Published in final edited form as:

Methods. 2012 August ; 57(4): 441–447. doi:10.1016/j.ymeth.2012.07.027.

YEAST ONE-HYBRID ASSAYS: A HISTORICAL AND TECHNICAL PERSPECTIVE

John S. Reece-Hoyes* and A.J. Marian Walhout*

Program in Systems Biology, University of Massachusetts Medical School, 364 Plantation St, Worcester, MA, USA, 01605.

Abstract

Since its development about two decades ago, the yeast one-hybrid (Y1H) assay has become an important technique for detecting physical interactions between sequence-specific regulatory transcription factor proteins (TFs) and their DNA target sites. Multiple versions of the Y1H methodology have been developed, each with technical differences and unique advantages. We will discuss several of these technical variations in detail, and also provide some ideas for how Y1H assays can be further improved.

1 INTRODUCTION

At the foundation of every biological process are complex networks of physical interactions between various biomolecules. Such networks control how an embryo develops into a multi-cellular organism, how that organism maintains homeostasis, and how it responds to pathogens or disease. The biomolecules that participate in these networks include DNA, RNA, proteins and small molecules such as metabolites. Here, we focus on a specific class of physical interactions: those between TFs and their DNA target sites within the genome. These protein-DNA interactions (PDIs) can increase or decrease transcript (and thus, protein) expression from nearby genes, and are therefore integral to most biological processes.

Multiple experimental approaches have been developed to detect PDIs (reviewed in [1]), and they can largely be divided into two conceptually complementary groups: TF-centered and DNA-centered [2] (Figure 1). TF-centered assays take a TF of interest and determine which DNA targets it binds, whereas DNA-centered assays take a region of DNA of interest and determine the repertoire of TFs with which it interacts. We have discussed the advantages and limitations of experimental techniques from both conceptual groups elsewhere [3,4]. Our major conclusion is that if the goal is to detect all PDIs that ever occur in an organism, no one technique can achieve this alone, and it will likely require combining the results gained by all the different PDI detection methods. The subject of this review is one such methodology: DNA-centered Y1H assays. We will discuss its development in the early 1990s, the consequences of several technical differences on assay performance, and some ideas for future modifications.

© 2012 Elsevier Inc. All rights reserved.

*Co-corresponding authors, marian.walhout@umassmed.edu; john.reece-hoyes@umassmed.edu, Telephone: +1 508 856 4364, Fax: +1 508 856 2105.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

2 DEVELOPMENT OF THE Y1H ASSAY

The basic Y1H assay involves two components (Figure 2): 1) a reporter construct with DNA of interest cloned upstream of a gene encoding a reporter protein that can be easily detected; and 2) an expression construct that generates a fusion (or “hybrid”) between a TF of interest and a yeast transcription activation domain (AD). The DNA of interest is commonly called the “bait”, while the hybrid protein is called the “prey”. Both components are introduced into a suitable yeast strain, and if the TF binds the DNA in the milieu of the yeast nucleus, the AD induces reporter protein expression. Importantly, the yeast AD activates the reporter regardless of whether the TF is an activator or repressor and thus Y1H assays only report on the physical interaction between protein and DNA. The yeast strain used for Y1H is usually mutated in multiple auxotrophic genes, making the strain unable to grow in the absence of various compounds. Genes commonly used in this way (and their associated compound) are: *trp1* (tryptophan), *leu2* (leucine), *his3* (histidine), and *ura3* (uracil). The researcher uses rescue of these auxotrophic phenotypes to either indicate the presence of constructs by including wild-type genes in the vector backbone, or to indicate a PDI by using the wild-type protein as the reporter. The bacterial gene *LacZ* is also commonly used in Y1H reporter constructs (see below).

The inspiration for using yeast as a “living test tube” to interrogate PDIs was largely to overcome the technical difficulties of purifying enough protein of interest, from either the host or recombinant bacteria, to perform biochemistry-based PDI detection methods of the 1980s (such as gel-shift and DNaseI protection assays). Early efforts to detect PDIs within yeast expressed only the endogenous TF (rat glucocorticoid receptor [5]; human oestrogen receptor [6]), and relied on the ability of the TF’s own AD to activate expression of the reporter protein upon binding to the bait DNA. Because not all TFs possess this ability, especially not repressors, only some TFs could be assayed in this way. At about the same time, however, Ptashne and colleagues demonstrated that the AD of the yeast TF *GAL4* was separable from its DNA binding region and that fusion of *GAL4* AD to either a bacterial [7] or a yeast repressor [8] converted them to activators in yeast. Therefore, expressing potentially any TF as a fusion with an AD like that from *GAL4* would make them amenable to testing PDIs in yeast, and this observation formed the basis for the Y1H assay.

