



# HHS Public Access

Author manuscript

*Nat Methods*. Author manuscript; available in PMC 2012 June 01.

Published in final edited form as:

*Nat Methods*. ; 8(12): 1059–1064. doi:10.1038/nmeth.1748.

## Enhanced yeast one-hybrid (eY1H) assays for high-throughput gene-centered regulatory network mapping

John S Reece-Hoyes<sup>1,2,3,\*</sup>, Alos Diallo<sup>1,2,3</sup>, Bryan Lajoie<sup>1,2,4</sup>, Amanda Kent<sup>1,2,3</sup>, Shaleen Shrestha<sup>1,2,3</sup>, Sreenath Kadreppa<sup>1,2,3</sup>, Colin Pesyna<sup>5</sup>, Job Dekker<sup>1,2,4</sup>, Chad L Myers<sup>5</sup>, and Albertha J M Walhout<sup>1,2,3,\*</sup>

<sup>1</sup>Program in Systems Biology, Worcester, MA, USA

<sup>2</sup>Program in Gene Function and Expression, Worcester, MA, USA

<sup>3</sup>Program in Molecular Medicine, Worcester, MA, USA

<sup>4</sup>Department of Biochemistry and Molecular Pharmacology University of Massachusetts Medical School, Worcester, MA, USA

<sup>5</sup>Department of Computer Science and Engineering, University of Minnesota–Twin Cities, Minneapolis, MN, USA

### Abstract

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 utilize 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 amount of time. eY1H assays will be an important tool for gene regulatory network mapping in *Caenorhabditis elegans* and other model organisms, as well as humans.

### INTRODUCTION

Gene expression is governed by sequence-specific transcription factors that bind to regulatory genomic DNA regions such as promoters. To understand the mechanisms of gene

---

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Correspondence and requests for materials should be addressed to A.J.M.W. (marian.walhout@umassmed.edu) or J.S.R-H. (john.reece-hoyes@umassmed.edu).

**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.M.; S.K. cloned additional transcription factor-encoding ORFs; J.S.R-H. and A.J.M.W. wrote the paper.

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 critical to know which transcription factors interact with which regulatory genomic DNA regions. These interactions can be represented as gene regulatory networks that can provide important insight into the “design principles” of gene control and, thereby, into the mechanisms of organismal development, growth, homeostasis and environmental responses<sup>1</sup>.

Several approaches can detect interactions between transcription factors and DNA<sup>1,2</sup>. 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. ChIP is well suited to detect interactions involving broadly or highly expressed proteins. However, it is more challenging to detect binding of factors for which suitable antibodies are not available, that bind DNA with low affinity or specificity, that are expressed at low levels, that are present only in a few cells or that bind DNA in a condition-dependent manner. Gene-centered (DNA-to-protein) methods such as Y1H assays 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 interactions that occur in a few cells or under highly specific conditions *in vivo* can effectively be discovered. An important conceptual advantage of a gene-centered approach includes the ability to directly identify the transcription factors that interact with regulatory elements of genes expressed in a tissue of interest<sup>3,4</sup>, involved in a particular biological process<sup>5</sup> or a set of genes belonging to a particular class, such as microRNAs<sup>6</sup>. Limitations of Y1H assays include the inability to retrieve heterodimeric transcription factors or proteins that require post-translational modifications in order to bind to DNA. Arguably, the chief problem currently facing both methodologies is achieving the throughput required to generate genome 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*<sup>7,8</sup>. 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 (see **Methods** for more details).

We previously demonstrated that large DNA fragments such as gene promoters can be used as DNA baits, and combined Y1H assays with Gateway cloning<sup>7</sup>. Gateway cloning is based on site-specific recombination, and can use existing open reading frame (ORF) and promoter clone resources<sup>9–11</sup>. We have used Gateway-compatible Y1H assays to delineate several medium-scale *C. elegans* regulatory networks<sup>3–6</sup>. Each of these contains 50–100 gene promoters and ~100 transcription factors, and took several years to complete. While 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 due to the low abundance of some transcription factors in cDNA libraries, and the lack of saturation in library-based Y1H screens<sup>12</sup>.

An alternative to screening complex libraries is to screen individual transcription factors directly for binding to DNA baits. This can be performed 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 express the prey protein (Fig. 1a). Both of these approaches detect more interacting transcription factors, take less time and reduce cost and effort as compared to screening complex libraries. However, mating assays only detect about half as many interactions compared to haploid transformation assays<sup>12</sup>.

Here we describe a novel eY1H pipeline in which almost all aspects of the Y1H assay are improved. We have increased coverage by mating with a new prey yeast strain carrying a high-copy prey expression vector, and by adjusting assay conditions. We have increased the throughput of the assay by generating a new yeast strain into which it is more efficient to integrate the DNA baits, by using a 1,536-colony robotic platform, and by standardizing readout analysis and interaction scoring. 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, identifying interacting transcription factors. eY1H assays will provide an important tool for the genome-scale delineation of gene regulatory networks in *C. elegans*, and, as shown in the accompanying papers<sup>13,14</sup>, can be efficiently applied to other model organisms and humans.

## RESULTS

### eY1H assays

eY1H assays incorporate a number of modifications that increase both throughput and coverage over previously reported Y1H methods (Table 1, see **Methods**). eY1H assays start with a lawn of a DNA bait strain that is mated with colonies of prey strains that express transcription factors (Fig. 1b). For the prey resource, we generated a collection of 865 *C. elegans* proteins (Supplementary Table 1) including 834 transcription factors (89% of the 937 present in 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<sup>3</sup> (Supplementary Table 3). This prey resource contains 85 additional factors compared to our previous collection<sup>12</sup>.

