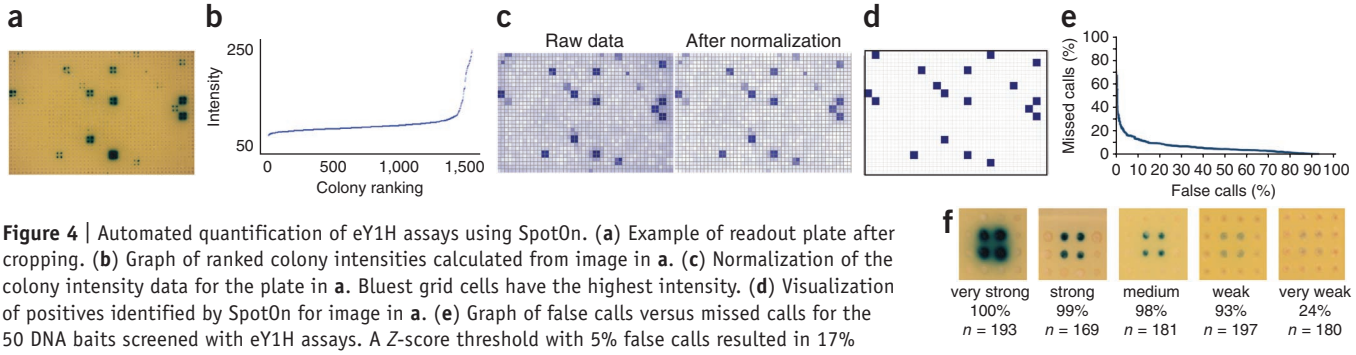

Figure 4 | Automated quantification of eY1H assays using SpotOn. (**a**) Example of readout plate after cropping. (**b**) Graph of ranked colony intensities calculated from image in **a**. (**c**) Normalization of the colony intensity data for the plate in **a**. Bluest grid cells have the highest intensity. (**d**) Visualization of positives identified by SpotOn for image in **a**. (**e**) Graph of false calls versus missed calls for the 50 DNA baits screened with eY1H assays. A *Z*-score threshold with 5% false calls resulted in 17% missed calls. With limited manual curation, the false-call rate was reduced to 1%. (**f**) Missed calls in SpotOn mostly correspond to very weak interactions that are barely detectable by eye. Percentage indicates proportion correctly found by SpotOn; *n* indicates the total number of interactions in each category.

each plate, as well as intrinsic differences between baits (**Fig. 4c** and **Supplementary Fig. 6**). Then, SpotOn uses a *Z*-score cutoff to identify positive colonies and removes false positives arising from bleed-over of blue compound from neighboring very strong positives (**Fig. 1c**). Finally, it identifies transcription factors for which two or more colonies score positively to produce a list of eY1H interactions (**Fig. 4d**).

We used the eY1H data from the 50 DNA bait strains assayed above to benchmark the performance of SpotOn (**Fig. 4e** and **Supplementary Table 6**). At a 5% false call rate, SpotOn detected 83% of the manual calls (that is, the false-negative rate was ~17%; **Fig. 4e**). Notably, the majority of missed manual calls had been scored as 'very weak' (**Fig. 4f**). False calls typically arose because of bleed-over from strong positives (**Fig. 1c**). By manually checking all positives that occurred next to highly blue quads (~20% of calls in this dataset; **Supplementary Table 8**), we eliminated this class of false calls, reducing the false-call rate from 5% to 1%. Individual users can tailor the SpotOn parameters to optimize the trade-off between sensitivity and specificity.

Another potential advantage of SpotOn is that generating quantitative eY1H data enables identification of transcription factors that are positive at different thresholds (that is, higher or lower *Z*-scores) or determination of average reporter activation for a given transcription factor across all DNA baits that it binds. SpotOn can also be applied to other plate-based assays such as Y2H (data not shown).

DISCUSSION

Understanding gene regulation at a systems level will require genome-scale datasets of DNA–transcription factor interactions for tens of thousands of regulatory genomic elements. eY1H assays provide a tremendous increase in throughput and performance compared to other types of Y1H assays and can be used to analyze *C. elegans* gene promoters as shown here as well as regulatory sequences in *Arabidopsis thaliana*[14] and *Homo sapiens*[13]. Although we developed mating-based assays, others have developed a transformation-based robotic Y1H platform[18]. Transformation-based assays provide high coverage[12] but are relatively cumbersome, as they require repeated preparations of prey DNA and handling of highly viscous solutions. Nonetheless, the availability of different Y1H approaches provides the research community with a selection of tools.