Soon after these early reports, several labs independently developed versions of Y1H screens expressing *GAL4* AD-TF hybrid proteins [9–13]. Interestingly, one of these approaches is TF-centered [9], where a library of bait reporter constructs is screened against a AD-TF hybrid prey, while the other four are DNA-centered, where a prey library is screened against a DNA bait. The protocols for all five of these screening methods are similar to each other, and also to what is still used today. First, a population of yeast that all harbor one of the Y1H components (either the bait reporter or prey expression construct) is transformed with a library of the other component, such that each cell of the population receives one construct from the library, and that enough library constructs are transformed to ensure that the library is efficiently interrogated. Then “positive” yeast colonies are selected that express sufficient amounts of the reporter protein, and the identity of the library construct within those yeast is determined to reveal the interactors.

3 TECHNICAL MODIFICATIONS TO Y1H ASSAYS

Notably, there are three technical variations among these first-reported Y1H screening techniques regarding their reporter constructs. Firstly, one uses two different reporter constructs [10], while the others use only one. Secondly, one method uses a reporter construct that is integrated into the yeast genome [11], while the others use episomal constructs. Thirdly, one detects TFs that bind a single copy of an endogenous promoter [12],

while the others use tandem repeats of a short (<30 bp) DNA sequence. Each of these technical differences has advantages and disadvantages in terms of assay performance. We shall now discuss in detail these, as well as some more recent, technical variations available for Y1H assays. A summary is presented in Table 1.

3.1 USING ONE OR MULTIPLE REPORTER CONSTRUCTS

The most commonly used Y1H reporter constructs express either beta-galactosidase or *HIS3*. Beta-galactosidase is detected using a colorimetric assay in which colorless X-gal is converted to a blue compound. *HIS3* is part the histidine biosynthesis pathway that enables yeast to grow in the absence of exogenously supplied histidine. It is much easier to screen many transformants using *HIS3* as the reporter as only colonies in which *HIS3* expression is activated will grow, whereas when using beta-galactosidase all colonies will grow and need to be tested colorimetrically. The activity of *HIS3* can be inhibited by adding a competitive inhibitor called 3-amino-1,2,4-triazole (3AT) to the growth media. 3AT can therefore be used to distinguish “strong” and “weak” PDI phenotypes by how much 3AT the PDI can withstand, but it is important to remember that there may be no correlation between PDI phenotype and DNA affinity of a TF.

Versions of Y1H that use multiple reporter constructs typically use two, one with *HIS3* as the reporter and the other with beta-galactosidase. Usually *HIS3* expression is tested first, then those positive colonies are tested for beta-galactosidase activity, and only the colonies that exhibit expression of both reporters (*i.e.* “double positives”) are considered further because the interaction is in essence recapitulated in the same cell. The Y1H version with multiple reporter constructs mentioned above has the reporters in two different yeast strains [10], but other versions have them in the same “bait yeast strain” [14,15]. Counting only double positive PDIs results in fewer technical false positive interactions than that observed when using a single reporter construct. Note that these technical false positives are due to some technical failure of the assay, and are distinct from biological false positives that are detected by the assay because they are biochemically real but cannot be demonstrated to occur in the endogenous organism [16]. Specific types of technical false positives that arise only with a single reporter system include proteins that are able to phenocopy increased reporter activity (like increased growth on media lacking histidine), or TFs that activate reporter expression despite binding outside the bait. These false positives do not occur when using two reporters because both the detection method for the two reporter proteins and the sequence of the reporter constructs are different. It is possible to identify these single reporter false positives when screening a short DNA sequence as they would remain positive even when the DNA sequence has been mutated. However, this extra step is challenging when screening larger regulatory regions as the PDI binding site(s) are typically unknown. Therefore, it is generally preferable to use more than one reporter in Y1H assays.

3.2 USING INTEGRATED OR EPISOMAL REPORTER CONSTRUCTS

Researchers introduce artificial plasmids into yeast that are either maintained as extra-chromosomally replicating episomes or are stably integrated into the genome [17]. The major complication of using episomal reporter constructs for Y1H assays is the occurrence of bait auto-activity, where a yeast TF (presumably) binds the DNA of interest, resulting in reporter activation. This was observed in the TF-centered version of Y1H mentioned above [9], where 1% of random 23-mers and 10% of randomly cloned small fragments (75 to 500 bp) of the rat genome exhibited reporter activity in the absence of a prey protein. Auto-activity is the major reason why TF-centered Y1H screens are challenging, with up to 10% of a bait reporter construct library causing technical false positives. The most common way to attenuate auto-activity is to use a reporter protein that can be inhibited, such as using 3AT to inhibit *HIS3*, so that only PDIs that activate reporter expression higher than the auto-

activity will be observed. If reporter inhibition is not possible, such as when using beta-galactosidase, only yeast that express more reporter protein than control yeast that lack a AD-TF fusion can be counted as positive. In some cases bait auto-activity is so high that no interactions can be detected using Y1H unless the yeast TF responsible for the auto-activation can be identified and disabled [18].