In eY1H assays, each DNA bait-prey combination is tested four times using “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 (see below). By using a high-density 1,536-colony format<sup>15,16</sup> all *C. elegans* factors are covered on only three plates. After mating a DNA bait yeast strain with the TF quad array and selecting for diploids, yeast are transferred to a single readout plate per quad array plate. Using a single readout plate halves the number of plates per assay and,

importantly, also removes the error-prone step of comparing two independent readouts. After seven days of incubation, a single image is taken per plate, which is manually or computationally evaluated for blue quads (see below). Capturing a single image at one time point for each readout plate increases the throughput of the eY1H pipeline and, typically, images captured after seven days of incubation display all interactions. For the minority of DNA bait strains that display 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 (*e.g.* 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, while reproducibility is defined as how many events detected by one screen are reproduced in a second<sup>17</sup>. 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 exhibit low background and can be bound by numerous transcription factors<sup>3,4</sup>. eY1H assays are essentially saturated by the third screen (Fig. 2a, Supplementary Fig. 3,4 and Supplementary Table 4). Importantly, 89% of all interactions are found in a single screen (sampling sensitivity), and 90% of the interactions found in a single screen are reproduced in a second (reproducibility), which indicates that eY1H assays have low technical false negative and false positive rates.

### eY1H assays provide high coverage

To compare the coverage of eY1H assays to previous Y1H methodologies, we focused on *Pvha-15* because we already had interaction data for this promoter from testing individual low-copy prey clones by both haploid transformation and mating as well as from screens of low-copy vector libraries<sup>12</sup>. These previous methods assayed interactions on multiple plates containing different concentrations of 3AT for up to two weeks, and were less likely to miss interactions only detectable outside the standardized readout conditions used by 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 only tested using 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 exclusively detected by eY1H assays, and 24 were not detected by eY1H assays (Fig. 2c). Recently, it has been demonstrated with yeast two-hybrid (Y2H) assays that analyzing the same set of baits or preys with different configurations of vectors and/or host yeast strains results in datasets that are only partially overlapping, but that interactions exclusive to any configuration can be as valid as those shared by all<sup>18,19</sup>. Our data suggest that a similar phenomenon occurs in Y1H assays.

To address whether interactions detected exclusively by eY1H assays can occur *in vivo*, we took advantage of publicly available ChIP data recently made available for 14 transcription factors ([modencode.org](http://modencode.org)) that can function in Y1H assays (*i.e.* that were detected at least

once). 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, even though the factors involved have been found to interact with other DNA baits in Y1H. Aside from being putative ChIP false positives, potential explanations for this include: 1) ChIP interactions involve isoforms not present in our clone collection; 2) the interactions occur in the distal portion of the promoter, which may be too far from the transcription start site to confer efficient yeast reporter activation; 3) the transcription factor requires post-translational modifications or cofactors not available in yeast for binding that specific promoter; 4) the chromatin context in the yeast nucleus may preclude the detection of some interactions; and 5) the interaction detected by ChIP may be indirect. Taken together, eY1H assays find more interactions for *Pvha-15* than the previous methods combined, and detect interactions that have *in vivo* support.

### eY1H assays provide a much higher throughput

We used 50 previously analyzed *C. elegans* DNA bait strains to evaluate the throughput of the eY1H pipeline, and also to further characterize eY1H coverage. These baits had been assayed by transforming haploid DNA bait strains with (cDNA- and transcription factor clone-) libraries of low-copy prey vectors, followed by limited directed experiments using individual clones<sup>3-6,12</sup>. All 50 DNA bait strains were screened once in a single eY1H batch, the interactions scored manually, altogether taking only 13 days. With library screens it took about two years to process a similar number of baits, identifying the interacting transcription factors by sequencing and retesting by transformation and/or gap-repair. eY1H assays are thus at least 50 times faster than library screens and can be configured for even higher throughput (discussed below).

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). This indicates that the quad array is of high quality and confirms the high rate of reproducibility. Previously, we reported 476 interactions for these 50 DNA baits (Supplementary Table 6). 221 of the published interactions (46%) were detected by eY1H assays, while 548 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 that are encoded by long ORFs, or that are 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 are uniquely detected in eY1H assays, possibly due 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<sup>5</sup>, which is enriched for nuclear hormone receptors (NHRs). The proportion of NHRs exclusively detected by eY1H assays is similar to that in the combined data (Fig. 3d), suggesting that eY1H data are of high quality.

There are several possible explanations why published Y1H interactions can be 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 published interactions may have been detectable only under different conditions. Third, some transcription factors may not work well in eY1H assays, because of the clone (*e.g.* 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 (*e.g.* GATA-type factors appear to work better in haploid experiments). Finally, it is of course possible that some of the previously published interactions are technical false positives that cannot be repeated. In most of our previous studies, we used a scoring system to classify interactions<sup>4</sup>. We found that interactions with the highest scores are more likely to be reproduced in eY1H assays (data not shown), suggesting that interactions not detected by eY1H assays may include technical false positives. Related to this may also be the fact that a higher proportion of “published only” interactions involve uDBPs (Fig. 3d), which may be more likely to be incorrect.