The individual technical advances compiled in the eY1H pipeline can also be applied in a modular fashion to suit different needs. For example, the high-coverage eY1H vectors and strains can be used without a robot; instead, mating or transformations can be performed in 96- or 384-colony formats, and yeast can be transferred with commonly used replica-plating or hand-held pinning tools. Similarly, this pipeline can be modified for use with any mating-based assay, including Y2H assays (data not shown).

Using library screens, it took about two years to identify interactions with confidence for 50 baits, whereas using the eY1H pipeline it took 13 d. Thus, eY1H assays are at least 50 times faster than library screens and could be configured for even higher throughput. For instance, the eY1H pipeline can be arranged to process staggered batches of bait strains whose assays are started on different days. Throughput can also increase with greater batch sizes and through use of multiple robots.

Although eY1H assays detect more interactions than any previous Y1H approach, they do not detect all the interactions previously found. It has been demonstrated with Y2H assays that detection of some interactions is specific to certain vector or yeast combinations and that assay coverage can be further improved with the use of different configurations of vectors (for example, C-terminal fusion of the activation domain to the prey)[19,20]. If coverage is paramount, the eY1H pipeline can easily incorporate such modifications, especially for preys that do not seem to function in the traditional vector that uses an N-terminal fusion.

Some interactions previously detected by ChIP were also not seen by eY1H. Apart from potential previous false positives, explanations for this include: (i) ChIP interactions may involve isoforms not present in the transcription factor clone collection; (ii) interactions may occur in the distal portion of the promoter, which may be too far from the transcription start site to confer efficient yeast reporter activation; (iii) the transcription factor may require post-translational modifications or cofactors not available in yeast for binding that specific promoter; (iv) the chromatin context in the yeast nucleus may preclude the detection of some interactions; and (v) the interaction detected by ChIP may be indirect.

We have previously shown that many different approaches can be used to validate interactions retrieved by Y1H assays *in vivo*. For instance, in *C. elegans*, one can use transgenic animals that express GFP under the control of the DNA bait and use genetic approaches to examine GFP expression in the presence or absence of an interacting transcription factor[8,10,11,21,22]. Similarly, expression of the endogenous gene may be examined after transcription factor RNAi or mutation. However, it is important to realize that a lack of validation does not necessarily invalidate an interaction, because validation assays have their own limitations and because the loss of a transcription factor can be masked by the redundant activity of another[23].

eY1H assays provide a tool for the gene-centered mapping of gene regulatory networks in model organisms, both for large-scale genome-wide studies and small-scale in-depth dissection of a single promoter region. This approach should also be applicable to additional systems for which genome sequences, transcription factor annotations and clones become available.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank members of the Walhout laboratory for discussions and critical reading of the manuscript, S. Lee for preparing media, K. Salehi-Ashtiani (Center for Cancer Systems Biology) for transcription factor Entry clones, J. Boeke for advice on the creation of Y1H-aS2, and C. Boone for general advice on the robotic pipeline and the use of different yeast strains. This work was supported by US National Institutes of Health (NIH) grant GM082971 to A.J.M.W. Research in the Dekker laboratory is supported by NIH grant HG003143 and a W.M. Keck Foundation Distinguished Young scholar award. C.L.M. and C.P. are supported by NIH grant HG005084 and US National Science Foundation grant DBI 0953881.