The combination of episomal reporter constructs and bait auto-activity causes problems for Y1H screens because episomes are subject to copy-number variation, meaning that auto-activity levels can vary from cell to cell. Thus, even when using measures to knock down or account for the background, falsely positive colonies are commonly selected that simply have more reporter constructs than others. Integrating the bait reporter construct into the genome fixes copy-number, making auto-activity attenuation methods substantially more effective, and removing this source of false positives. A further advantage of integrating reporter constructs is that the bait DNA is then chromatinized. This means that eukaryotic TFs are presented with regulatory sequences in a form that better reflects the endogenous situation than “naked” DNA. The chief disadvantage of using integrated reporter constructs is that integration is more cumbersome than transformation as it is less efficient, and because the integration must be confirmed from yeast genomic DNA, either by sequencing PCR products [19] or by Southern blotting.

3.3 INTERROGATING COMPLEX OR SIMPLE BAIT SEQUENCES

Regulatory sequences can be considered as either “simple” or “complex” according to how many potential binding sites they have. A simple bait has few potential binding sites as it is generally less than 30 base-pairs (bp) and may be cloned as a single copy or a tandem repeat, while a complex bait has more sites as it is generally longer and normally one copy is screened. Screening complex baits obviously enables a researcher to screen more genomic DNA than using simple baits, but there are more important reasons behind this choice. A complex bait is generally an entire promoter for a gene of interest, and detecting the repertoire of TFs that it binds will provide a broad view of the regulation of that gene. A simple bait usually has been linked to a regulatory role via evidence from other experiments (such as a functional activity screen or a multi-genome alignment to find highly conserved regions), and the researcher wants to determine which single TF (or a few) bind this small region.

A concern when generating reporter constructs for Y1H assays is that the junctions between the bait sequence and the vector sequence create new “artificial” potential binding sites. This problem is more pronounced when screening simple baits (especially in tandem where even more artificial sites occur between the repeats) as the junctions represent a much larger proportion of potential sites than for larger baits. Therefore, for simple baits it is important to also interrogate a mutated version in which only the junctions are maintained to identify PDIs that occur at these artificial sites [20].

3.4 GENERATING CONSTRUCTS USING RESTRICTION ENZYME- OR GATEWAY-CLONING

Traditionally, the reporter constructs and prey expression constructs for Y1H have been generated using restriction enzymes and ligase to clone a DNA bait or prey open reading frame (ORF) into a plasmid. However, the short DNA motifs (typically six base pairs) recognized by restriction endonucleases occur frequently enough to make generating constructs for many different baits or preys a time-consuming and complicated process, especially for large promoters or ORFs. We have developed a version of the Y1H assay for which the constructs are generated using the Gateway recombinational cloning system instead of restriction enzymes [15]. The Gateway enzymes utilize large sites (25 base pairs) [21] that rarely occur, enabling the same enzyme mix to be used to generate one construct or

hundreds of constructs in parallel (*e.g.* in 96-well plates) in one day [22,23], with inserts up to 10,000 base pairs successfully cloned. The Gateway-compatible Y1H assay is designed to take advantage of Gateway-based collections of cloned ORFs [24,25] and promoters [26] that are used to generate prey expression constructs and bait reporter constructs, respectively. In order to clone a sequence of interest, the Gateway recognition sites must be added to both ends, typically using PCR. Importantly, we have shown that these additions do not appear to interfere with the ability to detect PDIs [15], as for several bait-prey combinations, the recognition sites did not cause bait auto-activity, and the extra residues encoded by these sites did not affect prey function.

Gateway enzymes are more expensive than endonucleases, but this cost is mitigated if more than several constructs are required. As mentioned above, the Gateway system can be used to clone large fragments of DNA, however, cloning success is highest for DNA baits or prey ORFs that are 3,000 base pairs or smaller. Importantly, there may be a limitation on bait size that is linked to the distance over which a prey can activate reporter expression. Some researchers suggest this limit is 400 base pairs [27], while others have *in vivo* support for a PDI detected by Y1H that was 2,300 base pairs away [28]. While any such limit is likely dependent on the TF involved, we suggest breaking large regulatory regions into fragments of up to 1,000 base pairs for success of both cloning and detection of PDIs in Y1H.

3.5 SCREENING cDNA OR TF-ONLY LIBRARIES OF PREY CONSTRUCTS

Most Y1H screens use prey libraries in which cloned cDNA fragments are used to generate the hybrid proteins. Constructing a cDNA library from a whole organism or tissue of interest is a relatively common practice in molecular biology, and so a suitable library for a Y1H screen is often available. This ease of access to cDNA libraries, however, is offset by two major drawbacks for their use in Y1H screens. Firstly, due to limits in reverse transcriptase processivity, many cDNA clones are not full-length and do not encode the N-terminus of the protein, which may harbor the DNA binding domain. Secondly, because TFs account for only 5–10% of an organism's protein-coding genes [29,30] and they are generally expressed at low levels, the vast majority of cDNAs generated from a whole organism or tissue of interest do not encode TFs. This latter issue makes screening a library of cDNA clones for PDIs an inefficient process; a researcher has to screen many millions of library transformants to ensure that enough bait yeast actually receive a TF-encoding plasmid (Figure 3). "Normalized" cDNA libraries are sometimes available, in which the occurrence of each gene is more uniform than in non-normalized libraries. Screening a normalized library for PDIs will therefore require fewer transformants, but TF-encoding clones will at best make up 10% of a normalized cDNA library. Thus, a more suitable prey library for Y1H screens is one in which all the clones encode full-length TFs (Figure 3).