Overall, processing 50 baits using eY1H assays detected 60% more (769 *versus* 476) and higher quality interactions in 1/50<sup>th</sup> of the time. This vast improvement in assay throughput and robustness makes assaying large numbers of DNA baits, and therefore generating genome-scale interaction datasets, more feasible.

### SpotOn: a computational tool 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 (**Methods**). SpotOn imports a JPEG image of an eY1H readout plate (Fig. 4a) and performs the following major 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  $\beta$ -galactosidase expression (*i.e.* blueness) of each colony is determined (Fig. 4b). Third, these intensity values are normalized for noise due to growth differences resulting from uneven nutrient availability within 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 detects 83% of the manual calls (*i.e.* the false negative rate is ~17%, Fig. 4e). Importantly, the majority of missed manual calls were scored as “very weak” (Fig. 4f). False calls typically arise due to bleed-over from strong positives. By manually checking all positives that occur next to highly blue quads (~20% of calls in this dataset, Supplementary Table 8), this class of false calls is eliminated, reducing the false call rate from 5% to 1%. Finally, it is important to note that individual users can tailor the SpotOn parameters to optimize trade-off between sensitivity and specificity.

A further potential advantage of SpotOn is that quantitative eY1H data may enable further studies such as extracting transcription factors that are positive at a higher threshold, or determining the average reporter activation above background for a given transcription factor across all DNA baits that it binds. Finally, SpotOn can 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 regulatory genomic elements. eY1H assays provide a tremendous increase in throughput and performance compared to other types of Y1H assays. eY1H assays can effectively be used to analyze *C. elegans* gene promoters (this study), and *Arabidopsis* promoters<sup>14</sup> as well as for analysis of a variety of human regulatory sequences<sup>13</sup>.

While we chose to develop mating-based assays, others have developed a transformation-based robotic Y1H platform for the study of *Drosophila* DNA elements<sup>20</sup>. Transformation-based assays provide high coverage<sup>12</sup>, but are relatively cumbersome, as they require repeated costly preparations of prey DNA and the liquid handling of highly viscous solutions. Nonetheless, the availability of different Y1H approaches provides the community with a selection of tools that can be implemented in their own laboratories.

The technical advances outlined in this paper can 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 using 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). We envision that eY1H assays can be used with 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. Further, the throughput of the pipeline can increase with greater batch sizes, and of course by using 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<sup>18</sup>. This same study showed that the assay coverage can be further improved when using different configurations of vectors (*e.g.* C-terminal fusion of the activation domain to the prey). If

coverage is paramount, the eY1H pipeline can easily incorporate such modifications, especially for preys that do not appear to function in the traditional vector that uses an N-terminal fusion.

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 green fluorescent protein (GFP) under the control of the DNA bait, and examine GFP expression in the presence or absence of an interacting transcription factor, either by RNAi or by using mutant animals<sup>3,5,6,21,22</sup>. Similarly, quantitative RT-PCR or microarrays can be used to compare endogenous gene expression following 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<sup>23</sup>.

In summary, eY1H assays provide a convenient tool for the gene-centered mapping of gene regulatory networks in a variety of 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.

## ONLINE METHODS

### Yeast one-hybrid (Y1H) 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*<sup>7,8</sup>. 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 (*i.e.* the assay does not use “naked DNA”). We generally use genomic fragments between 300 bp and 2 kb<sup>8</sup>. 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  $\beta$ -galactosidase, which generates a blue compound from the colorless X-gal, while the *HIS3* reporter expression product permits growth on media lacking histidine and containing the competitive His3 inhibitor, 3-Amino-1,2,4-Triazole (3AT). The readout of the assay is therefore the ability of yeast to grow in the presence of 3AT and/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 by single-positive yeast that activate only one reporter because the physical interaction occurs twice within the same yeast nucleus<sup>8</sup>.

### Updated *C. elegans* transcription factor compendium (wTF2.2) and uDBP list

Previous iterations of the *C. elegans* transcription factor compendium contained 934 (wTF2.0)<sup>24</sup> and 940 (wTF2.1)<sup>12</sup> 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<sup>12</sup> 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 due to recent reports of sequence-specific DNA-binding ability of their protein products<sup>25–27</sup>.

Using Y1H cDNA-library screens, we have previously retrieved multiple *C. elegans* proteins that do not possess a recognizable DNA-binding domain<sup>3–6</sup>. We referred to these as “novel putative transcription factors”. Using proteins arrays, multiple human proteins were also uncovered that do not possess a known DNA-binding domain but that do bind DNA in a specific manner, and these are referred to as unconventional DNA-binding proteins, or uDBPs<sup>28</sup>. 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<sup>25,26</sup>.