AUTHOR CONTRIBUTIONS

J.S.R.-H. and A.J.M.W. conceived the project; A.K. and S.S. performed the eY1H assays; A.D. created the tools for automated eY1H assay quantification in collaboration with B.L., C.P., J.D., J.S.R.-H. and C.L.M.; S.K. cloned additional transcription factor–encoding ORFs; J.S.R.-H. and A.J.M.W. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturemethods/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Walhout, A.J.M. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Res.* **16**, 1445–1454 (2006).
2. Arda, H.E. & Walhout, A.J.M. Gene-centered regulatory networks. *Brief. Funct. Genomics* **9**, 4–12 (2009).
3. Deplancke, B., Dupuy, D., Vidal, M. & Walhout, A.J.M. A gateway-compatible yeast one-hybrid system. *Genome Res.* **14**, 2093–2101 (2004).
4. Deplancke, B., Vermeirssen, V., Arda, H.E., Martinez, N.J. & Walhout, A.J.M. Gateway-compatible yeast one-hybrid screens. *Cold Spring Harb. Protoc.* doi:10.1101/pdb.prot4590 (2006).
5. Walhout, A.J.M. *et al.* GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* **328**, 575–592 (2000).
6. Lamesch, P. *et al. C. elegans* ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions. *Genome Res.* **14**, 2064–2069 (2004).
7. Dupuy, D. *et al.* A first version of the *Caenorhabditis elegans* promoterome. *Genome Res.* **14**, 2169–2175 (2004).
8. Deplancke, B. *et al.* A gene-centered *C. elegans* protein-DNA interaction network. *Cell* **125**, 1193–1205 (2006).
9. Vermeirssen, V. *et al.* Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein-DNA interaction network. *Genome Res.* **17**, 1061–1071 (2007).
10. Martinez, N.J. *et al.* A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes Dev.* **22**, 2535–2549 (2008).
11. Arda, H.E. *et al.* Functional modularity of nuclear hormone receptors in a *C. elegans* gene regulatory network. *Mol. Syst. Biol.* **6**, 367 (2010).
12. Vermeirssen, V. *et al.* Matrix and Steiner-triple-system smart pooling assays for high-performance transcription regulatory network mapping. *Nat. Methods* **4**, 659–664 (2007).
13. Reece-Hoyes, J.S. *et al.* Yeast one-hybrid assays for gene-centered human gene regulatory network mapping. *Nat. Methods* doi:10.1038/nmeth.1764 (this issue).
14. Gaudinier, A. *et al.* Enhanced Y1H assays for *Arabidopsis*. *Nat. Methods* doi:10.1038/nmeth.1750 (this issue).
15. Tong, A.H.Y. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
16. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
17. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nat. Methods* **6**, 83–90 (2009).
18. Hens, K. *et al.* Automated protein-DNA interaction screening of *Drosophila* regulatory elements. *Nat. Methods* doi:10.1038/nmeth.1763 (30 October 2011).
19. Braun, P. *et al.* An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6**, 91–97 (2009).
20. Chen, Y.C., Rajagopala, S.V., Stellberger, T. & Uetz, P. Exhaustive benchmarking of the yeast two-hybrid system. *Nat. Methods* **7**, 667–668 (2010).
21. Ow, M.C. *et al.* The FLYWCH transcription factors FLH-1, FLH-2 and FLH-3 repress embryonic expression of microRNA genes in *C. elegans*. *Genes Dev.* **22**, 2520–2534 (2008).
22. Reece-Hoyes, J.S. *et al.* The *C. elegans* Snail homolog CES-1 can activate gene expression *in vivo* and share targets with bHLH transcription factors. *Nucleic Acids Res.* **37**, 3689–3698 (2009).
23. Walhout, A.J.M. What does biologically meaningful mean? A perspective on gene regulatory network validation. *Genome Biol.* **12**, 109 (2011).

ONLINE METHODS

Resource availability. All yeast strains, DNA bait constructs, transcription factor constructs and the code for SpotOn are available upon request.

Yeast one-hybrid assays. Y1H assays involve two components: DNA baits and protein preys (**Fig. 1a**). Briefly, a DNA bait is placed upstream of two Y1H reporter genes: *LacZ* and *HIS3* (refs. 3,4). Each DNA bait–reporter construct is integrated into a fixed location within the yeast genome to generate DNA bait strains, ensuring that the DNA bait is incorporated into yeast chromatin (that is, the assay does not use 'naked' DNA). We generally use genomic fragments between 300 base pairs (bp) and 2 kbp[4]. A plasmid that expresses a protein prey fused to the activation domain of the yeast Gal4 transcription factor (Gal4-AD) is then introduced into the DNA bait strain, and when the protein binds the promoter, the Gal4-AD moiety activates reporter gene expression. Because a heterologous activation domain is used, Y1H assays can detect physical interactions involving both transcriptional activators and repressors. The *LacZ* reporter encodes β-galactosidase, which generates a blue compound from the colorless X-gal, whereas the *HIS3* reporter expression product permits growth on medium lacking histidine and containing the competitive His3 inhibitor 3AT. The readout of the assay is therefore the ability of yeast to grow in the presence of 3AT or turn blue, with yeast able to do both termed double-positives. Traditionally, the two reporters are analyzed separately and the results combined. Interactions detected by double-positive yeast are regarded with higher confidence than those detected by single-positive yeast that activate only one reporter, because the physical interaction occurs twice within the same yeast nucleus[4].

Updated *C. elegans* transcription factor compendium (wTF2.2) and uDBP list. Previous iterations of the *C. elegans* transcription factor compendium contained 934 (wTF2.0)[24] and 940 (wTF2.1)[12] predicted transcription factors, respectively. The updated wTF2.2 compendium contains 937 predicted *C. elegans* transcription factors (**Supplementary Table 2**). Nineteen transcription factors have been removed from wTF2.1 (ref. 12) and 16 new transcription factors have been added. The majority of these changes are due to updates in gene model annotations (wTF2.2 is based on WS190). Several genes have been added owing to recent reports of sequence-specific DNA-binding ability of their protein products[25–27].