Generating individual full-length TF-encoding clones in order to construct a TF-only prey library is not trivial. Firstly, a complete list of which proteins are believed to be TFs is required for the organism of interest (*e.g.* [29]). Then the full-length TF ORFs must be individually cloned from a cDNA source, typically using PCR to amplify the ORFs for ligation or recombination into plasmids. Gateway-based collections of full-length ORFs are available for a few organisms (*e.g.* [24,25]), from which TF clones can be collected. Because a TF-only library contains hundreds of different clones, screening such a library requires only several thousand transformants. The major disadvantage of this approach is that a TF-only library will feature only the TFs collected by the researcher, and so will lack TFs yet to be cloned and TFs not in the initial TF list because they lack a recognizable DNA binding domain (known as "novels" [31], or "unconventional DNA binding proteins" [32]). PDIs involving such preys missing from a TF-only library can therefore only be detected using a cDNA library, and so the choice of screening either (or both) types of library depends on the goals of the researcher.

3.6 SCREENING LIBRARIES OR ARRAYS OF PREY CONSTRUCTS

When performing a Y1H screen using a library of preys, the complexity of the library is used to estimate how many transformants must be interrogated to ensure that every prey in the library is tested for a PDI. However, even with TF-only libraries, it is difficult to be sure that every TF clone was in fact transformed into yeast (Figure 3). An additional problem arises when selecting positive colonies to identify the interacting TF they harbor, because different PDIs induce different interaction phenotypes. For example, when screening for *HIS3* activity, colonies with strong interaction phenotypes will grow bigger and faster, making those colonies and their associated TFs more likely to be picked than colonies/TFs with weaker interaction phenotypes, resulting in missed PDIs. A final issue is that often the majority of positive colonies harbor clones for the same prey, which is poor return for the expense (commonly for PCR and sequencing) of determining the identity of those clones. A solution to all these problems is to screen an array of individual TF-encoding clones (Figure 3).

Typically, an array has a different TF clone at each coordinate (*e.g.* in 96-well plates) as well as a negative control plasmid that encodes only the AD. Each of these constructs is transformed individually in parallel, and activation of the reporters is then detected by comparing the yeast at each TF array position to those transformed with the negative control. Using this approach all TFs are individually interrogated, the interaction phenotype strength of one PDI does not affect detection of another because all TFs are interrogated separately, and there is no repeat sequencing of the same interacting prey as it is present once in the array. In fact, there is no need for sequencing at all as clone identity at every coordinate is already known. Screening an array detects significantly more PDIs than a library screen (Figure 4), suggesting that the issues listed above seriously reduce the ability of library-based screens to detect all possible PDIs. For some researchers, the benefit of detecting more PDIs will outweigh the challenges of generating an array. These challenges are similar to those for making a TF-only library, such as accessing individual TF clones and ensuring the completeness of the collection. Importantly, once new TF clones are made, or novel DNA binding proteins discovered, it is trivial to add these extra prey clones to the array.

3.7 PERFORMING Y1H BY TRANSFORMATION OR MATING

A Y1H assay requires that a prey construct is introduced into a bait yeast strain that harbors the reporter construct(s). This is commonly achieved by transforming the expression construct into a haploid bait yeast strain. However, this can also be achieved using the ability of haploid yeast strains to mate and generate diploid yeast. In this way, the haploid bait strain of one (usually **a**) mating type is mated with haploid “prey strains” of the opposing (usually **alpha**) mating type into which the prey constructs have been transformed, and the PDI is tested in the resulting diploid yeast. While mating could be used for library-based Y1H screens, it is more typically used for array-based screens.

Using mating for Y1H has the advantage that the yeast propagate the prey expression constructs, rather than the researcher having to extract these plasmids from bacteria for transformation into every bait strain. Mating detects somewhat fewer PDIs than transformation (Figure 4). However, PDIs detected using mating are more reproducible than those detected using transformation [33], meaning that an interaction seen by mating is more likely to be seen with another Y1H screening method (orange *versus* blue in Figure 4). Mating is also more easily adapted to robot-assisted screens (see below) because it involves simple transfers of yeast between media plates, whereas transformation requires multiple liquid-handling steps. Interestingly, we have observed instances of PDIs that are specific to

mating or transformation, suggesting that the different biology of haploid and diploid yeast has an effect on Y1H screens.

3.8 USING LOW-COPY OR HIGH-COPY PREY CONSTRUCTS

Plasmids used to express proteins in yeast typically either have a 2 μ origin of replication and are maintained as a high-copy episome with 50–100 copies per cell, or they feature an ARS/CEN origin and are low-copy with one or two copies per cell [34]. An increase in copy number results in an increase in protein expression.