### Generating the wTF2.2 clone array

The starting point for the wTF2.2 clone array was the wTF2.1 resource<sup>12</sup>. We added Entry clones for 85 newly cloned transcription factors to our collection (Supplementary Table 1). Seven were recent additions to our transcription factor list and existing clones were cherry-picked from the ORFeome<sup>10</sup>, 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<sup>29</sup>. Twenty-seven Entry clones were a kind gift from Dr. Salehi-Ashtiani (Center for Cancer Systems Biology), while the others were generated in-house. Overall, the wTF2.2 array contains 834 of the 937 (89%) of wTF2.2, as well as 31 uDBPs. The ORFs from all Entry clones in our collection were transferred into pDESTAD-2 $\mu$  (Invitrogen) by a Gateway LR reaction. The resulting AD-prey Destination clones were transformed into Yc1867 (see below) in 96-well format with empty pDEST-AD-2 $\mu$  in the H12 position. Frozen glycerol stocks of the resulting transcription factor prey yeast strains were generated by transferring a small amount (half a match-head) of yeast to 1 ml Sc-Trp liquid media in 96-well deep-well (2 ml) plates, incubating in an orbital shaker (MULTIFORS, 200rpm 30°C) for 48 hours, pelleting the yeast (2,000 g, 5 min), discarding the supernatant, resuspending each pellet in 200  $\mu$ l 15% (v/v) glycerol and transferring each yeast suspension to 96-well plates that are stored at –80°C. pDEST-AD-2 $\mu$  is a “high-copy” vector as it contains a 2 $\mu$  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<sup>30</sup>. Both pDEST-AD-2 $\mu$  and pDEST-22 use the full-length *ADHI* promoter to drive expression of the AD-prey fusion, therefore higher vector copy-number results in higher expression levels. While it is possible that expression of these AD-prey fusions would adversely affect the yeast, no difference in mating ability or growth of the haploid prey strains, or growth of resulting diploids, was observed (data not shown).

### Preparing 1536-colony “quad arrays”

The RoToR HDA robot (Singer Instruments) is used for every step in the generation of 1536 colony AD-wTF array plates. The process is outlined in Supplementary Figure 1. The

RoToR HDA uses disposable plastic pads with 96, 384 or 1536 pins to precisely transfer yeast between solid agar plates, or between a liquid source (*e.g.* yeast suspension) and a solid agar plate. Plates that are used on the robot require extra care to ensure that the agar surface is flat for efficient transfer of colonies using pads. We pour 65 mL media into each Singer Plusplate (prepared as stacks of no more than five), dry the plates overnight at room temperature on a flat surface, and store them at 4°C in air-tight plastic bags. To build a 1536-colony “quad array”, 96-colony plates of transcription factor prey yeast are first generated by spotting from 96-well plates of glycerol stocks onto solid agar plates. A single transfer from each of four separate 96-colony plates is then used to build a 384-colony plate in which each transcription factor prey yeast strain is present once. Finally, four transfers from the same 384-colony plate are used to create the 1536-colony plate in which each transcription factor prey yeast is present in quadruplicate. Transcription factor prey yeast strains are grown on Sc-Trp plates, with incubation steps of two days at 30°C after each transfer resulting in yeast colonies of ~3 mm, ~2 mm and ~1 mm on 96-, 384- and 1536-spot plates, respectively. The current collection of *C. elegans* transcription factor prey yeast strains fills eleven 96-well plates (Supplementary Table 1), so a total of three 1536-colony plates cover all available transcription factors and uDBPs, one using 96-well plates one to four, a second using 96-well plates five to eight, and a third using 96-well plates nine to eleven. Four of the 16 positions in the bottom right corner of each quad array are intentionally left without yeast so that plate orientation and identity can be verified (a different four colonies are omitted from each of the three quad array plates). The remaining 12 positions in the bottom right corner contain empty pDEST AD-2 $\mu$  and serve as negative controls in eY1H assays. The quad array is transferred every week to fresh Sc-Trp plates, and is stored at room temperature when not in use. New TF quad array plates are generated from glycerol stocks every eight weeks.

### Performing eY1H assays

The eY1H pipeline takes six days to generate diploid yeast from the haploid DNA bait and transcription factor prey strains, and typically another seven days to assay the diploids for reporter gene expression. An outline is provided in Figure 1. The RoToR HDA is used at every step unless specified. Between each step described below, the plates are incubated at 30°C. On day1, fresh copies of the quad arrays are generated using transfers from older copies to Sc-Trp plates, while a lawn of the DNA bait strain is generated on either a YAPD or Sc-Ura-His plate (yeast are mixed in sterile water, the suspension is spread onto solid agar plates using sterile glass beads). One plate of transcription factor prey yeast provides enough yeast to set up mating plates for four DNA baits. On day3, a mating plate is prepared 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 1536-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 day4, yeast are 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 day6, diploid yeast are transferred from the Sc-Ura-His-Trp plate to the eY1H “readout” plate (Sc-Ura-His-Trp + 5 mM 3AT, 160 mg/L 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 the 3AT will grow,

and only yeast that grow and express  $\beta$ -galactosidase from the *LacZ* reporter gene will metabolize X-gal into a blue compound. Thus, activation of both reporters is analyzed in the same media plate. All blue colonies that grow on readout plates are double positives, even though some positives appear 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 is captured using the spImager ([sprobotics.com](http://sprobotics.com)) that places each plate in a uniformly lit environment where a mounted Canon Rebel XSi digital camera with a EF-S 60mm F/2.8 macro lens controlled by spImager v1.0.2.0 software takes a high-resolution photograph that is converted to a 4272 pixel  $\times$  2848 pixel JPEG image and stored. If processing few or highly auto-active 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 seven days of incubation.