Using Y1H cDNA library screens, we have previously retrieved multiple *C. elegans* proteins that do not have a recognizable DNA-binding domain[8–11]. We refer to these as 'novel putative transcription factors'. Using protein arrays, multiple human proteins have also been uncovered that do not possess a known DNA-binding domain but that do bind DNA specifically, and these are referred to as uDBPs[28]. Here we adopt this nomenclature for 32 *C. elegans* proteins (**Supplementary Table 3**) that include those found in Y1H assays as well as two that have been annotated in the literature as regulating gene expression through specific DNA binding[25,26].

Generating the wTF2.2 clone array. The starting point for the wTF2.2 clone array was the wTF2.1 resource[12]. We added Entry clones for 85 newly cloned transcription factors to our collection (**Supplementary Table 1**). Seven were recent additions to our list of transcription factors; existing clones were picked from the ORFeome[6], and their identity was verified by sequencing. The remaining 78 ORFs were generated using primers designed with updated gene models that were largely based on recent RACE data[29]. Twenty-seven Entry clones were a kind gift from K. Salehi-Ashtiani (Center for Cancer Systems Biology), and the others were generated in-house. Overall, the wTF2.2 array contains 834 of the 937 factors (89%) in wTF2.2, as well as 31 uDBPs. The ORFs from all Entry clones in our collection were transferred into pDEST-AD-2μ (Invitrogen) by a Gateway LR reaction (Invitrogen). The resulting AD-prey Destination clones were transformed into Yα1867 in 96-well format with empty pDEST-AD-2μ in the H12 position. We generated frozen glycerol stocks of the resulting transcription factor prey yeast strains by transferring a small amount (half a match-head) of yeast to 1 ml liquid medium without tryptophan (–Trp) synthetic complete (SC) in 96-well deep-well (2 ml) plates, incubating in an orbital shaker (Multifors, 200 r.p.m., 30 °C) for 48 h, pelleting the yeast (2,000*g*, 5 min), discarding the supernatant, resuspending each pellet in 200 μl 15% (v/v) glycerol and transferring each yeast suspension to 96-well plates stored at −80 °C. pDEST-AD-2μ is a high-copy vector, as it contains a 2μ origin of replication that results in 50–100 copies per cell, whereas the low-copy vector (pDEST-22) has an ARS/CEN origin that results in only one or two copies per cell[30]. Both pDEST-AD-2μ and pDEST-22 use the full-length *ADH1* promoter to drive expression of the AD-prey fusion; therefore, a higher vector copy number results in greater expression. Although it is possible that expression of these AD-prey fusions might adversely affect the yeast, no difference was observed in mating ability or growth of the haploid prey strains, or growth of resulting diploids (data not shown).

Preparing 1,536-colony quad arrays. The RoToR HDA robot (Singer Instruments) was used for every step in the generation of 1,536-colony AD-wTF array plates. The process is outlined in **Supplementary Figure 1**. The RoToR HDA uses disposable plastic pads with 96, 384 or 1,536 pins to precisely transfer yeast between solid agar plates or between a liquid source (for example, yeast suspension) and a solid agar plate. Plates used on the robot require extra care to ensure that the agar surface is flat for efficient transfer of colonies using pads. We poured 65 ml media into each Singer Plusplate (prepared as stacks of no more than five), dried the plates overnight at room temperature on a flat surface and stored them at 4 °C in airtight plastic bags. To build a 1,536-colony quad array, we first generated 96-colony plates of transcription factor prey yeast by spotting from 96-well plates of glycerol stocks onto solid agar plates. A single transfer from each of four separate 96-colony plates was then used to build a 384-colony plate in which each transcription factor prey yeast strain was present once. Finally, four transfers from the same 384-colony plate were used to create the 1,536-colony plate, in which each transcription factor prey yeast was present in quadruplicate. Transcription factor prey yeast strains were grown on SC –Trp plates with incubation steps of 2 d at 30 °C after each transfer, resulting in yeast colonies of ~3 mm, ~2 mm and ~1 mm on 96-, 384- and 1,536-spot plates, respectively. Our current collection of *C. elegans* transcription factor prey yeast

strains fills eleven 96-well plates (**Supplementary Table 1**), so a total of three 1,536-colony plates cover all available transcription factors and uDBPs: one using 96-well plates 1 to 4, a second using 96-well plates 5 to 8 and a third using 96-well plates 9 to 11. Four of the 16 positions in the bottom right corner of each quad array were intentionally left without yeast so that plate orientation and identity could be verified (a different four colonies were omitted from each of the three quad array plates). The remaining 12 positions in the bottom right corner contained empty pDEST-AD-2μ and served as negative controls in eY1H assays. The quad array was transferred every week to fresh SC −Trp plates and stored at room temperature when not in use. New TF quad array plates were generated from glycerol stocks every 8 weeks.