A direct comparison of Y1H assays using transformation and mating revealed multiple PDIs that were consistently observed using transformation of the low-copy plasmid into haploid yeast but were undetectable with the same low-copy construct in diploid yeast generated by mating (Figure 4). Using the high-copy version of the Y1H prey expression vector to increase levels of the hybrid protein in diploid yeast made most of these missing PDIs detectable. It is unclear why a few of these PDIs remained undetectable using this approach, but one possibility is that the higher levels of prey protein adversely affected the yeast in some way, although no obvious phenotypic difference was observed. Similar reasons may also explain why several PDIs observed using low-copy constructs in diploid yeast were undetectable in diploids when using the high-copy construct. Transforming haploid yeast with high-copy plasmids leads to the most PDIs being detected (Figure 4), but this approach also detects the most interactions that cannot be reproduced by another Y1H screening method. It is important to note that no relationship between the reproducibility of an interaction and its *in vivo* relevance has been shown, largely because the methods used to determine biological consequences of an interaction have their own technical limitations [16].

3.9 USING SMART-POOLING OR ROBOTIC-ASSISTANCE TO INCREASE THROUGHPUT

One of the early Y1H protocols recognized that screening all the regulatory regions of a genome would require a significant increase in assay throughput [12]. Once the DNA bait yeast strain has been generated, screening a library of prey clones takes about six weeks, and screening an array either by transformation or mating takes about two weeks. The limiting factor when screening an array is the number of prey clones that can be interrogated per media plate. Using 96 preys per plate it is challenging for a researcher to screen more than five DNA baits per week. We have developed two methodologies that increase throughput by increasing this density: one transforms “smart-pools” of clones [33], while “enhanced Y1H” assays use robotic-assistance [35].

The smart-pool approach uses pools of a maximum of 25 prey clones, which provides a meaningful reduction in the number of required transformations over using individual clones while ensuring that every prey in the pool still gets the chance to demonstrate an interaction. The pools are also carefully designed with every TF clone in a unique combination of three pools so that the pattern of positive pools can be deconvoluted to reveal the TFs involved in PDIs without sequencing. The smart-pooling approach takes the same amount of time as a typical array screen, but uses a fifth of the media plates, thus increasing throughput to about 25 regulatory regions per week. Importantly, there is only a slight reduction in interactions detected using smart-pooling compared to screening an array of individual clones (Figure 4). The biggest drawback of this method is maintaining pool complexity, with the only way to be sure that every pool has the required clones being the time-consuming process of propagating all the prey clones individually and recreating the pools on a regular basis.

The enhanced Y1H (eY1H) assay was inspired by synthetic genetic array screens [36] that use robots to conduct 1,536 pairwise matings of haploid yeast strains on a single media

plate. Using a similar platform, eY1H assays mate a single bait strain with 1,536-colony arrays of prey yeast that feature up to 380 AD-TFs present four times (the remaining 16 colonies are negative controls that contain empty vector). Screening four colonies per TF interrogates each potential PDI four times, and by only considering interactions for which at least two of the four colonies are positive, the occurrence of both false negative and false positive interactions is reduced. eY1H assays are also different to previous Y1H matingbased protocols in that the TF preys are expressed from a high-copy vector, and different strains of yeast are used to host the bait and prey constructs. With this approach, 60 regulatory regions can be screened per week per robot, with slightly more interactions detected than with smart-pooling screens (Figure 4). Most of this throughput increase is afforded by the ability of the robot to handle 1,536 colonies where previous approaches used 96 colonies per media plate. However, another modification that increased throughput is testing the expression of both reporter proteins simultaneously instead of separately. This is done by adding both 3AT and X-gal to the same “readout” media plate, such that only yeast that activate enough *HIS3* will overcome the 3AT and grow, and only yeast that grow will be tested for beta-galactosidase expression and turn blue. Importantly, we have developed software that identifies PDIs from the eY1H readout plate images. Another robot-assisted Y1H assay has been reported that uses transformation to interrogate prey constructs in 1,536-colony format, and also uses software to call PDIs from assay images [37]. While this approach can also screen many baits per week, it uses only the *HIS3* reporter, and requires every bait be screened twice to identify interactions that are not reproducible. Performing Y1H assays with robots enables a laboratory to screen thousands of regulatory regions in a year, which is enough to address genome-scale questions of gene regulation. The major disadvantage of these approaches is the need for an expensive robot, and therefore this option is likely limited to researchers planning an extensive number of screens. It is important to note that these Y1H reagents can be used to conduct lower throughput mating screens if a robot is unavailable.

4 FUTURE DIRECTIONS

There are three aspects of the Y1H assay that we suggest targeting to further improve its performance: ease of bait strain generation, further increases in assay throughput, and the ability to detect heterodimeric interactions.

Before the development of eY1H, the bottleneck of a Y1H assay pipeline was performing the screens, but now it takes longer to create a bait strain than to screen it. Currently, generating a bait strain involves cloning two separate reporter constructs, integrating these at different locations in the yeast genome, and then confirming bait identity at both genomic locations in the final yeast strain. This process could be made easier by using one integration and one sequence confirmation. We are reluctant to use a single reporter construct for the reasons discussed above, so one option is to combine both reporters into a single plasmid that is integrated at one genomic location.