### The new prey strain: Y $\alpha$ 1867

Performing Y2H assays with different strains improves coverage<sup>19</sup>. We tested eight yeast strains (SY3002, Y287, Y1495, Y1864 of mating type “a”, and SY3003, Y288, Y1494, Y1867 of mating type “ $\alpha$ ”), and our original strains YM4271<sup>8</sup> and Y1H $\alpha$ 001<sup>12</sup> in all 25 pair-wise combinations in mating-based Y1H assays. All new strains had mutations in the *URA3*, *HIS3*, and *TRP1* genes, and were kind gifts from Dr. 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<sup>12</sup>. In addition to the number of interactions detected by each mating combination, we took into account mating compatibility of each pair, and the general phenotype of both haploid and diploid strains (*e.g.* some yeasts were “waxy” and difficult to transfer, some grew too fast or slow, some were more resistant to 3AT). The best combination was Y $\alpha$ 1867 (MAT $\alpha$  SUC2 gal2 mal mel flo1 flo8-1 hap1 ho bio1 bio6 ura3-52 ade2-101 trp1-901 his3- 200) as the host prey strain and YM4271<sup>8</sup> 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 integrate the two *DNA bait*:reporter constructs into different mutant loci within the genome of a host yeast strain<sup>7</sup>. Previously, we have used the yeast strain YM4271<sup>8</sup> with the *pMW#3 LacZ* construct integrated at the *URA3* locus, rescuing the *ura3-52* Ty insertion that disrupts the gene<sup>31</sup>, and the *pMW#2 HIS3* construct integrated at the *HIS3* locus, rescuing the *his3- 200* deletion that includes the entire ORF<sup>32</sup>. 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 1090 bp, the integration success rate for *pMW#2* is much lower than that for *pMW#3* (~100 and ~2,000 events per  $\mu$ g linear DNA, respectively, data not shown). This lower integration rate for *pMW#2* is the limiting factor when performing double integrations (*i.e.* 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- 200* locus (1040 bp deletion) within YM4271 with *his3- 1* (190 bp deletion<sup>33</sup>). This involved two yeast transformations performed using a standard protocol<sup>34</sup>. The *his3- 200* locus of the YM4271 strain was first replaced by a wild-type *HIS3* gene by transforming with *Bam*HI-digested *pPL97* (that contains *HIS3*), and selecting colonies that are able to grow in the absence of histidine (Sc-His). A resulting *HIS3+* strain was then transformed with *Xho*I-digested *pNN132* (that contains a *URA3* gene flanked by a wild type *HIS3* gene and a *his3- 1* 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 (5-fluo-orotic acid) agar plates (0.1% w/v 5-FOA). Yeast able to grow in the presence of 5-FOA have lost the *URA3* gene due to internal recombination between the *HIS3* genes, but have an equal chance of maintaining either the wild-type or 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 each other as well as to 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  $\mu$ g transformed linear vector for the YM4271 strain to ~4,000 for Y1H-aS2. Accordingly, the rate of double integration increased ten-fold (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 that anneals within the wild-type *HIS3* and *his3- 1* loci but not within *his3- 200* (Supplementary Table 9). Both pPL97 and pNN132 were kind gifts from Dr. Boeke (Johns Hopkins University).

### SpotOn: a program for automated colony quantification

Automated eY1H assay quantification involves three major steps: 1) drawing a grid to fit the quad array so that each colony can be associated with the transcription factor it expresses, 2) identifying which colonies are “positive” (*i.e.* those displaying significantly more reporter expression than background), and 3) removing systematic false calls to create a final list of eY1H DNA bait-transcription factor interactions.

**1) Drawing the grid**—Before drawing the grid, 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 ([libgd.org](http://libgd.org)). To compensate for local effects that arise from uneven media/nutrient distribution, each image is corrected by local median normalization (LMN) as follows. For each pixel, the median intensity of the 80 neighboring (9x9 square) pixels is calculated. The original intensity value is divided by this median value to get a 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. Since colonies are circular, their pixels will be densely clustered around a central core, so removing rough edges from (*i.e.* circularizing) all objects only significantly affects non-circular objects that are likely not colonies. Circularizing is performed by removing all pixels for which less than 18 of the surrounding 24 pixels (5x5 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 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 NxM 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 1536 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/column are detected, the complete grid can still be drawn. This is important for this application because the above filtering steps remove a significant number of true colonies. Finally, the grid lines are applied to the original cropped image and transcription factor identities are assigned to each colony using its grid coordinate (Supplementary Fig. 5f).

While 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 was 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. This manual grid drawing is necessary for difficult plates, but is more time-consuming, because it requires a manual component, whereas the SpotOn grid drawing is fully automated.

**2) Identifying positive colonies**—This process involves determining to what extent the reporters are expressed in each colony, followed by whether a colony exhibits significantly 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 standard deviation 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. From here onward, we will use the term “colony intensity” to represent that median red channel value.

Before determining whether a colony is positive (*i.e.* exhibits significantly higher reporter expression than background) two normalization steps are performed. Row/column normalization (RCN) neutralizes the effect of uneven media/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/column, the median of the colony intensities is determined, and the median of these row/column values is then calculated for the whole plate. A RCN factor is derived for each row/column by dividing each row/column median by the plate row/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 is divided into three groups according to the three TF quad array plates (“1to4”, “5to8”, “9to12”), 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 standard deviation is calculated using all colony intensities from all plates, and 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).

**3) 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 likely false positives, is low (data not shown). We observed two important situations where 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, we first identify putative bleed-over quads in which at least three



colonies have a raw intensity of  $\geq 200$  (this cutoff was empirically determined from the relevant quads; data not shown). We then automatically remove any neighboring colonies 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 (*e.g.* 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. 4a) and so the two distal colonies will remain positive and the transcription factor would 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 neighbor bleed-over quads removes too many true positives, so instead we flag 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 removing the colonies that are 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 a mean of the Z-scores and colony intensities for the positive colonies in each quad. The list also indicates which interactions are from quads that neighbor 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 are false and 17% of the manual calls are missed. Importantly, the majority of these missed calls exhibit a very weak Y1H phenotype (Fig. 5f). 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, the false call rate is reduced 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 the true calls would be missed (Fig. 5e).