Performing eY1H assays. The eY1H pipeline takes 6 d to generate diploid yeast from the haploid DNA bait and transcription factor prey strains, and typically another 7 d to assay the diploids for reporter gene expression. This is outlined in **Figure 1**. The RoToR HDA was used at every step unless specified. Between each step described below, the plates were incubated at 30 °C. On day 1, fresh copies of the quad arrays were generated using transfers from older copies to SC −Trp plates, and a lawn of the DNA bait strain was generated on either a rich yeast medium with adenine (YAPD) or SC without uracil or histidine (SC −Ura −His) plate (yeast were mixed in sterile water, and the suspension was spread onto solid agar plates using sterile glass beads). One plate of transcription factor prey yeast provided enough yeast to set up mating plates for four DNA baits. On day 3, we prepared a mating plate for each of the transcription factor prey strain plates by first transferring transcription factor prey yeast strains to a YAPD plate and then using a 1,536-pin pad to collect DNA bait yeast from the lawn and place this DNA bait yeast on top of the transcription factor prey yeast already on the YAPD plate. On day 4, yeast were transferred from the mating plate to an SC −Ura −His −Trp plate upon which only diploid yeast that contain both a DNA bait and a transcription factor prey can grow. On day 6, diploid yeast were transferred from the SC −Ura −His −Trp plate to the eY1H 'readout' plate (SC −Ura −His −Trp plus 5 mM 3AT, 160 mg l^{-1} X-gal, 26 mM disodium hydrogen phosphate, 25 mM sodium dihydrogen phosphate, pH 7.0), upon which only yeast that express enough His3 to overcome inhibition by 3AT will grow, and only yeast that grow and express β-galactosidase from the *LacZ* reporter gene will metabolize X-gal into a blue compound. Thus, activation of both reporters was analyzed in the same media plate. All blue colonies that grew on readout plates were considered double-positives, even though some positives appeared not to be larger than the negative background colonies. The color of a colony is a more reliable indicator of an interaction than colony size because yeast growth is somewhat inhibited at pH 7.0 compared to the usual pH 5.9 of yeast media (data not shown). An image of each readout plate was captured using the spImager-M (S&P Robotics), which places each plate in a uniformly lit environment where a mounted Canon Rebel XSi digital camera with a EF-S 60-mm F/2.8 macro lens controlled by spImager version 1.0.2.0 software takes a high-resolution photograph that is converted to a 4,272 pixel × 2,848 pixel JPEG image and stored. In projects that involve processing few or highly autoactive DNA baits (**Supplementary Fig. 2**), readout plates can be observed and photographed daily. However, when processing large numbers of DNA baits we typically photograph the readout plates just once after 7 d of incubation.

The new prey strain Yα1867. Performing Y2H assays with different strains improves coverage[19]. We tested eight yeast strains (SY3002, Y287, Y1495 and Y1864 of mating type **a**, and SY3003, Y288, Y1494 and Y1867 of mating type α), and our original strains YM4271 (ref. 4) and Y1Hα001 (ref. 12) in all 25 pairwise combinations in mating-based Y1H assays. All new strains had mutations in the *URA3*, *HIS3* and *TRP1* genes, and were kind gifts from C. Boone (University of Toronto). For the Y1H assays we integrated *Pvha-15*–reporter constructs into the bait strains and transformed into the prey strains 21 high-copy transcription factor clones corresponding to high-confidence interactors with this DNA fragment[12]. In addition to the number of interactions detected from each mating combination, we took into account the mating compatibility of each pair and the general phenotype of both haploid and diploid strains (for example, some yeasts were 'waxy' and difficult to transfer, some grew too fast or slow, and some were more resistant to 3AT). The best combination was Yα1867 (*MATα SUC2 gal2 mal mel flo1 flo8-1 hap1 ho bio1 bio6 ura3-52 ade2-101 trp1-901 his3-Δ200*) as the host prey strain and YM4271 (ref. 4) as the host bait strain (data not shown).