Despite the advances in assay throughput provided by eY1H, further increases in assay speed would further broaden the types of scientific questions that may be answered using this technique. To achieve this increase, we foresee modifications to the assay that utilize a combination of cell sorting and next generation sequencing technologies. Both of these have been used separately for detecting protein-protein interactions using yeast two-hybrid assays [38–39], supporting their feasibility for application to Y1H screens. We propose that with GFP and RFP as the two Y1H reporter proteins, a cell sorter can separate the population of bait yeast cells transformed with prey that activate both reporter constructs, and these positive prey clones then be identified using next generation sequencing. Different sequencing barcodes can be used for different baits so that a single sequencing run can

report PDIs for multiple baits. This approach will likely be well-suited to screening cDNA libraries of prey clones, as sequencing depth has the potential to overcome issues caused by low occurrence of TF-encoding clones, which would make this technique more widely accessible thanks to the availability of cDNA libraries.

The Y1H assay is currently designed to interrogate interactions involving one protein at a time, with the TF binding the bait DNA either as a monomer or a homodimer. However, many TFs are known to function as part of a heterodimer with another TF, with some TFs only capable of binding DNA when within a heterodimer. Current versions of Y1H assays are therefore blind to these biologically important interactions, and so adapting the Y1H assay to detect heterodimeric PDIs is important. The obvious approach is to simply transform two different prey constructs into the same bait strain, but the logistics of conducting such a heterodimeric PDI screen are daunting as it would require interrogating every pairwise combination of TFs in every bait strain. While improvements in assay throughput would likely make this approach less daunting, in the meantime we can use yeast two-hybrid assays to learn which TFs are capable of forming heterodimers and screen those specific pairs against many bait sequences.

5 SUMMARY

In this perspective we have highlighted several technical modifications to the Y1H assay that have been reported, and discussed their effect on assay performance. Perhaps a broader question that the average researcher would like answered is: Which of these multiple versions of Y1H assay is best for me? Firstly, we recommend that each bait strain feature two integrated reporter constructs; it does not matter whether these constructs were generated using restriction or recombination enzymes. How these baits are screened largely depends on three factors: how many baits are to be screened, whether individual TF clones are available for your organism of choice, and the desired balance between quantity and reproducibility of detected interactions. If the TF clones are available, we recommend generating an array of high-copy prey expression constructs. Screening this array by mating will yield fewer interactions than using transformation, but those detected by mating are more likely to be reproducible. If there are more than several hundred baits to be screened, then using a robot should be considered. If individual TF clones are not available, then cDNA library screens are the best/only option. However, if there are more than one hundred baits to be screened, then creating a TF clone resource should be seriously considered.

Acknowledgments

We thank members of the Walhout laboratory for discussions and critical reading of the manuscript. This work was supported by NIH grants GM082971 and DK06429 to A.J.M.W.

REFERENCES

1. Dey B, Thukral S, Krishnan S, Chakrobarty M, Gupta S, Manghani C, Rani V. DNA-protein interactions: methods for detection and analysis. *Mol. Cell. Biochem.* 2012; 365:279–299. [PubMed: 22399265]
2. Arda HE, Walhout AJ. Gene-centered regulatory networks. *Brief. Funct. Genomics.* 2010; 9:4–12. [PubMed: 20008400]
3. Walhout AJ. Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Res.* 2006; 16:1445–1454. [PubMed: 17053092]
4. Walhout AJ. Gene-centered regulatory network mapping. *Methods Cell. Biol.* 2011; 106:271–288. [PubMed: 22118281]
5. Schena M, Yamamoto KR. Mammalian glucocorticoid receptor derivatives enhance transcription in yeast. *Science.* 1988; 241:965–967. [PubMed: 3043665]