**4) Final comments**—SpotOn is designed to be robust 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. The code for SpotOn is available upon request.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

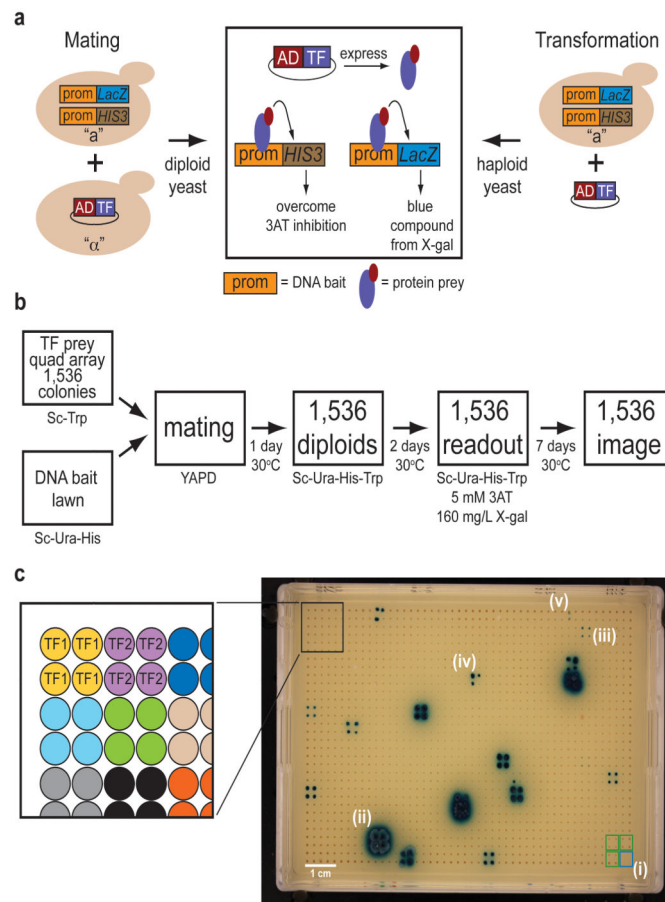
## ACKNOWLEDGMENTS

We thank members of the Walhout lab for discussions and critical reading of the manuscript. Special thanks to S. Lee for media preparation. We thank K. Salehi-Ashtiani (Center for Cancer Systems Biology) for transcription factor Entry clones, J. Boeke (Johns Hopkins University) for advice on the creation of Y1H-aS2 and C. Boone (University of Toronto) for general advice on the robotic pipeline and the use of different yeast strains. This work was supported by National Institutes of Health (NIH) grant GM082971 to A.J.M.W. Research in the Dekker lab 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 National Science Foundation grant DBI 0953881.