Creating the Y1H-aS2 DNA bait yeast strain. To generate DNA bait strains for Y1H assays, we integrated the two DNA bait–reporter constructs into different mutant loci within the genome of a host yeast strain[3]. Previously, we have used the yeast strain YM4271 (ref. 4) with the pMW#3 *LacZ* construct integrated at the *URA3* locus, rescuing the *ura3-52* Ty insertion that disrupts the gene[31], and the pMW#2 *HIS3* construct integrated at the *HIS3* locus, rescuing the *his3-Δ200* deletion that includes the entire ORF[32]. These integration events are mediated by DNA sequences shared by the reporter constructs and the yeast genome. Because pMW#2 shares only 463 bp with the genome of YM4271, whereas pMW#3 shares 1,090 bp, the integration success rate for pMW#2 is much lower than that for pMW#3 (~100 and ~2,000 events per μg linear DNA, respectively; data not shown). This lower integration rate for pMW#2 is the limiting factor when performing double integrations (that is, with both DNA bait–reporter plasmids simultaneously) with YM4271. We reasoned that increasing the amount of DNA sequence shared by pMW#2 and the yeast genome would increase the integration rate. To this end, we created Y1H-aS2 by swapping the *his3-Δ200* locus (1,040-bp deletion) in YM4271 with *his3-Δ1* (190-bp deletion[33]). This involved two yeast transformations performed using a standard protocol[34]. We first replaced the *his3-Δ200* locus of the YM4271 strain with a wild-type *HIS3* gene by transforming with BamHI-digested pPL97 (which contains *HIS3*) and selecting colonies able to grow in the absence of histidine (SC −His). A resulting $HIS3^+$ strain was then transformed with XhoI-digested pNN132 (which contains a *URA3* gene flanked by a wild-type *HIS3* gene and a *his3-Δ1* gene) and colonies able to grow in the absence of both histidine and uracil (SC −Ura −His) were selected. A $HIS3^+$,$URA3^+$ strain was grown in YAPD liquid overnight and plated on SC–5-FOA (SC with 0.1% (w/v) 5-fluo-orotic acid) plates. Yeast able to grow in the presence of 5-FOA have lost the *URA3* gene owing to internal recombination between the *HIS3* genes, but they have an equal chance of maintaining either the wild-type or

doi:10.1038/nmeth.1748

the mutant gene. Therefore, colonies that grew on SC–5-FOA were streaked to YAPD as well as SC –Ura and SC –His media to identify yeast unable to grow in the absence of both uracil and histidine. Three independent strains unable to grow in the absence of histidine or uracil were used to generate *Pvha-15* DNA bait strains and screened in eY1H assays. The interactions observed with all three strains were the same as one another and also the same as those observed with the YM4271 *Pvha-15* DNA bait strain (data not shown). One of the initial strains unable to grow in the absence of histidine or uracil was renamed Y1H-aS2. We observed no obvious phenotypic difference between the YM4271 and Y1H-aS2 strains, including the ability to be transformed by transcription factor prey vectors or mate with transcription factor prey yeast (data not shown). Integration rates at the *HIS3* locus increased from ~100 events per μg transformed linear vector for the YM4271 strain to ~4,000 for Y1H-aS2. Accordingly, the rate of double integration increased tenfold (data not shown). All yeast strains were genotyped and sequenced at each step using combinations of the primers F1 and R2, which flank the *HIS3* locus, and the primer R1, which anneals within the wild-type *HIS3* and *his3-Δ1* loci but not within *his3-Δ200* (**Supplementary Table 9**). Both pPL97 and pNN132 were gifts from J. Boeke (Johns Hopkins University).

SpotOn: a program for automated colony quantification. Automated eY1H assay quantification involves three major steps: (i) drawing a grid to fit the quad array so that each colony can be associated with the transcription factor it expresses, (ii) identifying which colonies are 'positive' (that is, which show substantially more reporter expression than background) and (iii) removing systematic false calls to create a final list of eY1H DNA bait–transcription factor interactions.

Drawing the grid. Before the grid is drawn, all objects on the plate need to be identified, and then those that are not yeast colonies need to be removed. eY1H readout JPEG images are cropped to remove the outer edge of the plastic plate and converted to PNG files (**Supplementary Fig. 5a**). Color intensity is extracted for the red, green and blue channels of every pixel in the image using the getPixel and rgb methods from the Perl GD library (http://libgd.org/). To compensate for local effects that arise from uneven distribution of media and nutrients, each image is corrected by local median normalization (LMN) as follows. For each pixel, the median intensity of the 80 neighboring (9×9 square) pixels is calculated. The original intensity value is divided by this median value to get an LMN factor, and the original intensity value is then multiplied by the LMN factor. LMN is performed for all three color channels (**Supplementary Fig. 5b**). To detect all objects in the image, the CIE-76 algorithm is used to calculate distances between pixels in terms of color, and this distance information is then used to separate the pixels into two groups (colony and background) using k-means clustering ($k = 2$). A flood fill algorithm is then applied to each pixel in the two resulting clusters, which recursively searches neighboring pixels for those belonging to the same k-means cluster, thus detecting all non-adjoined objects (**Supplementary Fig. 5c**). Sizes of all objects are then calculated, and it is assumed that the object of the largest size is the background agar of the plate, which is then removed from any further analysis.