6. Metzger D, White JH, Chambon P. The human oestrogen receptor functions in yeast. *Nature*. 1988; 334:31–36. [PubMed: 3290686]
7. Brent R, Ptashne M. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell*. 1985; 43:729–736. [PubMed: 3907859]
8. Ma J, Ptashne M. Converting a eukaryotic transcriptional inhibitor into an activator. *Cell*. 1988; 55:443–446. [PubMed: 3180218]
9. Wilson TE, Fahrner TJ, Johnston M, Milbrandt J. Identification of the DNA binding site for NGFI-B by genetic selection in yeast. *Science*. 1991; 252:1296–1300. [PubMed: 1925541]
10. Inouye C, Remondelli P, Karin M, Elledge S. Isolation of a cDNA encoding a metal response element binding protein using a novel expression cloning procedure: the one hybrid system. *DNA Cell Biol*. 1994; 13:731–742. [PubMed: 7772254]
11. Li JJ, Herskowitz I. Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. *Science*. 1993; 262:1870–1874. [PubMed: 8266075]
12. Dowell SJ, Romanowski P, Diffley JF. Interaction of Dbf4, the Cdc7 protein kinase regulatory subunit, with yeast replication origins *in vivo*. *Science*. 1994; 265:1243–1246. [PubMed: 8066465]
13. Wang MM, Reed RR. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature*. 1993; 364:121–126. [PubMed: 8321284]
14. Chong JA, Tapia-Ramírez J, Kim S, Toledo-Aral JJ, Zheng Y, Boutros MC, Altschuler YM, Frohman MA, Kraner SD, Mandel G. REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell*. 1995; 80:949–957. [PubMed: 7697725]
15. Deplancke B, Dupuy D, Vidal M, Walhout AJ. A gateway-compatible yeast one-hybrid system. *Genome Res*. 2004; 14:2093–2101. [PubMed: 15489331]
16. Walhout AJ. What does biologically meaningful mean? A perspective on gene regulatory network validation. *Genome Biol*. 2011; 12:109. [PubMed: 21489330]
17. Sherman F. Yeast genetics. In: Meyers, RA., editor. *The Encyclopedia of Molecular Biology and Molecular Medicine*. Vol. 6. VCH; Germany: 1997. p. 302–325.
18. Postma-Haarsma AD, Rueb S, Scarpella E, den Besten W, Hoge JH, Meijer AH. Developmental regulation and downstream effects of the knox class homeobox genes *Oskn2* and *Oskn3* from rice. *Plant Mol. Biol*. 2002; 48:423–441. [PubMed: 11908517]
19. Deplancke B, Vermeirssen V, Arda HE, Martinez NJ, Walhout AJ. Gateway-compatible yeast one-hybrid screens. *CSH Protocols*. 2006 pii: pdb.prot4590. doi: 10.1101/pdb.prot4590.
20. Reece-Hoyes JS, Deplancke B, Barrasa MI, Hatzold J, Smit RB, Arda HE, Pope PA, Gaudet J, Conradt B, Walhout AJ. 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]
21. Hartley JL, Temple GF, Brasch MA. DNA cloning using in vitro site-specific recombination. *Genome Res*. 2000; 10:1788–1795. [PubMed: 11076863]
22. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science*. 2000; 287:116–122. [PubMed: 10615043]
23. Walhout AJ, Temple GF, Brasch MA, Hartley JL, Lorson MA, van den Heuvel S, Vidal M. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol*. 2000; 328:575–592. [PubMed: 11075367]
24. Reboul J, Vaglio P, Rual JF, Lamesch P, Martinez M, Armstrong CM, Li S, Jacotot L, Bertin N, Janky R, Moore T, Hudson JR Jr, Hartley JL, Brasch MA, Vandenhautte J, Boulton S, Endress GA, Jenna S, Chevet E, Papanastasiopoulos V, Tolias PP, Ptacek J, Snyder M, Huang R, Chance MR, Lee H, Doucette-Stamm L, Hill DE, Vidal M. *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet*. 2003; 34:35–41. [PubMed: 12679813]
25. Rual JF, Hirozane-Kishikawa T, Hao T, Bertin N, Li S, Dricot A, Li N, Rosenberg J, Lamesch P, Vidalain PO, Clingingsmith TR, Hartley JL, Esposito D, Cheo D, Moore T, Simmons B, Sequerra R, Bosak S, Doucette-Stamm L, Le Peuch C, Vandenhautte J, Cusick ME, Albala JS, Hill DE,