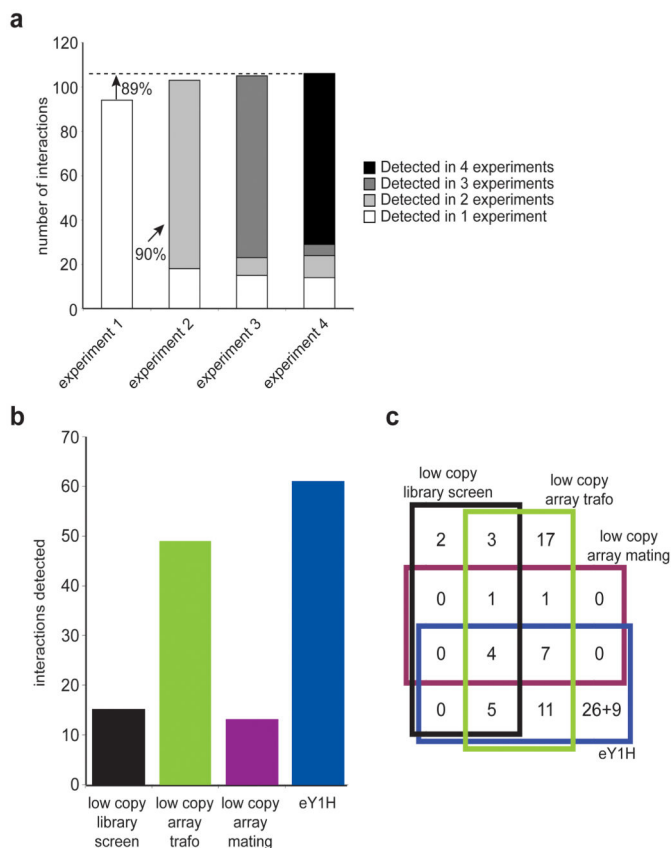
## REFERENCES

1. Walhout AJM. Unraveling Transcription Regulatory Networks by Protein-DNA and Protein-Protein Interaction Mapping. *Genome Res.* 2006; 16:1445–1454. [PubMed: 17053092]
2. Arda HE, Walhout AJM. Gene-centered regulatory networks. Briefings in functional genomics and proteomics. 2009 doi:10.1093/elp049.
3. Deplancke B, et al. A gene-centered *C. elegans* protein-DNA interaction network. *Cell.* 2006; 125:1193–1205. [PubMed: 16777607]
4. Vermeirssen V, et al. Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein-DNA interaction network. *Genome Res.* 2007; 17:1061–1071. [PubMed: 17513831]
5. Arda HE, et al. Functional modularity of nuclear hormone receptors in a *C. elegans* gene regulatory network. *Molecular Systems Biology.* 2010; 6:367. [PubMed: 20461074]
6. Martinez NJ, et al. A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes Dev.* 2008; 22:2535–2549. [PubMed: 18794350]
7. Deplancke B, Dupuy D, Vidal M, Walhout AJM. A Gateway-compatible yeast one-hybrid system. *Genome Res.* 2004; 14:2093–2101. [PubMed: 15489331]
8. Deplancke B, Vermeirssen V, Arda HE, Martinez NJ, Walhout AJM. Gateway-compatible yeast one-hybrid screens. *CSH Protocols.* 2006 doi:10.1101/pdb.prot4590.
9. Walhout AJM, et al. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods in enzymology: “Chimeric genes and proteins”.* 2000; 328:575–592.
10. Lamesch P, et al. *C. elegans* ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions. *Genome Res.* 2004; 14:2064–2069. [PubMed: 15489327]
11. Dupuy D, et al. A first version of the *Caenorhabditis elegans* promoterome. *Genome Res.* 2004; 14:2169–2175. [PubMed: 15489340]
12. Vermeirssen V, et al. Matrix and Steiner-triple-system smart pooling assays for high-performance transcription regulatory network mapping. *Nat. Methods.* 2007; 4:659–664. [PubMed: 17589517]
13. Reece-Hoyes JS, et al. Yeast one-hybrid assays for high-throughput human gene regulatory network mapping. *Nat. Methods.* 2011 In press.
14. Gaudinier A, et al. An automated eY1H approach to elucidate *Arabidopsis* stele gene regulatory networks. *Nat. Methods.* 2011 In press.
15. Tong AHY, et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science.* 2001; 294:2364–2368. [PubMed: 11743205]
16. Costanzo M, et al. The genetic landscape of a cell. *Science.* 2010; 327:425–31. [PubMed: 20093466]
17. Venkatesan K, et al. An empirical framework for binary interactome mapping. *Nat. Methods.* 2009; 6:83–90. [PubMed: 19060904]
18. Chen YC, Rajagopala SV, Stellberger T, Uetz P. Exhaustive benchmarking of the yeast two-hybrid system. *Nat. Methods.* 2010; 7:667–668. [PubMed: 20805792]
19. Braun P, et al. An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods.* 2009; 6:91–7. [PubMed: 19060903]
20. Hens K, et al. Automated protein-DNA interaction screening of *Drosophila* regulatory elements. *Nat. Methods.* 2011 In press.
21. Ow MC, et al. The FLYWCH transcription factors FLH-1, FLH-2 and FLH-3 repress embryonic expression of microRNA genes in *C. elegans*. *Genes Dev.* 2008; 22:2520–2534. [PubMed: 18794349]
22. Reece-Hoyes JS, 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.* 2009; 37:3689–3698. [PubMed: 19372275]
23. Walhout AJM. What does biologically meaningful mean? A perspective on gene regulatory network validation. *Genome Biol.* 2011; 12:109. [PubMed: 21489330]
24. Reece-Hoyes JS, et al. A compendium of *C. elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol.* 2005; 6:R110. [PubMed: 16420670]

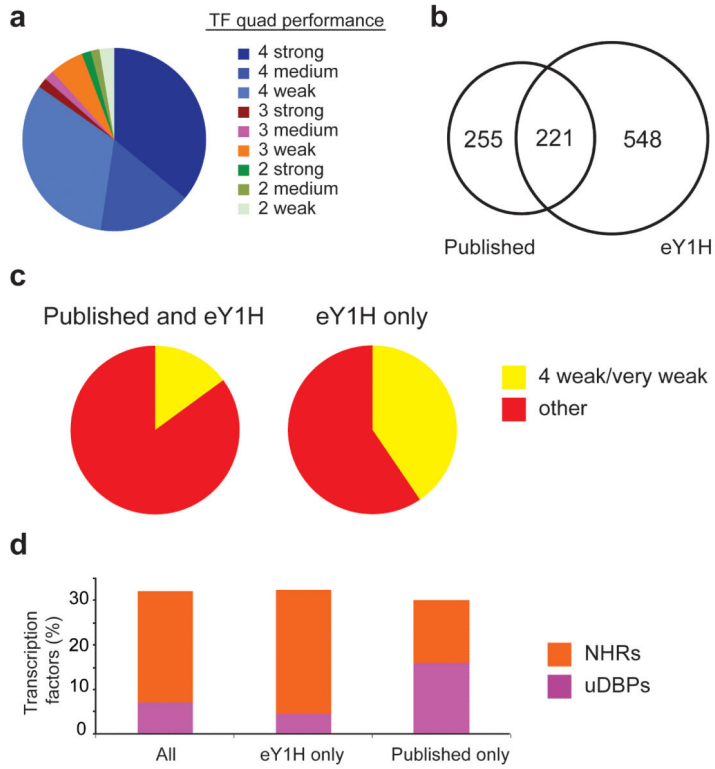
25. Lesch BJ, Gehrke AR, Bulyk ML, Bargmann CI. Transcriptional regulation and stabilization of left-right neuronal identity in *C. elegans*. *Genes Dev.* 2009; 23:345–358. [PubMed: 19204119]
26. Witze ES, Field ED, Hunt DF, Rothman JH. *C. elegans* pur alpha, an activator of end-1 synergizes with the Wnt pathway to specify endoderm. *Dev. Biol.* 2009; 327:12–23. [PubMed: 19084000]
27. Tabuchi T, et al. Chromosome-biased binding and gene regulation by the *Caenorhabditis elegans* DRM complex. *PLoS Genet.* 2011; 7:e1002074. [PubMed: 21589891]
28. Hu S, et al. Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell.* 2009; 139:610–622. [PubMed: 19879846]
29. Salehi-Ashtiani K, et al. Large-scale RACE approach for proactive experimental definition of *C. elegans* ORFeome. *Genome Res.* 2009; 19:2334–42. [PubMed: 19801531]
30. Rajagopala SV, Hughes KT, Uetz P. Benchmarking yeast two-hybrid systems using the interactions of bacterial motility proteins. *Proteomics.* 2009; 9:5296–5302. [PubMed: 19834901]
31. Rose M, Winston F. Identification of a Ty insertion within the coding sequence of the *S. cerevisiae* *URA3* gene. *Mol. Gen. Genet.* 1984; 193:557–560. [PubMed: 6323928]
32. Struhl K. Nucleotide sequence and transcriptional mapping of the yeast *pet56-his3-ded1* gene region. *Nucleic Acids Res.* 1985; 13:8587–8601. [PubMed: 3001645]
33. Scherer S, Davis RW. Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 1979; 76:4951–4955. [PubMed: 388424]
34. Walhout AJM, Vidal M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods.* 2001; 24:297–306. [PubMed: 11403578]