The remaining objects are subjected to noise reduction so that any object that is not a yeast colony is removed. This involves an initial 'circularization' step followed by size filtration. As colonies are circular, their pixels will be densely clustered around a central core, so removing rough edges from (that is, circularizing) all objects substantially affects only noncircular objects that are probably not colonies. Circularizing is performed by removal of all pixels for which less than 18 of the surrounding 24 pixels (5×5 square) are part of the same object. After circularization, the colony size is calculated and the size-filtration step removes any objects that are outside a specified limit (40 to 150 pixels in this study; **Supplementary Fig. 5d**).

All remaining objects should represent yeast colonies and are used to draw the grid. The (x, y) center of each object is first defined as the average x- and y-coordinates, respectively, of all pixels contained in the object. The coordinates of each object center are then clustered into $n \times m$ groups, where n is the number of rows (here 32) and m is the number of columns (here 48) within the original plate. A hierarchical clustering algorithm with a defined endpoint of 1,536 clusters is then used for detection of all rows and columns. Once the object centers have been clustered and grounded into 32 rows and 48 columns, lines drawn through these centers can be applied to the image (**Supplementary Fig. 5e**) and used to calculate the final grid lines as the distance between center i and center $(i - 1)$. A major benefit of this method is that even if only ~25% of all objects within each row or column are detected, the complete grid can still be drawn. This is important for this application because the above filtering steps remove a substantial number of true colonies. Finally, the grid lines are applied to the original cropped image and transcription factor identities are assigned to each grid cell using the coordinates from the 96-well plates used to store the prey yeast collection (**Supplementary Fig. 5f**).

Although SpotOn is fully automated for grid drawing from eY1H images, in a small number of cases (6 of the 150 plates in this study) it is not able to generate a grid. SpotOn informs the user which plates cannot be processed. To score the interactions from these images, we developed a Matlab-based interface that allows the user to visually inspect each image and define the grid by manually clicking the corner colonies. The resulting grid coordinates are then saved and the rest of the image processing is performed by the SpotOn pipeline. Although this manual grid drawing may be used for all images, we advise its use only for 'difficult' plates as it is more time-consuming.

Identifying positive colonies. This process involves determining to what extent the reporters are expressed in each colony, followed by whether a colony shows substantially more reporter expression than the background. First, each grid cell is examined to identify the colony. Using red channel intensity values, each colony is first distinguished from the surrounding agar using k-means clustering ($k = 2$) to separate the pixels within each grid cell into two clusters. Note that this clustering uses raw pixel-intensity values rather than the LMN-corrected intensities. The most central object relative to the grid cell is the colony, and the remaining object is the background media. For every colony, the mean, median and s.d. of intensities for each color channel for all the pixels that make up the colony are calculated, as well as the size of the colony. We found that the median red channel intensity for each colony is the most robust and representative measure for

the quantification of reporter expression. Hereafter we use the term 'colony intensity' for that median red channel value.

Before it is determined whether a colony is positive, two normalization steps are performed. Row-column normalization (RCN) neutralizes the effect of uneven media and nutrient distribution within each plate, whereas bait-to-bait normalization (BTBN) takes into account the fact that each DNA bait strain shows slightly different background levels of reporter expression. RCN is applied as follows. For each row or column, the median of the colony intensities is determined, and the median of these row or column values is then calculated for the whole plate. An RCN factor is derived for each row or column by dividing the individual row or column median by the plate row or column median value. Each colony intensity is then multiplied by the RCN factors for its row and column location in the grid. The resulting colony intensities are then subjected to BTBN as follows. The dataset was divided into three groups according to the three TF quad array plates (named '1 to 4', '5 to 8' and '9 to 12'), and each group is processed separately. The median of the colony intensities from each plate within the group is calculated, and the median of all of those medians is obtained. For each plate, the median of the plate medians is divided by the individual plate median to obtain a BTBN factor, and every colony intensity from that plate is then multiplied by this BTBN factor. These final normalized colony intensities are then used to determine which colonies are positive (**Supplementary Fig. 6**). A mean and s.d. is calculated using all colony intensities from all plates, and this is used to derive a *Z*-score for each colony. All colonies with a *Z*-score above a certain cutoff (selected by the user; in this study we used 1.32) are then deemed eY1H positives (**Supplementary Fig. 5g**).

Removing false calls and listing interacting transcription factors. For a transcription factor to be counted as an interactor, at least two of the four colonies in a TF quad must be positive. In our experience, the occurrence of spurious blue singletons, which are probably false positives, was infrequent (data not shown). We observed two noteworthy situations in which transcription factors were falsely identified as interactors, both caused by colonies of a quad expressing very high levels of the reporters and bleeding over into neighboring grid cells, resulting in colonies of neighboring quads (and thus their associated transcription factors) being incorrectly assessed as positive. In the first, more common situation, only the two colonies closest to these strongly expressing 'bleed-over quads' are affected. To mitigate this issue, SpotOn first identifies putative bleed-over quads in which at least three colonies have a raw intensity of ≥200 (we empirically determined this cutoff from the relevant quads; data not shown). Any neighboring colonies are then automatically removed from the positive list generated in step 2 (**Supplementary Fig. 5h**). We are aware that removing these types of false calls may result in false negatives (for example, the neighboring quad was a true interactor for which only two or three colonies were positive). However, in the majority of cases, all four colonies in a quad are positive (**Fig. 3a**), and so the two distal colonies will remain positive and the transcription factor will still be counted as an interactor. The second type of false call caused by bleed-over occurs where bleed-over has affected all four colonies of a neighboring quad. However, ignoring all quads that are adjacent to bleed-over quads removes too many true positives, so instead SpotOn flags these positive quads within the interaction list, and the user can choose to view them and manually edit the interaction list accordingly.

A list of eY1H interactions determined by SpotOn (**Supplementary Table 8**) is generated after removal of the colonies adjacent to bleed-over quads. This list includes information about the DNA bait, the interacting transcription factor, the number of colonies in each quad that scored positively, and means of the *Z*-scores and colony intensities for the positive colonies in each quad. The list also indicates which interactions are from quads adjacent to potential bleed-over quads. In this study we benchmarked the performance of SpotOn using an interaction list generated manually (**Supplementary Table 6**). Using a *Z*-score cutoff of 1.32, 5% of the calls were false and 17% of the manual calls were missed. Notably, the majority of these missed calls showed a very weak Y1H phenotype (**Fig. 4f**). The user may be content with this false call rate; however, by manually checking the interactions marked as neighbors of bleed-over quads (~20% of this dataset) and retaining only the true positives, we reduced the false call rate to 1.2%. A further option for the user to reduce the false call rate is to use a higher *Z*-score cutoff, but at the cutoff that would result in a 1% false call rate, 33% of true calls would be missed (**Fig. 4e**).

Comments. SpotOn is designed to be flexible enough to process other colony readout systems. The software can be customized for various image formats. Different grid sizes can be generated and alternative prey-identity coordinate files can be uploaded. Analysis with any of the measured variables is possible, including the three color channels and colony size.

24. Reece-Hoyes, J.S. *et al.* A compendium of *C. elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol.* **6**, R110 (2005).
25. Lesch, B.J., Gehrke, A.R., Bulyk, M.L. & Bargmann, C.I. Transcriptional regulation and stabilization of left-right neuronal identity in *C. elegans*. *Genes Dev.* **23**, 345–358 (2009).
26. Witze, E.S., Field, E.D., Hunt, D.F. & Rothman, J.H. *C. elegans* pur alpha, an activator of end-1 synergizes with the Wnt pathway to specify endoderm. *Dev. Biol.* **327**, 12–23 (2009).
27. Tabuchi, T. *et al.* Chromosome-biased binding and gene regulation by the *Caenorhabditis elegans* DRM complex. *PLoS Genet.* **7**, e1002074 (2011).
28. Hu, S. *et al.* Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* **139**, 610–622 (2009).
29. Salehi-Ashtiani, K. *et al.* Large-scale RACE approach for proactive experimental definition of *C. elegans* ORFeome. *Genome Res.* **19**, 2334–2342 (2009).
30. Rajagopala, S.V., Hughes, K.T. & Uetz, P. Benchmarking yeast two-hybrid systems using the interactions of bacterial motility proteins. *Proteomics* **9**, 5296–5302 (2009).
31. Rose, M. & Winston, F. Identification of a Ty insertion within the coding sequence of the *S. cerevisiae URA3* gene. *Mol. Gen. Genet.* **193**, 557–560 (1984).
32. Struhl, K. Nucleotide sequence and transcriptional mapping of the yeast pet56-his3-ded1 gene region. *Nucleic Acids Res.* **13**, 8587–8601 (1985).
33. Scherer, S. & Davis, R.W. Replacement of chromosome segments with altered DNA sequences constructed *in vitro*. *Proc. Natl. Acad. Sci. USA* **76**, 4951–4955 (1979).
34. Walhout, A.J.M. & Vidal, M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* **24**, 297–306 (2001).

doi:10.1038/nmeth.1748