- Vidal M. Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res.* 2004; 14:2128–2135. [PubMed: 15489335]
26. Dupuy D, Li QR, Deplancke B, Boxem M, Hao T, Lamesch P, Sequerra R, Bosak S, Doucette-Stamm L, Hope IA, Hill DE, Walhout AJ, Vidal M. A first version of the *Caenorhabditis elegans* Promoterome. *Genome Res.* 2004; 14:2169–2175. [PubMed: 15489340]
 27. Ouwerkerk PB, Meijer AH. Yeast one-hybrid screens for detection of transcription factor DNA interactions. *Methods Mol. Biol.* 2011; 678:211–227. [PubMed: 20931383]
 28. Brady SM, Zhang L, Megraw M, Martinez NJ, Jiang E, Yi CS, Liu W, Zeng A, Taylor-Teeple M, Kim D, Ahnert S, Ohler U, Ware D, Walhout AJ, Benfey PN. A stele-enriched gene regulatory network in the Arabidopsis root. *Mol. Syst. Biol.* 2011; 7:459. [PubMed: 21245844]
 29. Reece-Hoyes JS, Deplancke B, Shingles J, Grove CA, Hope IA, Walhout AJ. A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol.* 2005; 6:R110. [PubMed: 16420670]
 30. Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA. DBD-- taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res.* 2008; 36:D88–D92. [PubMed: 18073188]
 31. Deplancke B, Mukhopadhyay A, Ao W, Elewa AM, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Reece-Hoyes JS, Hope IA, Tissenbaum HA, Mango SE, Walhout AJ. A gene-centered *C. elegans* protein-DNA interaction network. *Cell.* 2006; 125:1193–1205. [PubMed: 16777607]
 32. Hu S, Xie Z, Onishi A, Yu X, Jiang L, Lin J, Rho HS, Woodard C, Wang H, Jeong JS, Long S, He X, Wade H, Blackshaw S, Qian J, Zhu H. Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell.* 2009; 139:610–622. [PubMed: 19879846]
 33. Vermeirssen V, Deplancke B, Barrasa MI, Reece-Hoyes JS, Arda HE, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Brent MR, Walhout AJ. Matrix and Steiner-triple-system smart pooling assays for high-performance transcription regulatory network mapping. *Nat. Methods.* 2007; 4:659–664. [PubMed: 17589517]
 34. 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]
 35. Reece-Hoyes JS, Diallo A, Lajoie B, Kent A, Shrestha S, Kadreppa S, Pesyna C, Dekker J, Myers CL, Walhout AJ. Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nat. Methods.* 2011; 8:1059–1064. [PubMed: 22037705]
 36. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science.* 2001; 294:2364–2368. [PubMed: 11743205]
 37. Hens K, Feuz JD, Isakova A, Iagovitina A, Massouras A, Bryois J, Callaerts P, Celniker SE, Deplancke B. Automated protein-DNA interaction screening of Drosophila regulatory elements. *Nat. Methods.* 2011; 8:1065–1070. [PubMed: 22037703]
 38. Chen J, Zhou J, Bae W, Sanders CK, Nolan JP, Cai H. A yEGFP-based reporter system for high-throughput yeast two-hybrid assay by flow cytometry. *Cytometry A.* 2008; 73:312–320. [PubMed: 18307271]
 39. Yu H, Tardivo L, Tam S, Weiner E, Gebreab F, Fan C, Svrikapa N, Hirozane-Kishikawa T, Rietman E, Yang X, Sahalie J, Salehi-Ashtiani K, Hao T, Cusick ME, Hill DE, Roth FP, Braun P, Vidal M. Next-generation sequencing to generate interactome datasets. *Nat. Methods.* 2011; 8:478–480. [PubMed: 21516116]

Highlights for

- i.** Multiple versions of the Y1H assay have been developed
- ii.** These technical differences are discussed from the perspective of assay performance
- iii.** Potential future modifications to Y1H assays are described
- iv.** An optimal Y1H assay version is suggested for various experimental situations

Figure 1. Cartoon illustrating the conceptual difference between TF-centered and DNA-centered approaches for detecting physical interactions. TFs are indicated by hexagons, and DNA by double helices.

Figure 2.
Cartoon illustrating the basic components of a Y1H assay.

Figure 3.

Cartoon illustrating the difference between screening a cDNA library, a TF-only library, and an array of TF clones. For each type of screen, a 15 cm round media plate is depicted with each circle representing a yeast colony transformed with an expression construct. When screening a library, a random portion of the clones present in the library is transformed into the bait yeast. Black circles indicate non-TF clones present in a cDNA library. The TF-only library has all the TFs but no clone for the red TF happened to be transformed. When screening an array, all the TFs in the array are interrogated.

Figure 4.

Number of interactions detected for the promoter of the *C. elegans* gene *vha-15* using different versions of Y1H assays. Interactions detected by multiple versions are in blue, while interactions unique to that version are in orange. It should be noted that no relationship between interaction reproducibility and *in vivo* relevance has been shown. “low-copy” and “high-copy” refers to the copy number of the vectors used to express prey proteins. “transfn” – transformation.

Table 1

The advantages and disadvantages of various technical modifications available for Y1H assays.

MODIFICATION	ADVANTAGES	DISADVANTAGES
using one or multiple reporter constructs	less false positives are detected when using multiple reporter constructs	detecting expression from multiple reporters takes longer
using integrated or episomal reporter constructs	less false positives are detected when using integrated reporter constructs	integrating reporter constructs takes longer
interrogating simple or complex bait sequences	complex baits interrogate more sequence, simple baits reveal precise regions of TF binding	tandem repeats of simple baits should be checked for PDIs at artificial junctions
generating constructs using restriction or Gateway enzymes	generating many constructs is faster using Gateway enzymes	Gateway enzymes are more expensive
screening cDNA or TF-only libraries of prey constructs	TF-only libraries require less clones be interrogated	uncloned or uncharacterised TFs will not be present in a TF-only library
screening libraries or arrays of prey constructs	array screens detect more interactions and take less time	uncloned or uncharacterised TFs will not be present in the prey array
performing Y1H by transformation or mating	transformation detects more interactions, mating uses the yeast to propagate the prey constructs	interactions detected by transformation are less likely to be reproducible
using low-copy or high-copy prey constructs	high-copy constructs reveal more interactions	some interactions detected by high-copy construct transformation are not reproducible
using smart-pooling or robotic-assistance	smart-pooling has higher throughput than 96-prey/plate array screens, robotic-assistance provides higher throughput than smart-pooling	difficult to maintain complexity of smart-pools, robots are expensive, smart-pooling detects fewer interactions than array transformation