**Figure 1.**

Enhanced Y1H (eY1H) assays. **(a)** Schematic illustration of Y1H assays by mating and by transformation. prom – promoter (or DNA bait); AD – Gal4 transcription activation domain; TF – transcription factor; 3AT - 3-Amino-1,2,4-Triazole. **(b)** eY1H pipeline in which the different steps are indicated. Ura – uracil; His – histidine; Trp – tryptophan. **(c)** Example of eY1H readout plate ("1,536 image" in (b)). Each transcription factor is tested in quadruplicate. (i) control quads that lack yeast indicate plate identity and orientation (blue square), while 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 exhibits 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 only consider quads in which at least two of the four colonies are positive because such an interaction is by definition retested.

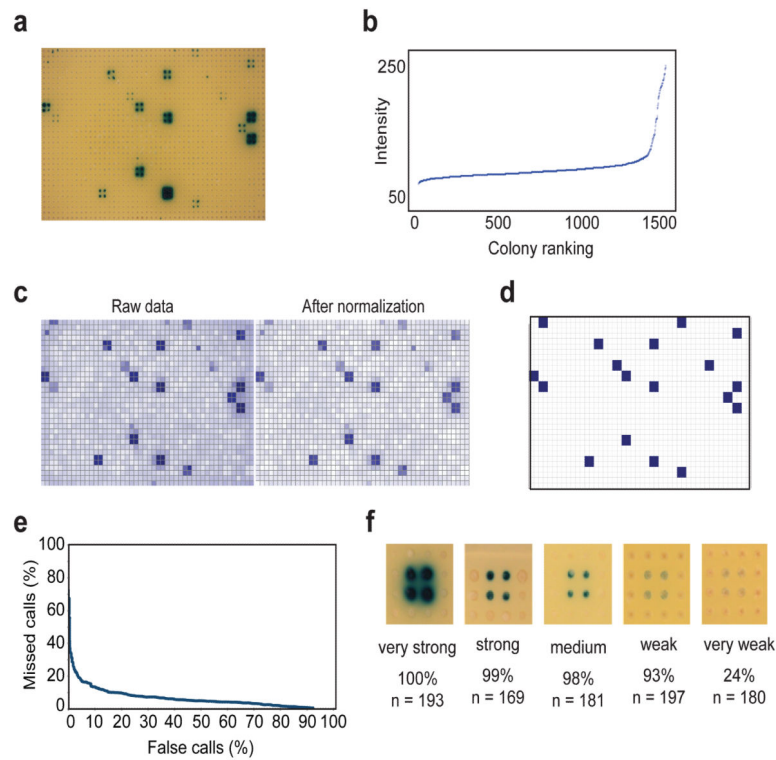


**Figure 2.** Sampling sensitivity and reproducibility of eY1H assays. **(a)** Two promoters, *Pvha-15* and *Pcog-1* were screened four times versus the worm transcription factor array, and the cumulative number of times an interaction was detected is indicated. Overall, a single experiment detects 89%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



**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 scored positively. Transcription factors for which only a single colony scored were not considered. **(b)** Venn diagram illustrating overlap between published and eY1H interactions. **(c)** Proportion of interactions detected with all four colonies scoring positively that were weak or very weak. **(d)** Percentage NHRs and uDBPs detected by different subsets of the data.



**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 shown in **(a)**. **(e)** Graph depicting false calls *versus* missed calls for the 50 DNA baits screened with eY1H assays. A Z-score threshold with 5% false calls results in 17% missed calls. With limited manual curation the false call rate is reduced to 1% (see main text). **(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.

**Table 1**

Summary of changes in eY1H assays

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

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript